ESCAPE FROM X CHROMOSOME INACTIVATION IS AN INTRINSIC PROPERTY OF THE MOUSE \textit{JARID1C} LOcus

A Dissertation in

Genetics

by

Nan Li

©2009 Nan Li

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

August 2009
The dissertation of Nan Li was reviewed and approved* by the following:

Laura Carrel  
Assistant Professor of Biochemistry and Molecular Biology  
Dissertation Advisor  
Chair of Committee

Sarah K. Bronson  
Associate Professor of Cellular and Molecular Physiology

Sergei A. Grigoryev  
Associate Professor of Biochemistry and Molecular Biology

Kent E. Vrana  
Elliot S. Vesell Professor and Chair of Pharmacology

Teresa L. Wood  
Professor and Vice Chair of Neurology and Neuroscience  
University of Medicine and Dentistry of New Jersey  
Special Member

Gavin Robertson  
Associate Professor of Pharmacology, Pathology, and Dermatology  
Co-chair, Intercollege Graduate Program in Genetics

*Signatures are on file in the Graduate School.
ABSTRACT

Although most genes on one X chromosome in mammalian females are silenced by X chromosome inactivation, some genes “escape” X inactivation and are expressed from both active and inactive Xs. How these escape genes are transcribed from a largely inactivated chromosome is not fully understood, but underlying genomic sequences are thought to be involved. We developed a transgene approach to ask whether an escape locus, that includes the mouse Jarid1c gene, is autonomous or is instead influenced by X chromosome location. Two BAC clones carrying Jarid1c and adjacent X-inactivated transcripts were randomly integrated into mouse XX embryonic stem cells. Four lines with single copy, X-linked transgenes were identified and we established that each integrated into regions that are normally X inactivated. As expected for genes that are normally subject to X inactivation, BAC transgene transcripts Tspyl2 and Iqsec2 were X inactivated. However, allelic expression and RNA/DNA FISH indicate that transgenic Jarid1c escapes X inactivation. Therefore, transgenes at four different locations on the X recapitulate endogenous inactive X expression patterns. We conclude that escape from X inactivation is an intrinsic feature of the Jarid1c locus and functionally delimit the Jarid1c escape domain as the 112 kb representing the maximum overlap of the BACs tested. Additionally, although extensive chromatin differences normally distinguish active and inactive loci, unmodified BACs are capable of directing proper inactive X expression patterns. Therefore, these studies establish that primary DNA sequence alone, in a chromosome position-independent manner, is sufficient to determine X chromosome inactivation status. This transgene approach will enable further dissection of key elements of escape domains and allow rigorous testing of specific genomic sequences on inactive X expression.
TABLE OF CONTENTS

LIST OF TABLES .......................................................................................................................... viii
LIST OF FIGURES ......................................................................................................................... ix
LIST OF ABBREVIATIONS ............................................................................................................ xi
ACKNOWLEDGEMENTS .................................................................................................................. xvi

CHAPTER I: LITERATURE REVIEW .......................................................................................... 1
1.1 X CHROMOSOME INACTIVATION ....................................................................................... 2
  1.1.1 Mammalian sex chromosomes ......................................................................................... 2
  1.1.2 Dosage compensation in mammals .................................................................................... 6
  1.1.3 Random and imprinted X chromosome inactivation ............................................................ 7
  1.1.4 Processes of X chromosome inactivation ............................................................................ 10
  1.1.5 X inactivation center .......................................................................................................... 10
  1.1.6 Epigenetic features of inactive X chromosomes ................................................................. 15
  1.1.7 Regulation on inactive X chromosomes ............................................................................ 16
    1.1.7.1 Inactivation in X;autosome translocations .................................................................. 16
    1.1.7.2 Inactivation on non-X chromosomes carrying an Xist/XIST transgene ....................... 17
    1.1.7.3 Lessons from translocation and transgene studies ..................................................... 18
1.2 ESCAPE FROM X CHROMOSOME INACTIVATION ............................................................ 20
  1.2.1 Discovery of escape genes ................................................................................................. 20
  1.2.2 Distribution and evolution of human escape genes ............................................................. 21
  1.2.3 Species difference of escape genes .................................................................................... 26
  1.2.4 Epigenetic features of escape genes ................................................................................... 27
  1.2.5 Expression pattern and level of escape genes .................................................................... 28
  1.2.6 Escape from spreading or maintenance of X chromosome inactivation ......................... 28
  1.2.7 Biological and medical significance of escape genes ....................................................... 29
1.3 REGULATION OF ESCAPE GENES ON INACTIVE X CHROMOSOMES ..................... 31
  1.3.1 Gene-specific control of escape genes ................................................................................ 31
  1.3.2 Regional control of escape genes ..................................................................................... 34
    1.3.2.1 Repetitive sequences ................................................................................................. 35
1.3.2.1.1 Long interspersed nuclear element 1 ............................ 35
1.3.2.1.2 Other repetitive elements .......................................... 36
1.3.2.2 Chromatin insulator CTCF and boundary elements ........... 37
1.3.2.3 Regional control on the mouse X .................................. 41
1.4 X-LINKED TRANSGENE STUDIES ........................................ 43
1.5 MOUSE JARID1C GENE ....................................................... 50
1.5.1 General features of Jarid1c .................................................. 50
1.5.2 Jarid1c and escape from X chromosome inactivation ........... 51
1.6 A MODEL SYSTEM TO STUDY X CHROMOSOME INACTIVATION .... 55
1.7 RESEARCH AIM AND PLAN .................................................. 57

CHAPTER II: MATERIALS AND METHODS ..................................... 59
2.1 MODIFICATION OF BAC DNA ................................................ 59
  2.1.1 BAC clones ........................................................................ 59
  2.1.2 Pulsed field gel electrophoresis (PFGE) .............................. 59
  2.1.3 Polymerase chain reaction (PCR) .................................... 59
  2.1.4 Recombineering .............................................................. 64
2.2 CELL LINES AND CULTURE CONDITIONS ............................ 64
  2.2.1 ES cell culture ............................................................... 64
  2.2.2 In vitro differentiation .................................................... 65
  2.2.3 Enriched differentiated ES cells ...................................... 65
2.3 TRANSFECTION .................................................................. 65
2.4 SOUTHERN ANALYSIS ........................................................... 66
2.5 FLUORESCENCE IN SITU HYBRIDIZATION (FISH) ............... 66
  2.5.1 FISH probes .................................................................... 66
  2.5.2 DNA FISH ...................................................................... 67
  2.5.3 RNA FISH ...................................................................... 67
  2.5.4 Sequential RNA and DNA FISH ..................................... 68
  2.5.5 Microscopy and image processing ................................. 68
2.6 INVERSE PCR .................................................................... 68
2.7 INACTIVE X EXPRESSION ANALYSIS ..................................... 69
2.7.1 RNA extraction .................................................................69
2.7.2 Quantitative-SNaPshot (Q-SNaPshot) assay ..........................69
2.7.3 Inactive X expression on normal X chromosomes .................72
2.7.4 Inactive X expression of transgenes .................................72

CHAPTER III: ESTABLISHMENT OF MOUSE FEMALE EMBRYONIC
STEM CELL LINES CARRYING AN X-LINKED Jarid1c BAC TRANSGENE ....76
3.1 INTRODUCTION .................................................................76
3.2 RESULTS .............................................................................81
   3.2.1 Generation of BAC transgene constructs .........................81
   3.2.2 Subcloning of stable ES cell lines .................................84
   3.2.3 ES cell transfection .....................................................87
   3.2.4 Screening for X-linked BAC transgenes .........................91
   3.2.5 BAC transgene integrity ..............................................96
   3.2.6 BAC transgene copy number ....................................103
   3.2.7 Transgene expression in undifferentiated ES cells ..........103
   3.2.8 Integration sites of X-linked BAC transgenes .................108
   3.2.9 X inactivation status of transgene integration sites .........108
   3.2.10 Strain origin of transgenic X chromosomes .................115
3.3 DISCUSSION .......................................................................123
   3.3.1 Transgene random integration ..................................123
   3.3.2 Transgene DNA integrity .........................................124
   3.3.3 Transgene copy number ..........................................124
   3.3.4 Transgene integration site effect ...............................125
   3.3.5 Subspecies difference of Xi expression status .............126

CHAPTER IV: AUTONOMOUS EXPRESSION OF Jarid1c BAC TRANSGENES
ON INACTIVATED X CHROMOSOMES .....................................127
4.1 INTRODUCTION .................................................................127
4.2 RESULTS .............................................................................128
   4.2.1 X inactivation of transgenic Igsec2 ............................128
4.2.2 X inactivation of transgenic Tspyl2 .................................................. 131
4.2.3 Allelic expression of transgenic Jarid1c .............................................. 138
4.2.4 Jarid1c transgene expression in single differentiated ES cells .............. 141
4.2.5 Candidate sequences in Jarid1c BACs and integration sites .................. 146
4.3 DISCUSSION .................................................................................................. 158
  4.3.1 Characterization of transgene expression upon X inactivation .............. 158
    4.3.1.1 Expression from active X chromosomes ...................................... 158
    4.3.1.2 X inactivation status of transgenic Jarid1c ................................. 158
  4.3.2 Mechanisms of escape from X inactivation ......................................... 159
    4.3.2.1 Transgene DNA size and escape of transgenic Jarid1c ................. 159
    4.3.2.2 BAC vector sequences and escape of transgenic Jarid1c ............... 159
    4.3.2.3 Chromosome location and escape of transgenic Jarid1c ............... 160
    4.3.2.4 Delineation of the Jarid1c escape domain ................................. 161
    4.3.2.5 Primary DNA sequences and escape from X chromosome
      Inactivation .................................................................................................. 162
  4.3.3 Inactive X expression of endogenous Jarid1c ...................................... 165

CHAPTER V: GENERAL DISCUSSION AND FUTURE DIRECTIONS .......... 167
  5.1 OVERVIEW .................................................................................................. 167
  5.2 MODEL OF ESCAPE FROM X CHROMOSOME INACTIVATION .......... 168
  5.3 X-LINKED TRANSGENE SYSTEM .......................................................... 171
  5.4 BROAD APPLICATION OF THE METHODOLOGY ................................. 174
  5.5 FUTURE DIRECTIONS .............................................................................. 174
    5.5.1 Key regulatory sequences at the Jarid1c locus ................................. 175
    5.5.2 CTCF binding sites at the 5’ end of Jarid1c ...................................... 176

REFERENCES .................................................................................................... 178
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Mouse genes that escape X chromosome inactivation and their human orthologues</td>
<td>25</td>
</tr>
<tr>
<td>1.2</td>
<td>X chromosome inactivation studies of transgenes randomly integrated into the mouse X chromosome</td>
<td>44</td>
</tr>
<tr>
<td>1.3</td>
<td>X chromosome inactivation studies of transgenes targeted at the mouse Hprt locus</td>
<td>46</td>
</tr>
<tr>
<td>2.1</td>
<td>Primers used in construction and characterization of BAC transgenes</td>
<td>62</td>
</tr>
<tr>
<td>2.2</td>
<td>SNPs assayed for inactive X expression</td>
<td>73</td>
</tr>
<tr>
<td>3.1</td>
<td>Establishment of stable subclones of mouse female embryonic stem cells</td>
<td>90</td>
</tr>
<tr>
<td>3.2</td>
<td>Mouse ES cell transfection experiments and screening of transgenic cell lines</td>
<td>97</td>
</tr>
<tr>
<td>3.3</td>
<td>Stability to maintain two X chromosomes in transgenic ES cell lines</td>
<td>98</td>
</tr>
<tr>
<td>3.4</td>
<td>Chromosomal location of transgene integration site</td>
<td>111</td>
</tr>
<tr>
<td>3.5</td>
<td>SNPs included in the integration site sequences of line A314</td>
<td>121</td>
</tr>
<tr>
<td>4.1</td>
<td>Chromosomal location and X inactivation status of genes near transgene integration sites</td>
<td>149</td>
</tr>
<tr>
<td>Figure</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1.1</td>
<td>X chromosome evolution in mammals</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>X chromosome inactivation in female mouse embryos</td>
<td>8</td>
</tr>
<tr>
<td>1.3</td>
<td>X inactivation center in mouse</td>
<td>12</td>
</tr>
<tr>
<td>1.4</td>
<td>Human X chromosome inactivation profile</td>
<td>23</td>
</tr>
<tr>
<td>1.5</td>
<td>Models of escape from X chromosome inactivation</td>
<td>32</td>
</tr>
<tr>
<td>1.6</td>
<td>CTCF protein specifically binds to several boundaries between escape and X-inactivated genes</td>
<td>38</td>
</tr>
<tr>
<td>1.7</td>
<td>Schematic map of the Jarid1c/JARID1C escape region in human and mouse</td>
<td>52</td>
</tr>
<tr>
<td>1.8</td>
<td>Research plan of my thesis work</td>
<td>58</td>
</tr>
<tr>
<td>2.1</td>
<td>Map of Jarid1c BAC clones RP23-330G24 and RP23-391D18</td>
<td>61</td>
</tr>
<tr>
<td>2.2</td>
<td>Scheme of the Q-SNaPshot assay</td>
<td>70</td>
</tr>
<tr>
<td>3.1</td>
<td>Mouse Jarid1c genomic region</td>
<td>78</td>
</tr>
<tr>
<td>3.2</td>
<td>Targeting of a neomycin/kanamycin cassette into the BAC backbone by recombineng</td>
<td>82</td>
</tr>
<tr>
<td>3.3</td>
<td>Pulsed field gel electrophoresis (PFGE) of wild-type and recombinant BAC DNA</td>
<td>85</td>
</tr>
<tr>
<td>3.4</td>
<td>Subcloning of ES cell line EL1/c16</td>
<td>88</td>
</tr>
<tr>
<td>3.5</td>
<td>FISH screening for X-linked BAC transgenes</td>
<td>92</td>
</tr>
<tr>
<td>3.6</td>
<td>Characterization of X-linked transgenes derived from BAC RP23-391D18-Neo</td>
<td>99</td>
</tr>
<tr>
<td>3.7</td>
<td>Characterization of X-linked transgenes derived from BAC RP23-330G24-Neo</td>
<td>101</td>
</tr>
<tr>
<td>3.8</td>
<td>Expression of transgene transcripts from active X chromosomes prior to ES cell differentiation</td>
<td>105</td>
</tr>
<tr>
<td>3.9</td>
<td>Scheme of inverse PCR method</td>
<td>109</td>
</tr>
<tr>
<td>3.10</td>
<td>Chromosomal location and transcription orientation of the four X-linked BAC transgenes</td>
<td>112</td>
</tr>
<tr>
<td>3.11</td>
<td>Allelic expression of genes adjacent to transgene integration sites and transcripts at the Jarid1c locus</td>
<td>116</td>
</tr>
<tr>
<td>3.12</td>
<td>Strain origin of transgenic X chromosomes</td>
<td>119</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>α-MHC</td>
<td>α-myosin heavy chain gene</td>
<td></td>
</tr>
<tr>
<td>Afp</td>
<td>alpha-fetoprotein gene</td>
<td></td>
</tr>
<tr>
<td>AGT</td>
<td>angiotensinogen gene</td>
<td></td>
</tr>
<tr>
<td>Atm</td>
<td>ataxia telangiectasia mutated</td>
<td></td>
</tr>
<tr>
<td>Atr</td>
<td>Atm and Rad3 related</td>
<td></td>
</tr>
<tr>
<td>BAC</td>
<td>bacteria artificial chromosome</td>
<td></td>
</tr>
<tr>
<td>Bcl-2</td>
<td>B-cell lymphoma 2 gene</td>
<td></td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
<td></td>
</tr>
<tr>
<td>°C</td>
<td>degrees Celsius</td>
<td></td>
</tr>
<tr>
<td>CAT</td>
<td>chloramphenicol acetyl transferase gene</td>
<td></td>
</tr>
<tr>
<td>CCD</td>
<td>cooled charge-coupled device</td>
<td></td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
<td></td>
</tr>
<tr>
<td>Cdx4</td>
<td>caudal type homeo box 4</td>
<td></td>
</tr>
<tr>
<td>Chic1</td>
<td>cysteine-rich hydrophobic domain 1</td>
<td></td>
</tr>
<tr>
<td>Ci</td>
<td>Curie</td>
<td></td>
</tr>
<tr>
<td>cLys</td>
<td>chicken lysozyme gene</td>
<td></td>
</tr>
<tr>
<td>CM</td>
<td>chloramphenicol resistance gene</td>
<td></td>
</tr>
<tr>
<td>CMV</td>
<td>cytomegalovirus</td>
<td></td>
</tr>
<tr>
<td>COL1A1</td>
<td>α1(I) collagen gene</td>
<td></td>
</tr>
<tr>
<td>cre</td>
<td>cre recombinase gene</td>
<td></td>
</tr>
<tr>
<td>CSK</td>
<td>cytoskeletal</td>
<td></td>
</tr>
<tr>
<td>CTCF</td>
<td>CCCTC binding factor</td>
<td></td>
</tr>
<tr>
<td>DAPI</td>
<td>4',6-diamidino-2-phenylindole</td>
<td></td>
</tr>
<tr>
<td>Ddx3x/DDX3X</td>
<td>DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked, gene</td>
<td></td>
</tr>
<tr>
<td>Ddx3y/DDX3Y</td>
<td>DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked, gene</td>
<td></td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s modified Eagle’s medium</td>
<td></td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
<td></td>
</tr>
<tr>
<td>Dnmt1</td>
<td>DNA methyltransferase 1 gene</td>
<td></td>
</tr>
<tr>
<td>Dnmt3b</td>
<td>DNA methyltransferase 3B gene</td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxynucleotide triphosphate</td>
<td></td>
</tr>
<tr>
<td>dpc</td>
<td>day post coitum</td>
<td></td>
</tr>
<tr>
<td>EB</td>
<td>embryoid body</td>
<td></td>
</tr>
<tr>
<td>E.coli</td>
<td><em>Escherichia coli</em></td>
<td></td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
<td></td>
</tr>
<tr>
<td>EGFP</td>
<td>enhanced green fluorescent protein gene</td>
<td></td>
</tr>
<tr>
<td>EIF1AX</td>
<td>eukaryotic translation initiation factor 1A, X-linked, gene</td>
<td></td>
</tr>
<tr>
<td>EIF2S3</td>
<td>eukaryotic translation initiation factor 2, subunit 3 gamma, gene</td>
<td></td>
</tr>
<tr>
<td>Eif2s3x</td>
<td>eukaryotic translation initiation factor 2, subunit 3 X-linked,</td>
<td></td>
</tr>
<tr>
<td>Eif2s3y</td>
<td>eukaryotic translation initiation factor 2, subunit 3 Y-linked</td>
<td></td>
</tr>
<tr>
<td>eNOS</td>
<td>endothelial nitric oxide synthase gene</td>
<td></td>
</tr>
<tr>
<td>Enox/ENOX</td>
<td>expressed neighbor of Xist/XIST</td>
<td></td>
</tr>
<tr>
<td>ES</td>
<td>embryonic stem</td>
<td></td>
</tr>
<tr>
<td>FCS</td>
<td>fetal calf serum</td>
<td></td>
</tr>
<tr>
<td>FISH</td>
<td>fluorescent <em>in situ</em> hybridization</td>
<td></td>
</tr>
<tr>
<td>GFP</td>
<td>green fluorescent protein gene</td>
<td></td>
</tr>
<tr>
<td>H3</td>
<td>histone 3</td>
<td></td>
</tr>
<tr>
<td>H3K4</td>
<td>histone 3 lysine 4</td>
<td></td>
</tr>
<tr>
<td>H3K27</td>
<td>histone 3 lysine 27</td>
<td></td>
</tr>
<tr>
<td>H4</td>
<td>histone 4</td>
<td></td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid</td>
<td></td>
</tr>
<tr>
<td>HIV</td>
<td>human immunodeficiency virus</td>
<td></td>
</tr>
<tr>
<td>HMG CoA</td>
<td>3-hydroxy-3-methylglutaryl coenzyme A reductase gene</td>
<td></td>
</tr>
<tr>
<td>HP1</td>
<td>heterochromatin protein 1</td>
<td></td>
</tr>
<tr>
<td>Hprt</td>
<td>hypoxanthine-guanine phosphoribosyltransferase gene</td>
<td></td>
</tr>
<tr>
<td>hr</td>
<td>hour(s)</td>
<td></td>
</tr>
<tr>
<td>HS4</td>
<td>hypersensitive site 4</td>
<td></td>
</tr>
<tr>
<td>HSV</td>
<td>herpes simplex virus</td>
<td></td>
</tr>
<tr>
<td>Igf2</td>
<td>Insulin-like growth factor 2</td>
<td></td>
</tr>
<tr>
<td>IgH</td>
<td>immunoglobulin heavy chain gene</td>
<td></td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
<td></td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>IQsec2</td>
<td>IQ motif and Sec7 domain-containing protein 2 gene</td>
<td></td>
</tr>
<tr>
<td>Jarid1c</td>
<td>Jumonji/ARID domain-containing protein 1C gene</td>
<td></td>
</tr>
<tr>
<td>Jarid1d</td>
<td>Jumonji/ARID domain-containing protein 1D gene</td>
<td></td>
</tr>
<tr>
<td>kV</td>
<td>kilo volts</td>
<td></td>
</tr>
<tr>
<td>L1</td>
<td>long interspersed nuclear element sub-class 1</td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td>long interspersed nuclear element sub-class 2</td>
<td></td>
</tr>
<tr>
<td>lacZ</td>
<td>β-galactosidase gene</td>
<td></td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani (medium)</td>
<td></td>
</tr>
<tr>
<td>LCR</td>
<td>local control region</td>
<td></td>
</tr>
<tr>
<td>LIF</td>
<td>leukemia inhibitory factor</td>
<td></td>
</tr>
<tr>
<td>LINE</td>
<td>long interspersed nuclear element</td>
<td></td>
</tr>
<tr>
<td>LTR</td>
<td>long terminal repeat</td>
<td></td>
</tr>
<tr>
<td>µF</td>
<td>microfarad</td>
<td></td>
</tr>
<tr>
<td>µg</td>
<td>microgram</td>
<td></td>
</tr>
<tr>
<td>µJ</td>
<td>microJoule</td>
<td></td>
</tr>
<tr>
<td>µl</td>
<td>microliter</td>
<td></td>
</tr>
<tr>
<td>µM</td>
<td>micromolar</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>molar</td>
<td></td>
</tr>
<tr>
<td>MAR</td>
<td>matrix attachment region</td>
<td></td>
</tr>
<tr>
<td>Mb</td>
<td>megabase</td>
<td></td>
</tr>
<tr>
<td>mg</td>
<td>milligram</td>
<td></td>
</tr>
<tr>
<td>Mid1/MID1</td>
<td>midline 1 gene</td>
<td></td>
</tr>
<tr>
<td>min</td>
<td>minute(s)</td>
<td></td>
</tr>
<tr>
<td>MIR</td>
<td>mammalian-wide interspersed repeat</td>
<td></td>
</tr>
<tr>
<td>ml</td>
<td>milliliter</td>
<td></td>
</tr>
<tr>
<td>mm</td>
<td>millimeter</td>
<td></td>
</tr>
<tr>
<td>mM</td>
<td>millimolar</td>
<td></td>
</tr>
<tr>
<td>MMLV</td>
<td>murine leukemia virus</td>
<td></td>
</tr>
<tr>
<td>MT</td>
<td>metallothionein gene</td>
<td></td>
</tr>
<tr>
<td>MTVP</td>
<td>metallothionein-vasopressin fusion gene</td>
<td></td>
</tr>
<tr>
<td>Myo7a</td>
<td>myosin VIIa gene</td>
<td></td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>n/a</td>
<td>not applicable</td>
<td></td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
<td></td>
</tr>
<tr>
<td><em>neo</em></td>
<td>neomycin resistance gene</td>
<td></td>
</tr>
<tr>
<td>NF-κB</td>
<td>nuclear factor kappa-light-chain-enhancer of activated B cells</td>
<td></td>
</tr>
<tr>
<td>ng</td>
<td>nanogram</td>
<td></td>
</tr>
<tr>
<td>OBR</td>
<td>origin of bidirectional replication</td>
<td></td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
<td></td>
</tr>
<tr>
<td>PAR</td>
<td>pseudoautosomal region</td>
<td></td>
</tr>
<tr>
<td>PARP-1</td>
<td>Poly(ADP-ribose) polymerase 1</td>
<td></td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate-buffered saline</td>
<td></td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
<td></td>
</tr>
<tr>
<td><em>PEPCK</em></td>
<td>phosphoenolpyruvate carboxykinase gene</td>
<td></td>
</tr>
<tr>
<td>PFGE</td>
<td>pulsed field gel electrophoresis</td>
<td></td>
</tr>
<tr>
<td><em>PGK</em></td>
<td>phosphoglycerate kinase gene</td>
<td></td>
</tr>
<tr>
<td>RBMX</td>
<td>RNA binding motif protein, X-linked, gene</td>
<td></td>
</tr>
<tr>
<td>REST</td>
<td>RE1-silencing transcription factor</td>
<td></td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
<td></td>
</tr>
<tr>
<td>RT-PCR</td>
<td>reverse transcription PCR</td>
<td></td>
</tr>
<tr>
<td>SAP</td>
<td>shrimp alkaline phosphatase</td>
<td></td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
<td></td>
</tr>
<tr>
<td>sec</td>
<td>second(s)</td>
<td></td>
</tr>
<tr>
<td>SINE</td>
<td>short interspersed nuclear element</td>
<td></td>
</tr>
<tr>
<td><em>SLN</em></td>
<td>sarcolipin gene</td>
<td></td>
</tr>
<tr>
<td>Smad3</td>
<td>Mothers against decapentaplegic homolog 3</td>
<td></td>
</tr>
<tr>
<td><em>Smchd1</em></td>
<td>structural maintenance of chromosomes hinge domain containing 1</td>
<td></td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
<td></td>
</tr>
<tr>
<td><em>SPRY3</em></td>
<td>sprouty homolog 3 gene</td>
<td></td>
</tr>
<tr>
<td><em>SRY</em></td>
<td>sex-determining Region Y gene</td>
<td></td>
</tr>
<tr>
<td>SSC</td>
<td>standard saline citrate</td>
<td></td>
</tr>
<tr>
<td><em>Sts/STS</em></td>
<td>steroid sulfatase gene</td>
<td></td>
</tr>
<tr>
<td>SV40</td>
<td>simian virus 40</td>
<td></td>
</tr>
</tbody>
</table>
SYBL1 synaptobrevin-like 1 gene
Tie2 angiopoietin receptor 2 gene
TGF-β transforming growth factor, beta
TIMP1 tissue inhibitor 1 of metalloproteinase gene
TK thymidine kinase gene
TNF transferrin gene
tRNA transfer RNA
Tsix/TISX Xist/XIST antisense gene
Tspyl2 testis specific protein, Y-linked like 2 gene
Tsx Testis specific, X-linked, gene
U unit or undifferentiated
Ube1 ubiquitin-activating enzyme E1 gene
UTR untranslated region
Utx/UTX ubiquitously transcribed tetratricopeptide repeat, X-linked, gene
Uty/UTY ubiquitously transcribed tetratricopeptide repeat, Y-linked, gene
UV ultraviolet
V volt(s)
Xa active X chromosome
XAR X added region
Xce X-controlling element
XCI X chromosome inactivation
XCR X conserved region
Xi inactive X chromosome
Xic/XIC X inactivation center
Xist/XIST X inactive specific transcript gene
Xite X-inactivation intergenic transcription element
Xm maternal X chromosome
Xp paternal X chromosome
Zfx/ZFX zinc finger protein, X-linked gene
I would like to express my gratitude to all of the people who contributed to this body of work and supported me in many different ways during my graduate study at Penn State. First, I would like to thank my advisor, Dr. Laura Carrel, whose guidance and insight in science have been a great source of inspiration for my Ph.D. training. I also deeply appreciate help and constructive discussions from everybody in her laboratory, Katie Prothero, Jill Stahl, Sarah Arnold-Croop, and Lindsay Horvath. I would also like to offer my sincere thanks to my thesis committee members: Drs. Sarah K. Bronson, Sergei A. Grigoryev, Kent E. Vrana, Teresa L. Wood, for providing suggestions on my research and critically reviewing the manuscript. Special thanks are extended to Dr. Sarah K. Bronson for her intellectual contribution to experimental design.

I am grateful for Dr. Hunt Willard (Duke) for valuable comments on this project, Dr. Neal Copeland for plasmids and bacterial strains necessary for recombineering experiments and Dr. En Li for the mouse female ES cell line EL1/c16.

Finally and most importantly, I would like to thank my family and friends for always staying close to me. My parents, Dr. Jiheng Li and Guifang Shen, and my brother, Su Li, always supported me with their love, confidence and encouragement despite the distance. My loving wife, Bing Dong, gave me endless encouragement and understanding, and share the good and bad moments of my study. I would not have been able to go through this long, tough journey without their unconditional support. I also wish to thank all my friends, previously or currently at Hershey, for making these years so much fun. Your friendship has been of great importance to me. It is to my family and everyone who supports me that I dedicate my work.
Although phenotypic traits are controlled by genetic information, many of them cannot be simply explained by the underlying DNA sequence. In many cases, changes in gene activity are mitotically or meiotically heritable but do not entail any change in DNA. For instance, paramutation in maize is induced by an interaction between two alleles of a kernel color gene, resulting in a heritable reduction in expression of one allele by the other allele that is only present in parental cells [1]. In Drosophila, the red eye coloration gene is silenced in some cells that carry the gene at loci adjacent to inactive heterochromatin, yielding a mosaic appearance of white and red sectors in eyes known as position effect variegation [2]. Such events are regulated by “epigenetic” mechanisms and are usually associated with characteristic marks on chromatin including DNA methylation and histone modifications, as well as small RNA molecules [3]. Epigenetic regulation is not only ubiquitous, but also essential for many biological processes such as development and tissue-specific gene expression [4]. Aberrant regulation of epigenetic events leads to a range of disorders including aging and cancer [5]. Thus, in addition to genetic control, there are epigenetic mechanisms that exert another critical level of regulation to mediate temporal and spatial gene expression.

Mono-allelic expression is a typical epigenetic event. In this case, although the two alleles of a gene may share essentially the same DNA sequence and reside in the same nucleus, they are present in two different transcriptional states. One allele is actively expressed while the other is silenced. For instance, the two alleles of some genes are expressed in a parent-of-origin specific manner, a process known as genomic imprinting [6]. X chromosome inactivation (XCI) is another extraordinary example of mono-allelic expression as hundreds of genes on one of the two X chromosomes in female mammals are silenced while the other X is unaffected. Studying XCI will facilitate our understanding of mono-allelic expression and, more generally, epigenetic regulation as well as the mechanisms of some X-linked diseases. Research in the past half century has revealed many mysteries in different aspects of XCI; however, a number of questions still remain unsolved including how gene expression is controlled on the inactivated X chromosome. Here I review the literature on X chromosome biology and gene regulation on the inactivated X.
1.1 X CHROMOSOME INACTIVATION

1.1.1 Mammalian sex chromosomes

There are two forms of mammalian sex chromosomes, X and Y, and their combinations determine gender in mammals. In therian mammals (marsupial and placental), the Y chromosome contains a testis-determinant gene, \( SRY \), \[7\] and individuals carrying the Y chromosome become males. Normal males have one Y and one single X chromosome, whereas females are characterized by two Xs and no Y. Due to the sex-determining function of the Y, some rare cases such as XXY individuals are males while XO and XXX individuals are females.

Besides the role in sex determination, X and Y chromosomes differ significantly in the size, chromatin composition, as well as content and function of their genes [8]. Y chromosomes are small, gene-poor, and contain many transcripts with male-specific functions. For instance, the human Y is \(~63\) million base pairs (Mb) long including a \(~40\) Mb heterochromatic block devoid of genes (Figure 1.1A) [9, 10]. Only \(~100\) genes or gene families have been identified in the remaining euchromatic region. Many of them are testis-specific and direct male development. In contrast, X chromosomes are large, gene-rich and absolutely essential for viability [8, 10]. The human X is \(~159\) Mb in size and contains \(~1000\) genes. These genes are distributed on the entire X and most of them are involved in housekeeping functions [8]. Similar to the human X and Y, the two sex chromosomes in other therian mammals such as mouse also significantly differ in gene content and overall DNA composition [11].

