Functional analysis of Xist-mediated silencing

Greta Pintacuda

Trinity College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2016

Abstract

The lncRNA Xist is the master regulator of X chromosome inactivation in female mammals. Its mono-allelic expression triggers a cascade of well characterised events leading to the stable silencing of the X chromosome. However, how Xist coordinates these events is still unclear. We developed a system to dissect Xist function by making a library of deletions of Xist cDNA. The cDNAs were then stably transfected into a male ES cell line under an inducible promoter. The parent cell line was derived from an interspecific cross between different mouse strains, allowing subsequent allele-specific analysis based on a previously described set of strain-specific SNPs. This approach allowed full and partial loss of silencing activity to be discriminated by isolating nascent transcripts upon prolonged Xist expression, and using them to make libraries for high-throughput sequencing. It was therefore possible to map regions of Xist required for specific functions, notably gene silencing and recruitment of Polycomb repressor proteins. Using this approach, the B-repeats were identified as the segment of the RNA crucial for *de novo* recruitment of both PRC1 and PRC2 Polycomb complexes, whereas the A-repeats were confirmed to be essential for Xist-mediated early silencing of genes. To complement these analyses, an *in vitro* assembly system was set-up to study the direct interaction between key regions of Xist RNA and defined nuclear proteins at the biochemical/proteomic level, obtaining a comprehensive list of functional candidates. Interestingly, the nuclear matrix protein hnRNP-K was found among the factors significantly enriched over the segment of Xist previously identified as essential for Polycomb recruitment. We thus investigated the functional role of hnRNP-K in mediating the interaction between Xist and Polycomb with a number of *in vivo* and *in vitro* assays, and found that hnRNP-K is required to mediate contact between PRC1 and Xist through its KI domain.

Functional analysis of Xist-mediated silencing



Greta Pintacuda

Trinity College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2016

This thesis is dedicated to
Michi, Gianfo, and Felice.

Acknowledgements

What I have always loved about Science is that it is not something you do by yourself. As in a well-tuned orchestra, you learn to become part of a collective intelligence, or shared brain, to say it in H.G.Wells's words. That is why I think whatever I have been doing during my PhD represents my personal contribution to the common effort of our Group, to understand how Nature works. Literally everyone in the Brockdorff lab had a part in the way I shaped my research, thanks to their comments, advice, and simply to their being inspiring Scientists to look up to.

In particular, I was lucky enough to work in close contact with many amazing people, that I would like to acknowledge individually. First, Andrea, who mentored me with passion and generosity since day one. Not only he taught me how to think and design a good experiment, but he also got me involved in a number of exciting projects and discussed with me so many challenging ideas that I never felt our scientific collaboration ever ended, even after he left the lab.

Then Benoît, my lab hero, as Tianyi once defined him. Simply witnessing the enthusiasm and commitment he put in his research, was an important lesson for me to learn, but he also actively participated in directing and shaping my project, through a myriad of daily-based advices, hour-long discussions, and patient explanations. That is how, between a lipofection and a bacterial transformation, we became friends, and the long hours spent together in tissue culture suddenly became great fun.

The same way, I met another amazing friend in the lab, Tatyana, whom I consider as my rock for virtually everything, science-related and non. She always makes sure that we all work in an efficient and well-oiled environment, with abnegation and inspiring passion. Not to mention the fact that her never-ending knowledge helped significantly to direct my experiments.

Mafalda and Tianyi, my fellow PhD students, definitely made the difference during these years. I am not even talking about all our scientific discussions, the reagents they spared with me, the support inside and outside the lab: the have

been truly great friends, and I could not imagine anyone else I would rather spend most of my days with.

Guifeng, although the most recent addition to the lab, became soon an essential part of our small community. I feel extremely grateful to him scientifically, because through his amazing data analysis, he made my data talk and tell their story, and even more personally, as his curiosity mixed with love for his home-country are for me a constant source of learning.

Last but definitely not least, I thank Neil for guidance on my project and for his supervision, as well as for giving me freedom and trust on how to develop and pursue my scientific questions and interests.

Furthermore, I acknowledge support from Shabaz Mohammed and Svenja Hassler, for the mass spectrometry, and Alfredo Castello Palomares, for the set up of the proteomics-related experiments. I have also made use of reagents received from A. Cerase, T. Nesterova, and A. Wutz. I thank Ilan Davies and Rob Klose for being in my thesis committee, the Wellcome Trust for funding me, and Matthew Whitby for administering the Chromosome and Developmental Biology programme.

Infine, grazie a Mamma, Papo, Felice, Giovi, Piergiorgio, Tata, e Andrea, che mi hanno supportata (e sopportata) in questi anni non sempre facili, e a Radio Tre che ce l'ha messa tutta per stimolare il mio cervello talvolta un po' anestetizzato per eccesso di Scienza.

Abstract

The lncRNA Xist is the master regulator of X chromosome inactivation in female mammals. Its mono-allelic expression triggers a cascade of well characterised events leading to the stable silencing of the X chromosome. However, how Xist coordinates these events is still unclear. We developed a system to dissect Xist function by making a library of deletions of Xist cDNA. The cDNAs were then stably transfected into a male ES cell line under an inducible promoter. The parent cell line was derived from an interspecific cross between different mouse strains, allowing subsequent allele-specific analysis based on a previously described set of strain-specific SNPs. This approach allowed full and partial loss of silencing activity to be discriminated by isolating nascent transcripts upon prolonged Xist expression, and using them to make libraries for high-throughput sequencing. It was therefore possible to map regions of Xist required for specific functions, notably gene silencing and recruitment of Polycomb repressor proteins. Using this approach, the B-repeats were identified as the segment of the RNA crucial for *de novo* recruitment of both PRC1 and PRC2 Polycomb complexes, whereas the A-repeats were confirmed to be essential for Xist-mediated early silencing of genes. To complement these analyses, an *in vitro* assembly system was set-up to study the direct interaction between key regions of Xist RNA and defined nuclear proteins at the biochemical/proteomic level, obtaining a comprehensive list of functional candidates. Interestingly, the nuclear matrix protein hnRNP-K was found among the factors significantly enriched over the segment of Xist previously identified as essential for Polycomb recruitment. We thus investigated the functional role of hnRNP-K in mediating the interaction between Xist and Polycomb with a number of *in vivo* and *in vitro* assays, and found that hnRNP-K is required to mediate contact between PRC1 and Xist through its KI domain.

Contents

List of Figures **xi**

List of Tables **xiii**

List of Abbreviations **xiv**

1 Introduction **1**

1.1 Sex determination and dosage compensation 2

1.1.1 Sex determination in mammals 2

1.1.2 Dosage compensation in mammals 3

1.2 X chromosome inactivation (XCI) 3

1.2.1 The X inactivation center (XIC) 4

1.2.2 XCI in mouse development 5

1.2.3 Chromatin structure of the inactive X (Xi) chromosome . . . 6

1.2.4 Regulation of Xist expression 8

1.2.5 XCI is a conserved mechanism 9

1.3 X inactive specific transcript (Xist) 10

1.3.1 Long non-coding RNAs (lncRNAs) 10

1.3.2 Xist is a lncRNA 11

1.3.3 Xist is the main trigger of XCI 11

1.3.4 Xist is conserved among eutherian mammals 11

1.3.5 Xist RNA comprises functional tandem repeats 12

1.3.6 Xist localisation and *spreading* 14

1.4 Polycomb in XCI 16

1.4.1 Polycomb group repressor proteins (PcGs) 16

1.4.2 The PRC1 and PRC2 complexes 16

1.4.3 Polycomb complexes contribute to XCI 19

1.4.4 *De novo* Polycomb recruitment 20

1.4.5 Non-canonical PRC1 initiates Polycomb recruitment on the Xi 21

1.5 Xist-interacting factors 22

1.5.1 The Ncor/HDAC pathway: Spen and Rbm15 24

1.5.2 Wtap and the m^6A complex 25

1.5.3 LBR and the nuclear lamina . . . 26
1.5.4 The nuclear matrix proteins: hnRNP-U and hnRNP-K . . . 27
1.6 Aims of this thesis . . . 28

2 Materials and methods **29**
2.1 Plasmids and cloning . . . 30
2.1.1 Plasmids . . . 30
2.1.2 Conventional cloning . . . 31
2.1.3 Ligation Independent Cloning (LIC) . . . 32
2.1.4 Gibson assembly . . . 32
2.1.5 Primers for cloning . . . 32
2.1.6 Bacterial transformation . . . 32
2.2 Polymerase Chain Reaction (PCR) . . . 33
2.3 Cell culture . . . 34
2.3.1 Cell lines . . . 34
2.3.2 Culturing conditions . . . 34
2.3.3 Preparation of feeders . . . 35
2.3.4 Doxycycline treatment . . . 36
2.3.5 ES cells differentiation . . . 36
2.3.6 Generation of stable cell lines . . . 36
2.3.7 siRNA knock down . . . 36
2.3.8 Metaphase spreads . . . 37
2.3.9 Extracting genomic DNA from ES cells . . . 37
2.4 Southern blotting . . . 38
2.5 Antibodies . . . 39
2.6 Cell extracts . . . 39
2.6.1 Whole cell extracts . . . 39
2.6.2 Nuclear cell extracts . . . 39
2.7 Immunoblotting . . . 40
2.8 Coupling GFP-nanobodies to magnetic beads . . . 40
2.9 Co-immunoprecipitation . . . 41
2.10 Immunofluorescence assays . . . 41
2.10.1 Protein immunofluorescence . . . 41
2.10.2 RNA-FISH . . . 42
2.10.3 DNA-FISH on metaphase spreads . . . 42
2.10.4 Immuno-FISH . . . 43
2.11 Microscopy . . . 43
2.12 Assessing RNA levels . . . 43
2.12.1 RNA extraction . . . 43

2.12.2 DNase treatment and quantitative real time PCR 44
2.13 4sU-seq 44
2.13.1 RNA labelling with 4-thiouridine 44
2.13.2 4sU-RNA biotinylation 45
2.13.3 Biotinylated 4sU-RNA isolation 45
2.13.4 4sU-RNA quality check 46
2.14 4sU-seq data analysis 46
2.14.1 Polymorphic sites between the genomes of 129S1 and CAST mouse strains 46
2.14.2 SNPs across the genome 46
2.14.3 Genomic distribution of the paired-end reads 46
2.14.4 4sU RNA isolation and RNA-seq library construction 47
2.14.5 Definition and Calculation of repression score for Xist and miniXist RNA 47
2.14.6 4sU-seq alignment, allelic split and calling of gene expression 48
2.15 RNA pull-downs 48
2.15.1 *In vitro* transcription 48
2.15.2 RNA purification 49
2.15.3 RNA slot blot 49
2.15.4 RNA co-IPs 49
2.16 Mass spectrometry 50
2.17 Mass spectrometry analysis 51

3 Establishing a system to analyse the function of Xist transgenes 52
3.1 Introduction 52
3.2 Results 55
3.2.1 Generation of a Tet-On hybrid cell line 55
3.2.2 Generation of an *in vivo* Xist deletion library 56
3.2.3 Xist transgenes trigger early exclusion of RNA PolII and histone H4 hypoacetylation 59
3.2.4 The A-repeats of Xist are required to trigger early exclusion of RNA PolII and histone H4 hypoacetylation 63
3.2.5 Assessment of Xist-mediated gene silencing by 4sU-seq . . . 63
3.2.6 FL Xist is able to ectopically silence genes upon 24 hours induction in ES cells 65
3.2.7 The A-repeats are essential for gene silencing 69
3.2.8 Ev-Xist is able to mediate gene silencing 69
3.3 Discussion 71

4 Xist localisation and domain formation 78
4.1 Introduction . . . 78
4.2 Results . . . 79
4.2.1 An element in Ev-Xist required for *in cis* localisation . . . 79
4.3 Discussion . . . 81

5 Mechanisms of Xist-mediated Polycomb recruitment by Xist RNA 85
5.1 Introduction . . . 85
5.2 Results . . . 86
5.2.1 The XN region of Xist is required to recruit both PRC1 and PRC2 . . . 86
5.2.2 The B-repeats of Xist are required for PRC1 and PRC2 recruitment . . . 87
5.2.3 The B-repeats are essential for PRC1 and PRC2 recruitment by Ev-Xist . . . 91
5.2.4 Dissection of the B-repeat function in Polycomb recruitment 94
5.3 Discussion . . . 94

6 Polycomb and Xist-mediated transcriptional silencing 101
6.1 Introduction . . . 101
6.2 Results . . . 102
6.2.1 The XN region contributes to Xist-mediated gene silencing . 102
6.2.2 Locus-directed targeting of Ev-Xist in a single copy . . . 104
6.2.3 Ev-Xist shows efficient gene silencing on chromosome 11 . . 105
6.2.4 The B-repeats contribute significantly to Ev-Xist-mediated gene silencing . . . 108
6.3 Discussion . . . 109

7 The B-repeat interactome 111
7.1 Introduction . . . 111
7.2 Results . . . 112
7.2.1 *In vitro* transcription of Xist variants . . . 112
7.2.2 Xist pull downs and mass spectrometry assays . . . 114
7.2.3 Identification of A- and B-repeat binding factors . . . 115
7.3 Discussion . . . 119

8 hnRNP-K function in the context of XCI **121**
8.1 Introduction . . . 121
8.2 Results . . . 124
8.2.1 hnRNP-K facilitates Polycomb recruitment on the Xi . . . 124
8.2.2 hnRNP-K is abundant in the nucleus, but not enriched on the Xi . . . 124
8.2.3 hnRNP-K interacts with PRC1 . . . 127
8.2.4 The KI domain of hnRNP-K is required for interaction with PRC1 . . . 128
8.2.5 Forced tethering of hnRNP-K to the inactive chromosome partially rescues the ΔXN Xist Polycomb phenotype . . . 129
8.3 Discussion . . . 134

9 General discussion **137**
9.1 Overview . . . 137
9.2 The A-repeats of Xist are essential for gene silencing . . . 139
9.3 Ev-Xist recapitulates key functions of Xist RNA . . . 140
9.4 The B-repeats of Xist mediate PRC1/2 recruitment . . . 141
9.5 Polycomb contributes to Xist-mediated gene silencing . . . 142
9.6 The nuclear matrix protein hnRNP-K mediates PRC1 interaction with the B-repeats of Xist RNA . . . 143

References **145**

List of Figures

1.1 Imprinted vs random XCI in mouse. 5
1.2 XCI and chromatin. 7
1.3 Structure of the Xist gene. 12
1.4 Conservation of Xist repeats. 13
1.5 Polycomb-mediated histone modifications. 17
1.6 The PRC2 complex. 18
1.7 Canonical vs non-canonical PRC1. 19
1.8 PRC1 *de novo* recruitment on the Xi. 22
1.9 Xist *bona fide* interacting partners. 24

3.1 SNPs density between the *domesticus* and *castaneus* strains of *Mus musculus*. 55
3.2 Generation of a doxycycline inducible ESC line 56
3.3 Generation of Xist transgene cell lines. 57
3.4 Xist cDNA deletion library 58
3.5 qPCR analysis of Xist expression 58
3.6 *In vivo* spreading of Xist deletion variants. 60
3.7 FL Xist establishes PolII/H4ac holes. 61
3.8 Ev-Xist establishes PolII/H4ac holes. 62
3.9 ΔSX Xist fails to establish with PolII/H4ac holes. 64
3.10 RNA fractions recovered by 4sU-seq. 66
3.11 Integration site of FL Xist. 67
3.12 Xist Repression Score (RS). 68
3.13 FL Xist time-course expression. 68
3.14 FL Xist-mediated gene silencing. 70
3.15 FL Xist RS time-course. 71
3.16 ΔSX Xist-mediated gene silencing. 72
3.17 Ev-Xist-mediated gene silencing on chromosome 9. 73
3.18 Ev-Xist-mediated gene silencing on chromosome 11. 74

4.1 Xist localisation sequences 79
4.2 Ev-Xist cDNA deletion library 80

4.3 qPCR analysis of Ev-Xist expression 81
4.4 *In vivo* spreading of Ev-Xist deletion variants 82
4.5 Ev-Xist localisation sequences 83

5.1 PRC1 and PRC2 recruitment by FL Xist RNA. 88
5.2 PRC1 and PRC2 recruitment by ΔSX Xist RNA. 89
5.3 PRC1 and PRC2 recruitment by ΔXN Xist RNA. 90
5.4 Quantification of co-localising domains. 91
5.5 PRC1 and PRC2 recruitment by ΔB Xist RNA. 92
5.6 Recruitment of Polycomb subunits by FL vs ΔB Xist. 93
5.7 PRC1 and PRC2 recruitment by Ev- and EvXb-Xist RNA. 95
5.8 PRC1 and PRC2 recruitment by invB-Xist RNA. 96
5.9 The B-repeats of Xist. 97
5.10 Proximal vs Distal Ev-Xist and Polycomb. 98

6.1 ΔXN Xist-mediated gene silencing. 103
6.2 Targeting of Ev-Xist to the col1a1 locus. 104
6.3 PRC1 and PRC2 recruitment by single copy Ev-Xist and ΔB Ev-Xist RNA. 106
6.4 RS of single copy Ev-Xist and ΔB Ev-Xist. 107
6.5 Single copy Ev-Xist and ΔB Ev-Xist RS analysis. 108

7.1 Xist *in vitro* transcription. 112
7.2 Xist pull-downs. 114
7.3 Xist A^+B^+ specific binders. 115
7.4 Correlation between Xist interactomes. 116
7.5 B-repeat binding factors. 117

8.1 Structure of the protein hnRNP-K. 122
8.2 siRNA knock-down of hnRNP-K. 125
8.3 hnRNP-K nuclear localisation. 126
8.4 hnRNP-K and Polycomb co-IPs. 127
8.5 hnRNP-K deletion analysis. 128
8.6 hnRNP-K deletion variants and Polycomb co-IPs. 129
8.7 Artificial tethering of hnRNP-K to Xist. 130
8.8 Forced accumulation of hnRNP-K on the Xi. 131
8.9 Rescue of Polycomb upon forced hnRNP-K and Pcgf3 tethering to Xist domains. 132
8.10 Rescue of Polycomb upon forced hnRNP-K tethering to Xist domains. 133
8.11 hnRNP-K-mediated Polycomb recruitment via the B-repeats of Xist. 135

List of Tables

1.1 Conservation of Xist repeats. 13
1.2 Xist functional interacting partners. 24

2.1 List of cloning primers. 33
2.2 PCR conditions. 33
2.3 List of gRNAs. 34
2.4 List of cell lines. 35
2.5 List of antibodies. 39
2.6 List of real time primers. 44

5.1 Xist variants and Polycomb recruitment. 99

7.1 A-repeat specific binding factors. 118
7.2 B-repeat specific binding factors. 118

List of Abbreviations

°C . . . . . Celsius degree.

4sU . . . . . 4-thiouridine.

μg . . . . . microgram.

μL . . . . . micromiter.

μM . . . . . micromolar.

μm . . . . . micrometer.

aa . . . . . amino acid.

ac . . . . . acetyl.

ATP . . . . . adenosine triphosphate.

bp . . . . . base pairs.

BSA . . . . . bovine serum albumin.

Da . . . . . dalton.

DAPI . . . . . 4',6-diamidino-2-phenylindole.

DMSO . . . . . dimethyl sulfoxide.

DNA . . . . . deoxyribonucleic acid.

DNase . . . . . deoxyribonuclease.

DTT . . . . . dithiothreitol.

dNTP . . . . . deoxyribonucleoside triphosphate.

E . . . . . embryonic day.

EDTA . . . . . ethylenediamine tetraacetic acid.

eGFP . . . . . enhanced green fluorescent protein.

ESC . . . . . embryonic stem cell.

FISH . . . . . fluorescence in situ hybridisation.

FL . . . . . full length.

GFP . . . . . green fluorescent protein.

h . . . . . hour.

H2AK119ub . ubiquitylation of lysine 119 on histone H2A.

H3K9me1 . . . monomethylation on lysine 9 on histone H3.

H3K9me3 . . . trimethylation on lysine 9 on histone H3.

H3K4me3 . . . trimethylation on lysine 4 on histone H3.

H3K27me3 . . trimethylation on lysine 27 on histone H3.

H4ac . . . . . acetylation on histone H4.

HDAC . . . . . histone deacetylase.

kb . . . . . kilobase.

kDa . . . . . kilodalton.

IF . . . . . immunofluorescence.

L . . . . . litre.

LIF . . . . . leukaemia inhibitory factor.

lincRNA . . . long intergenic non-coding RNA.

lncRNA . . . . long non-coding RNA.

M . . . . . molar.

m6a . . . . . N^6-Methyladenosine.

me . . . . . methyl.

mg . . . . . milligram.

min . . . . . minute.

N- . . . . . amino.

ng . . . . . nanogram.

n . . . . . nanometer.

Ncor . . . . . nuclear receptor co-repressor 1.

ncRNA . . . . non-coding RNA.

ORF . . . . . open reading frame.

PBS . . . . . phosphate buffered saline.

PCR . . . . . polymerase chain reaction.

pH . . . . . $-log_{10}[H^+]$.

PRC1 . . . . . polycomb repressive complex 1.

PRC2 . . . . . polycomb repressive complex 2.

PolII . . . . . RNA polymerase II.

RNA . . . . . ribonucleic acid.

RSX . . . . . RNA on the silent X.

s . . . . . second.

UV . . . . . ultra violet.

WT . . . . . wild type

X_m . . . . . Maternal X chromosome.

X_p . . . . . Paternal X chromosome.

XCI . . . . . X chromosome inactivation.

XIC . . . . . X inactivation center.

Xist . . . . . X inactivation specific transcript.

On ne naît pas femme: on le devient.

One is not born, but rather becomes, a woman.

— Simone de Beauvoir, *Le Deuxième Sexe*

1 Introduction

Contents

1.1 Sex determination and dosage compensation 2
1.1.1 Sex determination in mammals 2
1.1.2 Dosage compensation in mammals 3
1.2 X chromosome inactivation (XCI) 3
1.2.1 The X inactivation center (XIC) 4
1.2.2 XCI in mouse development 5
1.2.3 Chromatin structure of the inactive X (Xi) chromosome 6
1.2.4 Regulation of Xist expression 8
1.2.5 XCI is a conserved mechanism 9
1.3 X inactive specific transcript (Xist) 10
1.3.1 Long non-coding RNAs (lncRNAs) 10
1.3.2 Xist is a lncRNA 11
1.3.3 Xist is the main trigger of XCI 11
1.3.4 Xist is conserved among eutherian mammals 11
1.3.5 Xist RNA comprises functional tandem repeats 12
1.3.6 Xist localisation and *spreading* 14
1.4 Polycomb in XCI 16
1.4.1 Polycomb group repressor proteins (PcGs) 16
1.4.2 The PRC1 and PRC2 complexes 16
1.4.3 Polycomb complexes contribute to XCI 19
1.4.4 *De novo* Polycomb recruitment 20
1.4.5 Non-canonical PRC1 initiates Polycomb recruitment on the Xi 21
1.5 Xist-interacting factors 22
1.5.1 The Ncor/HDAC pathway: Spen and Rbm15 24
1.5.2 Wtap and the m^6A complex 25
1.5.3 LBR and the nuclear lamina 26
1.5.4 The nuclear matrix proteins: hnRNP-U and hnRNP-K . 27

1.6 Aims of this thesis 28

1.1 Sex determination and dosage compensation

A remarkable commonality of many multicellular animals, is their reliance on sexual reproduction for the propagation of the species, whereby offspring are generated through fusion of specialised haploid cells, called gametes. In order to produce gametes, a precise sexual development must occur, involving differential specification of individuals, or at least of reproductive tissues. Overall, sex determination is a complex molecular mechanism, which requires a fine degree of regulation to account for all sex-specific differences in gene expression that might also impact non sex-related features.

1.1.1 Sex determination in mammals

Until the twentieth century, it was generally believed that sexes were environmentally determined. This theory dated back to the 4th century BC, when the hellenistic philosopher Aristotle first promulgated the soon-to-become popular concept that women are simply mutilated men whose development was arrested too early, due to excessive coldness of their mother's womb (Aristotle, *De generatione*). Through the work of prominent physicians such as Galen (200AD), Vesalius (1543), and many others, the idea that environment mostly contributed to gender definition was perpetuated for centuries. In the 1880s, naturalists Geddes and Thomson rigorously collected and evaluated all existing evidence regarding environmental factors that were reported to be relevant for sex determination.

However, the discovery of sex chromosomes by zoologist McClung in 1902, combined with a thorough application of Mendel's work, soon lead to correlate genders with inheritable genetic traits (Stevens, 1905; Wilson, 1905). We now know that in mammalian sex determination there is no such thing as a "default state", as sexes are strictly chromosomal: in most cases, the female is XX and the male is XY.

In other words, homogametic females, carry two copies of the X chromosome, while heterogametic males are monosomic for both X and Y chromosomes [Gilbert 2000].

1.1.2 Dosage compensation in mammals

Given the marked divergence between X and Y chromosomes, the karyotypic difference between sexes translates into a gender-related imbalance in gene dose. This is especially relevant in the case of the X chromosome, which unlike the Y, is not sex-determining, but hosts in mouse over a thousand of genes [Mank et al. 2013]. Because eukaryotic genomes have complex interconnected network structures, gene dose difference, which often correlates with gene transcription and translation rates [Moghadam et al. 2012], could affect a large proportion of the genome, potentially leading to a remarkable gender-related difference in gene expression. Also, monosomy of an entire chromosome in usually catastrophic for an organism, raising the question of the biological mechanisms that permit X monosomy in males.

Susumu Ohno was the first to propose a model that envisioned differential regulation of sex-chromosomes to account for dose heterogeneity [Ohno 1967]. He indeed noticed how some species, like *Drosophila*, resolved imbalance in gene dose, through a compensatory mechanism, based on over-expression of the monosome in males. Subsequently, it was found that other species achieve compensation the opposite way, by reducing gene expression in females. This latter case applies to eutherian mammals, in which a complex system has evolved to reduce X-linked transcription in female individuals early in development. This process was termed X chromosome inactivation, and results in almost complete transcriptional silencing of one X chromosome of females [Lyon 1961].

1.2 X chromosome inactivation (XCI)

In a brief Nature report published in 1961, Harwell-based British geneticist Mary Lyon, first formulated her famous hypothesis about the mechanism of X

inactivation [Lyon 1961]. The key elements of the "Lyon hypothesis" were three: first, that each cell of a female mammal had only one active X chromosome; second, that this inactivation occurred early in embryonic development; and third, that the inactivated X of a particular cell might be of either maternal or paternal origin, and that this was randomly determined. All descendants of a particular cell line would reflect the inactivation of the parent cell, so that females would be mosaic in respect of their X chromosome constitution. The original paper did not provide any experimental data, but linked together a number of previously published observations [reviewed in Harper 2011]. Notably, Lyon referred to the discovery of sex chromatin bodies (Barr bodies) in the cells of most female mammals, including humans, by neuroanatomist Murray Barr [Barr 1959]; and to the observation by Susumo Ohno that one X chromosome was condensed and heterotypic, contrasting with the other one which was un-condensed and similar to autosomes [Ohno et al. 1960].

The "Lyon hypothesis" was soon confirmed experimentally by Ernest Beutler, who was studying Glucose-6-phosphate dehydrogenase (G6PD) deficiency in heterozygous females. He noticed that females had two populations of erythrocytes: either deficient or normal for G6PD, depending on whether the inactivated X chromosome contained the normal or defective allele [reviewed in Beutler 2008].

1.2.1 The X inactivation center (XIC)

In the years following the formulation of the "Lyon hypothesis", pioneer cytological and genetic studies were aimed at understanding the molecular mechanism of X inactivation. A region of the X chromosome between the androgen receptor and the phosphoglycerate kinase 1 (Pgk-1) gene, was identified as the minimal part of the X chromosome necessary for the inactivation process to occur *in cis* [Rastan 1983]. Subsequently, this region, termed X inactivation center (XIC), was shown to encompass the Xist locus [Brown et al. 1991a], producing a functional RNA, Xist (X Inactive Specific Transcript) [Brockdorff et al. 1992, Brown et al. 1992]. Tightly

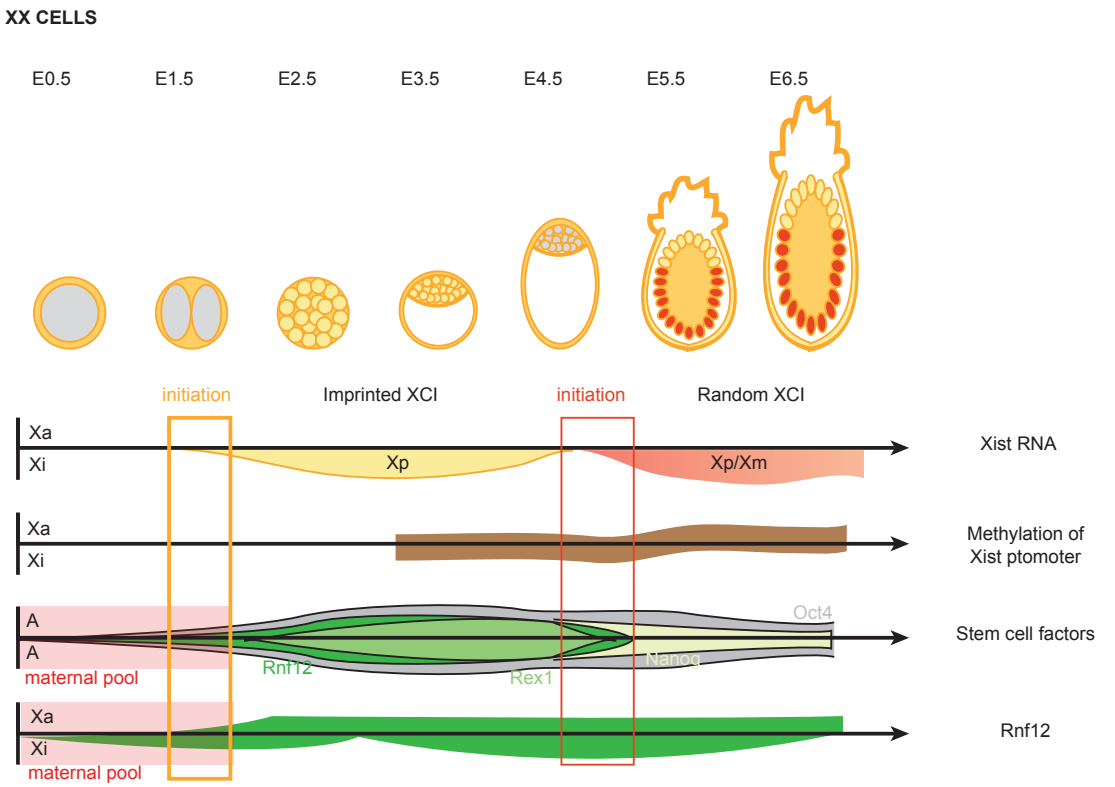


Figure 1.1: Imprinted vs random XCI. Imprinted XCI is initiated by Xist up-regulation (orange) from the paternal X at the 2-4-cell stage, when the Xist activator Rnf12 is present as a large maternal pool (rose). Imprinted XCI is reversed in the inner cell mass of the blastocyst (~E4.5), where Xist is down-regulated. Subsequently, Xist (red) is up-regulated from one randomly chosen X (X_m or X_p) in each cell, at least in part due to down-regulation of mostly autosomal (A), repressive stem cell factors (green). The methylation state of the Xist promoter (brown) is unknown, and was estimated from ES cell data. Modified from Schulz *et al* (2013).

regulated mono-allelic expression of Xist RNA was soon shown to trigger a cascade of molecular signals resulting in the stable, inheritable, and efficient silencing of one X chromosome of females [Penny et al. 1996, Lee et al. 1997].

1.2.2 XCI in mouse development

In mouse, XCI occurs in two phases during embryogenesis, an early wave of imprinted inactivation, and a later random wave. Imprinted XCI is initiated on the paternal X at the 2-4-cell stage [Okamoto et al. 2003], as shown in Figure 1.1. Subsequently, the paternal X (X_p) remains silent in the extra-embryonic tissue, while in the inner cell mass (ICM) of the blastocyst, which will give rise to the embryo proper, the X_p is reactivated, and random inactivation occurs during transition from

late blastocyst (E4.5) to the post-implantation epiblast (E5.5) [Okamoto et al. 2003, Mak et al. 2004]. This second wave of inactivation can be recapitulated *in vitro* in differentiating embryonic stem cell (ESC) systems. ESCs thus represent a powerful tool to dissect the molecular mechanisms underlying the random inactivation process. Interestingly, differentiation was shown to be delayed in the presence of two active X chromosomes, suggesting a tight and mutual coupling between X inactivation and normal development [reviewed in Chaligné et al. 2014].

1.2.3 Chromatin structure of the inactive X (Xi) chromosome

Several studies employing ESCs have focused on the early changes in chromatin structure that occur during the initiation of random XCI (Fig.1.2). Soon after Xist up-regulation, a first wave of loss of euchromatin-associated histone modifications, such as H3K9ac, H3K4me2, H3K4me3, is followed by de-acetylation of histone H4 and loss of factors linked to transcription, including RNA Polymerase II [Okamoto et al. 2003, Chaumeil et al. 2002, Heard et al. 2001, Keohane et al. 1996]. A few cell cycles later, new heterochromatin marks, associated with a silent state, for example H3K27me3, H3K9me2, H4K20me1, H2AK119ub1, are established [Escamilla-Del-Arenal et al. 2011]. The distribution of these modifications is not uniform across genomic features, and sometimes differ among mammals. For example, H3K9me3, is enriched on human but not on mouse Xi [Chadwick et al. 2004, Marks et al. 2009]. The enzymatic complexes responsible for establishing the hallmarks of the inactive state on the X, have been thoroughly studied in mouse, where the Polycomb group of proteins PRC1 and PRC2 were found to be responsible for catalysing deposition of H2AK119ub1 and H3K27me3, respectively [Napoles et al. 2004, Plath et al. 2003, Plath et al. 2004, Silva et al. 2003]. Moreover, Cdyl, a chromodomain-containing protein, was also identified as a potential H3K27me3 and H3K9me2 reader, able to promote association with other factors, such as MGA and the H3K9 methyltransferase (HMTase) G9a [Escamilla-Del-Arenal et al. 2013].

This complex network of factors recruited onto the Xi, determines the *in cis*

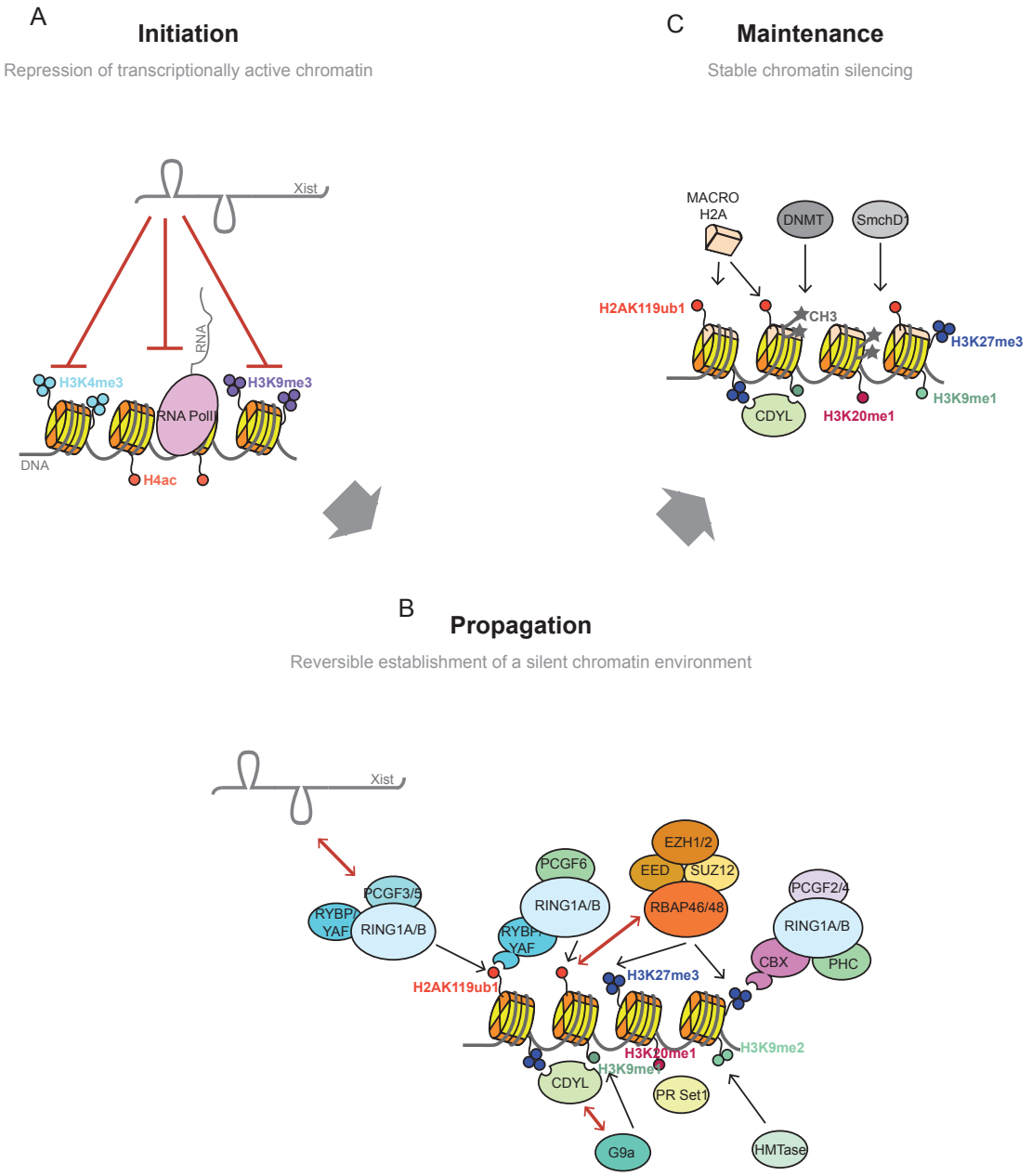


Figure 1.2: Changes in chromatin structure during XCI. (A) Xist is expressed *in cis* from the future Xi, causing loss of euchromatin-associated histone marks, H3K4me2/3, H3K9Ac and H4Ac, and initiation of gene silencing. (B) Several histone modifications also become enriched on the Xist RNA-coated chromosome at this time. These include H3K27me3, H3K9me2, H2AK119ub1 and H4K20me1. Cdyl may participate in propagating heterochromatin-associated marks such as H3K9me2 through G9a recruitment, and by binding to the H3K27me3/H3K9me2 enriched Xi. (C) At later stages of differentiation, in the maintenance phase, the PRC2 and PRC1 complexes are present at reduced levels on the Xi, and histone variant macroH2A also becomes associated with the Xi. Similarly, DNA methylation of promoters of X-linked genes, deposited by Dnmt3b, and recruitment of the chromosomal protein SmchD1, occur at later stages of the XCI process. Modified from Chaligné *et al* (2014).

spreading of silent chromatin features, often times due to the ability of chromatin modifiers to read their own deposited marks, generating a self-perpetuating positive feedback loop.

At later stages of differentiation, other factors are found to be associated with the Xi: amongst these are the histone variant macroH2A [Costanzi et al. 1998], which interferes with the transcriptional machinery, as well as with the SWI-SNF nucleosome-remodelling complex; Ash2L, part of the Trithorax group of proteins; and the nuclear matrix protein hnRNP-U/Saf-A [Costanzi et al. 2000, Pullirsch et al. 2010, Helbig et al. 2003]. Another late event is DNA methylation at CpG islands, mediated by the methyltransferase Dnmt3b, and the chromosomal protein SmchD1 [Gendrel et al. 2013, Blewitt et al. 2008, Sado et al. 2000, Nozawa et al. 2013]. Asynchronous replication timing was also described as one of the most conserved features of the Xi. Whether this contributes to the inactive state or it is a consequence, remains to be determined [Gomez et al. 2004, Chaumeil et al. 2002].

1.2.4 Regulation of Xist expression

The mono-allelic up-regulation of Xist from the future Xi is the trigger for the inactivation process, and therefore regulated through tightly-controlled molecular mechanisms, that act both *in cis* and *in trans*. As random XCI is a stochastic process, with no intrinsic bias for inactivation from either X chromosome [Monkhorst et al. 2008], the poised status of the future Xi, is due to local disruption of the equilibrium of X-encoded activators and repressors, otherwise perfectly balanced [reviewed in Gendrel et al. 2014, Galupa et al. 2015, Maduro et al. 2016]. Genes for some of these factors localise to the XIC, as in the case of the non-coding RNAs Tsix, Jpx, and Ftx. Tsix is transcribed antisense across Xist, and acts by repressing Xist expression on the active X, at the onset of X inactivation. [Lee et al. 1999]. Jpx (also known as Enox), and Ftx are both located in the XIC, and act as Xist activators. A mutation of Ftx in male cells results in reduced transcript levels of Xist and Jpx [Barakat et al. 2014], and a deletion of Jpx prevents XCI

and is female lethal [Jeon et al. 2011]. Regarding the ability of Jpx and Ftx to act also *in trans*, conflicting evidence has been reported [reviewed in Galupa et al. 2015]. Rnf12, also encoded from a locus in the XCI region, is an E3 ubiquitin ligase, and a potent *trans* activator of Xist [Makhlouf et al. 2014]. Further evidence has shown that Rnf12 activates Xist expression by targeting its repressor Rex1 for proteasome degradation [Bacher et al. 2006]. The repression mechanism through which Rex1 inhibits Xist expression is dual: on the one hand it activates Tsix, and at the same time competes for binding sites on the 5' end of Xist gene with YY1, which positively regulates Xist expression [Makhlouf et al. 2014]. The correct dose of Xist regulators, is developmentally regulated and directly linked to the level of expression of pluripotency factors, known to impact significantly on the XCI process [reviewed in Gendrel et al. 2014].

1.2.5 XCI is a conserved mechanism

It has been estimated that the X and Y chromosomes emerged approximately 181 million years ago, just prior to the divergence between placental (eutherian) and marsupial (metatherian) mammals, which occurred roughly 1 million years later [Cortez et al. 2014]. Subsequent divergent evolution led to substantial diversification of the X and Y chromosomes between therian mammals, although a common trend consisted in the progressive degradation and decay of the Y chromosome, concomitant with enlargement of the X [Ross et al. 2005]. This led to separate evolution of a gene dosage compensation mechanism, to account for sex-specific genetic imbalance. Both marsupial and placental mammals evolved an inactivation process to decrease the female-specific gene dose excess, based on similar mechanisms of epigenetic chromatin modifications [Pask et al. 2000]. Given this high degree of similarity, it was hypothesised that the trigger, Xist RNA, was also conserved. However, analysis of the putative XIC region in other vertebrates indicated that marsupials do not have an Xist homolog [Hore et al. 2007]. Absence of what was regarded as the master regulator of the eutherian inactivation process in marsupials, remained a conundrum until 2002, when the RSX (RNA on the silent X) gene

was discovered [Grant et al. 2012]. The RSX gene, like Xist, encodes for a non-coding RNA, which is not homolog to Xist, but does show some common features. Specifically, both Xist and RSX are enriched for tandem repeats that confer the ability to spread *in cis*, contact the nuclear matrix, and recruit the transcriptional silencing machinery [Shevchenko et al. 2013]. Thus, both marsupial and placental mammals, independently evolved a lncRNA-mediated inactivation process [reviewed in Whitworth et al. 2016]. The emerging picture from this example of convergent evolution, is that among the bulk of non-coding transcripts derived from ancient protein-coding genes, often truncated or infiltrated by repeat sequences, some might acquire functions that confer regulatory advantages [Migeon 2016].

1.3 X inactive specific transcript (Xist)

1.3.1 Long non-coding RNAs (lncRNAs)

The central dogma of molecular biology proposes a hierarchical model with a unidirectional flow of information from a unique coding molecule (DNA) to a plethora of functional molecules (proteins), through the activity of mediators (RNAs) [Watson et al. 1953]. We now know that this represents an over-simplified view of a much more complex information network, in which the boundaries between function and regulation of function are difficult to trace. In particular, the majority of transcripts were found not to be translated into proteins [Pertea 2012], partly because of general transcriptional noise, partly because RNA can be the repository of a biological function *per se*, separate and independent from any coding potential. As soon as new technologies permitted more thorough analysis of transcriptomes, non-coding RNAs (ncRNAs) were found to be widespread in the genome of species spanning from yeast to human [Rinn et al. 2012]. More refined analysis defined several classes of ncRNAs, based on their length, and genomic position. We thus distinguish between intergenic noncoding RNAs (lincRNAs) and intragenic ncRNAs, often arising from antisense transcription. Amongst the intergenic ncRNAs, mostly identified on the basis of specific histone modifications, those spanning more than 200bp were termed long non-coding RNAs (lncRNAs) [Hung et al. 2010]. lncRNAs are generally rapidly

evolving and not very abundant. A few examples, however, show a much pronounced conservation, suggesting a potentially relevant biological function [Rinn 2014].

1.3.2 Xist is a lncRNA

The X inactive specific transcript (Xist) is a prototypical example of a functional lncRNA, being fully-processed, longer than 200bp, and implicated in transcriptional regulation. Xist is encoded by a 20kb locus that comprises two very large and six small exons, giving rise to a 15-17kb transcript in mouse. Despite being capped, spliced, and polyadenylated, Xist RNA escapes the export machinery, and remains confined in the nucleus, where it spreads *in cis* onto the chromosome from which it is transcribed. Xist spreading leads to the cascade of epigenetic modifications that underlie the compact heterochromatic structure of the Xi, first described as a female-specific electron-dense nuclear structure, subsequently named Barr body [Barr 1959].

1.3.3 Xist is the main trigger of XCI

Classical genetic studies demonstrated that Xist RNA is the essential trigger for XCI, representing a *cis*-acting inactivation signal, necessary and sufficient for initiation and establishment of chromosome silencing. Knock-out studies, clearly showed that in the absence of Xist, XCI does not take place, whereas transgene studies demonstrated that the presence of an expressed form of Xist is sufficient to initiate the silencing cascade *in cis*, even at ectopic loci [Lee et al. 1996, Penny et al. 1996, Wutz et al. 2000]. However, the link between the observed role of Xist in mediating the inactivation process, and its primary/secondary structure, remains poorly understood, with the exception of few studies aimed at testing specific phenotypes linked to the loss of well characterised Xist-mediated functions.

1.3.4 Xist is conserved among eutherian mammals

Historically, even before identifying Xist as the master regulator of XCI, detailed molecular analysis had revealed complete synteny in the Xic/XIC region of the mouse and human X-chromosome, respectively [Brown et al. 1991a, Brockdorff et al. 1991].

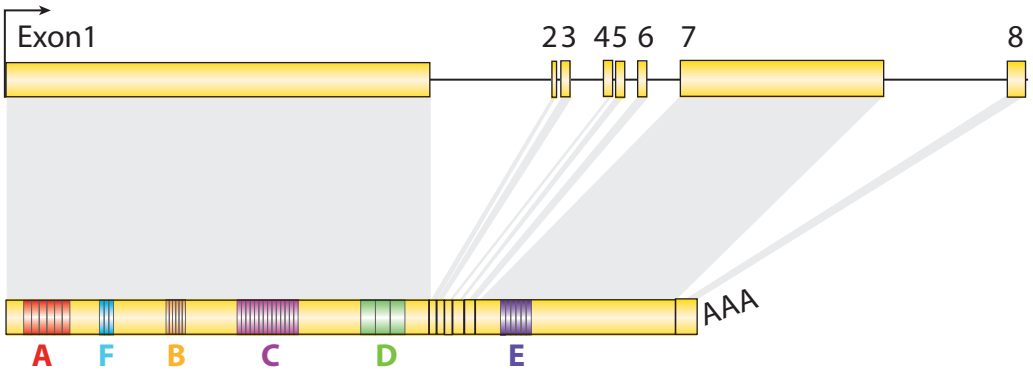


Figure 1.3: Structure of the Xist gene. Mouse Xist repeats and exon/intron structure. XIST/Xist share the same overall gene structure, including the position of intron/exon boundaries. Modified from Moindrot *et al* (2016).

When human XIST was observed to be expressed in a female-specific fashion [Brown et al. 1991b], the correspondent mouse gene was soon identified, using homology cloning [Brockdorff et al. 1992]. It was then determined that XIST/Xist share the same overall gene structure, including the position of intron/exon boundaries [Brockdorff et al. 1992, Brown et al. 1992]. Xist RNA was further shown not to comprise any ORF, and to be functional in the nucleus at the transcript level [Brockdorff et al. 1992, Brown et al. 1992].

1.3.5 Xist RNA comprises functional tandem repeats

A striking feature of Xist RNA is its highly repetitive nature (Fig.1.3), with large parts of its exons comprising tandem repeats of different length [Brockdorff et al. 1992]. Xist tandem repeats were annotated using consensus sequences. All of them can be related to conserved human XIST segments (Fig.1.4, Table 1.1). Since this first observation, several studies focused on their function in an evolutionary perspective. Hendrich *et al* (1993), found that the repeats are partially conserved among all eutherian mammals. An extensive study aimed at comparing Xist repeats between seven different mammalian species, found that the first and sixth (seventh in rodent) exons are always the largest and contain the repetitive elements that contribute for the most part to the size of the Xist gene. On the other hand, considerable variation in the copy number of the repeats was observed, with the

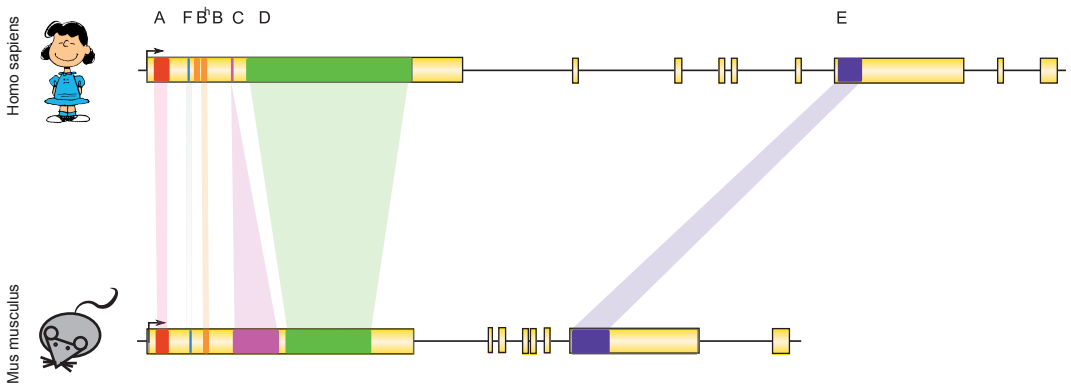


Figure 1.4: Conservation of Xist repeats. (A) Xist repeats are conserved between human and mouse, however considerable variation in the copy number of the repeats was observed, with the exception of the A-repeats, which are conserved both in terms of copy number and consensus sequence.

Table 1.1: A consensus motif was reconstructed for each repetitive segment and identified in several eutherian species. Adapted from Nesterova *et al* (2001).

Type of repeat	**A**	**F**	**Bh**	**B**	**C**	**D**	**E**
Monomer size (Consensus)	**Variable**	**16 bp motif**	**6-9 bp**	**6-7 bp**	**115 bp**	**290 bp**	**10-25 bp**
H. sapiens	429 (9)	42 (2)	92 (12)	104 (17)	105 (0.9)	6366 (26)	694
M. musculus	422 (8)	33 (2)		183 (32)	1611 (14)	3080 (10)	1351
M. arvalis	426 (8)	75 (4)		157 (26)	87 (0.8)	3122 (9)	1197
M. rossiaemeridionalis	431 (8)	85 (5)		159 (26)	88 (0.8)	3102 (9)	1014
M. kirgisorum	410 (8)	84 (5)		155 (26)	88 (0.8)	3078 (9)	1596
M. transcapicus	429 (8)	85 (5)		167 (28)	88 (0.8)	3166 (9)	1461

exception of the A-repeats, which are conserved both in terms of copy number and consensus sequence. In general, the authors of the study comment that for the repeats B-F, being the reduction of one repeat class concurrent with the expansion of another, a functional compensatory mechanism might be the main evolutionary drive [Yen et al. 2007].

Importantly, RSX, the marsupial Xist ortholog, is also highly enriched in tandem repeats. These were predicted to give rise to secondary structures similar to those modelled for some Xist RNA elements [Grant et al. 2012]. This observation reinforced the idea that the tandem repeats of Xist are the most likely repository of its general function. On the other hand, exception to the otherwise poor conservation of non-repetitive sequences of Xist is represented by exon IV. The primary sequence

of this region is highly conserved, compared to most of the Xist locus, revealing a similarity of 77% between mouse and human, where the overall conservation for all exons is 66% [Chureau et al. 2002]. Computational analysis also predicted a stable RNA stem-loop structure for this region, suggesting it might be functionally relevant [Zuker et al. 1999]. However, classical genetic studies reported only a minor phenotype upon deletion of exon IV [Caparros et al. 2002], suggesting that conservation might not necessarily be a good predictor of function. Alternatively, functional redundancy of the Xist molecule might mask more pronounced effects of this, and potentially other, deletions.

1.3.6 Xist localisation and *spreading*

Xist RNA can be easily detected by *in situ* techniques (RNA-FISH) [Duthie et al. 1999], or live-cell imaging [Ng et al. 2011]. The RNA exhibits a well defined localisation, confined to the Xi territory, where it forms a dense cloud around the chromosome. This specific localisation pattern indicates the presence of a molecular mechanism that ensures correct propagation and binding of Xist to the chromosome from which it is expressed [Clemson et al. 1996, Ng et al. 2011]. This process is currently poorly understood.

Cytogenetic analysis of Xist localisation on metaphase chromosomes demonstrated a non-linear distribution, with Xist being highly concentrated over gene rich domains [Duthie et al. 1999]. Support for this observation came from two recent studies, which characterised correlations between Xist spreading and chromatin features at different stages of the inactivation process [Simon et al. 2013, Engreitz et al. 2013]. In one study, maps of Xist binding were generated using allele-specific capture hybridisation analysis of RNA targets (CHART-seq), and showed how Xist targets gene-rich regions before eventually spreading to gene-poor domains, mirroring PRC2 and H3K27me3 distribution, as determined by chromatin immunoprecipitation (ChIP) analyses [Simon et al. 2013]. The other study, made use of RNA antisense purification (RAP-seq), and suggested a model whereby Xist

exploits the pre-existing 3D architecture of the Xi, to contact, possibly through its interaction with nuclear matrix proteins, a restricted number of distal regions (entry points) that do not correspond to any obvious sequence signature, but are localised close to the Xist gene in the 3D space [Engreitz et al. 2013]. According to the authors, only after this first wave of spreading, Xist starts to show a broad preference for gene-rich regions.

The zinc finger protein YY1 (Ying yang 1), a pleiotropic DNA-binding factor involved in transcription, chromosome conformation, and imprinting, was proposed to play a role in mediating proper Xist localisation, by bridging Xist RNA to Xist DNA. In detail, knock-down of YY1, as well as LNA interference, targeting putative YY1 binding sites on Xist RNA (located on the C-repeats), resulted in a visible defect in Xist clouds [Sarma et al. 2010], although deletion of the C-repeats was independently shown to be insufficient to dislocate Xist [Wutz et al. 2002].

Another insight into how Xist is constrained *in cis* comes from evidence that the nuclear matrix, defined as the biochemical fraction that is refractory to extraction with detergent and high salt, in combination with nucleases, facilitates Xist spreading [reviewed in Hasegawa et al. 2011]. The nuclear matrix contains a number of proteins involved in essential processes including DNA replication and RNA transcription, and has been proposed to regulate these events by providing a structural higher order organisation of the nucleus. Among the most abundant nuclear matrix components is the protein hnRNP-U (also known as Saf-A), which was shown to directly interact with Xist RNA [Hasegawa et al. 2011, McHugh et al. 2015, Chu et al. 2015]. Indeed, hnRNP-U is an obligate factor for Xist localisation to the Xi, as knock-down experiments caused disruption of Xist clouds and dispersion of Xist RNA in the nucleus [Hasegawa et al. 2010]. A recently published eCLIP map of hnRNP-U suggested that its binding sites are distributed along the whole length of Xist RNA [Van Nostrand et al. 2016], but its mode of engaging with chromatin to tether Xist to the Xi remains to be determined.

1.4 Polycomb in XCI

1.4.1 Polycomb group repressor proteins (PcGs)

Polycomb group (PcG) repressor proteins play a central role in epigenetic regulation of gene expression through differentiation and development, establishing a repressive chromatin environment at target gene promoters through their catalytic activity. They were first discovered in *Drosophila*, where they have a specific function in maintaining the normal segmental patterns of Hox gene expression through successive cell generations. In this context, the activity of PcG proteins is counterbalanced by Trithorax group (TrxG) factors, identified based on their capacity to maintain the expression of Hox gene loci [Ringrose et al. 2004]. Subsequent studies revealed that both PcG and TrxG proteins are highly conserved in multicellular organisms, where they are essential for normal patterns of gene expression throughout development and differentiation.

Biochemical and genetic studies have shown that PcG proteins generally associate with one of two multi-subunit chromatin modifying complexes, called Polycomb repressive complex (PRC) 1 and 2 [reviewed in Simon et al. 2013]. These complexes have a well characterised catalytic activity towards histone tails, with PRC1 mono-ubiquitylating histone H2A (H2AK119ub1) and PRC2 methylating histone H3 lysine 27 (H3K27me3), as shown in Figure 1.5.

1.4.2 The PRC1 and PRC2 complexes

PRC2 is the smaller of the two Polycomb complexes and has a well-defined core composition, which comprises Ezh2 (or its homolog Ezh1), Eed, Suz12, and RBAP46/48 (Fig.1.6). Ezh2 is a SET domain-containing protein that has HMT-specific activity towards histone H3 lysine 27 (H3K27) [Margueron et al. 2011]. Eed and Suz12 are essential for PRC2 complex integrity and for Ezh2/1 catalytic activity, while RBAP46/48 is most likely involved in nucleosome recognition, but is dispensable for PRC2 catalytic activity [Cao et al. 2002, Cao et al. 2004, Pasini

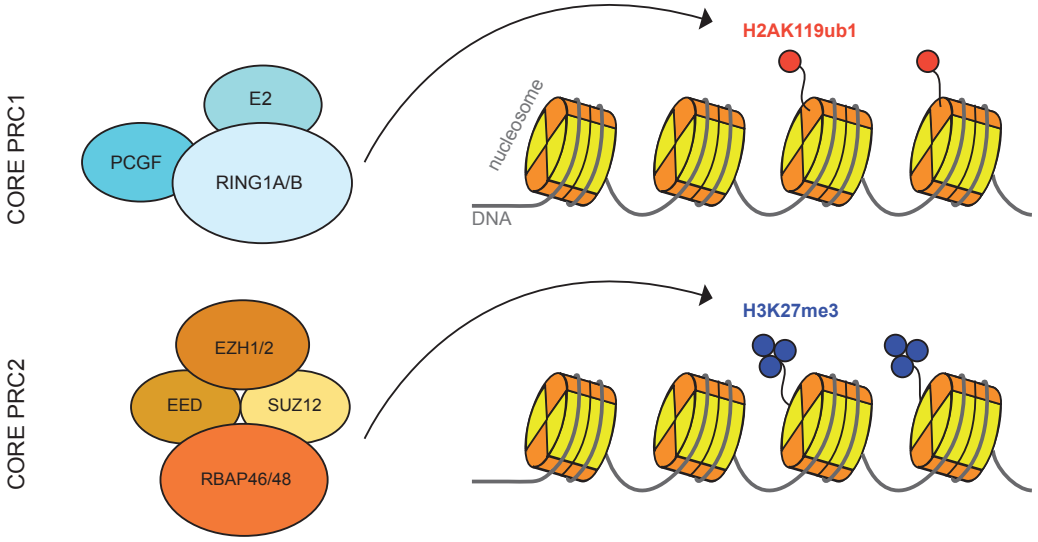


Figure 1.5: Polycomb-mediated histone modifications. Polycomb complexes have a well characterised catalytic activity aimed at modifying histone tails, with PRC1 mono-ubiquitylating histone H2A (H2AK119ub1) and PRC2 methylating histone H3 lysine 27 (H3K27me3). Adapted from Blackledge *et al* (2015).

et al. 2004].

Along with its core components, PRC2 also comprises several other accessory components, that have been reported to regulate its molecular and biological functions. Jarid2 has been demonstrated to globally regulate PRC2 binding to chromatin in ESCs [Pasini et al. 2010, Peng et al. 2009]. Recently, it was also shown that Jarid2 is substrate of PRC2, and methylated Jarid2 is a binding partner for Eed, acting in a positive feedback loop to mediate further stimulation of PRC2 activity at target sites [Sanulli et al. 2015]. Another component of PRC2 is the zinc finger protein Aebp2, which, similarly to Jarid2, can function to potentiate PRC2 catalytic activity towards unmodified nucleosome substrates *in vitro*. However, the precise role of Aebp2 in modulating PRC2 enzymatic activity *in vivo* is still unclear, and indeed it has been recently shown to antagonise Polycomb function [Son et al. 2013, Grijzenhout et al. 2016]. Other non-core components of PRC2, shown to regulate its catalytic function and recruitment, are the Polycomb-like proteins (PCL1-3), also known respectively as PHF1, MTF2, and PHF19 [Aranda et al. 2015].

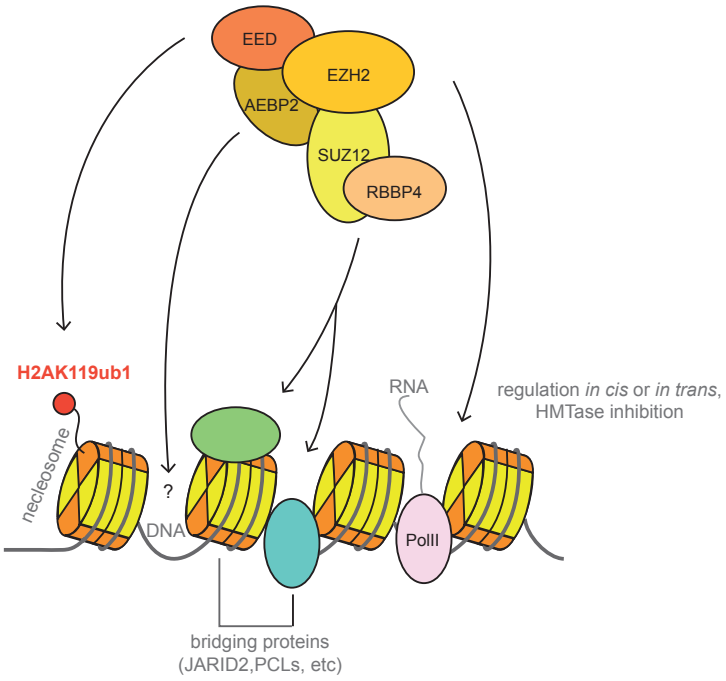


Figure 1.6: The PRC2 complex. PRC2 has a well-defined core composition, which comprises Ezh2, Eed, Suz12, and RBAP46/48, plus other accessory factors, such as Jarid2, Aebp2, and PCLs, that interact with chromatin in various ways. Modified from Davidovich *et al* (2015).

PRC1 is more dynamic in composition, particularly in mammalian cells. Ring1A and Ring1B are the central PRC1 components, and act as E3 ubiquitin ligases that mono-ubiquitylate lysine 119 of histone H2A (H2AK119ub1) [Wang et al. 2004]. In vertebrates, PRC1 can be assembled in two separate types of complex, referred to as canonical and non-canonical PRC1 (Fig.1.7). Canonical PRC1 comprises the catalytic subunits associated with a homolog of the *Drosophila* proteins Polycomb (CBX2/4/6/7/8) and Polyhomeotic (PH1/2/3), whereas the catalytic subunits of non-canonical PRC1 interact with one of the two closely related proteins Rybp or YAF2. Canonical PRC1 complexes mediate crosstalk with PRC2 complexes via interaction of a chromodomain in the CBX proteins with PRC2-mediated H3K27me3 [Fischle et al. 2003, Min et al. 2003], while recent evidence confirmed that non-canonical PRC1 can recruit PRC2 on chromatin through the H2AK119ub1 mark [Cooper et al. 2014, Blackledge et al. 2014]. This is consistent with findings demonstrating that PRC2 binds H2AK119ub1-modified nucleosomes, enhancing PRC2-mediated H3K27me3 deposition through Aebp2 and Jarid2 *in vitro* [Kalb

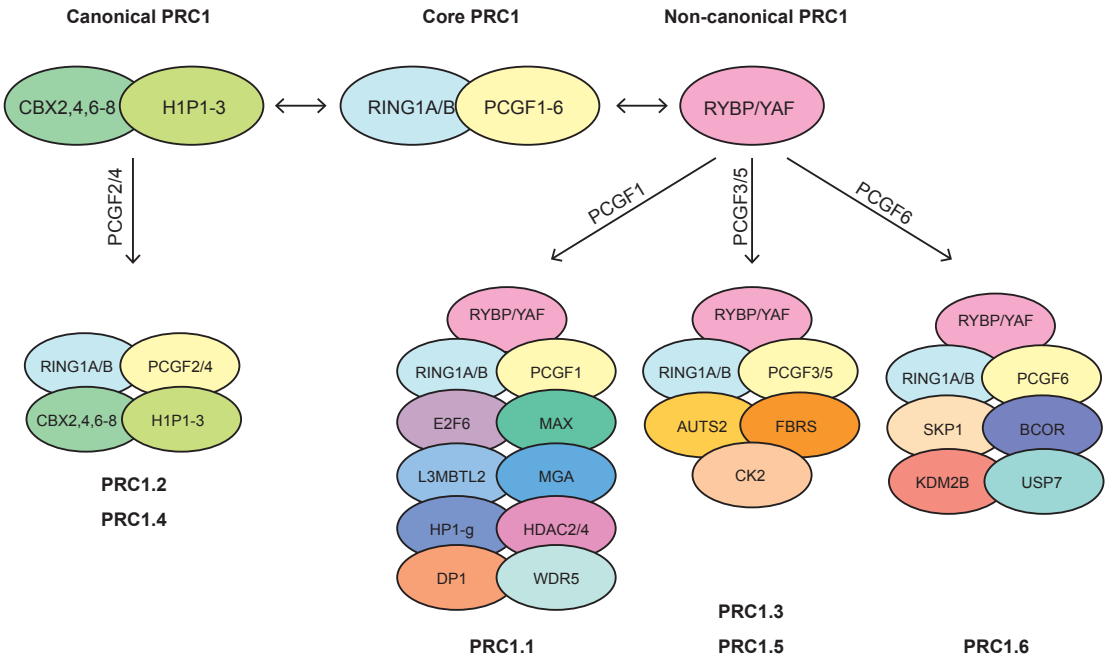


Figure 1.7: Canonical vs non-canonical PRC1. PRC1 can be assembled in two separate types of complex, termed canonical (PRC1.2, PRC1.4) and non-canonical (PRC1.1, PRC1.3, PRC1.5,, PRC1.6). Adapted from Gao *et al* (2012).

et al. 2014]. Although the biochemical composition of canonical- and non-canonical PRC1 have been clarified, their diverse biological and molecular activities remain to be fully investigated [Piunti et al. 2016].

1.4.3 Polycomb complexes contribute to XCI

The first evidence that Polycomb complexes contribute to XCI came from the observation that homozygous mutants for the gene *eed*, encoding a core subunit of PRC2, show a female-specific phenotype in trophoblast giant cells, an extra-embryonic lineage in early embryogenesis [Wang et al. 2001]. The protein Eed was subsequently shown to be highly enriched within the interphase Xi domain, and to localise on the X chromosome during metaphase, with a pattern closely resembling that of Xist RNA [Mak et al. 2002]. When Eed was defined as the core component of PRC2, together with Ezh2, this latter factor was also found to follow the same pattern of localisation on the Xi [Plath et al. 2003, Silva et al. 2003]. PRC1, the other major PcG complex, also accumulates over the Xi [Napoles et al. 2004].

1.4.4 *De novo* Polycomb recruitment

Genetic analyses have demonstrated that PcG targets are often co-regulated by PRC1 and PRC2. In fact PRC1/2 tend to occupy the same loci both in *Drosophila* and mouse genomes [Boyer et al. 2006, Ku et al. 2008, Schwartz et al. 2006]. This overlap has been explained as a consequence of the ability of canonical PRC1 to recognise the histone modification H3K27me3 deposited by PRC2. In detail, the chromodomain in the PRC1 protein PC (mammalian homologs CBX2/4/6/7/8) was shown to interact with PRC2-dependent H3K27me3, both biochemically [Cao et al. 2002, Fischle et al. 2003], and genetically [Boyer et al. 2006, Cao et al. 2002]. This model, envisioning a hierarchical mode of PRC2/PRC1 recruitment to target genes, was further corroborated in several studies demonstrating a direct link between H3K27me3 and PRC1 recruitment [Agger et al. 2007, Lee et al. 2007, Mujtaba et al. 2008]. The close dependency between Xist expression and Polycomb localisation on the Xi, seemed to suggest a simple recruitment model, whereby Xist would directly interact with PRC2, providing the physical trigger for its *de novo* accumulation over the Xi, with PRC1 recruitment occurring subsequently through H3K27me3. In agreement with this model, many studies found linear correlations between Xist and PRC2 localisation, and between Xist and PRC2-mediated H3K27me3 [Engreitz et al. 2013, Simon et al. 2013]. Moreover, a key set of *in vitro* studies, suggested that the PRC2 core component Ezh2 could provide the link between Polycomb and Xist, by binding to the A-repeat region.

Contrasting with the above findings, evidence emerged of a different mode of PRC1 *de novo* recruitment, potentially independent of PRC2. In fact, in PRC2-depleted mouse ESCs, global levels of H2AK119ub1 were similar to those of wild-type cells [Schoeftner et al. 2006], and PRC1 proteins have been detected at selected target loci both in ESCs [Leeb et al. 2010], and differentiated cells [Pasini et al. 2007]. These observations were confirmed in the context of X inactivation, where a PRC2-independent pathway for non-canonical PRC1 was described in mouse embryonic stem cells (mESCs) [Tavares et al. 2012]. Furthermore, showing that

PRC2 can be recruited through PRC1-mediated activity at target loci, raises the possibility that PRC1 recruitment precedes PRC2 recruiting [Blackledge et al. 2014, Cooper et al. 2014]. Further evidence arguing against the classical model for PRC2 recruitment by Xist RNA, came from the fact that the A-repeats are not essential for PRC2 recruitment [da Rocha et al. 2014].

Given the conflicting body of evidence regarding the simple "direct recruitment" model, the hypothesis that Polycomb accumulation might depend on a secondary response to Xist enrichment became more widespread. Interestingly, a recent study observed that the bulk of Xist RNA and all PRC2 components are clearly spatially separated as determined by 3D-SIM super-resolution microscopy [Cerase et al. 2014]. In agreement with this result, a previous study found that Xist mainly localises to the inter-chromatin/perichromatin regions (IC/PR), non DAPI-dense areas that display poor overlap with H3K27me3 domains [Smeets et al. 2014]. The idea that Xist does not contact Polycomb directly was corroborated by the lack of any Polycomb component in the recently published Xist proteomic studies, providing extensive Xist interactomes [McHugh et al. 2015, Chu et al. 2015]. McHugh *et al* (2015), interpret their result by ascribing PRC2 recruitment to an indirect effect of the activity of histone deacetylation. Differently, Chu *et al* (2015), suggested the presence of an intermediate factor mediating the contact between Xist and either PRC1 or PRC2.

1.4.5 Non-canonical PRC1 initiates Polycomb recruitment on the Xi

Recent unpublished findings in the lab, have revealed that a non-canonical PRC1 complex mediates Polycomb recruitment by Xist RNA. Recruitment of Pcgf3/5-PRC1 results in chromosome wide deposition of H2AK119ub1 at sites where Xist RNA accumulates, as shown in Figure 1.8. Following Pcgf3/5-PRC1 enrichment on the Xi, and early deposition of H2AK119ub1, other non-canonical PRC1 complexes are also recruited via interaction with the zinc finger domain of the RYBP/YAF2 subunit, thus reinforcing and amplifying levels of H2AK119ub1.

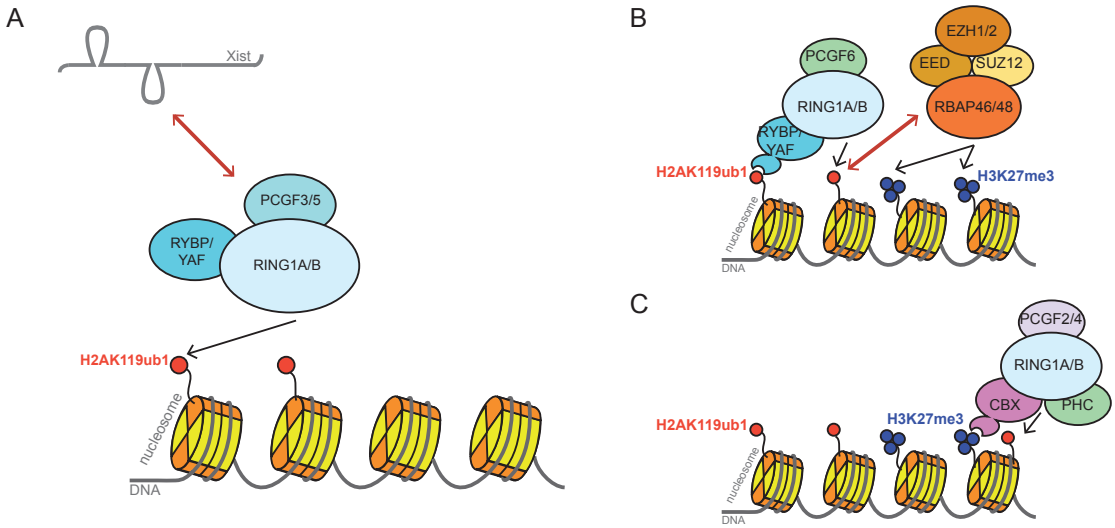


Figure 1.8: PRC1 *de novo* recruitment on the Xi. (A) Xist RNA is able to recruit Pcgf3/5-PRC1, resulting in chromosome wide deposition of H2AK119ub1. (B) Other non-canonical PRC1 complexes are also recruited via interaction with the zinc finger domain of the RYBP/YAF2 subunit, thus reinforcing and amplifying levels of H2AK119ub1. Then, H2AK119ub1 leads to recruitment of PRC2 complexes, and to deposition of H3K27me3 across the chromosome, which is in turn read by canonical PRC1 complexes through their CBX domains (C).

As previously shown, H2AK119ub1 is also able to recruit PRC2 complexes, leading to deposition of H3K27me3 across the chromosome [Blackledge et al. 2014, Cooper et al. 2014]. According to this model, canonical PRC1 is finally recruited through interaction of the CBX chromodomain with H3K27me3. How the interaction of Pcgf3/5-PRC1 with Xist RNA takes place at the molecular level, however, remains to be determined. Pcgf3/5-PRC1 could bind to Xist RNA directly, although no RNA-interaction domains are annotated in Pcgf3/5-PRC1, or via an RNA binding protein (RBP) co-factor.

1.5 Xist-interacting factors

Recently, systematic approaches were employed to profile the Xist interactome. Three proteomic screens used Xist antisense oligonucleotides to capture Xist RNA in complex with its interacting protein complexes, from soluble cell extracts [Chu et al. 2015, McHugh et al. 2015, Minajigi et al. 2015]. In order to stabilise the interactions between the nucleic acid and its binding partners that are *per se* weak,

all the aforementioned studies made use of cross-linking agents. In detail, Chu *et al* (2015), used formaldehyde crosslinking, which covalently links both RNA to proteins, and proteins to proteins, resulting in isolation of factors that interact with Xist either directly or through other proteins. In contrast, McHugh *et al* (2015), and Minajigi *et al* (2015), used UV-crosslinking, which excludes secondary interactions between proteins, resulting exclusively in the isolation of factors directly contacting nucleic acids. Another difference between these studies is the experimental system employed. While Chu *et al* (2015), and McHugh *et al* (2015), used male ES cells with inducible expression of an Xist transgene inserted in an autosomal ectopic locus, Minajigi *et al* (2015), carried out their study in female differentiated cells, expressing Xist RNA at endogenous levels. This difference is likely significantly impacting the nature of the profiled Xist interactome, as different factors might play different roles in subsequent phases of the silencing cascade. Importantly, all of these proteomic studies defined as Xist binders those proteins that were consistently enriched over Xist RNA but not over other non-coding RNAs, with the obvious caveat that ubiquitously bound factors are not necessary functionally less important.

In addition to the Xist interactome studies, two independent forward genetic screens were published, profiling extensive lists of factors required for Xist-mediated silencing [Moindrot et al. 2015, Monfort et al. 2015]. Moindrot *et al* (2015), made use of a male cell line in which a doxycycline-inducible Xist transgene and a GFP reporter were engineered *in cis* on chromosome 17. Using a lentiviral short hairpin RNA (shRNA) library, the authors screened for proteins whose knock-down following Xist induction resulted in increased GFP expression. Monfort *et al* (2015), used a male haploid ES cell system where Xist can be induced from the endogenous locus using doxycycline. Upon induction, cells die due to repression of their single X chromosome. Using a viral mutagenesis strategy, the authors screened for factors that, when mutated, increased cell viability, due to inhibition of silencing. Putting together the results for all proteomic and genetic screens, six factors emerged as

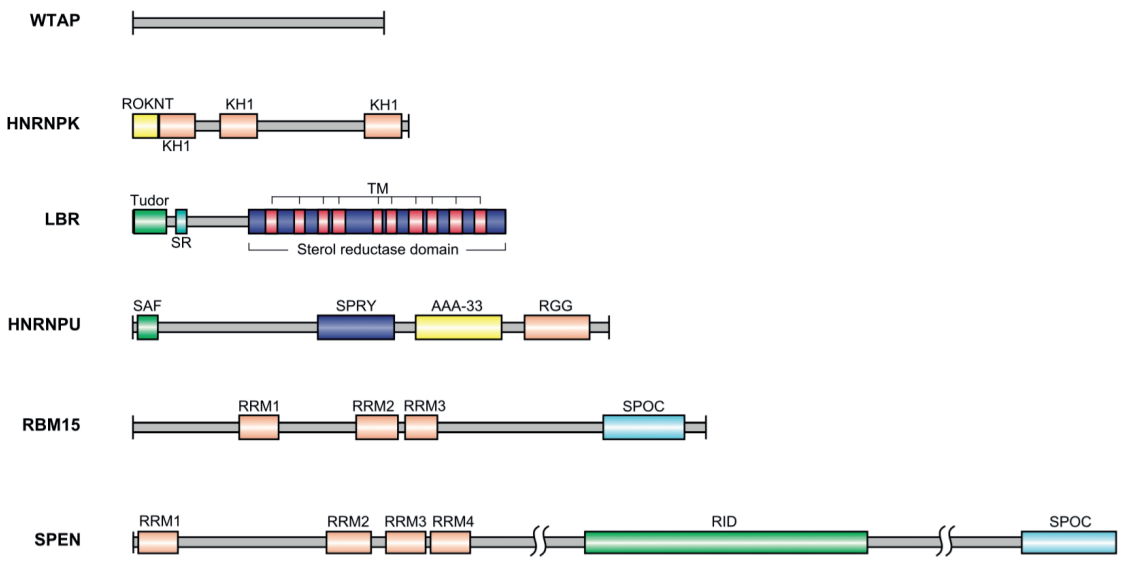


Figure 1.9: Xist *bona fide* interacting partners. Annotated functional domains of Xist interacting proteins, as listed in Table 1.2. Adapted from Moindrot *et al* (2016).

being *bona fide* functional Xist interacting partners: Spen, Rbm15, Wtap, LBR, hnRNP-U, and hnRNP-K (Table 1.2, Figure 1.9).

Table 1.2: Selection of *bona fide* Xist interactors and regulators of repression on the Xi. Adapted from Mira-Bontenbal *et al* (2016).

	Study				Function		
Protein ID	**McHugh *et al***	**Chu *et al***	**Moindrot *et al***	**Monfort *et al***	**Xist transcription**	**Xist localisation**	**Silencing**
Spen	+	+	+	+	-	-	+
Rbm15	+	+	+	-	-	-	+/-
Wtap	-	+	+	-	-	-	+/-
LBR	+	+	-	-	-	-	+
hnRNPU	+	+	-	-	-	+	+
hnRNPK	-	+	-	-	?	-	+

1.5.1 The Ncor/HDAC pathway: Spen and Rbm15

Spen, also known as SHARP in human and Mint in mouse, is the only factor identified in all of the aforementioned studies (Table 1.2), implying it is a high-confidence Xist interacting partner, with a key role in Xist-mediated silencing. A recent study, employing the PARIS method to identify and functionally characterise RNA secondary folds, found the A-repeats of Xist to be a Spen binding motif [Lu et al. 2016], and confirmed previously published eCLIP data on human XIST [Van Nostrand et al. 2016]. Functional characterisation of Spen using 3D SIM

super-resolution microscopy analysis, confirmed co-localisation between Xist and Spen, whilst knock-down experiments, identified defects in Xist-mediated silencing [Chu et al. 2015, McHugh et al. 2015, Moindrot et al. 2015, Monfort et al. 2015]. Moreover, a possible role of Spen in PRC1/2 recruitment was also hypothesised, as a defect in H3K27me3 deposition was observed upon Spen knock-down [McHugh et al. 2015, Monfort et al. 2015]. Whether this effect is direct or indirect, is still unclear, especially given that Moindrot *et al* (2015), did not report any H3K27me3 defect in Spen knock-down experiments. Spen-mediated transcriptional silencing was previously shown to depend upon its SPOC domain (Fig.1.9), which interacts with the C-terminus of the SMRT/Ncor complex (silencing mediator for retinoid and thyroid receptor co-repressor) [Shi et al. 2001], and can recruit Hdac1 and Hdac2 [Mikami et al. 2014]. Indeed, McHugh *et al* (2015), showed that the silencing defect observed following Spen knock-down was recapitulated through knock-down experiments targeting Hdac3 and Smrt.

A SPOC domain is also present in Rbm15, identified by Moindrot *et al* (2015), and in all of the proteomic screens (Table 1.2). Validations performed by McHugh *et al* did not report any defect in Xist-mediated silencing. On the other hand, Moindrot *et al* (2015), did report silencing defects in knock-down as well as in knock-out experiments (unpublished). In addition to its role in transcriptional silencing, Rbm15 was also reported to contribute to mRNA processing, via its interaction with the RNA export factors Nxf1 and Nxt1 [Uranishi et al. 2009, Lindtner et al. 2006], responsible for exporting most of the mature mRNAs; and Dbp5, an ATPase docked at the nuclear pore [Zolotukhin et al. 2008].

1.5.2 Wtap and the m^6A complex

In contrast to all other Xist binding factors, Wtap does not contain any canonical RNA-binding motif. Also, despite being identified as a biochemical Xist interactor by Chu *et al* (2015), it does not appear among the proteins directly bound to Xist according to McHugh *et al* (Table 1.2), suggesting that its interaction with

Xist could be mediated by another unknown factor. Wtap is implicated in the regulation of mRNA biogenesis, and is a regulatory subunit of the complex that establishes N^6-methyladenosine (m^6A) on RNA, also comprising Virilizer, Mettl3 and Mettl14 [Liu et al. 2014]. Interestingly, Moindrot *et al* (2015), identified both Wtap and Virilizer, as having a role in Xist-mediated silencing. In agreement with these findings, Xist was reported to be methylated in several regions [Dominissini et al. 2012, Patil et al. 2016], and its methylated state was shown to be crucial for gene silencing [Patil et al. 2016]. In a recent study, Rbm15 was suggested to represent the molecular adaptor between Xist RNA and the m^6A machinery, by interacting with Wtap, and Mettl3. According to the authors of the study, the role of Rbm15 in gene-silencing can be fully ascribed to its crucial activity in mediating the establishment of m^6A methylation on Xist, which can be subsequently read by YTHDC1, an m^6A binding protein. YTHDC1 was proposed to directly recruit other silencing factors, including Spen [Patil et al. 2016].

1.5.3 LBR and the nuclear lamina

LBR (Lamin B receptor), was first discovered as a Lamin B binding protein [Worman et al. 1988], closely associated with the nuclear lamina [Olins et al. 2014]. McHugh *et al* (2015), identified LBR as a high-confidence Xist binding factor (Table 1.2). Further validations performed by the same authors, showed that LBR depletion is linked to loss of Xist-mediated silencing and loss of exclusion of the transcriptional machinery from the inactive domain. In a recent study, LBR function in the context of XCI was linked to its ability to recruit the inactive chromosome to the nuclear lamina. In detail, LBR was proposed to elicit changes in the 3D structure of the Xi, enabling Xist and its silencing partners, to spread *in cis* and silence transcription [Chen et al. 2016].

1.5.4 The nuclear matrix proteins: hnRNP-U and hnRNP-K

Historically, a major challenge in identifying the Xist interactome, was the fact that Xist is retained in the insoluble nuclear matrix fraction, resulting in poor *ex vivo* isolation of intact ribonucleoprotein complexes. Recently, a 3D-SIM super-resolution microscopy analysis demonstrated that Xist co-localises with the nuclear matrix protein hnRNP-U/SAF-A [Smeets et al. 2014], already known to affect Xist localisation, possibly through direct interaction with Xist RNA [Hasegawa et al. 2010]. The idea of a direct link between Xist and the nuclear matrix, was also suggested by the observation that the nuclear matrix proteins SATB1/2 confer Xist responsiveness in tumour cells and adult somatic cells [Agrelo et al. 2009]. All of the recent Xist proteomic screens identified hnRNP-U, as one of their top hits [McHugh et al. 2015, Chu et al. 2015]. Furthermore, knock-down experiments showed loss of Xist-mediated *foci* and a diffuse nuclear localisation of Xist RNA in the absence of hnRNP-U, suggesting its involvement in the pathway for Xist localisation as discussed above [McHugh et al. 2015, Chu et al. 2015]. It is currently unclear why Moindrot *et al* (2015), and Monfort *et al* (2015), failed to identify hnRNP-U as important for Xist-mediated silencing in genetic screens, given the dramatic phenotype observed in the knock-down experiments; or why hnRNP-U enrichment is only observed late in the silencing cascade, when Xist localisation is already fully achieved [Pullirsch et al. 2010].

hnRNP-K is another major component of the nuclear matrix, and was proposed to act as a docking platform to mediate crosstalk between RNAs and several complexes related to a wide range of nuclear functions, notably gene activation/repression and nucleic acid processing [Bomsztyk et al. 2004]. hnRNP-K was also listed among the factors highly enriched on Xist compared to other non-coding RNAs (Table 1.2). Unlike other matrix proteins identified by Chu *et al* (2015), knock-down of hnRNP-K did not show any impaired Xist-coating of the Xi, suggesting that hnRNP-K is involved in Xist-mediated silencing functions, rather than in localisation. Despite

this, hnRNP-K was not found by McHugh *et al* (2015), who employed UV rather than formaldehyde as a cross-linking agent, implying that the protein does not directly contact Xist. Conversely, eCLIP data recently published by Van Nostrand *et al* (2016), show clear binding of hnRNP-K along human XIST RNA. Chu *et al* (2015), also tested the effects of hnRNP-K depletion on Xist-mediated *in cis* silencing and Polycomb recruitment, by siRNA-mediated knock-down of the protein. They reported a mild defect in the silencing ability of Xist (comparable to that observed for hnRNP-U), and also a clear reduction in Polycomb-mediated histone modifications.

1.6 Aims of this thesis

Previous studies applied a rigorous genetic approach to investigate the interface between Xist primary/secondary structure and its function, through serial deletions of the RNA either *in vitro* or *in vivo*. However, technological limitations, allowed testing of only a few phenotypes at a time, mostly cell/embryo viability and Polycomb recruitment. In this work I performed a comprehensive deletion analysis of Xist RNA with the aim of assessing silencing through allele-specific high-throughput sequencing of the transcriptome (Chapter 3). A quantitative analysis of gene silencing in combination with domain formation, defined a minimal functional Xist (Ev-Xist), spanning part of the first exon of FL Xist. Ev-Xist was subsequently employed to define the sequence of the RNA necessary to mediate proper coating of the future inactive chromosome (Chapter 4). The same system was also employed to precisely locate the segment of the RNA mediating the interaction between Xist RNA and Polycomb (Chapter 5), and to clarify the contribution of PRC1 and PRC2 in Xist-mediated gene-silencing, tested by measuring the silencing defects in the absence of the Polycomb interacting sequence of Ev-Xist (Chapter 6). In order to define the molecular mechanism of Xist interaction with Polycomb, an *in vitro* approach was adopted, where key Xist elements were tested for differential binding of molecular partners (Chapter 7). These experiments identified the nuclear matrix protein hnRNP-K as binding the Polycomb recruiting element, and further analysis demonstrated a role for this interaction in Polycomb recruitment (Chapter 8).

2
Materials and methods

Contents

2.1 Plasmids and cloning 30
2.1.1 Plasmids 30
2.1.2 Conventional cloning 31
2.1.3 Ligation Independent Cloning (LIC) 32
2.1.4 Gibson assembly 32
2.1.5 Primers for cloning 32
2.1.6 Bacterial transformation 32
2.2 Polymerase Chain Reaction (PCR) 33
2.3 Cell culture 34
2.3.1 Cell lines 34
2.3.2 Culturing conditions 34
2.3.3 Preparation of feeders 35
2.3.4 Doxycycline treatment 36
2.3.5 ES cells differentiation 36
2.3.6 Generation of stable cell lines 36
2.3.7 siRNA knock down 36
2.3.8 Metaphase spreads 37
2.3.9 Extracting genomic DNA from ES cells 37
2.4 Southern blotting 38
2.5 Antibodies 39
2.6 Cell extracts 39
2.6.1 Whole cell extracts 39
2.6.2 Nuclear cell extracts 39
2.7 Immunoblotting 40
2.8 Coupling GFP-nanobodies to magnetic beads 40
2.9 Co-immunoprecipitation 41
2.10 Immunofluorescence assays 41
2.10.1 Protein immunofluorescence 41

2.10.2 RNA-FISH 42
2.10.3 DNA-FISH on metaphase spreads 42
2.10.4 Immuno-FISH 43
2.11 Microscopy 43
2.12 Assessing RNA levels 43
2.12.1 RNA extraction 43
2.12.2 DNase treatment and quantitative real time PCR 44
2.13 4sU-seq 44
2.13.1 RNA labelling with 4-thiouridine 44
2.13.2 4sU-RNA biotinylation 45
2.13.3 Biotinylated 4sU-RNA isolation 45
2.13.4 4sU-RNA quality check 46
2.14 4sU-seq data analysis 46
2.14.1 Polymorphic sites between the genomes of 129S1 and CAST mouse strains 46
2.14.2 SNPs across the genome 46
2.14.3 Genomic distribution of the paired-end reads 46
2.14.4 4sU RNA isolation and RNA-seq library construction . 47
2.14.5 Definition and Calculation of repression score for Xist and miniXist RNA 47
2.14.6 4sU-seq alignment, allelic split and calling of gene expression 48
2.15 RNA pull-downs 48
2.15.1 *In vitro* transcription 48
2.15.2 RNA purification 49
2.15.3 RNA slot blot 49
2.15.4 RNA co-IPs 49
2.16 Mass spectrometry 50
2.17 Mass spectrometry analysis 51

2.1 Plasmids and cloning

2.1.1 Plasmids

Xist cDNA was inserted into a pTRE-tight promoter as in Moindrot et al. 2015. Deletions of Xist DNA were obtained by cutting it with relevant restriction enzymes, and, where required, re-inserting restriction fragments. For the ΔXN Xist-Bgl system, 18 repeats of the Bgl stem loop motif were fused to Xist, as in Moindrot et al. 2015. hnRNP-K (FL and deleted) and Pcgf3 coding sequences were obtained by PCR amplification from cDNA of a 129/1 WT cell line using primers listed in Table 2.1. In detail, primers LIC_hnRNP-K_F+LIC_hnRNP-K_R were used to generate FL hnRNP-K, LIC_hnRNP-K_F1+LIC_hnRNPK_R for Δ1,

LIC_hnRNPK_F2+LIC_hnRNPK_R for Δ2, LIC_hnRNPK_F3+LIC_hnRNPK _R for Δ3, LIC_hnRNPK_F4+LIC_hnRNPK_R for Δ4, LIC_hnRNPK_F+ LIC_hnRNPK_R2 for Δ5, LIC_hnRNPK_F1+LIC_hnRNPK_R2 for Δ6, and LIC_hnRNPK_F2+LIC_hnRNPK_R2 for Δ7. The sequences were inserted by LIC cloning into a modified pCAG-IRES-Puro mammalian expression plasmid between the coding sequences for a N-terminal enhanced GFP (eGFP) and a 3' IRES sequence that precedes a puromycin resistant gene. The same sequences were also cloned in a pCAG-BglG-IRES-Puro mammalian expression plasmid between the coding sequence for a N-terminal enhanced GFP (eGFP) and a 3' IRES sequence that precedes a puromycin resistance gene, previously obtained by in frame insertion of the BglG cDNA, derived from a BglG-mCherry fusion construct described in Moindrot et al. 2015 using primers listed in 2.1. For *in vitro* transcription studies, constructs XR (A-repeats, functional) and XNX (A-repeats, non-functional) [Wutz et al. 2002] kindly provided by Anton Wutz were used, cloned into a pBluescript plasmid (Stratagene) and fused to the B-repeats sense and antisense cut from Xist cDNA using restriction sites. In detail, the vector was cut with XhoI, the B-repeats with XbaI, blunt-ended and inserted in either direction. For the col1a1 targeting, the targeting vector described in Beard et al. 2006 was adapted by shortening the homology arms to 0.8kb on either side, adding MluI/PacI sites in the MCS (primers listed in Table 2.1) from pTre-tight, and inserting Ev-Xist in the MluI/PacI restriction sites, in the presence or absence of the B-repeats.

2.1.2 Conventional cloning

Assembly of vectors and inserts was obtained by either sticky-end or blunt-end cloning after gel extraction of purified plasmid-cut DNA using the Zymoclean Gel DNA Recovery Kit (Zymoresearch). Ligation was performed at 16 °C for 2h, 10 °C for 2h, and 4 °C overnight with T4 DNA Ligase (Promega). In case of non-directional sticky-end cloning, vector DNA was treated with 20U of CIAP (Gibco) at 37 °C for 5 minutes prior to gel extraction. Blunt-ending was obtained by treating DNA with 20U of T4 DNA Polymerase (ThermoScientific) for 30 min at 16 °C.

2.1.3 Ligation Independent Cloning (LIC)

5μg of vector plasmid was digested with the appropriate restriction enzyme in order to linearise it. The vector DNA was isolated by gel extraction using the Zymoclean Gel DNA Recovery Kit (Zymoresearch). Insert DNA was amplified by PCR and purified by gel extraction using the same kit. Vector and insert were processed by T4 DNA polymerase (ThermoScientific) in the presence of 2mM dGTP (vector) and 2mM dCTP (insert) for 30 min at 22 °C. After heat-inactivation of the polymerase (20 min, 75 °C), samples were purified using the DNA Clean & Concentrator-5 kit (Zymoresearch). Vector and insert were incubated using a range of ratios (3:1-1:6) in 2μL total volume for 30 min at 25 °C and introduced into bacteria.

2.1.4 Gibson assembly

Depending on the specific cloning, either two different PCR products, or a PCR product and a linearised vector were isolated by gel extraction using the Zymoclean Gel DNA Recovery Kit (Zymoresearch). 50ng of donor vector was incubated in a 1:3 ratio with the insert in the presence of 10μL of Gibson Assembly Master Mix (NEB) in 20μL total volume at 25 °C for 60 min. 5μL of the reaction were introduced into bacteria.

2.1.5 Primers for cloning

A complete list of cloning primers used in this study can be found in Table 2.1.

2.1.6 Bacterial transformation

Competent DH5α (fhuA2 lac(del)U169 phoA glnV44 Ψ80' lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17) cells were generated in-house using a protocol based on Inoue et al. 1990. Bacteria were thawed on ice and 2μL DNA of ligation mixture was added. DNA was mixed with bacteria by gentle flicking of the tube and incubated on ice for 25-30 min. Bacteria were heat-shocked at 42 °C for 30s and allowed to recover on ice for 2 min, before being incubated in 10 volumes of

Table 2.1: List of cloning primers used in this study. Overhangs are in lower case, annealed bases in upper case.

Name	Application	Sequence 5'→3'
LIC_hnRNPK_F	LIC cloning	tacttccaatccATGGAGACCGAACAGCCAGA
LIC_hnRNPK_R	LIC cloning	tatccacctttactTTAGAATCCTTCAACATCTGCAT
LIC_hnRNPK_F1	LIC cloning	tacttccaatccATGGAAATTCTGAAGAAAATCAT
LIC_hnRNPK_F2	LIC cloning	tacttccaatccATGAAACCTGATAGGGTTGTAGA
LIC_hnRNPK_F3	LIC cloning	tacttccaatccATGGGGTTCAGTGCTGATGAAAC
LIC_hnRNPK_F4	LIC cloning	tacttccaatccATGGAACCACAGGGTGGTTCTGG
LIC_hnRNPK_R2	LIC cloning	tatccacctttactCCAGAACCACCCTGTGGTTC
LIC_Pcgf3_F	LIC cloning	tacttccaatccATGTTGACCAGGAAGATTAAAC
LIC_Pcgf3_R	LIC cloning	tatccacctttactgTCAAGGGGGGCAAGGAGC
MluIPacI_F	LIC cloning	tacttccaatccGGAGAATTCGAGCTCGGT
MluIPacI_F	LIC cloning	tatccacctttactTTATTGCCAGAGTTTAAT
BglG_F	Gibson cloning	tcatcattttggcaaagaattccATGAACATGCAAATCACCAAA
BglG_R	Gibson cloning	ggattggaagtacaggttctccACGGTGGCGACCGGTAGCAA
5_external_F	Conventional cloning	CTGGTGAACAAGGCCCCT
5_external_R	Conventional cloning	CAGAGGATGTGGGGCTACAT
3_internal_F	Conventional cloning	CCGCCATCCCAACAATACAT
3_internal_R	Conventional cloning	ACGTGTCAATCAGTGGGTCT

Table 2.2: PCR conditions used in this study.

	Taq	Expand High Fidelity Plus	Phusion
DNA template	1μL	1μL	2μL
Buffer	1x Taq buffer + $MgCl_2$	1x HiFi buffer	1x HF buffer
Primers (each)	0.5μM	0.4μM	0.5μM
dNTPs	0.2μM	0.2μM	0.2μM
Enzyme	1.25U	2.5U	0.5U
Total volume	50μL	50μL	50μL

LB solution at 37 °C for 1 hour. Bacteria were then spread on high salt (10mg NaCl/mL) LB agar plates containing the appropriate antibiotic for selection.

2.2 Polymerase Chain Reaction (PCR)

For a summary of PCR conditions used for different enzymes, refer to Table 2.2. Standard cycling conditions consisted of a 3-step reaction, with initial incubation at 94 °C for 2 min, then 35 cycles of denaturation at 94 °C for 30s, annealing at 55 °C for 30s, and elongation at 72 °C for 1 min. Final elongation was performed at 72 °C for 5 min.

Table 2.3: List of gRNAs used in this study. PAM sequences are in lower case.

Name	Score	Off targets in genes	Sequence 5'→3'
ColA1_CrisprF_T1	85	12	caccgTAGAAGGGGCCTTCTAATGC
ColA1_CrisprF_T2	75	4	caccGAGGTTCATGAGCCCTCAAA
ColA1_CrisprF_T3	80	8	caccGCCCCTTCTATACTAAATTA

2.3 Cell culture

2.3.1 Cell lines

A modified version of pR26/P'nlsrtTA construct in which the puromycin resistance cassette was replaced by a hygromycin resistance cassette, was electroporated into hybrid C57BL/6JJcl x 129/SvJcl cells, to target rtTA into the constitutively active ROSA26 locus as previously described in Moindrot et al. 2015. Targeting was verified by Southern blot analysis using unique sequence probes as shown in Figure 3.2. All cell lines carrying an inducible Xist randomly integrated into the genome were generated by co-transfection into rtTA-expressing cells of the relevant vector and a puromycin expression cassette, as explained below. Correct integration was verified by RNA-FISH upon doxycycline treatment. hnRNP-K-eGFP fusion cell lines were obtained by transfection of the relevant construct into PGK 12.1 cells, and correct integration verified by western blot. BglG-hnRNP-K-eGFP and BglG-Pcgf3-eGFP cell lines were obtained by co-transfection into rtTA-expressing cells of either vector in combination with ΔXN-Xist-Bgl, and selected by GFP-Xist immuno-FISH upon doxycycline induction. Single copy Ev-Xist and Ev-Xist ΔB were obtained by co-transfection of targeting vector expressing a neomycin expression cassette and a cocktail of three gRNAs (Table 2.3). A complete list of cell lines used and generated in this study can be found in Table 2.4

2.3.2 Culturing conditions

ES cells were grown in ES cell medium, which consisted of Dulbecco's Modified Eagle Medium (DMEM, from Life Technologies) supplemented with 6.7% foetal calf serum (FCS, from Seralab), 2mM L-glutamine, 1x non-essential amino acids, 50M 2-mercaptoethanol, 50g/mL penicillin/streptomycin (all from Invitrogen) and LIF-

Table 2.4: List of cell lines used in this study.

Name	Constructs	Parental	Number of clones
pgk12.1	-	wt pgk (XX) x wt 129S1 (XY)	1
Dc7a2	-	wt 129S1 (XX) x wt Cast (XY)	1
P4D7	rtTA	Dc7a1	1
Xist-175	FL Xist	P4D7	2
Xist-739	Xist ΔSX	P4D7	2
Xist-822	Xist ΔXN	P4D7	2
Xist-1166	Xist ΔB	P4D7	2
Xist-923	Xist 3SX	P4D7	2
Xist-891	invSX	P4D7	2
Xist-1311	Xist ΔD	P4D7	2
Xist-1310	Xist ΔE	P4D7	2
Xist-977	Xist Ev	P4D7	2
Xist-980	Xist EvXb	P4D7	2
Xist-981	Xist 5Xb	P4D7	2
Xist-982	Xist 3Xb	P4D7	2
Xist-985	Xist invXb	P4D7	2
Xist-Nb	Xist Nb	P4D7	2
Xist-Bs	Xist Bs	P4D7	2
Xist-BlAc	Xist BlAc	P4D7	2
Xist-BlXb	Xist BlXb	P4D7	2
Xist-EvBsXb	Xist EvBsXb	P4D7	2
Xist-822-Bgl	Xist ΔXN-Bgl	P4D7	2
Ev-Xist	Ev-Xist (chr11)	P4D7	2
Ev-Xist ΔB	Ev-Xist ΔB (chr11)	P4D7	2
hnRNP-K-eGFP	hnRNP-K-eGFP	P4D7	2
eGFP-hnRNP-K	eGFP-hnRNP-K	P4D7	2
hnRNP-K-eGFP-FL	hnRNP-K-eGFP	pgk12.1	2
hnRNP-K-eGFP-Δ1	Δ1 hnRNP-K-eGFP	pgk12.1	2
hnRNP-K-eGFP-Δ2	Δ2 hnRNP-K-eGFP	pgk12.1	2
hnRNP-K-eGFP-Δ3	Δ3 hnRNP-K-eGFP	pgk12.1	2
hnRNP-K-eGFP-Δ4	Δ4 hnRNP-K-eGFP	pgk12.1	2
hnRNP-K-eGFP-Δ5	Δ5 hnRNP-K-eGFP	pgk12.1	2
hnRNP-K-eGFP-Δ6	Δ6 hnRNP-K-eGFP	pgk12.1	2
hnRNP-K-eGFP-Δ7	Δ7 hnRNP-K-eGFP	pgk12.1	2
BglG-hnRNP-K-eGFP	hnRNP-K-eGFP+Xist-ΔXN-Bgl	P4D7	2
BglG-Pcgf3-eGFP	hnRNP-K-eGFP+Xist-ΔXN-bg	P4D7	2

conditioned medium, made in house, at a concentration equivalent to 1000U/mL. Non-differentiating ES cells were grown on tissue culture dishes coated with PBS+1% gelatine. EC10 medium consisted of DMEM, supplemented with 6.7% FCS, 2mM L-glutamine, 1x non-essential amino acids, 50M 2-mercaptoethanol and 50g/mL penicillin/streptomycin. Cells were grown at 37 °C in a humid atmosphere with 5% CO_2. Cells were passaged using 0.05% trypsin-EDTA (Life Technologies) with 2% Chicken Serum (Life Technologies) and frozen in FCS + 10% DMSO.

2.3.3 Preparation of feeders

To prepare feeders for ES cell culture, SNL-P cells were grown to confluence in EC10 medium and inactivated by treating them with EC10 containing 10μg/mL of mitomycin C (Sigma) for 2h. SNL-P cells were used at $4x10^4/cm^2$.

2.3.4 Doxycycline treatment

All Xist transgene lines were treated with 1.5μg/mL doxycycline, 24 hours prior to relevant experiments, as previously described [Schoeftner et al. 2006], unless differently specified.

2.3.5 ES cells differentiation

ES cells were plated at a low density ($0.6 \cdot 10^6$cells) in 14cm non-gelatineised tissue culture dishes in EC10 medium, and grown in LIF-depleted conditions for 72 hours.

2.3.6 Generation of stable cell lines

All ES cells in this study are stable expressing lines. To generate stable ES cell lines, cells were plated in 6-well plates on feeders at a density of $1\text{-}2 \cdot 10^6$ cells/well, a day before transfection. Normal medium was replaced with antibiotics-free medium before transfection. Cells were transfected using Lipofectamine 2000 (Life Technologies, Invitrogen), according to the manufacturer's instructions. After 24 hours, transfected cells were passaged to 90mm gelatineised Petri dishes with feeders. Puromycin (1.5 to 2.5μg/mL) or G418 (400μg/mL) selections were applied 48 hours after lipofection. Cells were grown for 10-12 days under selection with medium being changed every day. Individual ES colonies were picked and expanded for further screening. For CRISPR-assisted homologous recombination, 1μg of each gRNA was added to the lipofection cocktail.