The differences between therian X and Y chromosomes may result from their unique evolutionary history. The third mammalian lineage, prootherian mammals (monotremes), has recently been found to contain multiple sex chromosomes (e.g. five Xs and five Ys in platypus) that have no \( SRY \) gene [12]. The X chromosomes lack homology to the eutherian X and at least one of these chromosomes have extensive sequence homology with the chicken sex chromosome Z that is unrelated to therian X and Y. These data suggest that therian sex chromosomes may descend from a pair of ancestral autosomes shortly after the divergence of therian and prootherian lineages, approximately 160-200 million years ago (mya) [13]. Sex chromosome evolution is believed to have involved a continuous process of adding autosomal
Figure 1.1  X chromosome evolution in mammals. (A) The evolution history of mammalian X and Y chromosomes proposed by Lahn and Page [14]. Eutherian sex chromosomes are homologous to an autosomal pair in birds and obtain the sex-determining genes SRY (red box) and its homologue SOX3. Non recombining regions on the Y (NRY, shown as grey box) emerge when these genes are differentiated on the proto-X/Y chromosomes and inhibit recombination between them. The X and Y chromosomes evolve with the addition of autosomal material and further expansion of NRYs, potentially mediated by large-scale chromosomal inversion and causing subsequent loss of Y chromosome sequences. These events may occur multiple times and affect different locations in various species. Main events and the consequent X-specific region including XAR and XCR are indicated. Five evolutionary strata on the human X are shown in number and different color. Blue regions are recombining areas. The PARs are present on both X and Y chromosomes (PAR2 is located at the distal end, not shown). A large region on the human Y is heterochromatic and depleted of genes (black box). The diagram is not shown to scale and centromeres are not marked as their location during these evolutionary events is unclear. Image is modified from [353]. (B) The chromosome organization, including evolutionary strata, of the mouse and human Xs are rearranged relative to each other, while the gene content is conserved. Relative positions of some genes on mouse and human X chromosomes and the strata on the human X are marked. Image source: [354].
A

SOX3  SRY

Ancestral autosomes

Recombination suppression and Y attrition

Emergence of sex chromosomes

Further Y attrition and differentiation, Marsupial X/Y

Non-Simian mammal X/Y

PAR1

5 4 3 2 1

human

heterochromatin

X

XAR

XCR

B

5.5 cM

5.7

24.5

32.0

34.8

64.0

Mouse

PAR

PAR

PAR

Smcx

(Rbmx)

Zfx

Sox3

Eif2s3x

Utx

Usp9x

Ube1x

ZfX

USP9X

UBX

SMCX

RBMX

SOX3

Human

segments to the proto-X and Y chromosomes (addition), repressing X-Y pairing and subsequent loss of sequences on the Y (attrition or degeneration) (Figure 1.1A) [14, 15]. Unlike the degenerative events on the Y, X-linked genes may have been under selective pressure for viability and thus remain highly conserved during evolution [11, 15].

According to the evolutionary history, X chromosome sequences are classified into several regions. The comparison of gene content and sequence between marsupial and human X chromosomes has revealed that a large proportion of the human X including the long arm and part of the short arm retain considerable homology, and is designated the X conserved region (XCR, Figure 1.1A) [16sp]. A region covering much of the short arm is an evolutionarily recent addition (80-130 mya) specific to eutherians (placental mammals) and is coined X added region (XAR, Figure 1.1A).

Further delineation of XAR sequences came through a comparison of sequence divergence for 19 human X-Y gene pairs [14]. This analysis led Lahn and Page to propose that XAR sequences were acquired in a stepwise fashion after the metatherian (marsupial) and eutherian (placental) lineages split. Thus, human X chromosome sequences are divided into “evolutionary strata” according to their acquisition time (Figure 1.1A) [14]. Identification of five strata on the human X (from the oldest S1 to the youngest S5) suggests that at least five major events suppressed X-Y pairing and allowed the X and Y to evolve independently without grossly disturbing gene order in each stratum [8]. These strata on the mouse X are reorganized relative to those on the human X, potentially due to independent chromosomal inversions during evolution (Figure 1.1B) [17].

Regions of high homology still exist between X and Y in order to undergo X-Y pairing that is essential for male meiosis [18]. These regions are thought to be added recently compared to the rest of sex chromosomes, and are called pseudoautosomal regions (PARs). The human X and Y share two PARs [19]. The major PAR (PAR1) extends 2.7 Mb at the end of the short arm and contains twenty-four genes. The minor PAR (PAR2), located at the extremity of the long arm, is significantly smaller (330 kb) and contains only five known genes. The mouse X and Y have only PAR1 at the distal tips with a much smaller size (720 kb) [20]. As PARs are small and carry only a few genes when compared to the entire X or Y, they cannot compensate for the differences in composition and functionality between the two sex chromosomes.
1.1.2 Dosage compensation in mammals

The vastly different composition of X and Y chromosomes results in genetic imbalance between male and female mammals, as X-linked genes are disomic in females but monosomic in males. This inequality is also present in other organisms in which gender is determined by sex chromosomes, such as worms and flies. These organisms have evolved mechanisms that can balance the disparity in DNA dosage, processes called dosage compensation. In *Caenorhabditis elegans*, XX hermaphrodites reduce the expression level from both Xs by half to maintain equality with transcription activity of XO males [21]. In *Drosophila melanogaster*, XY males double X chromosome expression to match the two Xs in females [22].

In therian mammals, dosage compensation is achieved by X chromosome inactivation (XCI) or lyonization, as it was first hypothesized by Mary Lyon [23]. It was observed in mouse that one chromosome is condensed in the diploid cells from only females, that one X is sufficient for development since XO females are viable and that some females carrying X-linked mutations show mosaic phenotypes [23]. Experiments have established that one of the two Xs in females is inactivated (Xi), and the genes on this chromosome are transcriptionally silenced [24, 25]. The other X is unaffected and genes are expressed from this active X (Xa). Thus, both sexes have one active copy of X-linked genes, rendering equal the amount of X-linked gene products between XX and XY individuals. Recent studies of dosage compensation in monotremes show that at least some X-linked genes show mono-allelic expression as well [26].

It has been proposed that XCI is linked to sex chromosome evolution. For specific X-Y gene pairs, inactivation of one X-linked copy in females makes the Y homologue unnecessary in males; therefore XCI may force Y chromosome degeneration (X-driven hypothesis) [27]. Alternatively, degradation of Y chromosome material may cause inequality of X-linked genes between two sexes and induce XCI (Y-driven hypothesis) [27]. Although the cause and effect relationship between XCI and Y degeneration is not clear, XCI may also play an important role in X chromosome evolution. As genes on one X are subject to inactivation, their translocation to autosomes should be detrimental due to loss-of-function and thus be selected against [28]. Therefore, genes are “locked in” after moving onto the X. Subsequently, the DNA sequence and gene content of X chromosomes are highly conserved in mammals.
(Ohno’s law), especially in eutherian mammals [10].

While my thesis focuses on XCI, as a matter of completeness on dosage compensation that affects X chromosomes, an additional process has been recently recognized to balance X-linked and autosomal transcripts. Compared to disomic autosomal genes, X-linked genes are functionally monosomic if Xa alleles retain their expression levels after XCI. Quantitative allele-specific analysis and microarray studies in human, mouse, and rat have revealed that X-linked genes in males and Xa alleles in females are globally up-regulated compared to those in germ line cells to maintain X:autosomal balance [29-31]. Such a process may protect mammals from deleterious effects due to X chromosome haploinefficiency. Both the Xa hyperactivation and XCI processes are great examples of large-scale gene regulation.

1.1.3 Random and imprinted X chromosome inactivation

There are two forms of XCI. In the cells that form the embryo proper of eutherian mammals, XCI occurs randomly and silences either the paternal (Xp) or maternal X (Xm) chromosome [32]. In metatherian mammals such as marsupial, imprinted XCI occurs with only the Xp inactivated [33]. Imprinted inactivation is poorly maintained and CpG island methylation, a reliable marker of random XCI, is not associated with the marsupial Xi [34, 35]. Although the marsupial X is highly conserved with much of the eutherian X, some key regions that are required for random XCI (see section 1.1.5) have been disrupted on the marsupial X [36]. This finding suggests that a unique XCI pathway may exist in metatherian mammals.

Imprinted XCI also takes place in mouse. Recent studies have demonstrated that mouse embryos carry an inactive Xp at the early stage of development (8-16 cells) (Figure 1.2) although the inactivation is not pan-chromosomal [37]. It is not clear whether zygotes receive this inactive Xp from germ line cells [37] or the Xp is active at fertilization but undergoes “de novo” inactivation before implantation [38, 39]. This inactive Xp is preserved in mouse extraembryonic tissues (trophectoderm and primitive endoderm) [40]; however, it is reactivated in the inner cell mass of blastocysts, followed by the random XCI event [38].

The imprinted XCI during preimplantation may be regulated differently from random XCI. Although the inactivation is stably maintained, it is associated with a wide range of different epigenetic features from random XCI (e.g. [41, 42]). Differences in these two
Figure 1.2 X chromosome inactivation in female mouse embryos. Imprinted X chromosome inactivation with the paternal X chromosome (Xp) being silenced is seen at the four-cell stage during preimplantation of mouse development. The inactive Xp is subject to epigenetic modifications and maintained to the early blastocyst stage. One study also suggested that the zygote might inherit an inactive X from the male germline cells [37]. At the late blastocyst stage, the fate of the Xp diverges depending on the tissue type. In extraembryonic tissues including trophectoderm (TE) and primitive endoderm (PE), the Xp remains inactive. However, the Xp is reactivated in the inner cell mass (ICM) and releases the epigenetic modifications acquired in imprinted XCI. Thus, cells that contribute to embryo proper (epiblast) contain two active Xs. These cells subsequently inactivate one of the two Xs on a random basis. The inactive X acquires many epigenetic modifications in a sequential order and is stably transmitted during somatic cell division. This inactive X reactivates in the female germline prior to meiosis. Image source: [43].
processes are also apparent in gene response to the inactivation events as some transgenes are inactivated by only one of these processes [44-46]. Further, human extraembryonic tissues may not undergo imprinted XCI, as one study, testing a gene in cytotrophoblasts from a number of human female placentas, demonstrated that XCI was random in most samples [47]. Therefore, imprinted XCI occurs at unique developmental stages or in specific cell lineages. In contrast, random XCI affects placental mammals predominantly and has been studied more extensively. Here I will focus my review on random XCI.

1.1.4 Processes of X chromosome inactivation

Random XCI is generally divided into several stages in a sequential order: counting and choice, initiation, spreading, and maintenance [48]. First, each cell senses the X:autosome ratio and determines which X is to be inactivated. This step follows an “n-1” rule and only a single X remains active per diploid set of autosomes [49]. Therefore, individuals with one X (such as XY and XO) do not have the Xi, and individuals with two or more Xs (such as XX, XXY, X trisomy and tetrasomy) have all but one X to be inactivated.

In normal females, both Xs are active prior to XCI and either can be chosen to be silenced [50]. Once the choice is made, inactivation occurs at the late blastocyst stage (32-64 cells) of embryogenesis [50]. Inactivation initiates at a specific locus on the X called X inactivation center, then spreads in cis along the chromosome and silences genes on it within one or two cell cycles [48]. Meanwhile, the Xi undergoes multiple epigenetic modifications that stabilize the inactive state [51] (see section 1.1.6). Random XCI is believed to be complete at 6.5 day post coitum (dpc) in the post-implantation embryo [52]. During subsequent cell divisions, the Xi is stably maintained and faithfully duplicated. Random XCI followed by clonal inheritance gives rise to female adults that are mosaics for cell populations with either the Xp or Xm inactivated, visible in the color pattern of calico cats [53]. Nevertheless, the Xi can reactivate during oogenesis [54], allowing random XCI to occur in the next generation and exhibiting the plasticity of such an epigenetic event.

1.1.5 X inactivation center

X inactivation center is a major genetic locus on the X that controls XCI. It has been identified in both mouse (Xic) and human (XIC) and contains multiple genes and functional
elements (Figure 1.3) [55-58]. Xic/XIC is required for XCI to occur as X chromosomes lacking this region are unable to be inactivated [59, 60]. When the X is truncated or autosomally translocated, XCI occurs only on derivative Xs that contain Xic/XIC [61-69]. Further, Xic/XIC sequences integrated on autosomes and Y chromosomes are sufficient to initiate inactivation of juxtaposed non-X sequences [57, Migeon, 1999 #116, 60, 70]. Therefore, it is proposed that Xic/XIC is the initiation site of XCI.

Xic/XIC is also involved in other XCI processes. Deletion in Xic impairs counting mechanisms and causes XCI in XO and XY cells [71]. This region also regulates the choice of which X is inactivated. In mouse, choice depends on a locus on the X called X-controlling element (Xce) that influences the probability of inactivation in cis [72]. Four strain-specific Xce alleles have been identified, from the weakest Xce^a (most likely to be inactivated) to strongest Xce^d [73]. Female cells carrying two identical Xce alleles inactivate either X at an equal chance, while heterozygotes show skewed XCI with a ratio that can skew to 80:20. It is not clear whether there is a human equivalent of the Xce locus. However, familial skewing has been observed in human [74, 75]. A number of studies have mapped the Xce function within or close to Xic/XIC [74, 76-79]. The counting and choice processes may be regulated through a transient interaction of Xic/XIC sequences between the two Xs [80-82]. Intriguingly, Xic is not required to maintain inactivation, as removing this region after XCI is established did not alter the inactive state of the Xi [83].

X inactivation specific transcript (Xist in mouse/XIST in human) is a gene that maps within Xic/XIC and absolutely essential for XCI [56, 84, 85]. Instead of producing a protein, this gene encodes a large, untranslated RNA product that is 17 kb in mice and 19 kb in human [86]. Deletion and cDNA transgene studies demonstrated that Xist/XIST is necessary and sufficient to induce long-range inactivation in cis [57, 70, 87-91]. Functional studies of mouse Xist, as well as other aspects of XCI in mouse, benefit from an ex vivo model system, mouse female embryonic stem (ES) cells, which recapitulate XCI upon in vitro differentiation (see section 1.6). Xist expression at the initial stage of ES cell differentiation can induce chromosome silencing in cis while late expression cannot, suggesting a role of this RNA as an initiator of XCI [91].

The expression pattern of mouse Xist is also consistent with the idea that Xist initiates XCI. Before the onset of XCI, Xist is transcribed at a low level from both active Xs [92].
Figure 1.3  

**X inactivation center in mouse.** The genes and sequence elements that regulate the XCI process include non-coding genes *Xist, Tsix, Xite* and a repetitive sequence DXPas34. Protein coding genes *Tsx, chic1* and *Cdx4* do not show function in XCI regulation. The regions that are involved in counting and choice are shown as the black bars under the genomic sequence. CEN: centromere. Image modified from [43].
Once the X to be inactivated is chosen, the earliest events known are the up-regulation of *Xist* expression on the future Xi, followed by spreading of the RNA product in *cis* to encompass the entire chromosome [92]. Thereafter, *Xist* is constitutively transcribed from the Xi and stably maintained *in situ*. This unique pattern also provides a useful marker to detect the Xi in mouse cells. Through fluorescent *in situ* hybridization (FISH), *Xist* RNA is shown as two pinpoint dot signals in nuclei prior to XCI and a large accumulation on the Xi upon XCI [93]. Although human *XIST* RNA binds to the Xi as well, its expression pattern may be complicated as some undifferentiated human ES cell lines have already shown *XIST* up-regulation [94, 95]. XCI may have taken place in human ES cells, or the cell lines tested may have initiated the differentiation process.

*Xist/XIST* also mediates the propagation of inactivation along the X chromosome. *Xist/XIST* RNA localization to the Xi provides evidence of a spatial interaction with the X at the onset of XCI. The *Xist* coating in mouse occurs immediately after the up-regulation of this gene but shortly before an initial chromosome inactivation event, allowing time for this RNA to spread over the X and induce silencing [91, 96]. Although this initial silencing is transient and reversible, it is required for establishing a stable inactive state on the Xi [91]. Importantly, the transcriptional silencing is dependent on *Xist* RNA. An *Xist* mutant with deletion of a repeat sequence at the 5’ end, known as A-repeat, specifically abolishes the initial silencing event, whereas mutations at other regions disrupt the RNA association [97]. The *Xist*-mediated silencing is likely achieved through a repressive nuclear compartment confined by this RNA [96].

Stabilization and maintenance of the initial silencing is independent of mouse *Xist*, as conditional deletion of A-repeat after XCI is established does not alter chromosome silencing although the RNA is still associated with the Xi [97]. However, conditional knockout of *Xist* after XCI is finished causes sporadic gene reactivation [98]. Thus, *Xist* is not required for but still contributes to the inactive state of the Xi. This is accomplished potentially by epigenetic modifications and heterochromatin formation associated with the Xi (see section 1.1.6), because cells with *Xist* mutations fail to localize proteins to the Xi that are involved in chromatin remodeling and heterochromatinization, such as histone variant macroH2A [98]. Additionally, the repressive chromatin territory established by *Xist* RNA may help restrain genes to more internal locations of the *Xist* domain [96, 99].
In mouse, Xist expression is regulated by another non-coding gene that is antisense to Xist, named Tsix [58]. This gene originates 15 kb downstream of Xist and encompasses the entire Xist genomic region (Figure 1.3). Similar to Xist, Tsix is expressed from both alleles at a low level prior to XCI. However, these two genes are transcribed in a mutually exclusive manner after XCI initiates; Tsix is active solely on the future Xa until inactivation is established and both alleles are repressed thereafter. It is proposed that, by repressing Xist transcription via modification of chromatin structure at the Xist promoter, Tsix regulates the choice process and prevents the initiation of XCI in cis [100]. The sequence and transcription process of Tsix also play a role in counting the number of the X [101, 102]. Tsix expression is controlled by another non-coding gene Xite and a developmentally regulated enhancer DXPas34 [103]. Human XIST may be controlled differently, as human TSIX is truncated compared to the mouse homologue and does not seem to regulate XIST [104].

1.1.6 Epigenetic features of inactive X chromosomes

The Xi acquires a number of epigenetic features distinguishable from its active counterpart, including peripheral nuclear location, delayed replication timing and chromatin modifications characteristic of inactive transcription and heterochromatin. At interphase, the Xi organizes into a highly condensed structure next to the nuclear membrane, cytologically identifiable as the “Barr body” [105]. During S phase, the Xi is targeted to the nucleolar periphery [106]. The Xi replicates late following Xist induction, while the Xa is synchronous with autosomes [107].

Chromatin modifications of the Xi closely follow Xist coating and occur in a sequential manner. Early events include the loss of many features associated with euchromatin, such as acetylation of histone H3 and H4 [108-110], and methylation of H3K4 (di, tri-), H3R17 (di-), H3K36 (di-), H3R2 (di-) and H3R26 (di-) [111, 112]. Xist specifically and transiently recruits repressive proteins, such as polycomb repressor complex 2 (PRC2), which in turn catalyzes H3K27 tri-methylation [113], and PRC1, which mediates H2AK119 mono-ubiquination [114]. The Xi is also associated with methylated H3K9 (di-and tri-) and H4K20 (mono- and tri-) [115-118]. Methylated H3K9 deposits heterochromatin protein 1 (HP1) to the Xi, at least in human, and the latter can promote heterochromatin formation [112]. Occurring relatively late, macroH2A is deposited to the Xi by Xist and forms a macrochromatin body [98, 119].
protein may be involved in the heterochromatinization of the Xi and contribute to gene silencing as well. Finally, CpG islands at the promoter region of silenced genes are methylated, although the rest of Xi is hypomethylated compared to the Xa [120].

The epigenetic features identified to date may not be required to initiate XCI, as initial silencing can take place in the absence of some epigenetic modifications (e.g. [121]). However, disruption of these features sporadically reactivates X-linked genes in both human and mouse (e.g. [41, 122, 123]). Thus, they are important for establishing and maintaining a stable inactive state. The stability of inactivation seems to increase gradually with epigenetic modification of the Xi. Some features such as DNA methylation are progressively intensified during XCI [124], and expression of many genes from the Xi is inversely correlated with the level of DNA methylation [125]. Additionally, when only one feature is impaired, reactivation occurs at a single gene level rather than globally, whereas simultaneous repression of two or more features increase the reactivation frequency [113, 126, 127]. Therefore, multiple epigenetic modifications operate in a redundant and synergistic manner to “lock in” a transient inactive state that is initiated by Xist/XIST RNA, and maintain the memory of inactivation by retaining the Xi as facultative heterochromatin [126].

1.1.7 Regulation on inactive X chromosomes

Some epigenetic marks of the Xi predominantly localize to gene promoters and could suggest that each gene is independently regulated by XCI [110, 116, 120]. Nonetheless, insight into the long-range regulation of XCI has come from transgene and translocation studies.

1.1.7.1 Inactivation in X;autosome translocations

Long-range inactivation can be induced on autosomal material after translocation to X chromosome fragments. In females who carry a balanced X;autosome translocation, the translocated Xs remain active while the normal X is inactivated in order to maintain normal dosage of autosomal and X-linked genes. However, derivative Xs containing the Xic/XIC region can be inactivated. Using these unbalanced X;autosome translocations, a number of studies assessed transcriptional activity of X-linked and autosomal genes spanning the break points and demonstrated that XCI could propagate into adjoining autosomal regions [66, 68, 128-131]. Investigation of epigenetic marks, such as replication timing [128, 132-135], H4
hypoacetylation and DNA methylation[67-69], yielded largely consistent results and demonstrated that autosomal sequences can be inactivated.

Despite the ability of XCI to propagate into non-X sequences, inactivation of autosomal material is not as complete and efficient as that of X-specific DNA. First, autosomal inactivation is short and discontinuous. Although silencing extends 45 Mb from the translocation point in one case [68], it generally covers small regions adjacent to the X and never includes entire autosomal fragments [129, 136-138]. Within the silenced autosomal sequences, it is frequently seen that some interior regions remain active [66, 68, 131, 137]. Secondly, autosomal inactivation is unstable. Some inactivated regions reactivate over time during somatic development or cell proliferation in culture systems [132, 138]. Additionally, the distance of propagation is sometimes variable between cells [128, 139, 140]. The unstable inactivation may be due to reduced levels of epigenetic modifications. Autosomal inactivation can take place in the absence of $Xist/XIST$ coating and some epigenetic marks of Xi [68, 69, 136, 137, 140-145]. On the translocated X chromosome, X-specific sequences are usually replicated latest and inactivated autosomal regions are slightly earlier but later than normal autosomes, suggesting autosomal silencing may be in an intermediate state and not as complete as XCI [128]. Moreover, it is found that XCI does not spread to autosomal sequences at all in some cases [68, 136, 137, 140-146].

1.1.7.2 Inactivation on non-X chromosomes carrying an $Xist/XIST$ transgene

$Xist/XIST$ transgenes on autosomes or $Y$ chromosomes are sufficient to initiate long-range inactivation in cis; however, similar to $X$;autosome translocations, inactivation is incomplete [60, 70, 87, 88, 147, 148]. Compared to the extensive inactivation on the Xi, silencing induced by the transgenes is more localized to integration sites, or is only found in a subset of cells [60, 87, 88, 147]. The transgenic chromosomes replicate earlier than the Xi and associated histones are not extensively hypoacetylated [60, 147]. Some inactivated regions show increased histone acetylation over time and are progressively reactivated [88]. Therefore, these transgene studies also demonstrate the incomplete and inefficient autosomal inactivation compared to XCI.

1.1.7.3 Lessons from translocation and transgene studies
The *cis* propagation of XCI suggests that inactivation may be specific for X chromosome sequences. However, the X;autosome translocation and Xist/XIST transgene studies clearly show that XCI, instead of being restricted to the X, can also inactivate autosomal material. Autosomal genes are active under normal circumstances, but can be inactivated after translocating to the X or when physically linked to the Xist/XIST gene. Inactivated autosomal regions are randomly distributed in the genome, unlikely to share specific sequence elements responsible for inactivation. Therefore, genes silenced by XCI do not require specific *cis* regulatory sequences. Moreover, autosomal genes do not experience X chromosome evolution or acquire any potential regulatory mechanisms specific for X-linked genes. It is unlikely that acquisition of gene-specific regulatory elements during evolution causes X-linked genes to be inactivated. Instead, inactivation may be controlled at a regional level, as silenced genes are immediately adjacent to one another on the Xi and inactivated autosomes [25, 66]. The Xi in human is composed of two types of heterochromatin that are cytogenetically different and spatially segregated, also suggesting the inactivation is regulated regionally [149]. One of them is characterized by trimethylated H3K9, trimethylated H4K20, HP1 and latest replication, while another is associated with XIST RNA, trimethylated H3K9 and macroH2A. Thus, the human Xi is organized into regions with distinct heterochromatic features, and genes within each region may be controlled coordinately.

Then why is inactivation more complete on the X than on autosomes? Differences between these chromosomes include underlying sequences and the chromosomal organization resulting from the primary sequences, but how they influence inactivation remain elusive. To explain incomplete autosomal silencing, Gartler and Riggs put forward that the spreading and maintenance of XCI were facilitated by sequence elements present on the X, referred to as "booster elements" or “way stations” [150]. These elements may act as high-affinity sites to interact with inactivation machinery, thereby enhancing gene silencing and/or heterochromatin formation. These sequence elements are not unique to the X as autosomal material can be inactivated as well. However, they may be prevalent or specially organized on the X to ensure efficient propagation of inactivation by stabilizing or amplifying silencing factors [150]. Due to the cytogenetic and functional features of Xist/XIST RNA, it may serve as such silencing factors during the XCI spreading. Other factors, such as epigenetic modifications, may play a similar role in the maintenance. The “booster element” model is compatible with the regional
control of inactivation because it proposes that genes near the booster elements are prone to be silenced.

Mary Lyon postulated that long interspersed nuclear element (LINE) sub-class 1 (L1) was involved in XCI by behaving as the “booster elements” (L1 hypothesis) [151]. L1s are endogenous retrotransposons specific for mammals and play a significant role in gene evolution and genomic instability [152]. By DNA FISH using L1 sequences as the probe, it has been shown that L1s are highly concentrated on the X in both human and mouse when compared to the autosomes [153, 154]. Bioinformatic approaches verified a nearly two-fold enrichment of L1s, especially a much higher level of evolutionarily younger L1s, on the human X [8, 155]. Lyon also proposed that Xist/XIST RNA might facilitate the spreading of XCI by mediating intra-chromosomal pairing of L1 repetitive sequences [151]. However, questions remain unsolved whether and how L1 sequences assist Xist-mediated gene silencing because the cytogenetic distribution of L1s and Xist RNA are not consistent. L1s accumulate predominately at AT-rich G-positive bands [153], while Xist RNA localizes mainly to GC-rich R bands [65]. Moreover, L1 content is also dramatically increased on the human Y chromosome and at certain regions of chromosome 21 that have low recombination rate [156]. Thus, L1 enrichment on the X may result from restricted recombination between X and Y chromosomes and are not related to XCI.
1.2 ESCAPE FROM X CHROMOSOME INACTIVATION

Despite the inactive state of the Xi, not all genes on this chromosome are silenced. Some genes are expressed from both the Xa and Xi, and therefore “escape” XCI. With the transcriptional activity contrary to most genes on the Xi, escape genes are of great importance and help for understanding XCI regulation. Study of these genes will also provide insights into their roles in some X chromosome diseases, such as X chromosome aneuploidies, and the gender differences related to these genes.

1.2.1 Discovery of escape genes

The idea that some X-linked genes are expressed from the Xi was originally proposed to explain the abnormalities seen for patients with X chromosome aneuploidies [157]. The rationale is that, if the Xi is completely silenced, it would not contribute to normal physiological functions. Consequently, individuals carrying one or more Xs should always be phenotypically normal because only one X is active. However, human females with the karyotype 45,X have an extremely high frequency of in utero lethality (99%) and survivors suffer from Turner Syndrome, which includes short stature, ovarian failure and lymphoedema [158]. 47,XXY males are aberrant in development and fertility (Klinefelter Syndrome) although one X is shown to undergo inactivation [159]. Compared to Triple X Syndrome (47,XXX) that shows slightly abnormal phenotypes, Tetrasomy and Pentasomy X Syndromes (48,XXXX and 49,XXXXX) display more severe phenotypes while only one Xa is present in all cases [160]. These observations strongly suggest that the Xi is, at least partially, functional.

A number of approaches have been used to determine the XCI status (the transcriptional activity on the Xi) of X-linked genes. Initial efforts focused on the protein product. Difference in the protein level between individuals with one, two or more X chromosomes infers escape from XCI [161-163]. Additionally, protein polymorphisms affecting expression levels, biochemical properties (such as migration distance in electrophoresis) [164], enzymatic activities [165] or epitope presentation [166-168] could distinguish mono-allelic (potentially from Xa only) and bi-allelic (from both Xa and Xi) expression. With the development of molecular biology techniques, the transcriptional activity of X-linked genes could be directly monitored by examining their RNA transcripts using reverse transcribed PCR (RT-PCR). Advances in discovery and detection of transcribed single nucleotide polymorphisms (SNPs)
allow qualitative and quantitative analysis of allelic expression, facilitating the comparison of
gene activity between the Xa and Xi [25]. Moreover, microarray analysis has also been
applied to study the large-scale transcription profile of X chromosomes [169-172]. Methods to
monitor gene activity at foci of transcription, such as RNA FISH, also provide a direct and
decisive way to determine XCI status [96].

Approaches to analyze polymorphic genes are complicated by the random nature of XCI
that gives rise to mixed cell populations with either X being inactivated. Therefore,
polymorphic genes are usually analyzed in cells with the same X inactivated. For instance,
non-random XCI occurs in cell lines derived from females who carry a structurally abnormal
X [173]. Mice with an X;autosome translocation such as T(X;16)16H (or T16H, Searle’s
translocation) inactivate the normal X in 100% of cells (e.g. [174]). Another useful system to
study human X chromosomes is mouse somatic cell hybrids that contain a single human Xi or
Xa [173, 175, 176]. In mouse, cell lines carrying the same Xi in all cells have been isolated by
subcloning [177], or by selection for or against the function of X-linked genes (such as \textit{Hprt})
in heterozygous animals [178].

The locus encoding the human blood cell antigen \textit{XG} was the first gene discovered to
escape XCI [24]. Notably, escape from XCI is not a rare event, at least on the human Xi. By an
extensive survey of 624 human X-linked genes and EST transcripts using the mouse-human
somatic hybrids, Carrel and Willard reported that approximately 15% of them always escape
XCI [25]. Microarray analyses to compare transcript levels between normal individuals and
patients with X chromosome aneuploidy [169], as well as between males and females
[170-172], also found some escape genes that have been observed in hybrid cells. However,
expression of the vast majority of X-linked genes is not significantly higher in female samples
compared with male samples [170-172]. Many escape genes were not identified by microarray
analysis likely because allelic expression levels vary between individuals [179]. Alternatively,
escape genes are expressed lower from the Xi than from the Xa, yielding subtle changes in the
transcript level that are not detectable using the microarray approach.

\subsection{1.2.2 Distribution and evolution of human escape genes}

Escape genes are located along the entire length of the human X; however, their
distribution is not random (Figure 1.4) [25]. The majority of them cluster and map to the XAR
on the short arm [25], in which the sequence is evolutionarily young [14]. Among the five evolutionary strata defined, the youngest S5 has the highest density of escape genes [25]. This density decreases at regions with longer history on the X, with the oldest S1 showing the lowest level [25]. The most recent regions, PARs, seem nearly resistant to XCI, as only two out of the sixteen genes tested are inactivated [25]. The correlation between the density of escape genes and evolutionary history of regions on the X reflects a potential influence of X chromosome evolution on XCI status. It is likely that escape genes are associated with one or more features promoting Xi expression whereas loss of these features during evolution causes inactivation.