2.3.7 siRNA knock down

iGENOME SMARTpools siRNAs against hnRNP-K were purchased from Dharmacon. Transfection into cells was performed on coverslips previously inserted in 6-well plates with RNAiMAX (Invitrogen), according to the manufacturer's instructions.

2.3.8 Metaphase spreads

Cells were grown in T25 flasks until nearly confluent, and incubated at 37 °C with fresh medium containing 1.5μg/mL ethidium bromide (Roche). After 1hr and 20 min, colchicine (KaryoMAX Colcemid, Life Technologies) was added to a final concentration of 0.1μg/mL and incubated for a further 40 min. Then cells were washed in PBS and harvested by trypsinisation at room temperature. After inactivation of the trypsin by addition of medium, cells were pelleted by centrifugation (400g, 3 min, 25 °C). A pellet of 1-2 mm thickness was carefully resuspended in 1mL hypotonic solution (75 mM KCl) for no more than 5 min. Then 200μL freshly prepared fixative (75% methanol, 25% Acetic Acid, 4 °C) was added drop-wise. The tube was not agitated in any way, but carefully placed in the centrifuge for pelleting (400g, 3 min, 25 °C). Supernatant was removed leaving about 100μL, used to re-suspend the cells in by gentle flicking. 1mL of fixative was added to the resuspended cells and incubated overnight at 4 °C without agitation. The following day the cells were carefully resuspended in the same fixative and pelleted as before. The pellet was resuspended in 1.5mL fixative and pelleted again. This step was repeated twice more. Then cell suspension was dropped onto clean microscope slides and air dried.

2.3.9 Extracting genomic DNA from ES cells

Cells from a confluent T75 were harvested and resuspended in 1-3mL lysis buffer (10 mM NaCl, 10 mM Tris-HCl pH 7.5, 10 mM EDTA-NaOH pH 8.0, 0.5% Sodium lauroyl sarcosinate) and proteinase K was added to a final concentration of 200μg/mL. Samples were incubated overnight at 55 °C. An equal volume of phenol:chloroform:isoamyl alcohol 25:24:1 (Sigma) was added and incubated for 1 hr at 25 °C with agitation. After centrifugation (3600g, 10 min, 25 °C), the aqueous phase was transferred to a clean tube. 1/25 volume of 5 M NaAcetate and 2.5 volume of 100% ethanol was added. After mixing, a visible white cloud of DNA was extracted using a bent pipette tip and transferred to a clean tube containing 1 ml 70% ethanol. DNA was pelleted (16,100g, 5 min, 4 °C) and air

dried. Subsequently the pellet was resuspended in 300-400μL 10 mM Tris pH 8.5 and the concentration measured by Nanodrop.

2.4 Southern blotting

20μg of DNA was digested overnight in a 50μL reaction with 15U PstI and 15U EcoRI. Digestions were loaded onto a 1xTAE 0.8% agarose gel containing 0.1μg/mL ethidium bromide (Fisher Scientific). The gel was run at 80V for a minimum of 6 hours. All gel treatments were performed at 25 °C with agitation. The gel was rinsed in water and then washed twice in depurination solution (0.125M HCl) for 5 min and then twice with water for 4 min. Subsequently the gel was incubated for 45 min with denaturation buffer (0.5M NaOH, 1.5M NaCl). The gel was then washed twice with neutralisation buffer (0.5M Tris-HCl pH 7.5, 1.5M NaCl) and rinsed in water. DNA was then transferred to Hybond-XL (GE Healthcare) by capillary transfer with 20xSSC (3M NaCl, 0.3M Sodium Citrate dehydrate (Fisher Scientific), pH 7.0) for a minimum of 16 hrs. Next, the membranes were washed in 2xSSC for 5 min, and then allowed to dry for 1 hr, after which DNA was cross-linked to the membrane using the optimal crosslink setting on Stratalinker (which is 120,000 $\mu J/cm^2$). Membranes were then pre-incubated with Church buffer (0.25M sodium phosphate buffer pH 7.2, 7% SDS, 1mM EDTA-NaOH pH 8.0) at 65 °C with rotation. Probes were generated by PCR and purified by QIAquick gel extraction followed by QIAquick PCR purification (Qiagen) (Probes were generated by PCR on genomic DNA, using primers listed in Table 2.1). 30ng of each probe was labelled with 32P using the Prime-It II kit (Stratagene). Radio-labelled probes were added in fresh Church buffer and incubated overnight at 65 °C with rotation. Subsequently blots were washed once at 65 °C with low stringency wash solution (2xSSC, 0.1% SDS) and once at 25 °C with low Stringency wash solution, before being exposed to a Fuji MS-type imaging plate and imaged using a FLA-7000 scanner (GE Healthcare).

Table 2.5: List of antibodies used in this study. IF: immunofluorescence, IB: immunoblotting

Name	Source	Product ID	In IF	In IB
H3K27me3	Active motif	61017	1:1000	-
H3	Abcam	1791	-	1:10000
Ciz1	Gift	-	1:1000	-
Suz12	Cell Signalling	3737	1:1000	-
Ezh2	Diagenode	C15410039	1:1000	1:1000
GFP	Abcam	ab13970	1:2000	1:20000
H2AK119u1	NEB	8240	1:1000	-
H4 pan	Diagenode	pAB-156-050	1:1000	1:1000
hnRNP-K	NEB	4675	1:1000	1:1000
Jarid2	Cell signalling	13594	1:1000	-
RNA Pol II 8WG16	Covance	MMS-126R	1:500	-
Anti-mouse Ig, HPR	Amersham	NA934V	-	1:20000
Anti-rabbit Ig, HPR	Amersham	NA931V	-	1:20000
Alexa 568 anti-mouse IgG	Life technologies	A11031	1:400	-
Alexa 488 anti-mouse IgG	Life technologies	A11029	1:400	-
Alexa 568 anti-rabbit IgG	Life technologies	A11034	1:400	-
Alexa 488 anti-rabbit IgG	Life technologies	A11008	1:400	-
Alexa 488 anti-chicken IgG	Life technologies	A11039	1:400	-
Neutravidin-HRP	Life technologies	31030	-	1:500

2.5 Antibodies

A complete list of the antibodies used in this study can be found in Table 2.5.

2.6 Cell extracts

2.6.1 Whole cell extracts

Whole cell extracts were prepared by harvesting cells, washing, and re-suspending them in PBS. The cell suspension was diluted twofold with loading buffer (2% SDS, 20% glycerol, 0.1M Tris-HCl pH 8.0, 0.002% Bromophenol Blue, 1.4M β-mercaptoethanol (Sigma)). This preparation was boiled for 10 min at 100 °C.

2.6.2 Nuclear cell extracts

Nuclear cell extracts were prepared by harvesting cells and either processing them immediately or snap-freezing them on dry ice and storing them at −80 °C. In both cases, cell pellets were washed with PBS and resuspended in 10 packed cell volume (PCV) buffer A (10mM HEPES-KOH pH 7.9, 1.5mM $MgCl_2$, 10mM KCl,

with 0.5mM DTT, 0.5mM PMSF, and complete protease inhibitors (Roche) added fresh). After a 10 min incubation at 4 °C, cells were collected by centrifugation (1500g, 5 min, 4 °C) and resuspended in 3 PCV of buffer A + 0.1% NP-40 (Sigma). After another 10 min incubation at 4 °C, nuclei were collected by centrifugation (400g, 5 min, 4 °C) and resuspended in 1 PCV buffer C (250mM NaCl, 5mM HEPES-KOH (pH 7.9), 26% glycerol, 1.5mM $MgCl_2$, 0.2mM EDTA-NaOH, pH 8.0 with complete protease inhibitors (Roche) + 0.5mM DTT added fresh). 5M NaCl was added drop-wise to bring the concentration to 350mM and the mixture was incubated for 1 hr at 4 °C with occasional agitation. After centrifugation (16,100g, 20 min,4 °C), the concentration of the supernatant was quantified using the Bio-Rad Bradford assay and stored at −80 °C.

2.7 Immunoblotting

Samples were separated on a polyacrylamide gel and transferred onto a nitrocellulose membrane by semi-dry transfer (for all other protein; 15 V for 50 min). Membranes were blocked by incubating them for 1 hour at room temperature in 10 ml TBS, 0.1% Tween (TBST) with 5% w/v Marvell milk powder. Blots were incubated overnight at 4 °C with the primary antibody, washed 4 times for 10 min with TBST and incubated for 40 min with secondary antibody conjugated to horseradish peroxidase. After washing 4 times for 5 min with TBST, bands were visualised using ECL (GE Healthcare).

2.8 Coupling GFP-nanobodies to magnetic beads

165μL of M-280 Tosyl-activated Dynabeads (ThermoScientific) were washed twice in buffer A (0.1M H_2BO_3 pH 9.5), before adding 100μL of GFP-nanobodies (made in house), for a total volume of 150μL in buffer A. After addition of 100μL of buffer C (0.1M H_2BO_3, 3M $(NH_4)_2SO_4$), the mixture was incubated at 37 °C overnight. The day after, 1mL of buffer D (0.01M BSA in PBS) was added, and beads incubated at 4 °C for a further 1 hour. Coupled beads were washed twice in

buffer E (0.005M BSA in PBS), resuspended in buffer E to a final concentration of 20mg/mL, and stored at 4 °C.

2.9 Co-immunoprecipitation

500μg of nuclear extract were used and salt concentration was adjusted to 150mM NaCl in a 1mL total reaction volume. Extracts were treated with 250 U Benzonase nuclease (Millipore) for 30 min at 4 °C. 100μg of nanobodies were added and incubated at 4 °C overnight. The flow through was collected and beads were washed 3 times with 1 ml wash buffer (5mM HEPES-KOH pH 7.9, 1.5mM $MgCl_2$, 26% glycerol, 0.2mM EDTA, with complete protease inhibitors (Roche) + 0.5mM DTT added fresh) with 150mM NaCl, and 3 times with the same wash buffer with 450mM NaCl. Subsequently beads were boiled in 50μL SMASH buffer (50mM Tris.HCl pH 6.8, 10% Glycerol, 2% SDS, 0.02% bromophenol blue, 1% β-mercaptoethanol) for 10 min at 95 °C. 5μL of the supernatant was loaded as the IP sample.

2.10 Immunofluorescence assays

2.10.1 Protein immunofluorescence

Cells were plated on slides or coverslips (13mm diameter cover slips from VWR), at least a day before the experiment. On the day of the experiment, cells on slides were washed with PBS and then fixed with 2% formaldehyde for 15 mins followed by 5 mins of permeabilisation using 0.4% Triton X-100. Cells were briefly washed with PBS before blocking with a 0.2% PBS-based solution of fish gelatine for three periods of 5 mins. Primary antibody dilutions were prepared in fish gelatine solution with 5% normal goat or normal donkey serum depending on the secondary antibodies used. Primary antibody dilutions are listed in Table 2.5. Cells on slides were incubated with primary antibodies for 2 hours in a humid chamber at room temperature. Slides were washed three times in fish gelatine solution. Secondary antibodies were diluted 1:400 in fish gelatine solution and incubated with cells on slides for 30 min in a humid chamber at 37 °C. After incubation, slides were washed

twice with fish gelatine and one time with PBS before mounting using Vectashield mounting medium with 4,6-diamidino-2-phenylindole (DAPI). Excess mounting medium was removed and the coverslips were sealed to slides using nail varnish.

2.10.2 RNA-FISH

Cells were plated on either slides or coverslips. After washing twice with PBS, cells were fixed for 10 mins with 2.6% formaldehyde followed by permeabilisation with 0.42% Triton X-100 for 5 mins at 4 °C. After fixation, cells on slides were washed twice with 70% ethanol and dehydrated using a series of concentrated ethanol solutions (80%, 95% and 100%). Slides were briefly air dried before incubation with probe in a humid chamber overnight at 37 °C. Xist RNA probe was generated from an 18 kb fragment spanning the whole Xist transcript using a nick translation kit (Abbott Molecular) as previously described [Moindrot et al. 2015]. Labelled RNA probes (1.5μL) were co-precipitated with 10μg salmon sperm DNA, 1/10 volume 3M sodium acetate (pH 5.2) and 3 vol ethanol. After washing in 75% ethanol, the pellet was dried, resuspended in 6μL formamide and denatured at 75 °C for 7 min before flash cooling on ice. Probe was diluted in 6μL 2x hybridisation buffer (5x SSC, 12.5% dextran sulfate, 2.5mg/mL BSA (NEB)), added to the slide and incubated overnight at 37 °C in a humid chamber. After incubation, slides were washed three times with a solution of 2xSSC/50% formamide followed by three washes with 2xSSC in a water bath at 42 °C. Slides were mounted and sealed as for immunofluorescence.

2.10.3 DNA-FISH on metaphase spreads

Metaphase spreads prepared on slides were de-hydrated in 75%, 80%, and 100% ethanol for 2 min at 25 °C, denatured in 50% formamide/2%SSC at 65 °C for 2 hours, and de-hydrated again in ice-cold 75% ethanol for 2 min, followed by another 2 min in 80%, and 100% at 25 °C. Xist probe was prepared as for RNA-FISH. Once the probes were added to the slides, another 3 minutes of denaturation at 65 °C were performed before over-night incubation at 37 °C in a humid chamber. After incubation, slides were washed three times with a solution of 2xSSC/50%

formamide followed by three washes with 2xSSC in a water bath at 42 °C. Slides were mounted and sealed as for immunofluorescence.

2.10.4 Immuno-FISH

RNA-FISH was performed essentially as previously described. Cells grown on slides or coverslips were washed in PBS, and fixed with 4% formaldehyde for 10 min before permeabilisation in 0.4% Triton X- 100 for 5 min. After washing in PBS, slides were incubated with 50% formamide in 2xSSC for 2 h. Washes were performed twice in 50% formamide in 2x SSC for 3 min at 42 °C, twice in 2xSSC for 3 min at 42 °C and twice in PBS at room temperature. Slides were blocked in 0.2% fish gelatine (Sigma) three times for 2 min each, and primary antibodies were added in 0.2% fish gelatine and 5% normal goat serum in PBS. The slides were then processed in an identical manner to the immunofluorescence protocol detailed above.

2.11 Microscopy

Z stack images were acquired with a DeltaVision system (Applied Precision) with a 1003/1.40 NA objective (Olympus) and de-convolved using the SoftWorx software algorithm (Applied Precision, conservative ratio method, 10 iterations). Best exposure time for each field and channel was manually determined and kept fixed among experiments. A minimum of 100 cells were generally analysed per slide and each experiment was performed in a minimum of 3 replicates. Further image editing and refinement was achieved through Fiji/ImageJ.

2.12 Assessing RNA levels

2.12.1 RNA extraction

Cells from 90mm dishes were harvested and washed twice in PBS after which the cell pellet was resuspended in 1-2mL Trizol reagent (Life Technologies). Samples were incubated for 2 min at 25 °C and then 0.2 volume chloroform was added to each and mixed by vigorous shaking for 15 s. Samples were incubated for 2 min at 25 °C and

Table 2.6: List of primers used for real time quantitative PCR.

Name	Sequence 5'→3'
qXist1endF	TCCAGGCAATCCTTCTTCTTG
qXist2R	GATCCTGCTTGAACTACTGCT
Cdx2_F_RT	TCAACCTCGCCACAACCTTCCC
Cdx2_R_RT	tatccacctttactTTATTGCCAGAGTTTAAT
DXN_F	TCTTCTCCTTAGCCCATCGG
DXN_R	AACCATATTTCCATCCACCAA
rtTA_F	GACTCGAGATGCCAAAGAGACCCAGACCT
rtTA_R	GCGCGGCCGCCTACCCACCGTACTCGT

centrifuged at 12,000g for 15 min at 4 °C. The upper aqueous phase was transferred to a clean pre-chilled tube and an equal volume of isopropanol, pre-chilled to 4 °C, was added. The sample was incubated at 4 °C for 10 min and RNA was pelleted at 12,000g at 4 °C for 30-60 min. Subsequently the pellet was washed twice in 75% ethanol and then air-dried in 3-5 min and resuspended in 100μL RNase-free water.

2.12.2 DNase treatment and quantitative real time PCR

Contaminating DNA was removed using the Ambion DNA-free DNase Treatment kit (Life Technologies) according to the manufacturer's instructions. cDNA was synthesised using the SuperScript III kit (Life Technologies) using Oligo dT primers (Life Technologies), according to the manufacturer's instructions. Quantitative PCR assays were carried out using the Rotor-Gene Q (Qiagen) and the iQ SYBR GREEN Supermix (BioRad) using the primers listed in Table 2.6.

2.13 4sU-seq

2.13.1 RNA labelling with 4-thiouridine

4-thiouridine (4sU, Sigma, T4509) was dissolve in sterile PBS and stored at −20 °C. 4sU was thawed just before use and added to the cells in the growing media at a concentration of 500μM. Cells were incubated with the 4sU-implemented medium for 12 minutes. Cell culture medium was rapidly removed from cels and 5mL of Trizol reagent (Life Technologies) were added.

2.13.2 4sU-RNA biotinylation

Total RNA was extracted as described above, treated with DNase using the Ambion DNA-free DNase Treatment kit according to the manufacturer's instructions, and resuspended in water. For each μg of total RNA, 2μL of Biotin-HPDP (Pierce, 50mg EZ-Link Biotin-HPDP), previously dissolved in DMF at a concentration of 1mg/mL, and 1μL of 10xBiotinylation buffer (100mM Tris HCl pH 7.4, 10mM EDTA), were added. The reaction was incubated with rotation for 15 min at 25 °C. RNA was transferred to Phase Lock Gel Heavy Tubes (Eppendorf), and an equal volume of chloroform was added. After vigorously mixing, tubes were left incubating for 3 min at 25 °C and then centrifuged at 13,000rpm for 5 min at 4 °C. The upper phase was transferred to new Phase Lock Gel Heavy Tubes, and chloroform added again. After further centrifugation, the upper phase was transferred to a tube containing an equal volume of isopropanol and 1/10 volume of 5M NaCl. After inversion, the tubes were centrifuged at 13,000rpm for 20 min at 4 °C. Supernatant was washed in 75% ethanol and resuspended in water.

2.13.3 Biotinylated 4sU-RNA isolation

Biotinylated 4sU-RNA was recovered using the μMacs Streptavidin Kit (Miltenyi), with a modified protocol. Per μg of recovered biotynilated 4sU-RNA, 0.5μL of streptavidin beads were added, in a total volume of 200μL. Samples were incubated with rotation for 15 min at 25 °C. μMacs columns supplied with the μMacs Streptavidin Kit were equilibrated in 1mL of washing buffer (100mM Tris HCl pH 7.5, 10mM EDTA, 1M NaCl, 0.1% Tween 20) at 65 °C. Samples were added to the columns that were then washed 6 time with washing buffer, 3 times at 65 °C and three times at 25 °C. RNA was eluted in freshly-prepared 100mM DTT. RNA was further purified using the DNA Clean & Concentrator-5 kit (Zymoresearch) according to the manufacturer's guidelines.

2.13.4 4sU-RNA quality check

1μL of 4sU-labelled RNA was quality-checked using the Agilent RNA 6000 Pico kit (Agilent technologies) according to the manufacturer's instructions, and run on a 2100 Bioanalyzer Instrument (Agilent). Quantification of recovered RNA provided by the instrument was also recorded.

2.14 4sU-seq data analysis

2.14.1 Polymorphic sites between the genomes of 129S1 and CAST mouse strains

Known polymorphic sites (just SNPs, not indels) for the individual 129S1 and Cast mouse strains were collected from the Sanger Mouse Genome Project. The GRC38/mm10 genome assembly (dbSNP142) was used, with 5,878,742 and 22,603,978 high quality SNPs annotated for 129S1 and Cast, respectively. Polymorphic sites between 129S1 and Cast were compared, resulting in a total of 23,005,850 occurrences between the genomes. By comparison with C57BL/6J reference genome, 25,726,959 polymorphic sites were found in both the 129S1 and Cast genome sequences, which were employed to generate an "N-masked" genomic sequence.

2.14.2 SNPs across the genome

Absolute number of SNPs was counted for every megabase region across each chromosome, then the last bin region shorter than one megabase was excluded. The heatmap was drawn with R using the z-score value. Although only 41.17% (9471317) SNPs are in the annotated intergenic regions (the mouse genome annotation was downloaded from GENOCDE_vM8), 96.74% of annotated transcripts contained at least one SNPs site. Among the intragenic SNPs, 8,803,708 (92.95%) were located in intronic regions.

2.14.3 Genomic distribution of the paired-end reads

After alignment, bedtools were employed to annotate the reads with genomic features. For instance, if at least one of the read mates was mapped in a spliced

region, the read was defined as "spliced", indicating that it originated from a mature transcript. Meanwhile, if both read mates were not spliced and overlapped introns, the read was identified as "intron fraction", most probably originating from pre-mRNA. The reads from exons cannot be distinguished from mature or nascent transcripts. These reads were assigned as "others". The four full-length Xist 4sU-seq libraries generated in this study were compared with the downloaded RNA-seq datasets, one is same protocol (SRR2040596) and the other is ENCODE protocol.

2.14.4 4sU RNA isolation and RNA-seq library construction

The libraries were constructed using Illumina TruSeq Stranded Total RNA Library Prep Kit, with the incorporation of dUTP in the second strand synthesis.

2.14.5 Definition and Calculation of repression score for Xist and miniXist RNA

According to the 4sU RNA-seq data, the Repression Score (RS) for Xist RNA was calculated to compare the repression score in different chromosomes and different regions across the chromosomes. RS of each gene was defined as:

$$RS = \frac{Y_i - Y_0}{Y_0} - \frac{X_i - X_0}{X_0} \tag{2.1}$$

Here, X and Y represent the gene expression value of 129S1 and Cast alleles, respectively. The labels 0 represent the no dox-induced time point of experiment, while i represent the different time points (24 hours, 48 hours and 72 hours treatment of dox). In theory, the RS value is comprised between 0 and 1. Therefore all experimentally measured RS larger than 1, were set to 1, and lower than 0, were set to 0. For the silencing ability across the chromosome, a 10 MB bin window was used. The RS distribution was compared for bin windows across the selected chromosome. All the custom scripts used in this study are available at https://github.com/guifengwei/XCI.

2.14.6 4sU-seq alignment, allelic split and calling of gene expression

The raw fastq files of read pairs were mapped to previously build rRNA, then the remaining unmapped reads were aligned to the "N-masked" genome with STAR (2.4.2a) using the parameters: "$--outFilterMultimapNmax 1 --outFilterMismatchNmax 2 --alignEndsTypeEndToEnd$" for all the 4sU-seq libraries. The unique alignments were reserved for further analysis. Then, SNPsplit (http://www.bioinformatics.babraham.ac.uk/projects/SNPsplit/) was employed to separate the alignment into distinct alleles (Cast and 129S1) with the paired mode. To call the allelic gene expression, the numbers of unique mapped read pairs for each annotated gene (CPM) were counted using the HTseq3 with the parameters "-t transcript -s reverse", then normalised to million mapped read pairs (library size). The CPM value was split by the G1 (Cast) or G2 (129S1) ratio. Cast Allelic expression was defined as $CPM * G1/(G1 + G2)$; 129S1 allelic expression as $CPM * G2/(G1 + G2)$. Here G1 or G2 represent the number of allele-specific read pairs. For libraries generated for the time-course 4sU-seq experiment of full-length Xist RNA, only the genes whose expression value sum over the complete time course from Cast and 129S1 alleles were more than 8 were retained. Otherwise, for the experiments of various truncated Xist RNA, the threshold was set to 6. For visualisation of RNA-seq data in UCSC genome browser, the coverage was normalised to 10 million mapped reads.

2.15 RNA pull-downs

2.15.1 *In vitro* transcription

RNAs were transcribed from a T7 promoter, following linearisation with excess of PvuI. 1μg of DNA was *in vitro* transcribed using the HiScribe T7 High Yield RNA Synthesis Kit (NEB), according to the manufacturer's guidelines. In the reaction 1μL of Biotin-16-dUTP (Sigma) was added, at a concentration of 50μM. The size of the RNA was checked on a 1% agarose gel electrophoresed in MOPS buffer (0.2M

MOPS pH 7.0, 0.05M sodium acetate, 0.005M EDTA pH 8.0). Before loading on the gel, 0.5μg of each RNA sample, were diluted into 5xMOPS buffer, in the presence of 20μL of formamide and 7μL of 37% formaldehyde, and heated at 68 °C for 15 min.

2.15.2 RNA purification

In vitro transcribed RNA was diluted in a final volume of 100μL, treated with DNase using the Ambion DNA-free DNase Treatment kit according to the manufacturer's instructions, and resuspended in water. DNA-free RNA was subsequently added to Phase Lock Gel Heavy Tubes (Eppendorf), with an equal volume of chloroform. After vigorously mixing, tubes were left incubating for 3 min at 25 °C and then centrifuged at 13,000rpm for 5 min at 4 °C. The upper phase was transferred to normal tubes containing an equal volume of isopropanol and 1/10 volume of 5M NaCl. After inversion, the tubes were centrifuged at 13,000rpm for 20 min at 4 °C. Supernatant was washed in 75% ethanol and resuspended in water.

2.15.3 RNA slot blot

500ng of each RNA sample was diluted into 10μL, and mixed with 20μL of formamide, 7μL of 37% formaldehyde, 2μL of 20xSSC. Tubes were heated at 68 °C for 15 min, chilled on ice and diluted in 2 volumes of 20xSCC. Samples were then loaded on a blotter, and transferred onto Amersham Hybond-XL nylon membrane (GE Healthcare). RNA was then cross-linked to the membrane using the optimal crosslink setting on Stratalinker (120,000$\mu J/cm^2$). The membrane was blocked by incubation for 1 hour at room temperature in 10 ml TBS, 0.1% Tween (TBST) with 5% w/v Marvell milk powder, and incubated for 1 hour at 25 °C with Neutravidin-HRP antibody, washed 4 times for 10 min with TBST and visualised using ECL (GE Healthcare).

2.15.4 RNA co-IPs

15μg of nuclear extract from PGK 12.1 cells was incubated with 10μg of each RNA in the presence of 5μL RNasin Ribonuclease Inhibitors overnight at 4 °C in

1mL total volume of incubation buffer (150mM NaCl, 15mM Tris HCl (pH 7.5), 0.5mM EDTA-NaOH, pH 8.0, 0.025%NP40, with complete protease inhibitors (Roche) + 1mM DTT added fresh). Samples were transferred to a tissue culture 6-well plate on ice, and cross-linked using a Stratalinler at 250,000$\mu J/cm^2$. Samples were transferred back to protein LoBind micro-centrifuge tubes (Eppendorf), and 50μL of M-280 Streptavidin Dynabeads (ThermoFisher) were added to each of them. Tubes were incubated with rotation at 4 °C overnight. Flow through was collected for further analysis, whereas beads were washed 2 times in wash buffer A (150mM NaCl, 15mM Tris HCl (pH 7.5), 0.5mM EDTA-NaOH, pH 8.0 with complete protease inhibitors (Roche) + 1mM DTT added fresh), 2 times in wash buffer B (450mM NaCl, 15mM Tris HCl pH 7.5, 0.5mM EDTA-NaOH, pH 8.0, 0.025%NP40, 0.05%SDS, with complete protease inhibitors (Roche) + 0.1mM DTT added fresh), and 2 times in wash buffer C (150mM NaCl, 15mM Tris HCl pH 7.5, 0.5mM EDTA-NaOH, pH 8.0, with complete protease inhibitors (Roche)). Elution was performed in elution buffer (15mM Tris HCl pH 7.5, 0.02%SDS) for 15 min at 65 °C.

2.16 Mass spectrometry

Peptides were resuspended in 5% Formic acid and 5% DMSO and then trapped on a C18 PepMap100 pre-column (300 μm i.d. x 5 mm, 100 Å, Thermo Fisher Scientific) using 0.1% Formic acid in water at a pressure of 500 bar and analysed on an Ultimate 3000 UHPLC system (Thermo Fischer Scientific) coupled to a QExactive mass spectrometer (Thermo Fischer Scientific). The peptides were separated on an in-house packed analytical column (360 μm x 75 μm i.d. packed with ReproSil-Pur 120 C18-AQ, 1.9 μm, 120 Å) and then electrosprayed directly into an QExactive mass spectrometer (Thermo Fischer Scientific) through an EASY-Spray nano-electrospray ion source (Thermo Fischer Scientific) using a linear gradient (length: 60 min, 7% to 28% solvent B (0.1% Formic acid in acetonitrile), flow rate: 200 nL/minute). The raw data was acquired on the mass spectrometer in a data-dependent mode (DDA). Full scan MS spectra were acquired in the Orbitrap (scan range 350-2000 m/z, resolution 70000, AGC target 3xe^6, maximum injection time 100 ms). After

the MS scans, the 20 most intense peaks were selected for HCD fragmentation at 30% of normalised collision energy. HCD spectra were also acquired in the Orbitrap (resolution 17500, AGC target $5xe^4$, maximum injection time 120 ms) with first fixed mass at 180 m/z. The raw data files generated were processed using MaxQuant (Version 1.5.0.35), integrated with the Andromeda search engine as previously described. To identify protein groups, peak lists were searched against mouse database as well as list of common contaminants by Andromeda. Trypsin with a maximum number of missed cleavages of 2 was chosen. Acetylation (Protein N-term, i.e. only the n-terminus of the protein), Oxidation (M) and Phosphorylation (S, T and Y) were used as variable modifications while Carbamidomethylation (C) was set as a fixed modification. Protein and PTM false discovery rate (FDR) were set at 0.01 and a minimum score of 40 and localisation probability of $\geq$0.7 for phospho-peptides. Match between runs was applied.

2.17 Mass spectrometry analysis

For the Xist A^+B^+ interactome, all hits annotated as contaminants by the Andromeda research engine were rejected. Subsequently all 280 identified hits were compared with the ones identified as Xist A^-B^- interactors. MS abundance was subtracted between the lists, and only the hits still passing the threshold of significance were kept. An identical approach was used to identify the A- and B-repeat interactomes, where all Xist A^+B^+ hits were compared with the ones identified in the Xist A^-B^+, and Xist A^+B^- experiments respectively. All remaining hits were classified according to the assigned score after subtraction.

3

Establishing a system to analyse the function of Xist transgenes

Contents

3.1 Introduction 52
3.2 Results 55
3.2.1 Generation of a Tet-On hybrid cell line 55
3.2.2 Generation of an *in vivo* Xist deletion library 56
3.2.3 Xist transgenes trigger early exclusion of RNA PolII and histone H4 hypoacetylation 59
3.2.4 The A-repeats of Xist are required to trigger early exclusion of RNA PolII and histone H4 hypoacetylation . . . 63
3.2.5 Assessment of Xist-mediated gene silencing by 4sU-seq . 63
3.2.6 FL Xist is able to ectopically silence genes upon 24 hours induction in ES cells 65
3.2.7 The A-repeats are essential for gene silencing 69
3.2.8 Ev-Xist is able to mediate gene silencing 69
3.3 Discussion 71

3.1 Introduction

Xist is a 15-17kb transcript in mouse, comprising two large and six small exons, highly enriched for conserved tandem repeats. Many classical genetic studies showed that the presence of an expressed form of Xist is sufficient to initiate the silencing

cascade *in cis*, even at ectopic loci [Lee et al. 1996, Wutz et al. 2000]. However, the link between the role of Xist in mediating the inactivation process, and its sequence/secondary folds, remains poorly understood.

An effort to functionally characterise the Xist molecule was made in 2002, in the context of a systematic deletion analysis of mouse Xist RNA [Wutz et al. 2002]. In this study, a comprehensive panel of Xist cDNA deletion variants was artificially expressed from the X chromosome of male cultured ES cells. The survival of cells under forced Xist expression was used as an indirect measure of silencing. The binary outcome of this experiment clearly showed the involvement of the most 5' end of Xist (the A-repeats) in gene silencing. The A-repeats consist of a conserved AT-rich motif, predicted to fold in a well-defined secondary structure [Lu et al. 2016], and previously proposed to be a good candidate for functional studies, based on conservation arguments [Brown et al. 1992, Nesterova et al. 2011].

The method employed by Wutz *et al* (2002), to assign function to Xist segments is nevertheless relatively crude and it cannot be ruled out that other regions of Xist RNA contribute more subtly to silencing. Indeed, a small insertion in a region of Xist RNA downstream of the A-repeats was reported to abrogate silencing, at least to some degree [Hoki et al. 2011]. Subsequent studies used more refined model systems to confirm the involvement of the A-repeats in gene silencing, by deleting the most 5' end of the RNA (ΔSX Xist) in the endogenous locus of female cells, and assessing Xist defects by analysis of selected transcripts or RAP-seq [da Rocha et al. 2014, Engreitz et al. 2013]. However these studies were not specifically aimed at assigning function to any other Xist segments, and ΔSX Xist was mostly used as a negative control for gene silencing assays often in the context of Xist-independent genetic perturbations of cells.

The observed accumulation of the Polycomb group of proteins on the Xi soon after Xist up-regulation, suggested a simple correlation between transcriptional

silencing and Polycomb-mediated establishment of a silent chromatin state. In agreement with their crucial role in gene silencing, the A-repeats of Xist were proposed to mediate PRC2 recruitment [Zhao et al. 2008]. However, later studies showed that PRC2 recruitment depends on the region of Xist defined by the XN deletion, and not on the A-repeats [da Rocha et al. 2014], but no mechanisms of interactions were proposed, nor further studies conducted on the XN region.

Efforts were also made to characterise the small exon IV of Xist, highly conserved between human and mouse, and predicted to fold into a stable RNA stem-loop structure [Caparros et al. 2002]. However only a mild phenotype was observed *in vivo* upon loss of exon IV, suggesting that a compensatory mechanism might have masked the effects of the deletion. In this perspective, a wider panel of deletions could disentangle potentially redundant elements, revealing their separate functions.

In summary, Wutz *et al* (2002), identified elements of Xist that recapitulate its silencing ability, but the employed experimental system did not allow sensitive analysis of the deletion variants that do not impair silencing. In the same way, due to technological limitations, many other studies were restricted to candidate-driven analyses of specific regions of Xist, and relied on individual phenotypes or indirect effects upon simple deletions [Caparros et al. 2002, Senner et al. 2011, Yamada et al. 2015]. Work in this section was aimed at setting up an experimental system to dissect Xist-mediated function *in vivo*, through high-throughput sequencing analysis of transcriptional changes, following deletion of potentially key elements of the RNA. In this context, the ΔSX deletion of Xist could be used as a positive control, given the reported involvement of the A-repeats and their binding partners in the silencing pathway.

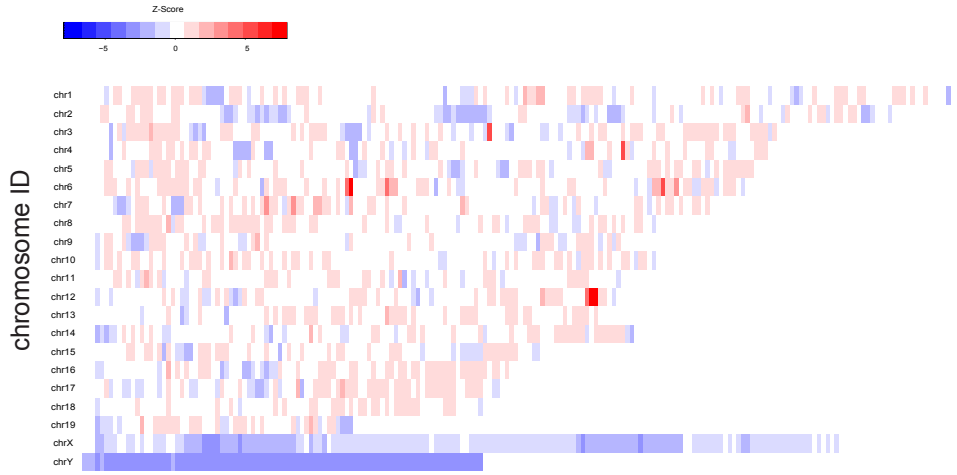


Figure 3.1: SNPs density between the *domesticus* and *castaneus* strains of *Mus musculus*. Density of annotated SNPs between *M.m.domesticus* and *M.m.castaneus*. All mouse chromosomes are listed on the Y-axis. The X-axis represents the length of each chromosome, segmented in 1Mb windows. The z-score represents the density of SNPs for each window, according to the formula $z = (x - \mu)/\sigma$ and is colour-coded according to the legend. Figure generated in collaboration with Dr. Guifeng Wei.

3.2 Results

3.2.1 Generation of a Tet-On hybrid cell line

The first comprehensive study of genetic variations in commonly used mouse strains, found 8.27 million high quality SNPs densely distributed along the genome [Frazer et al. 2007]. The haplotype map resulted in three distinct types mainly derived from the genetic contribution of four *Mus musculus* subspecies: *M.m.domesticus*, *M.M.molossinus*, *M.m.musculus*, and *M.m.castaneus* (68%, 6%, 10%, 3% respectively). Since this pioneer study, the number of annotated inter-specific SNPs increased, reaching the number of 23,005,850 between *M.m.domesticus* and *M.m.castaneus* in the latest *mm10* mouse genome release. In the context of prior studies in the lab, Dr. Tatyana Nesterova isolated an XY ES cell line from blastocysts resultant from the cross of a *M.m.domesticus* female and a *M.m.castaneus* male (unpublished). Sequence analysis of these cells demonstrated a homogeneous distribution of SNPs across the genome, with few exceptions (blue bins in Figure 3.1). An rtTA transactivator was targeted into the ROSA26 locus on chromosome 6 (in collaboration with Dr. Andrea Cerase), to subsequently allow the expression

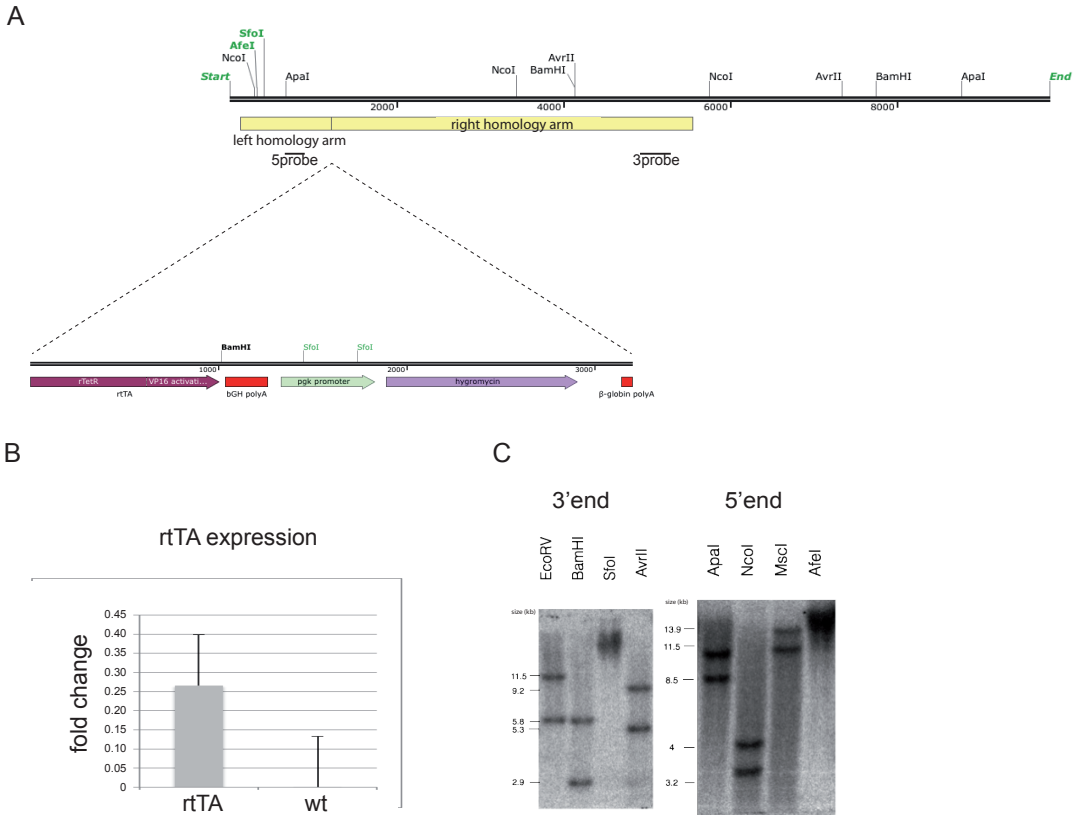


Figure 3.2: Generation of a doxycycline inducible mouse ES cell line. (A) An rtTA cassette was inserted into the ROSA26 locus on chromosome 6 by homologous recombination. Expression levels of the rtTA transgene were assessed by qPCR (B). Correct integration of the transgene was checked by Southern blot (C). Digestions are shown in the Figure, probes are depicted in (A).

of doxycycline-inducible transgenes. Correct integration of the rtTA cassette was assessed by Southern blot and qPCR (Fig.3.2).

3.2.2 Generation of an *in vivo* Xist deletion library

Mouse Xist cDNA was used as reference for wild-type, and a large panel of Xist deletions was generated, with a particular focus on exon I, which contains most of the segments that were previously shown to partially recapitulate Xist-mediated functions [Yen et al. 2007, Wutz et al. 2002]. All Xist variants were cloned under a pTre promoter, responsive to doxycycline treatment. An experimental pipeline was set-up in order to generate at least two independent mouse ES cell lines, harbouring each Xist variant in the panel. In detail, each construct was randomly integrated into the genome by lipofection in the presence of a selective marker. Lipofected

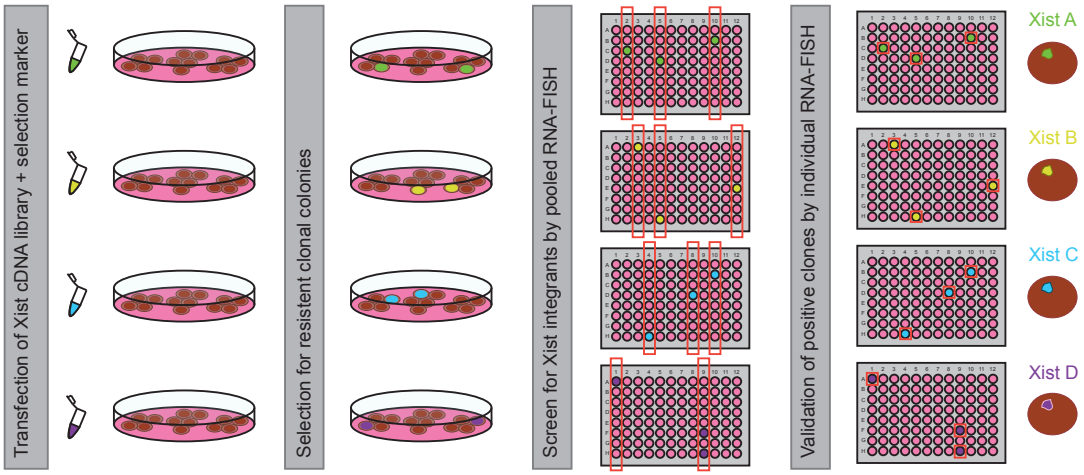


Figure 3.3: Generation of Xist transgene cell lines. Different Xist cDNA variants (here schematically represented in different colours) were lipofected in previously generated rtTA-expressing cells, in presence of a selection marker. After enrichment of selection-resistant clones, several were pooled together and tested for Xist expression by RNA-FISH. Individual clones were validated by RNA-FISH and RT-qPCR.

cells were put under selective pressure, and screened in pools for Xist expression upon 24 hours doxycycline treatment. Individual cell lines harbouring different Xist cDNAs were subsequently assessed for correct transgene expression and spreading using RNA-FISH, and only cell lines in which the proportion of individual cells expressing the Xist transgene was above 80% were used for further analysis (Fig.3.3).

The ability of different transgenes to form domains was tested by RNA-FISH (all tested Xist variants are listed in Figure 3.4). Xist domains were defined as localised clouds of RNA that were visibly larger than the punctuate Tsix endogenous *focus* stained in at least 80% of cells across three independent cell lines. All Xist variants that showed local enrichment upon doxycycline treatment, but failed to form RNA clouds distinct from Tsix in terms of size and intensity of the signal in any of the 50 cells analysed across at least 3 independent cell lines harbouring the same construct, were classified as non-localising. The observation was subsequently corroborated by pulled RNA-FISH analysis of all clones isolated for each non-localising construct. A list of observed phenotypes for each cell line is reported in Figure 3.4, and individual examples are shown in Figure 3.6. Real-time quantitative PCR was also employed

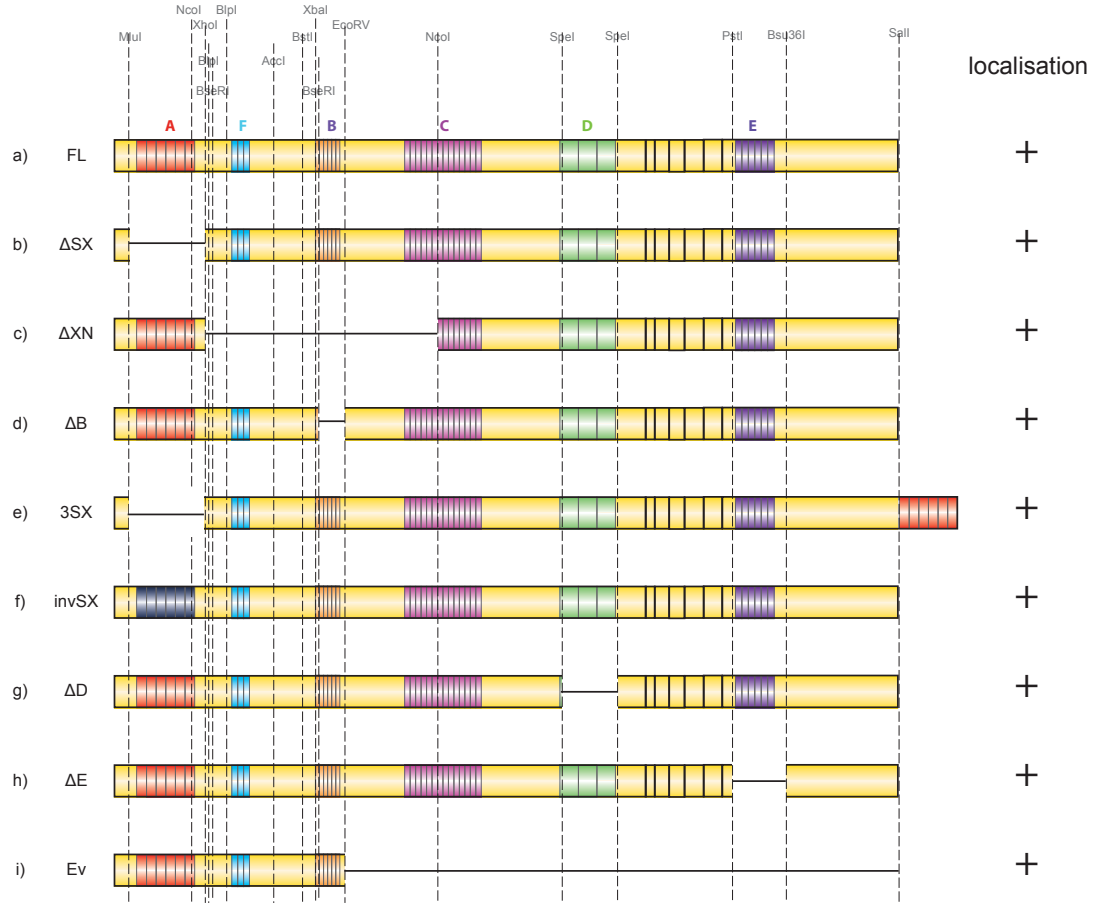


Figure 3.4: Xist cDNA deletion library. Schematic of Xist cDNA deletion series. Restriction sites used to generate deletions are indicated in the figure. Localisation was classified as +, if Xist clouds appeared in more than 80% of cells, in response to 24h doxycycline treatment.

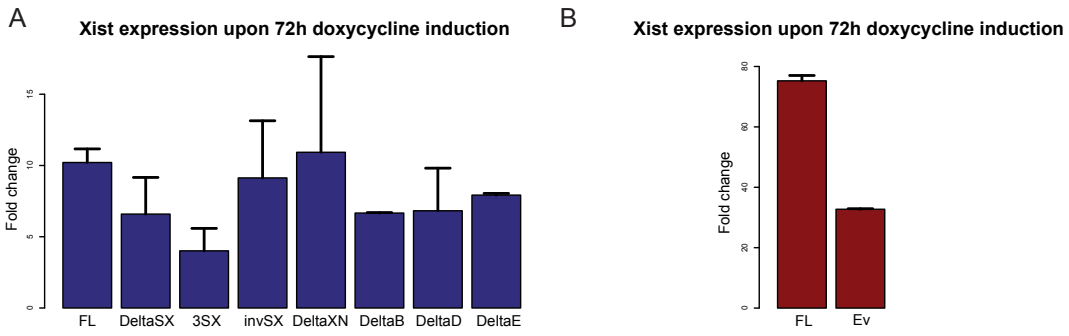


Figure 3.5: qPCR analysis of Xist expression. (A) and (B) represent Xist fold change upon doxycycline induction, measured using two separate sets of primers, qXist1endF+R (A) and DXN_F+R (B), described in Chapter 2, Section 2.12. Error bars represent standard deviation of technical replicates averaged between two biological replicates. Xist expression was normalised to GAPDH.

to test for Xist steady-state levels (Fig.3.5). These findings are broadly consistent with those reported by Wutz *et al* (2002).

3.2.3 Xist transgenes trigger early exclusion of RNA PolII and histone H4 hypoacetylation

I went on to characterise selected transgene lines. PolII exclusion from the Xi is an early event in X inactivation and has been considered one of its hallmarks. The ability of ectopically expressed Xist transgenes to mediate PolII exclusion was measured, following 24 hours of doxycycline treatment (Fig.3.7 A). The acetylation state of histone H4 following Xist forced expression was also tested (Fig.3.7 D), as it is generally regarded as a reliable cytogenetic marker for X inactivation [Jeppesen et al. 1993]. It should be noted that neither PolII exclusion nor lack of H4 hypoacetylation at the level of Xist domains can be considered as a proxy for Xist-mediated gene silencing because both markers are seen in the presence of gene-silencing deficient forms of Xist [Chaumeil 2006]. In order to better interpret the observed effect, two independent methods of quantification were applied. First, lines were drawn across individual Xist domains, and the intensity of either PolII or H4ac along the line was reported for a number of cells (Fig.3.7 A, B, D, E). This method of quantification has been employed in several studies aimed at evaluating PolII distribution across discrete nuclear domains, and particularly in the context of Xist-mediated transcriptional silencing [Clemson et al. 2006]. Using the line profile method, a large number of cells exhibited a dramatic loss of both PolII and H4ac in the core of Xist domains, as represented in the case shown in Figure 3.7 B and E. The effect was quantified in a more unbiased way by comparing the intensity of the PolII (or H4ac) signal in the Xist domain with that coming from another equivalent region of the nucleus (excluding the nucleolus). The results, shown in Figure 3.7 C and F suggest that FL Xist is indeed able to trigger transcriptional silencing as evidenced by exclusion of the transcriptional machinery from the inactive domain. For Ev-Xist (Fig.3.8 A, D), the line profile method also revealed an over-representation of cells with a clear depletion of PolII and H4ac (Fig.3.8 B, E). However, when the unbiased

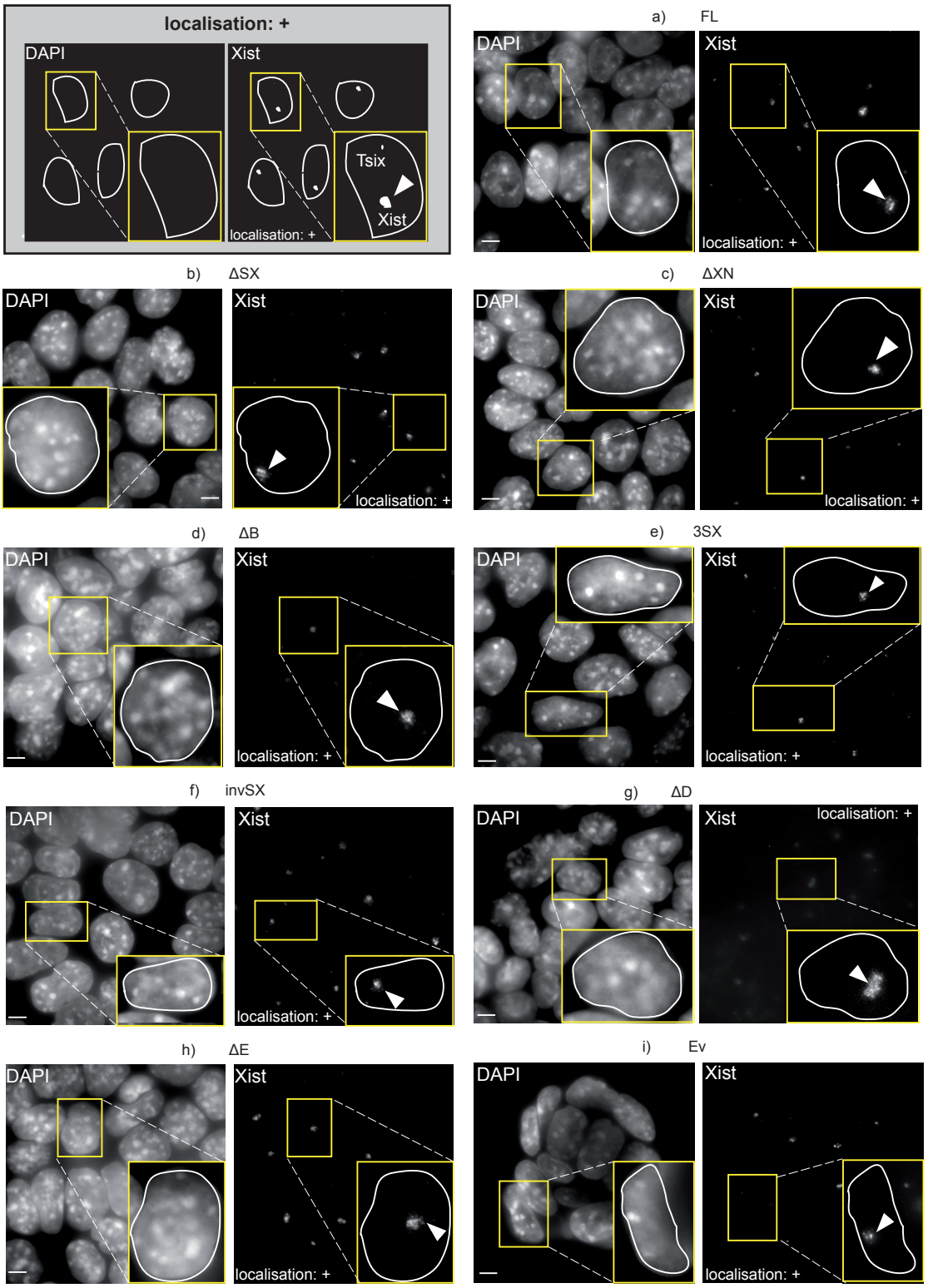


Figure 3.6: *In vivo* spreading of FL Xist deletion variants. Xist deletion variants are labelled as in Figure 3.4. The schematic (grey background) represents an example of Xist clouds classified as correctly localising (+). Cells were induced with doxycycline for 24 hours, and Xist was visualised by RNA-FISH. Wide-field images are representative examples of averages of 10 stacks. Enlarged images of single cells (in yellow boxes) are single stacks, drawn profiles of nuclei are based on DAPI signal. Arrows point to Xist domains. Scale bar is 5μm.