A feature that may influence Xi expression status is the retention of a functionally equivalent Y homologue. The purpose of XCI is to balance the dosage of X-linked genes that are disomic in females and monosomic in males. Thus, there is no need of XCI for genes that have two copies in both sexes. For instance, human PARs are located on both X and Y chromosomes and largely escape XCI [25]. Among the 1,098 genes annotated on the human X, only 54 have a functional homologue on the Y [8]. 35 of them have been tested for the XCI status and only four are X-inactivated, including two pseudoautosomal genes, SPRY3 and SYBL1, and two X-specific genes, TSPYL2 and RBMX [25]. Intriguingly, SPRY3 and SYBL1 are inactivated on both Xi and Y chromosomes [180]; the Y homologues of TSPYL2 and RBMX have differentiated from their X-linked copies and acquire male-specific functions [181, 182]. Hence, all human genes examined that have a functionally equivalent Y homologue escape XCI. Among the mouse genes tested, all of the five genes with a Y homologue that is functional and ubiquitously expressed escape XCI (Table 1.1) [183]. Furthermore, mouse Zfx and Ube1 are X-inactivated and their Y homologues are only transcribed in testis, whereas human ZFX and UBE1 escape XCI and their Y homologues are ubiquitously expressed [177]. Therefore, dosage compensation appears to be important for all X-linked genes with a Y homologue. Importantly, genes that have a functionally equivalent Y homologue always escape XCI, suggesting these genes may be under selection for Xi expression. This is consistent with the Y-driven hypothesis that loss or divergence of Y homologues induces the inactivation of X-linked genes [27].
Figure 1.4  Human X chromosome inactivation profile. The XCI status of 624 genes or EST transcripts on the human X chromosome were determined in mouse somatic cell hybrids containing inactive human X. The results from nine independent hybrid cell lines are shown here and each column represents one cell line. The XCI status of each gene is displayed linearly with the color indicated in the box. Positions of centromere (CEN) and XIST are also shown. Image source: [25].
Table 1.1  Mouse genes that escape X chromosome inactivation and their human orthologues. Chromosomal positions are obtained from UCSC genome browser. (Mouse: Build 37, July 2007 assembly; Human: Build 36.1, March 2006 assembly). Modified from

<table>
<thead>
<tr>
<th>Gene</th>
<th>XCI status</th>
<th>Chromosome location (Mb)</th>
<th>Y homologue</th>
<th>Expression from Y</th>
<th>Reference</th>
<th>Gene</th>
<th>XCI status</th>
<th>Chromosome location (Mb)</th>
<th>Y homologue</th>
<th>Expression from Y</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ddx3x</td>
<td>Escape</td>
<td>13</td>
<td>Ddx3y</td>
<td>Ubiquitous</td>
<td>[183]</td>
<td>DDX3X</td>
<td>Escape</td>
<td>41</td>
<td>DDX3Y</td>
<td>Ubiquitous</td>
<td>[184]</td>
</tr>
<tr>
<td>Utx</td>
<td>Escape</td>
<td>18</td>
<td>Uty</td>
<td>Ubiquitous</td>
<td>[185]</td>
<td>UTX</td>
<td>Escape</td>
<td>44</td>
<td>UTY</td>
<td>Ubiquitous</td>
<td>[185]</td>
</tr>
<tr>
<td>Enox</td>
<td>Escape</td>
<td>47</td>
<td>Not found</td>
<td>-</td>
<td>[191]</td>
<td>ENOX</td>
<td>Escape</td>
<td>130</td>
<td>Not found</td>
<td>-</td>
<td>[186]</td>
</tr>
<tr>
<td>Eif2s3x</td>
<td>Escape</td>
<td>91</td>
<td>Eif2s3y</td>
<td>Ubiquitous</td>
<td>[187]</td>
<td>EIF2S3</td>
<td>Escape</td>
<td>24</td>
<td>Not found</td>
<td>-</td>
<td>[187]</td>
</tr>
<tr>
<td>Jarid1c</td>
<td>Escape</td>
<td>149</td>
<td>Jarid1d</td>
<td>Ubiquitous</td>
<td>[188]</td>
<td>JARID1C</td>
<td>Escape</td>
<td>53</td>
<td>JARID1D</td>
<td>Ubiquitous</td>
<td>[188]</td>
</tr>
<tr>
<td>Mid1</td>
<td>Escape</td>
<td>166</td>
<td>Mid1y</td>
<td>Loss of function</td>
<td>[189]</td>
<td>MID1</td>
<td>Inactivated</td>
<td>11</td>
<td>Not found</td>
<td>-</td>
<td>[189]</td>
</tr>
<tr>
<td>Sts</td>
<td>Escape</td>
<td>167</td>
<td>Sts</td>
<td>Ubiquitous</td>
<td>[192]</td>
<td>STS</td>
<td>Escape</td>
<td>7</td>
<td>STSP</td>
<td>Loss of function</td>
<td>[190]</td>
</tr>
</tbody>
</table>
On the other hand, not all escape genes have a functional Y homologue. Among ~100 human escape genes, only 31 have a functional counterpart on the Y [25]. The remainder of human escape genes and two mouse escape genes have a truncated copy on the Y or do not have a Y homologue at all (Table 1.1). These genes presumably have a higher expression level in females than in males and thus may contribute to gender differences and female-specific functions [193]. Alternatively, equal transcript dosage between two sexes is not important for these genes [194].

1.2.3 Species difference of escape genes

Mouse escape genes appear to be arranged quite differently from human. In contrast to the large number of human escape genes, only seven genes on the mouse X have been found to escape [183]. As only a limited number of mouse X-linked genes have been tested and the selection is biased to those that have their human orthologues escaping XCI, it is difficult to conclude how frequently mouse genes escape XCI. Nevertheless, the fact that many orthologous genes escape in human but are X-inactivated in mouse suggests a more complete XCI in mouse. This is also consistent with the observation that X monosomy (XO) mice have relatively normal phenotypes while human patients with Turner syndrome (45,X) show severe effects [195]. Another difference between species is the distribution of escape genes. Most of human escape genes reside within clusters each of which contains multiple continuous escape genes [25]; however, none of the seven mouse escape genes are adjacent to each other [51, 196].

There are several possibilities that can explain the species variation of escape profiles. Xist/XIST RNA and the Xic/XIC regions appear to be different in mouse and human [86, 104]. They may affect the efficiency of Xist/XIST RNA spreading, as well as gene silencing mediated by these RNA molecules [197]. The spreading of XCI may also be influenced by specific chromosomal structures, such as the centromere, a heterochromatic and highly compact structure associated with multiple layers of proteins. The centromere on the human X separates two arms and may impede the cis spreading of XIST RNA that is expressed from its genomic locus on the long arm, causing more genes to escape on the short arm [25, 198]. The mouse X centromere lies at one end of the chromosome and perhaps allows more efficient spreading of Xist and XCI. However, centromere location cannot explain why a few human
escape genes are still located on the long arm of the human X or why other short arm genes get silenced.

Mouse and human X chromosomes are also organized differently, which may explain the different escape profiles between species. Independent evolutionary history after species divergence leads to varied sequence features and chromosome organization between human and mouse Xs. Despite high gene conservation, the sequence and evolutionary strata on the mouse and human Xs are highly rearranged with respect to each other [17]. These rearrangements, as well as different sequence composition, may change global or local chromosomal structures that affect heterochromatinization and long-range gene silencing. Investigating the role of chromosome organization in escape regulation is an important goal of my thesis research.

Despite the different escape profiles between species, major similarities may still exist in the escape mechanisms. Six mouse escape genes have their human orthologues also escaping XCI (Table 1.1). Mouse Mid1 is the only exception. Unlike its human counterpart present at an X-specific location, Mid1 encompasses the pseudoautosomal boundary of laboratory mouse X chromosome and may be regulated by a different mechanism controlling PAR genes [189]. Therefore, orthologous genes that are X-specific and escape XCI in mouse also escape in human. Many genes, including both escape and X-inactivated, retain the same XCI status in different species such as human, mouse, bovine and mole [199, 200]. Moreover, as a well established system for XCI studies, mouse somatic cell hybrids can stably maintain the human Xi. The genes on it, including both escape and X-inactivated genes, share the same XCI status and epigenetic features as those maintained in primary human cell lines [25, 201, 202]. The human Xi is faithfully duplicated in a mouse genetic background, suggesting common XCI regulation between species. Therefore, studying the escape mechanism in one species will help understand how genes may escape in others.

1.2.4 Epigenetic features of escape genes

Escape genes on the Xi, similar to their Xa alleles, carry the epigenetic features that are characteristic of active transcription and euchromatin. For instance, they are enriched for H3K4 tri-methylation and H4 hyperacetylation while depleted of H3K9 methylation at their promoters [110, 116]. Their CpG islands are also hypomethylated [27, 203]. Escape genes
generally replicate earlier than X-inactivated genes, although later than their Xa alleles [134]. In contrast, X-inactivated genes, which occupy much of the Xi, are linked to the features typical of transcriptional silencing and endow an overall heterochromatic appearance for this chromosome. Escape and X-inactivated genes also differ in the distribution of certain modifications. Monoallelically expressed genes, including those that are X-inactivated and imprinted, are associated with H3K4 di-methylation at their promoters only. Biallelically expressed genes, such as escape and autosomal genes, have this mark throughout their promoters and exonic regions [204]. Therefore, XCI status correlates with specific epigenetic marks and histone modifications, which constitute a “histone code” for Xi gene expression.

### 1.2.5 Expression pattern and level of escape genes

Some escape genes do not always escape XCI. In the extensive survey of human X-linked transcripts, ~10% are expressed from the Xi in only a proportion of hybrid cell lines and human fibroblast lines [25]. The variability of Xi expression for some escape genes has been confirmed using different approaches and model systems [25, 169-171, 173, 205]. These studies showed that certain genes escape XCI depending on individuals, tissues and cell lineages [202, 205] and some genes can even alter the XCI status with time of development [206, 207]. The escape genes with a variable expression status tend to lack CpG islands [25]. Thus, deficient DNA methylation at the 5’ end of escape genes may fail to stabilize the inactivation and cause frequent reactivation.

Does an escape gene on the Xi behave just like its Xa allele? In fact, Xi expression levels are usually reduced as compared to Xa levels, such that most escape genes partially escape XCI. Lower levels of Xi expression were first seen for the human gene \textit{STS} by quantitatively measuring enzymatic activity of the two allelic products [190]. Another escape gene, mouse \textit{Jarid1c}, is found to be transcribed at a level of 25-50% from the Xi relative to the Xa, depending on tissues and development stages [174, 208]. As mentioned earlier, reduced Xi expression may explain why the gender difference for many escape genes are not readily seen in microarray analyses (see section 1.2.1) [170-172].

### 1.2.6 Escape from spreading or maintenance of chromosome X inactivation
As XCI involves spreading of gene silencing and subsequent maintenance of the inactive state, escape from XCI can take place at either stage. The spreading of initial silencing may skip some genes, or these genes are initially inactivated but reactivate as a result of failure to maintain the silencing. One study reported that an escape gene was inactivated specifically during early development and concluded that escape genes are subjected to an initial silencing and progressive reactivation [209]. However, another group did not observe the reactivation process as the inactivation retained in the same proportion of cells during development [96]. Additionally, two other studies did not see the inactivation of the escape gene in adult cells [174, 208]. Moreover, inactivation of the autosomal boundary on a translocated X chromosome is limited [210]. This attenuation appears from the onset of XCI, suggesting escape occurs at the initial spreading of XCI. However, it has not been definitively determined whether escape genes undergo a transient inactivation, and different escape genes may utilize different strategies.

1.2.7 Biological and medical significance of escape genes

Escape genes are candidates for the phenotype observed in individuals with X chromosome aneuploidies. As described in section 1.2.1, human females with a single X (45,X) are largely lethal in fetuses and live-born offspring have Turner syndrome [158]. As 46,XX females and 46,XY males are normal, it is likely that escape genes with functional Y homologues are responsible for this disease and other X-linked haploinsufficiency disorders. A gene that can be expressed from both Xs, *SHOX*, has been related to the short stature conditions of Turner syndrome [19]. A candidate region responsible for the neurocognitive phenotype has been mapped to a locus containing many genes that escape XCI [211]. Escape genes may also be involved in Klinefelter Syndrome (47,XXY) and multiple X Syndrome (47,XXX, 48,XXXX and 49,XXXXX) because patients manifest more severe phenotypes with the presence of additional Xs. Thus, appropriate dosage of escape genes plays an important role in ensuring a normal phenotype in human.

Escape genes have important implications for some X-linked skin disorders as well. A well-known example is X-linked recessive ichthyosis caused by steroid sulphatase (STS) deficiency. In this disorder, males are severely affected, whereas females heterozygous for the *STS* gene are normal and do not even display a mosaic pattern because this gene escapes XCI
In some other skin diseases, female carriers are either unaffected or less severe compared to males and lack the mosaic distribution, indicating that the underlying gene loci escape XCI [213].

Escape genes may also play a role in potential differences between two sexes. Many sex differences have been attributed to hormonal differences between males and females. However, some sexually dimorphic phenotypes occur before the emergency of hormonal differences [214, 215] or independent of gonadal sex [216, 217], suggesting that sex chromosome genes contribute directly to sexual differentiation, especially in brain development and function. As escape genes are expressed at different levels between two sexes and their Y partners, when present, often differ in function or expression levels despite sequence similarities [183, 193], these genes may participate in sex differences in development, as well as behavior and disease susceptibility.
1.3 REGULATION OF ESCAPE GENES ON INACTIVE X CHROMOSOMES

How genes acquire proper expression patterns on the Xi is a key question to understand XCI mechanisms. Escape genes provide a unique window to study Xi gene regulation as any explanation for the establishment of XCI should incorporate genes that are not affected by inactivation. While epigenetic modifications are important to maintain XCI status, a critical question is how escape genes are initially differentiated from the genes to be silenced on the same chromosome. As described earlier, inactivation on the Xi is likely regulated at a level of chromosomal regions. Evidence exists for both gene-specific and regional control of escape genes; however, it is more likely escape from XCI is regulated regionally.

1.3.1 Gene-specific control of escape genes

A potential factor to determine escape from XCI is regulatory elements that constitutively promote the expression of individual genes regardless of the transcriptional activity of adjacent regions (Figure 1.5A). Such elements are usually located in close proximity to target genes, such as strong promoters and enhancers. As described in section 1.2.4, escape and X-inactivated genes differ in many epigenetic modifications, such as elevated levels of histone acetylation and unmethylated CpG islands at escape genes [110, 194]. Notably, analysis of the sub-genic localization of these epigenetic features by chromatin immunoprecipitation revealed that many modifications were promoter-specific and absent in gene body or intergenic regions [110], suggesting the importance of promoter regions in influencing XCI status.

However, studies on the promoter regions did not find their role in escape. Comparison of the promoter of human ZFX, which escapes XCI, and mouse Zfx, which is X-inactivated, identified highly conserved sequences and completely identical CpG islands [218]. Investigation of the 5’ region of the mouse escape gene Jarid1c revealed a promoter typical of housekeeping genes and no unique features can explain the different XCI status between Jarid1c and nearby genes (see section 1.5) [219]. Thus, the overall structure and primary sequence of promoter regions seem irrelevant to the ability to escape XCI. Moreover, X-linked transgenes containing a variety of strong promoters and enhancers are subject to XCI (see section 1.4). Although only a few promoters and enhancers have been studied respect to XCI, these regulatory elements may not determine whether genes escape or not.
Figure 1.5 Models of escape from X chromosome inactivation. (A) Gene-specific regulatory elements such as unique promoters or enhancers (green triangle). (B) Regional control mechanisms. (i) Escape regions lack, or are depleted in, sequences that propagate inactivation (black oval). (ii) Escape regions contain, or are enriched for, sequences that prevent inactivation or heterochromatin formation (red star). (iii) Boundary elements prevent heterochromatin spreading and maintain active domains (black sandglass).
A. Gene-specific control

B. Regional control

---

Genomic sequence

Yellow: X-inactivated gene

Blue: escape gene

Gene-specific cis regulatory element

Sequence that propagate inactivation

Sequence that prevent inactivation

Boundary element
Instead, these elements may be responsible for transcription after the decision to escape is made. This is more likely to be true given the facts that promoter-specific modifications are present at X-inactivated genes and these genes are likely controlled coordinately.

Gene-specific control of escape has also been suggested by the observation that some escape genes are scattered on the X and lie quite close to adjacent inactivated transcripts [219]. For instance, the human escape gene *EIF1AX* is flanked by two X-inactivated transcripts, which are only 11 kb and 13.5 kb away, respectively [25]. However, this escape gene organization is extremely rare on the human X [25]. Mouse escape genes that have been identified are not adjacent to each other, and three of them are flanked by X-inactivated genes ([220] and Laura Carrel, unpublished data). Although escape genes may be regulated differently in mouse and human, the XCI profile, especially for mouse, has not been extended to all X-linked genes. Escape genes may reside in areas that are small in size but contain multiple escape genes, which are less likely to be regulated individually. Even the genes flanked by X-inactivated transcripts may be located within a domain that is regulated by regional mechanisms (see section 1.3.2.3).

### 1.3.2 Regional control of escape genes

The vast majority of human escape genes cluster, suggesting expression of these genes from the Xi is controlled at a regional level [25]. Two extensive assays using mouse somatic cell hybrids containing a human Xi and non-randomly X-inactivated primary cell lines clearly demonstrated that most human escape genes are immediately adjacent to each other and form large blocks of active transcripts [25]. The continuous distribution of human escape genes gives rise to a chromosome with alternating domains of inactive and active transcription. Thus, it is likely that certain regions of the X are more susceptible to escaping and genes within such regions are controlled coordinately. Intriguingly, most human escape genes are located within a domain that contains at least one escape gene with a functionally equivalent Y homologue [25], suggesting that, if retention of a functional X-Y pair is the mechanism for escape, it may affect nearby genes on the X.

Cytogenetic analysis also revealed that epigenetic features on the Xi were present regionally. Differences in replication timing, localization of H4 acetylation, tri-methylated H3K4, and MacroH2A1 are seen at a regional level on the human Xi [108, 110, 221-223].
Thus, the entire human Xi is separated into heterogenous chromatin domains, within each of which chromatin marks are consistent with transcriptional activity [149]. Large regions associated with epigenetic marks of active transcription are also frequently seen on the autosomal material translocated to the human or mouse X [128-131, 136-138, 146, 224-226], suggesting some chromosomal regions do not undergo long-range inactivation. Hence, genomic environment may functionally distinguish escape and X-inactivated regions on the Xi and contribute to the regional control of Xi expression, but the underlying mechanism that establishes the epigenetic domains on the Xi remains unknown.

Sequence organization along human X chromosomes also correlates with Xi expression status. As mentioned earlier, the proportion of escape genes within each evolutionary stratum on the human X correlates with the evolutionary history of the chromosomal region (see section 1.1.1). Importantly, different X strata have been subjected to restricted recombination and other X chromosome evolution events for different lengths of time, resulting in different sequence composition [8]. Sequence analysis also revealed that human X-inactivated and escape regions differs significantly. Escape regions have unique sequence features and specific regulatory elements such as repetitive sequences and binding sites for chromatin insulator protein CTCF (CTCCC binding factor), consistent with a regional control of escape genes.

1.3.2.1 Repetitive sequences

1.3.2.1.1 Long interspersed nuclear element 1

The L1 hypothesis (see section 1.1.7.1) proposes that L1s act as “booster elements” to promote XCI propagation along the X. An extension of this model is that escape genes are located at regions deficient of L1s. The L1 content in escape regions is concordant with this scenario on the human X [25, 155]. Distribution of repeat elements including L1s has not been systematically investigated on the mouse X in relation to XCI status. However, L1s are deficient at the Jarid1c genomic region, the only mouse escape gene tested [196]. Notably, in a mouse X;autosome translocation, the spreading of Xist RNA is limited at the autosomal boundary with low L1 content [210].

Recently, two groups performed independent analyses of human X chromosome sequences and searched for sequence motifs that are associated with genes of certain XCI
status [227, 228]. They found that the oligonucleotides overrepresented in the vicinity of X-inactivated genes mapped predominantly to repetitive elements including L1s, particularly the evolutionarily younger L1s. These motifs are highly effective in predicting both X-inactivated and escape genes along the entire human X, strengthening the correlation between specific genomic sequences and XCI status. Surprisingly, most informative motifs provide their best prediction in the largest window tested (250 kb), suggesting regulatory sequences important for Xi expression may be located within relatively large distances from each gene [227].

If L1s are important for heterochromatinization and gene silencing as suggested by Lyon [151], chromosomal locations depleted of L1s may lack the capacity to promote inactivation, allowing genes to be expressed at these loci (Figure 1.5B). However, whether L1s plays any role in XCI has not been determined. Even though they do, the mechanism may be complex since a fine mapping analysis found that the L1 level was not reduced throughout an entire escape region [196]. Thus, L1 distribution does not completely correlate with XCI status.

1.3.2.1.2 Other repetitive elements

Escape and X-inactivated regions on human Xs also vary in the content of other sequence elements. Another sub-class of LINEs, L2s, is also deficient in escape regions and serves as one of the most informative features to predict XCI status [228]. Escape regions are also depleted of mammalian-wide interspersed repeats (MIRs), a sub-family of short interspersed nuclear elements (SINEs), and long terminal repeats (LTRs) [228, 229]. One LTR sub-family, MLT1Ks, is virtually excluded from escape regions and the absence predicts escape perfectly [228]. However, the overall content of L2s, MIRs or LTRs on the human X is approximately equal or slight lower compared to autosomes, which cannot explain the incomplete inactivation of autosomal material [8].

Unlike the sequence elements mentioned above, some repetitive sequences are enriched in human escape regions, such as Alus, another sub-family of SINEs, and [GATA]n repeat, a microsatellite DNA sequence [228, 230]. Additionally, the mouse X has a dramatically lower level of B1 elements (mouse Alus) than the human X, consistent with the potentially complete XCI in mouse [227]. How can they regulate escape gene expression? Alu sequences contain CpG islands that do not undergo de novo methylation [231] and thus may allow Xi expression
by preventing methylation at their CpG islands [227]. The \([\text{GATA}]_n\) repeat can recruit transcription factors GATA family and SATB1, both of which have been shown to be expressed in early embryonic cells and to integrate gene regulation with high-order chromatin architecture that involves looping and chromatin domains [232-235]. Thus, the sequence elements specifically enriched in escape regions provide an additional possibility for the regional control of escape. Escape from XCI may be due to the existence or enrichment of elements that prevent heterochromatinization or gene silencing, rather than the lack of sequences that promote inactivation (Figure 1.5B) [230].

If repeat content influences \(X_i\) expression status, it is likely that XCI status depends on more than one sequence element [228]. A combination of sequence motifs from multiple repeats, including LINEs, SINEs, LTRs, DNA transposons and simple repeats, provides the most effective predictor for XCI status in a computational study of the human X [228]. These sequence elements may act together to organize the X chromosome into domains with varied susceptibility to inactivation. Some of these elements may serve to stabilize the \(XIST\) RNA association with the X, and others may facilitate the propagation of this RNA along the chromosome or may be involved in different procedures of heterochromatinization such as histone modifications, DNA methylation and architectural changes. While studies on L1s and other repeats strengthen their correlation with XCI status, a functional system to delimit the amount of genomic sequence that may influence XCI status has been missing. Importantly as a first step, my thesis research aims to functionally define the maximum size of an escape domain that may be influenced by such elements.

1.3.2.2 Chromatin insulator CTCF and boundary elements

Several boundary regions between escape and X-inactivated genes in both human and mouse contain sequences with insulator and CTCF binding activities (Figure 1.6) [220]. Mouse \(\text{Eif2s3}\) and human \(\text{EIF2S3}\) are located at regions of conserved synteny and both escape XCI. CTCF protein binds to their 5’ ends, each immediately adjacent to an X-inactivated gene. At another syntenic locus, both mouse \(\text{Jarid1c}\) and human \(\text{JARID1C}\) escape XCI. However, CTCF preferentially binds to the 5’ end of mouse \(\text{Jarid1c}\), between this gene and an X-inactivated gene \(\text{Iqsec2}\). In contrast, the highly conserved region at the human \(\text{JARID1C}\) locus, which is embedded within a large escape domain containing multiple escape
Figure 1.6  CTCF protein specifically binds to several boundaries between escape and X-inactivated genes. (A) CTCF sites are present at the 5’ end of mouse Eif2s3x and human EIF2S3, between genes of different XCI status. (B) CTCF binds to the 5’ end of mouse Jarid1c, which is adjacent to an X-inactivate gene. However, CTCF cites are absent at the conserved region of human JARID1C that is located between two escape genes. Image is modified from [220].
A

i. mouse

\[\text{Zfx} \quad \text{Eif2s3x} \quad \text{Klh15}\]

\[\text{mouse}\]

ii. human

\[\text{ZFX} \quad \text{EIF2S3} \quad \text{KLHL15}\]

\[\text{human}\]

B

i. mouse

\[\text{Iqsec2} \quad \text{Jarid1c} \quad \text{Tspyl2}\]

\[\text{mouse}\]

ii. human

\[\text{IQSEC2} \quad \text{JARID1C} \quad \text{TSPYL2}\]

\[\text{human}\]

Legend:
- **Yellow**: X-inactivated gene
- **Blue**: Escape gene
- **Red**: CTCF binding sites
- **Black dot**: Centromere
genes, is absent of CTCF binding sites and does not elicit an insulator activity (Figure 1.6). Notably, the role of CTCF in regulating Xi expression is supported by specific CTCF binding at the boundaries on the Xi, but not on the Xa [220].

Although CTCF is a transcription activator among other roles, it is less likely to directly promote Xi expression because the 5' end of Jarid1c does not show any repressive activity [219]. Additionally, mouse Jarid1c and human JARID1C are highly conserved in the genomic sequence and have similar expression patterns, except that the human gene does not bind CTCF at the 5' end but still escapes XCI [220].

CTCF protein is heavily involved in gene regulation by preventing the spreading of heterochromatin. CTCF is a ubiquitously expressed nuclear protein and contains eleven zinc finger motifs [236]. Through the combination of different zinc fingers, this protein can bind to diverse DNA sequences and control expression of numerous genes as an enhancer blocker or boundary element [237]. It has been shown that CTCF regulates some epigenetic loci such as genomic imprinting of H19/Igf2 and Tsix/Xist [238, 239]. CTCF binding sites can shield from position effect in yeast and block the spreading of repressive telomeric chromatin [240]. At the β-globin locus in a range of vertebrates, CTCF plays an important role in long-range gene regulation by establishing an active chromatin domain and protecting genes from being silenced due to proximity to inactive heterochromatin [241]. CTCF is also implicated in several well-characterized chromatin insulators of higher eukaryotic organisms [236, 242]. These CTCF-bound insulators can protect genes from the activity of nearby enhancers, or act as chromatin barriers to prevent the spreading of adjacent heterochromatin [236, 237, 243].

In the case of XCI, CTCF sites may function as boundary elements to establish chromatin domains and protect genes within these domains from silencing due to proximity to inactive chromatin or inactivation machinery (Figure 1.5B). As mentioned earlier, CTCF specifically recognizes several boundaries between escape and inactivated regions, [220], suggesting DNA sequences that recruit CTCF may insulate domains on the Xi and thus escape genes are regulated at the level of chromatin domains. Additionally, CTCF may also shield X-inactivated genes from activation by nearby escape domains. Recent studies suggested that CTCF mediated the formation of chromatin loops by protein dimerization or tethering to the nucleolus [244, 245]. Immunostaining analysis showed an intense focus of CTCF protein within the Barr body [112], suggesting that it may form a subnuclear compartment to confine
boundaries. Hence, escape and inactivated domains may be separated into chromatin loops by CTCF and occupy different nuclear compartments. By this manner, genes located on each loop can be coordinately regulated [96, 220].

The DNA binding activity of CTCF is methylation-sensitive and plays an important role in preserving an unmethylated status at the binding sites [241, 245]. As CpG island methylation is absent in escape domains but necessary to maintain silencing in inactivated regions [27, 203], CTCF may contribute to the regional control of escape through binding to unmethylated domain boundaries, such as the 5’ end of Jarid1c, and preventing the spreading of DNA methylation [220].

Despite CTCF binding between escape and X-inactivated genes, its function in establishing escape domains has not been proven. A GFP transgene flanked by CTCF binding sites is not resistant to XCI at the inactivated mouse Hprt locus, even though CTCF clearly functions on the Xa to insulate this transgene from surrounding chromatin effects compared to uninsulated controls [246]. Hence, CTCF binding alone is not sufficient for Xi expression and escape may require additional boundary elements that have not been identified.

In fact, whether boundary elements insulate escape domains on the Xi has not been addressed. In addition to CTCF sites, other sequence elements that may insulate chromatin domains include matrix attachment regions (MARs), which have been shown to organize chromatin into structural domains and protect genes from position effect [178, 247]. However, a reporter transgene flanked by MARs is not protected from XCI or CpG island methylation [178]. Several potential binding sites for transcription factor YY1 have been discovered at the 5’ end of Jarid1c [219, 220]. YY1 has been proposed to be involved in chromatin insulation [248]. However, unlike CTCF, potential YY1 binding sites are also present at human JARID1C but not at Eif2s3x/EIF2S3. If boundary elements exist, they are not marked by significantly divergent sequences since a boundary region in human that is located within an inactivated domain in mouse remains high sequence similarity [220]. This makes it somewhat difficult to identify boundary elements by simple sequence comparison. Thus, defining an escape domain containing potential boundary elements is of great help to dissect for these elements.

1.3.2.3 Regional control on the mouse X
Despite the apparent organizational differences between mouse and human escape genes, important features are conserved and argue that escape genes are not simply under a gene-specific control in mouse. As described earlier, escape genes in mouse and human share many characteristics that are consistent with a regional control of Xi expression, such as the distribution of repetitive sequences and potential boundary elements [196, 220]. Human orthologues of mouse genes that are X-specific and escape XCI are also expressed from the Xi, suggesting orthologous genes are regulated by similar mechanisms [183]. Although mouse escape genes identified are not adjacent to each other, in extension of the domain model presented above, such an event may occur if a boundary element shifted or if heterochromatin formation was limited in one species. For instance, if CTCF contributes to boundary formation on the Xi, the absence of CTCF binding sites at \textit{JARID1C} may lead to a much larger domain in human. Alternatively, if any sequence features determine domain formation on the Xi, the regions with such features may diminish in size in mouse. It should be noted that the gene-specific and regional control mechanisms are not mutually exclusive. They may regulate different genes on the X chromosome and each escape gene may be controlled by either or both of these two pathways.
1.4 X-LINKED TRANSGENE STUDIES

A useful approach to study Xi gene regulation is to examine how X-linked transgenes respond to the chromosome-wide gene silencing. A number of randomly integrated transgenes have been investigated for their XCI status (Table 1.2). Most transgenes are subject to random XCI [44, 249-256]. Only four transgenes escape XCI, including two that showed escape only in extraembryonic tissues, which undergo imprinted XCI [44, 46], and one escaped in 3% of somatic cells [45]. The only transgene that completely escapes random XCI is a large DNA fragment of 187 kb consisting of 11 tandem copies of the chicken transferrin gene [257]. It is not clear why it escapes, although the multiple copies and large size have been proposed as the potential reason [178]. Silencing of most transgenes on the Xi supports the conclusions made in the X;autosome translocation and Xist/XIST transgene studies that genes can be X-inactivated after transfer to the X and that inactivation of individual genes may not require any specific sequences. X-inactivated transgenes include some endogenous autosomal genes, but most are minigenes that are constructed using promoters, coding sequences and polyadenylation signals of various origins and sizes (Table 1.2). These prokaryotic and eukaryotic fragments can respond to XCI regulation in mouse, regardless of their primary sequences and transgene composition.

Another advantage of the transgene approach is the ability to assess whether a certain chromosomal location can influence Xi expression status, by monitoring a variety of transgenes at this locus. The only locus that has been extensively investigated is the mouse Hprt gene, which is subject to XCI [258, 259]. Thanks to a well-established homologous recombination technology [260], single-copy genes of interest can be inserted into Hprt. To date, numerous genes have been targeted into this site and ten studies investigated transgene XCI status using cells or tissues from transgenic mice (Table 1.3). These transgenes include synthetic transcription units, as well as endogenous genes and even large genomic fragments. In nine studies, transgene expression was uniform in homozygous females and hemizygous males, but variegated in heterozygous females, consistent with transgene expression only in cells in which the transgenes are located on the Xa [178, 246, 260-266]. Two studies found that the transgenes were expressed at a similar level in males and homozygous females, but at approximately half of the level in heterozygous females, which is also consistent with XCI of the transgenes [260, 261]. Thus, it is concluded that these transgenes are subject to random
Table 1.2  X chromosome inactivation studies of transgenes randomly integrated into the mouse X chromosome.

Abbreviation: Tg: transgene; Chr: chromosome; kb: kilobase; #: number; R: random; I: imprinted; -: inactivated; +: escape; Ref: reference; n/a: not applicable; abbreviation of genes see page viii.

Notes:
1 Only mammalian chromosomes are shown.
2 Integration sites are shown as the cytogenetic band location.
3 Random XCI was tested in somatic cells or tissues, or differentiating ES cells when noted (ES). Imprinted XCI was tested in extraembryonic cells or tissues.
4 The XCI status for this transgene was inferred from its methylation status.
5 two independent transgenes were tested.
6 inactivated in some extraembryonic lineages but escape in trophoblastic giant cells (extraembryonic).
<table>
<thead>
<tr>
<th>Promoter</th>
<th>Coding region</th>
<th>Other cis element</th>
<th>Insert size (kb)</th>
<th>Copy #</th>
<th>Insertion sites</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>mouse <em>Afp</em> (~5 kb, chr 5)</td>
<td>mouse <em>Afp</em> minigene (first 3 and last 2 exons, chr 5)</td>
<td></td>
<td>7.4 or 14.4</td>
<td>2-3</td>
<td>n/a</td>
<td>-/+ [44]</td>
</tr>
<tr>
<td>chicken <em>TNF</em> (2.2 kb)</td>
<td>chicken <em>TNF</em> (17 kb genomic)</td>
<td></td>
<td>17</td>
<td>11</td>
<td>D</td>
<td>+ n/a [257]</td>
</tr>
<tr>
<td>viral SV40 early (~300 bp)</td>
<td>bacterial <em>CAT</em> (~1 kb)</td>
<td></td>
<td>~7-5</td>
<td>5-6</td>
<td>n/a</td>
<td>-4/-4 [249]</td>
</tr>
<tr>
<td>mouse <em>protamine 1</em> (~5 kb, chr 16)</td>
<td>viral SV40 large T antigen (~2.5 kb)</td>
<td>mouse <em>IgH</em> enhancer (chr12)</td>
<td>~9</td>
<td>n/a</td>
<td>n/a</td>
<td>- n/a [250], [251]</td>
</tr>
<tr>
<td>mouse <em>tyrosinase</em> (~2 kb, chr 7)</td>
<td>mouse <em>tyrosinase</em> (cDNA, ~2 kb, chr 7)</td>
<td></td>
<td>~7</td>
<td>n/a</td>
<td>n/a</td>
<td>-5 n/a [251], [252]</td>
</tr>
<tr>
<td>human <em>COL1A1</em> (chr 17)</td>
<td>human <em>COL1A1</em> (18 kb genomic, chr 17)</td>
<td></td>
<td>40</td>
<td>1</td>
<td>D/E (3%)</td>
<td>- [45]</td>
</tr>
<tr>
<td>mouse <em>HMG-CoA</em> (1.4 kb, chr 13)</td>
<td><em>E.coli lacZ</em> (~3 kb)</td>
<td></td>
<td>8.9</td>
<td>14</td>
<td>A6</td>
<td>- (ES also) - [255], [267], [279]</td>
</tr>
<tr>
<td>Viral <em>HSV-TK</em> (~2 kb)</td>
<td>bacterial <em>neo</em> (~1 kb)</td>
<td>v-<em>myc</em> gene with LTR promoter</td>
<td>~8</td>
<td>1</td>
<td>D/E</td>
<td>- - [252]</td>
</tr>
<tr>
<td>mouse <em>Mt</em> (1 kb, chr 8)</td>
<td>mouse/rat <em>MTVP</em> fusion (2 kb cDNA, chr 8/3)</td>
<td></td>
<td>3</td>
<td>10</td>
<td>C</td>
<td>- n/a [253]</td>
</tr>
<tr>
<td>chicken <em>β-actin</em> (350 bp)</td>
<td>jellyfish <em>GFP</em> (~1 kb)</td>
<td>CMV enhancer</td>
<td>~7</td>
<td>22-24</td>
<td>A3.1-3.3</td>
<td>- - [+/- 6, 254, 279, 279, 279]</td>
</tr>
<tr>
<td>human <em>β-globin</em> (6.5 kb for promoter and LCR, chr 11)</td>
<td>mouse <em>Gata1</em> (4.3 kb genomic, chr X)</td>
<td>human <em>β-globin</em> LCR</td>
<td>~13</td>
<td>n/a</td>
<td>n/a</td>
<td>- n/a [255]</td>
</tr>
<tr>
<td>chicken <em>β-actin</em> (350 bp)</td>
<td>jellyfish <em>GFP</em> (~1 kb)</td>
<td>CMV enhancer</td>
<td>~7</td>
<td>n/a</td>
<td>C2</td>
<td>- (ES) n/a [256, 281]</td>
</tr>
</tbody>
</table>
Table 1.3  X chromosome inactivation studies of transgenes targeted at the mouse *Hprt* locus

1 The random XCI status was assayed in somatic tissues.
2 The transgene was tested in both directions relative to the *Hprt*.
3 Only five cells were examined for both males and heterozygous females.
<table>
<thead>
<tr>
<th>Year</th>
<th>Promoter</th>
<th>Coding sequence</th>
<th>Other cis element</th>
<th>Random XCI status</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>chicken β–actin</td>
<td>mouse Bcl-2 (cDNA, 850 bp, chr 1)</td>
<td></td>
<td>inactivated</td>
<td>[260]</td>
</tr>
<tr>
<td>1996</td>
<td>human β–actin</td>
<td>mouse Bc-12 (cDNA, 850 bp, chr 1)</td>
<td></td>
<td>inactivated</td>
<td>[260]</td>
</tr>
<tr>
<td>2000</td>
<td>human AGT</td>
<td>human AGT (genomic, 13.8 kb, chr 1)</td>
<td></td>
<td>inactivated</td>
<td>[261]</td>
</tr>
<tr>
<td>2000</td>
<td>mouse Tie2</td>
<td>E. coli LacZ (~3 kb)</td>
<td></td>
<td>inactivated</td>
<td>[262]</td>
</tr>
<tr>
<td>2000</td>
<td>mouse Tie2</td>
<td>E. coli LacZ (~3 kb)</td>
<td>mouse Tie2 enhancer (1.7 kb, downstream of polyA, chr 4)</td>
<td>inactivated</td>
<td>[262]</td>
</tr>
<tr>
<td>2000</td>
<td>human eNOS</td>
<td>E. coli LacZ (~3 kb)</td>
<td></td>
<td>inactivated^2</td>
<td>[263]</td>
</tr>
<tr>
<td>2002</td>
<td>endogenous chicken cLYs</td>
<td>chicken cLYs (genomic, 21.4 kb)</td>
<td>flanked by MARs, contain enhancer, silencer and OBR</td>
<td>inactivated^2</td>
<td>[178]</td>
</tr>
<tr>
<td>2004</td>
<td>human c-fos</td>
<td>EGFP (729 bp)</td>
<td>three HIV NF-κB cis elements (12 bp each) upstream of promoter</td>
<td>escape^3</td>
<td>[282]</td>
</tr>
<tr>
<td>2004</td>
<td>mouse α-MHC</td>
<td>rabbit SLN (cDNA, 210 bp)</td>
<td></td>
<td>inactivated</td>
<td>[264]</td>
</tr>
<tr>
<td>2006</td>
<td>rat PEPCK</td>
<td>phage cre (~1 kb)</td>
<td></td>
<td>inactivated</td>
<td>[265]</td>
</tr>
<tr>
<td>2006</td>
<td>human β–actin</td>
<td>jellyfish GFP (~1 kb)</td>
<td></td>
<td>inactivated</td>
<td>[246]</td>
</tr>
<tr>
<td>2006</td>
<td>human β–actin</td>
<td>jellyfish GFP (~1 kb)</td>
<td>Flanked by one or two chicken β-globin HS4 insulator (1.2 kb)</td>
<td>inactivated</td>
<td>[246]</td>
</tr>
<tr>
<td>2008</td>
<td>mouse Myo7a</td>
<td>mouse Myo7a (BAC RP23-138G2, 219 kb, chr 7)</td>
<td></td>
<td>inactivated</td>
<td>[266]</td>
</tr>
<tr>
<td>2008</td>
<td>mouse Myo7a</td>
<td>mouse Myo7a (BAC RP24-233J11, 150 kb, chr 7)</td>
<td></td>
<td>inactivated</td>
<td>[266]</td>
</tr>
</tbody>
</table>

^1 Random XCI status indicates the inactivation status of the X chromosome in males. ^2 Some studies report inactivation, while others report escape. ^3 Escape occurs when the X chromosome is active.
XCI. One study claimed a transgene escaping XCI since the reporter $\text{EGFP}$ gene was expressed in all, rather than half, of the five cells examined from heterozygous females [282]. This conclusion needs to be validated in a larger sample size that can give statistically significant results. Inactivation of all, or at least nearly all, transgenes at the $\text{Hprt}$ locus implies that this X-inactivated locus has a profound influence on inactivation at this site.

The transgene approach is also useful to assess the role of specific sequences in Xi gene regulation. Using this approach, it has been demonstrated that some candidates for boundary elements, such as CTCF sites and MARs, alone cannot establish an escape domain on the Xi (see section 1.3.2.2). Locus control regions (LCRs) are also involved in long-range gene regulation and can enhance expression of many linked genes [283]. However, the transgene regulated by a LCR is X-inactivated [255], indicating it is not sufficient for escape. Some X-inactivated transgenes also contain other cis elements, such as strong promoters or enhancers, but they are not capable of escaping either [178, 249, 254, 256, 262, 280]. For the $\text{EGFP}$ transgene that may escape XCI [282], if additional experiments confirm its Xi expression, it will be important to determine whether the three binding sites for the transcription factor NF-κB present in the transgene lead to escape and also whether these sites explain the regulation of any endogenous escape genes.

Despite the power of transgene analysis, important questions remain elusive. Many of the transgenes tested have chimeric transcripts and are derived from mammalian autosomes or the genome of vertebrates that do not carry Xs or even prokaryotic organisms. It is not clear how non-mammalian sequences should respond to XCI and whether mammalian autosomal sequences are regulated similarly to X-linked genes. Only one transgene contains sequences from mammalian X chromosome, the X-inactivated mouse $\text{Gata1}$ [255]. However, this transgene contains only the $\text{Gata1}$ coding region and it is not clear whether it contributes to the inactivation [255].

The transgenes used also diverge in size, ranging from several to hundreds of kb. Some of them have different copy numbers, even as many as eleven tandem repeats, and many have not been characterized. Except the transgenes that were targeted into the $\text{Hprt}$ locus, others integrate into random and ill-defined locations on the X such that it is difficult to evaluate whether their XCI status can be influenced by insertion sites. The $\text{Hprt}$ locus is the only site that has been extensively studied, but it is unclear whether it is representative for all
X-inactivated loci. Because of these drawbacks, conclusions that can be made from these transgene studies are limited and preliminary at best.

The transgene approach has some other applications as well. X-linked GFP and LacZ reporter genes have been used to indicate the XCI status of transgenic X chromosomes in different tissues and developmental stages [46, 254, 256, 269-278, 280]. As exogenous reporters, these transgenes can also be used to screen for tran-acting factors that influence the XCI event. For instance, GFP transgene reactivation caused by gene mutations allowed the identification of proteins important for XCI maintenance, such as DNA methyltransferase, Dnmt1 [126], and cell cycle checkpoint/DNA repair proteins, Atm and Atr [284].
1.5 MOUSE JARID1C GENE

1.5.1 General features of Jarid1c

A well-studied locus on the mouse X chromosome that is the focus of my thesis work includes the escape gene Jarid1c (Jumonji AT-rich interactive domain 1c), formerly known as Xel69 and Smcx (selected mouse cDNA on the X). The Jarid1c genomic region is approximately 40.5 kb in length and contains 26 exons that encode a protein of 1,551 amino acids [196]. This mouse gene is evolutionarily conserved, with its human homologue JARID1C showing 90% nucleotide similarity and 94% amino acid identity [196]. Human JARID1C was recently identified as a demethylase for di- and tri-methylated H3K4, marks of active transcription, and therefore results in transcription repression [285, 286]. By recruiting histone deacetylases and silencing factor REST at the promoter region as well, JARID1C participates in the chromatin remodeling and transcriptional repression of specific neuronal genes in stem cells and non-neural tissues to promote neuronal differentiation [285-287]. Mutations of JARID1C are associated with X-linked mental retardation, autism and epilepsy, indicating that this gene is essential for normal brain development and function [285, 286, 288]. In addition, JARID1C is a novel oncogene and can antagonize the tumor suppressing activity of TGF-β/Smad3 signaling pathway [289].

The Y homologue JARID1D, first identified as the H-Y antigen, is highly similar to JARID1C and encodes an H3K4 demethylase as well [285]. The Y homologue in mouse, Jarid1d, is also ubiquitously transcribed [285]. Therefore, these Y homologues constitute functional X-Y gene pairs, with their X partners escaping XCI.

The Jarid1c/Jarid1d gene pair in mouse is expressed in a sex-specific manner. Females express a significantly higher level of Jarid1c in brain than males due to escape from XCI, and males do not generate sufficient Jarid1d to compensate their lower amount of Jarid1c protein [193, 290, 291]. Therefore, this gene pair may not be fully equivalent in function. The higher expression of Jarid1c is associated with the presence of two X chromosomes, irrespective of gonadal sex [291], and thus do not result from hormonal differences. Notably, Jarid1c is highly expressed only in adult brain, not in neonatal brain or adult liver [291]. The unique expression pattern of Jarid1c, as well as its specific regulatory function at neuronal genes, may lead to sex-specific brain development and behaviors such as increased aggression [193, 290, 291]. Understanding the escape mechanism of Jarid1c may therefore help elucidate...
reasons of sex differences in normal brain function and related neuropsychiatric disorders, as well as perhaps cancer susceptibility [289].

1.5.2  *Jarid1c* and escape from X chromosome inactivation

The Xi expression pattern of *Jarid1c* has been well characterized. As the first escape gene identified in mouse, *Jarid1c* has been demonstrated independently by a number of groups using different methods that the Xi allele is transcribed and associated with epigenetic marks characteristic of active transcription [27, 96, 188, 194, 208]. Xi expression levels of *Jarid1c* are reduced compared to Xa alleles, and vary depending on the cell lineage and developmental stage (see section 1.2.5, [194, 208]). Although it is still controversial whether *Jarid1c* is subject to initial silencing during XCI, it clearly escapes inactivation in all cells where XCI has been completely established [194, 208, 209]. For the purposes of my thesis work, developmental studies have established that Xi expression of *Jarid1c* is relatively easy to detect on ~12.5 dpc as the Xi expression level is constant and high at this time point, even higher than adulthood [193, 194, 209].

The *Jarid1c* escape region is relatively small and contains a single escape gene. *Jarid1c* maps to the position 148.6 Mb on the mouse X chromosome of 167 Mb (Build 37, July 2007 assembly at UCSC genome browser website http://genome.ucsc.edu/) and is located within the X-specific portion. This gene is flanked by X-inactivated transcripts, *Iqsec2* that is 8 kb upstream and EST *AK013346* that is 21 kb downstream (Figure 1.7) [196]. The XCI status of this locus is different from that of the human syntenic region. Despite highly conserved sequences and genomic organization, human *JARID1C* and four upstream transcripts escape XCI and thus generate an uninterrupted 286 kb escape domain at the human Xp11.2 locus (Figure 1.7) [196].

A number of studies have examined the *Jarid1c* region in order to understand escape mechanisms. *Jarid1c* is the only gene that escapes XCI in this region, raising the possibility that gene-specific regulatory elements within or near this gene may determine the Xi expression [219]. However, sequence analysis and functional characterization of the 2.3 kb promoter region did not identify any unique features that could explain the distinctive XCI status of *Jarid1c* compared to nearby genes [219]. In contrast, this promoter is typical of housekeeping genes, containing a CpG island and no TATA box [219]. There are two
**Figure 1.7  Schematic map of the Jarid1c/JARID1C escape region in human and mouse.** Escape genes are shown in blue and X-inactivated transcripts in yellow. Orthologous regions in genes are illustrated in shade. The transcripts included in parentheses may not be independent transcripts.
consensus binding sites for the transcription factor Sp1 in the promoter and 5’ UTR [219]. Although Sp1 can block de novo methylation at the CpG island of some genes [292], it has not been shown that this protection occurs on the Xi and contributes to escape. It should be noted that these studies do not exclude the possibilities that gene-specific elements are located elsewhere in this region or cannot be detected by the assays used. However, no evidence to date supports a gene-specific regulation of Jarid1c Xi expression.

Sequence analysis of the entire Jarid1c region has been performed to define the correlation of sequence features and XCI status. Compared with the X chromosome and entire genome, this region has a unique repeat composition. L1s are depleted at this locus and the human JARID1C gene, consistent with the L1 hypothesis (see section 1.1.7.3 and 1.3.2.1.1) [196]. However, L1 density is also low at the adjacent inactivated Igsec2 gene, and the reduction is not throughout the entire human JARID1C escape domain. Alu/B1 and MIR elements are enriched at Jarid1c/JARID1C genes, while this increase extends to adjacent regions that escape in human but is X-inactivated in mouse [196]. Additionally, MIRs are depleted at other human escape regions examined [228]. The LTR level is reduced specifically in mouse and human escape domains [196]. However, LTRs are not enriched on the X chromosome compared to the whole genome and cannot explain the incomplete inactivation of autosomal material in X;autosome translocations [196]. Therefore, repeat distribution does not correlate with XCI status perfectly at the Jarid1c locus and none of these repeat elements alone can explain Xi expression of Jarid1c. It is still unclear whether and how these repeat elements influence the Xi expression, although the composition of several repeat elements, instead of any single one of them, may be responsible to establish escape domains [228].

As described in section 1.3.2.2, CTCF binding sites are found at the promoter and 5’UTR of the mouse Jarid1c but not at the human ortholog [220]. These data suggest that CTCF protein may mark the boundaries between escape and X-inactivated genes. However, CTCF alone is not sufficient to establish an escape domain [246]. Other factor(s) may cooperate with CTCF or function alone to render Xi expression. Alternatively, CTCF may function at specific loci on the Xi, such as Jarid1c.
1.6 A MODEL SYSTEM TO STUDY X CHROMOSOME INACTIVATION

Mouse female (XX) embryonic stem (ES) cells provide a powerful *ex vivo* system for XCI studies. ES cells are pluripotent and can differentiate into all types of somatic cells [293, 294]. They are derived from the inner cell mass of embryos at the blastocyst stage, at which random XCI has not yet occurred [295]. Thus, the two X chromosomes of female mouse ES cells are active at the undifferentiated stage, and genes are expressed from both Xs [295]. Importantly, XCI coincides with ES cell differentiation. One X chromosome silences genes in *cis* upon differentiation in cell culture, recapitulating the random XCI event that occurs *in vivo* during development [295].

XCI induced in differentiating ES cells reflects *in vivo* events. A number of studies have revealed that XCI during ES cell differentiation proceeds in a stepwise manner [92, 296, 297, Brockdorff, 2002 #413]. During the first two days, *Xist* RNA starts to accumulate on the X to be inactivated, concurrent with the counting and choice steps [82, 298]. They are followed by changes in chromatin structure and gene activity on the Xi. Late replication and down regulation of genes appear on day 2-4 [296]. Histone H4 hypoacetylation on the Xi is then observed from day 4 and reaches the maximal level on day 6 [296]. The recruitment of macroH2A, formation of macrochromatin bodies and most of other epigenetic changes finish on day 7-10, suggesting a nearly complete XCI by day 10 of differentiation. CpG island methylation is a relatively late event and usually begins from day 10 [297]. The time frame of these events parallels closely those that occur in early mouse embryos [299, 300], allowing the successive processes of random XCI to be followed. Differentiated XX ES cells faithfully duplicate XCI status *in vivo* as well [51]. The epigenetic features associated with X-inactivated and escape genes in ES cells mimic those seen in cultured cells or tissues [96, 117, 296, 301]. Importantly, for my thesis, *Jarid1c* shares the same chromatin structure in tissues and differentiated ES cells [96, 117, 208].

Combined with the transgene approach, this model system has been widely used to study the molecular mechanisms of random XCI [76, 302]. By introducing genes onto active X chromosomes that can be induced to undergo XCI, such as those in ES cells, X-linked transgenes can be monitored upon *in vitro* differentiation of ES cells and concomitant XCI. Additionally, undifferentiated ES cells have not undergone XCI and therefore the insertion sites are unrelated to XCI response. As transgenic approaches are not available yet for human
ES cells, all X-linked transgene experiments were carried out in mouse ES cells or embryos, which also allow whole animals to be made to observe tissue-specific responses to XCI.

Two different procedures of in vitro ES cell differentiation have been employed to study XCI, including a spontaneous process by withdrawal of the growth factor leukemia inhibitory factor (LIF) [92, 296, 297] and an induced process by applying retinoic acid (RA) [303]. However, RA induction does not always lead to XCI, as the Xi was identified in less than 20% of differentiated cells after RA treatment for ten days and reporter genes were not always X-inactivated [256]. In contrast, spontaneous differentiation inactivates one X and the reporter gene on the Xi in nearly all cells and thus is a reliable approach to mimic the early embryonic development event [256]. Thus, our studies will examine Xi expression status in spontaneously differentiated ES cells.
1.7 RESEARCH AIM AND PLAN

How are escape gene regulated? Compared to X-inactivated genes, escape genes are associated with unique genomic and epigenetic features, such as local sequence composition, \textit{cis} regulatory elements and chromatin modifications. While epigenetic modifications are believed to be responsible for maintaining the XCI status, the key regulator(s) differentiating the ability to express from the Xi is missing. Several models have been proposed to explain escape, including repetitive sequence distribution and chromatin domain boundaries. However, boundary element candidates are not sufficient to establish Xi expression, and the repeat sequence model has not been functionally tested. Whether these factors influence Xi expression is still poorly understood. Direct investigation of existing models is somewhat difficult as the genomic regions containing candidate sequences have not been defined. Searching for the determining factors of escape will benefit from a system that definitively incorporate these factors and allow further dissection of them.

Escape from XCI is likely determined at the level of chromosomal domains. Before characterization of the elements responsible for escape, it is essential and helpful to delimit an escape domain that potentially contains all regulatory sequences necessary for escape. Such an escape domain should function as an autonomous unit on the Xi. Therefore, the hypothesis of my thesis work is that DNA fragments incorporating the entire escape region include all \textit{cis}-regulatory sequences necessary for escape and thus are expressed from the Xi independent of chromosomal location. Specifically, we will test whether an escape gene, mouse \textit{Jarid1c}, in the context of adjacent sequences and X-inactivated transcripts, can still escape XCI after it is relocated at different sites on the X.

In chapter 3, we will explain the establishment of mouse transgenic cell lines that carry an X-linked \textit{Jarid1c} transgene and the characterization of the transgenes and their integration sites (Figure 1.8). In chapter 4, we will look at the Xi expression status of the transgenes using sequential RNA and DNA FISH and allelic expression assays. Then finally we conclude in chapter 5 with significant findings of this research project and future directions that will have impacts on XCI studies.
Figure 1.8  Research plan of my thesis work.
CHAPTER II
MATERIALS AND METHODS

2.1 MODIFICATION OF BAC DNA

2.1.1 BAC clones

BAC clones RP23-330G24 (Genebank accession #: AC083816) and RP23-391D18 (Genebank accession #: AC084236) were obtained from BACPAC resources (http://bacpac.chori.org/). These two clones contain the genomic sequences of *Jarid1c* and adjacent transcripts from mouse strain C57BL/6 that were constructed into pBACe3.6 vector (Genebank accession #: U80929) (Figure 2.1). *E.coli* strains containing each BAC clone were cultured in Luria-Bertani (LB) medium with chloramphenicol (25 μg/ml). BAC DNA was prepared using the Genopure BAC maxi kit (Roche) following manufacturer's instructions and stored in Tris buffer (pH8.0) at 4°C.

There are two forms of mammalian sex chromosomes, X and Y, and their combinations determine gender in mammals. In therian mammals (marsupial and placental), the Y chromosome contains a testis-determinant gene, *SRY*, [7] and individuals carrying the Y chromosome become males. Normal males have one Y and one single X chromosome, whereas females are characterized by two Xs and no Y. Due to the sex-determining function of the Y, some rare cases such as XXY individuals are males while XO and XXX individuals are females.

2.1.2 Pulsed field gel electrophoresis (PFGE)

Approximately 0.5 μg of circular or restriction digested BAC DNA was electrophoresed through 1% agarose in 0.5× TBE for 24 hours (hr) at 200 volts (V) and 4°C with an initial switching time of 1 second (sec), a final switching time of 25 sec and a switching ratio of one. Fragments were stained by ethidium bromide for visualization.

2.1.3 Polymerase chain reaction (PCR)

A ~2 kb neomycin cassette was amplified from plasmid PL451 [304] using Platinum Pfx DNA polymerase (Invitrogen) and primer pair BACneo5/BACneo3 (Table 2.1). PCR cycles were as follows: 94°C for 1 minute (min), 35 cycles (94°C for 30 sec, 58°C for 30 sec, 68°C
Figure 2.1 Map of *Jarid1c* BAC clones RP23-330G24 and RP23-391D18. Each BAC clone contains several mouse X-linked transcripts (open arrows) in the genomic insert (thick line) and selective markers (black arrows) on the backbone (thin line). Linearization sites for each BAC are shown as vertical lines on the backbone. The location and orientation of neomycin cassette inserted by recombineering are also indicated in the maps. Maps are not drawn to scale.
Table 2.1 Primers used in construction and characterization of BAC transgenes.

<table>
<thead>
<tr>
<th>Primer pair</th>
<th>Sequence (5’ to 3’)</th>
<th>Annealing temperature (°C)</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BACneo5</td>
<td>TAGCAGCAGCCCATAGTGACGAGTTTACGTCCAGCCAGCAAGC</td>
<td>58</td>
<td>1836</td>
</tr>
<tr>
<td>BACneo3</td>
<td>GTCAAGGCAACCGGTGTATGAAGAATGTACCTGAAGTGATGAAGGTTCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BACad1</td>
<td>GCATAGAAATTGCATCAACGCATATAGCGCTAGCAGACGCGCCATAGTGAC</td>
<td>55</td>
<td>1896</td>
</tr>
<tr>
<td>BACad2</td>
<td>GTAGTTTATCAGTTAAAATTGCTAACGCAGTCAGGCACCGGTGTATGAAGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BACad3</td>
<td>GCAGGCTGGACAGTGCTCCGAGAAGGCGGTGCGCATAGAAATTGCGATCAAC</td>
<td>52</td>
<td>1956</td>
</tr>
<tr>
<td>BACad4</td>
<td>GACAGCTTTATCATCGAATGCTTAATGCGGTAGTTATCGACAGTTAAAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAC1</td>
<td>TTTGCGCGCAAGCTTCTTTATG</td>
<td>58</td>
<td>2471</td>
</tr>
<tr>
<td>BAC2</td>
<td>AACGGAATACCCTCGGTGTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BACcloxp1L</td>
<td>ATTTGCATCGCATTTGCTGTG</td>
<td>58</td>
<td>558</td>
</tr>
<tr>
<td>BACcloxp1R</td>
<td>AACGGAATACCTCGGTGTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BACcloxp2L</td>
<td>CCGACATAGATAAAACTTCGT</td>
<td>58</td>
<td>187</td>
</tr>
<tr>
<td>BACcloxp2R</td>
<td>GCAAAATTATTAGAGCAATATAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SgrAL</td>
<td>CTTGGTTGGCGATAATCGTT</td>
<td>58</td>
<td>1529</td>
</tr>
<tr>
<td>SgrAR</td>
<td>GCGCTGGGAATAGGTGAAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BACpbL</td>
<td>GGCCAAGGCGATTTTATTCTCA</td>
<td>58</td>
<td>1088</td>
</tr>
<tr>
<td>BACpbR</td>
<td>GGCGAAGAAGTTGTCCTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BACpsL</td>
<td>CTTAAACGCCTGGTGCTA</td>
<td>55</td>
<td>338</td>
</tr>
<tr>
<td>BACpsR</td>
<td>TTCAACCGCTAGCTCCTT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2.1 Primers used in construction and characterization of BAC transgenes (continued).

<table>
<thead>
<tr>
<th>Primer pair</th>
<th>Sequence (5’ to 3’)</th>
<th>Annealing temperature (°C)</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smcxpb2L</td>
<td>TACCAGGTTTTTGGCTCACC</td>
<td>58</td>
<td>545</td>
</tr>
<tr>
<td>Smcxpb2R</td>
<td>TGGGCTTTGGTAGTTTCACC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL109</td>
<td>TAGATCCACAGGACGGGTGT</td>
<td>55</td>
<td>N/A²</td>
</tr>
<tr>
<td>NL110</td>
<td>CGATCCTCCCGAAATGACTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL117</td>
<td>GTCATATGGCGTGCTGCTGA</td>
<td>55</td>
<td>N/A²</td>
</tr>
<tr>
<td>NL118</td>
<td>TCAGACCGACGATACGAGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL176</td>
<td>TATGAAATTCGCTCGACGTAGTGTTGGCAAGCGGATT</td>
<td>58</td>
<td>5814</td>
</tr>
<tr>
<td>NL177</td>
<td>ATAGAATTCGCTCGACCAGGACGAAAGACTCCAGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M13 forward</td>
<td>GTAAAACGACGGCCAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M13 reward</td>
<td>CAGGAAACAGCTCTGAC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1the product is 2471 bp for targeted clones and 714 bp for non-targeted vectors.
2the product size is dependent on transgene integration sites, usually several kb.
for 2 min) and an additional extension at 68 °C for 5 min. PCR product was purified with QIAquick PCR purification kit (QIAGEN). Sequential rounds of PCR were performed using overlapping primer pairs BACad1/BACad2 and BACad3/BACad4 (Table 2.1) to add 80 bp of BAC homology to each end. Thermal cycle programs were as follows: 94°C for 1 minute (min), 10 cycles (94°C for 30 sec, 52/55°C for 30 sec, 68°C for 2 min), 25 cycles (94°C for 30 sec, 68°C for 2.5 min) plus an extension at 68 °C for 5 min. PCR product was purified and concentration was determined.

2.1.4 Recombineering

The PCR product was inserted into BAC backbone by recombineering using the recombination efficient E.coli strain, DY380 [305]. Constitutive expression of the phage recombination genes beta, exo and gam is lethal. Thus, E.coli strains for recombineering are usually grown at low temperatures, at which these genes are inhibited by temperature sensitive repressors. By a heat shock, these genes are induced to express proteins and are capable of promoting DNA recombination. DY380 was first grown at 32°C to an OD_{600} = 0.5-0.6 in LB medium and then shifted to 42°C for 15 min, followed by chilling on wet ice for 10 min. Electrocompetent cells were prepared by washing the cells three times with ice-cold water. The final cell pellet was resuspended at a 1:1 volume with ice-cold water. 100 ng of BAC DNA was electroporated into 50 μl of ice-cold competent cells in a cooled cuvette (1 mm) using Bio-Rad gene pulser set at 1.8 kV, 25 μF and 200 Ω. After electroporation, 1 ml of LB medium was added to the cuvette, and the culture was incubated at 32°C for 1 hr with shaking. 100-200 μl of cells were plated on LB agar medium containing chloramphenicol (25 μg/ml) and grown overnight. Electrocompetent DY380 harboring the BAC of interest was then prepared by the same procedures and 300 nanograms (ng) of PCR products were electroporated into heat shocked cells. Kanamycin (25 μg/ml) and chloramphenicol (25 μg/ml) dual-resistant clones were screened by PCR using the primer pair that flanks the targeting sequence, BAC1/BAC2 (Table 2.1).