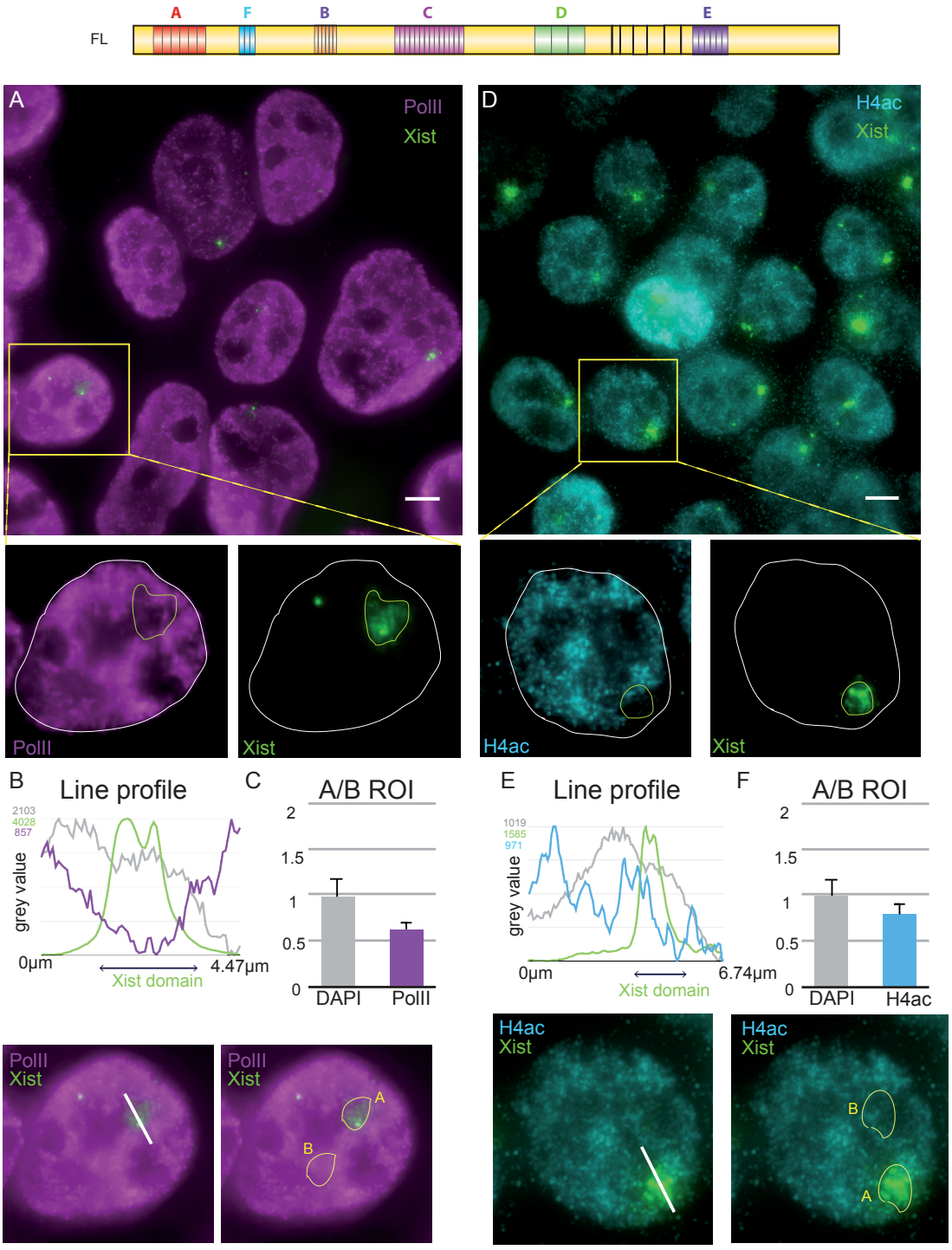


Figure 3.7: FL Xist determines PolII/H4ac holes. For cells in (A) and (D), representing PolII and H4ac immunofluorescence staining in combination with Xist RNA-FISH upon 24 hours doxycycline induction (channels are also separately shown below, with nuclei circled in white based on DAPI signal, and Xist domains in green), lines were placed across the Xist domain (B, E), and intensity values scaled to 100%. Matching plots are shown above. For each cell, ROI intensity within the Xist domain (C, F: yellow ROI A) was calculated as the average for each channel of the 3 slices with the brightest Xist signal. Values were compared with a matching ROI (C, F: yellow ROI B) outside the Xist domain. Intensity for the Xist channel (not shown) was always above 2. Quantification is shown above; average of 3 replicates, 10 cells per replicate. Error bars represent standard deviation. Scale bar for A, D is 5μm.

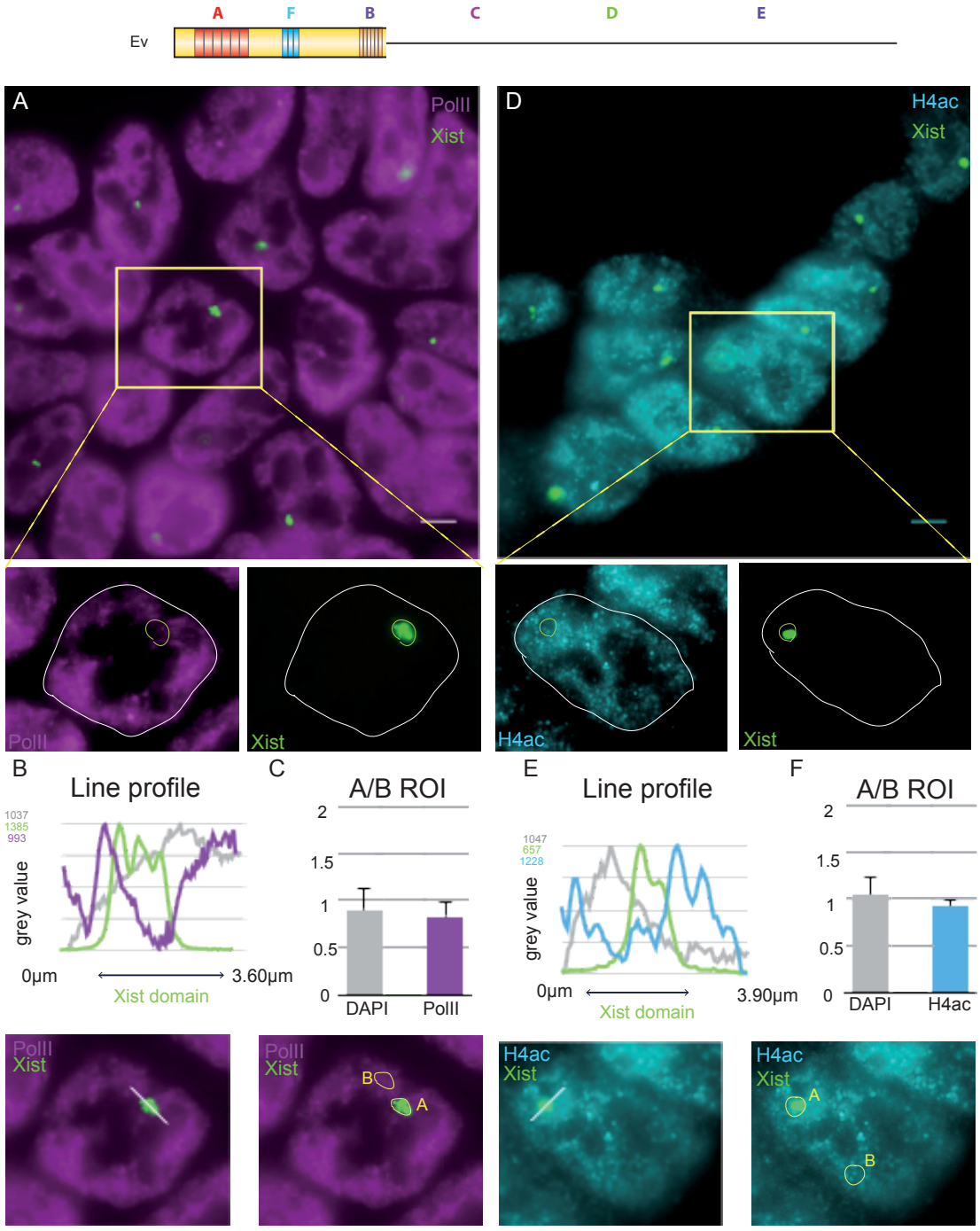


Figure 3.8: Ev-Xist determines PolII/H4ac holes. For cells in (A) and (D), representing PolII and H4ac immunofluorescence staining in combination with Xist RNA-FISH upon 24 hours doxycycline induction (channels are also separately shown below, with nuclei circled in white based on DAPI signal, and Xist domains in green), lines were placed across the Xist domain (B, E), and intensity values scaled to 100%. Matching plots are shown above. For each cell, ROI intensity within the Xist domain (C, F: yellow ROI A) was calculated as the average for each channel of the 3 slices with the brightest Xist signal. Values were compared with a matching ROI (C, F: yellow ROI B) outside the Xist domain. Intensity for the Xist channel (not shown) was always above 2. Quantification is shown above; average of 3 replicates, 10 cells per replicate. Error bars represent standard deviation. Scale bar for A, D is 5μm.

quantification method was applied, the reported effect was weaker compared to that observed for FL Xist (Fig.3.7 C, F). This result implies that the silencing ability of Ev-Xist may be reduced relative to FL Xist.

3.2.4 The A-repeats of Xist are required to trigger early exclusion of RNA PolII and histone H4 hypoacetylation

Xist-mediated effects on the early exclusion of the transcriptional machinery were also tested with the ΔSX Xist construct, lacking the A-repeats. Previously, PolII and H4ac holes were observed to overlap the Xi, following expression of a form of Xist lacking the A-repeats [Chaumeil 2006], despite the reported defect in transcriptional silencing linked to their absence [Wutz et al. 2002]. Loss of PolII and histone H4 hypoacetylation was here re-evaluated, using the same quantification methods described above (Fig.3.9 A, D). Although line profiles indicated that there was a small proportion of cells (∼10%) in which depletion of either PolII or H4ac could be reported, most cells showed little or no exclusion of these marks (Fig.3.9 B, E). The results were then quantified using signal intensities overlapping Xist domains. The result indicated that Xist lacking the A-repeats was not able to mediate significant exclusion of PolII and H4ac (Fig.3.9 C, F). In the light of this observation, early exclusion of the transcriptional machinery was confirmed to be somehow dependent on factors interacting with the A-repeats of Xist, in agreement with recent reports [McHugh et al. 2015].

3.2.5 Assessment of Xist-mediated gene silencing by 4sU-seq

Assessment of transcriptional activity based on *in situ* staining of RNA PolII and other cytological marks, has the clear limitation of providing an average view of overall transcription, without giving clues as to the transcriptional state and histone modification of individual genes or other genomic elements. In fact, measuring general transcription could be misleading, as local enhanced transcriptional activity of otherwise silent elements might reflect altered gene expression. In this sense, several studies reported how transcription from repetitive and non-coding regions

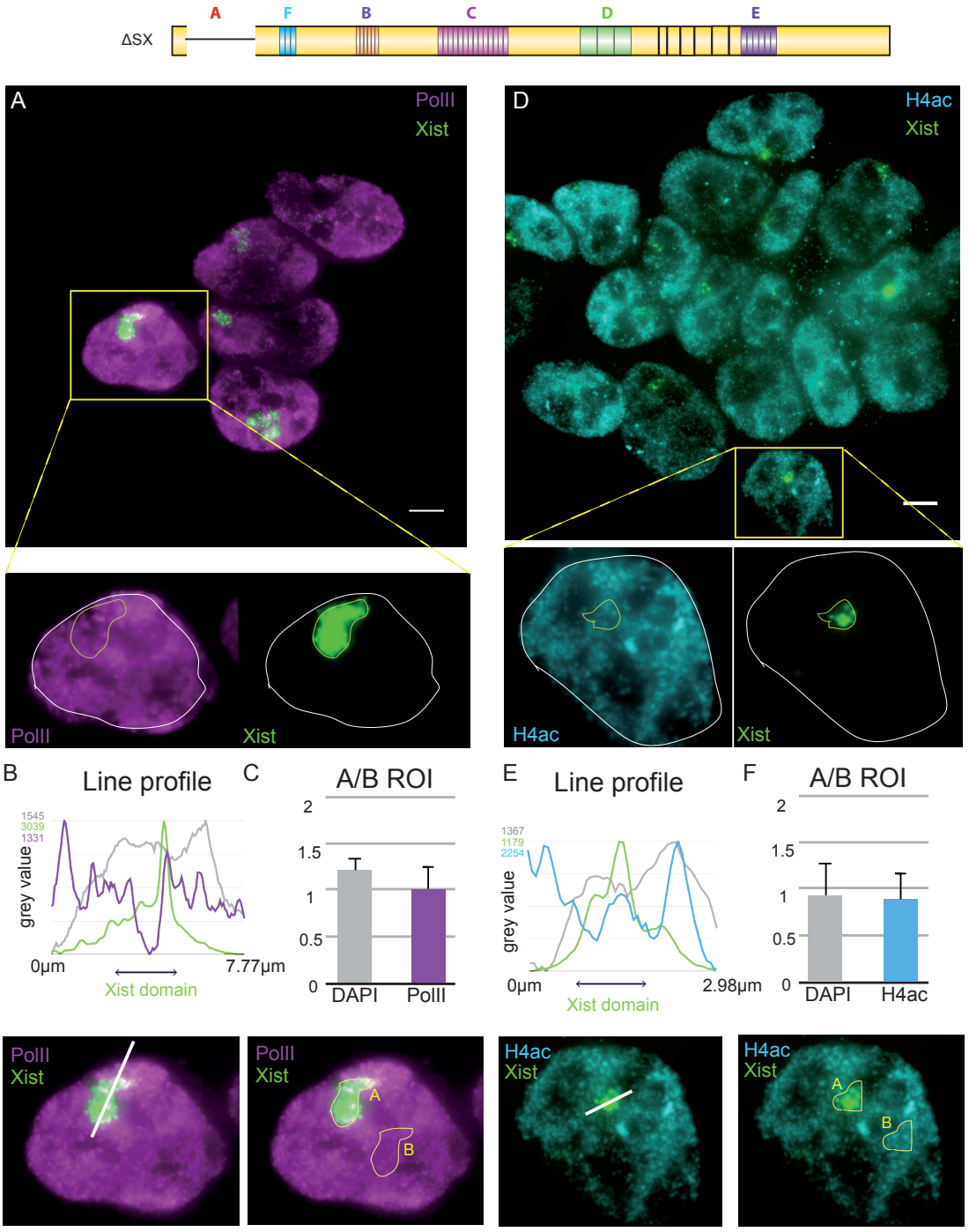


Figure 3.9: ΔSX Xist co-localises with PolII/H4ac. For cells in (A) and (D), representing PolII and H4ac immunofluorescence staining in combination with Xist RNA-FISH upon 24 hours doxycycline induction (channels are also separately shown below, with nuclei circled in white based on DAPI signal, and Xist domains in green), lines were placed across the Xist domain (B, E), and intensity values scaled to 100%. Matching plots are shown above. For each cell, ROI intensity within the Xist domain (C, F: yellow ROI A) was calculated as the average for each channel of the 3 slices with the brightest Xist signal. Values were compared with a matching ROI (C, F: yellow ROI B) outside the Xist domain. Intensity for the Xist channel (not shown) was always above 2. Quantification is shown above; average of 3 replicates, 10 cells per replicate. Error bars represent standard deviation. Scale bar for A, D is 5μm.

of the Xi might be enhanced upon Xist up-regulation [Hall et al. 2002]. In order to gain more detailed insights into the establishment of Xist-mediated gene silencing, a transcriptome-based analysis was employed. Given that Xist only acts *in cis*, silencing was measured by comparing gene expression from paired alleles upon Xist up-regulation.

This type of analysis requires unequivocal mapping of the sequencing reads to either allele, which was made possible by the inter-allelic SNPs carried by all cell lines employed in this study. In detail, 23,005,850 SNPs are annotated between the *Mus musculus castaneus* and *Mus musculus domesticus* strains. With the purpose of maximising SNP-containing reads, which cluster in introns (Fig.3.10 A), an RNA isolation method that enriches for nascent transcripts, 4sU-seq, was adopted [Rabani et al. 2011]. In 4sU-seq, metabolic labelling of RNA is used to capture nascent transcription. This technique makes use of 4-thiouridine (4sU), a naturally occurring modified uridine, to distinguish recently transcribed RNA from the total RNA population. The modified base is added to the growing media for a given time, and incorporated into the RNA chain in place of uridine. 4sU can be subsequently used to attach a biotin tag for easy separation of newly transcribed RNA from the total RNA population. In order to maximise the abundance of pre-processed RNA, cells were incubated with 4sU for 12 minutes. A 12 minutes 4sU-pulse was estimated to be the shortest compatible with isolation of enough material for good quality sequencing. This method provided an abundant fraction of pre-processed and un-spliced RNAs, as shown in Figure 3.10 B.

3.2.6 FL Xist is able to ectopically silence genes upon 24 hours induction in ES cells

First, it was important to test whether the transcriptional silencing of genes mediated by FL Xist could be detected in this experimental setup. In order to identify a timepoint that represents the highest level of Xist-mediated silencing, a time-course experiment was carried out, inducing Xist for 24h, 48h, and 72h

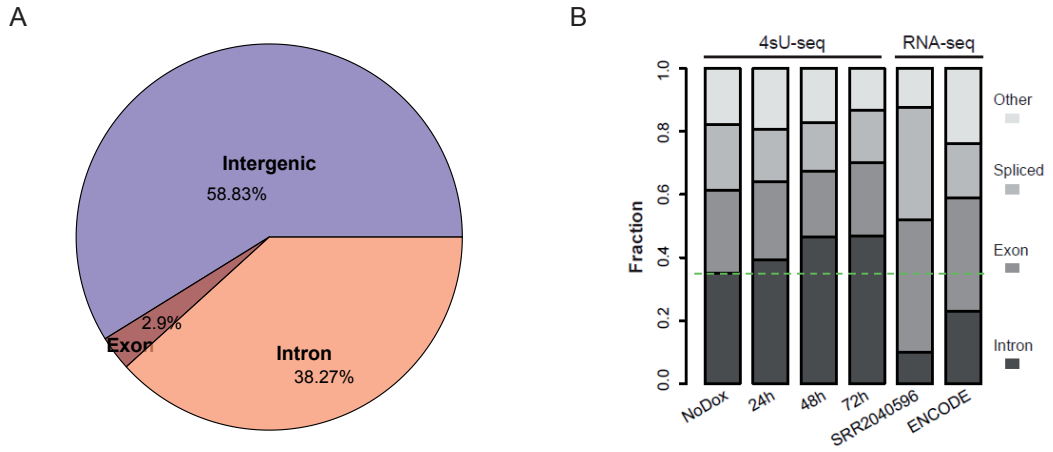


Figure 3.10: RNA fractions recovered by 4sU-seq. (A) Fraction of annotated interspecific SNPs between *Mus musculus castaneus* and *Mus musculus domesticus* strains, in Introns, Exons, and Intergenic regions. (B) Fraction of reads mapping to Introns, Exons, and Spliced transcripts for all four datasets of the time-course experiment: FL Xist untreated, FL Xist 24h dox, FL Xist 48h dox, FL Xist 72h dox. Public available RNA-seq data were also classified for direct comparison. Figure generated in collaboration with Dr. Guifeng Wei.

before isolating actively-transcribed RNA (Fig.3.13 A). In addition, since previous microarray studies showed that differentiation facilitated Xist-mediated silencing [Tang et al. 2010], cells were grown in LIF-withdrawal conditions.

4sU-labelled RNA was isolated, depleted from ribosomal RNA, and sequenced. Only reads containing annotated SNPs were kept for subsequent analysis. Informative SNPs were then aligned allele-specifically to the genome. As the cell line harbouring FL Xist was generated by random integration of the inducible transgene into the genome, the locus of integration was not known. It was therefore necessary to assess all pairs of alleles, in order to identify and eventually define a chromosome with a significant allele-specific transcriptional effect. Using this approach, the pericentromeric region of chromosome 6 was defined as the site of integration for FL Xist, as shown in Figure 3.11 A. This was confirmed by combining DNA-FISH for Xist transgene and DNA banding on metaphase spreads (Fig.3.11 B). The extent of silencing was quantified (in collaboration with Dr. Guifeng Wei) by assigning

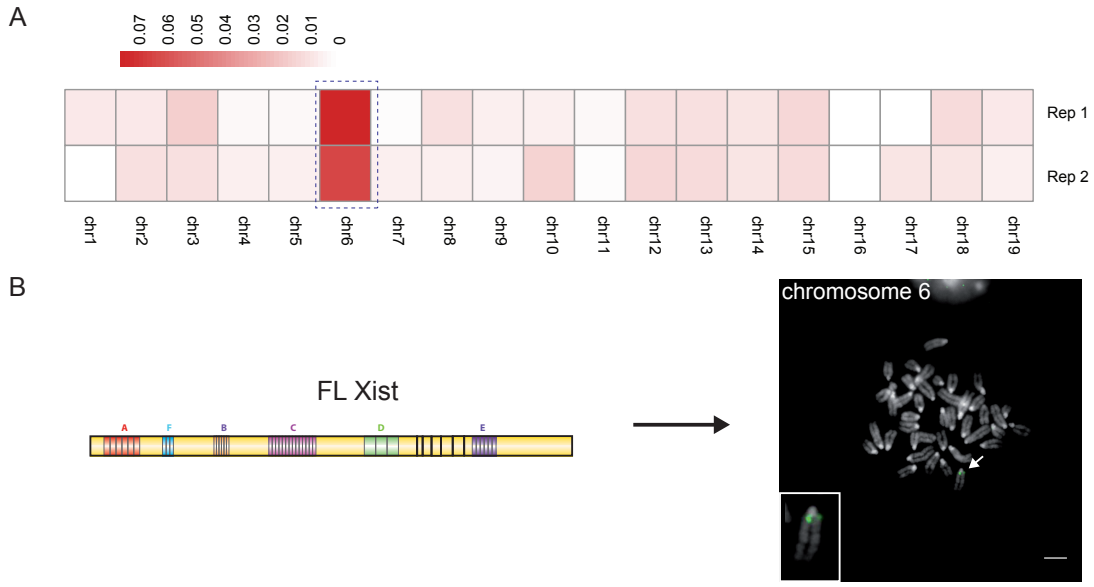


Figure 3.11: Integration site of FL Xist. (A) Significantly different levels of gene expression between alleles are plotted for each chromosome over 2 biological replicates. Each box represents a chromosome; overall allelic difference in gene expression is colour-coded from white (no difference), to red (mono-allelic expression). (B) DNA-FISH for Xist (green) was performed on metaphase spreads. Chromosomes were distinguished based on their banding. Figure generated in collaboration with Dr. Guifeng Wei.

a Repression Score (RS) for all genes, as follows:

$$RS = \frac{Y_i - Y_0}{Y_0} - \frac{X_i - X_0}{X_0} \tag{3.1}$$

where Y_i and X_i represent gene expression levels at each time point for the two alleles, and Y_0 and X_0 represent the level of transcription in the untreated (no doxycycline) sample, for the two alleles. A graphic representation of RS is provided in Figure 3.12. First, the 4sU-seq data were analysed to confirm that Xist expression during the time-course experiment was stable and doxycycline-responsive (Fig.3.13 A, B). Xist expression levels remained almost unchanged during the time-course, and were strictly doxycycline-dependent. Chromosome 6 was then split into windows spanning 10 Mbases across its length, and a cumulative RS was evaluated for each window (Figure 3.14 A). A higher level of silencing was reported around the site of integration of the transgene, after only 24 hours of transgene induction, with RS values decreasing progressively further away from the integration site (Figure 3.14 A). During the time-course, the silencing becomes more pronounced, peaking at

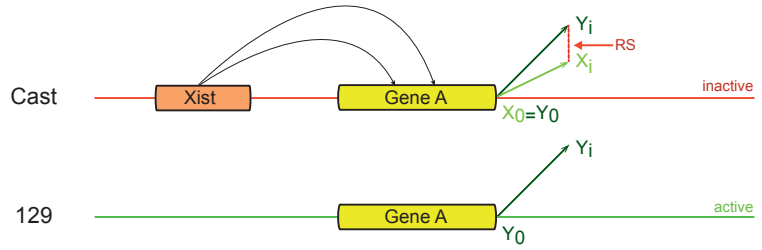


Figure 3.12: Xist Repression Score (RS). A graphic representation of the RS calculated for each gene. Levels of transcription are measured before and after doxycycline treatment and compared between alleles. If Xist interfered with transcription, a change in levels of transcription was only due to the reads assigned to the allele that expressed the transgene. Figure generated in collaboration with Dr. Guifeng Wei.

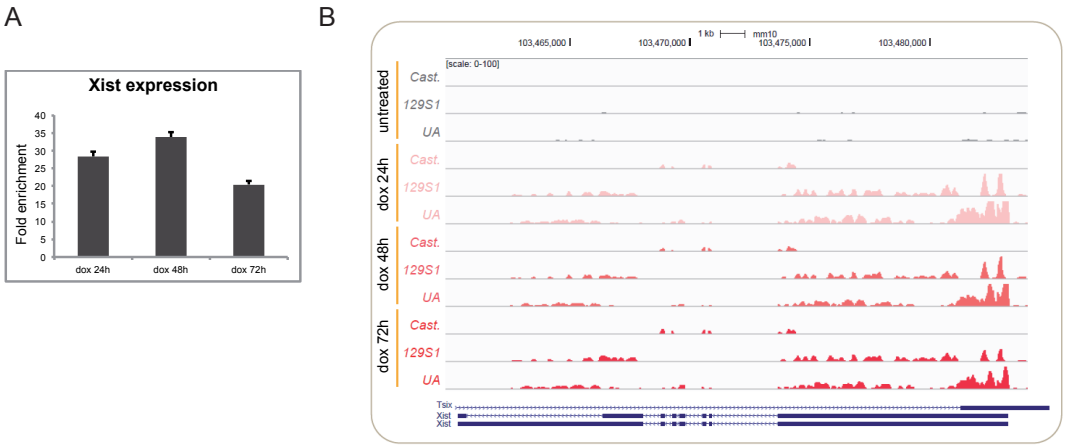


Figure 3.13: FL Xist time-course expression. Induction of Xist was forced by treating cells with doxycycline for 24h, 48h, and 72h. Levels of Xist expression were checked by real-time qPCR (A), and by analysing the reads mapping to Xist (B). In (B) a UCSC snapshot of the Xist locus is shown; cast, 129S1, and UA represent the castaneus, 129S1, and unassigned reads respectively. Figure generated in collaboration with Dr. Guifeng Wei.

72h. Interestingly, all the genes that were found to be silenced at 72h, were already partially or completely silenced after 24h, suggesting that the different silencing timing does not reflect an ordered preference but rather a stochastic process. An example of a gene affected by Xist expression is shown in Figure 3.14 B. Overall changes of RS across the time-course, and compared to other autosomes, are shown in Figure 3.15 A and B. However, at some sites transcriptional silencing fluctuated at the 48h time point, showing an apparent reversion compared to the 24h, then overcome again at 72h (Fig.3.15 C). With this in mind, the 72h time point was selected as the best representation of complete gene silencing.

3.2.7 The A-repeats are essential for gene silencing

Next, a cell line harbouring ΔSX Xist, lacking the A-repeats, was tested. Deletion of the A-repeats was shown in this study to abolish RNA PolII exclusion, and therefore expected to abolish, at least partially, Xist-mediated silencing. Indeed, when blind analysis of gene silencing was performed, the locus of integration could not be located, as only background fluctuations in transcriptional levels were registered across the whole genome (Fig.3.16 A). After *in situ* analysis, the site of integration was defined as being chromosome 2 (Fig.3.16 B), but despite further allele-specific analysis along that chromosome, no significant difference in terms of transcription from either allele could be detected, even in proximity of the integration site (Fig.3.16 C). In conclusion, ΔSX Xist fails to establish gene-silencing *in cis*.

3.2.8 Ev-Xist is able to mediate gene silencing

As discussed in Section 3.2.3, Ev-Xist, encompassing only part of Xist exon I, recapitulated some features of FL Xist. It was therefore interesting to test the ability of this construct to mediate gene silencing. To increase the confidence in measuring silencing, two independent cell lines of Ev-Xist were tested. Interestingly, in both cases the locus of integration could be easily defined, located (chromosome 9 and 11 respectively), and subsequently verified by DNA-FISH (Fig.3.17 and Fig.3.18 A, B). As shown in Figure 3.18 A, chromosome 14 could also be considered a candidate for

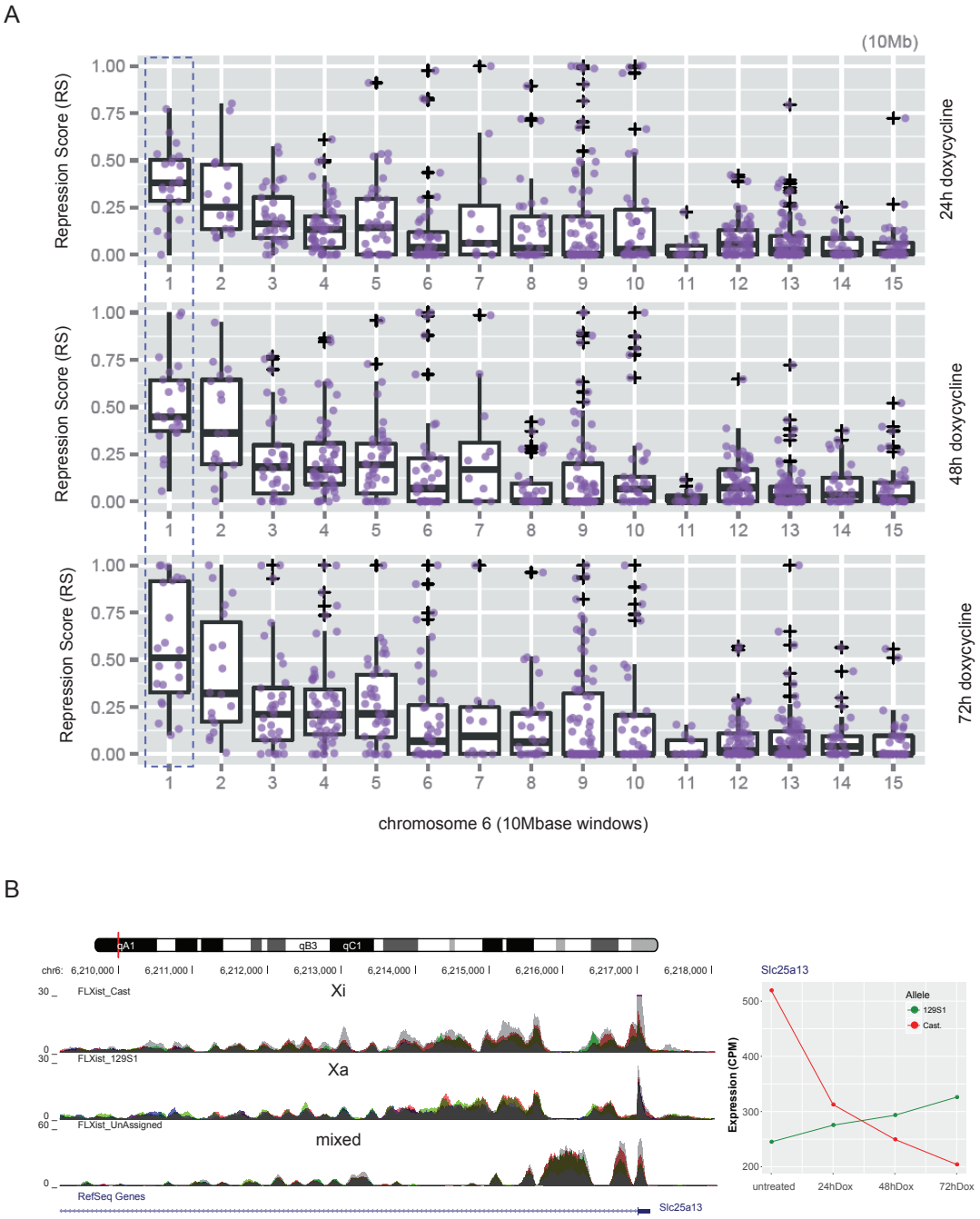


Figure 3.14: FL Xist-mediated gene silencing. (A) Box plots represent the median value of all Repression Scores for each 10Mbase window spanning the length of chromosome 6. RS is represented in purple for each gene within individual windows. The first window (boxed), where RS is the highest, encompasses the site of integration of the transgene. (B) UCSC snapshot comprising the Slc25a13 locus on chromosome 6; cast, 129S1, and UA represent the castaneus, 129S1, and unassigned reads respectively. All reads per each time point are plotted for each track: untreated: grey; 24h dox: red; 48h dox: green; 72h dox: blue. Overall expression levels are also plotted for the same gene. Figure generated in collaboration with Dr. Guifeng Wei.

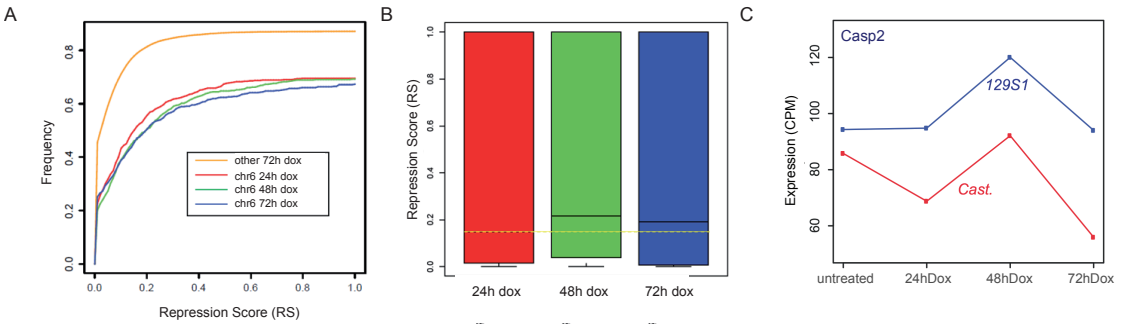


Figure 3.15: FL Xist RS time-course. (A) Lines represent the occurrence of each value of RS in each dataset. Values for all autosomes are also reported in orange. (B) Box plots represent median value of RS for each individual dataset along chromosome 6. (C) Overall expression levels are plotted for the Casp2 gene; cast and 129S1, represent the castaneus and 129S1 reads, respectively. Figure generated in collaboration with Dr. Guifeng Wei.

transgene integration, but further analysis of the chromosome, revealed randomness in the mono-allelic expression of genes located on that chromosome, distinct from what is shown for chromosome 11 (Fig.3.18 A, B).

Despite being in different parts of the chromosome, and likely different chromosome environments, both Ev-Xist constructs showed a similar behaviour in terms of gene-silencing, with an almost identical RS (0.2671 and 0.2133, respectively). These values correspond to over 50% of that calculated for FL Xist (Fig.3.17 and Fig.3.18 C). This result is of particular interest from two points of view; first, the fact that the RS for the same construct is comparable, shows the consistency of the employed method, as Xist behaves similarly despite its genomic location and context-dependent silencing efficiency; second, Ev-Xist was shown to be sufficient to mediate silencing, suggesting that the other parts of the RNA (not included in the Ev-Xist construct) might contribute to an increased and perhaps more stable maintenance of the silent state, but are not crucial for its establishment.

3.3 Discussion

Work described in this Chapter established a model to study the function of Xist transgenes. A deletion library of mouse Xist cDNA was used to generate independent

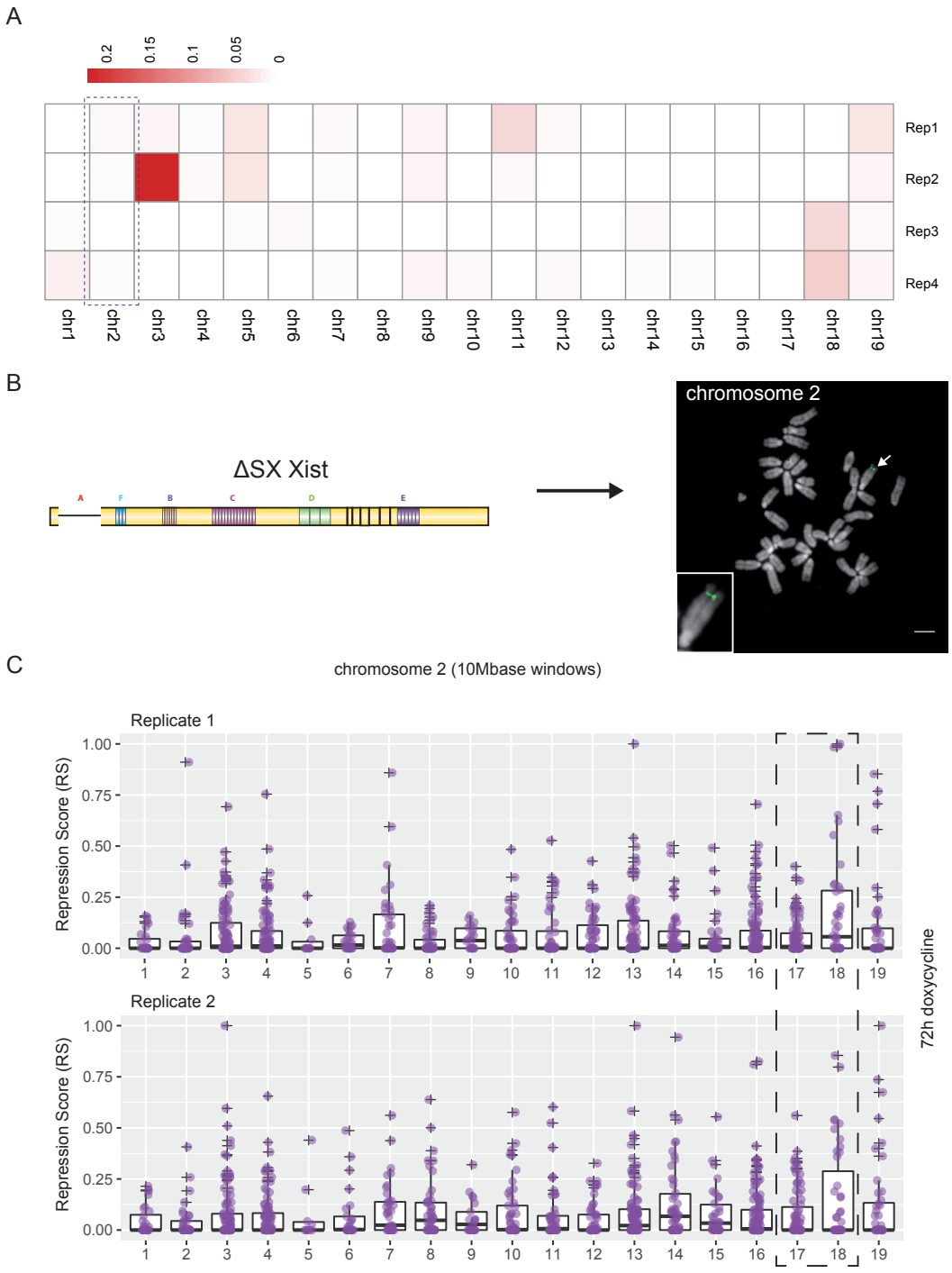


Figure 3.16: ΔSX Xist-mediated gene silencing. (A) Significantly different levels of gene expression between alleles are plotted for each chromosome over 4 biological replicates. Each box represents a chromosome; overall allelic difference in gene expression is colour-coded from white (no difference), to red (mono-allelic expression). No consistency could be observed among replicates, and mono-allelic expression is locally due to randomness. (B) DNA-FISH for Xist (green) was performed on metaphase spreads. Chromosomes were distinguished based on their banding, and the transgene located in the distal region of chromosome 2, corresponding in (C) to bins 17-19. (C) Box plots represent median value of RS for each 10Mbase window spanning the length of chromosome 2. RS is represented in purple for each gene within individual windows. Two independent biological replicates are represented, and a dotted box placed in correspondence of the windows surrounding the site of transgene integration. Figure generated in collaboration with Dr. Guifeng Wei.

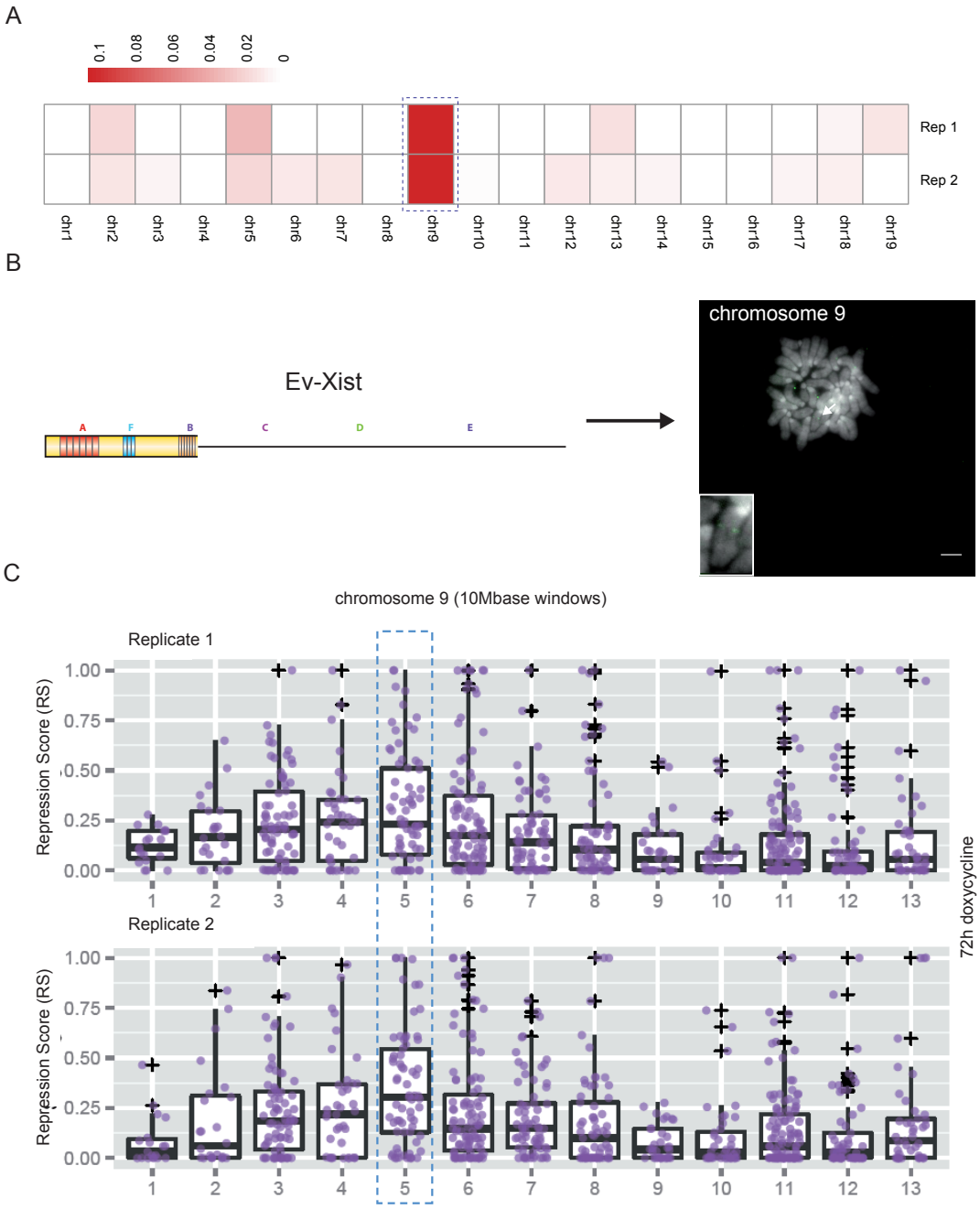


Figure 3.17: Ev-Xist-mediated gene silencing on chromosome 9. (A) Significantly different levels of gene expression between alleles are plotted for each chromosome over 2 biological replicates. Each box represents a chromosome; overall allelic difference in gene expression is colour-coded from white (no difference), to red (mono-allelic expression). (B) DNA-FISH for Xist (green) was performed on metaphase spreads. Chromosomes were distinguished based on their banding. (C) Box plots represent RS median value for each 10Mbase window spanning the length of chromosome 9. RS is represented in purple for each gene within individual windows. The fifth window (boxed), where RS is the highest, encompasses the site of integration of the transgene. Two independent biological replicates are represented. Figure generated in collaboration with Dr. Guifeng Wei.

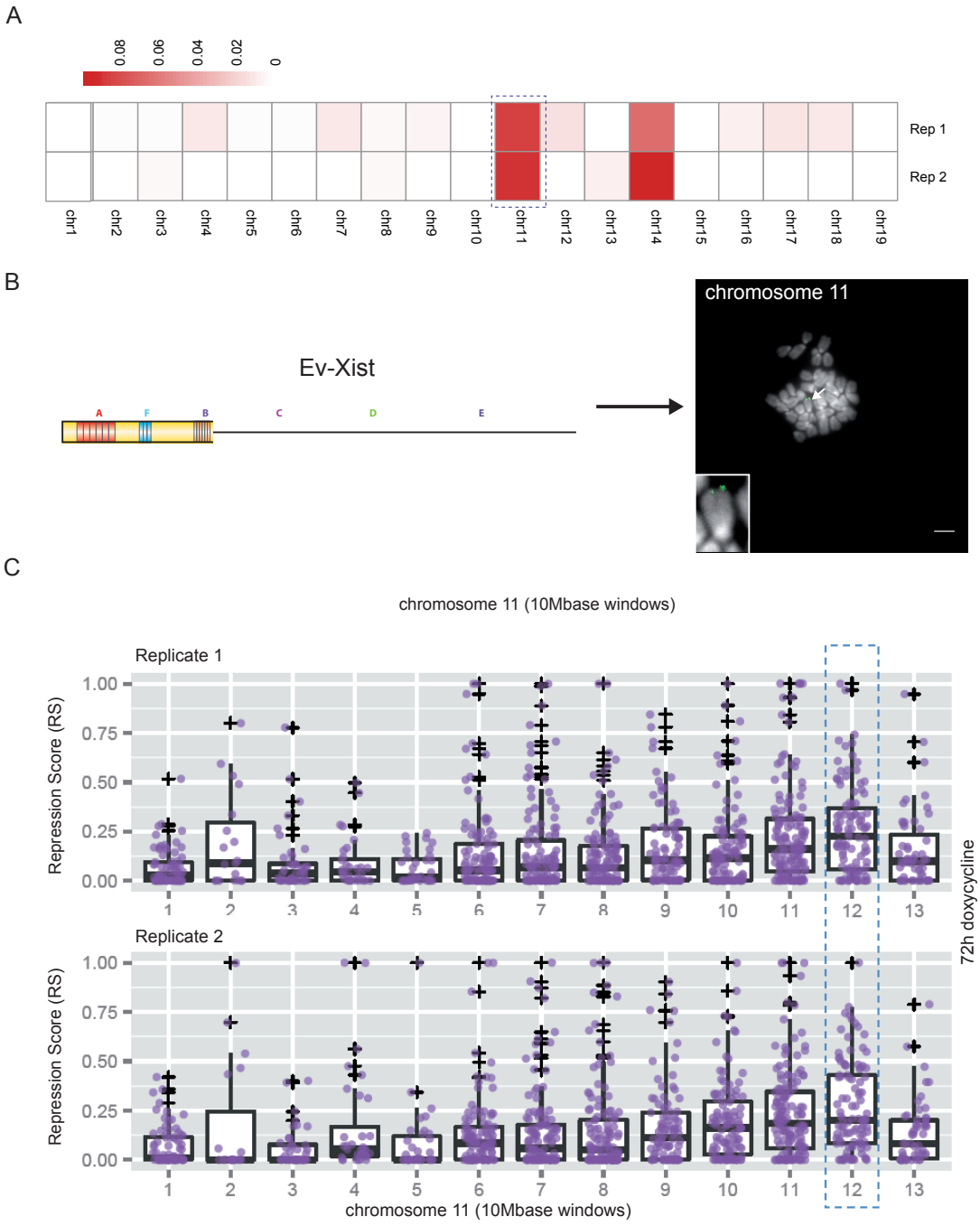


Figure 3.18: Ev-Xist-mediated gene silencing on chromosome 11. (A) Significantly different levels of gene expression between alleles are plotted for each chromosome over 2 biological replicates. Each box represents a chromosome; overall allelic difference in gene expression is colour-coded from white (no difference), to red (mono-allelic expression). (B) DNA-FISH for Xist (green) was performed on metaphase spreads. Chromosomes were distinguished based on their banding. (C) Box plots represent RS median value for each 10Mbase window spanning the length of chromosome 11. RS is represented in purple for each gene within individual windows. The twelfth window, where RS is the highest, encompasses the site of integration of the transgene. Two independent biological replicates are represented. Figure generated in collaboration with Dr. Guifeng Wei.

male ES cell lines in which Xist variants were randomly integrated into the genome, and whose expression could be induced upon doxycycline treatment. They were subsequently analysed for their ability to silence genes. An obvious limitation of this experimental approach is that random integration of Xist constructs, gives rise to variables that cannot be controlled, notably the locus of integration, and the copy number of the transgene. Nevertheless, the method provides a fast readout to assess simple phenotypes, such as Xist domain formation, protein complexes recruitment, and gene silencing. In fact, despite not being ideal for direct comparison of independent cell lines, analysis of Ev-Xist at different sites provided consistent phenotypes. Further refinement of the approach based on controlled integrations, is described in Chapter 6, Section 6.2.

The model system described was used to assess the silencing capability of three Xist deletion variants (FL Xist, ΔSX Xist and Ev-Xist), previously tested by Wutz *et al* (2002). This allowed a proof-of-principle for the reliability of the experimental setup, and also a re-evaluation of previously-published data, using a more sensitive and direct analysis. A quantitative measure of allele-specific transcriptional efficiency was used to precisely measure the degree of silencing achieved by each Xist deletion variant. This approach gives a better idea of subtle transcriptional effects, possibly overlooked using cruder assays, that either focus on small sets of genes or on indirect effects, that could be influenced by compensatory mechanisms.

Assessment of RNA PolII and H4 hypoacetylated domains indicated a significant loss of transcriptional activity overlapping Xist clouds, except in the case of ΔSX Xist, lacking the A-repeats. This suggests that the A-repeats are crucial to establish early hallmarks of the inactivation process. However, the presence of the transcriptional machinery does not shed light on the sites of transcription that are affected. In fact, several reports showed that during the inactivation process, transcription of repetitive elements might take place, to the detriment of gene expression [Chow et al. 2010, Hall et al. 2002]. Moreover, conflicting evidence

is reported in the literature regarding the ability of ΔSX Xist to repress overall transcription on autosomes [Chaumeil 2006, Chow et al. 2010, Engreitz et al. 2013], suggesting that cytogenetic markers may be differently affected by Xist activity depending on the genomic features of its locus of integration. In this sense, a direct measure of gene silencing following Xist expression is a more consistent and reliable readout of Xist activity.

A key innovation in this work is the development of a method to consistently measure Xist-mediated gene silencing in an ectopic location. Because SNPs are found in the majority of genes, only a small number of genes were excluded from the RS analysis. First, the silencing ability of FL Xist was evaluated during a time-course experiment. Silencing was observed to become progressively more pronounced, with its peak at 72 hours within the analysed time window. Importantly, all the genes that were found to be repressed at 72 hours, were partially or completely silenced after 24 hours, during the early phase of the inactivation process, suggesting that gene silencing might be concomitant to the earliest phases of epigenetic modification of the Xi.

After defining the 72 hours of doxycycline induction as the time-point in which efficient FL Xist-dependent gene silencing occurs, the silencing ability of the A-repeats deletion was analysed. The ΔSX Xist construct was found to be unable to establish gene silencing, even at low levels. This is in agreement with previous studies, showing that the presence and proper folding of the A-repeats are important for overall Xist-mediated silencing [Wutz et al. 2002, Engreitz et al. 2013, da Rocha et al. 2014]. However, until recently, it was not clear whether the observed defect in silencing upon deletion of the A-repeats was due to a general destabilisation of Xist RNA, possibly linked to its inefficient processing, or could rather be ascribed to loss of specific binding partners. The fact that highly specific binding modes were observed for some A-repeats interacting factors [Lu et al. 2016], and that some of these also contact general RNA biogenesis machineries [Moindrot et al. 2015], seem

to suggest that both hypothesis could be true and several overlapping pathways may contribute to the A-repeat-mediated function of Xist. In the future, it will be important to clarify the molecular mechanism of A-repeat mediated gene-silencing. On the other hand, the idea that it depended on the interaction with Polycomb [Zhao et al. 2008], was recently disproved, in favour of a different region of Xist [da Rocha et al. 2014]. The interaction between Xist and Polycomb will be further addressed in Chapter 5.

Finally, it should be noted that the fundamental role of the A-repeats in Xist-mediated gene silencing, does not rule out the potential contribution from other parts of Xist. To address this, the Ev-Xist construct was employed. The ability of this construct to spread *in cis* meant that it was possible to quantify its RS. The results of the RS analysis of Ev-Xist, clearly showed its ability to silence genes, albeit at lower levels compared to FL Xist. The reduced levels of silencing can explain why expression of this construct did not cause cell lethality in the study of Wutz *et al* (2002). Definition of a minimal silencing construct, enabled further analysis of functional elements in Xist exon I, as reported in Chapter 4.

4

Xist localisation and domain formation

Contents

4.1 Introduction 78
4.2 Results 79
4.2.1 An element in Ev-Xist required for *in cis* localisation . 79
4.3 Discussion 81

4.1 Introduction

A distinctive feature of Xist is its ability to spread onto the chromosome from which it is expressed, forming domains that can be visualised by RNA-FISH [Brown et al. 1992, Beletskii et al. 2001]. In a classical genetic study, Xist spreading ability was assessed and compared between different deletions [Wutz et al. 2002]. Three separate regions were identified as being redundantly involved in domain formation, even in the absence of a 5' element (Fig.4.1). However, a more recent paper identified the F- and C-repeat regions of Xist as crucial for proper localisation of the RNA on the inactive X chromosome, their targeting with LNA being sufficient to dramatically reduce Xist clouds [Sarma et al. 2010, Jeon et al. 2011]. The authors followed up on this result, proposing a localisation mechanism based on dual binding of the factor YY1 to the C-repeats of Xist RNA and to the F-repeats of

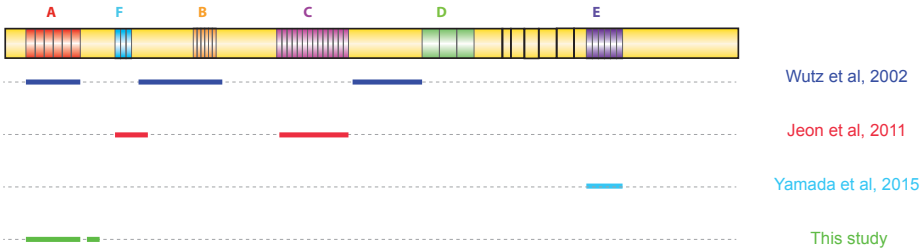


Figure 4.1: Xist localisation sequences. Schematic of the localisation sequences identified in previous studies. Results from the present study are also included for comparison (green).

Xist DNA [Jeon et al. 2011]. Another contrasting piece of evidence comes from the observation that a deletion of exon 7 (the E-repeats region) affects Xist localisation, in a manner that the previous studies failed to detect [Yamada et al. 2015]. The authors explained their result by relating the deleted region to the presence of binding sites for hnRNP-U/Saf-A, a nuclear matrix protein known to directly bind Xist and crucial for its accumulation on the Xi [Hasegawa et al. 2010]. In Chapter 3, an experimental system was set up with the purpose of dissecting Xist-mediated function *in vivo*. High-throughput sequencing analysis of the transcriptome upon Xist expression, revealed how a short form of Xist, named Ev-Xist, retains most of the silencing ability of FL Xist, despite lacking exons II-VII, and part of exon I. In agreement with previous studies, Ev-Xist was also shown to form apparently normal Xist clouds [Wutz et al. 2002].

Given the current debate regarding the contribution of different parts of Xist RNA to *in cis* spreading (Fig. 4.1), Ev-Xist was further deleted to test whether specific segments are key for mediating proper Xist cloud formation *in vivo*.

4.2 Results

4.2.1 An element in Ev-Xist required for *in cis* localisation

As discussed in Chapter 3, all tested deletions in the context of FL Xist, and also Ev-Xist, did not show any obvious defect in forming RNA clouds following

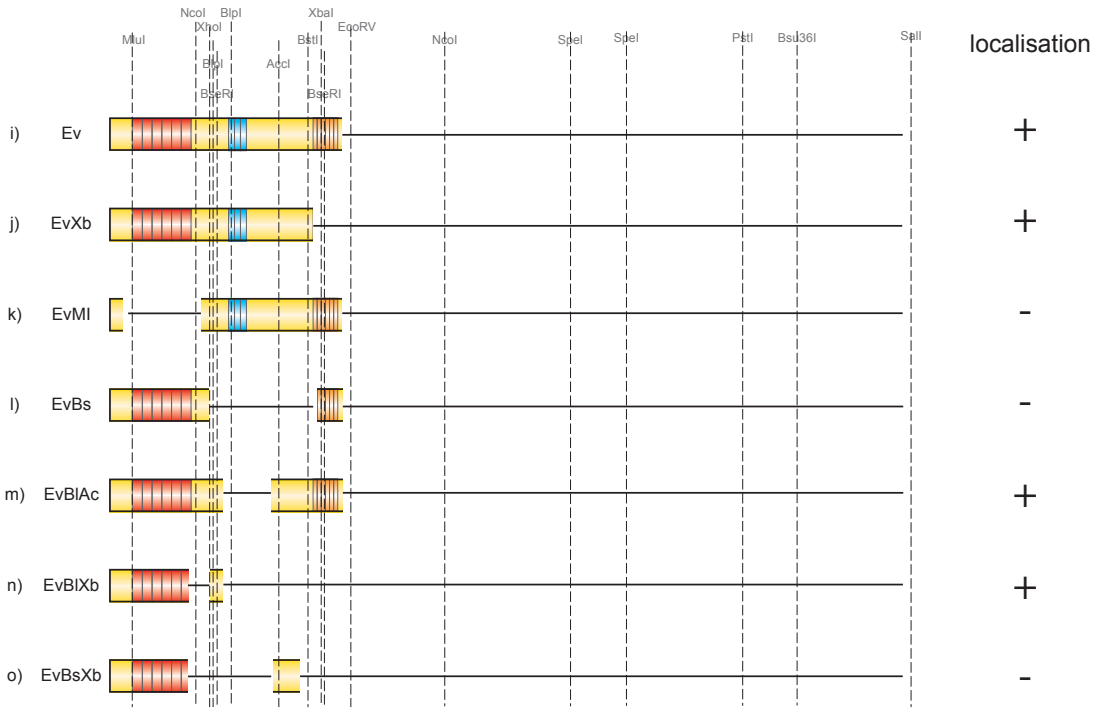


Figure 4.2: Ev-Xist cDNA deletion library. Schematic of Ev-Xist cDNA deletion series. Restriction sites used to generate the deletions are indicated in the figure. Localisation was classified as +, if Xist clouds appeared in more than 20% of cells across all biological replicates, in response to 24h doxycycline induction. If no cells with Xist domains could be observed in response to 24h doxycycline induction, the construct was defined as non-localising (-).

doxycycline induction. Further dissection of the Ev-Xist region was performed using the constructs illustrated in Figure 4.2. qPCR analysis of the integrated cell lines was also performed (Fig.4.3) Deletion of the B-repeats did not affect Xist domain formation (Fig.4.4, construct j). In contrast, the A-repeats were essential for proper Xist spreading in this context (Fig.4.4, construct k).

Interestingly, an element between the A- and B-repeats was also required to mediate proper localisation of Ev-Xist (Fig.4.4, construct l). This segment encompasses one of the YY1 binding sites, previously implicated in Xist localisation [Jeon et al. 2011]. To test this more definitively, the YY1 binding region was specifically deleted. Surprisingly, Xist domains were observed upon induction of several deletion variants spanning the putative YY1 binding site (one example is shown in Fig.4.4, construct m). On the other hand, comparison of several deletion

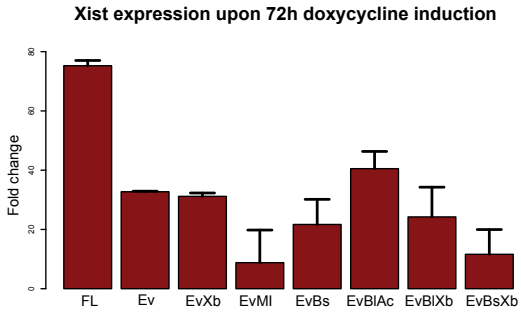


Figure 4.3: qPCR analysis of Ev-Xist expression. Xist fold change upon doxycycline induction, was measured using DXN_F+R primers, described in Chapter 2, Section 2.12. Error bars represent standard deviation of technical replicates averaged between two biological replicates. Xist expression was normalised to GAPDH.

variants, defined a short segment downstream of the A-repeats as responsible for complete loss of Xist spreading. To test the contribution of this segment to Xist ability to spread onto the inactive chromosome, it was fused to a truncated form of Ev-Xist comprising only the A-repeats (Fig.4.4, construct n). The experiment was controlled in parallel by linking another sequence of the same length (taken from just downstream of the putative localisation sequence), to the truncated Ev-Xist (Fig.4.4, construct o). Interestingly, the identified localisation sequence was indeed specifically required for generating proper Xist clouds as determined by RNA-FISH.

4.3 Discussion

The ability of different Ev-Xist constructs to form visible clouds upon forced transgene up-regulation was assessed by RNA-FISH. In this context, two regions of the RNA appeared to be crucial for its correct localisation: the A-repeats and a 188bp sequence located immediately downstream of the A-repeats, defined by the restriction sites BseRI and BlpI. In the absence of other functionally redundant elements located downstream of the C-repeats, loss of either of these sequences abrogated the ability of Xist to form functional domains (Fig.4.5).

For the constructs in which localisation was deficient, accumulation of Xist RNA at the transcription locus was observed in response to doxycycline treatment, but

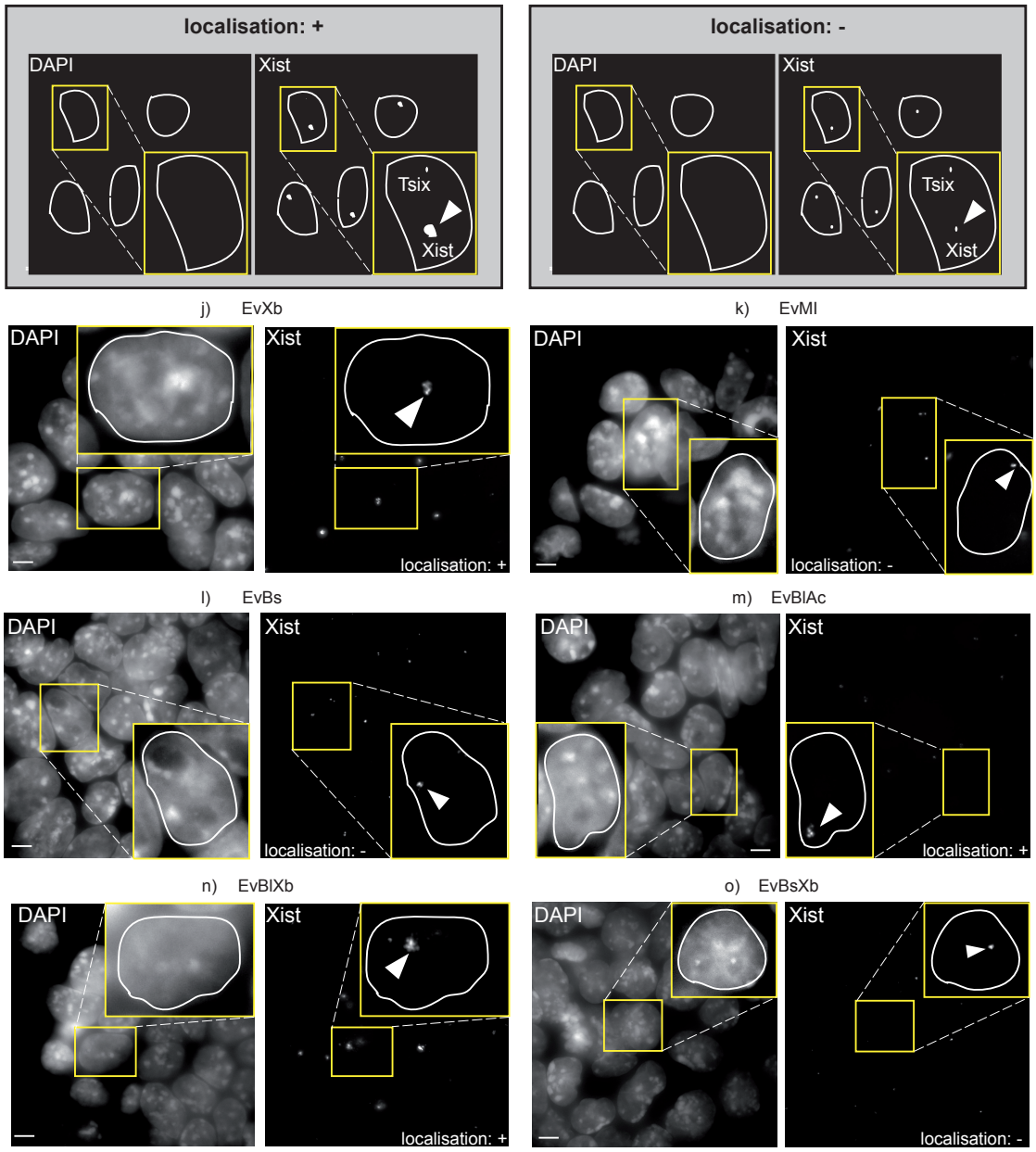


Figure 4.4: *In vivo* spreading of Ev-Xist deletion variants. Ev-Xist was obtained truncating FL Xist at the first EcoRV restriction site. All Ev-Xist deletion variants are named as in Figure 4.2. The schematics (grey background) represent examples of Xist clouds classified as correctly localising (+) or non-localising (-). Cells were induced with doxycycline for 24 hours, and Xist was visualised by RNA-FISH. Wide-field images represent average of 10 stacks. Enlarged images of single cells (in yellow boxes) are single stacks, drawn profiles of nuclei are based on DAPI signal. Arrows point to Xist domains. Scale bar is 5μm.

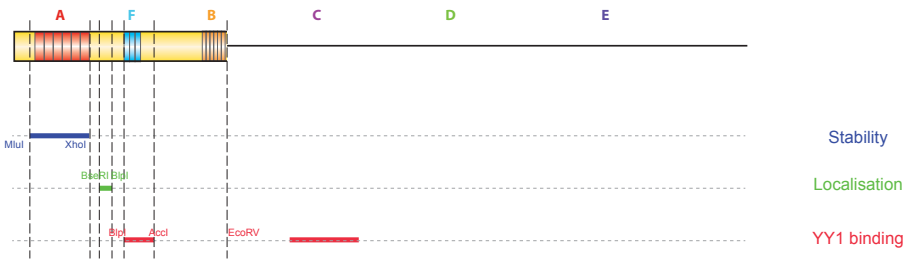


Figure 4.5: Ev-Xist localisation sequences. Schematics of the localisation sequences identified in the context of Ev-Xist, and their potential function. YY1 binding sites are also reported according to [Jeon et al. 2011].

expression did not result in domain formation. At this stage it is not known if these sequences affect transcript localisation *per se* or RNA stability. However, the relatively high levels of the EvBs transcript that were detected by qPCR (Fig.4.3), as well as its visible accumulation observed by RNA-FISH (Fig.4.4, construct l), seem to point to a defect in the spreading ability of the RNA. Conversely, the EvMl construct (Fig.4.4, construct k), appears to transcribe RNA that is less stable (Fig.4.3). Based on this, it is tempting to speculate that the A-repeats, known to fold into a complex secondary structure [Wutz et al. 2002, Zarnegar et al. 2016], act to stabilise the RNA, while the BseRI-BlpI segment is a spreading signal. Previous studies, did not functionally characterise this segment, however Wutz *et al* (2002), reported poor localisation for an Xist variant (Xist deletion 18), that despite the presence of the A-repeats and most of Xist central exons, lacked among other regions, also the BseRI-BlpI region. Interestingly, Hoki *et al* (2011), observed a defect in Xist function following insertion of a short segment in a site of Xist adjacent to the BseRI-BlpI region. It could be that deletion of the BseRI-BlpI region leads to destabilisation of the site identified by Hoki *et al* (2011), or conversely, that the insertion tested by Hoki *et al* (2011), interfered with the novel spreading signal.

This novel localisation element may function either as a binding site for specific factors, or harbour a structure that facilitates Xist domain formation. Both hypotheses appear equally plausible for a number of reasons. First, although this segment does not contain any known protein binding motif, sequence conservation

between human and mouse is higher compared to the A- and and F-repeats, and most of the rest of Xist RNA [Nesterova et al. 2011]. Second, recent studies aimed at elucidating RNA folds *in vivo*, found the same region as being highly structured, at least in human, although no direct experiments have been conducted to validate the presence of a well-defined secondary fold potentially involving the novel localisation sequence [Spitale et al. 2015, Lu et al. 2016]. Additionally, this region was identified as being enriched for m^6A methylation sites [Patil et al. 2016], therefore perturbation of its sequence/fold might lead to interference with the m^6A machinery.

In the future it will be interesting to test whether the combined deletion of the A-repeat and the BserI-BlpI region could partially overcome the redundancy observed in the FL Xist context, allowing further investigation of the functional contribution of other localising elements.

5

Mechanisms of Xist-mediated Polycomb recruitment by Xist RNA

Contents

5.1 Introduction 85
5.2 Results 86
5.2.1 The XN region of Xist is required to recruit both PRC1 and PRC2 86
5.2.2 The B-repeats of Xist are required for PRC1 and PRC2 recruitment 87
5.2.3 The B-repeats are essential for PRC1 and PRC2 recruitment by Ev-Xist 91
5.2.4 Dissection of the B-repeat function in Polycomb recruitment 94
5.3 Discussion 94

5.1 Introduction

Several different Polycomb complexes, encompassing PRC2 [Wang et al. 2001, Silva et al. 2003, Plath et al. 2003], canonical PRC1 [Napoles et al. 2004], and variant PRC1 [Schoeftner et al. 2006, Tavares et al. 2012], have been implicated in the X inactivation process. They accumulate on the Xi rapidly in response to Xist RNA, and ongoing recruitment requires continued Xist expression [Plath et al. 2003, Silva et al. 2003, Napoles et al. 2004]. These observations suggest a direct

relationship between Xist RNA and Polycomb recruitment. A classical model to explain this, proposes that RepA, a short transcript arising from the Xist promoter, recruits the PRC2 core-component Ezh2 [Zhao et al. 2008]. Observations made in the context of this study mainly rely on RNA immuno-precipitation experiments (RIP) in which RepA was found to co-IP with several PRC2 components (mostly Ezh2 and Suz12), and EMSA assays in which Ezh2 was identified as the most plausible candidate for mediating the RepA-PRC2 interaction. Further related work, also showed that phosphorylation of Ezh2 enhanced its binding to Xist RNA [Kaneko et al. 2010]. However, da Rocha *et al* (2014), found that the A-repeat region is dispensable for PRC2 recruitment, and that this function instead maps to an element defined by the XN deletion of Xist RNA [da Rocha et al. 2014]. Indeed, studies from ourselves and others, have challenged the model for PRC2 interaction with Xist RNA [Kohlmaier et al. 2004, Tavares et al. 2012]. Moreover, as outlined in Chapter 1, Section 1.4, the Brockdorff lab has determined that a PRC1 complex, Pcgf3/5-PRC1, initiates Polycomb recruitment by Xist RNA [Almeida et al. 2016].

Experiments in this section are aimed at defining the region of Xist RNA required to recruit Pcgf3/5-PRC1 therefore initiating Polycomb *de novo* recruitment on the Xi.

5.2 Results

5.2.1 The XN region of Xist is required to recruit both PRC1 and PRC2

Given the published evidence of a potential involvement of either the A-repeats or the region of Xist defined by the XN deletion in Polycomb recruitment, ΔSX and ΔXN Xist were both analysed. First, it was confirmed that the experimental system was suitable for assessing *de novo* Polycomb recruitment on the Xi. Three independent cell lines harbouring FL Xist were analysed by RNA-FISH in combination with immunofluorescence, to assess co-localisation between Xist clouds and Polycomb proteins, after 24h doxycycline induction. Staining of the

histone modifications H2AK119ub1 and H3K27me3 was used as a surrogate for PRC1 and PRC2 recruitment, respectively. Examples are shown in Figure 5.1 A and Figure 5.1 B, and the data for one of the cell lines are quantified in Figure 5.4 A. Following Xist expression, high co-localisation of Xist clouds and both H2AK119ub1 (Fig.5.1 A) and H3K27me3 was observed (Fig.5.1 B), indicating Xist-mediated recruitment of catalytically active PRC1 and PRC2 complexes on the Xi. The same experiment was therefore performed for ΔSX Xist, lacking the A-repeats. In agreement with previously published results [da Rocha et al. 2014], no significant reduction of H3K27me3 could be observed after 24 hours of Xist expression (Fig.5.2 B). Importantly, PRC1 levels on the inactive chromosome were also unaffected (Fig.5.2 A). ΔXN Xist, which formed normal Xist clouds, failed to establish H3K27me3 domains (Fig.5.3 B), in agreement with previously published evidence [da Rocha et al. 2014]. Moreover, there was also a complete absence of PRC1-mediated H2AK119ub1 (Fig.5.3 A). In conclusion, the A-repeats of Xist are not required for either PRC1 or PRC2 recruitment on the Xi, which instead depends on the region defined by the XN deletion. The observed phenotype was highly reproducible among different cell lines and different times of transgene induction (data not shown), and is quantified in Figure 5.4 B and C. The following section sets out to identify the key element(s) within this region of Xist.

5.2.2 The B-repeats of Xist are required for PRC1 and PRC2 recruitment

Within the XN deletion (∼4kb), the B-repeats represented a good candidate for mediating PRC1/PRC2 recruitment, as this element is highly conserved in different species [Nesterova et al. 2011]. To test this, the ΔB Xist construct was stably transfected into ES cells, and H2AK119ub1 and H3K27me3 were scored following induction of the transgene for 24 hours. Neither H2AK119ub1 (Fig.5.5 A), nor H3K27me3 (Fig.5.5 B) domains could be observed in the absence of the B-repeats. Absence of H2AK119ub1 and H3K27me3 suggests that neither PRC1 nor PRC2 are recruited in the absence of the B-repeats. Moreover, formally it is possible

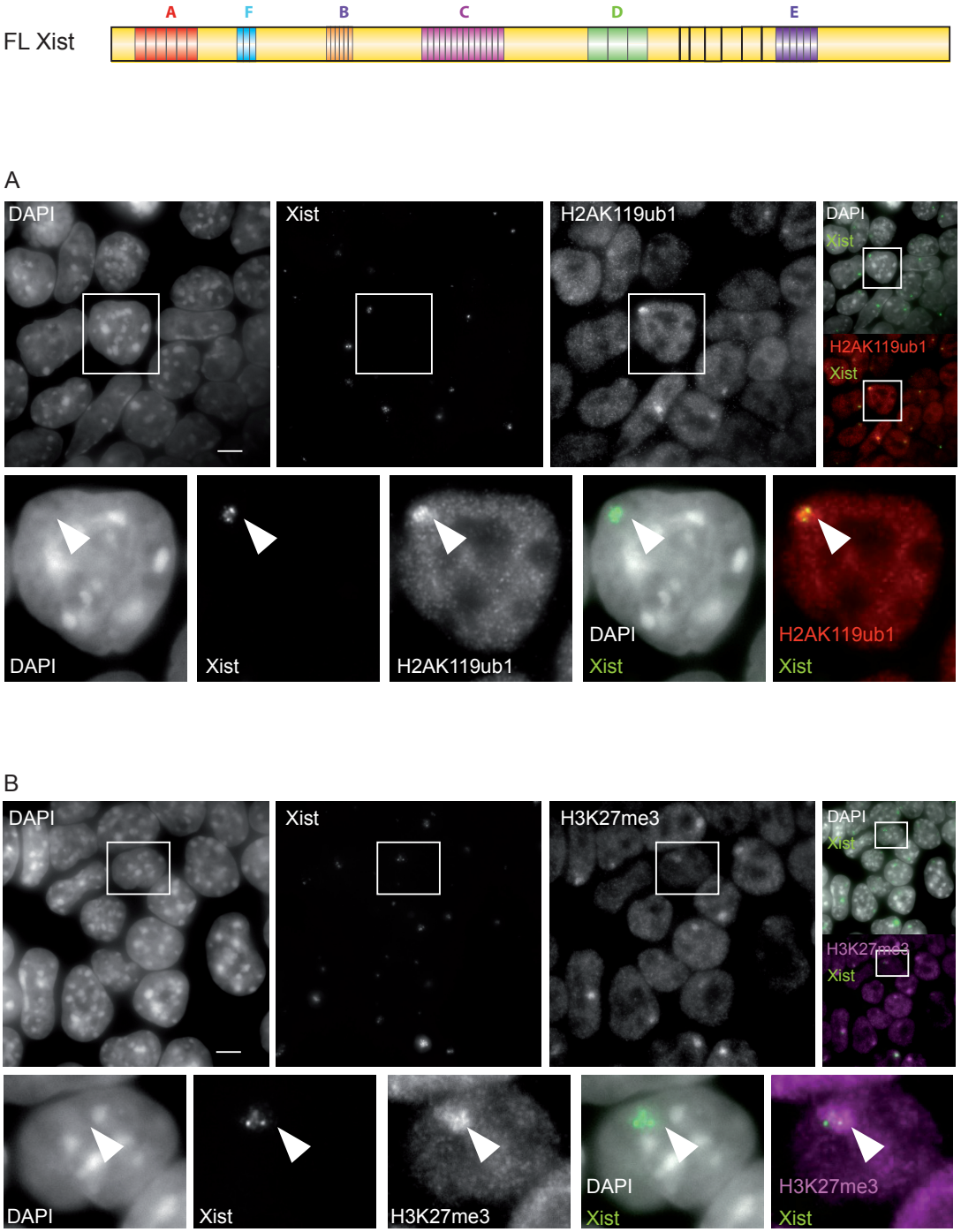


Figure 5.1: PRC1 and PRC2 recruitment by FL Xist RNA. FL Xist, shown above, was detected by RNA-FISH, after inducing cells with doxycycline for 24 hours. Cells were co-stained by immunofluorescence for H2AK119ub1 (A) and H3K27me3 (B). Wide-field images represent average of 10 stacks. Enlarged images of cells are single stacks. Arrows indicate Xist domains. Scale bar is 5μm.

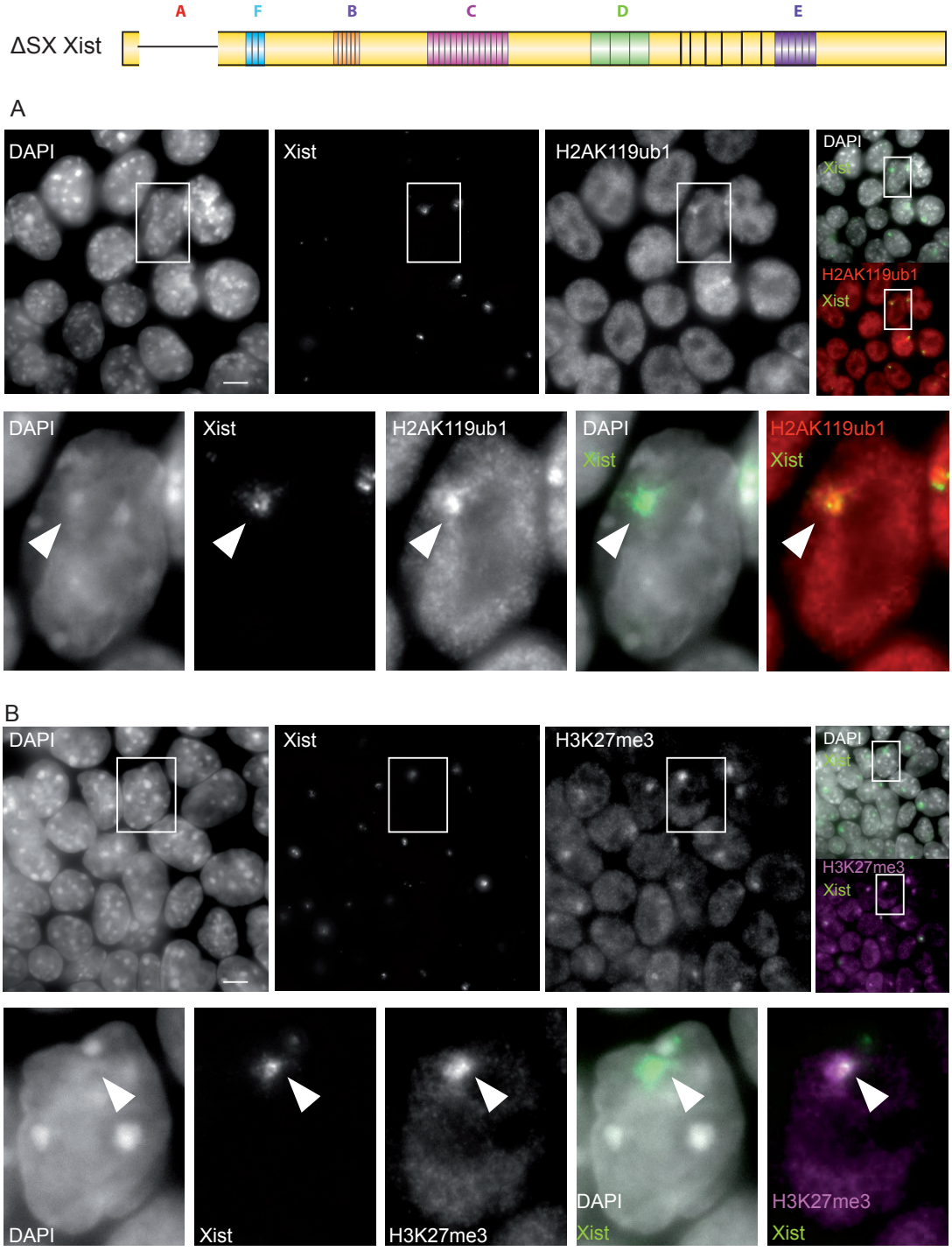


Figure 5.2: PRC1 and PRC2 recruitment by ΔSX Xist RNA. ΔSX Xist, shown above, was detected by RNA-FISH, after inducing cells with doxycycline for 24 hours. Cells were co-stained by immunofluorescence for H2AK119ub1 (A) and H3K27me3 (B). Wide-field images represent average of 10 stacks. Enlarged images of cells are single stacks. Arrows indicate Xist domains. Scale bar is 5μm.

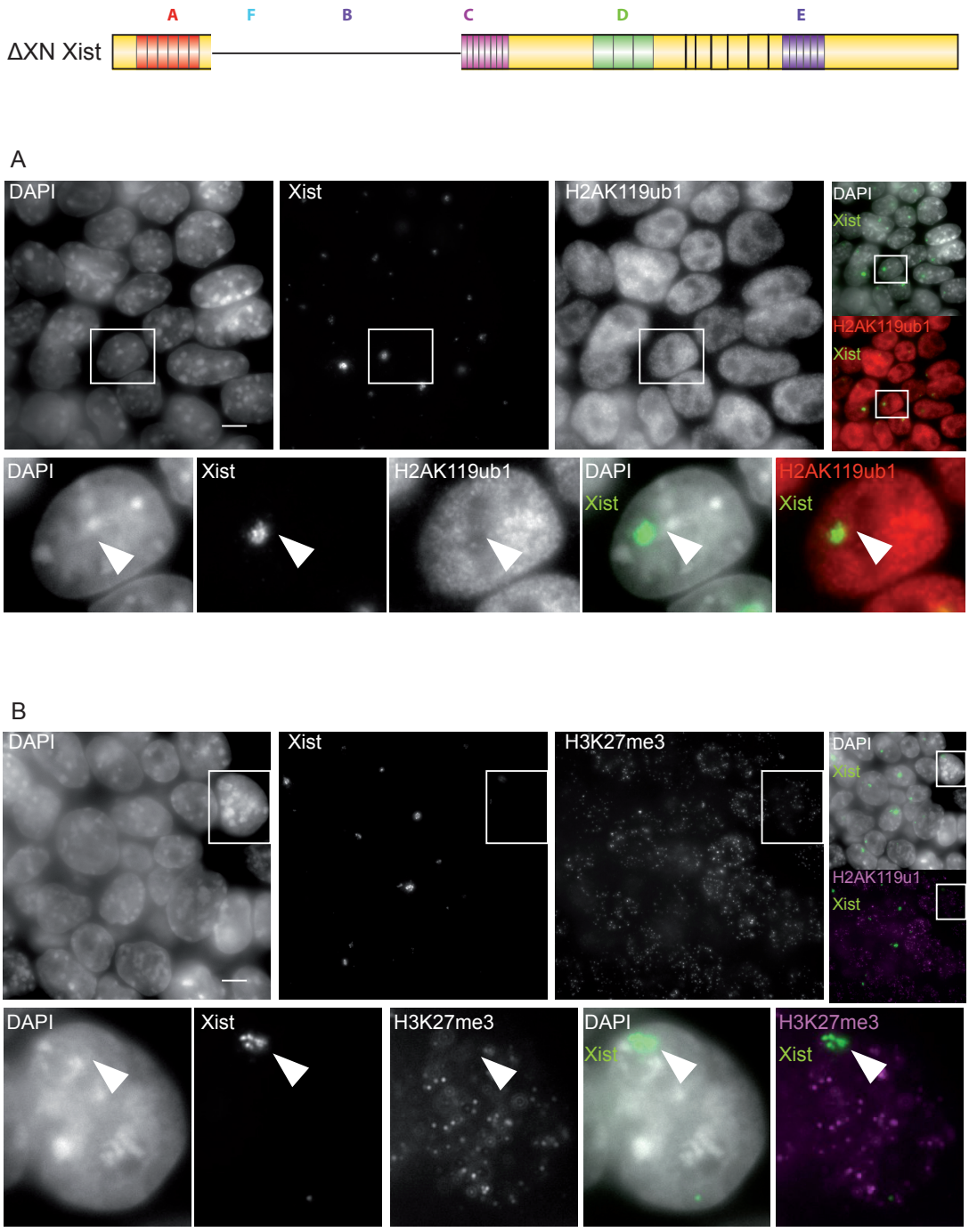


Figure 5.3: PRC1 and PRC2 recruitment by ΔXN Xist RNA. ΔXN Xist, shown above, was detected by RNA-FISH, after inducing cells with doxycycline for 24 hours. Cells were co-stained by immunofluorescence for H2AK119ub1 (A) and H3K27me3 (B). Wide-field images represent average of 10 stacks. Enlarged images of cells are single stacks. Arrows indicate Xist domains. Scale bar is 5μm.

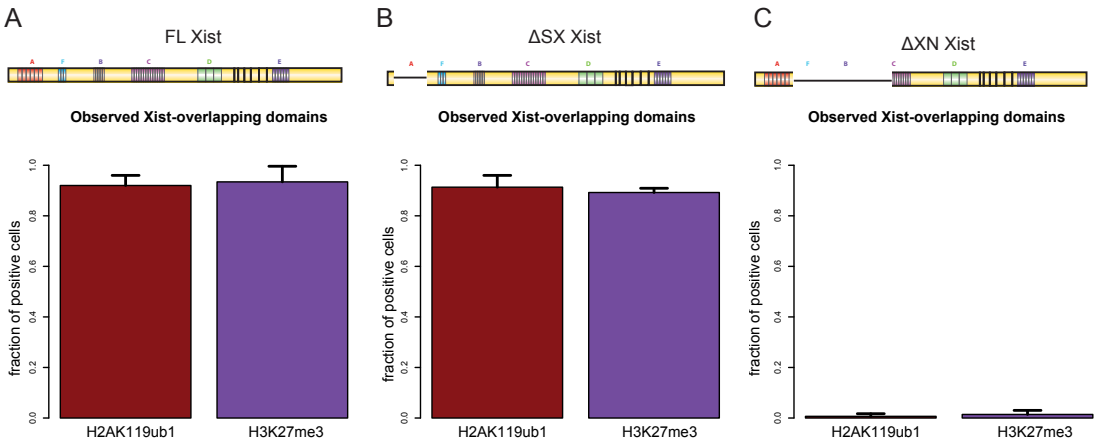


Figure 5.4: Quantification of co-localising domains. Combined RNA-FISH/IF experiments were quantified by counting H2AK119ub1 (red bars) or H3K27me3 (purple bars) that co-localise with Xist domains 24 hours after doxycycline induction. Results are shown for FL (A), ΔSX (B), and ΔXN (C) Xist. Scale bars represent sd over 3 biological replicates. Over 150 cells were counted for each replicate.

that activity of the complexes, rather than recruitment is affected. To test this, immunofluorescence assays were performed for several Polycomb proteins upon Xist induction. A cell line harbouring FL Xist provided a positive control. For FL Xist, well-defined domains of enrichment were observed for all tested Polycomb subunits (Fig.5.6 A). However, no domains could be observed in the ΔB-repeats Xist cell line (Fig.5.6 B). In conclusion, the B-repeats are crucial for Polycomb *de novo* recruitment on the Xi.

5.2.3 The B-repeats are essential for PRC1 and PRC2 recruitment by Ev-Xist

To further define Xist sequences required for Polycomb recruitment, the short Ev-Xist construct, which includes the B-repeats (Fig.4.2, constructs i), was analysed. First, the ability of Ev-Xist to recruit PRC1/PRC2 was analysed in the experimental setup previously described, and compared with that of EvXb-Xist (Fig.4.2, constructs j), the matching construct missing the B-repeats. After 24 hours of transgene induction, Xist and Polycomb-mediated histone modifications were co-stained. The results perfectly mirrored the outcome of the experiment carried out using the matching FL Xist constructs, confirming the crucial role of

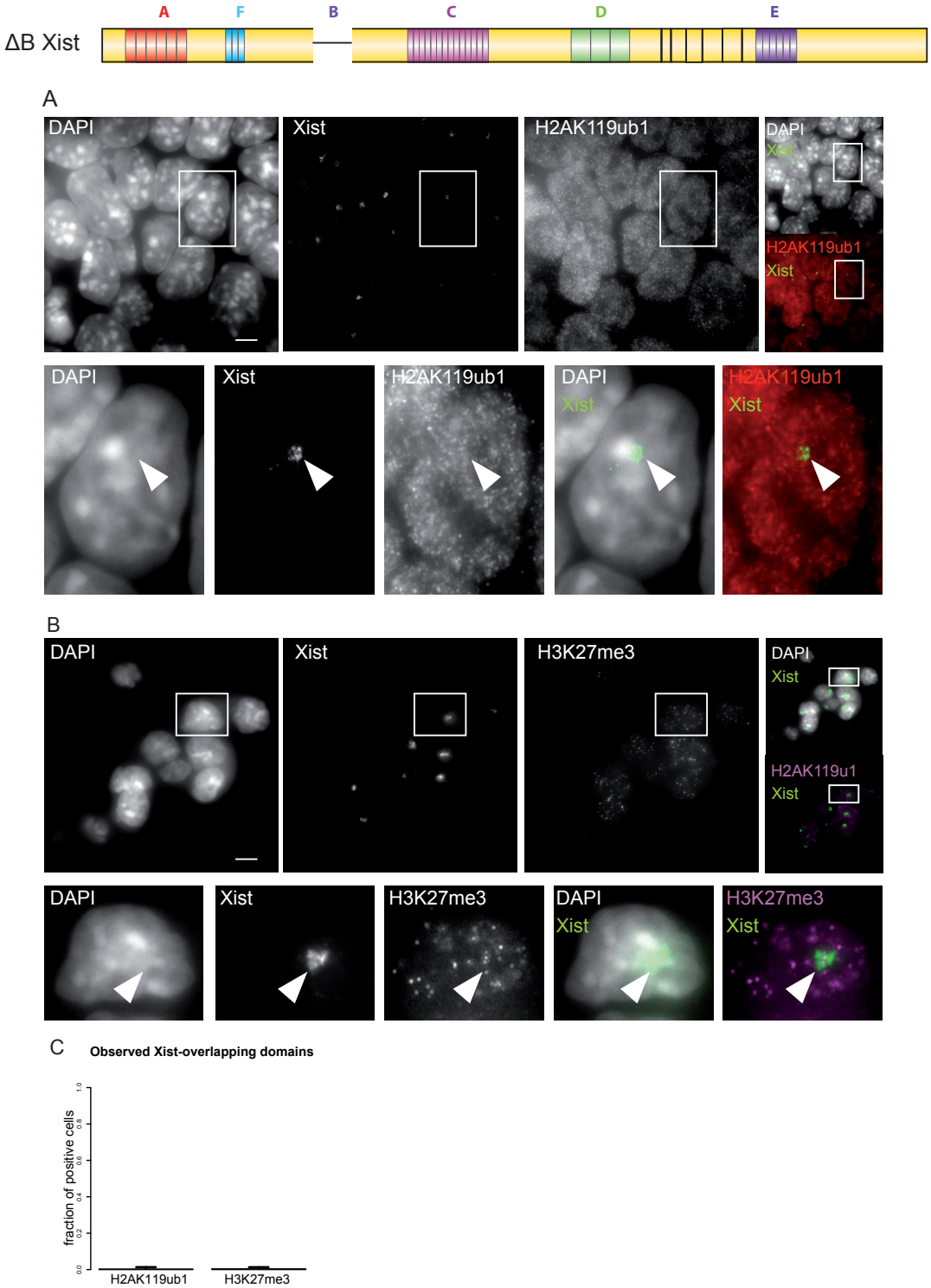


Figure 5.5: PRC1 and PRC2 recruitment by ΔB Xist RNA. ΔB Xist, shown above, was detected by RNA-FISH, after inducing cells with doxycycline for 24 hours. Cells were co-stained by immunofluorescence for H2AK119ub1 (A) and H3K27me3 (B). Wide-field images represent average of 10 stacks, cells in boxes were enlarged below and represented as single stacks. Arrows indicate Xist domains. Scale bar is 5μm. In (C), combined RNA-FISH/IF experiments were quantified by counting H2AK119ub1 (red bar) or H3K27me3 (purple bar) domains overlapping Xist after 24 hours doxycycline induction.

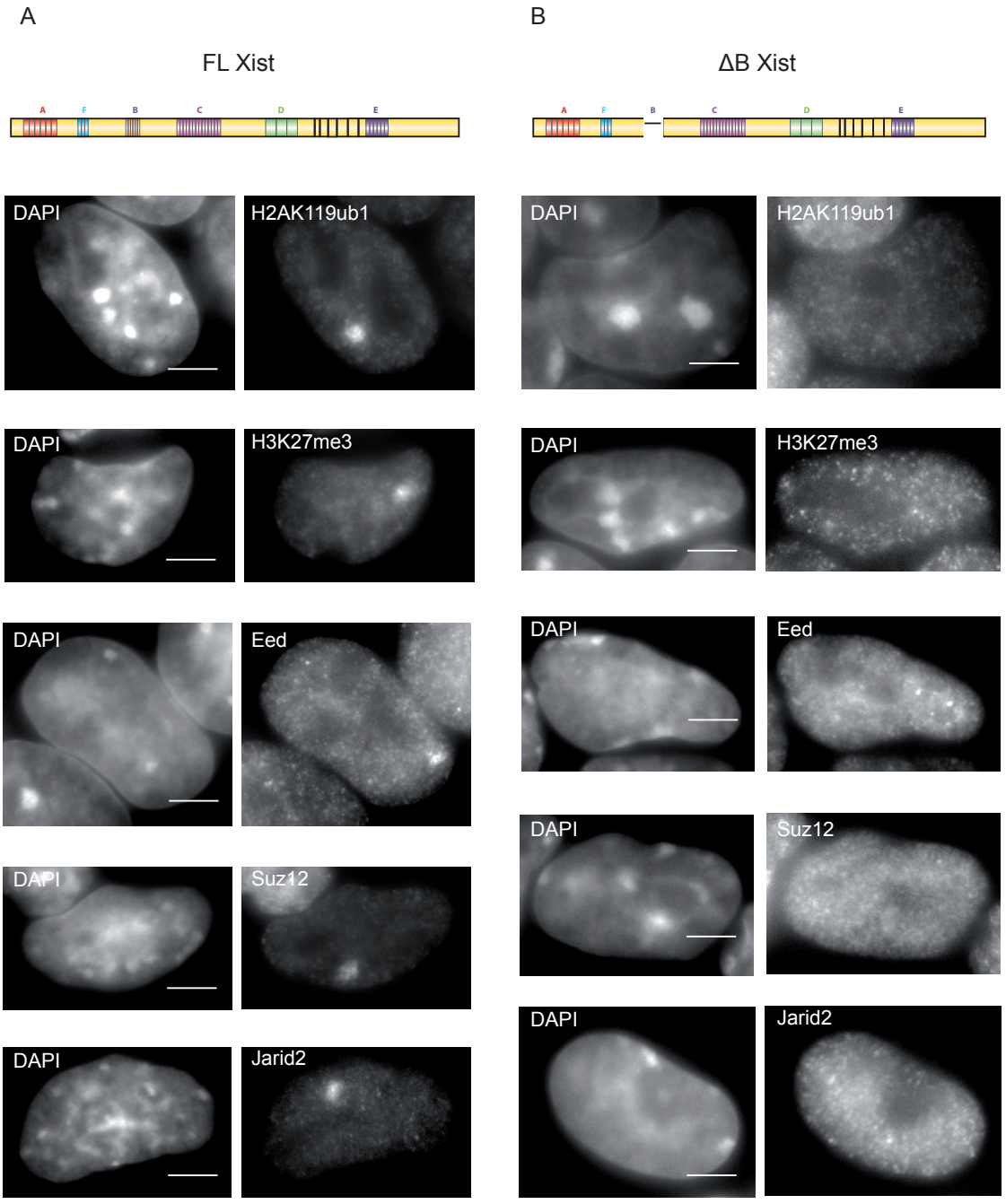


Figure 5.6: Recruitment of Polycomb subunits by FL vs ΔB Xist. Immunofluorescence assays were performed on cells expressing FL (A) vs ΔB (B) Xist following 24 doxycycline induction. Images represent average of 10 stacks. Scale bar is 5μm

the B-repeats (Fig.5.7 A and Fig.5.7 B). The specificity of the B-repeat sequence in recruiting Polycomb was further tested by inverting the sequence of the repetitive region. As shown in Figure 5.8, the phenotype is identical to that observed in the absence of the B-repeats. In conclusion, these results demonstrate that Ev-Xist encompasses all of the sequences required for Polycomb recruitment and that the B-repeats are critical for this function.

5.2.4 Dissection of the B-repeat function in Polycomb recruitment

The B-repeats of mouse Xist can be divided in two distinct segments. The first spans 300bp and comprises multiple repetitions of a short A/TGCCCC sequence, while the second segment, which is similar in terms of total length, comprises a longer and more complex C-rich sequence repeated fewer times (Fig.5.9). A similar organisation of the B-repeats is also observed in other mammals. To investigate if the bipartite structure of the sequence reflects functional heterogeneity, two distinct Ev-Xist variants harbouring either half of the B-repeats were generated, and termed Proximal and Distal Ev-Xist. Both constructs were found to be deficient in recruiting Polycomb (Fig.5.10). In the case of Proximal Ev-Xist, recruitment of PRC1 (Fig.5.10 A,C) and PRC2 (Fig.5.10 C) was reduced around 50%. The effect was dramatic in the case Distal Ev-Xist, where no PRC1 or PRC2 were detected following Xist induction (Fig.5.10 B,D). In conclusion, the Proximal part of the B-repeats is critical to mediate the Xist-Polycomb interaction, while the Distal region may play a role in stabilising already established Polycomb domains, through a yet undefined mechanism.

5.3 Discussion

Recently, work by da Rocha *et al* (2014), challenged the proposal that Xist RNA directly recruits PRC2 via the A-repeats, ascribing the ability to recruit PRC2 via the cofactor Jarid2, to the XN region of Xist [da Rocha et al. 2014]. In agreement with this conclusion, PRC2 recruitment was found unaffected by deletion of the

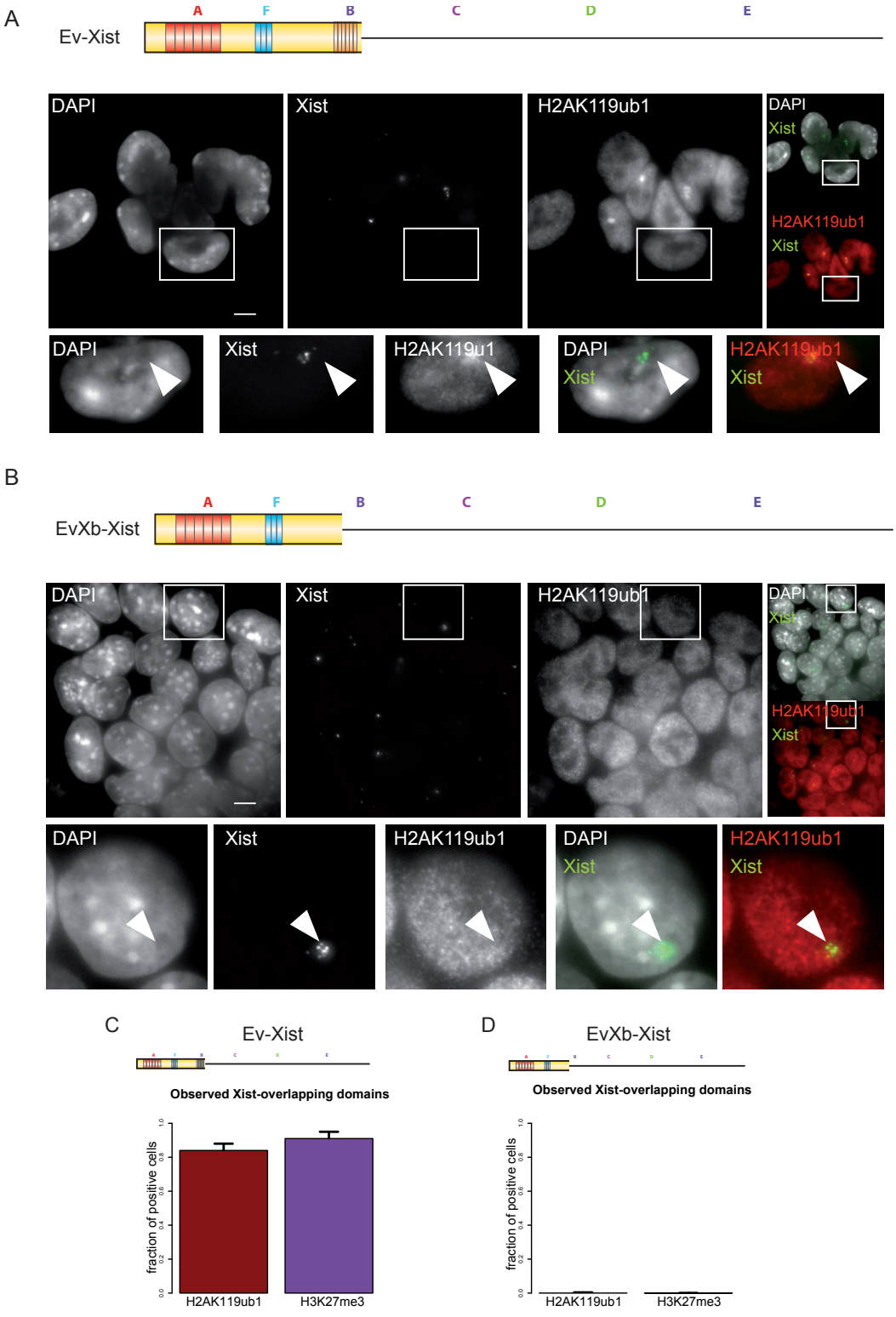


Figure 5.7: PRC1 and PRC2 recruitment by Ev- and EvXb-Xist RNA. Ev-Xist (A) and EvXb-Xist (B), both schematically represented above, were detected by RNA-FISH, after inducing cells with doxycycline for 24 hours. Cells were co-stained by immunofluorescence for H2AK119ub1. Wide-field images represent average of 10 stacks. Enlarged images of cells are single stacks. Arrows indicate Xist domains. Scale bar is 5μm. In (C) and (D), combined RNA-FISH/IF experiments for Ev-Xist (C) and EvXb-Xist (D) were quantified by counting H2AK119ub1 (red bars) or H3K27me3 (purple bars) domains overlapping Xist following 24 hours doxycycline induction.