2.2 CELL LINES AND CULTURE CONDITIONS

2.2.1 ES cell culture

The female mouse ES cell line EL1/c16 was kindly provided by En Li (Novartis Institute)
and carries X chromosomes from two different mouse strains, 129 and *M.m.castaneus* (CAST) [76]. Undifferentiated ES cells were maintained on 0.2% gelatin-coated plates in Dulbecco’s minimal Eagle medium (DMEM, reagents were from Invitrogen unless specified) containing 15% fetal calf serum (FCS, ES cell grade, Hyclone), 1 mM sodium pyruvate, 100 μM non-essential amino acids, 100 μM 2-mercaptoethanol, 2 mM L-Glutamine, 100 U/ml penicillin, 0.1 mg/ml streptomycin, and 1000 U/ml leukemia inhibitory factor (LIF, Chemicon). All cells were grown at 37°C in a 10% CO₂ environment.

2.2.2 *In vitro* differentiation

To induce differentiation, ES cells were grown at 1×10⁴/ml in suspension on bacterial petri dishes for 3 days in the absence of LIF. The resulting embryoid bodies (EBs) were then plated on glass slides to allow outgrowth. Differentiated ES cells were cultured in DMEM supplemented with 10% FCS (cell culture grade, Hyclone), 100 μM 2-mercaptoethanol, 2 mM L-Glutamine, 100 U/ml penicillin and 0.1 mg/ml streptomycin.

2.2.3 Enriched differentiated ES cells

Ten-day differentiated cells were trypsinized, resuspended, and plated for 10 min. Subsequently, media was changed to remove non-adherent cells and the remainder cells were grown for an additional five to ten days prior to harvesting.

2.3 TRANSFECTION

BAC DNA was transfected into undifferentiated ES cells by electroporation [306]. ES cells were first passaged 1:5 on gelatin-treated plates and cultured for two days. On the day prior to electroporation, cells were split at 1:2 and cultured overnight. On the next day, cells were fed with fresh ES cell medium for 4 hours prior to transfection. The BAC DNA was linearized within the BAC backbone at the *Sgr*AI or *I-Sce*I sites between loxP and lox511 (Figure 2.1) and 12 μg of DNA for 10⁷ ES cells was electroporated in antibiotic-free medium at 240 V and 500 μF (GenePulser Xcell, Bio-Rad). Drug selection (250 μg/ml G418) was initiated at 24 hr and resistant colonies were picked after 8 days. After one additional passage, drug selection was removed for all subsequent experiments.
2.4 SOUTHERN ANALYSIS

BAC transgenes were examined by Southern analysis using a 1 kb fragment in the pBACe3.6 vector backbone (9,917-11,004 bp). The probe was PCR amplified with the primer pair BACpbL/BACpbR (Table 2.1) from BAC DNA and random labeled with [α-32P] dCTP (3000 Ci/mmol; Perkin Elmer) under the condition of 20 μg/ml probe DNA, 100 μg/ml random hexanucleotides, 200 mM HEPES (pH6.4), 50 mM Tris (pH8.8), 5 mM MgCl, 10 mM b-Mercaptoethanol, 0.02 mM dT/G/ATP, 5 uCi [α-32P] dCTP and 5 units (U) of Klenow polymerase at 37°C for 1 hour. Unincorporated nucleotides were removed with the QIAquick column (QIAGEN). Southern analysis was essentially as described [307]. Briefly, genomic DNA was isolated using Wizard Genomic DNA purification kit (Promega), restriction digested with KpnI, NheI or SacI overnight at 37°C, electrophoretically fractionated on 1% agarose gel and stratalinked in gel by crosslinking (900uJ/cm²). The gel was treated with denaturizer (1.5 M NaCl, 0.5M NaOH) for 50 min and then neutralizer (1M Tris pH7.4, 1.5 M NaCl) for 20 min twice with shaking in both steps. DNA was transferred to nitrocellulose membranes and immobilized by crosslinking (1200uJ/cm²). Hybridization was performed at 65°C overnight in 1 M NaCl, 1% SDS, 10% Dextran Sulfate, and 70 μg/ml sheared salmon sperm DNA. The blots were briefly rinsed in 2×SSC (standard saline citrate)/0.1% SDS at room temperature, followed by stringency washes (2×20 min) with each solution A (1% SDS, 1 M NaCl, 50 mM Tris-Cl pH8.6, 2 mM EDTA pH8), B (0.5% SDS, 0.5 M NaCl, 50 mM Tris-Cl pH8.6) and C (0.1% SDS, 0.1×SSC) at 65°C. Membranes were exposed to phosphor screens and scanned with a Molecular Imager PharosFX plus system (Bio-Rad).

2.5 FLUORESCENCE IN SITU HYBRIDIZATION (FISH)

2.5.1 FISH probes

Probes for DNA FISH included the mouse X chromosome pericentromeric repeat DXWas70 [308, 309] and BAC DNA. Probes for RNA FISH were an Xist subclone that includes a 7.6 kb genomic XbaI fragment covering most of exon 1, Jarid1c genomic sequence from a 19kb EcoRI subclone that encompasses exons 5-12, and an Iqsec2 genomic probe from three overlapping SpeI fragments that include exons 3-8, a Tspyl2 probe that encompasses all genomic sequence for this gene and was amplified from BAC RP23-391D18 using the primer pair NL176/NL177, and mouse Ube1 BAC RP24-99E9. For both RNA and DNA FISH,
double-stranded plasmid and BAC DNA were directly labeled with Alexa dyes 488, 546, 594, or 647 by nick translation using ARES DNA labeling kits (Invitrogen) following the manufacturer’s specifications. All probes were denatured for 10 min at 75°C before applying on slides. To block repetitive sequences, genomic probes were preannealed for 45 min at 37°C in the presence of 1 mg/ml mouse Cot-1 DNA (Invitrogen) after denaturation.

### 2.5.2 DNA FISH

Metaphase spreads for DNA FISH analysis were prepared by standard techniques. Cells were blocked at the mitotic stage by adding 50 ng/ml Colcemid and incubating 45 min at 37°C. All cells were harvested, washed with 1×PBS, fixed in three rounds of 3:1 methanol:acetic acid and then dropped on glass slides. Slides were first dehydrated in an ice-cold ethanol series (70, 80, and 100% for 3 min), denatured at 72°C for 2 min in 70% formamide, 2×SSC, pH 7.0 and immediately dehydrated in the alcohol series again. 10 μl of hybridization mixture was applied to each slide under a sealed coverslip and placed in a humidified chamber overnight at 37°C. FISH was essentially performed as described [310]. Briefly, 60 ng of denatured probe DNA in 10 μl hybridization buffer (50% formamide, 10% dextran sulfate, 500 μg/ml salmon sperm, 2×SSC) was applied to denatured slides and hybridized overnight at 37°C in a humid chamber. Post-hybridization washes were performed in 50% formamide, 2×SSC, pH 7.0 (2 × 8 min at 42°C) followed by 2×SSC (8 min at 37°C). Slides were counterstained and mounted with Vectashield antifade (Vector Laboratories) containing 200 ng/ml 4, 6-diamido-2-phenylindole dihydrochloride (DAPI). For each slide a minimum of 20 metaphase spreads were examined.

### 2.5.3 RNA FISH

RNA FISH was performed essentially as described [93]. Undifferentiated ES cells were harvested and cytospun onto L-polylysine (Sigma) treated glass slides. For differentiated cells, EBs were plated and grown directly on glass slides on day 3. All slides were briefly washed in PBS and cytoskeletal (CSK) buffer, permeabilized for 5 min with 0.5% Triton X-100 in CSK, fixed in situ with 4% paraformaldehyde in PBS (10 min, 25°C), and stored in 70% ethanol. To ensure hybridization was specific to RNA, nuclei were not denatured prior to hybridization. Hybridization and wash conditions were as described above except that the hybridization
buffer also included unlabeled competitor tRNA (2 mg/ml, Sigma) and 20 mM vanadyl ribonucleoside complex (New England Biolabs).

To determine whether these hybridization conditions specifically detected RNA but not DNA, RNA FISH was performed using slides that were pretreated with RNaseA (Sigma) or to detect X-inactivated genes, such as mouse *Iqsec2* and *Ube1*.

### 2.5.4 Sequential RNA and DNA FISH

Sequential RNA and DNA FISH experiments were performed by slightly modifying established methods [93]. Briefly, after RNA FISH hybridization and washes were performed as described above, hybridization signals were fixed in 4% paraformaldehyde in PBS (15 min, 25°C). Subsequently, slides were denatured in 70% formamide, 2× SSC for 5 min at 75°C. Samples were then treated with 100 mg/ml RNaseA in 2×SSC at 37°C for 1 hr, rinsed four times in 2×SSC and DNA FISH was performed as described above. Each RNA or sequential RNA and DNA FISH experiment was performed in duplicate and 100–200 interphase nuclei were analyzed. Results were evaluated using the *Chi*-square statistic.

### 2.5.5 Microscopy and image processing

Slides were analyzed using 60X or 100X lenses on a Nikon ECLIPSE E1000 epi-fluorescence microscope equipped with a Hamamatsu cooled charge-coupled device (CCD) camera and ImagePro3 software (MediaCybernetics) or a Nikon TE2000-U microscope outfitted with a Roper Scientific CCD camera and NIS elements software. Images were taken at a 406 nm filter to visualize DAPI-stained nuclei and 488, 546, 594 and 647 nm filters to visualize fluorescently labeled probes. Each fluorophore was captured individually without image shift, pseudocolored and merged in Photoshop (Adobe). To ensure appropriate scoring, the focus was adjusted for each image to confirm that any out-of-plane signals were not omitted. Nonspecific background fluorescence was measured to determine the threshold of FISH signals.

### 2.6 INVERSE PCR

Transgene integration sites were identified using inverse PCR by modifying established protocols [311]. 10 μg of genomic DNA from each line was digested to completion with *NheI*,
HindIII or appropriate restriction enzymes and purified with phenol:chloroform. Digested DNA was self-ligated for 2 hr at room temperature in a volume of 200 μl using 3 U of T4 DNA ligase (Invitrogen). Transgene ends and adjacent integration site sequences were amplified from the circular ligation products using complementary strand primers within the BAC vector, NL109/NL110, NL117/NL118 or appropriate primer pairs (Table 2.1). PCR amplification conditions were optimized for large insert sizes using LA polymerase (TaKaRa) and 100 ng of ligation products under the following conditions: 20 pmol of each primer, 200 μmol of each dNTP, 2.5 U of LA DNA polymerase (TaKaRa), and 5 μl of 10× LA DNA polymerase buffer in a 50 μl reaction volume. PCR cycles were as follows: 94°C for 30 sec, 35 cycles (94°C for 30 sec, 55°C for 30 sec, 72°C for 5 min); plus additional prolonged extension at72°C for 10 min. PCR products were directly sequenced or cloned into vector pCR2.1 using TA cloning kit (Invitrogen). After transformation into electrocompetent GeneHog cells (Invitrogen), sequencing reactions were performed with the M13 or appropriate primers (Table 2.1) by an ABI 3100 automatic sequencer (Applied Biosystems) in Core Facility of Penn State College of Medicine. Sequence homology searches were conducted on UCSC Genome Browser to identify the integration sites on the mouse X chromosome.

2.7 INACTIVE X EXPRESSION ANALYSIS

2.7.1 RNA extraction

Total RNA was isolated with Trizol (Invitrogen) and treated with RNase-free DNaseI for 15 min at 25°C. First-strand cDNA was synthesized from 5 μg of total cellular RNA by random priming at 42°C for 2 hrs. Samples were reverse transcribed and amplified with Moloney Murine Leukemia Virus (MMLV) reverse transcriptase and Platinum Taq polymerase (Invitrogen). Controls lacking reverse transcriptase were processed in parallel to verify the absence of genomic DNA contamination.

2.7.2 Quantitative-SNaPshot (Q-SNaPshot) assay

XCI status was examined using a quantitative allele-specific primer extension method, Q-SNaPshot, that incorporates fluorescent dideoxynucleotides at the polymorphic site (Figure 2.2) [25]. The Q-SNaPshot assay was performed using a PRISM SNaPshot Multiplex kit
Figure 2.2 Scheme of the Q-SNaPshot assay. A primer first binds to the sequences that abut the SNP. Fluorescently labeled ddNTP complementary to the nucleotide at the polymorphic site are attached to the primer in a primer extension reaction. Fluorescence intensity is measured using a genetic analyzer and the peak heights correspond to the relative allelic expression level after normalizing the RNA level to the DNA level.
(Applied Biosystem) according to the manufacturer’s recommendations except that two modifications were required to establish quantitative conditions [25]. To remove dNTPs and primers, PCR reactions were treated with 2 U of each shrimp alkaline phosphatase (SAP, Promega) and Exonuclease I (Fermentas) at 37°C for 1 hr and heat inactivated at 72°C for 15 min. SNaPshot reactions were prepared using 1.25 μM specific primer with the concentration increased to 1.25 μM to ensure that oligonucleotide primers were not limiting, as well as 1 μl of PCR product, 2.5 μl of SNaPshot Multiplex Ready Reaction Mix, 0.5× PCR buffer, 0.75mM MgCl₂ and deionized water to a final reaction volume of 10 μl. Reactions were carried out for only a single cycle in the PTC-100 (MJ Research) thermocycler at 96°C for 2 min, 56°C for 5 sec and 60°C for 5 min. Following cycling, samples were treated with 2 U of SAP at 37°C for 1 hr and heat inactivated. In an ABI optical plate, 0.5 μl of the reaction, 9 μl HI-DI formamide and 0.5 μl of Genescan 120 LIZ size standard were combined, denatured at 95°C for 5 min, and immediately placed on ice. Samples were then analyzed on an ABI 3100 Genetic Analyzer and the peak heights were determined using GeneMapper 4.0 software. Genomic DNA was included as a control of presentation of SNPs on two different alleles and as a standard to normalize the expression level of the genes on both Xa and Xi.

2.7.3 Inactive X expression on normal X chromosomes

To assess the normal Xi expression pattern of transcripts from lines without transgenes, expression was tested in the B119 cell line. This cell line is an early passage primary fibroblast cell line derived from a (T16H × CAST) F1 female newborn mouse [174]. Because T16H carries a balanced X:16 translocation, the CAST X chromosome is inactive in all cells [174] and allele-specific expression of this X was assessed by the Q-SNaPshot assay. SNPs between different mouse strains were identified from the Perlegen Mouse Genome Browser (http://mouse.perlegen.com/mouse/browser.html) and NCBI SNP database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp). Primer sequences for all allele specific assays are indicated in Table 2.2.

2.7.4 Inactive X expression of transgenes

A different strategy was used to determine Xi expression for the Jarid1c and Tspyl2 transcripts on the BAC transgenes. Enriched populations of differentiated cells were isolated
Table 2.2  SNPs assayed for inactive X expression.

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>Primer sequence (5’ to 3’)</th>
<th>Annealing temperature (°C)</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AK139935</td>
<td>rs29852013</td>
<td>AAGAAAAGCAGGAGACAGCG* GGCCTCCGTCATTACCTTCT</td>
<td>54</td>
<td>231</td>
</tr>
<tr>
<td>Atp1b4</td>
<td>rs29842888</td>
<td>GCTGAAGGGCAAAGGCAT* GTCCCAACAATATCACTTCATACACAC</td>
<td>56</td>
<td>263</td>
</tr>
<tr>
<td>Birc4</td>
<td>rs29962995</td>
<td>GGCATTATGTGAAGCCCAAAAA GCATTAGGAACCTCAAAACACACAAAN</td>
<td>54</td>
<td>211</td>
</tr>
<tr>
<td>Cul4b</td>
<td>rs29854158</td>
<td>TCCTAAGGGCAAGATATTTGAAGA* TCAACCCTTTCTTTCATCTGG</td>
<td>56</td>
<td>111</td>
</tr>
<tr>
<td>EG547215</td>
<td>rs29075800</td>
<td>GGATTCTTTTTTCACACCGAGACTAG* CCATGAAACGTTCAAGAAGG</td>
<td>56</td>
<td>489</td>
</tr>
<tr>
<td>Gpr173</td>
<td>rs29295779</td>
<td>CCCCTTCACAATCCTTTAGC CAGCTTATACCCACATCACCTCTA*</td>
<td>58</td>
<td>425</td>
</tr>
<tr>
<td>Gprasp1</td>
<td>rs29265645</td>
<td>CTGGGGATGAAGACAGGTTT* CCCAGGAGACAGACTCAAAAG</td>
<td>56</td>
<td>155</td>
</tr>
<tr>
<td>Gria3</td>
<td>rs29941705</td>
<td>GGAGCTTTCAAGAAGGCCTG* TTGCAACACGTTCTGTCTC</td>
<td>56</td>
<td>239</td>
</tr>
<tr>
<td>Iqsec2</td>
<td>rs33850012</td>
<td>CAGCACTGTGGTCTGAGGA AAGAGGTGTAAGAGCCTCTCAG*</td>
<td>56</td>
<td>330</td>
</tr>
</tbody>
</table>
Table 2.2  SNPs assayed for inactive X expression (continued).

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>Primer sequence (5’ to 3’)</th>
<th>Annealing temperature (°C)</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jarid1c</td>
<td>rs29296190</td>
<td>GCCTCACAGAGAGAGCCATC GGTCACAAGTGCCCCCATCTGTAAGGGA*</td>
<td>58</td>
<td>489</td>
</tr>
<tr>
<td>Lamp2</td>
<td>rs29850796</td>
<td>CGGAATAGTGCCAAAATAATGATACA* TGTTGTAGGAAAAGCAGATGTC</td>
<td>56</td>
<td>112</td>
</tr>
<tr>
<td>Mid1</td>
<td>rs31818284</td>
<td>CGCCACACCCATGTTCCTC GGTCACAGTACATGTTCCACCTTCTC*</td>
<td>55</td>
<td>247</td>
</tr>
<tr>
<td>Tspyl2</td>
<td>rs29297026</td>
<td>TTTGCCCTCTCTTCAGCTA CTCGTACAGGAAAGTCTGGCTGCA*</td>
<td>57</td>
<td>185</td>
</tr>
<tr>
<td>Zmat1</td>
<td>rs29267919</td>
<td>TATCACTGACAGATCTATCTGATTGAGA* GTTCACTGCTCTGCATGTT</td>
<td>55</td>
<td>156</td>
</tr>
</tbody>
</table>

* Indicates the primer that abuts the SNP and was used for the Q-SNAPshot reaction.
from both non-transgenic and transgene-containing lines as described above. SNPs were identified between the C57Bl6/J, 129, and CAST alleles. Expression levels of individual alleles were then examined by the Q-SNaPshot assay and normalized to the DNA level. Allelic expression values could be used to determine Xi expression by normalizing for the percentage of cells that were clearly differentiated and the frequency that either X is inactivated. The percentage of cells in the “enriched differentiated” populations that are differentiated and retain two Xs, as well as the percentage of differentiated cells in each line that inactivate the X carrying the transgene were determined by sequential RNA and RNA FISH. Assays were repeated three times and evaluated by two-tailed t-tests. Expected escape values were calculated that correspond to varying levels (from 0% to 100%) of Xi expression relative to levels on the Xa. Predicted values were computed separately for each line.
CHAPTER III
ESTABLISHMENT OF MOUSE FEMALE EMBRYONIC STEM
CELL LINES CARRYING AN X-LINKED Jarid1c BAC TRANSGENE

3.1 INTRODUCTION

Transgenes located on mammalian X chromosomes are important tools to study gene regulation on inactivated X chromosomes and understand establishment and maintenance of XCI. Previous X-linked transgene studies have provided valuable insight into Xi gene regulation, such as the ability of diverse sequences to be inactivated, the influence of the mouse \( Hprt \) locus on XCI status, and the potential role of specific sequence elements on XCI (see section 1.4). However, as discussed previously, conclusions from the aforementioned studies are limited and preliminary at best. Moreover, studies of transgenes that escape XCI are fairly limited. None of these studies has given insight into what sequences and functional elements are needed for escape, and the mechanisms involving how Xi genes are regulated still remain largely mysterious.

Identification of sequences important for escape by the transgene approach would benefit from the use of genes that originate on the X with known XCI responses, especially ones that escape XCI. Previously established X-linked transgenes were derived from species other than mammals, or mammalian autosomes. As most of the transgenes are subject to XCI (see section 1.4), escape genes are likely associated with specific regulatory elements or sequence composition that are absent in non-mammalian genomes, most mammalian autosomal regions and X-inactivated genes. The mouse escape gene \( Jarid1c \) is ideal for studying Xi expression because of the well-established XCI status and well-characterized promoter and insulator elements (see section 1.5).

To locate a region that carries regulatory information necessary for escape, we chose to generate X-linked transgenes using large genomic fragments including genes and adjacent sequences. As large DNA fragments in the form of bacterial, phage or yeast artificial chromosomes (BACs, PACs or YACs) potentially contain all regulatory elements of a gene, they are of great help for defining transcription units, modeling inherited disease states, acquiring development- or tissue-specific expression and dissecting candidate loci for regulatory elements [312, 313]. Importantly, studies of the human X suggest that sequences
several hundred kilobases away from a gene can predict XCI status [227]. Thus, large constructs likely include potential distant regulatory sequences that are important for escape. Studies on Xi regulation of large constructs are inadequate to date. Only one study reported that two BACs containing the mouse autosomal gene *Myo7a* were X-inactivated at the *Hprt* locus as the complementation of the *Myo7a* null mutant phenotype by the X-linked BAC transgenes produced a fine mosaic of two cell types in hemizygous transgenic females [266]. However, Xi expression status was not directly tested and inactivation of a large autosomal fragment does not necessarily represent the regulation of X-linked genes, especially those that escape XCI. Another advantage of using large constructs is that the chromosomal interval between the two X-inactivated genes flanking *Jarid1c* can be included in single constructs (Figure 3.1), allowing investigation of the entire *Jarid1c* escape region. Furthermore, *Jarid1c* is the only mouse escape gene that can be included in BACs with genes of opposite XCI expression status. Transgenes derived from these BACs allow direct comparison between X-inactivated and escape genes at the same integration site.

As an escape domain including all regulatory sequences necessary for escape is predicted to retain Xi expression independent of chromosomal location, an effective approach to delimit such a domain is to analyze the XCI status of the same transgene at multiple X loci. This approach requires insertion of the candidate escape region into different, unbiased locations on the mouse X chromosome. Transgene DNA can be introduced onto the X by site-specific targeting into previously investigated loci (e.g. [246]). However, these loci were studied because of specific regulatory mechanisms. They are not representative for the multitude of the X. Further, X-linked loci that are currently available for targeting are limited in quantity and the construction of a targeting vector for each site is rather time-consuming. More importantly, results will be more convincing if transgene locations are random and characterized by different XCI status and repeat sequence content. Therefore, we chose to introduce BAC DNA into the mouse genome by random integration. Approaches to integrate exogenous DNA into the mammalian genome include viral delivery systems and transposon elements such as the sleeping beauty and PiggyBac [314, 315]. These systems have been applied to mouse transgenesis and allow stable, long-term transgene expression. However, extensive studies on the integration site in these transportation systems revealed non-random transportation systems revealed non-random selectivity, with the preference for genes and
**Figure 3.1 Mouse *Jarid1c* genomic region.** A 300-kb region on the mouse X chromosome at 148.52-148.82 Mb (UCSC Genome Browser, July 2007 assembly) is shown. *Jarid1c* escapes XCI (blue) and is surrounded by X-inactivated genes (yellow). Two BAC clones containing the entire *Jarid1c* region are selected for the transgene studies. The genomic sequences included in each clone are indicated by green lines below the map of this region. Modified from [316].
Mouse Chr X (q) | Xist/Xic

CEN A1.1 A2 A4 A5 BC1 BC3 D E1 E3 F1 F2 F3 F4 F5

---

20 kb

Iqsec2 → Jerid1c

RP23-330G24 (186 kb)

AK013346 BB54222 Tspyl2 Gpr173

RP23-391D18 (175 kb)

---

Genomic sequence
Yellow: X-inactivated gene
Blue: escape gene
BAC genomic DNA
CEN: centromere
upstream regulatory elements as well as certain repetitive sequences [317, 318]. Therefore, we chose to integrate DNA molecules that are not constructed with any of these systems.

In the process of generating stable transgenic cell lines, one of the most important experimental considerations is to obtain intact, single-copy transgenes. Features frequently observed in random integration include truncated transgene DNA and copy number variation [313], which can influence transgene activity and complicate analysis of transgene Xi expression. Transgenes with different copy numbers may differ in their expression levels as transcription may accumulate from multiple copies, or be reduced and even shut down due to copy number-dependent gene silencing mechanisms [319]. To directly relate transgene Xi activity to XCI status, we sought to examine cell lines with a single, intact BAC integrant.

In this study, we sought to establish transgenic ES cell lines in order to determine whether the Jarid1c escape domain is autonomous or dependent on chromosome location. We first introduced Jarid1c BAC DNA into the genome of undifferentiated mouse female ES cells by random integration and screened transgenic cell lines to identify X-linked integrants. We then characterized transgenes to evaluate features that may influence their expression, such as transgene integrity, copy number, and insertion site location. Intact, single-copy transgenes were chosen for future analysis of Xi expression. Transgene activity prior to XCI, XCI landscape of each integration site and the probability to observe the transgene on the Xi in each cell line were also determined. These transgenic cell lines will not only help understand the regulation of escape from XCI in mouse, but also provide a transgene system to analyze Xi expression of any genes at well-characterized locations and assess the function of candidate regulatory sequences.
3.2 RESULTS

3.2.1 Generation of BAC transgene constructs

BACs containing mouse Jarid1c were selected to generate transgenes. Two BAC clones were chosen to incorporate the entire Jarid1c genomic region and adjacent X-inactivated transcripts at either side of Jarid1c (Figure 3.1). BAC clone RP23-330G24 includes Jarid1c and the upstream gene Iqsec2. Clone RP23-391D18 contains Jarid1c and the downstream genes Tspyl2 and Gpr173. Tsuchiya et al. reported two additional X-inactivated transcripts at this locus, expressed sequence tags (ESTs) AK013346 and BB854232 [196]. They have been recently assigned to hypothetical genes 2810454L23Rik and 2900056M20Rik, respectively.

Jarid1c BACs needed to be modified prior to the introduction into ES cells as common BAC vectors contain only prokaryotic resistance genes. In order to select stable transgenic cell lines after transfection, it was necessary to insert a mammalian selective marker into the BAC DNA. Compared with smaller constructs such as plasmids, BACs are relatively difficult for molecular cloning due to the large size and lack of unique restriction sites. A method known as recombineering or recombinogenic engineering [320] has been successfully used to modify DNA constructs regardless of their sizes and restriction sites (e.g. [321, 322]). Recombineering is a process of homologous recombination directed by λ phage genes (Figure 3.2A). The λ phage genome encodes three genes that are sufficient for a DNA crossover event, including an antagonist of endogenous nucleases (gam), an endonuclease that removes the 5’ overhang of DNA fragments (exo) and a recombinase that stabilizes single-stranded DNA and facilitates annealing of complementary sequences (bet) [320]. Sequences of interest flanked by regions homologous to the targeting site can be inserted into this site. This highly efficient system allows recombination to occur even with homology limited to 40-50 bp. To facilitate the recombineering process, λ phage DNA has been integrated into E.coli genomes and DNA recombination can be achieved in the engineered bacterial strains [305].

Using the recombineering approach, a neomycin (neo) cassette with 80 bp of homology at each side was inserted into the BAC backbone between the loxP site and the SacB coding sequence, with the transcription orientation towards loxP (Figure 2.1). This cassette contains the coding sequence of a neomycin resistance gene that is regulated by a eukaryotic Pgk1 and a prokaryotic EM7 promoters, which render resistance to drug G418 in mammalian cells and to kanamycin in E.coli, respectively. As the BAC vector also carries a chloramphenicol
Figure 3.2 Targeting of a neomycin/kanamycin cassette into the BAC backbone by recombineering. (A) Scheme of Recombineering. A modifying DNA fragment contains the gene of interest flanked by sequences homologous to regions in the target construct (region of homology 1 and 2, RH1/RH2), which can be 40-bp to several kb in size. In the target construct, RH1 and RH2 flank the sequence that will be replaced by recombination. RH1 and RH2 pair between the modifying fragment and target construct during the phage-mediated homologous recombination, allowing the gene targeted into a specific site in the construct. A neomycin/kanamycin (neo/kan) cassette is shown here to be inserted into the BAC backbone. Location of the primer pair to amplify sequences flanking the targeting site (BAC1/BAC2, Table 2.1) is marked by arrows and the size of PCR products from wild-type (wt) and recombinant BACs is also indicated. (B) PCR results of wild-type (wt) and recombinant BAC clones.
resistance gene, following recombineering kanamycin and chloramphenicol dual resistant clones were screened for proper insertion by amplification of sequences flanking the site of recombination (Figure 3.2A). Among eight clones examined for the two BACs, only one yielded a PCR product with a size similar to that from non-targeted BACs (Figure 3.2B). The remaining seven clones generated a ~2 kb longer product, as expected for clones that had properly undergone recombineering. As the false recombinant clone was resistant to both antibiotics, it is possible that the \textit{neo} cassette randomly integrated into the bacterial genome or BAC DNA. Therefore, this recombineering experiment generated positive recombinant clones at a frequency of 87.5%.

For each BAC, a modified clone was analyzed by pulsed field gel electrophoresis (PFGE) to verify proper recombination and BAC integrity. The \textit{neo} cassette has one \textit{Eag}I site, allowing wild-type and recombinant BAC DNA to be differentiated by their restriction patterns after \textit{Eag}I digestion. The original vector region has two \textit{Eag}I sites separated by a distance of 8.7 kb, whereas the added \textit{Eag}I site is 8.0 and 2.4 kb distant to these two sites, respectively (Figure 3.3A). By electrophoresis, \textit{Eag}I digestion generated a single ~9 kb band for wild-type BACs but two smaller bands for recombinant BACs with the sizes consistent with the expectation, implying that the \textit{neo} cassette was inserted correctly (Figure 3.3B). Moreover, PFGE analysis did not identify any change in the restriction pattern of BAC genomic sequences, suggesting that recombineering did not affect the genomic insert. Thus, restriction digest and PFGE analysis confirm that each of the recombinant BACs, named as RP23-330G24-neo and RP23-391D18-neo, contains a neomycin cassette on the backbone and the genomic inserts remain intact.

### 3.2.2 Subcloning of stable ES cell lines

Mouse female ES cells were chosen to insert BAC DNA and assess transgene Xi expression. ES cell line EL1/c16 and subclones have been used in other XCI studies (e.g. [76]) and were extended to my thesis work. This cell line was produced from a cross between mouse strains 129 and \textit{M.m.castaneus} (CAST) and contains one X chromosome from each of the two strains [323].