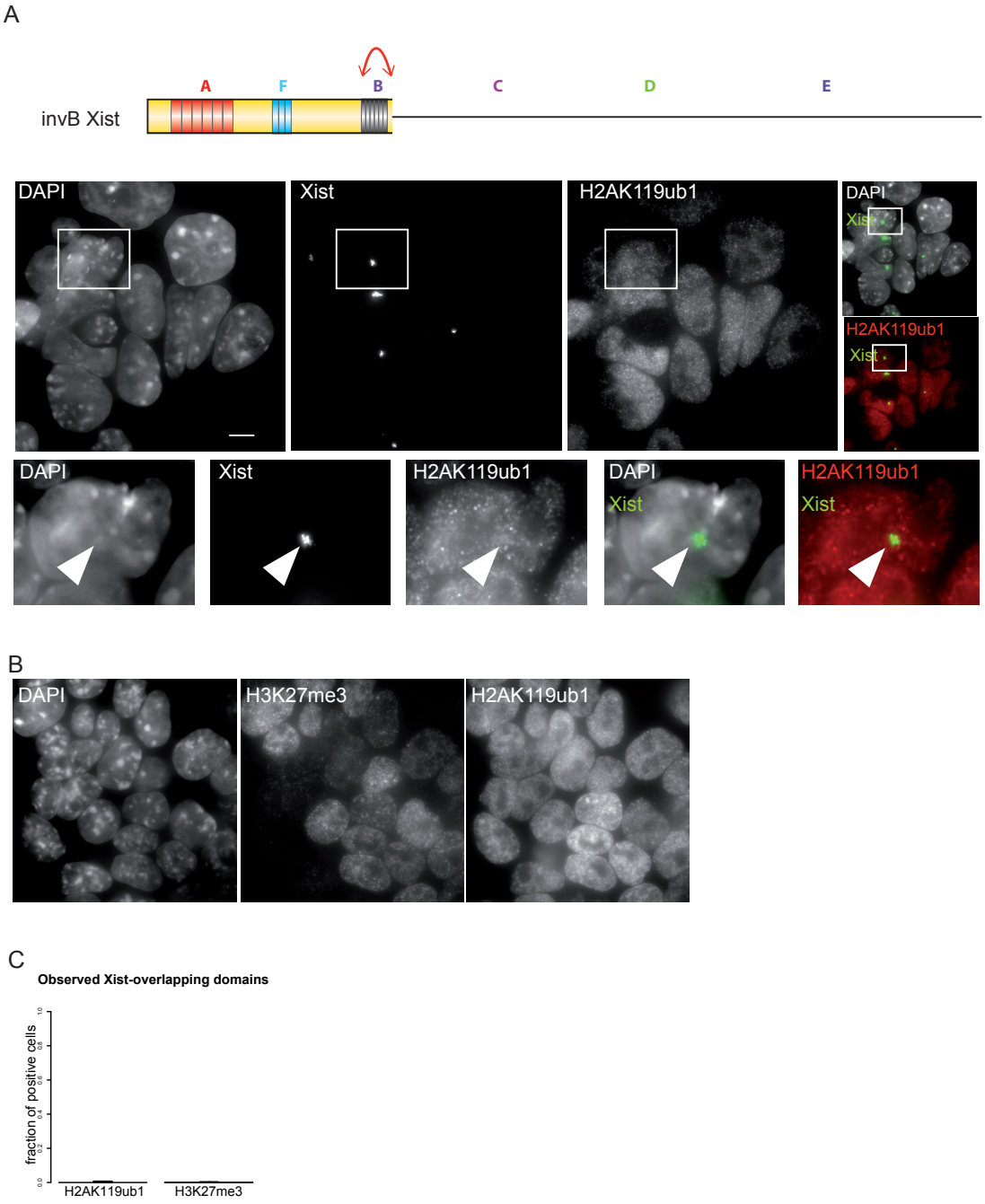


Figure 5.8: PRC1 and PRC2 recruitment by invB-Xist RNA. (A) Ev-Xist with inverted B-repeats was detected by RNA-FISH, after inducing cells with doxycycline for 24 hours. Cells were co-stained by immunofluorescence for H2AK119ub1. (B) Upon 24 doxycycline induction, cells expressing invB-Xist were co-stained for H3K27me3 and H2AK119ub1. Wide-field images represent average of 10 stacks. Enlarged images of cells are single stacks. Arrows indicate Xist domains. Scale bar is 5μm. (C) RNA-FISH/IF co-stained cells in (A) were quantified by counting H2AK119ub1 (red bars) or H3K27me3 (purple bars) domains overlapping Xist following 24 hours doxycycline induction.

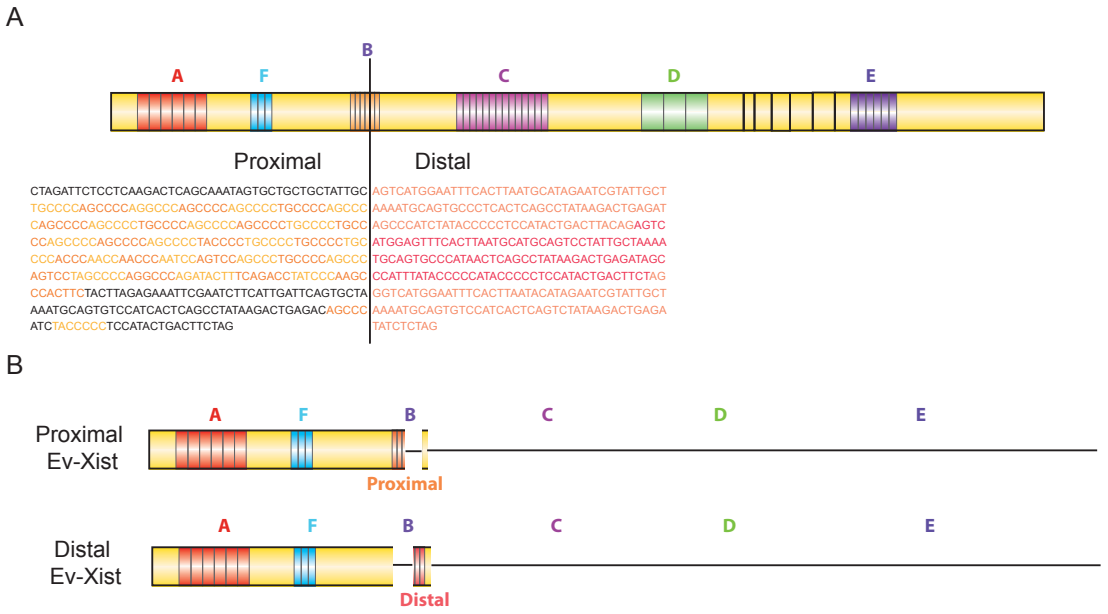


Figure 5.9: The B-repeats of Xist.(A) Schematic of the B-repeats of Xist, with full sequence of the Proximal/Distal regions. Tandem repeated regions are in coloured letters. (B) Schematic of Ev-Xist carrying either the Proximal or the Distal sequence represented in (A).

A-repeats (ΔSX Xist), but was completely lost in the context of the XN deletion (ΔXN Xist). Importantly, the same region of Xist was also found to be essential for *de novo* recruitment of PRC1, as inferred from the absence of H2AK119ub1 in response to ΔXN Xist expression. Previously, Xist sequence requirements for PRC1 recruitment on the Xi had not been directly tested, as PRC2 was believed to be the first Polycomb complex to be enriched on the Xi. Given the recent findings, showing that initiation of Polycomb recruitment on the Xi requires Pcgf3/5-PRC1 [Almeida et al. 2016], loss of H2AK119ub1, not H3K27me3, likely accounts for absence of Polycomb on the Xi. Consistent with this, an independent study in the lab, has shown that Jarid2 recognises H2AK119ub1 modifications [Cooper et al. 2016], explaining the importance of Jarid2 for PRC2 recruitment by Xist, reported by da Rocha et al (2014).

Thus, results in this Chapter, together with our observations on Pcgf3/5-PRC1, delineate a new model whereby the XN region of Xist is the crucial element mediating recruitment of Polycomb in XCI.

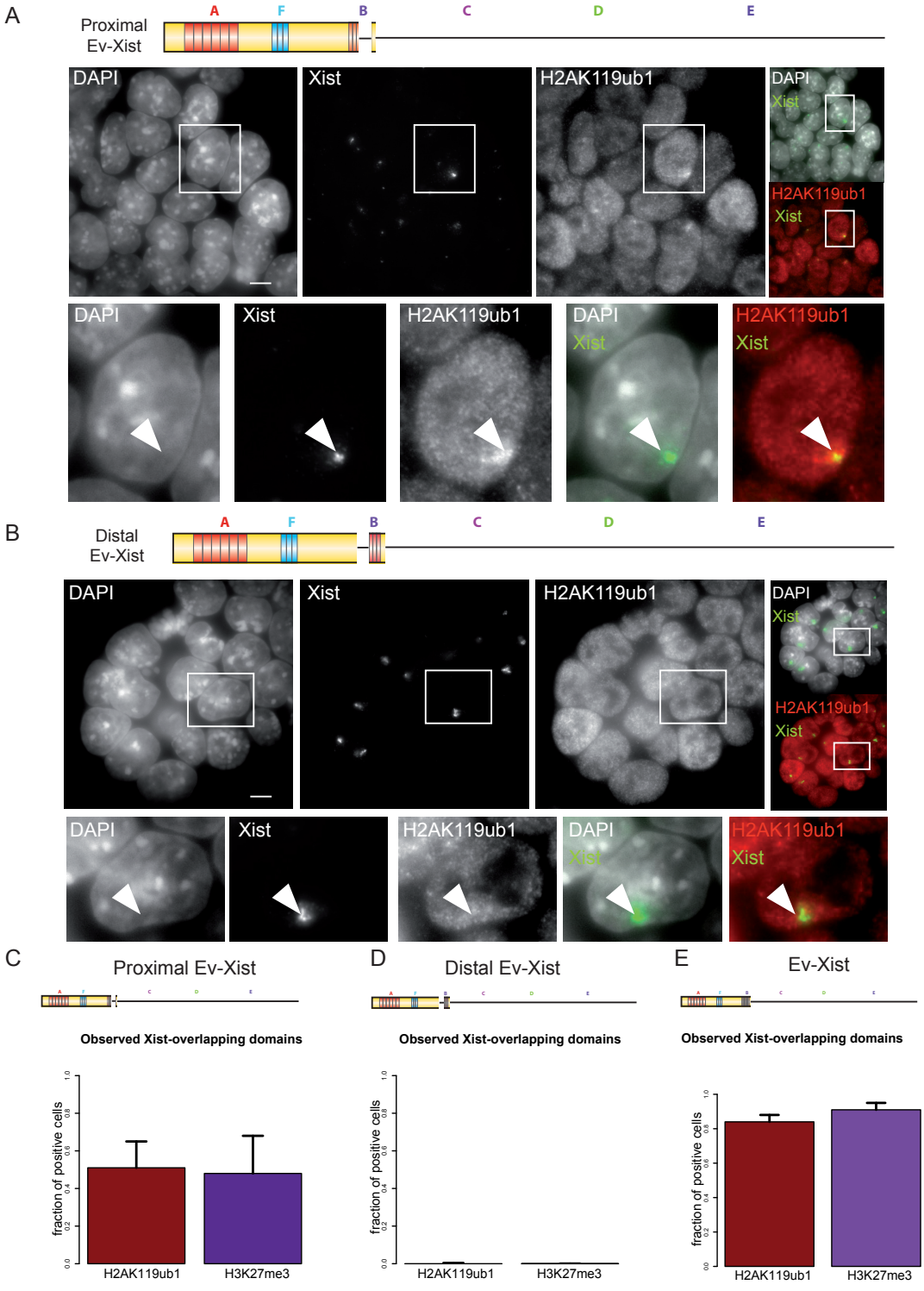


Figure 5.10: Proximal vs Distal Ev-Xist and Polycomb. Proximal (A) and Distal Ev-Xist (B), both depicted in the schematics at the top of the figure, were detected by RNA-FISH, after inducing cells with doxycycline for 24 hours. Cells were co-stained by immunofluorescence for H2AK119ub1. Wide-field images represent average of 10 stacks. Enlarged images of cells are single stacks. Arrows indicate Xist domains. Scale bar is 5μm. In (C) and (D), combined RNA-FISH/IF experiments for Proximal Ev-Xist (C) and Distal Ev-Xist (D) were quantified by counting H2AK119ub1 (red bars) or H3K27me3 (purple bars) overlapping Xist domains 24 hours after doxycycline induction. Data for Ev-Xist are also shown for comparison (E).

Table 5.1: Summary of Xist-mediated Polycomb recruitment observed in each of the analysed cell lines. Numbers represent averages for at least two replicates for each cell line.

Xist ID	Number of independent cell lines	H2AK119ub1 (%)	H3K27me3 (%)
FL	3	92	93
ΔSX Xist	4	91	89
ΔXN Xist	3	0.7	0.1
ΔB Xist	5	0.2	0.4
Ev-Xist	6	84	91
EvXb-Xist	6	0	0.1
Proximal Ev-Xist	2	51	48
Distal EvXist	2	0.1	0.1
InvB Ev-Xist	2	0	0

A brief summary of the most significant results obtained by co-staining Xist and the marks deposited by either PRC1 and PRC2 (H2AK119ub1 and H3K27me3), can be found in Table 5.1. The fact that PRC1 and PRC2 enrichment appeared unaffected in response to ΔSX Xist expression, in comparison with FL Xist, does not rule out that the A-repeats could play a subsidiary role in contacting Polycomb. Genome-wide techniques, such as ChIP-seq, are required to exclude possible redistribution of Polycomb complexes across the chromosome, and to quantify levels of PRC1/PRC2 enrichment at target sites, difficult to evaluate by immunofluorescence-based techniques.

Further work in this Chapter mapped the segment within the XN region of Xist required for Polycomb recruitment to a small region encompassing the B-repeats. This observation is consistent with the reported high conservation of the B-repeats between human and mouse and also in other eutherian species [Yen et al. 2007]. In the future it will be interesting to explore the possibility of a common mechanism of Polycomb recruitment by lncRNAs relying on related sequences or similar secondary structures. An interesting example could be the Xist ortholog, RSX. Rsx RNA was shown to mediate Polycomb recruitment during the inactivation process in marsupials [Grant et al. 2012], but whether this depends on sequence motives

potentially similar to the B-repeats remains to be determined.

Dissection of the the B-repeat region showed that it can be naturally split into two parts of similar length. The proximal part of the B-repeats that was shown to be essential for Polycomb recruitment is composed of ∼30 repeats of a simple C-rich short sequence (Fig.5.9). This sequence could be functionally active either by constituting a defined secondary structure, able to fold into a docking platform for interacting factors as was hypothesised for the A-repeats; or by acting as a recognition motif for specific RNA-binding factors. The second possibility is perhaps more likely given that the B-repeat motif, unlike the A- and E-repeats, does not seem to be very structured, as recently determined through the PARIS method for mapping RNA structure *in vivo* [Lu et al. 2016]. It therefore seems more likely that the B-repeats, and primarily their proximal part, function as a sequence-specific binding motif, which directly or indirectly recruits Polycomb. In this respect, it was relevant to define potential B-repeat specific binding factors, that could shed light on the mechanism of Polycomb recruitment (addressed in Chapter 7). In addition, identification of a short motif responsible for PRC1 and PRC2 recruitment, provided a tool to further investigate the contribution of Polycomb to Xist-mediated transcriptional silencing (Chapter 6).

6

Polycomb and Xist-mediated transcriptional silencing

Contents

6.1 Introduction 101
6.2 Results 102
6.2.1 The XN region contributes to Xist-mediated gene silencing 102
6.2.2 Locus-directed targeting of Ev-Xist in a single copy . . 104
6.2.3 Ev-Xist shows efficient gene silencing on chromosome 11 105
6.2.4 The B-repeats contribute significantly to Ev-Xist-mediated gene silencing 108
6.3 Discussion 109

6.1 Introduction

The purpose of the X inactivation process is to convert into a facultative heterochromatic state one X chromosome of female cells, in order to establish and stably maintain silencing of the underlying genes. How gene silencing is established remains poorly understood, but clearly involves several overlapping pathways. The known properties of Polycomb complexes suggest that recruitment of PRC1 and PRC2 could be an important mechanism for Xist-mediated silencing. In support of this, the Polycomb complex Eed-Enx1 was implicated in the maintenance of

imprinted XCI in the mouse trophectoderm lineage [Wang et al. 2001]. Whilst there is evidence to support that this could also be the case for random XCI [Silva et al. 2003], it has been reported that disruption of either PRC1 or PRC2 does not lead to gene reactivation or to a defect in the initiation of gene silencing [Kalantry et al. 2006, Leeb et al. 2007]. One could argue that PRC1 and PRC2 have redundant functions compensating for each other in knock-out experiments, but loss of silencing ability was observed when the A-repeats of Xist were deleted [Wutz et al. 2002], and Polycomb was still present [da Rocha et al. 2014]. Also, no Polycomb components were listed in any of the genetic screens for Xist silencing factors [Moindrot et al. 2015, Monfort et al. 2015], whereas several other factors identified in the same studies were subsequently directly related to the establishment of gene silencing, linked to histone deacetylation and subsequent exclusion of the transcriptional machinery from the Xi, or to recruitment of the m^6A methylation complex [McHugh et al. 2015, Chu et al. 2015, Patil et al. 2016]. In this context, Polycomb might have a more subtle role in maintaining the silent state achieved by these other factors.

Work in this Chapter was set out to dissect the role of Polycomb in gene silencing, first by RS analysis of ΔXN Xist, previously shown to be incapable of PRC1/PRC2 recruitment (Chapter 5, Section 5.2). Then, using a single copy targeting strategy subtle differences in gene expression in the presence and absence of the Polycomb-recruiting element were directly compared.

6.2 Results

6.2.1 The XN region contributes to Xist-mediated gene silencing

The ΔXN Xist variant (Fig.4.2, construct c) was shown to correctly localise and spread *in cis* (Chapter 4, Section 4.2), as previously reported [Wutz et al. 2002]. As the location of the transgene was unknown *a priori*, the RS for ΔXN Xist was determined using a blind analysis approach. The integration locus in a specific ΔXN Xist line was identified based on a significant down-regulation of

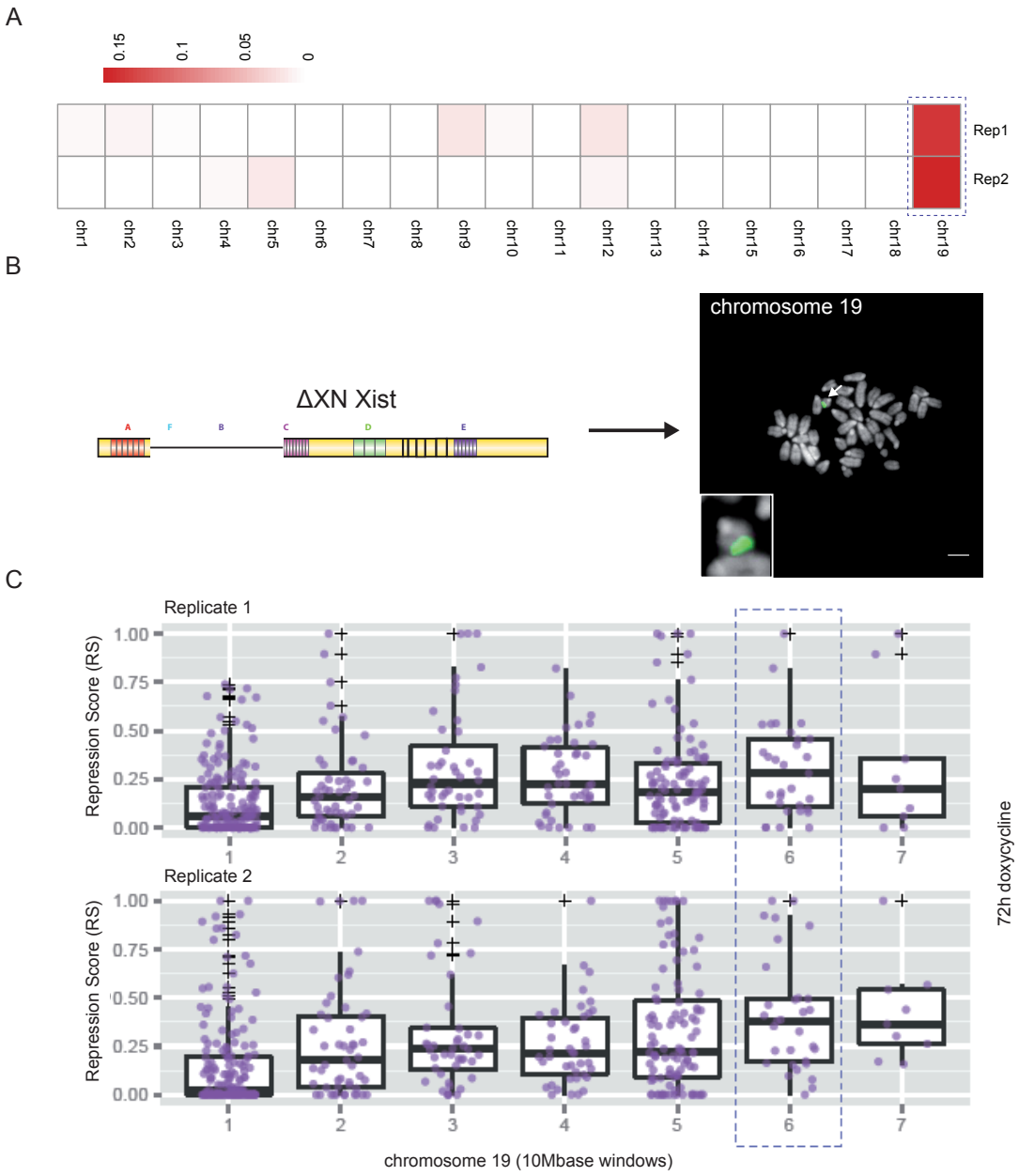


Figure 6.1: ΔXN Xist-mediated gene silencing. (A) Significantly different levels of gene expression between alleles are plotted for each chromosome over 2 biological replicates. Each box represents a chromosome; overall allelic difference in gene expression is colour-coded from white (no difference), to red (mono-allelic expression). (B) DNA-FISH for Xist (green) was performed on metaphase spreads. Chromosomes were distinguished based on their banding. (C) Box plots represent the median value of all RS for each 10Mbase window spanning the length of chromosome 19. RS is represented in purple for each gene in individual windows. The sixth window (boxed), where RS is the highest, encompasses the site of integration of the transgene. Two independent biological replicates are represented. Figure generated in collaboration with Dr. Guifeng Wei.

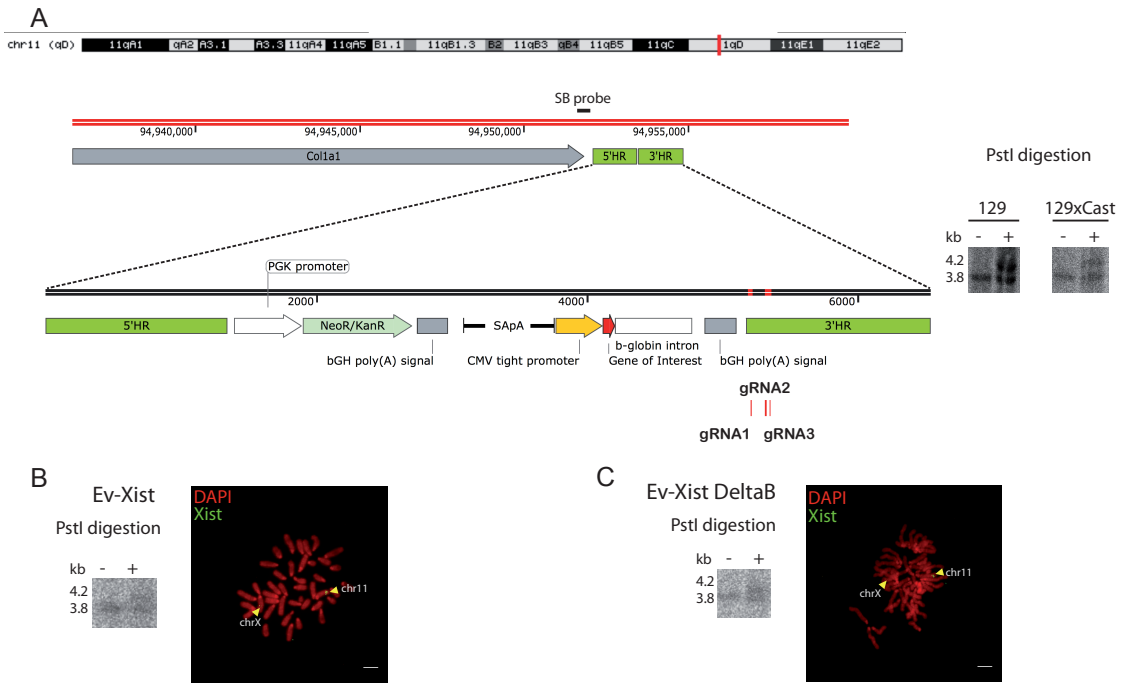


Figure 6.2: Targeting of Ev-Xist to the col1a1 locus. (A) A transgene containing a CMV inducible promoter controlling the gene of interest, was cloned between homology arms designed to match the region of chromosome 11, just downstream of the col1a1 gene. The transgene was co-transfected with gRNAs also depicted in the Figure. Correct integration of a mock transgene was verified by PstI digestion in two different cell lines (probe location is shown in (A)), in a classical *Mus musculus domesticus* (129) and in the hybrid *Mus musculus domesticus* x *Mus musculus castaneus* (129xCast) background. The cassette comprised either Ev-Xist (B) or ΔB Ev-Xist (C). The integration was checked by Southern blot (probe location is shown in (A)) and DNA-FISH on metaphases. DNA-FISH images represent single stacks, and scale bar represent 5μm.

genes located on one allele of chromosome 19 (Fig.6.1 A), and confirmed by *in situ* analysis (Fig.6.1 B). Interestingly, a significant transcriptional effect was detected on several genes, with an overall RS of 0.3294 matching ∼65% of that calculated for FL Xist (0.4917), shown in Figure 6.1. Although the numbers are not directly comparable, it can be inferred that gene-silencing was partially achieved, although possibly not to the extent achieved by FL Xist.

6.2.2 Locus-directed targeting of Ev-Xist in a single copy

Given the limitations of a random integration system (variable site of integration and copy number), the above results were substantiated using a targeting system in which a single copy of Ev-Xist was integrated at a defined locus. For this purpose,

a targeting strategy previously employed by Beard *et al* (2002), for targeting the col1a1 locus on chromosome 11, was adapted. The efficiency of targeting with this system was relatively low, especially in the case of long integrants, and proved to be technically challenging, given the need of two subsequent steps of genetic manipulation in order to establish the final working cell line. The targeting strategy was therefore modified with the use of CRISPR/Cas9 technology [Platt et al. 2014]. gRNAs were designed to cut the col1a1 locus around the site of integration, and co-transfected with the gene of interest cloned between short homology arms (Fig.6.2 B). Development of this strategy was in collaboration with Dr. Tatyana Nesterova. Correct integration of the transgene was assessed by DNA-FISH on metaphase chromosomes and by Southern blot [Beard et al. 2006]. The col1a1 locus was separately targeted with two constructs (Fig.6.2 B and C), both controlled by a doxycycline-inducible promoter, identical to that used in the random integration strategy:

1. Ev-Xist

2. ΔB Ev-Xist

Once positive clones were selected, they were tested by RNA-FISH for correct Xist expression upon doxycycline treatment. Furthermore, PRC1/PRC2 recruitment, was observed in the presence (Fig.6.3 A) but not in the absence (Fig.6.3 B) of the B-repeat region.

6.2.3 Ev-Xist shows efficient gene silencing on chromosome 11

Ev-Xist integrated in the col1a1 locus was then tested for its ability to mediate allele-specific gene silencing with the RS analysis. In detail, the RS was calculated for each gene and used to compute a median RS for each 10Mb window across the whole chromosome (Fig.6.4 A). The single copy Ev-Xist transgene was able to mediate silencing of many genes on chromosome 11, notably around the site of integration, at similar levels to those observed for the other two Ev-Xist transgenes,

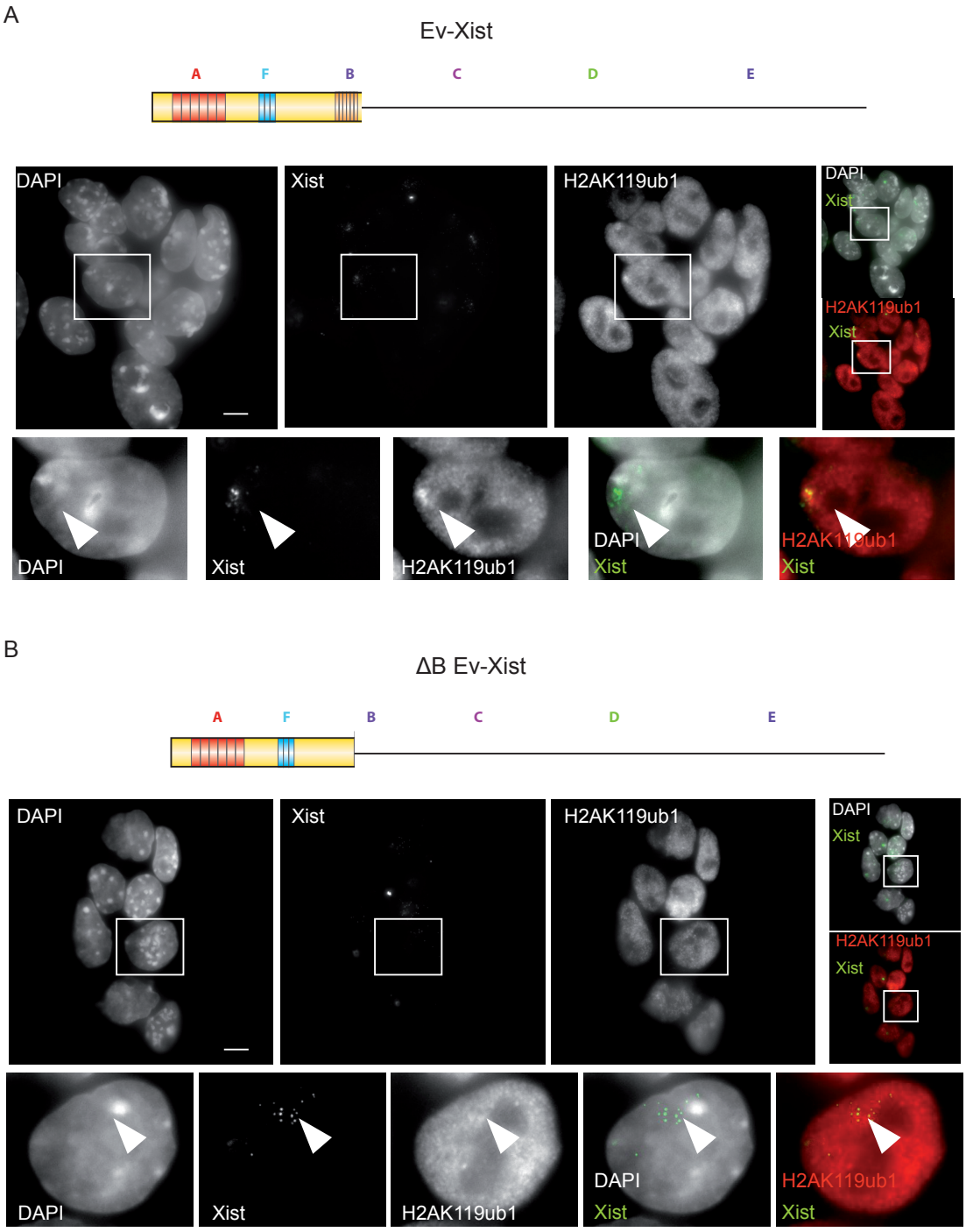


Figure 6.3: PRC1 and PRC2 recruitment by single copy Ev-Xist and ΔB Ev-Xist RNA. Ev-Xist (A) and ΔB Ev-Xist (B), generated by targeting the col1a1 locus on chromosome 11, were stained by RNA-FISH, after inducing cells with doxycycline for 24 hours. Cells were co-stained by immunofluorescence for H2AK119ub1. Wide-field images represent average of 10 stacks. Enlarged cells are single stacks. Arrow indicate Xist domains. Scale bar is 5μm.

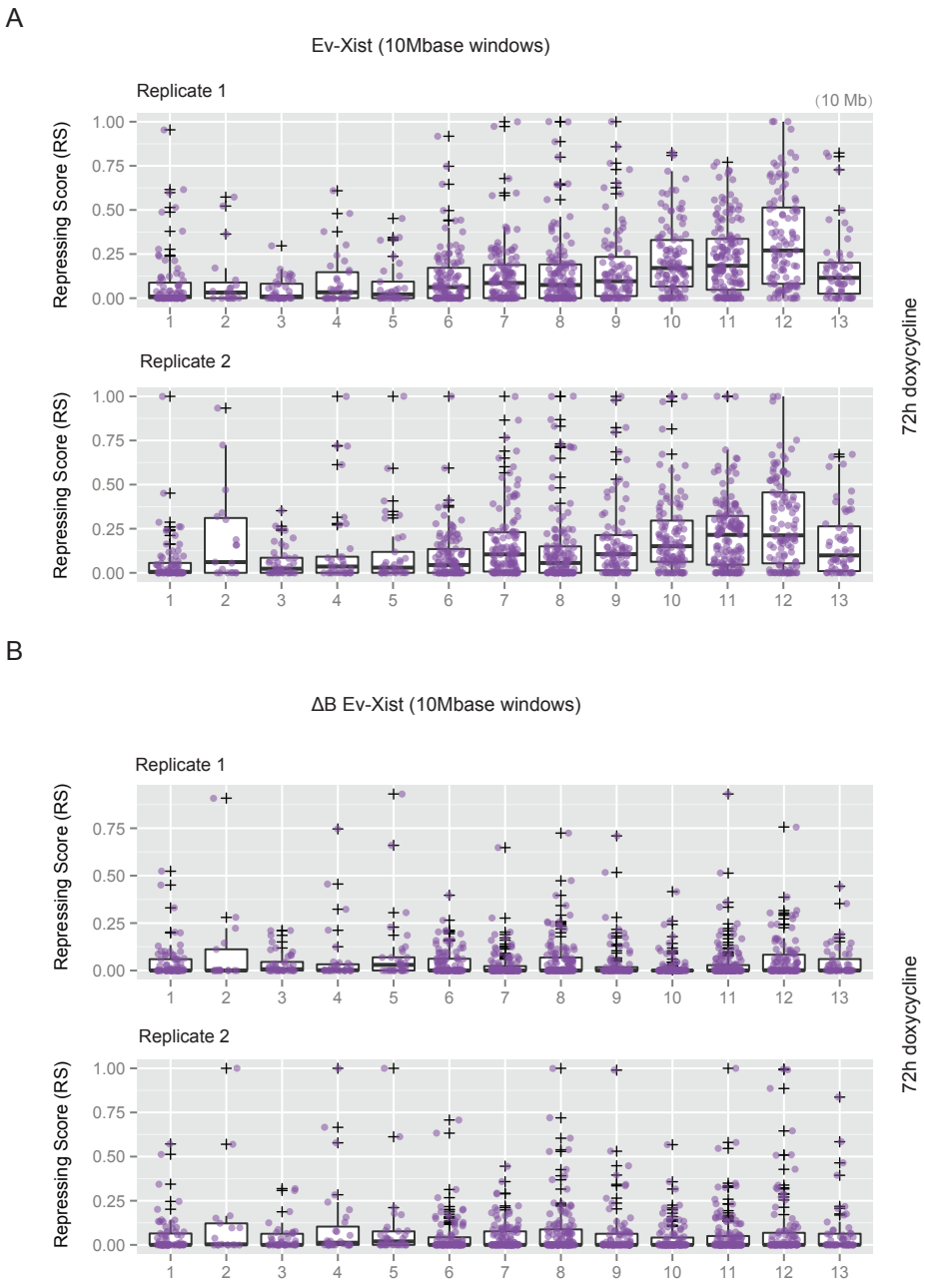


Figure 6.4: RS of single copy Ev-Xist and ΔB Ev-Xist. Box plots represent the median value of all RS for each 10Mbase window spanning the length of chromosome 11. RS is represented in purple for each gene within individual windows. Two independent biological replicates are represented. In (A) data are shown for Ev-Xist, in (B) for ΔB Ev-Xist, both integrated in the col1a1 locus. Figure generated in collaboration with Dr. Guifeng Wei.

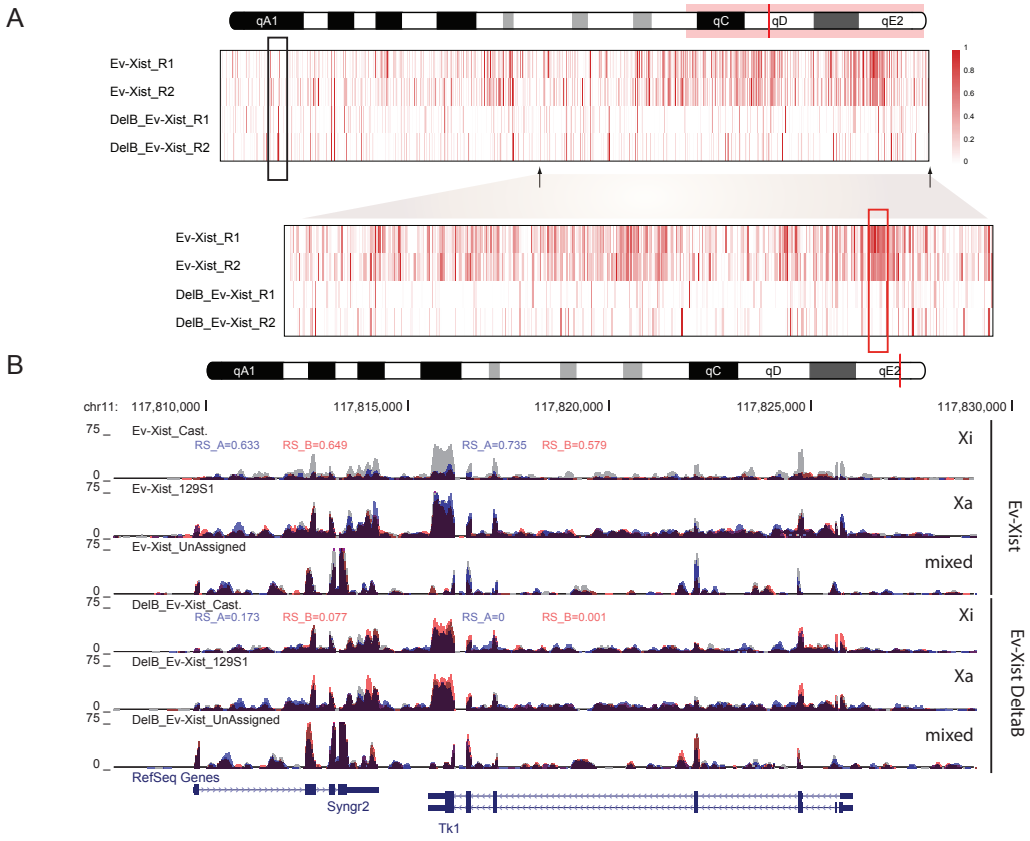


Figure 6.5: Single copy Ev-Xist and ΔB Ev-Xist RS analysis. (A) RS of each gene of the region surrounding the col1a1 site is colour-coded according to the legend for each of the two replicates of both FL and ΔB Ev-Xist. A region with high silencing is enlarged below. (B) The UCSC snapshot shows the distal region on chromosome 11, boxed in red in (A); cast, 129S1, and UA represent the Castaneus, 129S1, and unassigned reads respectively. All reads at each time point are plotted for each track: untreated: grey; replicate 1: red; replicate 2: blue. Figure generated in collaboration with Dr. Guifeng Wei.

randomly integrated elsewhere on the same chromosome, and on chromosome 9 (Chapter 3, Section 3.2).

6.2.4 The B-repeats contribute significantly to Ev-Xist-mediated gene silencing

The silencing ability of Ev-Xist was compared with that of ΔB Ev-Xist, both integrated in the same locus. Interestingly, upon induction of ΔB Ev-Xist, only very low levels of silencing were measured, based on RS analysis (Fig.6.4 B). Silencing was moreover inconsistent between biological replicates, suggesting that even sporadic silencing events occurring for few genes, tend to be rapidly reverted (Fig.6.4 B,

compare Replicate 1 and 2). RS across chromosome 11 for all genes in two biological replicates of both Ev-Xist and ΔB Ev-Xist is represented in Figure 6.5 A. Using Ev-Xist most genes were repressed, notably in the distal portion of the chromosome where the transgene was integrated. However, the same region was largely unaffected by expression of ΔB Ev-Xist, for which only random mono-allelic expression could be reported for few genes (Fig.6.5 A). Specific examples, are the Syngr2 and Tk1 genes in Figure 6.5 B, for which Ev-Xist represented a powerful silencing trigger, whereas ΔB Ev-Xist had little or no effect. Only few exceptions, exhibiting a consistent mild silencing due to ΔB Ev-Xist (highlighted in Figure 6.5 A, top panel, black box), could be reported. Interestingly, these genes are located in a proximal position, far away from the integration site, where the silencing process is generally less efficient.

6.3 Discussion

Work in this section was set out to clarify the contribution of Polycomb to gene silencing. First, RS analysis, described in Chapter 3, was applied to Xist lacking the PRC1/PRC2 recruitment region (ΔXN Xist). Using the same time point after Xist induction, in which efficient FL Xist-dependent gene silencing was observed (72 hours), a defect in silencing was reported, although to a lesser extent compared to what was observed for Xist lacking the A-repeats. In fact, ΔSX Xist was unable to establish gene silencing, even at low levels, while Xist lacking the XN region (ΔXN Xist) is still capable of silencing genes, even if less strongly. This result is in partial agreement with previously published work, in which Polycomb was found to significantly contribute to gene silencing in trophoblast giant cells, during imprinted XCI [Wang et al. 2001], and partially also during random XCI [Silva et al. 2003]. On the other hand, the degree of silencing loss was unexpected, given the reported silencing defects for the A-repeats Xist deletion, in which Polycomb is still present [Wutz et al. 2002].

In order to directly compare the silencing ability of Xist in the presence and absence of Polycomb, a random integration system was not suitable, as potential

subtle differences in gene silencing could be due to locus-specific variables. Therefore, conventional gene targeting was combined with CRISPR technology, with the purpose of integrating Ev-Xist into the col1a1 locus on chromosome 11, and assess transcription arising from the integrated allele. Although a good level of gene silencing was detected when Ev-Xist was expressed from the integrated locus, this was completely lost upon deletion of the XN region. This result was particularly surprising given that silencing was detected for the ΔXN construct in the context of FL Xist (Chapter 5, Section 5.2), and, notably, the A-repeats are still present in the ΔB Ev-Xist construct. There are two alternative hypotheses to explain why so little silencing occurs. It could be that the A-repeats are necessary, but not sufficient to mediate gene-silencing, requiring other sequences not present in the Ev-Xist construct. This hypothesis is partially corroborated by the fact that a ∼50% decrease of the observed RS was reported for Ev-Xist compared to FL Xist. Good candidates for assisting the A-repeats are, for instance, the E-repeats, providing binding sites for the protein LBR, whose knock-down was shown to affect Xist-mediated silencing [Chen et al. 2016]. Alternatively, the reported defect in silencing using ΔB Ev-Xist, might reflect secondary effects due to impaired stability of the RNA, that were not detectable by RNA-FISH. In this sense, different deletions, and perhaps inversion of the deleted region, could also be tested in the col1a1-targeted system. Although more experiments are needed to precisely define the extent of its contribution to gene silencing, data from this Chapter collectively suggest that Polycomb is partially responsible for transcriptional silencing during XCI. In particular, it seems likely that Polycomb exerts its silencing function downstream of the early pathways, maintaining the silent state initially established through exclusion of the transcriptional machinery, and most likely dependent on the A-repeats. It would be interesting to understand how Polycomb stabilises the previously established silent state. One hypothesis is that PRC1/PRC2-mediated histone modifications act by changing overall chromatin accessibility and compaction.

7

The B-repeat interactome

Contents

7.1 Introduction 111
7.2 Results 112
7.2.1 *In vitro* transcription of Xist variants 112
7.2.2 Xist pull downs and mass spectrometry assays 114
7.2.3 Identification of A- and B-repeat binding factors 115
7.3 Discussion 119

7.1 Introduction

Data shown in Chapter 5, together with unpublished observations, identified the B-repeats of Xist as the segment mediating Polycomb recruitment via interaction with the non-canonical PRC1 complex Pcgf3/5-PRC1 [Almeida et al. 2016]. However the molecular nature of this interaction remains to be defined: it could be that Xist directly contacts Pcgf3/5-PRC1, or instead that an unknown RBP adaptor links Pcgf3/5-PRC1 to Xist RNA via the B-repeats. Indirect evidence points towards the latter hypothesis, as no known RNA-binding domains are annotated in any Pcgf3/5-PRC1 subunits. Additionally, Pcgf5 was found to interact with Xist in formaldehyde crosslinking experiments, but not in UV-crosslinking ones, implying an indirect contact between Pcgf3/5-PRC1 and Xist [Chu et al. 2015, McHugh et al.

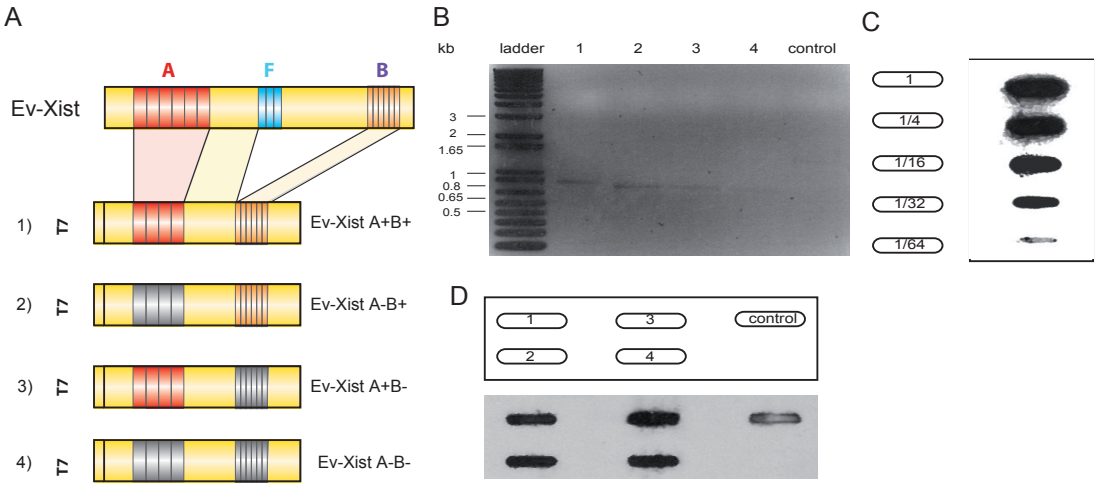


Figure 7.1: Xist *in vitro* transcription. (A) All combinations of intact (red and orange) vs non-functional (grey) Xist constructs were *in vitro* transcribed, employing a T7 RNA polymerase. The corresponding sequence of endogenous Ev-Xist cDNA is also shown. (B) Size of the constructs (numbered as in (A)) was checked on denaturing gel (0.8% agarose). FLuc Control RNA, with a known size of 1.3kb was used as a positive control. (C) Titration of dUTP-biotin with respect to the recommended concentration of 2.5mM. (D) Slot blot assay of all constructs (numbered as in (A)), 1μg of RNA was loaded.

2015].

Recently, an extensive list of Xist binding partners has been identified, as discussed in Chapter 1, Section 1.5. In particular, Chu *et al* (2015), performed an *in vivo* formaldehyde-based crosslinking study, and identified factors (for instance, hnRNP-K) that were shown to affect Polycomb activity when knocked-down. However, the involvement of these factors as specific RBP adaptors for Polycomb, was not further explored. With this in mind, an *in vitro* approach was set out in this Chapter, to define a B-repeat specific interactome.

7.2 Results

7.2.1 *In vitro* transcription of Xist variants

Immuno-precipitation assays were developed using *in vitro* transcribed Xist B-repeats. The A-repeat region was included as a control, as A-repeat specific binding

factors have been defined previously. Transcribed RNAs were UV-crosslinked with nuclear extract from ES cells, followed by label-free quantitative mass spectrometry, to identify RNA-bound factors. The main goal was to compare Xist constructs, either with or without functional elements, and test for gain/loss of binding factors. Four different constructs were *in vitro* transcribed, using a T7 promoter (Fig.7.1 A):

1. Xist A^+B^+ (intact A- and B-repeats)
2. Xist A^-B^+ (non-functional A-repeats and intact B-repeats)
3. Xist A^+B^- (intact A-repeats and non-functional B-repeats)
4. Xist A^-B^- (non-functional A- and B-repeats)

For the intact vs non-functional A-repeats (A^+ vs A^-), constructs described in Wutz et al. 2002 *et al* (2002), were employed. In detail, a synthetic version of the A-repeats, which was shown to be functional, and a mutant derivative, with point mutations that disrupt the secondary structure of the RNA and abolish the A-repeats mediated silencing activity, were used. For the intact B-repeats (B^+), the full mouse sequence, as annotated in the UCSC mm10 genome, was employed. An inverted sequence was used to achieve loss-of-function (B^-), based on the observation that InvB Ev-Xist is not able to recruit PRC1/2 (as shown in Chapter 5, Section 5.2). The size of the *in vitro* transcribed RNA was confirmed on a denaturing agarose gel (0.8% m/v, 0.5μg of RNA were loaded per each lane), as shown in Figure 7.1 B. Biotin molecules were randomly incorporated into the transcript, by addition of biotin-dUTP to the reaction. This way, labelling of the RNA molecules at either end could be avoided. Specifically, adding a relatively large moiety to the 5' end of Xist could disrupt the folding of the A-repeats, whereas adding it to the 3' end of the transcript, could cause premature termination of transcription. On the other hand, it was necessary to minimise the amount of biotin molecules per transcript, in order to avoid disruption of internal structured RNA motifs. Therefore an assay was set up to evaluate the minimum amount of biotin that was compatible with RNA detection after pull-downs, using avidin magnetic

beads. 1μg of RNA synthesised with different concentrations of biotin-dUTP was pulled down, and a slot blot assay was performed using a peroxidase-coupled biotin antibody to visualise the incorporation of biotin into the transcripts (Fig.7.1 C). 1/32 of the recommended concentration of biotin was sufficient to obtain minimal labelling for all input RNA molecules, giving consistently detectable signal. This concentration was used for all subsequent labelling reactions. The same assay was also used to test for correct incorporation of the biotin (as well as to check consistency between samples) in all constructs, including a control RNA (Fig.7.1 D).

7.2.2 Xist pull downs and mass spectrometry assays

Nuclear extract from undifferentiated PGK 12.1 XX mouse ES cells was employed for each assay. 20μg of nuclear extract was incubated with 10μg of each RNA species in the presence of RNAse inhibitors at 4 °C over night. UV-crosslinking

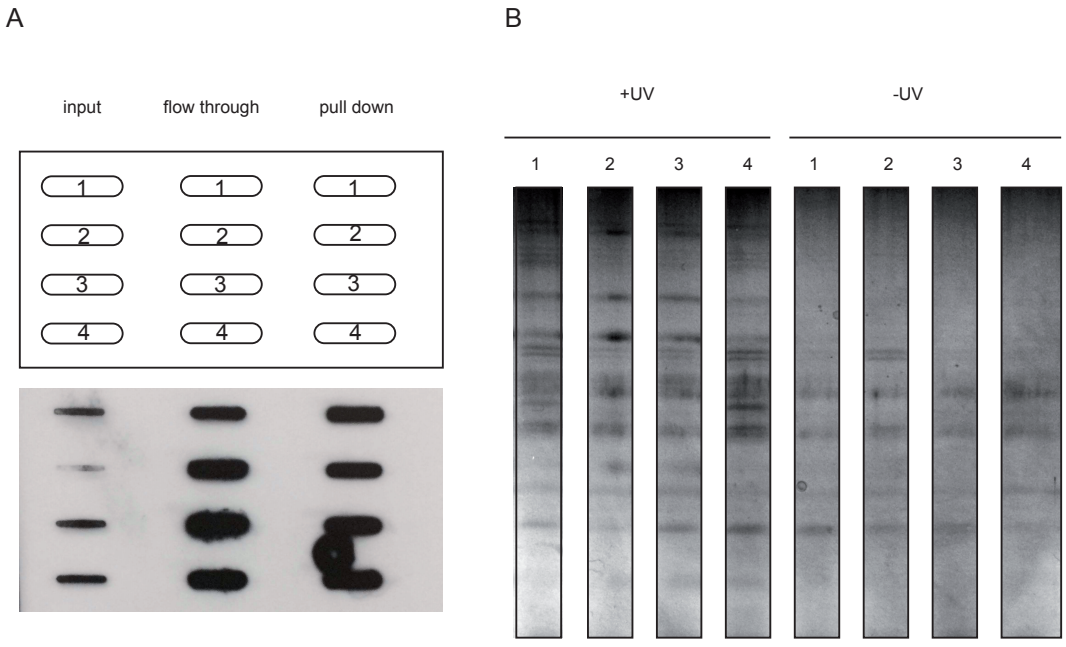
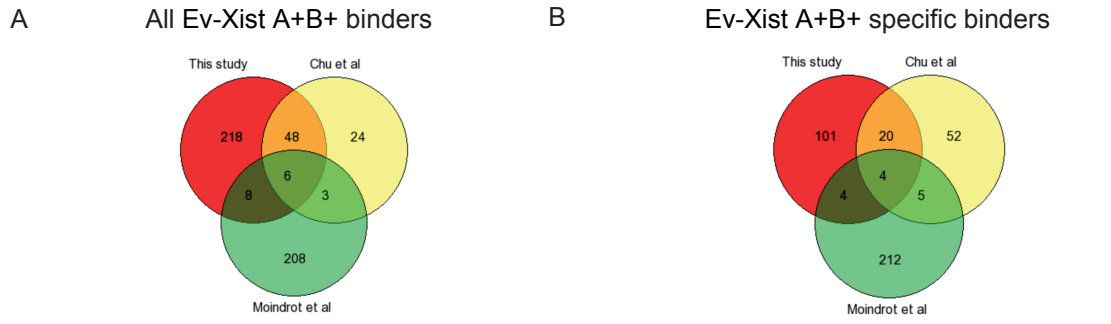


Figure 7.2: Xist pull-downs. (A) Recovery of transcribed RNAs was verified by slot blot. 1/10 of the input, 1/10 of the flow-through, and 1/50 of the pull-down were loaded. (B) 10% silver-stained SDS gel, with 1/50 of the pulled-down fractions loaded in each well. Constructs are numbered as in Figure 7.1 A.

was performed subsequently, in order to achieve stable binding of the nucleic acid



All studies (6)	This study & Chu et al (45)		This study & Moindrot et al (7)	Chu et al & Moindrot et al (3)
rbm15	celf1	khsrp	anxa11	sap18
ptbp1	ddx17	l1td1	luc7l3	spen
matr3	ddx5	lin28a	nop56	wtap
hnrnpl	dhx9	mybbp1a	pcbp2	
hnrnpc	elavl1	myef2	pou5f1	
hnrnpa0	fubp3	poldip3	sf1	
	fus	raly	smap2	
	hnrnpa1	rbfox2		
	hnrnpa2b1	rbm14		
	hnrnpa3	sfpq		
	hnrnpab	srrt		
	hnrnpd	srsf2		
	hnrnpk	srsf3		
	hnrnpm	srsf5		
	hnrnpr	srsf7		
	hnrnpu	srsf9		
	hnrnpul2	ssb		
	hnrpdl	syncrip		
	igf2bp1	tardbp		
	igf2bp3	tra2b		
	ilf2	xrn2		
	ilf3	zfr		
	khdrbs1			

All studies (4)	This study & Chu et al (26)		This study & Moindrot et al (4)	Chu et al & Moindrot et al (5)
rbm15	celf1	l1td1	luc7l3	hnrnpa0
ptbp1	ddx17	lin28a	nop56	hnrnpc
matr3	ddx5	mybbp1a	pcbp2	sap18
hnrnpl	eif4a3	myef2	sf1	spen
	fubp3	poldip3		wtap
	fus	rbm14		
	hnrnpab	srsf2		
	hnrnpk	srsf3		
	hnrnpm	srsf5		
	hnrpdl	srsf7		
	igf2bp1	srsf9		
	igf2bp3	ssb		
	khsrp	tardbp		

Figure 7.3: Xist A^+B^+ specific binders. Overlap between Xist binders identified in this study, and Xist interacting partners identified in two independent *in vivo* screens, for biochemical interactors [Chu et al. 2015], and genetic interactors [Moindrot et al. 2015]. (A) All Xist A^+B^+ binders are shown, (B) All Xist A^+B^+ binders are shown after subtracting the hits obtained for Xist A^-B^-.

and its interacting partners, followed by pull-downs with avidin-coupled magnetic beads. After elution from beads, presence of labelled RNA was verified by slot blot using 1/10 of the input, 1/10 of the flow-through, and 1/50 of the pull-downs (Fig.7.2 A). Pull-downs were also performed without the crosslinking step (Fig.7.2 B). Samples 1-4 were used for further mass spectrometric analysis.

7.2.3 Identification of A- and B-repeat binding factors

280 factors were identified using Xist A^+B^+ (after filtering out known contaminants). 20% of these overlapped the Xist interactome defined by Chu *et al* (2015), as shown in Figure 7.3 A. The hits from Xist A^-B^- were then subtracted, giving a revised list of 129 proteins. The identified proteins overlapped significantly, and could therefore be compared with, the Xist interactomes published by Chu *et al*

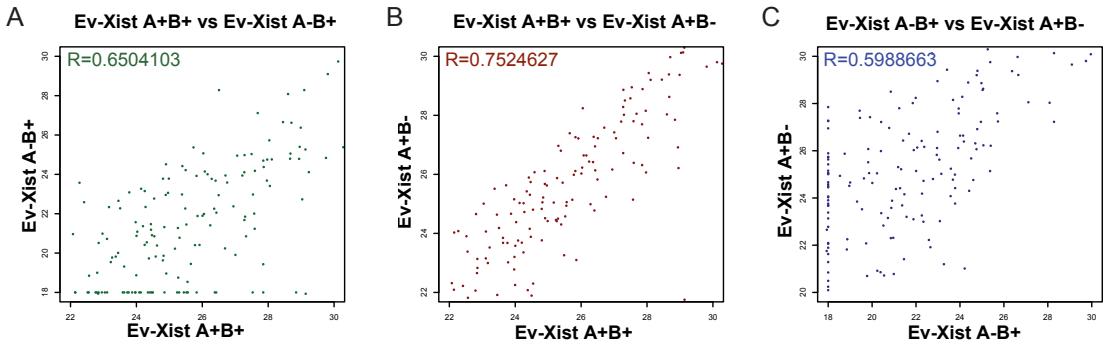


Figure 7.4: Correlation between Xist interactomes. Each plot in the Figure represents the correlation between samples, in pairs: (A) Xist A^+B^+ vs Xist A^-B^+; (B) Xist A^+B^+ vs A^+B^-; (C) Xist A^-B^+vs Xist A^+B^-. Overall Pearson's correlation coefficient R was calculated and reported for each pair.

(2015), and Moindrot *et al* (2015), (Fig.7.3 B). Several proteins, despite being present in both comparisons, showed larger enrichment in the presence of functional repeats. This is best represented in the comparison between Xist A^+B^+ vs Xist A^+B^-, as highlighted in Figure 7.5.

Next, a subtractive comparison of the hits obtained using Xist A^+B^- and Xist, A^+B^+ was performed, with the aim of identifying B-repeat specific binders. It was noticed that the correlation was significantly higher between Xist A^+B^+ and Xist A^+B^-, compared Xist A^+B^+ and Xist A^+B^-, or Xist A^-B^+ and Xist A^+B^-, suggesting that most of the differently bound proteins are A-repeat specific binders (Fig.7.4 compare values on diagonals of A, B, and C). All identified hits that were found to be significantly different between samples in the presence of functional A- or B-repeats, (highlighted in red in Figure 7.5), are listed in Table 7.1 and Table 7.2, respectively.

Among the most prominent A-repeat binders is Rbm15, which was recently shown to be involved in Xist-mediated silencing [Moindrot et al. 2015], possibly via recruiting the m^6A machinery to Xist RNA [Patil et al. 2016]. Additionally, Luc7l3, a rather poorly characterised RNA-binding factor, which was also found to be important for Xist-mediated silencing [Moindrot et al. 2015], and Myef2 (Myelin

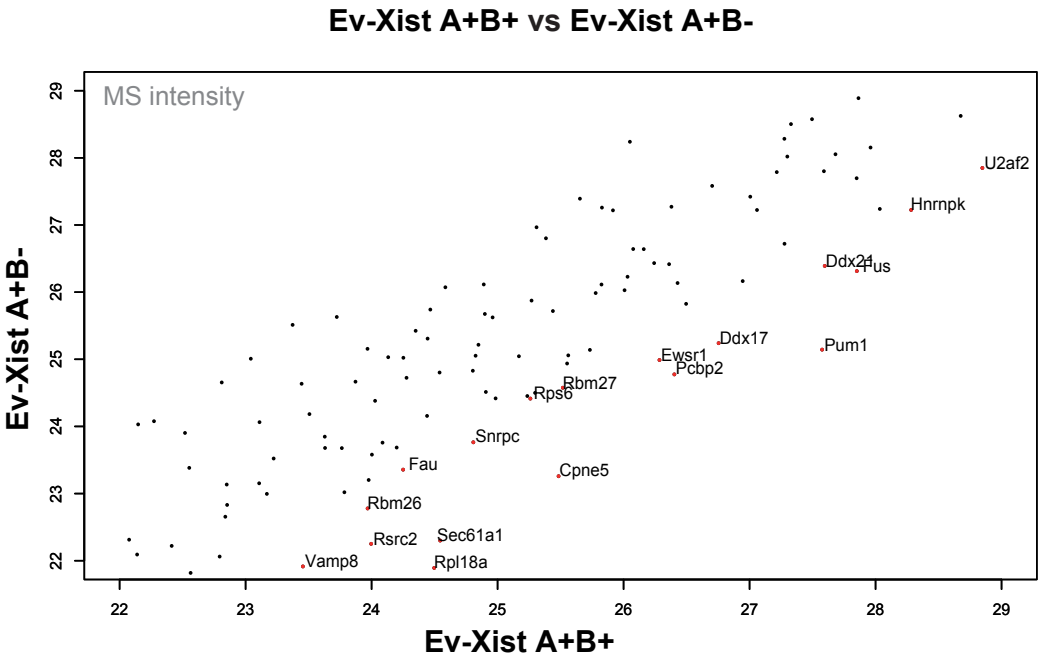


Figure 7.5: B-repeat binding factors. Scatter plot representing changes in abundance of factors interacting with Xist A^+B^+ vs Xist A^+B^-. Red dots represent all significantly different hits between samples. Values on both axes represent MS abundance.

Expression Factor 2), also known to interact with Xist [McHugh et al. 2015, Chu et al. 2015], were present. The identification of known Xist interactors amongst the A-repeat interactome suggests that the method employed is valid.

In the case of the B-repeat interactome, the most prominent hit was the protein hnRNP-K, known to bind to C-rich sequences as exemplified by the Xist B-repeats. Interestingly, hnRNP-K was also identified as an Xist binding partner in the interactome analysis of Chu *et al* (2015), and moreover knock-down of hnRNP-K was shown to reduce recruitment of PRC1 and PRC2 by Xist RNA [Chu et al. 2015]. On the other hand, it should be noted that hnRNP-K binding is also found in the absence of functional B-repeats, albeit at lower levels.

Table 7.1: A-repeat specific binding factors. Summary of factors that were found to be highly significantly enriched in Xist A^+B^+ over Xist A^-B^+, therefore *bona fide* A-repeats binders. All factors with a q-value $\leq$0.03 are listed.

Protein ID	Differential q-value	Ranking by abundance
Rbm15	0.0162602	119
Lin28a	0.0172414	114
Dkc1	0.0191083	92
Myef2	0.0194805	71
Luc7l3	0.0196562	36
Rbmx2	0.0199683	62
Sf3b2	0.02	31
Pcbp1	0.0204082	63
Ssb	0.0206186	26
Snrpd1	0.0229885	97
Rbms1	0.0234419	102
Snrnp70	0.0236244	56
Snrpd3	0.0239619	65
Srek1	0.024	5
Srsf11	0.0244364	64
Poldip3	0.025641	37
Pum1	0.0259841	30
Thrap3	0.0263455	55
Hnrnpm	0.0263652	12
Rsrc2	0.0264803	7
Snrpb	0.0265035	121
Stau1	0.0266243	69
L1td1	0.0269511	74
Sf3b1	0.0274638	78

Table 7.2: B-repeat specific binding factors. Summary of factors that were found to be significantly enriched in Xist A^+B^+ over Xist A^+B^-, therefore *bona fide* B-repeats binders. All factors with a q-value $\leq$0.5 are listed.

Protein ID	Differential q-value	Ranking by abundance
hnRNP-K	0.0262972	15
Ewsr1	0.0437105	42
Poldip3	0.047875	104
Pum1	0.0494194	24
Gar1	0.0598769	116
Slc2a3	0.0602632	25
Cpne5	0.076436	58
Fus	0.0949298	19
Rsrc2	0.113271	94
Sec61a1	0.122742	77
Pcbp2	0.130888	39
Rbm27	0.131799	57
Ddx21	0.135467	22
U2af2	0.148221	11
Rbm26	0.169879	96
Snrpc	0.181994	74
Ddx3y	0.191398	109
S100a14	0.197912	84
Ddx17	0.19817	35
Fau	0.352694	87
Rpl17	0.365113	89
S100a16	0.39876	115
Rpl18a	0.423265	79
Rps6	0.442404	64

7.3 Discussion

In this Chapter, Xist interactomes were generated by UV-crosslinking nuclear extract with synthetic constructs harbouring combinations of either non-functional or intact A- and B-repeats. This approach identified many of the previously published Xist binders/interactors (Fig.7.3). Interestingly, all but two of the Xist-interacting proteins isolated in McHugh *et al* (2015), who employed a similar UV-crosslinking approach *in vivo* [McHugh et al. 2015], were identified. The absent proteins are LBR and Spen, both of which are known A-repeat interactors [Chu et al. 2015, Chen et al. 2016], and therefore expected to bind to the construct used. A possible explanation is that post-transcriptional modification of the RNA (that was not reproduced *in vitro*) is required for binding of these factors. Indeed, it was recently shown that Xist RNA has m^6A methylation sites, some of which are located within the A-repeat region [Patil et al. 2016].

Interestingly, many among the identified factors showed a strong preference for binding Xist in the presence of a particular motif, either the A- or the B-repeats (Figure 7.5), thus defining A-repeat and B-repeat specific interactomes (Table 7.1, Table 7.2). Conversely, other Xist binding partners were found to interact with the RNA rather promiscuously, not exclusively in the presence of a specific element, with only few factors being absolutely depleted in the presence of non-functional elements. This observation may suggest that Xist repeats act *in vivo* as nucleation sites for proteins that are potentially able to interact with other segments of Xist. In this respect Xist may act as a molecular sponge, leading to high enrichment of functional protein complexes over its surface.

Importantly, some of the identified A-repeat specific binders, were already shown to have a functional link to gene silencing, or at least to interact with the A-repeats [Moindrot et al. 2015, McHugh et al. 2015, Chu et al. 2015]. An example is Rbm15, which was identified in recently published interactome studies as a *bona fide* Xist interactor [McHugh et al. 2015, Chu et al. 2015, Moindrot et al. 2015, Minajigi et al.

2015]. Recently, eCLIP data published by Van Nostrand *et al* (2016), confirmed that Rbm15 binds specifically to the A-repeats of human XIST [Van Nostrand et al. 2016].

The identification of known Xist binders amongst the A-repeat interactome, also suggests that the method employed is valid, and adds to the confidence of the newly-identified B-repeat interactome, where the nuclear matrix protein hnRNP-K was found among the most significantly enriched factors (Table 7.2). hnRNP-K was previously identified as a mouse Xist binder [Chu et al. 2015], and reported to be highly enriched on the human XIST RNA molecule [Van Nostrand et al. 2016]. Of note, eCLIP data showed that hnRNP-K indiscriminately favours C-rich regions of Xist, with no clear binding preference. There was nevertheless a distinct peak corresponding to the B-repeats [Van Nostrand et al. 2016]. Given that hnRNP-K is known to bind C-rich nucleic acids non-specifically [Makeyev et al. 2002], it is not surprising that it might also bind to neighbouring C-rich regions of the RNA.

In conclusion, work in this Chapter established a new *in vitro* method to generate differential Xist interactomes, that were broadly consistent with previous *in vivo* studies [McHugh et al. 2015, Chu et al. 2015]. In the future this system could be further employed to rapidly detect precise binding sites for key factors, potentially at nucleotide resolution. On the other hand, obvious drawbacks derive from the artificial setup of the experimental system. Specifically, folding of the RNA *in vitro*, will not necessarily reflect the real secondary structures existing *in vivo*. Additionally, artificial RNAs do not bear post-transcriptional modifications (for instance, m^6A methylation), that could confer structural constrains as well as novel binding sites to RNA molecules. All these limitations, together with the sensitivity limits of mass spectrometry, may undermine identification of crucial RNA binding factors or underestimate their abundance.

8

hnRNP-K function in the context of XCI

Contents

8.1 Introduction 121
8.2 Results 124
8.2.1 hnRNP-K facilitates Polycomb recruitment on the Xi . 124
8.2.2 hnRNP-K is abundant in the nucleus, but not enriched on the Xi 124
8.2.3 hnRNP-K interacts with PRC1 127
8.2.4 The KI domain of hnRNP-K is required for interaction with PRC1 128
8.2.5 Forced tethering of hnRNP-K to the inactive chromosome partially rescues the ΔXN Xist Polycomb phenotype . . 129
8.3 Discussion 134

8.1 Introduction

Given that hnRNP-K was the best candidate for a B-repeat interacting RBP (Chapter 7, Section 7.2), its potential role in PRC1 and PRC2 recruitment was further tested. The protein hnRNP-K is encoded in mammals by a conserved gene harbouring at least five alternative spliced transcripts. Among these, the most prominent variant is translated into a 65kDa protein that was first discovered as a component of the 40S heterogenous nuclear ribonucleoprotein (hnRNP) particle [Bomsztyk et al. 1997, Matunis et al. 1992]. Subsequent studies, however, indicated

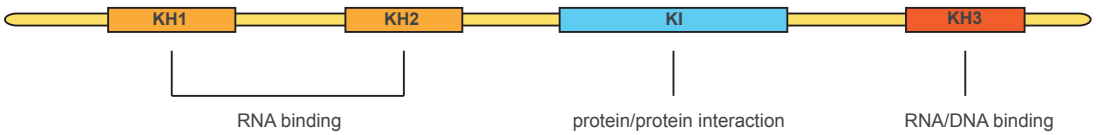


Figure 8.1: Structure of the protein hnRNP-K. Mouse hnRNP-K comprises three separate KH domains, involved in nucleic acid binding; and a KI domain, that was proposed to mediate protein/protein contacts [Makeyev et al. 2002].

a wider nuclear distribution for hnRNP-K, suggesting that a large fraction of the protein might be independent from hnRNP particles [Dejgaard et al. 1994]. Although a specific function for the protein, either in the hnRNP particle or in any other protein complex, has not yet been defined.

hnRNP-K contains the evolutionarily conserved KH repeats (Fig.8.1) that are found in RNA-binding proteins among species as diverse as *Escherichia coli* and *Saccharomyces cerevisiae* [Gibson et al. 1993]. Via its KH domains (Fig.8.1), hnRNP-K can bind RNA (with a preference for Poly(C) strands) and single- and double-stranded DNA, in a sequence-selective fashion [Bomsztyk et al. 1997]. On the other hand, hnRNP-K protein-protein interactions are in large part mediated by its KI region (Fig.8.1). Interestingly, unlike KH domains, the KI domain is not conserved in other Poly(C) binding proteins, suggesting that it is required to mediate hnRNP-K specific functions [Makeyev et al. 2002].

Functionally, hnRNP-K has been implicated in a wide and diverse range of signal transduction and gene expression pathways [for a review, Bomsztyk et al. 2004]. A proposed dual RNA/DNA-binding capacity together with the presence of a protein-protein interaction motif, suggests that it might serve as a docking platform at sites of nucleic acid directed processes where it senses, connects and perhaps initiates signals from multiple cascades.

hnRNP-K was found to be among the high-confidence Xist interacting partners using ChIRP, an *in vivo* RNA formaldehyde crosslinking approach [Chu et al. 2015].

In detail, Chu *et al* (2015), identified hnRNP-K was among the factors highly enriched on Xist compared to other non-coding RNAs. On the other hand, hnRNP-K was not found by McHugh *at al*, who employed UV rather than formaldehyde as a cross-linking agent, arguing against direct contact with Xist. However, eCLIP data show binding of hnRNP-K to human XIST RNA [Van Nostrand et al. 2016]. The same study failed to identify a consensus binding motif in XIST, with hnRNP-K showing a broad distribution along the C-rich regions of the RNA, especially the B- and C-repeats. This observation raises the question of whether hnRNP-K may bind RNA in a sequence-oriented way, or whether its binding, although tenacious, is relatively non-specific. Both scenarios have been reported in the literature, where hnRNP-K was described to engage in a sequence-selective fashion with RNA, but also to capture Poly(C) RNA non-specifically [reviewed in Bomsztyk et al. 2004].

Chu *et al* (2015), tested the effects of hnRNP-K depletion on Xist-mediated *in cis* silencing and Polycomb recruitment, following siRNA-mediated knock-down [Chu et al. 2015]. Unlike other nuclear matrix proteins identified in the study (such as hnRNP-U/Saf-A), knock-down of hnRNP-K did not affect Xist-coating of the Xi, but a defect in the silencing ability of Xist was reported, and, importantly, a reduction in the PRC1 and PRC2-mediated histone modifications H2AK119ub1 and H3K27me3. This latter finding is in agreement with previous reports in which hnRNP-K was shown to interact with members of the Polycomb group of proteins Ezh2 and Eed [Denisenko et al. 1997, Cao et al. 2002]. On the other hand, it should also be noted that hnRNP-K is a common contaminant in mass spectrometry experiments due to non-specific binding to agarose and magnetic beads used to purify protein complexes from nuclear extracts [Mellacheruvu et al. 2013]. Hence, the functional relevance of these interactions remains to be defined.

8.2 Results

8.2.1 hnRNP-K facilitates Polycomb recruitment on the Xi

In the previous Chapters, a convergence of phenotypes was reported for the B-repeats of Xist, that were found to mediate both Polycomb recruitment (Chapter 5, Section 5.2), and hnRNP-K binding (Chapter 7, Section 7.2), suggesting potential interdependence of these events. In Chu *et al* (2015), knock-down of hnRNP-K resulted in impaired PRC1 and PRC2-mediated histone modifications on the inactive chromosome [Chu et al. 2015]. To confirm this result, CRISPR/Cas9-mediated knock-out of hnRNP-K was attempted. However, despite extensive attempts, only heterozygous knock-out lines were recovered, suggesting that hnRNP-K is required for viability of ES cells. As an alternative, a siRNA knock-down approach was used. Maximum hnRNP-K knock-down efficiency (70%) was obtained after 24h of siRNA treatment, as shown in Figure 8.2 A. Cells were therefore treated with siRNAs, and Xist expression was induced at the same time for 24 hours (Figure 8.2 B). When Polycomb-mediated histone marks were stained, the percentage of Polycomb domains overlapping Xist clouds dropped, compared to the control (Fig.8.2 C, D, E). The measure of the effect is consistent with the level of knock-down that was achieved (Fig.8.2 B, D).

8.2.2 hnRNP-K is abundant in the nucleus, but not enriched on the Xi

To determine if hnRNP-K is enriched over Xist domains, an eGFP tagged version of the protein was stably expressed in ES cells, also harbouring a doxycyline-responsive FL Xist. After establishment of 3 independent cell lines with eGFP fused to either the N- or C-terminus of the protein, Xist and eGFP were co-stained (Fig.8.3). Enrichment of hnRNP-K could be detected in the nucleus, suggesting that the protein is targeted to the right compartment, but no clear accumulation over the inactive domain was observed in either case (Fig.8.3 A, B). This result does not rule out that there is association of hnRNP-K with Xist, as the high abundance of the protein in the nucleus may mask localisation to specific sites.

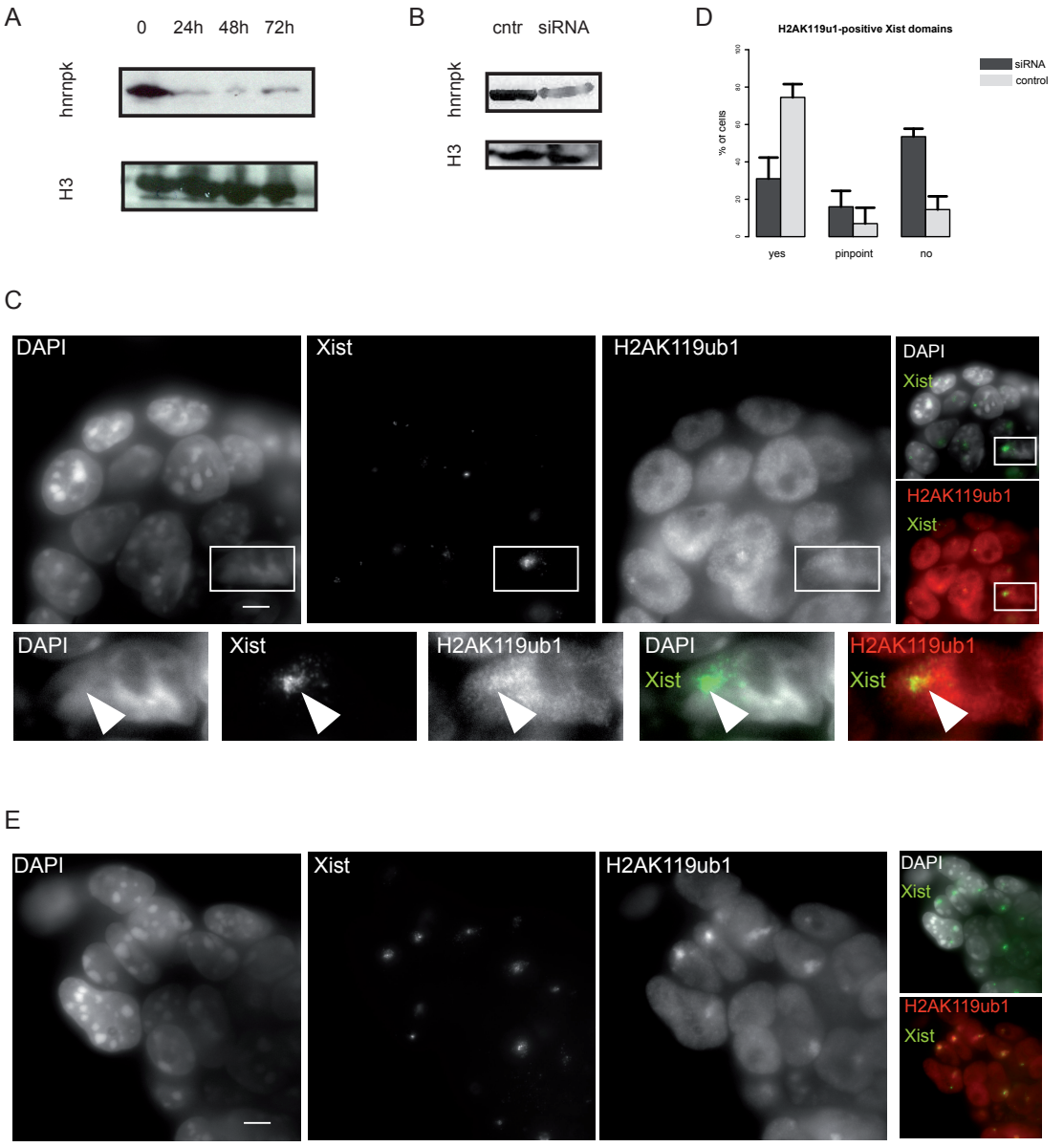


Figure 8.2: siRNA knock-down of hnRNP-K. (A) Cells were treated with a siRNA smartpool for 24h, 48h, and 72h, harvested in equal number for each condition, and checked for the level of expression of hnRNP-K by western blot. Histone H3 was used as a loading control. (B) The 24h time point was selected for the actual experiment, and the specificity of the siRNA-mediated reduction of hnRNP-K was confirmed. (C) In the siRNA treated cells, reduction of H2AK119ub1 domains was observed by combining RNA-FISH and immunofluorescence. The effect was quantified in (D). Error bars represent variation over two biological replicates. (E) Normal levels of H2AK119ub1 were observed in the mock control. Wide-field images represent average of 10 stacks. Zoomed images are single stacks. Arrows point to Xist domains. Scale bar is 5μm.

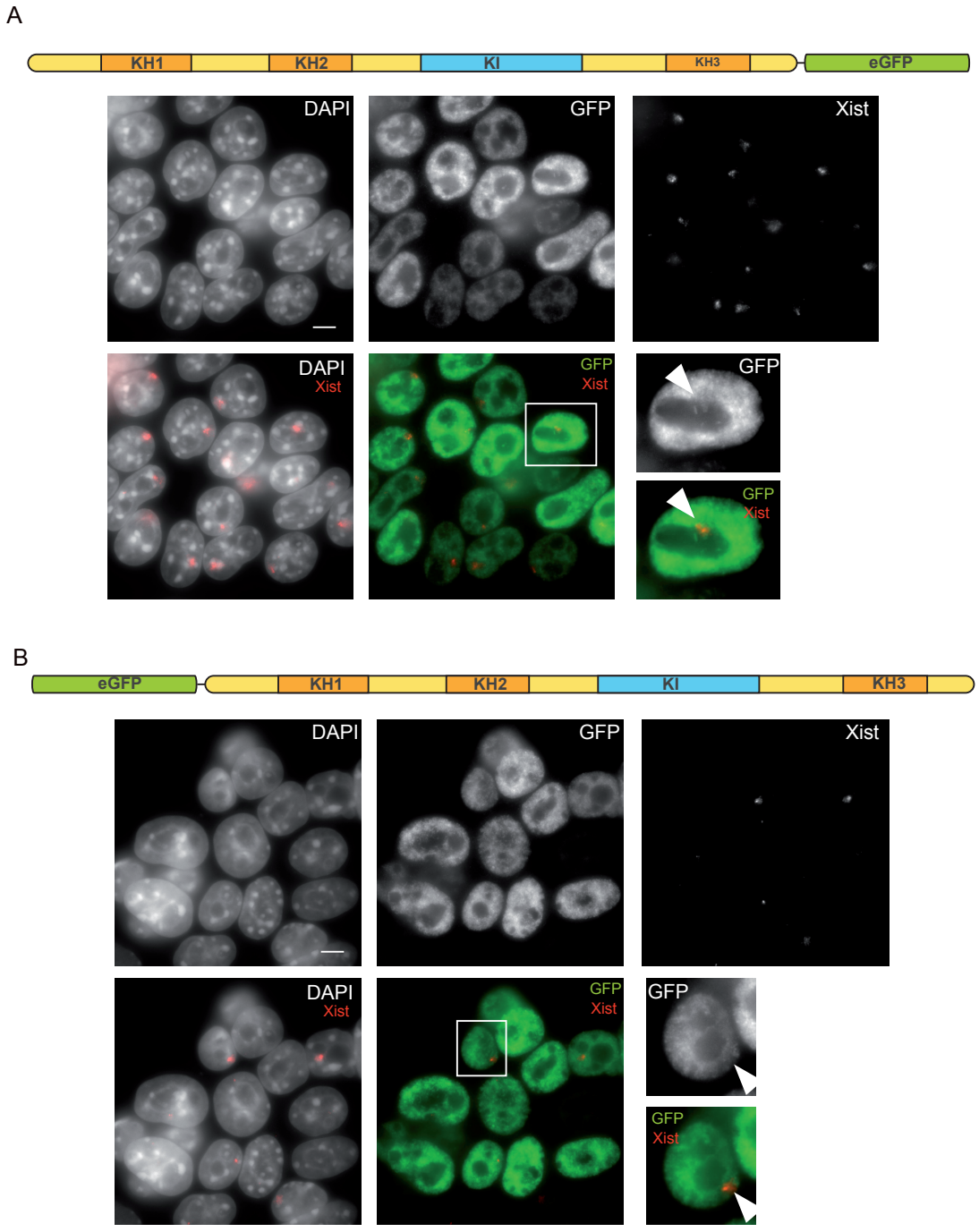


Figure 8.3: hnRNP-K nuclear localisation. Two different constructs were stably transfected in mouse ES cells expressing an inducible form of FL Xist, where the eGFP cDNA was fused either to the N-terminus (A), or to the C-terminus (B) of hnRNP-K. In both cases nuclear enrichment of the protein was observed by immunofluorescence combined with RNA-FISH. Wide-field images represent average of 10 stacks. Zoomed images are single stacks. Scale bar is 5μm.

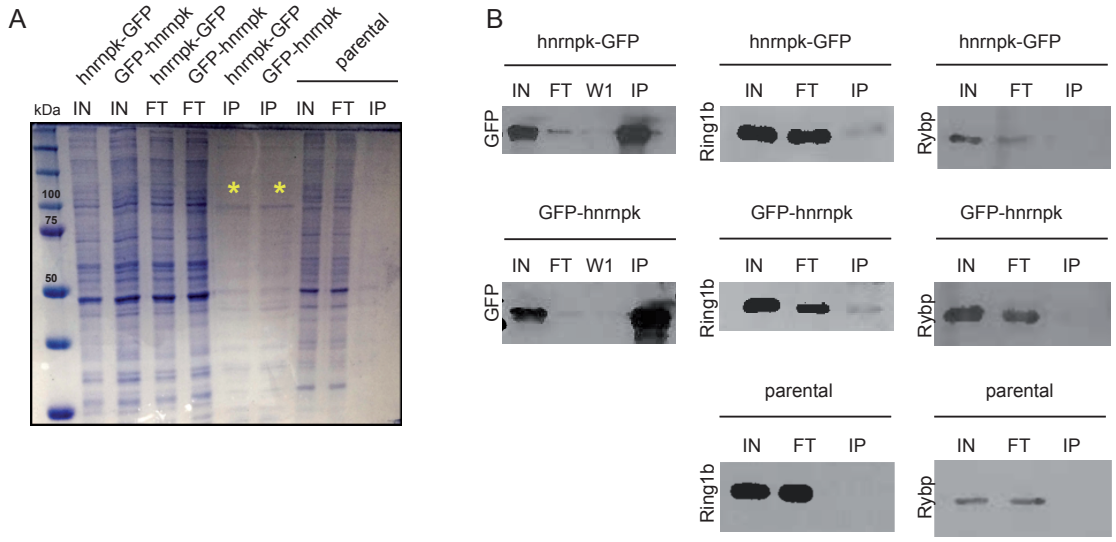


Figure 8.4: hnRNP-K and Polycomb co-IPs. hnRNP-K fused to eGFP was pulled down from nuclear extracts with GFP-nanobodies coupled to magnetic beads. IP efficiency is shown in (A) on a Coomassie Brilliant Blue stained gel, 0.1% of the input (IN), 0.1% of the flow-through (FT), and 10% of the pull down (IP) were loaded. Yellow asterisks represent the band corresponding to the molecular weight of tagged hnRNP-K. (B) Co-IP assays on the same fractions. The parental cell line expresses endogenous hnRNP-K only.

8.2.3 hnRNP-K interacts with PRC1

Although hnRNP-K is not obviously enriched on the Xi, it was nevertheless interesting to test whether Polycomb and hnRNP-K interact. In order to answer this question, co-immunoprecipitation assays were performed, pulling down hnRNP-K from the nuclear extract of cells expressing an eGFP-tagged version of the protein, and assaying for the presence of Polycomb components in the eluted sample by western blot. GFP-nanobodies were used in these experiments, as they are compatible with washes of hnRNP-K-bound beads in high salt conditions, crucial to ensure specificity of the assay. Correct pull-down of the protein was checked by Coomassie Brilliant Blue staining (Fig.8.4 A), and blotted for Polycomb components (Fig. 8.4 B). Interestingly, presence of Ring1b was consistently observed in the eluate, albeit weakly, whereas it could not be detected in any of the negative controls. A second PRC1 subunit, Rybp, however, was not detected. PRC2 subunits were also not detected in the eluate (not shown). Interaction of hnRNP-K with PRC1 and not with PRC2 is consistent with recent findings demonstrating

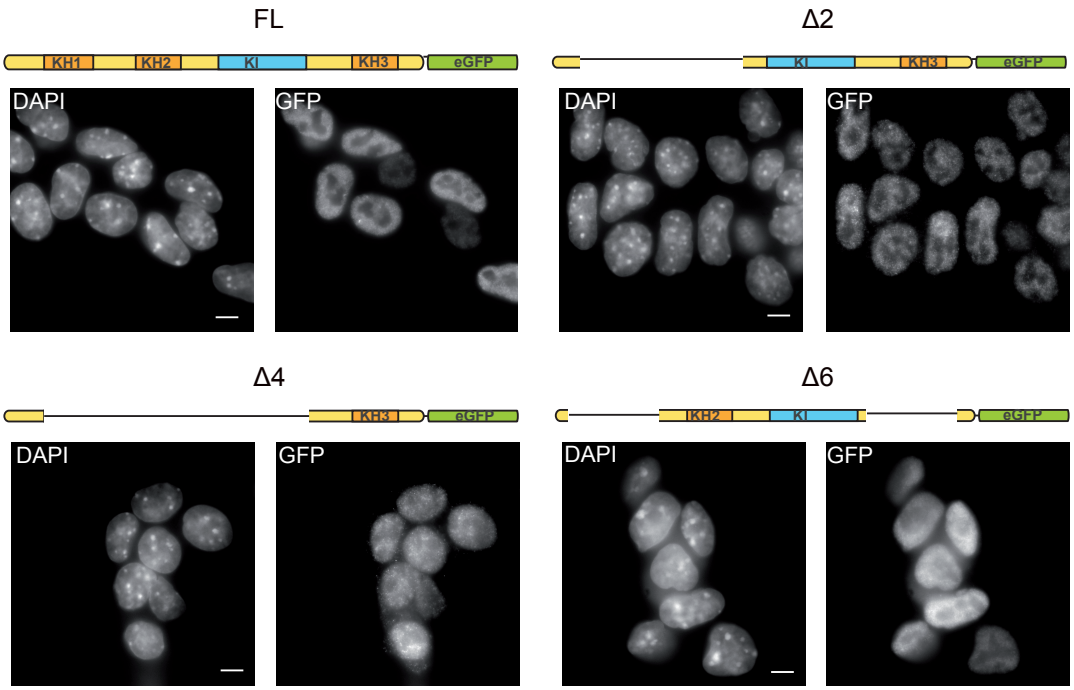


Figure 8.5: hnRNP-K deletion analysis. All hnRNP-K deletion variants, depicted in the schematics, were stably expressed in mouse ES cells, and tested for level of expression and localisation of the protein by immunofluorescence. All constructs are nuclear and expressed at similar levels. Images represent single stacks. Scale bar is 5μm.

that the non-canonical Pcgf3/5-PRC1 complex initiates Polycomb recruitment by Xist RNA [Almeida et al. 2016].

8.2.4 The KI domain of hnRNP-K is required for interaction with PRC1

As a further step to validate the interaction between hnRNP-K and PRC1, the specificity of the reported contact was tested by serially deleting domains of hnRNP-K, and repeating the co-IP assay. Deletion variants were first checked for correct localisation in the nucleus and expression by immunofluorescence (Fig.8.5). Previously reported co-immunoprecipitation of Ring1b could be consistently repeated with three of the four tested constructs (Fig.8.6 A, B). Analysis of the overlap of the reported regions, indicates that the KI domain, deleted in the Δ4 construct, specifically mediates the interaction between hnRNP-K and PRC1, as no PRC2 components were absorved in any other IP (Fig.8.6 C). This is in agreement

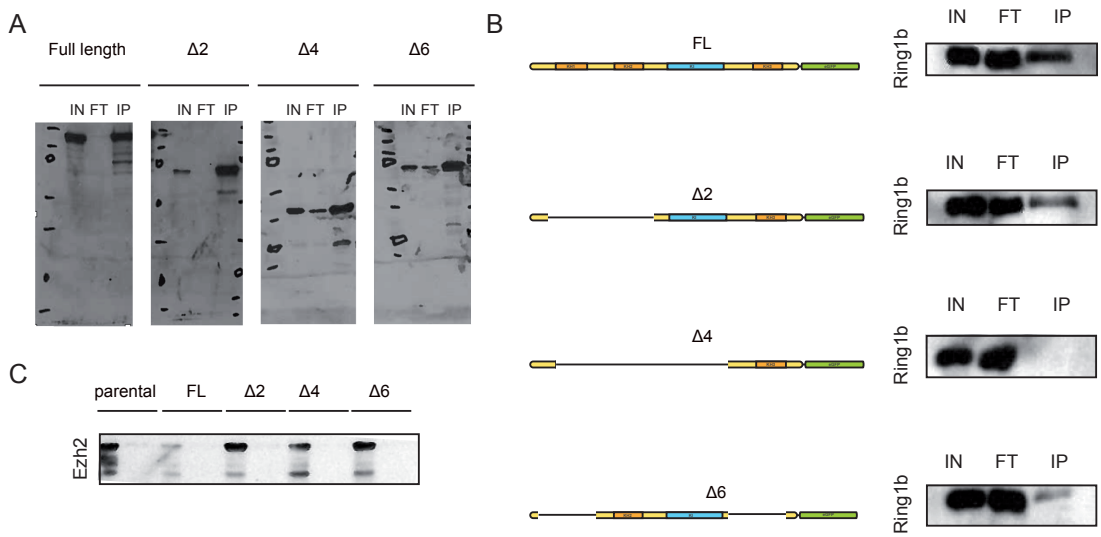


Figure 8.6: hnRNP-K deletion variants and Polycomb co-IPs. Four different hnRNP-K constructs were used. (A) IP efficiency is shown in western blot. 0.1% of the input (IN), 0.1% of the flow-through (FT), and 10% of the pull down (IP) were loaded for each of the named constructs. (B) Co-IP assays on the same fractions for each construct represented in the schematics. (C) Comparative Co-IP on the same amount of protein loaded for each of the named constructs, and for the parental cell line, with just endogenous levels of hnRNP-K.

with previous reports that suggest that the KI domain is involved in protein-protein interactions and not in nucleic acid binding [Makeyev et al. 2002].

8.2.5 Forced tethering of hnRNP-K to the inactive chromosome partially rescues the ΔXN Xist Polycomb phenotype

In order to unequivocally demonstrate that hnRNP-K plays a role in PRC1 recruitment by Xist RNA, through binding of the B-repeats, the protein was artificially tethered to ΔXN Xist, which does not normally recruit Polycomb complexes (Fig.8.7 A). A tethering system, based on binding of BglG fusion protein to Bgl stem loops [Moindrot et al. 2015], was adapted by fusing inducible ΔXN Xist with a Bgl stem loop RNA motif, that can be efficiently and specifically recognised by the protein BglG, expressed in fusion with hnRNP-K (Fig.8.7 B). An eGFP ORF was also fused to the protein, in order to better track its expression.

Independent cell lines, co-expressing ΔXN Xist-Bgl and hnRNP-K-BglG-eGFP,

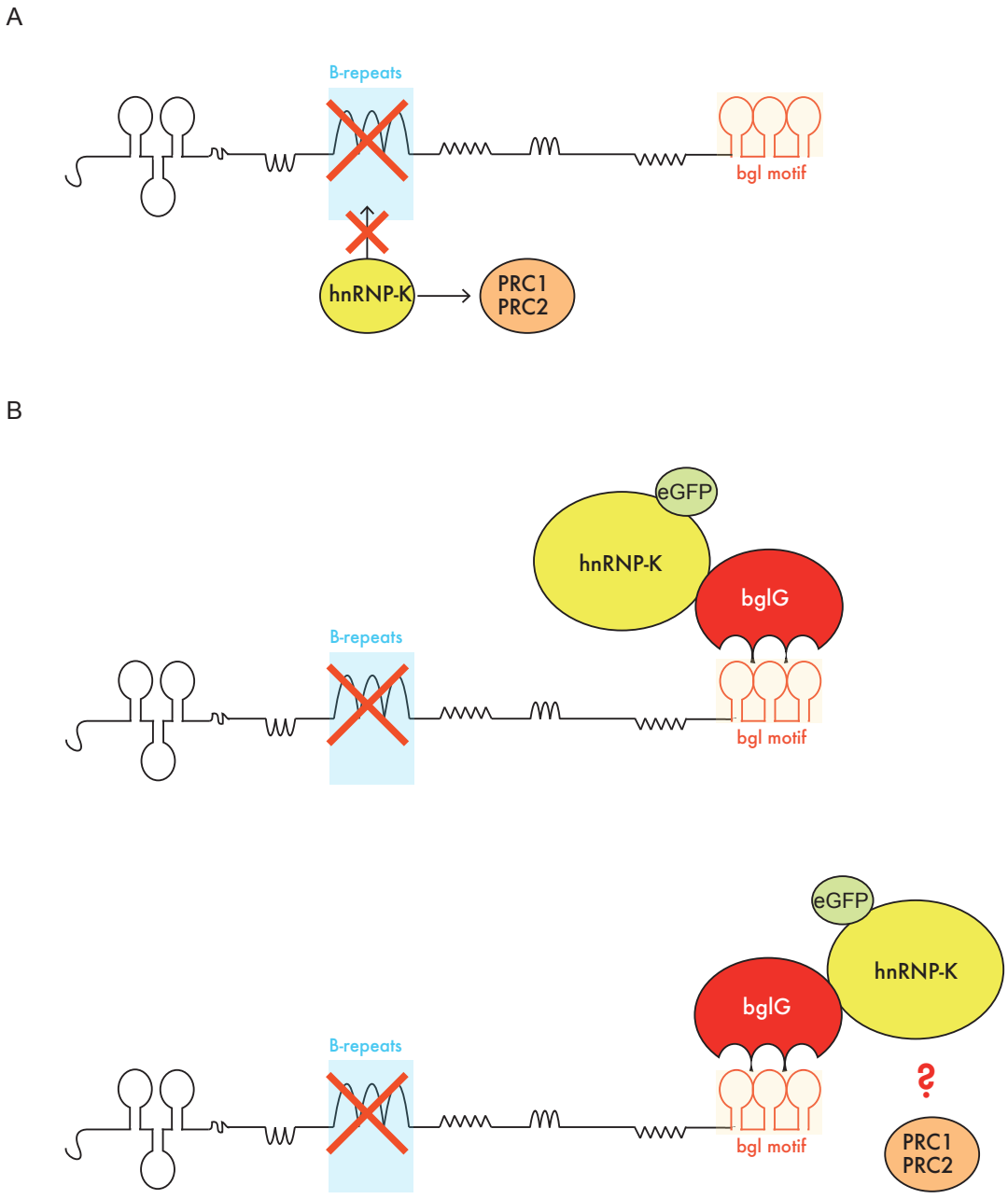


Figure 8.7: Artificial tethering of hnRNP-K to Xist. (A) In the absence of the B-repeats, hnRNP-K is not able to bind Xist and therefore to mediate Polycomb recruitment. (B) hnRNP-K was fused to a BglG protein able to bind a Bgl motif previously linked to Xist lacking the XN region. Artificial tethering of hnRNP-K to Xist RNA is predicted to mediate artificial accumulation of hnRNP-K on the Xi, and consequent rescue of Polycomb recruitment.

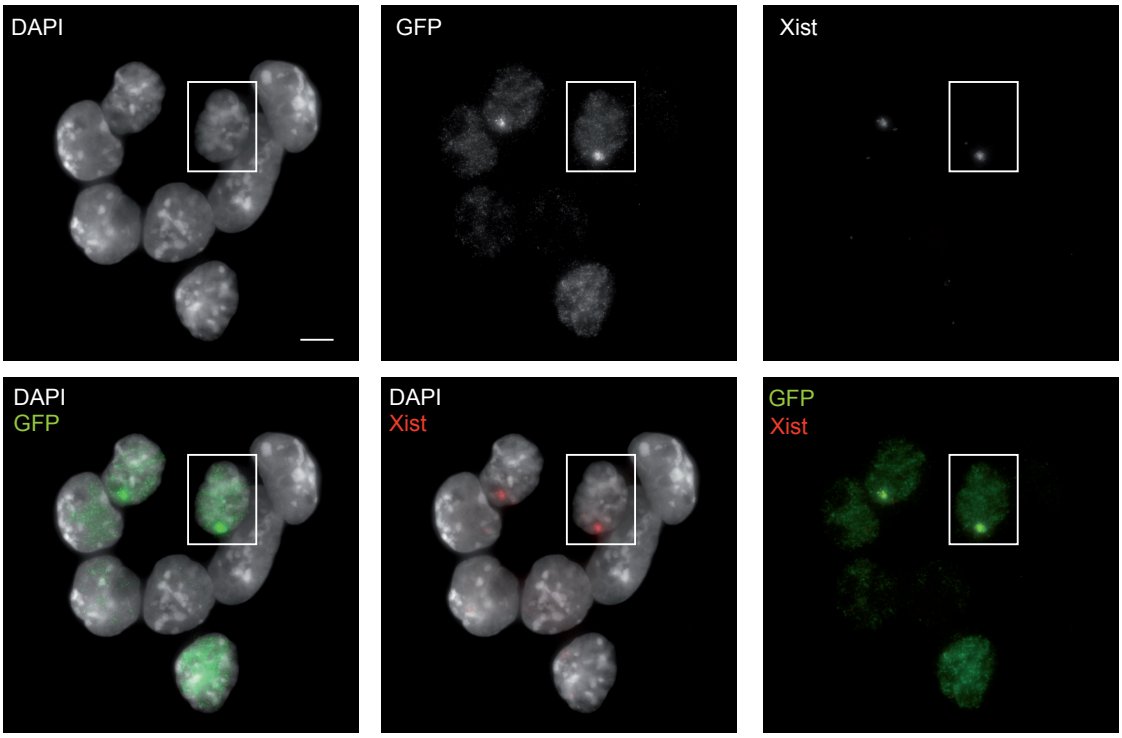


Figure 8.8: Forced accumulation of hnRNP-K on the Xi. Immunofluorescence combined with RNA-FISH in the Bgl tethering system, shows that hnRNP-K-eGFP was forced to accumulate over the inactive chromosome upon 24 hours doxycycline induction. Wide-field images are average of 10 stacks. The boxed cell is an example of eGFP/Xist-positive nucleus. Scale bar is 5μm.

were generated. After doxycycline induction of cells for 24 hours, accumulation of eGFP in live cell imaging could be observed, demonstrating hnRNP-K targeting to Xist RNA (not shown). Co-localisation between the eGFP signal and Xist was further confirmed by combining immunofluorescence and RNA-FISH, as shown in Figure 8.8. Rescue of Polycomb recruitment was tested by co-staining eGFP and either H2AK119ub1 or H3K27me3. In both cases, a subset of the total population of cells with eGFP domains clearly showed a rescue phenotype (Figure 8.9 A, B). As a parallel control, BglG was also fused to Pcgf3, a core subunit of the complex that initiates *de novo* Polycomb recruitment [Almeida et al. 2016], and independent cell lines were generated, expressing ΔXN Xist-Bgl and Pcgf3-BglG-eGFP. Also in this case, discrete eGFP domains could be observed upon doxycycline treatment in live cell imaging, although a much lower number of eGFP-positive cells was detected, suggesting that the accumulation of eGFP *foci* on the inactive

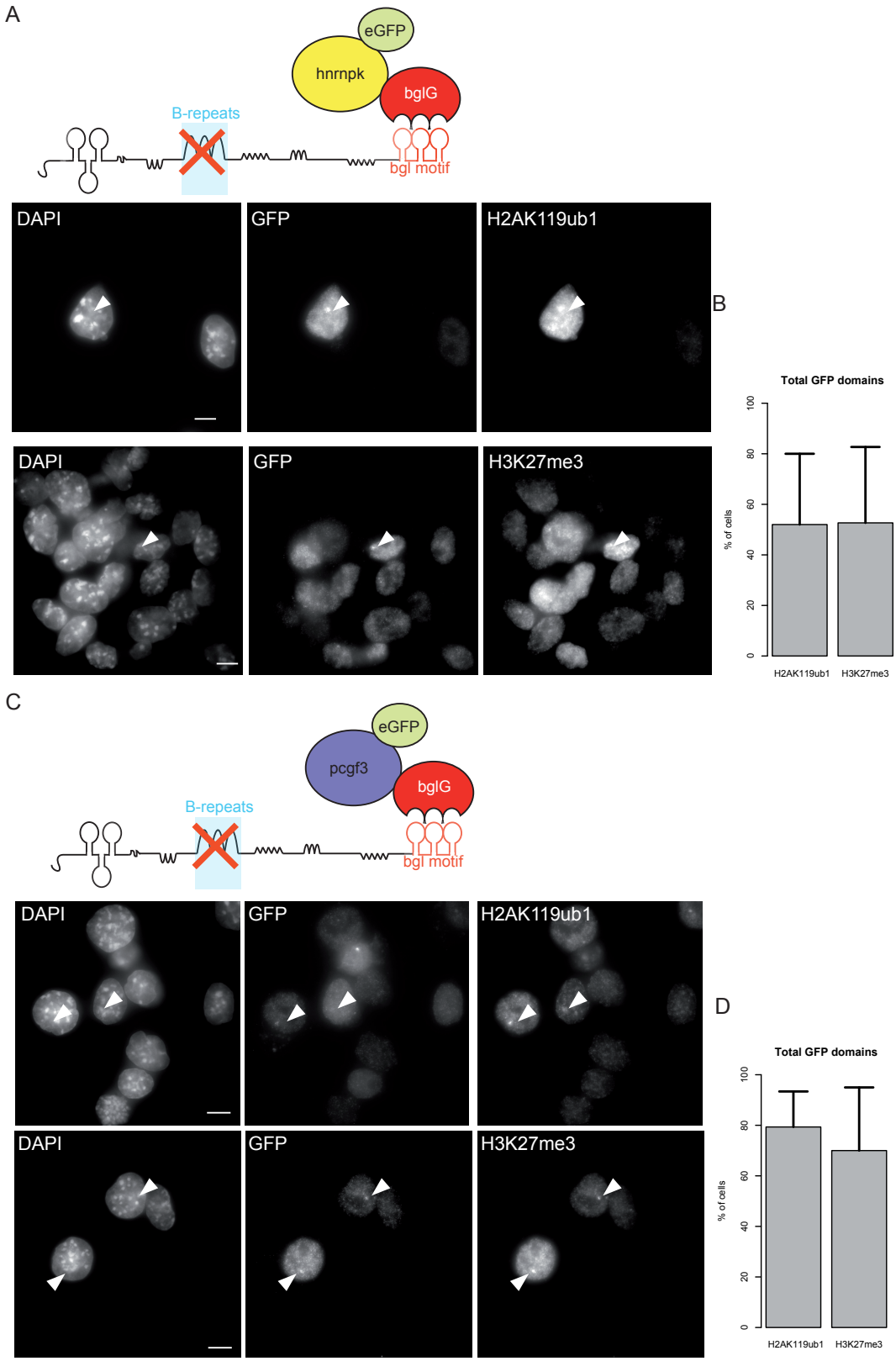


Figure 8.9: Rescue of Polycomb upon forced hnRNP-K and Pcgf3 tethering to Xist domains. Using the Bgl tethering system, hnRNP-K (A) and Pcgf3 (C) were tethered on the inactive chromosome to force recruitment of PRC1 and PRC2, as shown in a number of cells by immunofluorescence using H2AK119ub1 and H3K27me3 staining. Quantification is reported in (B) and (D), respectively. Error bars represent sd across three biological replicates. Wide-field images represent average of 10 stacks. Arrows indicate eGFP domains. Scale bar is 5μm.