As an essential step, ES cell subclones that stably maintain two X chromosomes were isolated. The cell line EL1/c16 is karyotypically unstable and stochastically loses one X
Figure 3.3 Pulsed field gel electrophoresis (PFGE) of wild-type and recombinant BAC DNA. (A) Restriction pattern of BAC clones after *Eag*I digestion. Vertical lines indicate individual *Eag*I sites in the BAC vector region (thin line) and genomic insert (thick line). The length of each fragment (kb) is marked, but is not drawn to scale. (B) PFGE analysis of BAC clones.
during cell proliferation, giving rise to a mosaic population of 40,XX/39,XO cells [323]. In order to generate stable, uniform 40,XX ES cells, Dr. Laura Carrel (Penn State College of Medicine) performed limiting dilution on this cell line and isolated four subclones, SA13, SA14, SA17 and SA22. We carried these subclones for 20 passages and performed FISH on both interphase and metaphase cells to determine the number of X chromosomes in each genome. FISH is a powerful technique in which fluorochrome conjugated single-stranded nucleic acids specifically interact with complementary sequences to detect and localize DNA fragments (DNA FISH) or mRNA molecules (RNA FISH). DNA FISH using a probe, DXWas70, recognizing repetitive sequences near the centromere of mouse X chromosomes [324] established that the karyotype of all subclones was stable and ≤5% of cells lost one X over 40 doubling times in culture (Figure 3.4A and Table 3.1). Subclone SA13 maintained two X chromosomes at the highest frequency and was used in subsequent experiments.

The active state of the two Xs in undifferentiated SA13 cells was verified by RNA FISH. As described in the introductory chapter (see section 1.1.4), Xist RNA is transcribed at a low level from both Xs during early development, but upon XCI is highly up-regulated on the Xi and coats the chromosome in cis [92]. The RNA FISH experiment identified two pinpoint dots of Xist RNA in undifferentiated cells (Figure 3.4B) and one large accumulation of this RNA upon differentiation (Figure 3.4C), revealing that subclone SA13 has two active Xs at the undifferentiated stage and retains the ability to undergo XCI upon differentiation.

### 3.2.3 ES cell transfection

BACs were linearized prior to transfection into ES cells. Previous studies have shown that circular DNA molecules are randomly linearized before integration into the mammalian genome [312]. The large genomic insert constitutes the majority of BAC DNA and is the most likely region to be cut if circular BAC DNA is used for transfection. To increase the likelihood that the genomic insert remains intact during integration, BAC DNA was linearized within the vector region prior to transfection. RP23-391D18-neo was digested with SgrAI, which removed a ~1.2 kb vector DNA fragment (Figure 2.1A). As RP23-330G24-neo has an additional SgrAI site within the genomic insert, we linearized this BAC at a unique I-SceI site (Figure 2.1B). Additionally, it was important to linearize the BAC DNA at these sites to not disturb the specific palindromic motifs, loxP and lox511, that flank the genomic inserts.
**Figure 3.4 Subcloning of ES cell line EL1/c16.** (A) DNA FISH using the probe DXWas70 that detects pericentric repetitive sequences of the mouse X. An example of interphase nuclei and metaphase chromosome spreads for subclone SA13 is shown here. (B) RNA FISH of undifferentiated SA13 cells using the *Xist* probe. Two pin-point dots of *Xist* RNA signals in each interphase nucleus indicate that two active X chromosomes are present before *in vitro* differentiation of ES cells. (C) RNA FISH of differentiated SA13 cells using the *Xist* probe. A large accumulation of *Xist* RNA signals in each interphase nucleus indicates that one X chromosome has undergone XCI.
Table 3.1 Establishment of stable subclones of mouse female embryonic stem cells.

<table>
<thead>
<tr>
<th>Subclone</th>
<th>% of interphase cells(^1) with</th>
<th>% of metaphase cells(^2) with</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>39,X</td>
<td>40,XX</td>
</tr>
<tr>
<td>SA13</td>
<td>4</td>
<td>94</td>
</tr>
<tr>
<td>SA14</td>
<td>4</td>
<td>91</td>
</tr>
<tr>
<td>SA17</td>
<td>5</td>
<td>91</td>
</tr>
<tr>
<td>SA22</td>
<td>4</td>
<td>90</td>
</tr>
</tbody>
</table>

\(^1\)100 interphase cells were examined for each subclone.
\(^2\)20 metaphase cells were examined for each subclone.
\(^3\)80,XXXX cells are tetraploids.
The lox sites will allow the DNA sequences between them to be replaced through cre-mediated recombination if we later choose to test additional sequences at the same integration site (see section 5.4.1).

As an essential step to integrate exogenous DNA fragments into the genome, the DNA delivery process can influence transgene integrity and copy number [313]. Common DNA transfection techniques include chemical (calcium phosphate and liposome), physical (electroporation and microinjection) and viral infection methods. Among these methods, lipofection (liposome-mediated) and electroporation have been optimized and widely used for ES cells (e.g. [281, 325]). These two methods differ in the transportation mechanism and the properties of transgene fragments generated [313]. Lipofection delivers DNA molecules through a complex of liposome coated DNA; however, it frequently gives rise to multiple-copy transgenes [313]. Electroporation transfers DNA into the nucleus by emitting electric pulses that affect cell permeability and efficiency is not dependent of the size of DNA molecules. This method generates intact, single-copy integrants at a frequency of ~85%, which is significantly higher than lipofection [281]. Moreover, electroporation minimizes the period of time that transferred DNA is together with chemical compounds that may increase the chance to break large DNA molecules. Therefore, we chose to transfer DNA by electroporation.

Linearized BAC DNA was transferred into the mouse XX ES cells by electroporation. Using the conditions described in Chapter II, electroporation of 12 μg BAC DNA into $10^7$ ES cells yielded ~200 G418-resistant colonies. Following the isolation of stable transgenic cell lines, drug selection was maintained for one additional passage to limit false positives.

### 3.2.4 Screening for X-linked BAC transgenes

X-linked transgenes were identified by DNA FISH to metaphase chromosomes from each transgenic cell line. X chromosomes were marked by the pericentric probe DXWas70 [324], and labeled BAC DNA identified both endogenous and transgenic Jarid1c loci (Figure 3.5). Endogenous Jarid1c is located at 148.6 Mb on the mouse X chromosome that is 167 Mb in length. BAC DNA signals at other chromosomal sites indicate the transgene location. As the two X chromosomes comprise ~5% of the total female mouse genome, we expect that one in 20 randomly integrated transgenes will be X-linked. Indeed, 185 cell lines from three
Figure 3.5 FISH screening for X-linked BAC transgenes. (A) DNA FISH screening of metaphase chromosomes from 185 transgenic ES cell lines identified twelve X-linked BAC transgenes. A representative image is shown for each X-linked transgene. The probe DXWas70 (red) detects mouse X chromosome pericentric sequences. BAC DNA probes (green) hybridize to transgenic (yellow arrowhead) and endogenous loci (white arrow). The name of each line is shown in top left corner of the image. Line B051 contains a truncated X chromosome missing the endogenous Jarid1c region. All transgenes were derived from BAC RP23-391D18-Neo except A314 and A319 (RP23-330G24-Neo). (B) A representative image of autosomal BAC transgenes. (C) Enlarged X chromosomes from 12 independent ES cell lines with an X-linked transgene (arrowheads). Some images were published in [316].
independent transfection experiments were screened and twelve lines (6.5%) carried a BAC transgene on the X chromosome (Table 3.2 and 3.3, Figure 3.5). Lines A314 and A319 were generated from RP23-330G2-neo and the remaining ten lines from RP23-391D18-neo.

Line B051 contains one normal X chromosome with a transgene and one truncated X that is missing the distal end including the endogenous Jarid1c. As a complete complement of X-linked genes are essential for cell viability [11, 15], this cell line may inactivate the truncated X chromosome in all cells upon differentiation. Thus, this line was excluded from subsequent experiments.

3.2.5 BAC transgene integrity

As random integration frequently results in truncated transgene fragments [313], transgene integrity of the eleven cell lines was assessed. BAC vector sequences flank the genomic insert upon proper linearization such that intact transgene DNA should retain the vector sequences (Figure 3.6A and 3.7A). The transgene linearization sites and the two lox sites, loxP and lox511, flanking the BAC genomic inserts were then examined by PCR amplification of these sequences from the genomic DNA.

PCR analysis was first applied to the nine cell lines derived from RP23-391D18-neo. The BAC was linearized properly because the genomic DNA from all lines did not contain the sequences spanning the linearization site (Figure 3.6B). However, the loxP and lox511 sites were absent in five cell lines (Figure 3.6B), suggesting that these BAC transgenes do not have an intact genomic insert or truncation occurred between the insert and the vector sequences tested. However, the latter is less likely as lox511 is only several hundred bp distant from the genomic insert and, although loxP is ~4 kb away from the insert, the neo gene within this region must be intact leaving ~2 kb between the neo and the insert. Although these lines were rearranged or deleted, four transgenes were apparently intact. In lines B079, B181, B202 and B259, the BAC transgene linearized within the backbone region and maintained both ends of the linear product. It is highly possible that the genomic insert is intact.

The two cell lines derived from RP23-330G24-neo were analyzed similarly. This BAC did not linearize correctly because the I-SceI sequence was still present in both A314 and A319 genomic DNA (Figure 3.7B). Although loxP and lox511 were also present (Figure 3.7B), the linearization of BAC DNA likely occurred within the genomic insert.
Table 3.2 Mouse ES cell transfection experiments and screening of transgenic lines.

<table>
<thead>
<tr>
<th>Transfection series</th>
<th>BAC</th>
<th># screened</th>
<th># with an X-linked transgene</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>RP23-330G24-neo</td>
<td>42</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td>RP23-391D18-neo</td>
<td>116</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>C</td>
<td>RP23-391D18-neo</td>
<td>27</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>185</td>
<td>12</td>
<td>6.5</td>
</tr>
</tbody>
</table>
Table 3.3 Stability to maintain two X chromosomes in transgenic ES cell lines.

<table>
<thead>
<tr>
<th>Subclone</th>
<th>% of interphase cells(^1) with</th>
<th>% of metaphase cells(^2) with</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>39,X</td>
<td>40,XX</td>
</tr>
<tr>
<td>A314</td>
<td>2</td>
<td>47</td>
</tr>
<tr>
<td>B079</td>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td>B202</td>
<td>2</td>
<td>46</td>
</tr>
<tr>
<td>B259</td>
<td>1</td>
<td>48</td>
</tr>
</tbody>
</table>

\(^1\) 50 interphase cells were examined for each subclone.
\(^2\) 20 metaphase cells were examined for each subclone.
\(^3\) 80,XXXX cells are tetraploids.
Figure 3.6 Characterization of X-linked transgenes derived from BAC RP23-391D18-Neo. (A) Cartoon of BAC RP23-391D18-Neo (not to scale) indicates relevant genes in BAC vector (thin line) and genomic insert (thick line). Location of PCR products (grey bars), linearization sites SgrAl (hash mark), Southern probe (black bar), relevant SacI (S) and NsiI (N) restriction sites are also indicated. Numbered PCR products are used to examine the linearization sites (1), lox sites flanking the genomic insert (2 and 3) and a control fragment within the Jarid1c gene (4). Primers used to amplify these fragments are 1: SgrAL/SgrAR, 2: BACloxp1L/BACloxp1R, 3: BACloxp2L/BACloxp2R, 4: Smcxpb2L/Smcxpb2R, Southern probe: BACpbL/BACpbR (Table 2.1). (B) PCR analysis of transgene DNA. Transgenic lines, B079, B168, B176, B181, B202, B259, C014, C048, C138, are compared to BAC DNA (BAC) and the parental ES cell line (ES). (C and D) Southern analysis of SacI- (C) or NsiI- (D) digested genomic DNA to evaluate transgene copy number in six subclones from each line that may have an intact BAC transgene. Intensity differences between subclones reflect DNA loading differences. Some images were published in [316].
Figure 3.7 Characterization of X-linked transgenes derived from BAC RP23-330G24-Neo. (A) Cartoon of BAC RP23-330G24-Neo (not to scale) indicates relevant genes in BAC vector (thin line) and genomic insert (thick line). Location of PCR products (grey bars), linearization site I-SceI (double hash mark), Southern probe (black bar), relevant SacI (S) and NsiI (N) restriction sites are also indicated. Numbered PCR products are used to examine the linearization sites (1), lox sites flanking the genomic insert (2 and 3) and a control fragment within the Jarid1c gene (4). Primers used to amplify these fragments are 1: BACpsL/BACpsR, 2: BACloxp1L/BACloxp1R, 3: BACloxp2L/BACloxp2R, 4: Smcxpb2L/Smcxb2R, Southern probe: BACpbL/BACpbR (Table 2.1). (B) PCR analysis of transgene DNA. Transgenic lines, A314 and A319, are compared to BAC DNA (BAC) and the parental ES line (ES). (C) Southern analysis of SacI- (C) or NsiI- (D) digested genomic DNA to evaluate transgene copy number of each line. (D) Cartoon of A314 transgene. The BAC transgene linearized at a position between the NsiI and SacI sites flanking the neo gene. Some images were published in [316].
Transgene linearization in lines A314 and A319 was further characterized by Southern analysis using a BAC vector probe. *Sac*I digestion generated a product with the same size for BAC and A314 genomic DNA, whereas *Nsi*I yielded a slightly larger product from A314 genomic DNA (Figure 3.7C), inferring that A314 transgene linearized at a position between the *Nsi*I and *Sac*I sites flanking the *neo* gene (Figure 3.7D). As the *Nsi*I site is only ~5 kb away from vector sequences, majority of the genomic insert is intact. This conclusion was further confirmed by experiments to identify integration site location (see section 3.2.8). In contrast to line A314, southern analysis of A319 genomic DNA digested with *Sac*I or *Nsi*I detected fragments that were similar to digested BAC DNA. Intact BAC vector sequences in A319 suggests that the transgene linearized within the genomic insert, likely some distance from vector sequences although the precise location remains unknown.

### 3.2.6 BAC transgene copy number

Transgene copy number was examined as it may cause variation in transgene expression that is unrelated to XCI. Copy number of the BAC DNA was inferred by Southern analysis using a BAC vector probe (Figure 3.6A and 3.7A). This probe hybridizes to a fragment that contains BAC vector sequences and also is expected to include adjacent endogenous DNA at the integration site. Multiple hybridizing bands are indicative of transgene rearrangement or multiple copies that have inserted. Southern analysis of *Sac*I- or *Nsi*I- digested genomic DNA showed only one hybridizing band in B079, B202, B259, A314 and A319, suggesting a single BAC insert in each line (Figure 3.6C, D and 3.7C). In contrast, line B181 contained two bands suggestive of at least two copies of BAC DNA (Figure 3.6C and D).

### 3.2.7 Transgene expression in undifferentiated ES cells

To determine whether the transgenic X chromosomes are stable during cell proliferation without selection, the ability of each cell line to maintain two Xs was tested by DNA FISH in both interphase nuclei and metaphase chromosomes. X chromosomes were detected using the X chromosome control probe DXWas70. After twenty passages, ≥95% of cells from each line were 40,XX (Table 3.3), suggesting that both the transgenic and normal X chromosomes are stable and transgene integration does not seem to affect the stability of the X chromosomes.
Successfully maintaining a proper karyotype also suggests that the genetic background is stable during cell proliferation.

As many transgenes are silenced by position effects that are unrelated to XCI response [326], Jarid1c transgene expression was assessed in undifferentiated cells carrying two active X chromosomes. By RNA FISH, Jarid1c nascent transcripts containing intronic sequences were detected using a Jarid1c genomic DNA probe and diploid cells were confirmed by a probe that identifies Xist/Tsix RNA products from two Xs (Figure 3.8A). The Jarid1c probe does not distinguish endogenous and transgene transcripts. As expected for lines carrying an expressed transgene, three Jarid1c RNA foci were detected in >95% of diploid cells scored from lines A314, B079, B202 and B259. Additionally, two Jarid1c RNA signals were always close to one another and were in the vicinity of one Xist/Tsix signal, consistent with a transgenic X chromosome carrying two copies of Jarid1c, one endogenous and one transgenic. As a control, non-transgenic ES cells showed only two signals of Jarid1c expression, consistent with the presence of only two endogenous alleles (data not shown). The Jarid1c transgene in the four cell lines, similar to the endogenous alleles, is transcribed in undifferentiated ES cells and thus is active prior to XCI.

Transgene activity in line A319 was also tested. Undifferentiated A319 cells showed only two dot-like Jarid1c signals in RNA FISH (data not shown), suggesting that the transgene or one endogenous allele was not expressed. To determine which allele was inactive, sequential RNA and DNA FISH was performed to localize transgene DNA (using a BAC vector probe) relative to the Jarid1c nascent transcripts. As a control, line A314 showed three Jarid1c RNA foci representing expression from two endogenous loci and the transgene (Figure 3.8B). Two foci were always close and transgene DNA corresponded to one of the two signals, confirming an expressed Jarid1c transgene. In contrast, line A319 contained two Jarid1c RNA foci and the transgene DNA signal did not colocalize to any of the RNA signals (Figure 3.8B). This indicated that the Jarid1c transgene is silenced on an active undifferentiated X in A319. Given that this BAC transgene linearized within the genomic insert, it may be truncated within the Jarid1c genomic region leading to inactivation of the transgene. Therefore, line A319 was excluded from subsequent analysis.

Using RNA FISH, two genes that are also included in the BAC transgenes, Igsec2 and Gpr173, were found to be transcribed in undifferentiated cells as well. The Igsec2 transgene
Figure 3.8 Expression of transgene transcripts from active X chromosomes prior to ES cell differentiation. (A) RNA FISH to detect Jarid1c nascent transcripts and Xist/Tsix RNA in undifferentiated cells of lines A314, B079, B202 and B259. Representative nuclei are shown. Of >100 nuclei scored for each line, three Jarid1c foci (two endogenous and one transgene) were detected in >95% of diploid cells. (B) Sequential RNA and DNA FISH to detect Jarid1c nascent transcripts and BAC DNA locus in lines A314 and A319. Representative nuclei are shown. Only two Jarid1c signals were present in line A319 and BAC DNA does not colocalize to any of them, not like line A314. (C) RNA FISH to detect Iqsec2 nascent transcripts and Xist/Tsix in line A314. Of >100 nuclei scored, three Iqsec2 foci were detected in >97% of diploid cells. (D) RNA FISH to detect Gpr173 nascent transcripts and Xist/Tsix in lines B079, B202 and B259. Of >100 nuclei scored for each line, three Tspyl2 foci (two endogenous and one transgene) were detected in >90% of diploid cells. Some images were published in [316].
in line A314 and the Gpr173 transgene in lines B079, B202 and B259 demonstrated similar FISH patterns to that of the Jarid1c transgene (Figure 3.8C and D). While RP23-330G24-neo contains only Jarid1c and Iqsec2, RP23-391D18-neo includes five transcripts with Jarid1c and Gpr173 flanking other genes. These data not only confirm BAC transgene activity prior to XCI, but also suggest that these BAC transgenes are absent of substantial deletions and rearrangements. Altogether, these data suggest that each of the cell lines B079, B202, B259 and A314 carry an intact or largely intact, single-copy BAC transgene. Xi expression would be examined only in these four cell lines.

3.2.8 Integration sites of X-linked BAC transgenes

In order to evaluate whether features of the chromosomal integration site may affect transgene Xi expression, the precise integration site of each transgene was identified by inverse PCR (Figure 3.9) [327]. Transgene location and orientation are indicated in Table 3.4 and Figure 3.10. The insertion sites are consistent with the cytogenetic location observed in the DNA FISH experiment. B079 and B202 transgenes reside at gene-poor regions and endogenous genes are quite distant from the integration sites, whereas B259 transgene is located in a relatively gene-rich region. Intriguingly, A314 transgene integrated into the second exon of the Mid1 gene. Additionally, genomic sequences of several kb were deleted at the site of integration in lines B079 and B259, and transgene integration in line B202 caused a substantial deletion of ~250 kb of endogenous sequences although this region does not contain any known transcripts. We were unable to determine the location of the proximal end of A314 transgene due to the improper linearization, and could not assess whether endogenous sequences was deleted in this line.

Integration site sequencing also identified that A314 transgene linearized at a position 4.0 kb distant from the end of the genomic insert. Therefore, all transcripts in this BAC transgene, including Jarid1c and adjacent Iqsec2, are unaffected (Figure 3.7D).

3.2.9 X inactivation status of transgene integration sites

To determine whether BAC transgenes integrated into normally X-inactivated regions, we evaluated the XCI landscape of integration sites by examining the XCI status of adjacent genes on a non-transgenic X chromosome. Transcripts from annotated genes and ESTs
**Figure 3.9 Scheme of inverse PCR method.** a) Genomic DNA from each transgenic line is digested using an appropriate restriction enzyme (E) to generate DNA fragments that contain BAC vector sequences (thin lines) and endogenous DNA of the integration site (thick lines). b) These chimeric DNA fragments are self-ligated to generate circular products. c) Integration site sequences are PCR amplified using primers located within the BAC vector regions (arrows). d) PCR products are cloned into the TA vector (from Invitrogen). e) Sequence the TA constructs and blast the results to identify integration sites.
a. restriction digest

b. self ligation

c. PCR

d. TA cloning

e. Sequencing and homology search of each end

X chromosome DNA
BAC backbone
Genomic DNA in BAC
Restriction site
Primer pair
Table 3.4 Chromosomal location of transgene integration sites.

<table>
<thead>
<tr>
<th>Transgene</th>
<th>Proximal end (bp)†</th>
<th>Distal end (bp)†</th>
<th>Cytogenetic band</th>
<th>Topoisomerase I consensus sites*</th>
</tr>
</thead>
<tbody>
<tr>
<td>B079</td>
<td>36,658,864</td>
<td>36,659,365</td>
<td>G</td>
<td>Yes</td>
</tr>
<tr>
<td>B202</td>
<td>87,604,183</td>
<td>87,848,273</td>
<td>G</td>
<td>Yes</td>
</tr>
<tr>
<td>B259</td>
<td>131,946,410</td>
<td>131,947,351</td>
<td>R</td>
<td>Yes</td>
</tr>
<tr>
<td>A314</td>
<td>N/A‡</td>
<td>166,365,036</td>
<td>R</td>
<td>Yes</td>
</tr>
</tbody>
</table>

†Map locations are from the UCSC genome browser (http://genome.ucsc.edu/), build 37, July 2007 assembly.
‡This side does not contain BAC vector sequences and is not eligible for inverse PCR analysis.

*Topoisomerase type I strong consensus sites (5’-A/T-G/C-T/A-T-3’) were searched in the 20 bp sequences at the chromosome/transgene junctions.
Figure 3.10 Chromosomal location and transcription orientation of the four X-linked BAC transgenes. The distance of adjacent genes to the integration site is shown in parenthesis. A314 transgene is located in the second exon (thick black bar) of the Mid1 gene.
adjacent to each integration site were obtained from the UCSC genome browser (http://genome.ucsc.edu/, July 2007 assembly) and the Perlegen database (http://mouse.perlegen.com/mouse/browser.html). To test Xi expression status, we applied the Q-SNaPshot assay to a non-randomly inactivated fibroblast cell line, B119 [174]. This cell line was derived from F1 animals of an inter-subspecies cross between the Searle’s translocation (T16H) carrier and the *M.m.castaneous* (CAST). To maintain proper dosage of the genes on the translocated chromosomes, the intact CAST X chromosome is inactivated in all cells [174]. Using these cells, the XCI status of a gene, as well as the Xi expression level if it escapes, can be determined directly from the CAST allelic expression. As a control, endogenous *Jarid1c* clearly showed transcription from both Xa and Xi, with an escape level of ~33% when compared to its Xa level (Figure 3.11). Expression from endogenous *Iqsec2* and *Tspyl2* was absent from the inactive CAST X chromosome (Figure 3.11).

Integration site genes were then analyzed. Importantly, ratios between Xi and Xa expression for all integration site genes tested are less than 0.1% after normalization to DNA signals (data not shown). This mono-allelic expression indicates that they are expressed from the Xa only and therefore are X-inactivated (Figure 3.11). With the Q-SNaPshot assay, the closest transcripts that are expressed in fibroblasts and contain transcribed SNPs were assayed. Genes that are closer to the insertion sites that were excluded from analysis showed tissue-specific expression patterns or lacked exonic SNPs. Therefore, the genes tested may lie some distance from the integration sites, especially in lines B079 and B202. The transgene in line B079 inserted within a large gene-poor region, with the closest assayable transcripts more than 1 Mb away from the insertion site. B202 carries the transgene in another gene-poor region with a number of *Mageb2* gene duplications in which SNPs are not unique in the genome. The only transcript available for the Q-SNaPshot analysis (EST BY674816) is approximately 900 kb at the proximal end of the integration site; the inactivation status of genes on the distal end is not clear. Nevertheless, endogenous genes are located at or close to the transgene integration site in lines A314 and B259. Altogether, these data suggest that the four BAC transgenes have integrated into normally X-inactivated regions, with the caveat that the closest testable transcripts lie some distance from the insertion site in two lines.

### 3.2.10 Strain origin of transgenic X chromosomes
Figure 3.11 Allelic expression of genes adjacent to transgene integration sites and transcripts at the Jarid1c locus. The primary fibroblast line B119 was derived from a (T16H x CAST) F1 mouse and carries an active *M. domesticus*-derived X (DOM) and an inactive *M. castaneus* X (CAST). Q-SNaPshot of genomic DNA (gDNA) confirms that B119 is informative for each SNP. Mono-allelic expression in cDNA only from the active DOM-derived X chromosome indicates that each gene is subject to X inactivation except *Jarid1c*. Primers used in these assays are shown in Table 2.2. Some images were published in [316].
As the two X chromosomes in our ES cell lines are inactivated with different probabilities, prior to analyzing Xi expression of BAC transgene transcripts it is necessary to determine how often each transgene is located on the Xi. While either X chromosome can undergo XCI, the probability that a particular X will be chosen to be inactivated is influenced by strain origin of the X. As mentioned previously, the ES cells used in our studies are hybrids with X chromosomes from two different mouse strains, 129 and CAST [76]. The 129 X carries an \(Xce^a\) allele, while the CAST X contains an \(Xce^c\) allele. Genetic variations between these two \(Xce\) alleles results in skewed XCI upon ES cell differentiation. The 129 X is inactivated in a higher percentage of cells (~75%) [76, 328]. Therefore, the frequency that each transgene is located on the Xi is influenced by strain origin of the transgenic X. A transgene integrated on the 129 X will be on the Xi in ~75% of cells but it will be on the Xi in only 25% of cells if it integrated onto the CAST X.

Strain origin of the transgenic X chromosomes was inferred by identifying the percentage of cells with the transgene on the Xi. To calculate this frequency, the four transgenic cell lines were \textit{in vitro} differentiated for ten days and fully differentiated cells at the periphery of outgrown embryoid bodies were analyzed by sequential RNA and DNA FISH. Interphase nuclei were hybridized under non-denaturing conditions to detect the \(Xist\) RNA that demarcates the Xi. Following denaturing and hybridizing nuclei with labeled BAC DNA, which detects the transgene locus and also two endogenous \(Jarid1c\), three DNA loci were identified in ≥95% of cells. The proportion of cells with two BAC signals colocalizing to \(Xist\) RNA (transgene on the Xi) was compared to the percentage of cells with two BAC signals clearly separated from the \(Xist\) RNA (transgene on the Xa). Consequently, the transgene in lines B079 and B259 was on the Xi in >70% of cells (Figure 3.12), indicating that the transgene integrated onto the 129 X. However, the transgene in lines A314 and B202 was on the Xi in < 30% of cells (Figure 3.12), suggesting insertion onto the CAST X.

Strain origin of transgenic X chromosomes also can be determined by polymorphisms within integration site sequences. The endogenous sequence adjacent to A314 transgene includes five single nucleotide polymorphisms (SNPs) between mouse strains 129 and CAST. Our integration site sequencing results demonstrated a CAST allele at all of the five sites, which confirms integration into the CAST X (Table 3.5). Importantly, by directly establishing that A314 transgene integrated on the CAST X and by demonstrating that this line inactivated
**Figure 3.12 Strain origin of transgenic X chromosomes.** Sequential RNA and DNA FISH was performed on differentiated cells to analyze Xist RNA, marking the inactive X, and BAC DNA. Only nuclei with three BAC DNA signals (two endogenous and one transgene) were scored to ensure that replicated chromatids did not complicate analysis. Expected results reflect that the transgene on the CAST or 129 X is located on the inactive X (Xi) in a proportion of cells but are on the active X (Xa) in all remaining cells. Images and tables were published in [316].
<table>
<thead>
<tr>
<th>Expectation</th>
<th>% cells</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tg on CAST X</td>
<td>~75</td>
<td>~25</td>
</tr>
<tr>
<td>Tg on 129 X</td>
<td>~25</td>
<td>~75</td>
</tr>
</tbody>
</table>

**Observation**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A314</td>
<td>72</td>
<td>28</td>
<td>Cast X</td>
</tr>
<tr>
<td>B079</td>
<td>24</td>
<td>76</td>
<td>129 X</td>
</tr>
<tr>
<td>B202</td>
<td>74</td>
<td>26</td>
<td>Cast X</td>
</tr>
<tr>
<td>B259</td>
<td>21</td>
<td>79</td>
<td>129 X</td>
</tr>
</tbody>
</table>
Table 3.5 SNPs included in the integration site sequences of line A314

<table>
<thead>
<tr>
<th>SNP</th>
<th>Strain 129</th>
<th>Strain CAST</th>
<th>Integration site sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs31818289</td>
<td>T</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>rs31818287</td>
<td>T</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>rs31816669</td>
<td>A</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>rs31816666</td>
<td>T</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>rs31816665</td>
<td>T</td>
<td>C</td>
<td>C</td>
</tr>
</tbody>
</table>
the CAST X in a proportion of cells that is equivalent to non-transgenic cells, these data indicate that BAC transgene insertion did not influence the ability of either X chromosome to be inactivated. Unfortunately, the transgene in lines B079, B202 and B259 is not near any informative SNPs and therefore transgenic Xs could not be confirmed by this approach.
3.3 DISCUSSION

We have generated a series of transgenic ES cell lines that carry *Jarid1c* BAC DNA on the X chromosome. Two BACs containing *Jarid1c* and surrounding genes were introduced into the genome of undifferentiated mouse XX ES cells by random integration. As expected, transgene DNA in some cell lines are rearranged. However, four lines incorporated an intact or largely intact, single-copy BAC transgene. While two transgenes are located on the 129 X, two others are on the CAST X. Further, these four transgenes integrated into different X-inactivated loci on the X that potentially represent diverse Xi environment upon XCI. Regardless of the different integration sites, all transgene transcripts tested including *Jarid1c* are active prior to XCI and thus are not influenced by position effects. Thus, these four lines are ideal to determine whether the *Jarid1c* escape domain embedded in the BAC DNA is autonomous or whether it is dependent on chromosome location.

3.3.1 Transgene random integration

As mentioned earlier, randomly integrated transgenes are helpful to determine whether an escape domain is autonomous. Several lines of evidence suggest that our X-linked transgenes are randomly distributed. DNA FISH screening identified X-linked transgenes at a frequency consistent with random integration (6.5% versus 5%). Cytogenetically, these transgenes are well spaced along the X. Integration site sequencing further determined distinctive locations of the transgene in the four cell lines. However, transgene integration may not be completely random. Statistical analysis of the distribution of integration sites has suggested the presence of “hot areas” for random integration, but the frequency of integrating into such areas is only slightly higher than elsewhere in the genome [281]. Previous studies also reported that short homologies of one to three nucleotides and the consensus sequence for topoisomerase I cleavage sites were present at the sites of random integration [281, 329, 330]. As the eukaryotic type I topoisomerase induces single strand nicks in DNA and ligates non-homologous ends, it may catalyze the DNA integration event [329]. At the chromosome level, transgenes tend to integrate into G-positive band regions, which have specific sequence composition, such as AT enrichment, and a low gene density [281]. Intriguingly, all of the four cell lines established in this study contain the topoisomerase I recognition sequences (5’-A/T-G/C-T/A-T-3’)) at the integration sites although only two transgenes are located at
G-positive band regions (Table 3.4). Nevertheless, the consensus sites are present in the genome with the likelihood of 1/32; therefore, they should span the genome and do not significantly affect the location of transgene integration at the chromosome level.