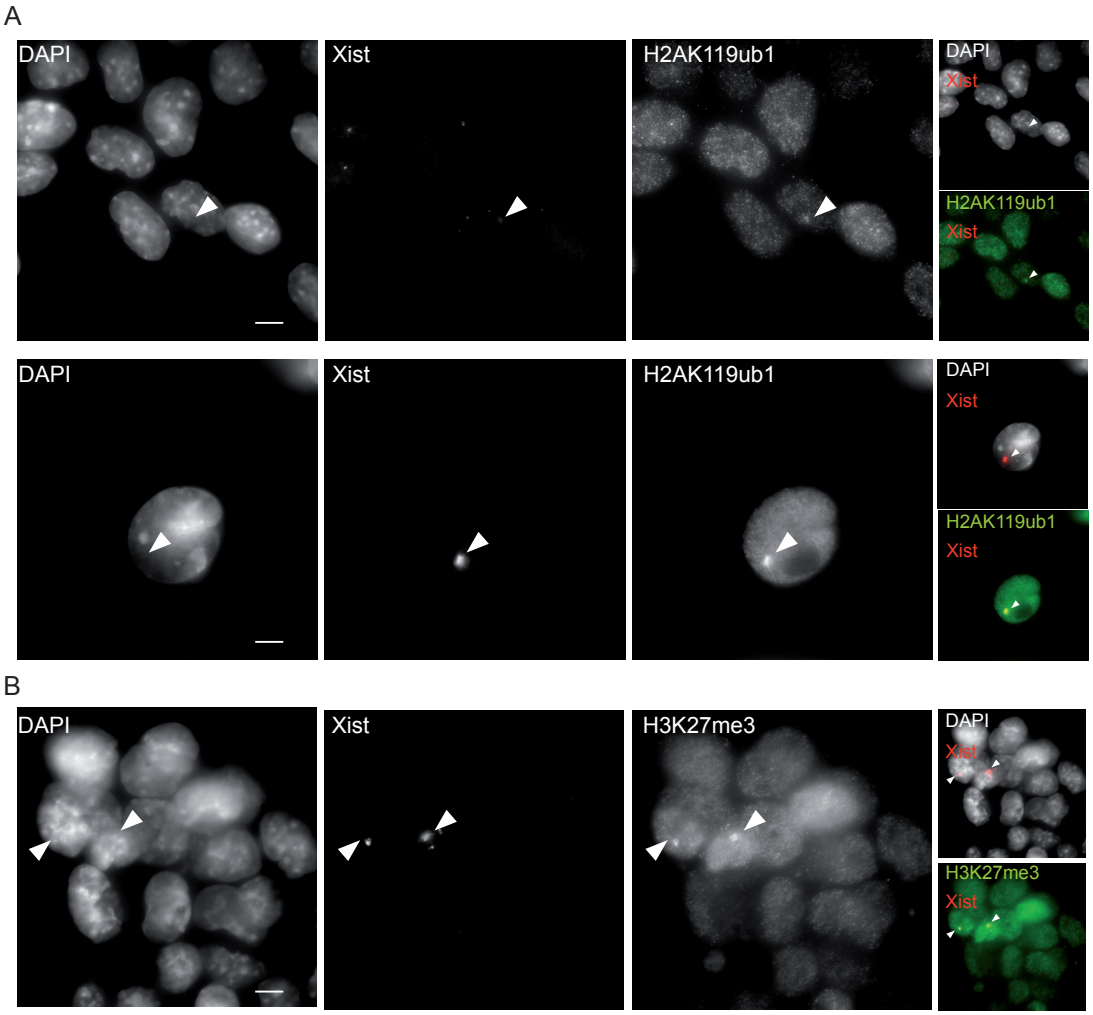


Figure 8.10: Rescue of Polycomb upon forced hnRNP-K tethering to Xist domains. Using the Bgl tethering system, hnRNP-K was recruited to the inactive chromosome and determined recruitment of PRC1 (A) and PRC2 (B) in a number of cells, as shown by immunofluorescence for H2AK119ub1 and H3K27me3, combined with RNA-FISH. Wide-field images are averages of 10 stacks. Arrows indicate Xist domains. Scale bar is 5μm.

chromosome is rather transient, and difficult to capture through formaldehyde fixation, employed for the immunofluorescence protocol. Although the absolute number of cells with eGFP domains was lower than for the previous experiment with hnRNP-K, the co-localisation between eGFP and H2AK119ub1 and H3K27me3 was significant (Figure 8.9 C, D). Co-staining of Xist RNA and PRC1/2-deposited marks was also performed. The experiment confirmed the result obtained by immunofluorescence (Fig.8.10 A, B).

8.3 Discussion

hnRNP-K is a nuclear matrix protein abundant in the nucleus of vertebrates, although its biological function is not well-defined [Mellacheruvu et al. 2013]. Recently, hnRNP-K was found to interact with Xist RNA in two independent studies, that made use of different techniques. Chu *et al* (2015), conducted a ChIRP study aimed at producing an exhaustive list of Xist interacting partners, and found hnRNP-K as one of the most abundant factors specifically enriched on Xist *in vivo* [Chu et al. 2015], while in a vast large-scale eCLIP analysis of nuclear RNA-binding proteins, hnRNP-K was also found among the binders of human XIST [Van Nostrand et al. 2016]. More precisely, the binding seemed to span the B-/C-repeats region of XIST. These results were confirmed by the *in vitro* data presented in this study (Chapter 7, Section 7.2), where hnRNP-K was found to be enriched in the presence of the B-repeats. Importantly, Chu *et al* (2015), also reported a defect in Polycomb activity on the Xi, following depletion of hnRNP-K. This result was confirmed here, with a significant reduction of Polycomb-mediated histone modifications observed upon loss of hnRNP-K following siRNA treatment. Importantly, the effect was observed for both PRC2 and PRC1, that was recently shown to initiate *de novo* recruitment of Polycomb on the Xi [Almeida et al. 2016].

Interaction between hnRNP-K and PRC1 but not PRC2 was also confirmed by coIP experiments, where hnRNP-K was shown to specifically associate with Ring1b and not with PRC2 subunits. It should be noted that the interaction between hnRNP-K and PRC1 appeared rather weak, given that Rybp did not coIP with hnRNP-K. On the other hand, previous data generated in the lab, showed interaction between Pcgf3/5-PRC1 and hnRNP-K, and not between PCGF2/4-PRC1 and hnRNP-K by mass spectrometry (unpublished). Given the role of Pcgf3/5-PRC1 in initiating recruiting of Polycomb on the Xi [Almeida et al. 2016], this result adds confidence that the observed interactions are specific, which could be questioned considering the tendency of hnRNP-K to bind non-specifically to agarose and magnetic beads. Additionally, the observed interaction between hnRNP-K and

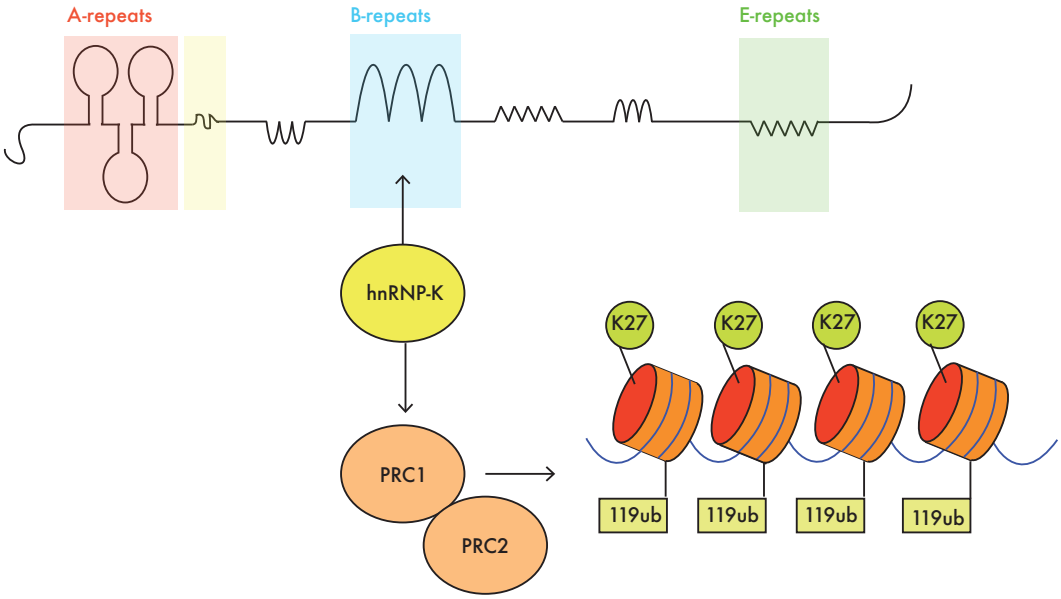


Figure 8.11: hnRNP-K-mediated Polycomb recruitment via the B-repeats of Xist. According to the proposed model, hnRNP-K binding to the B-repeats of Xist, and subsequent spreading onto the neighbouring C-rich region of the RNA, determines *de novo* recruitment of PRC1 to the Xi. This triggers subsequent recruitment of PRC2 and canonical PRC1. Once recruited, the Polycomb proteins depose their distinctive histone marks, H2AK119ub1 and H3K27me3, and establish a silent chromatin state.

PRC1 was specifically lost in the absence of the KI domain of hnRNP-K, previously proposed to mediate protein-protein contacts [Makeyev et al. 2002].

To further validate the interaction between hnRNP-K and Xist RNA, an eGFP-tagging strategy was attempted. However, hnRNP-K was not visibly enriched over the inactive chromosome following Xist expression, appearing abundantly present in all nuclear compartments with the exception of the nucleolus. This result does not rule out that there is association of hnRNP-K with Xist as the high abundance of the protein in the nucleus may mask localisation to specific sites. This is the case of other well-characterised Xist-binders, such as Rbm15 and Spen [Moindrot et al. 2015].

Whilst genetic and biochemical evidence are not definitive in support of a

model whereby hnRNP-K mediates the contact between Xist and PRC1 (Figure 8.11), the system to artificially tether the protein to Xist in the absence of the B-repeats does add considerable insights to the hypothesis. The observed rescue of Polycomb-mediated histone modifications on the Xi, suggests that hnRNP-K is indeed involved in the B-repeat mediated *de novo* recruitment of PRC1. However, at this stage, it cannot be ruled out that other factors could also contribute to recruiting PRC1 by Xist either independently or together with hnRNP-K.

In the future it will be interesting to further explore the mechanism of the interaction of hnRNP-K with both PRC1 and Xist. In particular, conditional knock-out ES cell lines could be used to test the effects of complete depletion of the protein on Polycomb recruitment. On the other hand, the failure to observe enrichment of eGFP-tagged hnRNP-K on the inactive chromosome (when not artificially tethered) suggests that few molecules are bound to Xist at any given time. It will therefore be interesting to better understand the mode of hnRNP-K binding to the B-repeats. This could also help to generalise hnRNP-K contribution to Polycomb recruitment genome-wide, via interaction with other non-coding RNAs.

9 General discussion

Contents

9.1 Overview 137
9.2 The A-repeats of Xist are essential for gene silencing . 139
9.3 Ev-Xist recapitulates key functions of Xist RNA . . . 140
9.4 The B-repeats of Xist mediate PRC1/2 recruitment . 141
9.5 Polycomb contributes to Xist-mediated gene silencing 142
9.6 The nuclear matrix protein hnRNP-K mediates PRC1 interaction with the B-repeats of Xist RNA 143

9.1 Overview

Since Mary Lyon's first formulation of the "X inactivation hypothesis" [Lyon 1961], many advances have been made towards the understanding of the molecular mechanism of X chromosome inactivation (XCI) [Brockdorff et al. 1992, Brown et al. 1992, Penny et al. 1996, Lee et al. 1997]. XCI was confirmed to be an efficient way of ensuring dosage compensation between genders, through random transcriptional silencing of either the maternal or paternal X chromosome of female individuals early in development [reviewed in Harper 2011]. The general mechanism was shown to be conserved among therian mammals, although in marsupials the master regulator, the lncRNA RSX, evolved independently [Pask et al. 2000]. Evolutionary convergence,

notably in utilising *in cis* acting lncRNAs, makes it tempting to speculate that some of the silencing strategies involved in XCI could be employed elsewhere in the genome. In fact, recent studies have highlighted that lncRNAs, derived from pseudogenisation or insertion events, are prevalent in most species. Many of these molecules may be non-functional, but nevertheless provide for evolution of new functions, facilitated by a high rate of non-disruptive mutations [Rinn 2014]. Most likely, this is how Xist and Rsx independently acquired their modular function in co-ordinating the inactivation process in eutherians and metatherians, respectively [Grant et al. 2012].

In this perspective, a better understanding of the mechanism of Xist-mediated silencing becomes relevant on two different levels: it can help shed light on the regulation and detailed mechanism of X inactivation *per se*, but it could also provide a paradigm to study potentially conserved regulatory strategies epigenetically coordinated by a lncRNA.

Soon after Xist was first characterised and identified as the main trigger of the inactivation process, many attempts were made to define the modes of interaction between Xist sequence elements and the molecular machinery responsible for epigenetic silencing of the chromosome [reviewed in Gendrel et al. 2014]. However, this has been challenging for a number of reasons. First of all Xist is a complex molecule, rich in local secondary structures and repeats, and tightly bound to the nuclear matrix. This means native Xist RNA is refractory to extraction, preventing the application of classical biochemical analyses. Second, different assays performed in a range of cellular systems or partially based on *in vitro* experiments, have revealed inconsistencies, with conflicting reports on key functional elements and the factors they bind. Third, most of the phenotypes observed in the context of classical genetic studies rely on indirect or partial assays to test for loss- and gain-of-functions, as allele-specific silencing, which is the outcome of the inactivation process, is difficult to measure at the technical level. For all these reasons, a paradoxical divergence

has emerged between detailed knowledge of many of the mechanisms intervening to ensure efficient inactivation, and poor understanding of the interplay between these pathways and the sequence/structure of Xist RNA.

The main goal of this thesis work, was to address some of these shortcomings using a comprehensive deletion study of Xist RNA, coupled to quantitative analysis of the outcome of allele specific silencing. The experimental strategy, was based on the work of Wutz *et al* (2002), which defined the importance of the A-repeats of Xist for its silencing activity. In this study, ectopic expression of Xist was driven in an autosomal system, in which the alleles could be distinguished based on a well-known set of interspecific SNPs, allowing Xist-mediated silencing defects to be directly quantified through high-throughput approaches. In this perspective, the Repression Score (RS) was defined, as a means to quantify Xist silencing ability for all expressed genes located *in cis* with the Xist transgene.

9.2 The A-repeats of Xist are essential for gene silencing

Using the aforementioned strategy, it was confirmed that Xist lacking the A-repeats (ΔSX Xist) is largely incapable at establishing gene silencing, even at low levels (Chapter 3). In contrast to published findings [Chaumeil 2006, Wutz et al. 2002], induction of ΔSX Xist, revealed persistence of RNA Polymerase II and other active chromatin marks, at least to some degree, on the inactive chromosome territory (Chapter 3). Defects in early exclusion of the transcriptional machinery could be explained through the binding to the A-repeats of factors recruiting the Hdac/Ncor complex, responsible for histone deacetylation and RNA PolII inhibition [McHugh et al. 2015]. In particular, Rbm15, a SPOC-domain containing RNA-binding factor [Shi et al. 2001], was found to specifically contact the A-repeats in a recent eCLIP analysis [Van Nostrand et al. 2016]. Moreover, silencing assays performed following Rbm15 knock-down in ES cells, showed a defect, albeit not dramatic, in Xist-mediated silencing [Moindrot et al. 2015]. The mild effect of

Rbm15 knock-down compared to the more prominent phenotype observed upon deletion of the A-repeats, could be explained by partial redundancy with Spen, another SPOC-domain containing protein shown elsewhere to bind specifically to the A-repeats of Xist [Lu et al. 2016]. In fact, Spen knock-down also results in partial silencing defects in gene silencing [McHugh et al. 2015].

In the future, it will be important to investigate whether knock-out of either Rbm15 or Spen, or both factors, recapitulates the dramatic loss of gene repression obtained with the ΔSX deletion. Specifically, RS analysis could be used to directly compare Xist-mediated silencing between different knock-out cell lines, with Xist trangene located in the same locus. Alternatively, other yet-unknown pathways could also be involved in the early exclusion of the transcriptional machinery and gene-silencing. Very recently, Rbm15 was suggested to represent the molecular adaptor between Xist RNA and the m^6A machinery, responsible for RNA methylation, by interacting with Wtap, and the m^6A methyltransferase like 3 (Mettl3) [Patil et al. 2016]. However, whether the role of Rbm15 in gene-silencing can be fully ascribed to its involvement in establishing m^6A methylation on Xist, is still not clear.

9.3 Ev-Xist recapitulates key functions of Xist RNA

One of the most striking findings from this work, is that a short form of Xist RNA, Ev-Xist, comprising only part of exon I between the A- and the C-repeats, recapitulates Xist function. Previously shown to properly localise on the Xi, Ev-Xist indeed partially overcame the functional redundancy of FL Xist in cell killing assays [Wutz et al. 2002]. In this work, Ev-Xist was found to recapitulate chromosome wide silencing, albeit less efficiently compared to FL Xist (Chapter 3). This result confirmed the idea that Xist-mediated silencing ability mostly resides in exon I, which can be therefore envisioned as a short functional version of the full length form of Xist, and a suitable tool for *in vitro* assays (Chapter 7).

Previous studies have shown that multiple elements in Xist RNA contribute to *in cis* localisation in a redundant manner, but to date it has not been possible to define individual localisation elements. The small size of Ev-Xist facilitated this task, and indeed it was possible to narrow down two elements, the A-repeats, previously shown to play a role in Xist localisation [Wutz et al. 2002], and a novel 188bp element located immediately downstream. The novel element is distinct from the YY1 binding site previously indicated as being required for proper Xist localisation [Jeon et al. 2011], and moreover coincides with a region to which hnRNP-U binds [Lu et al. 2016]. In the future it will be interesting to identify factors that bind the novel localisation signal and test whether its presence is sufficient to mediate Xist cloud formation in a physiological context, using endogenous FL Xist.

9.4 The B-repeats of Xist mediate PRC1/2 recruitment

Among the early events observed at the onset of the inactivation process, is the establishment of histone tail modifications mediated by both canonical and non-canonical PRC1, and PRC2, responsible for mono-ubiquitylating histone H2A (H2AK119ub1) and methylating histone H3 lysine 27 (H3K27me3), respectively [reviewed in Brockdorff 2013]. The close dependency between Xist expression and Polycomb localisation on the Xi, was first explained in a model whereby the A-repeats of Xist recruit PRC2 directly, and provide the physical tether for its *de novo* accumulation over the Xi, with *de novo* recruitment of PRC2 then signalling recruitment of PRC1 [Zhao et al. 2008]. However, subsequent studies challenged this model. Notably, *in vivo* observations pointed towards a different part of Xist, the XN region, spanning the F-,B-, and C-repeats, being responsible for recruiting PRC2, with an important role for the Jarid2 subunit [da Rocha et al. 2014]. Moreover, unpublished work from the Brockdorff lab has demonstrated that the non-canonical PRC1 complex Pcgf3/5 initiates Polycomb recruitment by Xist RNA, with H2AK119ub1 catalysed by Pcgf3/5 accounting for subsequent recruitment of Jarid2-PRC2 [Almeida et al. 2016]. Here, it was shown that the XN

region of Xist RNA is required to recruit both PRC1 and PRC2, presumably via Pcgf3/5 binding. Moreover, the B-repeats were shown to be specifically required to mediate Polycomb recruitment, as upon their deletion, no PRC1 or PRC2 complexes could be observed in response to Xist RNA expression (Chapter 5).

In the future it will be interesting to further explore the conservation of the mechanism of Polycomb recruitment by B-repeat like sequences of non-coding RNAs. For instance, RSX RNA, the marsupial Xist ortholog, also triggers Polycomb recruitment on the Xi [Grant et al. 2012]. Although no clear homology with the B-repeats of Xist could be observed in the RSX sequence, there are several GC-rich tandem repeats, that could provide similar docking sites. Sequence swapping experiments using Xist and RSX could potentially shed light on the sequence/structure elements of the RNA that facilitate contact with Polycomb during XCI but also in other biological contexts. Interestingly, recent studies showed how Polycomb is involved in maintaining the latent state during HSV-1 viral infection of human sensory neurones, through establishment and maintenance of transcriptional repression of lytic genes [reviewed in Watson et al. 2013]. One of the proposed models for Polycomb recruitment in the HSV-1 infection model, envisions its binding to viral RNA, in which B-repeat like tandem repeats are highly enriched.

9.5 Polycomb contributes to Xist-mediated gene silencing

PRC2 function has been shown to be important for XCI, although principally in a specific extra-embryonic lineage, trophoblast giant cells [Wang et al. 2001]. In the embryo proper there are contradictory findings regarding PRC2 function in gene silencing [Silva et al. 2003, Kalantry et al. 2006, Williams et al. 2011]. Moreover the contribution of PRC1, as well as of both PRC1 and PRC2, has never been assessed. Identifying the B-repeats provided a tool to dissect the impact of Polycomb on Xist-mediated gene silencing. Moreover, by analysing gene silencing chromosome wide using inter-specific mouse ES cells, subtle defects could be detected. In this

study, the RS calculated for ΔXN Xist was lower than for FL Xist, implying partial loss of gene silencing, although not as dramatic as in the case of the ΔSX deletion (Chapter 6). Because calculation of RS for random autosomal integrants is subject to important variables, notably integration site and copy number, the short form of Xist, Ev-Xist, and a variant with the B-repeats deleted, were targeted to a specific locus in single copy. The results obtained using the RS analysis in this experimental setup, confirmed that Polycomb contributes to efficient gene-silencing (Chapter 6). In fact, the observed loss of silencing was dramatic, a surprising finding given that the A-repeats were still present and presumably able to recruit silencing factors, such as Rbm15 and Spen. Among the potential explanations for this apparent inconsistency, two seem more plausible. First, the A-repeat mediated silencing does not function properly in the absence of specific segments of Xist RNA missing in the Ev-Xist construct, or in the context of the B-repeat deletion used in this study. Second, initial silencing mediated by the A-repeats is achieved, but absence of Polycomb proteins leads to prompt reactivation of genes and general failure to maintain the silent state. In order to further investigate these hypotheses, it will be interesting to dissect the exact nature of Polycomb contribution in facilitating gene repression, by analysing other epigenetic features that could be affected by its action, notably chromatin compaction and accessibility.

9.6 The nuclear matrix protein hnRNP-K mediates PRC1 interaction with the B-repeats of Xist RNA

Unpublished work in the lab has shown that Polycomb recruitment on the Xi is initiated by Pcgf3/5-PRC1, with PRC2 recruitment occurring due to chromosome wide deposition of H2AK119ub1 at sites where Xist RNA accumulates. Following non-canonical Pcgf3/5-PRC1 enrichment on the Xi, and early deposition of H2AK119ub1, other non-canonical PRC1 complexes are also recruited, thus reinforcing and amplifying levels of H2AK119ub1, that mediate recruitment of PRC2, which in turn leads to deposition of H3K27me3 across the chromosome,

and eventual recruitment of canonical PRC1 [Almeida et al. 2016]. However, how Pcgf3/5-PRC1 is recruited by Xist RNA remains an open question. The results in this study demonstrate that the B-repeats of Xist mediate this interaction, that it is most likely indirect, via the RNA binding protein hnRNP-K. It will be important to establish whether this interaction represents a paradigmatic mode of Polycomb recruitment by lncRNAs conserved genome-wide, or rather a strategy uniquely evolved for the XCI process.

True originality consists not in a new manner but in a new vision.

— Edith Wharton, *The writing of fiction*

References

Agger, Karl et al. (2007). "UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development". In: *Nature* 449.7163, pp. 731–734.

Agrelo, Ruben et al. (2009). "SATB1 Defines the Developmental Context for Gene Silencing by Xist in Lymphoma and Embryonic Cells". In: *Developmental cell* 16.4, pp. 507–516.

Almeida, Mafalda et al. (2016). "PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation". In: *Manuscript under Revision*.

Aranda, S, G Mas, and L Di Croce (2015). "Regulation of gene transcription by Polycomb proteins". In: *Science Advances* 1.11, pp. 1–15.

Bacher, Christian P et al. (2006). "Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation". In: *Nature Cell Biology* 8.3, pp. 293–299.

Barakat, Tahsin Stefan et al. (2014). "The Trans-Activator RNF12 and Cis-Acting Elements Effectuate X Chromosome Inactivation Independent of X-Pairing". In: *Molecular cell* 53.6, pp. 965–978.

Barr, Murray Llewellyn (1959). "Sex chromatin and phenotype in man." In: *Science* 130.3377, pp. 679–685.

Beard, Caroline et al. (2006). "Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells". In: *Genesis* 44.1, pp. 23–28.

Beletskii, Anton et al. (2001). "PNA interference mapping demonstrates functional domains in the noncoding RNA Xist". In: *Proceedings of the National Academy of Sciences of the United States of America* 98.16, pp. 9215–9220.

Beutler, Ernest (2008). "Glucose-6-phosphate dehydrogenase deficiency: a historical perspective". In: *Blood* 111.1, pp. 16–24.

Blackledge, Neil P et al. (2014). "Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation." In: *Cell* 157.6, pp. 1445–1459.

Blewitt, Marnie E et al. (2008). "SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation". In: *Nature genetics* 40.5, pp. 663–669.

Bomsztyk, Karol, Oleg Denisenko, and Jerzy Ostrowski (2004). "hnRNP K: One protein multiple processes". In: *BioEssays* 26.6, pp. 629–638.

Bomsztyk, Karol et al. (1997). "Diverse molecular interactions of the hnRNP K protein". In: *FEBS Letters* 403.2, pp. 113–115.

Boyer, Laurie A et al. (2006). "Polycomb complexes repress developmental regulators in murine embryonic stem cells". In: *Nature* 441.7091, pp. 349–353.

Brockdorff, Neil (2013). "Noncoding RNA and Polycomb recruitment". In: *RNA* 19.4, pp. 429–442.

Brockdorff, Neil et al. (1991). "Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome". In: *Nature* 351, pp. 1–3.

Brockdorff, Neil et al. (1992). "The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus." In: *Cell* 71.3, pp. 515–526.

Brown, C J et al. (1991a). "A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome." In: 349.6304, pp. 38–44.

Brown, C J et al. (1991b). "Localization of the X inactivation centre on the human X chromosome in Xq13." In: *Nature* 349.6304, pp. 82–84.

Brown, C J et al. (1992). "The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus." In: *Cell* 71.3, pp. 527–542.

Cao, Ru and Yi Zhang (2004). "SUZ12 Is Required for Both the Histone Methyltransferase Activity and the Silencing Function of the EED-EZH2 Complex". In: *Molecular cell* 15.1, pp. 57–67.

Cao, Ru et al. (2002). "Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing". In: *Science* 298, pp. 1039–1043.

Caparros, M L et al. (2002). "Functional analysis of the highly conserved exon IV of XIST RNA." In: *Cytogenetic and Genome Research* 99.1-4, pp. 99–105.

Cerase, Andrea et al. (2014). "Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy". In: *Proceedings of the National Academy of Sciences of the United States of America* 111.6, pp. 2235–2240.

Chadwick, Brian P and Huntington F Willard (2004). "Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome". In: *Proceedings of the National Academy of Sciences of the United States of America* 101.50, pp. 17450–17455.

Chaligné, Ronan and Edith Heard (2014). "X-chromosome inactivation in development and cancer". In: *FEBS Letters* 588.15, pp. 2514–2522.

Chaumeil, J (2006). "A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced". In: *Genes & development* 20.16, pp. 2223–2237.

Chaumeil, J et al. (2002). "Integrated kinetics of X chromosome inactivation in differentiating embryonic stem cells". In: *Cytogenetic and Genome Research* 99.1-4, pp. 75–84.

Chen, C K et al. (2016). "Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing". In: *Science*, pp. 1–9.

Chow, Jennifer C et al. (2010). "LINE-1 Activity in Facultative Heterochromatin Formation during X Chromosome Inactivation". In: *Cell* 141.6, pp. 956–969.

Chu, Ci et al. (2015). "Systematic Discovery of Xist RNA Binding Proteins". In: *Cell* 161.2, pp. 404–416.

Chureau, Corinne et al. (2002). "Comparative Sequence Analysis of the X-Inactivation Center Region in Mouse, Human, and Bovine". In: *Genome Research* 12, pp. 894–908.

Clemson, Christine Moulton et al. (1996). "XIST RNA Paints the Inactive X Chromosome at Interphase: Evidence for a Novel RNA Involved in Nuclear/Chromosome Structure". In: *The Journal of cell biology* 132.3, pp. 259–275.

Clemson, Christine Moulton et al. (2006). "The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences". In: *Proceedings of the National Academy of Sciences of the United States of America* 103.20, pp. 7688–7693.

Cooper, Sarah et al. (2014). "Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment." In: *Cell reports* 7, pp. 1456–1470.

Cooper, Sarah et al. (2016). "Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2". In: *Nature Communications* 7, pp. 1–8.

Cortez, D et al. (2014). "Origins and functional evolution of Y chromosomes across mammals". In: *Nature* 508.7497, pp. 488–493.

Costanzi, Carl and John R Pehrson (1998). "Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals". In: *Nature* 393, pp. 599–601.

Costanzi, Carl et al. (2000). "Histone macroH2A1 is concentrated in the inactive X chromosome of female preimplantation mouse embryos". In: *Development* 127, pp. 2283–2289.

Dejgaard, Kurt et al. (1994). "Identification, Molecular Cloning, Expression and Chromosome Mapping of a Family of Transformation Upregulated hnRNP-K Proteins Derived by Alternative Splicing". In: *Journal of Molecular Biology* 236, pp. 33–48.

Denisenko, Oleg and Karol Bomsztyk (1997). "The Product of the Murine Homolog of the". In: *Molecular and cellular biology* 17.8, pp. 4707–4717.

Dominissini, Dan et al. (2012). "Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq". In: *Nature* 485.7397, pp. 201–206.

Duthie, Sarah M et al. (1999). "Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis". In: *Human Molecular Genetics* 8.2, pp. 195–204.

Engreitz, J M et al. (2013). "The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome". In: *Science* 341.6147, pp. 1237973–1237973.

Escamilla-Del-Arenal, Martin, Simão Teixeira da Rocha, and Edith Heard (2011). "Evolutionary diversity and developmental regulation of X-chromosome inactivation". In: *Human Genetics* 130.2, pp. 307–327.

Escamilla-Del-Arenal, M et al. (2013). "Cdyl, a New Partner of the Inactive X Chromosome and Potential Reader of H3K27me3 and H3K9me2". In: *Molecular and cellular biology* 33.24, pp. 5005–5020.

Fischle, Wolfgang et al. (2003). "Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains". In: *Genes & development* 17, pp. 1870–1881.

Frazer, Kelly A et al. (2007). "A sequence-based variation map of 8.27 million SNPs in inbred mouse strains". In: *Nature* 448.7157, pp. 1050–1053.

Galupa, Rafael and Edith Heard (2015). "ScienceDirect X-chromosome inactivation: new insights into cis and trans regulation". In: *Current Opinion in Genetics & Development* 31, pp. 57–66.

Gendrel, Anne-Valerie and Edith Heard (2014). "Noncoding RNAs and Epigenetic Mechanisms During X-Chromosome Inactivation". In: *Annual Review of Cell and Developmental Biology* 30.1, pp. 561–580.

Gendrel, Anne-Valerie et al. (2013). "Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes." In: *Molecular and cellular biology* 33.16, pp. 3150–3165.

Gibson, Toby J, Julie D Thompson, and Jaap Heringa (1993). "The KH domain occurs in a diverse set of RNA-binding proteins that include the antiterminator NusA and is probably involved in binding to nucleic acid". In: *FEBS Letters* 324.3, pp. 361–366.

Gilbert, Scott F (2000). *Developmental Biology, 6th edition*. Chapter 17. Sunderland (MA): Sinauer Associates.

Gomez, Maria and Neil Brockdorff (2004). "Heterochromatin on the inactive X chromosome delays replication timing without affecting origin usage". In: *Proceedings of the National Academy of Sciences of the United States of America* 101.18, pp. 6923–6928.

Grant, Jennifer et al. (2012). "Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation". In: *Nature* 487.7406, pp. 254–258.

Grijzenhout, Anne et al. (2016). "Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ESCs". In: *Development* 143.15, pp. 2716–2723.

Hall, Lisa L et al. (2002). "An ectopic human XIST gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells". In: *Proceedings of the National Academy of Sciences of the United States of America* 99.13, pp. 8677–8682.

Harper, Peter S (2011). "Mary Lyon and the hypothesis of random X chromosome inactivation". In: *Human Genetics* 130.2, pp. 169–174.

Hasegawa, Yuko and Shinichi Nakagawa (2011). "Revisiting the function of nuclear scaffold/matrix binding proteins in X chromosome inactivation". In: *RNA Biology* 8.5, pp. 735–739.

Hasegawa, Yuko et al. (2010). "The Matrix Protein hnRNP U Is Required for Chromosomal Localization of Xist RNA". In: *Developmental cell* 19.3, pp. 469–476.

Heard, Edith et al. (2001). "Methylation of Histone H3 at Lys-9 Is an Early Mark on the X Chromosome during X Inactivation". In: *Cell* 107, pp. 727–738.

Helbig, Roger and Frank O Fackelmayer (2003). "Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain". In: *Chromosoma* 112.4, pp. 173–182.

Hoki, Y et al. (2011). "Incomplete X-inactivation initiated by a hypomorphic Xist allele in the mouse". In: *Development* 138.13, pp. 2649–2659.

Hore, Timothy A et al. (2007). "The region homologous to the X-chromosome inactivation centre has been disrupted in marsupial and monotreme mammals". In: *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* 15.2, pp. 147–161.

Hung, Tiffany and Howard Y Chang (2010). "Long noncoding RNA in genome regulation". In: *RNA Biology* 7.5, pp. 582–585.

Inoue, Hiroaki, Hiroshi Nojima, and Hiroto Okayama (1990). "High efficiency transformation of *Escherichia coli* with plasmids". In: *Gene* 96, pp. 23–28.

Jeon, Yesu and Jeannie T Lee (2011). "YY1 Tethers Xist RNA to the Inactive X Nucleation Center". In: *Cell* 146.1, pp. 119–133.

Jeppesen, Peter and Bryan M Turner (1993). "The Inactive X Chromosome in Female Mammais Is Distinguished by a Lack of Histone H4 Acetylation, a Cytogenetic Marker for Gene Expression". In: *Cell* 74, pp. 281–289.

Kalantry, Sundeep et al. (2006). "The Polycomb group protein Eed protects the inactive X-chromosome from differentiation-induced reactivation". In: *Nature Cell Biology* 8.2, pp. 195–202.

Kalb, Reinhard et al. (2014). "Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression". In: *Nature* 21.6, pp. 569–571.

Kaneko, Syuzo et al. (2010). "Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA." In: *Genes & development* 24.23, pp. 2615–2620.

Keohane, Ann M et al. (1996). "X-Inactivation and Histone H4 Acetylation in Embryonic Stem Cells". In: *Developmental Biology* 180, pp. 618–630.

Kohlmaier, Alexander et al. (2004). "A Chromosomal Memory Triggered by Xist Regulates Histone Methylation in X Inactivation". In: *PLoS biology* 2.7, e171–13.

Ku, Manching et al. (2008). "Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains". In: *PLOS Genetics* 4.10, e1000242–14.

Lee, J T and R Jaenisch (1997). "Long-range cis effects of ectopic X-inactivation centres on a mouse autosome." In: 386.6622, pp. 275–279.

Lee, J T et al. (1996). "A 450 kb transgene displays properties of the mammalian X-inactivation center." In: *Cell* 86.1, pp. 83–94.

Lee, Jeannie T, Lance S Davidow, and David Warshawsky (1999). "Tsix, a gene antisense to Xist at the X-inactivation centre". In: *Nature genetics* 21, pp. 400–404.

Lee, Min Gyu et al. (2007). "Demethylation of H3K27 Regulates Polycomb Recruitment and H2A Ubiquitination". In: *Science* 318.5849, pp. 447–450.

Leeb, Martin and Anton Wutz (2007). "Ring1Bis crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells". In: *The Journal of cell biology* 178.2, pp. 219–229.

Leeb, M et al. (2010). "Polycomb complexes act redundantly to repress genomic repeats and genes". In: *Genes & development* 24.3, pp. 265–276.

Lindtner, Susan et al. (2006). "RNA-binding Motif Protein 15 Binds to the RNA Transport Element RTE and Provides a Direct Link to the NXF1 Export Pathway". In: *Journal of Biological Chemistry* 281.48, pp. 36915–36928.

Liu, Liang and Xiu-Duo Fan (2014). "CRISPR–Cas system: a powerful tool for genome engineering". In: *Plant Molecular Biology* 85.3, pp. 209–218.

Lu, Zhipeng et al. (2016). "RNA Duplex Map in Living Cells Reveals Higher- Order Transcriptome Structure". In: *Cell* 165.5, pp. 1267–1279.

Lyon, M F (1961). "Gene action in the X-chromosome of the mouse (Mus musculus L.)." In: *Nature* 190, pp. 372–373.

Maduro, Cheryl, Bas de Hoon, and Joost Gribnau (2016). "Fitting the Puzzle Pieces: the Bigger Picture of XCI". In: *Trends in Biochemical Sciences* 41.2, pp. 138–147.

Mak, Winifred et al. (2002). "Mitotically stable association of polycomb group proteins eed and enx1 with the inactive x chromosome in trophoblast stem cells." In: *Current biology* 12.12, pp. 1016–1020.

Mak, Winifred et al. (2004). "Reactivation of the Paternal X Chromosome in Early Mouse Embryos". In: *Science* 303.5658, pp. 666–669.

Makeyev, Aleksandr V and Stephen A Liebhaber (2002). "The poly(C)-binding proteins: A multiplicity of functions and a search for mechanisms". In: *RNA* 8, pp. 265–278.

Makhlouf, Mélanie et al. (2014). "A prominent and conserved role for YY1 in Xist transcriptional activation". In: *Nature Communications* 5, pp. 1–12.

Mank, J E, N Wedell, and D J Hosken (2013). "Polyandry and sex-specific gene expression". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 368.1613, pp. 20120047–20120047.

Margueron, Raphaël and Danny Reinberg (2011). "The Polycomb complex PRC2 and its mark in life". In: *Nature* 469.7330, pp. 343–349.

Marks, H et al. (2009). "High-resolution analysis of epigenetic changes associated with X inactivation". In: *Genome Research* 19.8, pp. 1361–1373.

Matunis, Michael J, W Matthew Michael, and Gideon Dreyfuss (1992). "Characterization and Primary Structure of the Poly(C)-Binding Heterogeneous Nuclear Ribonucleoprotein Complex". In: *Molecular and cellular biology* 12.1, pp. 164–171.

McHugh, Colleen A et al. (2015). "The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3". In: *Nature* 521.7551, pp. 232–236.

Mellacheruvu, Dattatreya et al. (2013). "The CRAPome: a contaminant repository for affinity purification–mass spectrometry data". In: *Nature Methods* 10.8, pp. 730–736.

Migeon, Barbara R (2016). "An overview of X inactivation based on species differences". In: *Seminars in Cell and Developmental Biology* 56, pp. 111–116.

Mikami, Suzuka et al. (2014). "Structural Insights into the Recruitment of SMRT by the Corepressor SHARP under Phosphorylative Regulation". In: *Structure* 22.1, pp. 35–46.

Min, Jingrong, Yi Zhang, and Rui-Ming Xu (2003). "Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27". In: *Genes & development* 17, pp. 1823–1828.

Minajigi, Anand et al. (2015). "A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation". In: *Science* 349.6245.

Moghadam, Hooman K et al. (2012). "W chromosome expression responds to female-specific selection". In: *Proceedings of the National Academy of Sciences of the United States of America* 109.21, pp. 8207–8211.

Moindrot, Benoit et al. (2015). "A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing". In: *Cell reports* 12.4, pp. 562–572.

Monfort, Asun et al. (2015). "Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem Cells". In: *Cell reports* 12.4, pp. 554–561.

Monkhorst, Kim et al. (2008). "X Inactivation Counting and Choice Is a Stochastic Process: Evidence for Involvement of an X-Linked Activator". In: *Cell* 132.3, pp. 410–421.

Mujtaba, Shiraz et al. (2008). "Epigenetic transcriptional repression of cellular genes by a viral SET protein". In: *Nature Cell Biology* 10.9, pp. 1114–1122.

Napoles, Mariana de et al. (2004). "Polycomb Group Proteins Ring1A/B Link Ubiquitylation of Histone H2A to Heritable Gene Silencing and X Inactivation". In: *Developmental cell* 7.5, pp. 663–676.

Nesterova, Tatyana B et al. (2011). "Pluripotency factor binding and Tsix expression act synergistically to repress Xist in undifferentiated embryonic stem cells." In: *Epigenetics & chromatin* 4.1, p. 17.

Ng, Karen et al. (2011). "A system for imaging the regulatory noncoding". In: *Molecular Biology of the Cell* 22, pp. 2634–2645.

Nozawa, Ryu-Suke et al. (2013). "Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway". In: *Nature structural & molecular biology* 20.5, pp. 566–573.

Ohno, Susumo (1967). "Sex Chromosomes and Sex-Linked Genes." In: *New York Springer-Verlag*.

Ohno, Susumo and T S Haushka (1960). "Allocycly of the X-chromosome in tumors and normal tissues." In: *Cancer research* 20, pp. 541–545.

Okamoto, Ikuhiro et al. (2003). "Epigenetic Dynamics of Imprinted X Inactivation During Early Mouse Development". In: *Science* 303.5658, pp. 644–649.

Olins, Ada L et al. (2014). "Lamin B receptor". In: *Nucleus* 1.1, pp. 53–70.

Pasini, Diego et al. (2004). "Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity". In: *The EMBO Journal* 23, pp. 4061–4071.

Pasini, Diego et al. (2010). "JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells". In: *Nature* 464.7286, pp. 306–310.

Pasini, D et al. (2007). "The Polycomb Group Protein Suz12 Is Required for Embryonic Stem Cell Differentiation". In: *Molecular and cellular biology* 27.10, pp. 3769–3779.

Pask, Andrew J, Marilyn B Renfree, and Jennifer A Marshall Graves (2000). "The human sex-reversing ATRX gene has a homologue on the marsupial Y chromosome, ATRY: Implications for the evolution of mammalian sex determination". In: *Proceedings of the National Academy of Sciences of the United States of America* 97.24, pp. 13198–13202.

Patil, Deepak P et al. (2016). "m6A RNA methylation promotes XIST-mediated transcriptional repression". In: *Nature*, pp. 1–25.

Peng, Jamy C et al. (2009). "Jarid2/Jumonji Coordinates Control of PRC2 Enzymatic Activity and Target Gene Occupancy in Pluripotent Cells". In: *Cell* 139.7, pp. 1290–1302.

Penny, Graeme D et al. (1996). "Requirement for Xist in X chromosome inactivation ". In: *Nature* 379, pp. 131–138.

Pertea, Mihaela (2012). "The Human Transcriptome: An Unfinished Story". In: *Genes* 3.4, pp. 344–360.

Piunti, A and A Shilatifard (2016). "Epigenetic balance of gene expression by Polycomb and COMPASS families". In: *Science* 352.6290, aad9780–aad9780.

Plath, Kathrin et al. (2003). "Role of Histone H3 Lysine 27 Methylation in X Inactivation". In: 300.5616, pp. 131–135.

Plath, Kathrin et al. (2004). "Developmentally regulated alterations in Polycomb repressive complex 1 proteins on the inactive X chromosome". In: *The Journal of cell biology* 167.6, pp. 1025–1035.

Platt, Randall J et al. (2014). "CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling". In: *Cell* 159.2, pp. 440–455.

Pullirsch, D et al. (2010). "The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation". In: *Development* 137.6, pp. 935–943.

Rabani, Michal et al. (2011). "Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells". In: *Nature Biotechnology* 29.5, pp. 436–442.

Rastan, Sohila (1983). "Non-random X-chromosome inactivation". In: *Development* 78, pp. 1–22.

Ringrose, Leonie and Renato Paro (2004). "Epigenetic Regulation of Cellular Memory by the Polycomb and Trithorax Group Proteins". In: *Annual Review of Genetics* 38.1, pp. 413–443.

Rinn, J L (2014). "lncRNAs: Linking RNA to Chromatin". In: *Cold Spring Harbor Perspectives in Biology* 6.8, a018614–a018614.

Rinn, John L and Howard Y Chang (2012). "Genome Regulation by Long Noncoding RNAs". In: *Annual Review of Biochemistry* 81.1, pp. 145–166.

Rocha, Simão Teixeira da et al. (2014). "Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome". In: *Molecular cell* 53.2, pp. 301–316.

Ross, Mark T et al. (2005). "The DNA sequence of the human X chromosome". In: *Nature* 434.7031, pp. 325–337.

Sado, Takashi et al. (2000). "X Inactivation in the Mouse Embryo Deficient for Dnmt1: Distinct Effect of Hypomethylation on Imprinted and Random X Inactivation". In: *Developmental Biology* 225.2, pp. 294–303.

Sanulli, Serena et al. (2015). "Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation". In: *Molecular cell* 57.5, pp. 769–783.

Sarma, Kavitha et al. (2010). "Locked nucleic acids (LNAs) reveal sequence requirements and kinetics of Xist RNA localization to the X chromosome". In: *Proceedings of the National Academy of Sciences of the United States of America* 107.51, pp. 22196–22201.

Schoeftner, Stefan et al. (2006). "Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing". In: *The EMBO Journal* 25, pp. 3110–3122.

Schwartz, Yuri B et al. (2006). "Genome-wide analysis of Polycomb targets in Drosophila melanogaster". In: *Nature genetics* 38.6, pp. 700–705.

Senner, C E et al. (2011). "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". In: *Development* 138.8, pp. 1541–1550.

Shevchenko, A I, I S Zakharova, and S M Zakian (2013). "The Evolutionary Pathway of X Chromosome Inactivation in Mammals". In: *Acta Naturae* 5.2, pp. 40–53.

Shi, Yanhong et al. (2001). "Sharp, an inducible cofactor that integrates nuclear receptor repression and activation". In: *Genes & development* 15, pp. 1140–1151.

Silva, Jose et al. (2003). "Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes." In: *Developmental cell* 4.4, pp. 481–495.

Simon, Matthew D et al. (2013). "High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation". In: *Nature* 504.7480, pp. 465–469.

Smeets, Daniel et al. (2014). "Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci". In: *Epigenetics & chromatin* 7.1, pp. 8–27.

Son, J et al. (2013). "Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin". In: *Genes & development* 27.24, pp. 2663–2677.

Spitale, Robert C et al. (2015). "Structural imprints in vivo decode RNA regulatory mechanisms". In: *Nature* 519.7544, pp. 486–490.

Tang, Y Amy et al. (2010). "Efficiency of Xist-mediated silencing on autosomes is linked to chromosomal domain organisation." In: *Epigenetics & chromatin* 3.1, p. 10.

Tavares, Ligia et al. (2012). "RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3." In: *Cell* 148.4, pp. 664–678.

Uranishi, Hiroaki et al. (2009). "The RNA-binding Motif Protein 15B (RBM15B/OTT3) Acts as Cofactor of the Nuclear Export Receptor NXF1". In: *Journal of Biological Chemistry* 286.38, pp. 26106–26116.

Van Nostrand, Eric L et al. (2016). "Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP)". In: *Nature Methods* 13.6, pp. 508–514.

Wang, Hengbin et al. (2004). "Role of histone H2A ubiquitination in Polycomb silencing ". In: *Nature* 431.7010, pp. 873–878.

Wang, Jianbo et al. (2001). "Imprinted X inactivation maintained by a mouse". In: *Nature genetics* 28.4, pp. 371–375.

Watson, James D and Francis H Crick (1953). "Genetical implications of the structure of deoxyribonucleic acid." In: *Nature* 171.4361, pp. 964–967.

Watson, Zachary et al. (2013). "Role of Polycomb Proteins in Regulating HSV-1 Latency". In: *Viruses* 5.7, pp. 1740–1757.

Whitworth, Deanne J and Andrew J Pask (2016). "The X factor: X chromosome dosage compensation in the evolutionarily divergent monotremes and marsupials". In: *Seminars in Cell and Developmental Biology* 56, pp. 117–121.

Williams, L H et al. (2011). "Transcription precedes loss of Xist coating and depletion of H3K27me3 during X-chromosome reprogramming in the mouse inner cell mass". In: *Development* 138.10, pp. 2049–2057.

Worman, Howard J et al. (1988). "A lamin B receptor in the nuclear envelope". In: *Proceedings of the National Academy of Sciences of the United States of America* 85, pp. 8531–8534.

Wutz, A and R Jaenisch (2000). "A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation." In: *Molecular cell* 5.4, pp. 695–705.

Wutz, Anton, Theodore P Rasmussen, and Rudolf Jaenisch (2002). "Chromosomal silencing and localization are mediated by different domains of Xist RNA". In: *Nature genetics* 30.2, pp. 167–174.

Yamada, Norishige et al. (2015). "Xist Exon 7 Contributes to the Stable Localization of Xist RNA on the Inactive X-Chromosome". In: *PLOS Genetics* 11.8, pp. 1–21.

Yen, Ziny C et al. (2007). "A cross-species comparison of X-chromosome inactivation in Eutheria". In: *Genomics* 90.4, pp. 453–463.

Zarnegar, Brian J et al. (2016). "irCLIP platform for efficient characterization of protein–RNA interactions". In: *Nature Methods* 13.6, pp. 489–492.

Zhao, Jing et al. (2008). "Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome." In: *Science* 322.5902, pp. 750–756.

Zolotukhin, Andrei S et al. (2008). "The RNA transport element RTE is essential for IAP LTR-retrotransposon mobility". In: *Virology* 377.1, pp. 88–99.

Zuker, M., D. H. Mathews, and D. H. Turner (1999). "Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide". In: *RNA Biochemistry and Biotechnology*. Ed. by Jan Barciszewski and Brian F. C. Clark, pp. 11–43.