3.3.2 Transgene DNA integrity

As transgene integrity is important to determine whether the Jarid1c BACs contain an autonomous escape domain, in addition to selecting an optimal approach to transfer BAC DNA, other efforts also have been taken to ensure intact transgene DNA. Because BAC DNA is relatively large and easy to shear or degrade compared to smaller plasmids, extra care was taken during the preparation and linearization of the BAC DNA. After insertion of a neo cassette into the BAC backbone, no detectable change in the restriction pattern was observed for the genomic insert between modified and wild-type BACs, suggesting the absence of gross rearrangements and deletions during BAC modification. To avoid truncation of the genomic insert resulted from random digestion of circular BAC DNA during transfection, BACs were linearized within the vector region prior to introduction into ES cells. Complete BAC inserts were supported by the expression of two transcripts in each BAC transgene from active X chromosomes. Intact transgene DNA in lines B079, B202 and B259 was also inferred by proper linearization on the BAC backbone and presence of the vector sequences flanking the genomic insert, as truncated transgene DNA usually loses the end sequences either by DNA breakage or exonuclease digestion. However, DNA rearrangement and deletion may occur within the BAC genomic insert [331]. Additionally, while A314 transgene linearized at 4 kb of the distal end of the BAC genomic insert, whether any sequences were lost at the opposite end of the transgene DNA is unknown. Although my thesis work did not test these likelihoods, they can be addressed by some of the approaches described in Chapter II. For instance, gross rearrangement can be revealed by Southern analysis using BAC genomic sequences as probes. As the BAC genomic insert contains abundant SNPs that do not distinguish the transgene from the endogenous 129 allele but both vary from the CAST allele, allelic ratios at these polymorphisms can be determined by an allele-specific primer extension assay to locate truncations in the BAC transgenes.

3.3.3 Transgene copy number
The copy number was assessed by detecting the DNA fragments in the genome that contain a particular BAC vector sequence. These fragments also contain additional sequences, such as integration site sequences if the transgene is adjoining to endogenous genomic DNA or another copy of transgene DNA if transgenes form concatamers. A multiple-copy transgene generates several fragments likely differing in size and thus can be visualized by Southern analysis. By this method, we identified four X-linked transgenes with single-copy, intact BAC integrants. However, a transgene that produces fragments with similar length or loses the vector sequence will be miscalculated by this method. To definitively address this problem, Southern analysis can be used to detect BAC genomic sequences and quantify transgene DNA relative to endogenous sequences or serial dilutions of BAC DNA. Alternatively, the allelic ratio at the polymorphisms carried in the BAC genomic DNA can be used to quantify transgene copy number through a comparison between transgenic and non-transgenic lines. Using this approach, the experiments described in the next chapter demonstrate that line B259 includes one copy of transgenic \textit{Tspyl2} and \textit{Jarid1c} and that line B079 contains one copy of transgenic \textit{Tspyl2} and likely more than one copy of transgenic \textit{Jarid1c} (see section 4.2.2 and 4.2.3).

3.3.4 Transgene integration site effect

Notably, a negative effect on transgene expression that is frequently observed during random integration was not seen in our transgenic lines. Random integration into mammalian genomes sometimes results in unstable and unpredictable expression due to position effects [313]. For instance, transgene incorporation into regions of transcriptional silencing, such as heterochromatin or in close proximity to the regulatory elements of other genes can have subtle or dramatic impact on transgene transcription [332, 333]. Large genomic sequences constructed in BACs, PACs and YACs are common strategies to avoid these position effect variegations. As these large constructs likely contain all necessary elements to control gene expression and ample flanking sequences to buffer the influence of adjacent regions, they have been found to be auto-regulated at many genomic loci [312, 313]. Nevertheless, large inserts still can be susceptible to integration site effects [334]. We tested this likelihood in the four transgenic lines and found that none of them was silenced by position effects as two
transcripts in each transgene were expressed in nearly all undifferentiated cells. Hence, transgene expression upon XCI is unrelated to position effect.

3.3.5 Subspecies difference of Xi expression status

During the characterization of integration site genes, we identified a gene with strain-specific XCI status. In line A314, BAC DNA integrated into the second exon of Mid1 on the CAST X chromosome. This gene was subject to XCI when it was tested in the cell line B119 that carries an Xi from the CAST strain (Figure 4.8). Intriguingly, Mid1 has been demonstrated to escape XCI in the C57BL/6J strain [189]. This is the first example of a gene that shows different Xi expression patterns in different mouse strains. The C57BL/6J Mid1 spans the boundary between the X-specific area and the PAR that is also present on the Y and escapes XCI [335]. Thus, the XCI status of this gene may be influenced by the adjacent escape region. It is interesting that the Mid1 gene in another mouse stain, M.spretus, is essentially located in the X-specific region although its XCI status remains unknown [335]. The structure of CAST Mid1 and the CAST X chromosome has not been well studied, and it is not clear why this gene has a differential XCI status. Investigation of gene structure and location may be helpful to answer the question whether the XCI status of a gene is associated with its chromosomal location.
CHAPTER IV
AUTONOMOUS EXPRESSION OF Jarid1c BAC TRANSGENES
ON INACTIVATED X CHROMOSOMES

4.1 INTRODUCTION

Upon XCI, inactivation initiates from the X inactivation center, spreads along the length of the X and transcriptionally silences genes on this chromosome [51]. Answering the question of how X-linked genes acquire proper Xi regulation is key to understand how the chromosome-wide inactivation is established and maintained. Importantly, gene silencing is not complete on the Xi. A subset of genes escape XCI and are expressed from both the Xa and Xi [25, 183]. These genes provide a valuable experimental tool to study Xi gene regulation as an appropriate model of X inactivation should also explain how escape genes remain expressed on the largely inactivated X. Previous studies suggested that escape from XCI was controlled at the level of chromosomal domains [8, 25, 196, 220]. However, such domains have not been defined and regulatory elements necessary for escape remain elusive.

In order to functionally delineate the chromosomal domain responsible for escape, we asked whether a domain that escapes XCI is autonomous or instead takes on properties of the region on the X in which it is located. Specifically, we sought to determine whether the mouse escape gene Jarid1c, in the context of adjacent genomic sequences, would retain Xi expression upon relocation on the X.

Chapter III described the establishment of four mouse ES cell lines that carry a Jarid1c BAC transgene on one of the two X chromosomes. These intact, single-copy transgenes reside at different X loci and are expressed prior to XCI, making them ideal for characterizing the Jarid1c escape domain. The BACs used in our studies include Jarid1c and adjacent X-inactivated transcripts, Iqsec2 and Tspyl2. We hypothesized that the BAC DNA contains an autonomous Jarid1c escape domain and therefore (1) the Jarid1c transgene would escape XCI in all transgenic cell lines; (2) adjacent transgenes would be X-inactivated. To test our hypothesis, we differentiated the transgenic cell lines to induce XCI and evaluated the Xi expression pattern of transgene transcripts by sequential RNA and DNA FISH and allelic expression assays. The XCI status and candidate regulatory sequences at each integration site were also investigated to understand their roles in Xi gene regulation.
4.2 RESULTS

4.2.1 X inactivation of transgenic *Iqsec2*

We first tested whether genes that are normally X-inactivated were properly regulated when present in large BAC transgenes. The X-inactivated gene *Iqsec2* is fully contained in BAC RP23-330G24-neo, which was used to generate the transgenic line A314. Thus, *Iqsec2* transgene expression was analyzed in this line. The A314 transgene is on the Xi in ~25% of differentiated cells as it is located on the CAST X chromosome (Figure 3.11).

Fully differentiated cells where XCI has been well established were chosen to assess Xi expression status. ES cells aggregate and form embryoid bodies (EBs) during spontaneous differentiation. However, substantial undifferentiated cells and randomly differentiated cells with different developmental stages still accumulate in these multi-cellular structures [336]. As differentiated cells outgrow from EB colonies at the initiation of differentiation and continue to spread out during culturing [337], cells at the periphery of the EB colonies are fully differentiated after ten-day differentiation and were chosen to examine in FISH experiments.

Sequential RNA and DNA FISH was used to identify cells with the transgene on the Xi and assess *Iqsec2* transcription in comparison to cells with the transgene on the Xa. First, interphase nuclei were hybridized under non-denaturing conditions with an *Iqsec2* probe that detects nascent transcripts from both endogenous and transgene loci and an *Xist* probe that marks the Xi. Subsequently, after signal fixation and denaturation, labeled BAC DNA was hybridized to identify both endogenous and transgene DNA loci. Transgene DNA location was determined by DNA foci and *Xist* territory; when the transgene was on the Xi, two DNA foci were within or in vicinity of the *Xist* domain (Figure 4.1A). Transgene expression status was then revealed by determining whether *Iqsec2* RNA signals overlap with any transgene DNA signals. This approach provides a manner to examine gene expression in single cells from all genomic loci, including both endogenous and transgenic. Further, as the endogenous Xa allele is expected to be expressed, hybridization to endogenous Xa transcripts serves as an internal control for transgene expression. When the transgene was on the Xa, two Xa transcript signals were present in ≥90% of cells (data not shown) and corresponded to the endogenous allele and the transgene, indicating transgenic *Iqsec2* was transcribed from the Xa (Figure 4.1A). To ensure that diploid nuclei with optimal hybridization were analyzed, only
Figure 4.1 X inactivation status of transgenic *Iqsec2* by sequential RNA and DNA FISH. (A) Representative nuclei (pseudocolored grey) of line A314 with the transgene (Tg) on the Xa or Xi. Sequential RNA and DNA FISH was performed using probes to detect *Iqsec2* RNA (green), transgene DNA (red) and *Xist* RNA (blue). (B) Summary of *Iqsec2* FISH results in line A314. Only *Xist* RNA positive cells with three clear BAC DNA foci and *Iqsec2* RNA transcripts from at least one Xa locus were scored. Hybridization patterns are denoted with transgene (Tg) location, XCI status of the transgene and/or endogenous (Edg) locus. The percentage of cells showing each pattern was compared to expectations for a transgene that is inactivated or escapes XCI (>100 nuclei scored). The statistical significance (P-value) of comparison to an escape gene by Chi square statistic is indicated. Experiments were performed in duplicate and the average numbers are shown. Images and tables were published in [316].
### A

- **Tg on Xa**
  - lqsec2 RNA
  - BAC DNA
  - Xist RNA

- **Tg on Xi**
  - lqsec2 RNA
  - BAC DNA

### B

<table>
<thead>
<tr>
<th>Results</th>
<th>% cells</th>
<th>Conclusion</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expectation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tg inactive</td>
<td>~75</td>
<td>~25</td>
<td>0</td>
</tr>
<tr>
<td>Tg escape</td>
<td>~75</td>
<td>0</td>
<td>~25</td>
</tr>
<tr>
<td><strong>Observation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A314</td>
<td>78</td>
<td>19</td>
<td>2</td>
</tr>
</tbody>
</table>

**FISH signals**
- [ ] Xist RNA
- [ ] BAC DNA
- [ ] lqsec2 RNA
Xist positive cells with three clear BAC DNA foci and Iqsec2 RNA transcripts from all Xa loci were scored. When the transgene was on the Xa, the endogenous Xi expression was absent in nearly all cells (99%, Figure 4.1B pattern 1 and 4), suggesting endogenous Iqsec2 was silenced on the Xi. Further, observed proportions for all FISH patterns were statistically different from the expectation for a transgene that escape XCI (Figure 4.1B). These data were consistent with the XCI status established previously ([196] and Figure 3.11) and also validated this approach to test Xi expression.

By this method, the expression pattern of transgenic Iqsec2 was similar to that of the endogenous gene. Cells carrying the transgene on the Xi did not express Iqsec2 from either the endogenous or transgene loci on the Xi (Figure 4.1A), and the proportion of these cells closely matched the percentage of cells that inactivated the transgenic X (Figure 4.1B pattern 2). Thus, Iqsec2 expression was detected from the Xa but not from the Xi in nearly all cells (97%, Figure 4.1B patterns 1 and 2). Iqsec2 RNA signals were associated with Xi loci in a small proportion of cells (3%, Figure 4.1B patterns 3 and 4). This observation may represent non-specific hybridization or suggest that the transgene or the endogenous allele escapes XCI in a very small percentage of cells. Altogether, our results demonstrate that the Iqsec2 transgene, similar to the endogenous alleles, is X-inactivated in all, or at least the vast majority, of cells. We conclude that the normally X-inactivated Iqsec2 remains subject to XCI when placed at an ectopic transgene location.

4.2.2 X inactivation of transgenic Tspyl2

The Tspyl2 gene is included in BAC RP23-391D18-neo, which was used to generate transgenic lines B079, B202 and B259. We tested Tspyl2 expression in these cell lines by sequential RNA and DNA FISH using the same criteria as those for the Iqsec2 experiments. Tspyl2 transcripts for endogenous and transgene loci were absent specifically on the Xi in some cells (Figure 4.2A), suggesting both were X-inactivated. However, Tspyl2 was not detected on the Xa in a large percentage of cells (data not shown). Even by RNA FISH where RNA signals are not compromised by DNA FISH procedures that require denaturing samples at high temperatures, Tspyl2 transcripts were detected from the Xa only in a small proportion of cells (Figure 4.2B). With few scorable cells, our results did not approach statistical significance. Similarly, only a subset of undifferentiated cells exhibited Tspyl2 RNA signals in
Figure 4.2 X inactivation status of transgenic Tspyl2 by FISH. (A) A representative nucleus (pseudocolored grey) of line B079 with the transgene on the Xi. Sequential RNA and DNA FISH was performed using probes to detect Iqsec2 RNA (green), transgene DNA (red) and Xist RNA (blue). (B) Summary of Tspyl2 RNA FISH results in lines B079, B202 and B259. All hybridization patterns were scored and are denoted with transgene (Tg) location and XCI status of transgene and/or endogenous (Edg) locus. The percentage of cells showing each pattern was compared to expectations for a transgene that is inactivated or escapes XCI. 100 nuclei were scored for lines B079 and B259, and 69 nuclei were scored for B202.
A

B

<table>
<thead>
<tr>
<th>FISH signals</th>
<th>Tg on Xa Edg Xi off</th>
<th>Tg on Xi Tg &amp; Edg Xi off</th>
<th>Tg on Xi Tg or Edg Xi on</th>
<th>Tg on Xa Edg Xi on</th>
<th>No Tspyl2 RNA signal from Xa or Xi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xist RNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tspyl2 RNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Results</th>
<th>% cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tg on 129 X</td>
<td></td>
</tr>
<tr>
<td><strong>Expectation</strong></td>
<td></td>
</tr>
<tr>
<td>Inactivated</td>
<td>~25</td>
</tr>
<tr>
<td>Escape</td>
<td>~25</td>
</tr>
<tr>
<td><strong>Observation</strong></td>
<td></td>
</tr>
<tr>
<td>B079</td>
<td>1</td>
</tr>
<tr>
<td>B259</td>
<td>0</td>
</tr>
</tbody>
</table>

| Tg on CAST X |         |
| **Expectation** |     |
| Inactivated | ~75 | ~25 | 0 | 0 | 0 |
| Escape     | ~75 | 0 | ~25 | 0 | 0 |
| **Observation** |     |
| B202       | 0 | 14 | 3 | 0 | 83 |
FISH experiments (data not shown). According to the NCBI gene expression database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=geo), Tspyl2 is ubiquitously transcribed; however, the inability to detect Tspyl2 transcripts by FISH may be due to low expression levels. Alternatively, hybridization to the nascent transcripts may not be efficient as the Tspyl2 genomic sequence is only ~5 kb in length. An additional strategy was required to test this transgene.

Tspyl2 transgene expression was then assessed by comparing allelic expression in transgenic and non-transgenic cell lines. The BAC DNA used in these studies was derived from the mouse strain C57BL/6 and contains transcribed SNPs that are the same as the endogenous 129 allele but both differ from the endogenous CAST allele, such as the Tspyl2 SNP rs29297026 (Figure 4.3A). Relative allelic expression levels can be measured by an allele-specific primer extension assay, Q-SNaPshot at these polymorphic sites (Figure 2.2) [25]. These allelic ratios reflect the level of 129 expression in non-transgenic cells, but the combined level of 129 and transgene expression in transgenic cells. Using this strategy, transgene expression can be inferred by the difference in the allelic expression ratio between transgenic and non-transgenic lines.

The two cell lines with the transgene on the 129 X chromosome, B079 and B259, were tested by this approach. As the CAST X chromosome does not carry the transgene in these lines, the level of CAST expression can be used as an internal control of the level from the 129 X. Additionally, the transgene is on the 129 X and both alleles can be monitored whether on the Xa or Xi. This allows us to compare the transgene and 129 expression in mixed differentiated cells with either X being inactivated, as long as cell populations with the same composition are tested for non-transgenic and transgenic lines. Furthermore, in both lines, the transgene is on the Xi in a high percentage of differentiated cells (~75%, Figure 3.11). This facilitates the detection of transgene Xi expression since most escape genes are expressed from the Xi at levels significantly lower than Xa levels (see section 1.2.5).

The allelic expression level of the 129 X relative to the CAST X was first quantitated in undifferentiated cells. As expected for cells with two active X chromosomes, equivalent expression was seen for the 129 and CAST alleles in non-transgenic ES cells (Figure 4.3B and C). In both transgenic cell lines, the combined level of the 129 allele and transgene was approximately twice that of the CAST allele (Figure 4.3B and C), suggesting that transgenic
Figure 4.3 X inactivation status of transgenic Tspyl2 examined by Q-SNaPshot assays. (A) Atspyl2 SNP differentiates the 129 allele and transgene from the CAST allele. The color of the nucleotides indicates the fluorescently labeled di-deoxynucleotides used in Q-SNaPshot. (B) Tspyl2 expression in non-transgenic ES cells (ES) and transgenic lines B079 and B259 was evaluated by allelic expression at the SNP. Undifferentiated (U) and enriched differentiated (D) cells were tested and the levels are indicated by arbitrary units. (C) Relative allelic expression levels from Q-SNaPshot assays. The ratio between U and D for each line is indicated. The expected allelic expression ratios corresponding to different levels of transgene inactive X expression (%esc, indicated for 0%, 25%, 50%, 75%, and 100% of Xa levels) are shown for comparison to observed results. Experiments were performed in triplicate and standard deviations are indicated. Data were compared to expected values using two-tailed t-test. As results indicated some variation in Xa expression between the two transgenic lines, predicted values were computed separately for each line. (D) The formula to calculate expected values (Exp) that correspond to percentage transgene escape levels (%esc) using the Xa expression level of the 129 allele (L₁₂₉), the CAST allele (L₀⁰⁰), and the transgene (L₁₂₉). Some images were published in [316].
### Table A

<table>
<thead>
<tr>
<th></th>
<th>SNP</th>
<th>129</th>
<th>CAST</th>
<th>Transgene</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs29297026</td>
<td></td>
<td>A</td>
<td>T</td>
<td>A</td>
</tr>
</tbody>
</table>

### Table B

<table>
<thead>
<tr>
<th></th>
<th>ES</th>
<th>B079</th>
<th>B259</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table C

<table>
<thead>
<tr>
<th></th>
<th>U</th>
<th>D</th>
<th>%esc</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B079</td>
<td>0.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B259</td>
<td>0.46</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Equation D

For non-transgenic cells:

\[
\text{Exp} = \frac{L_{129} \times (17\% + 83\% \times 25\% + 83\% \times 75\% \times \%esc)}{L_{\text{CAST}} \times (17\% + 83\% \times 75\% + 83\% \times 25\% \times \%esc)}
\]

For transgenic lines:

\[
\text{Exp} = \frac{L_{129} \times (17\% + 83\% \times 25\%) + L_{\text{Tg}} \times (17\% + 83\% \times 25\% + 83\% \times 75\% \times \%esc)}{L_{\text{CAST}} \times (17\% + 83\% \times 75\%)}
\]
*Tspyl2* is active in both lines prior to XCI and expressed at levels similar to endogenous loci. Also, these data imply that transgene integration does not influence endogenous expression.

The Xi expression pattern of transgenic *Tspyl2* can be assessed in fully differentiated cells with the transgene on the Xi. However, single differentiated cells did not proliferate in our tissue culture conditions. The differentiated cells used in FISH experiments were enriched by plating cells harvested from EBs for a short period of time [338]. Although the cell populations retained a small percentage of undifferentiated cells, the proportion was quantitated by FISH and could be discounted from ratios. To characterize the mixed cell populations, sequential RNA and DNA FISH to detect *Xist* RNA and X chromosome pericentric repetitive DNA was performed to determine the percentage of undifferentiated cells in population and to ensure the population had not lost an X. In both non-transgenic and transgenic lines, nearly all cells retain two X chromosomes and most (83%) have undergone XCI. The “enriched” differentiated cell populations were used to determine the allelic expression ratios by the Q-SNaPshot assay and these ratios were compared to the expected values corresponding to expected transgene escape levels. The expected values were calculated from the undifferentiated cell population (17%) and the differentiated cells (83%) normalized for the frequency that the 129 X chromosome is active (25%) or inactive (75%) (Figure 4.3D).

The effectiveness of this approach was initially tested in non-transgenic ES cells. The observed allelic expression ratio of the 129 X relative to the CAST X was not statistically different from the expected value for X-inactivated genes using two-tailed *t*-test (Figure 4.3C). Thus, the level of 129 expression reflects the transcription from undifferentiated cells and from the Xa in 25% of differentiated cells, but not from the Xi. This result is consistent with previous conclusions that endogenous *Tspyl2* is subject to XCI ([196] and Figure 3.11) and thus validates this approach to infer the XCI status of X-linked genes.

The allelic expression ratio was next measured in the two transgenic cell lines. Importantly, the ratio of enriched differentiated cells compared to undifferentiated cells was highly similar to the ratio seen for the non-transgenic line (~0.46, Figure 4.3C), implying the transgene had the same XCI status as the endogenous 129 allele. Further, comparison of the observed to expected values revealed that, by two-tailed *t*-test, the transgene expression levels were not statistically different from the level of X-inactivated genes (Figure 4.3C). Therefore,
we conclude that transgenic *Tspyl2* in lines B079 and B259 is subject to XCI.

The Q-SNaPshot assay also provided useful information on transgene copy number. The allelic ratio of the genomic DNA from both transgenic lines increased ~1.9 times compared to non-transgenic cells (Figure 4.3B), suggesting that each of the transgenic lines contains one copy of the *Tspyl2* transgene. These data support our conclusion that the transgenic lines contain single-copy BACs.

### 4.2.3 Allelic expression of transgenic *Jarid1c*

As transgenic *Iqsec2* and *Tspyl2* are X-inactivated like the endogenous loci, the question remains whether *Jarid1c* in the same BAC transgene is also subject to XCI, or whether like the endogenous loci, transgenic *Jarid1c* escapes XCI. To address this question, the *Jarid1c* transgene in lines B079 and B259 was analyzed in the same manner as *Tspyl2*, by comparing allelic expression in non-transgenic and transgenic cell lines. A *Jarid1c* polymorphism distinguishing the transgene and the 129 allele from the CAST allele was used in the Q-SNaPshot assay (Figure 4.4A). As seen for *Tspyl2*, comparison of undifferentiated transgenic and non-transgenic cells suggested that, prior to XCI, transgenic *Jarid1c* was expressed at levels that are equal to or somewhat exceed either endogenous allele (Figure 4.4B and C). These data verify the results obtained from RNA FISH experiments that the *Jarid1c* transgene is active prior to differentiation (section 3.2.7).

The allelic expression ratio was then examined in enriched differentiated cells of the non-transgenic ES line. The ratio of 129 to CAST *Jarid1c* expression was significantly higher than for *Tspyl2* in two-tailed student *t*-test (Figure 4.3C and 4.4C), consistent with *Jarid1c* escaping XCI. However, this ratio was not close to that in undifferentiated cells where both alleles are fully active, suggesting reduced expression from the Xi (Figure 4.4C). The Xi expression level was calculated to be 40% of the Xa level (Figure 4.4C and D). Our results are consistent with previous reports that *Jarid1c* partially escapes XCI with Xi expression levels at 25-50% of Xa levels depending on the cell lineage and differentiation stage [174, 208]. Importantly, the escape level is ~40% in 12.5 dpc embryonic cells, which share a similar developmental stage to the differentiated ES cells used in our studies [174]. Hence, we demonstrate again that this approach is feasible to test the XCI status of genes.

The allelic expression ratio was examined in lines B079 and B259 after differentiation.
Figure 4.4 X inactivation status of transgenic \textit{Jarid1c} examined by Q-SNaPshot. \textbf{(A)} \textit{Jarid1c} SNP differentiates the 129 allele and transgene from the CAST allele. The color of the nucleotides indicates the fluorescently labeled di-deoxynucleotides used in Q-SNaPshot. \textbf{(B)} \textit{Jarid1c} expression in non-transgenic ES cells (ES) and transgenic lines B079 and B259 was evaluated by allelic expression using the SNP. Undifferentiated (U) and enriched differentiated (D) cells were tested and the levels are indicated by arbitrary units. \textbf{(C)} Relative allelic expression levels from Q-SNaPshot assays. The ratio between U and D for each line is indicated. The expected allelic expression ratios corresponding to different levels of transgene inactive X expression (\%esc, indicated for 0\%, 25\%, 50\%, 75\%, and 100\% of Xa levels) are shown for comparison to observed results. Experiments were performed in triplicate and standard deviations are indicated. Data were compared to expected values using two-tailed $t$-test. As results indicated some variation in Xa expression between the two transgenic lines, predicted values were computed separately for each line. \textbf{(D)} The formula to calculate expected values (Exp) that correspond to percentage transgene escape levels (\%esc) using the Xa expression level of the 129 allele ($L_{129}$), the CAST allele ($L_{CAST}$) and the transgene ($L_{Tg}$). Some images were published in [316].
A

<table>
<thead>
<tr>
<th>SNP</th>
<th>129</th>
<th>CAST</th>
<th>Transgene</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs29296190</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th></th>
<th>ES</th>
<th>B079</th>
<th>B259</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
<td><img src="image3" alt="Graph" /></td>
</tr>
<tr>
<td>D</td>
<td><img src="image4" alt="Graph" /></td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
<tr>
<td>cDNA RT-</td>
<td><img src="image7" alt="Graph" /></td>
<td><img src="image8" alt="Graph" /></td>
<td><img src="image9" alt="Graph" /></td>
</tr>
<tr>
<td>cDNA RT+</td>
<td><img src="image10" alt="Graph" /></td>
<td><img src="image11" alt="Graph" /></td>
<td><img src="image12" alt="Graph" /></td>
</tr>
<tr>
<td>gDNA</td>
<td><img src="image13" alt="Graph" /></td>
<td><img src="image14" alt="Graph" /></td>
<td><img src="image15" alt="Graph" /></td>
</tr>
</tbody>
</table>

C

For nontransgenic cells:

\[
\text{Exp} = \frac{L_{129} \times (17\% + 83\% \times 25\% + 83\% \times 75\% \times \%\text{esc})}{L_{\text{CAST}} \times (17\% + 83\% \times 75\% + 83\% \times 25\% \times \%\text{esc})}
\]

For transgenic lines:

\[
\text{Exp} = \frac{L_{129} \times (17\% + 83\% \times 25\% + 83\% \times 75\% \times 40\%) + L_{\text{Tg}} \times (17\% + 83\% \times 25\% + 83\% \times 75\% \times \%\text{esc})}{L_{\text{CAST}} \times (17\% + 83\% \times 75\% + 83\% \times 25\% \times 40\%)}
\]
The ratios were about double that of non-transgenic cells (Figure 4.4C), implying the transgene and endogenous alleles were transcribed similarly. Importantly, the ratios between enriched differentiated and undifferentiated cells were highly similar to the non-transgenic line (~0.71, Figure 4.4C), supporting the conclusion that transgenic and endogenous Jarid1c are expressed similarly on the Xi. Statistical analysis also established that transgenic Jarid1c Xi expression levels were significantly different from that of X-inactivated genes or the Tspyl2 transgenes (Figure 4.4C and 4.3C). Observed levels were calculated to be 36% for line B079 and 37% for line B259 (Figure 4.4C and D). These data suggest that transgenic Jarid1c escapes XCI, at a level comparable to the endogenous gene.

4.2.4 Jarid1c transgene expression in single differentiated ES cells

To confirm and extend the results obtained from the analysis of mixed cell populations, Jarid1c transgene expression was directly examined in individual differentiated cells by sequential RNA and DNA FISH. Interphase nuclei from ten-day differentiated cells were first hybridized with probes to detect nascent Jarid1c RNA transcripts and the Xist domain that marks the Xi. Following signal fixation, cells were denatured and hybridized with a BAC DNA probe to identify all endogenous and transgenic Jarid1c genomic loci.

Prior to analyzing the Jarid1c transgene, endogenous Jarid1c was investigated in non-transgenic cells. Using FISH, nearly all cells (≥98%) showed Jarid1c RNA signals from the Xa (Figure 4.5 patterns 1 and 2), indicating Xa expression is fully detectable. Importantly, ≥85% of cells showed Jarid1c nascent transcripts from the Xi (Figure 4.5 pattern 1). The ability to detect Xi expression in vast majority of differentiated cells by FISH allows us to use this approach to determine XCI status of the Jarid1c transgene. Jarid1c RNA signals were absent on the Xi in a small proportion of cells (≤13%, Figure 4.5 pattern 2), either reflecting inactivation in those cells or, perhaps more likely, reduced hybridization efficiency due to decreased Jarid1c expression from the Xi relative to from the Xa.

Two control experiments were performed to assess the possibility that Jarid1c probes detected the genomic DNA fragments instead of RNA transcripts in our FISH experiments. First, differentiated cells were treated with or without RNaseA before sequential RNA and DNA FISH. RNaseA treatment abolished both Xist and Jarid1c RNA signals but maintained BAC DNA foci (Figure 4.6A), demonstrating the RNA specificity of Jarid1c probes. RNA
Figure 4.5 X inactivation status of endogenous *Jarid1c* by sequential RNA and DNA FISH. FISH was performed in non-transgenic ES cells to analyze *Jarid1c* RNA, *Xist* RNA and BAC DNA. The percentage of cells showing each hybridization pattern was compared to expectations for a gene that is subject to or escapes XCI. The statistical significance (P-value) of comparison to an inactivated gene by Chi square statistic is indicated. Results indicate that *Jarid1c* escapes XCI in a large percentage of cells. Figure was published in [316].
<table>
<thead>
<tr>
<th>Results</th>
<th>1) Xα + Xι expression</th>
<th>2) Xα only expression</th>
<th>3) Xι only expression</th>
<th>FISH signals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Xιst RNA</td>
<td>BAC DNA</td>
<td>Jarid1c RNA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Expected &amp; Observed</th>
<th>% cells</th>
<th>Conclusion</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inactivated</strong></td>
<td></td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td><strong>Escape</strong></td>
<td></td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

| **Observation** | **Experiment 1** | 89 | 10 | 1 | Escape | < 0.01 |
|                | **Experiment 2** | 85 | 13 | 2 | Escape | < 0.01 |
Figure 4.6 FISH control experiments. (A) Differentiated cells were treated without (-) or with (+) RNaseA before sequential RNA and DNA FISH. Interphase nuclei (pseudocolored gray) were hybridized to Jarid1c probes (green) that detected Jarid1c nascent transcripts, a Xist probe (blue) and BAC DNA (red). Merged images are shown for cells from line B259 as an example. All RNA and DNA signals were present without RNaseA treatment, while RNaseA treatment abolished both Xist and Jarid1c RNA signals but maintained BAC DNA foci. (B) BAC (RP23-362P12) DNA was labeled and used to detect the mouse Ube1x nascent transcripts in differentiated ES cells using RNA FISH. As expected for X-inactivated genes, Ube1x expression was detected from the Xa but not from the Xi. Images were published in [316].
FISH of an additional X-inactivated gene, *Ube1*, detected Xa transcripts only (Figure 4.6B), further confirming the hybridization conditions were indeed specific for RNA transcripts. Moreover, the *Iqsec2* FISH experiments also identified the Xa expression only (Figure 4.1) and validate the FISH method to detect RNA molecules. Thus, the FISH results truly represent the transcription activity.

With the hybridization conditions established using non-transgenic cells, transgenic lines were tested by sequential RNA and DNA FISH. Similar to transgenic *Iqsec2*, the *Jarid1c* transgene was expressed from the Xa in nearly all cells (data not shown). Cells were scored using criteria that ensured only diploid nuclei with optimal hybridization were analyzed. Specifically, only *Xist* domain positive cells with three clear BAC DNA foci and *Jarid1c* RNA transcripts from all Xa loci and at least one Xi locus were scored. For all four transgenic lines, nascent *Jarid1c* RNA transcripts were identified at each of the three DNA foci (two endogenous and one transgene) in ≥93% of cells (Figure 4.7B pattern 1 and 3), suggesting each *Jarid1c* transgene was expressed whether present on the Xa or Xi. Furthermore, the cells that carried the transgene on the Xi clearly showed two *Jarid1c* RNA signals colocalizing to the *Xist* domain, implying expression from both endogenous and transgenic *Jarid1c* when located on the Xi (Figure 4.7A). Importantly, the proportion of these cells closely matched the percentage of cells that inactivated transgenic X chromosomes (Figure 4.7B pattern 3). One *Jarid1c* RNA signal was absent on the Xi in a subset of cells from each line (≤7%, Figure 4.7B pattern 2), consistent with the observation in non-transgenic cells. Nonetheless, statistical analysis demonstrated that the pattern distribution for each line was different from that of X-inactivated transgenes (Figure 4.7B). These results clearly demonstrate that transgenic *Jarid1c* in the four lines escapes XCI in all, or at least the vast majority of, cells. Results are summarized in Table 4.1.

### 4.2.5 Candidate sequences in *Jarid1c* BACs and integration sites

To assess the correlation between local sequence features and transgene XCI status, repeat composition of the *Jarid1c* BACs and transgene integration sites were investigated. The genomic context of a gene may play an important role in its Xi regulation since XCI status has been correlated with the content of repeats such as L1s, MIRs, Alu/B1s and LTRs [151, 155, 227, 228, 230]. Specifically, L1s, MIRs and LTRs are enriched in X-inactivated
Figure 4.7 Sequential RNA and DNA FISH analysis of *Jarid1c* transgene expression. (A) Representative nuclei (pseudocolored grey) with the transgene on the Xi. Sequential RNA and DNA FISH was performed using probes to detect *Jarid1c* RNA (green), transgene DNA (red) and *Xist* RNA (blue). (B) Summary of FISH results. To ensure that only diploid cells with optimal hybridization signals were considered, scored cells showed one *Xist* RNA signal, three *Jarid1c* DNA foci (one transgenic and two endogenous), and *Jarid1c* transcripts from all Xa and at least one Xi. The percentage of cells showing each hybridization pattern was compared to expectations for a transgene that is subject to or escapes XCI. The statistical significance (P-value) of comparison to an inactivated gene by Chi square statistic is indicated. Images and tables were published in [316].
### Results

#### Expected & Observed

<table>
<thead>
<tr>
<th>1) Tg on Xa</th>
<th>2) Tg on X1</th>
<th>3) Tg on X1</th>
<th>Conclusion</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tg on Cast X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactivated</td>
<td>~75</td>
<td>~25</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Escape</td>
<td>~75</td>
<td>0</td>
<td>~25</td>
<td></td>
</tr>
<tr>
<td>Observation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A314</td>
<td>71</td>
<td>6</td>
<td>23</td>
<td>Escape</td>
</tr>
<tr>
<td>B202</td>
<td>78</td>
<td>7</td>
<td>15</td>
<td>Escape</td>
</tr>
<tr>
<td>Tg on 129 X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactivated</td>
<td>~25</td>
<td>~75</td>
<td>0</td>
<td>Escape</td>
</tr>
<tr>
<td>Escape</td>
<td>~25</td>
<td>0</td>
<td>~75</td>
<td></td>
</tr>
<tr>
<td>Observation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B079</td>
<td>23</td>
<td>4</td>
<td>73</td>
<td>Escape</td>
</tr>
<tr>
<td>B259</td>
<td>27</td>
<td>6</td>
<td>68</td>
<td>Escape</td>
</tr>
</tbody>
</table>
Table 4.1 Chromosomal location and X inactivation status of genes near transgene integration sites.

<table>
<thead>
<tr>
<th>Transgene Line</th>
<th>Transgenic X Strain</th>
<th>BAC Genes Assayed</th>
<th>Adjacent Genes Assayed*</th>
<th>Location on X (Mb) †</th>
<th>XCI Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>B079</td>
<td>129</td>
<td>36.67</td>
<td>AK139935</td>
<td>35.83</td>
<td>Inactivated</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Cul4b</td>
<td>35.93</td>
<td>Inactivated</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Jarid1c</strong></td>
<td></td>
<td><strong>Escape</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tspyl2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Gria3</td>
<td>38.75</td>
<td>Inactivated</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Birc4</td>
<td>39.45</td>
<td>Inactivated</td>
</tr>
<tr>
<td>B202</td>
<td>CAST</td>
<td>88.19</td>
<td>EG547215</td>
<td>87.31</td>
<td>Inactivated</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Jarid1c</strong></td>
<td></td>
<td><strong>Escape</strong></td>
</tr>
<tr>
<td>B259</td>
<td>129</td>
<td>131.95</td>
<td>Zmat1</td>
<td>131.54</td>
<td>Inactivated</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Jarid1c</strong></td>
<td></td>
<td><strong>Escape</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tspyl2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Gprasp1</td>
<td>132.28</td>
<td>Inactivated</td>
</tr>
<tr>
<td>A314</td>
<td>CAST</td>
<td>166.36‡</td>
<td>Mid1</td>
<td>166.12</td>
<td>Inactivated</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Jarid1c</strong></td>
<td></td>
<td><strong>Escape</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Igsec2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The closest single-copy annotated genes with expressed SNPs that were expressed in fibroblasts were tested. Using these criteria, very few transcripts were excluded except in the B259 line. At least six transcripts near this gene-rich integration site showed tissue-restricted expression or were multi-copy. †Map locations are from the UCSC genome browser (http://genome.ucsc.edu/), build 37, July 2007 assembly. ‡A314 integrated into exon 2 of the Mid1 gene. Table was published in [316].
regions, while Alus/B1s tends to associate with genes that escape XCI. We compared levels of these repetitive sequences between the BACs, integration site sequences in windows of 100 kb, 200 kb and 500 kb, the mouse X chromosome and the entire mouse genome. Results are similar for the windows of different sizes and an example of 200 kb is shown in Figure 4.8.

The two BACs as well as the 112 kb overlap region of the two BACs have a unique composition of repetitive sequences [196], indicating a distinctive sequence feature of the Jarid1c genomic locus. For instance, L1s and LTRs are depleted whereas Alu/B1s are enriched relative to the whole X and mouse genome (Figure 4.8), consistent with other escape genes [155, 227, 228, 230]. The only exception is MIRs. They are enriched in the Jarid1c BACs but not human escape genes (Figure 4.8), and thus their correlation with XCI status needs to be further investigated.

Notably, similar repeat distribution is not seen at the transgene integration sites. Instead, repeat composition at these sites is quite distinct (Figure 4.8). First, the content of L1s, MIRs and LTRs does not correlate with transgene XCI status, with the levels higher, equal or lower when compared to the X and genome. Alu/B1s show a somewhat skewed distribution, with lower levels at three sites and a similar level at one site compared to the average of the X. However, Alu/B1 sequences were found enriched, instead of depleted, in human escape regions [228]. If human and mouse share the same correlation, Alu/B1 content cannot explain why transgenic Jarid1c escapes XCI. Thus, we did not observe any correlation between repeat levels of the integration sites and transgene XCI status.

Another candidate regulator of escape is the insulator protein CTCF, as binding sites have been mapped to three boundaries between genes with different XCI status [220]. A plausible role for the CTCF sites in escape gene regulation is to isolate active genes from surrounding inactive heterochromatin [220]. This model suggests that escape genes reside in the chromatin domains flanked by CTCF binding sites [220, 339]. It has been shown that sequences at the promoter and 5’-UTR of Jarid1c bind CTCF and act as a CTCF-dependent DNA insulator [220]. A prediction of this model is that additional sites downstream of Jarid1c will be present either in the BAC transgenes or integration sites.

Recent whole-genome mapping and computational studies of CTCF binding sites in human have reexamined and refined the consensus sequences for CTCF binding [340, 341]. As CTCF is evolutionarily conserved among higher eukaryotes, [342], these consensus
Figure 4.8 Repetitive element composition of the endogenous *Jarid1c* locus and transgene integration sites. The repeat content of the mouse genome, X chromosome, two BACs, a 112 kb overlap region of the two BACs and a 200 kb region surrounding each integration site with repeat masker data from the UCSC genome browser (http://genome.ucsc.edu/), build 37 (July, 2007 assembly), are analyzed. Repeat levels are shown in comparison to the total mouse genome. Images were published in [316].
these consensus motifs, sequences in the intervening region between *Jarid1c* and downstream
sequences have been extended to predict sites in other genomes such as mouse [343]. Using these consensus motifs, sequences in the intervening region between Jarid1c and downstream X-inactivated transcripts were analyzed by the prediction program at the CTCFBSDB database (http://insulatordb.utmem.edu/). This algorithm incorporates four motifs derived from the two independent studies [340, 341], and computationally predicts several CTCF sites with different levels of likelihood. The best prediction using each motif is showed in Figure 4.9. The predicted CTCF sites at the integration sites are showed in Figure 4.10.
Figure 4.9 CTCF sites within the \textit{Jarid1c} genomic region and the intervening area between \textit{Jarid1c} and adjacent X-inactivated transcripts. Experimentally identified sites are indicated at the 5’ end of \textit{Jarid1c}. Predicted sites within \textit{Jarid1c} and the boundary between \textit{Jarid1c} and the X-inactivated transcript \textit{AK013346} were obtained from CTCFBSDB. This algorithm incorporates four motifs derived from two independent studies. The top site for each consensus motif is shown, including two that overlap. Image was published in [316].
Figure 4.10 X inactivation status and CTCF sites at transgene integration sites. XCI status of genes are shown in yellow (X-inactivated) and blue (escape). Experimentally identified CTCF sites (filled triangles) are indicated at the 5’ end of Jarid1c. Predicted sites (open triangles) at integration sites were obtained from CTCFBSD. This algorithm incorporates four motifs derived from two independent studies. The top site for each consensus motif is shown. The transgene in line A314 is located within the Mid1 gene and downstream pseudoautosomal sequences are not available for CTCF site prediction. The distance to the integration site is shown in parenthesis.
A314

(191 kb)  (128 kb)  Mid1

Jarid1c  Igsec2  Neo

B079

Alp1b4  (969 kb)  Lamp2  (850 kb)  AK139935  (831 kb)  Cul4b  (730 kb)  Gria3  (2.1 Mb)  Birc4  (2.8 Mb)

Tspyl2  Gpr173

Jarid1c

B202

EG547215  (295 kb)  (325 kb)

Tspyl2  Gpr173

Jarid1c

B259

Zmat1  (386 kb)  Gprasp1  (331 kb)  (563 kb)  (716 kb)

Tspyl2  Gpr173

Jarid1c

△ CTCF-Experimentally verified

△ CTCF-Predicted
4.3 DISCUSSION

To better understand the mechanisms governing escape from XCI, studies presented in this chapter characterized the Xi expression pattern of mouse Jarid1c at multiple ectopic loci on the X. An understanding of Xi regulation of the Jarid1c transgene is critical to define an escape domain and identify the regulatory elements necessary for escape. In these studies, we tested four mouse XX ES cell lines that were established in the previous chapter and carry an intact, single-copy Jarid1c BAC transgene on one X chromosome and to determine whether transgenic Jarid1c still escapes XCI.

4.3.1 Characterization of transgene expression upon X inactivation

4.3.1.1 Expression from active X chromosomes

Transgene Xa expression status is important to evaluate gene silencing events that may occur during ES cell differentiation but are unrelated to XCI. Most genes are expressed in undifferentiated ES cells as chromatin structures are open at the pluripotent state [344]. During early embryonic development and ES cell differentiation, a number of loci are silenced and chromatin state may substantially change [344]. This alteration can affect not only endogenous genes but also randomly integrated transgenes. However, Xa expression of transgenic Iqsec2 and Jarid1c were observed in nearly all cases (Figure 4.1 and 4.7), indicating the transgenes were not silenced by developmental regulation during ES cell differentiation. Hence, Xi expression status of the transgenes directly reflects their XCI status.

4.3.1.2 X inactivation status of transgenic Jarid1c

We have demonstrated that transgenic Jarid1c escapes XCI in all four cell lines, whereas two additional transcripts in BAC transgenes, Iqsec2 and Tspyl2, are properly X-inactivated. Intriguingly, the two Jarid1c transgenes tested (B079 and B259) appear to escape at levels similar to endogenous Jarid1c. Thus, the Jarid1c BAC transgenes recapitulate the endogenous locus with respect to XCI status. As transgene DNA integrates onto the 129 X in lines B079 and B259 and onto the CAST X in lines A314 and B202, Jarid1c transgene Xi expression does not appear to be influenced by strain or parental origin of transgenic X chromosomes. Transgene Xi expression is also independent of transcription orientation, as
transgenic \textit{Jarid1c} is transcribed towards the centromere in line A314 while in the reverse direction in the remaining three lines.

4.3.2 Mechanisms of escape from X inactivation

4.3.2.1 Transgene DNA size and escape of transgenic \textit{Jarid1c}

As mentioned earlier, the only transgene previously reported that clearly escapes random XCI is the chicken transferrin gene embedded in a large DNA fragment (187 kb) [257]. Nearly all smaller transgenes are X-inactivated (see section 1.1.4). It has been proposed that the large size of transgene DNA may cause escape from XCI, possibly by buffering the influence from adjacent X-inactivated regions [178]. However, this hypothesis was questioned by the finding that the two \textit{Myo7a} BAC transgenes on the mouse X showed an expression pattern consistent with X-inactivated genes [266]. Our studies rule out the possibility that large DNA fragments block transgenes from endogenous inactivated regions and cause transgenic \textit{Jarid1c} to escape XCI. While the \textit{Jarid1c} transgene escapes XCI in all of the four cell lines, the adjacent \textit{Iqsec2} and \textit{Tspyl2} transgenes undergo XCI. All these transgene transcripts are approximately 30-40 kb away from the end of BAC integrants. Therefore, transgenes that are located within large DNA fragments and quite distant to endogenous sequences do not always escape XCI.

4.3.2.2 BAC vector sequences and escape of transgenic \textit{Jarid1c}

In our transgenic cell lines, BAC vector sequences remain in the genome. Vector sequences have been reported to influence gene expression at integration sites, although the ability is dependent on the sequences that have integrated and the distance from the regulatory regions of a gene [333, 345]. The BAC vector region is approximately 10 kb in length, significantly smaller than the genomic insert, and includes several prokaryotic genes and functional elements (Figure 2.1). As previous X-linked transgenes studies and experiments presented in my thesis demonstrate both transgenes that are subject to XCI and escape XCI (see section 1.1.4), we can conclude that vector sequences do not strongly interfere with transgene \textit{Xi} expression status. Alternatively, if they have any strong effect, they are more likely to promote gene silencing as the majority of X-linked transgenes are X-inactivated. Hence, \textit{Xi} expression of transgenic \textit{Jarid1c} is not due to the BAC vector sequences.
Although a eukaryotic recombinant gene, the neomycin selective marker (neo), is present within the BAC vector region, it may not influence *Jarid1c* transgene Xi expression either. In our studies, transgene expression was examined in ES cells that were differentiated in the absence of G418. Upon XCI, there was no requirement for Xi expression of the *neo* gene. Transgenic cell lines were also tested for *neo* transcription activity after culturing for several passages. Substantial cell death upon reintroduction of G418, despite retention of two Xs in a high percentage of cells, argues that *neo* is not constitutively expressed in all cells and unlikely to influence transgene expression. In mammalian cells, *neo* is driven by a *Pgk1* promoter instead of sequences with strong promoter activities. While *neo* is adjacent to the BAC genomic insert, it is at least 30 kb distant from the closest transgene transcripts. In line A314, the X-inactivated *Iqsec2* transgene separates the *Jarid1c* transgene and *neo*. Escape from XCI, at least in this line, is not due to *neo* transcription. Moreover, a number of X-linked transgenes containing *neo* were X-inactivated [263, 265, 266]. Overall, it is unlikely that BAC vector sequences or genes within this region cause Xi expression of transgenic *Jarid1c*.

### 4.3.2.3 Chromosome location and escape of transgenic *Jarid1c*

Xi expression of transgenic *Jarid1c* is not determined by the XCI status of integration sites. All integration site genes that were testable are X-inactivated (Figure 3.11), suggesting all transgenes have integrated into normally X-inactivated regions, in contrast to the XCI status of transgenic *Jarid1c*. Notably, A314 transgene is inserted into the genomic region of the X-inactivated gene *Mid1* and is valuable to conclude that X-inactivated regions do not prevent the *Jarid1c* transgene from escaping inactivation. A314 transgene is on the CAST X chromosome and organization of this X chromosome has not been well studied. Assuming the CAST X shares a structure similar to the C57BL/6J X where *Mid1* is adjacent to the PAR, one can speculate that transgenic *Jarid1c* in this line escapes XCI because of the proximity to an active region. However, the X-inactivated *Iqsec2* transgene is located between the *Jarid1c* transgene and the PAR. Therefore, it is unlikely that the PAR skips over a gene to render *Jarid1c* resistant to XCI.

In fact, only the A314 transgene is certainly located within an X-inactivated region. The other three BAC transgenes are not located at or immediately adjacent to the genes tested, with the distance ranging from a hundred kilobases to one megabases. It is possible that the
transgenes inserted into small escape regions that were not identified by our studies, but the likelihood that all three transgenes reside within escape regions is quite low. As mentioned earlier, seven mouse genes have been reported to escape XCI [183]. Notably, none of these is close to the Jarid1c transgenes. The mouse escape genes were identified among ~40 genes tested, with the percentage about one sixth [183]. If this percentage is representative of the entire mouse X, the likelihood that all three transgenes are located at escape loci is about 0.5% and the possibility for any two transgenes is less than 3%. Importantly, XCI may be more complete in mouse (see section 1.2.3) and these likelihoods may be even lower. Hence, it is most likely that all or most transgene integrants are located within X-inactivated regions, and therefore the Jarid1c transgenes escape XCI not due to integration into permissive locations. More importantly, the Iqsec2 and Tspyl2 transgenes that are included within the same BAC integrants as the Jarid1c transgene are X-inactivated, indicating the transgenes do not reside at permissive locations or they are not affected by active integration sites. Either of these two possibilities supports the conclusion that Xi expression of transgenic Jarid1c is not influenced by integration sites.

The Jarid1c transgene escapes XCI at four random loci on the mouse X chromosome. If mouse genes escape XCI at the probability of one sixth, the likelihood that four random genes all escape XCI is less than 0.1%. Therefore, Xi expression of transgenic Jarid1c is not a statistically random event. Moreover, each of the four integration sites represents a unique genomic context indicated by distinctive cytogenetic banding patterns and repeat composition. X-inactivated environment at the integration sites does not prevent transgenic Jarid1c from escaping XCI. Combining these results, we conclude that Jarid1c transgene Xi expression is independent of chromosome location. Hence, escape from XCI is an intrinsic property of the Jarid1c locus. The BAC transgenes exhibit the XCI status identical to their endogenous alleles, suggesting they are properly regulated. These results have significance for understanding escape gene regulation.

4.3.2.4 Delineation of the Jarid1c escape domain

Our data suggest the BAC transgenes contain all cis-regulatory sequences necessary for escaping XCI. Thus, the sequence included in each BAC is adequate to direct Xi expression. Although this may not be surprising because the BAC transgenes are quite large, these studies
help functionally define a mouse escape domain. The transgenes used in these studies were generated from two 175 and 186 kb BACs that overlap to share 112 kb. Therefore, this 112 kb region of overlap defines the maximum Jarid1c escape domain. Alternatively, the domain could extend to include all sequences of either individual BAC if overall sequence composition is essential for escape. Thus, the amount of genomic sequence that could influence Xi expression can be delimited to at most the size of the smallest BAC used here, 175 kb. To our knowledge, these studies are the first to delineate an escape domain derived from an endogenous X-linked locus and give important insight into mechanisms of XCI.

If the BAC overlap region includes an escape domain, what could the domain look like? This region contains the entire Jarid1c genomic sequence and at least part of X-inactivated transcripts at both 5’ and 3’ ends (Figure 3.1). An escape domain likely extends to the immediately adjacent X-inactivated genes. Using this definition, the escape domain for mouse Jarid1c is approximately 70 kb, including the Jarid1c genomic region (~41 kb) and intergenic sequences at both sides (~8 kb and ~21 kb, respectively). The ultimate definition of a complete escape domain can be addressed through deletion analysis aided by the transgene system established in my thesis work (see section 5.4.2).

It should be noted that the mouse Jarid1c region contains only one escape gene and may not be regulated by sequences responsible for chromosomal domain formation. However, autonomous Xi expression of our Jarid1c BAC transgenes suggest they include cis-regulatory sequences sufficient for escape and therefore are of great help to dissect for important sequences. Whether mouse Jarid1c is regulated by domain mechanisms can be tested by inserting a reporter gene adjacent to Jarid1c and monitoring its XCI status.

4.3.2.5 Primary DNA sequences and escape from X inactivation

Our transgene studies also demonstrate that escape from XCI at the Jarid1c locus is determined at the level of primary DNA sequences. As described in the introductory chapter, both escape and X-inactivated genes/regions are associated with a number of epigenetic features that are important to maintain their XCI status. However, the BAC DNA utilized to generate the transgenes was isolated from organisms lacking eukaryotic DNA modification systems and chromatin associated proteins. These naked DNA fragments were able to establish appropriate Xi expression patterns when inserted onto the mouse X. This provides
the first, direct evidence that DNA sequence alone, in a chromosome position-independent manner, is sufficient to direct Xi gene expression and presumably establish important epigenetic modifications.

What transgene sequences drive proper Xi gene regulation? While X-inactivated and escape genes are distinguished by chromatin differences including DNA methylation and histone modifications [51, 346], much less is understood about the initial events to demarcate the active and inactive transcripts. In the case of Jarid1c, reports differ as to whether this gene escapes from the onset of XCI or is initially inactivated [96, 209], although the promoter remains unmethylated throughout the onset and establishment of XCI [220]. It is unclear what initially differentiates the Jarid1c chromatin domain from others and why its promoter behaves differently from other CpG rich sequences on the Xi that will become heavily methylated and transcriptionally inactive. Importantly, our studies are the first step in locating the sequences controlling escape from XCI, and we now demonstrate that all sequences necessary for escape are included in the BACs and the escape domain defined here. Our studies can also help evaluate the X inactivation model in which escape genes are regulated by long-range gene regulation mechanisms, such as repeat sequence composition and/or boundary elements [51, 196, 219, 227].

(i) **Repetitive sequences**

We found that repeat sequence features at transgene integration sites do not correlate with the XCI status of transgenic Jarid1c. Thus, they do not seem to influence transgene XCI status, or at minimum do not have a major impact. Notably, the two Jarid1c BACs, as well as the 112 kb overlap region, had a unique repeat composition similar to human escape regions. The L1 and LTR content of these BACs is dramatically lower than the average of the X chromosome and the genome, while the Alu/B1 content is much higher (Figure 4.8). If sequence features play a role in Xi gene regulation, they may contribute to establish an active domain that is autonomous on the Xi, allowing the Jarid1c transgene to escape XCI regardless its specific location. It should be noted that our sequence analysis is preliminary at best. More thorough analysis to survey various sequence motifs such as oligonucleotides may reveal better correlation of repetitive elements or other key regulatory sequences.

If the sequence composition of the BAC DNA is sufficient for escape from XCI, why are the Iqsec2 and Tspyl2 transgenes still X-inactivated? It is likely that Iqsec2 and Tspyl2 are not
associated with special sequence organization, but the composition in the *Jarid1c* genomic regions is fairly remarkable. Thus, although the whole BACs still show overall distinctive sequence features, transgenic *Iqsec2* and *Tspyl2* are still X-inactivated. If this is the case, it suggests that the sequence composition of an escape domain does not influence adjacent X-inactivated genes, even when they are quite close.

It is difficult to determine the boundary of escape domains if they are established by special sequence composition. The repeat distribution is generally calculated by sliding-window analyses, which compute repeat content in chromosomal intervals of relatively large sizes (e.g. 50 kb) with a short slide (e.g. 5 kb) [196, 227]. As a result, repeat levels gradually change at the intervening regions between X-inactivated and escape genes. It is unknown what repeat level serves as the threshold of differential XCI status. Additionally, the observed level at a certain site reflects the level in a sequence range, instead of the specific location. A solution to this problem may rely on the identification of the core region sufficient for escape by deletion analysis of our BAC transgenes.

**(ii) CTCF binding sites**

The CTCF sites at the 5’ end of *Jarid1c* are present in all of the four BAC transgenes, consistent with the idea that they are critical for *Jarid1c* escaping XCI. However, the CTCF sites downstream of *Jarid1c* have not been identified. Although a putative site has been suggested [220], it is located between two X-inactivated transcripts, EST *BB854232* and *Tspyl2*, contradictory to the role as boundary elements. Additionally, this site is absent in A314 transgene and thus, even if present, is not necessary for escaping inactivation. Using the computational programs established in other studies [343], we predict several CTCF sites downstream of *Jarid1c*, between this gene and the X-inactivated EST *AK013346*. As these sites reside within the BAC transgenes, they may be important for transgenic *Jarid1c* to escape XCI by establishing an escape domain with the CTCF sites at the 5’ end of *Jarid1c*. Although predicted CTCF sites also exist at transgene integration sites, X-inactivated genes are located between these sites and experimentally established sites, which conflicts with the boundary model (Figure 4.10). Additional experimentation, aided by the transgene system reported here, will be useful to determine whether the putative binding sites in the BACs or other long-range sites are important for escape gene regulation.
4.3.3 Inactive X expression of endogenous \textit{Jarid1c}

The ability of our system to examine expression in single cells also help evaluate the controversy as for the Xi expression level of \textit{Jarid1c}. Lingenfelter \textit{et al.} performed RT-PCR analysis in single embryonic cells and did not detect bi-allelic expression of \textit{Jarid1c} in some cells [209]. A recent report using FISH was able to observe \textit{Jarid1c} Xi transcripts in only about 50% of differentiated ES cells [96]. As Xi expression levels are significantly lower than Xa levels in tissues or cells tested [174, 208], authors concluded that \textit{Jarid1c} was expressed from the Xi only in some cells and levels were similar to those from Xa alleles [209]. However, neither of the studies quantified \textit{Jarid1c} allelic expression on the basis of single or clonal cells. Non-uniform XCI may be due to multiple cell lineages tested. Some studies have demonstrated that \textit{Jarid1c} escapes in all single-cell subclones of certain adult tissues, at decreased levels from the Xi in comparison to the Xa [174, 208]. In addition, although the Xi expression patterns and levels of another escape gene, human \textit{RPE1}, also vary among different cell lines, it escapes XCI in all subclones of certain cell lines, supporting the idea that the XCI status is stable during somatic cell division [173].

Our results are in apparent contrast to the previous FISH study. Using non-transgenic cells, our sequential RNA and DNA FISH method identified the Xi expression of endogenous \textit{Jarid1c} in a high percentage (≥85%) of cells. Perhaps the discrepancy between the two FISH studies indicates that low transcript levels are difficult to detect in all cells using particular FISH probes or hybridization conditions. Alternatively, the differences may reflect the difference between cell lines as previous study utilized the mouse ES line LF2 that contains two homozygous X chromosomes from chinchilla [347]. It has not been determined whether X chromosomes from other rodents are X-inactivated faithfully in mouse cells. Additionally, a high proportion of cells in this line (84%) are present as tetraploids [347], suggesting the genomic stability is compromised. In contrast, our cell lines carry two mouse X chromosomes and have been demonstrated to maintain a diploid karyotype and stable genetic background.

The high efficiency in detecting \textit{Jarid1c} Xi expression not only facilitates the analysis of transgene activity on the Xi, but also suggests that \textit{Jarid1c} escapes XCI with the Xi allele expressed at lower levels than the Xa allele. In our FISH experiments, \textit{Jarid1c} RNA signals were absent on the Xi in approximately 10% of cells. Importantly, the Xi expression of \textit{Jarid1c} in at least the majority of differentiated cells, combined with the result obtained from
populations of differentiated cells that Xi expression levels are about 40% of Xa levels, argues that *Jarid1c* escapes XCI in majority of or all cells but the Xi level in each cell is reduced relative to the Xa level. Moreover, we noticed that the intensity of *Jarid1c* RNA FISH signals on the Xi was weaker than that on the Xa, supporting a partial escape in each cell as opposed to escaping at a full level in a proportion of cells. However, the signal intensity was not quantified. Definitive evidence of whether and how much the Xi level is lower than the Xa level needs further investigation.
CHAPTER V
GENERAL DISCUSSION AND FUTURE DIRECTIONS

5.1 OVERVIEW

In order to functionally define a chromosomal domain that contains all information necessary for escape from XCI, we asked whether an escape gene is autonomous or instead takes on properties of the region on the X chromosome in which it is located. Specifically, we sought to determine whether an escape gene, in the context of adjacent genomic sequences, would retain Xi expression upon relocation on the X. We proposed that Jarid1c BAC transgenes would contain an autonomous escape domain and therefore (1) the Jarid1c transgene would escape XCI at multiple ectopic locations; (2) adjacent transgenes would be X-inactivated.

Four ES cell lines were generated that carry an intact or largely intact, single-copy BAC transgene at different loci on the X. Upon ES cell differentiation, the Jarid1c transgene escapes XCI in all four lines, whereas the other two transgene transcripts, Igsec2 and Tspyl2, are X-inactivated. Therefore, the Jarid1c BAC transgenes recapitulate the endogenous locus with respect to XCI status. The large size of transgene DNA and BAC vector sequences are unlikely to cause transgenic Jarid1c to escape XCI. X-inactivated environment and diverse chromosomal context at the randomly chosen loci do not prevent transgenic Jarid1c from escaping XCI. Hence, Jarid1c transgene Xi expression is independent of chromosome location and escape from XCI is an intrinsic property of the Jarid1c locus.

By establishing that Jarid1c Xi expression is autonomous and chromosome-position independent, we have narrowed sequences necessary for Jarid1c expression and functionally defined a Xi escape domain. As transgenic Jarid1c always escapes XCI, the BAC transgenes contain all regulatory information important for Xi expression. Escape from XCI at the Jarid1c locus is determined at the level of primary DNA sequences since naked DNA fragments are sufficient to establish appropriate Xi expression status. The 112 kb overlap region of the two BACs used defines the maximum Jarid1c escape domain, or the domain could extend to include all sequences of either individual BAC.

As the first functional analysis to establish an escape domain, these studies provide significant insight into how genes are regulated on the Xi. The transgene approach described
in my thesis also established a system to functionally test the effect of genomic sequences on Xi expression and as such has broad applications for understanding gene regulation on the Xi. Future work using truncated or modified transgenes can address the role of specific sequences in escape gene regulation. These and proposed studies should aid our understanding of long-range control of gene expression and the impact of genomic sequences.

5.2 MODEL OF ESCAPE FROM X CHROMOSOME INACTIVATION

Recent studies have suggested that three-dimensional architecture of the Xi influences gene expression status [96, 99, 348]. Chromatin organization is important not only for DNA packaging and genome replication but also for transcription regulation [349]. In the case of XCI, transcriptional status of a gene is found to be related to its subnuclear positioning [96, 99]. Using a high-resolution three-dimensional analysis, the Xi allele of the X-inactivated gene \textit{ANT2} was found to be located inside of Xi chromosomal territory, while the Xa allele of \textit{ANT2}, as well as both Xi and Xa alleles of the escape gene \textit{ANT3}, showed a peripheral location [348].

The spatial regulation of Xi expression status may be accomplished by nuclear compartmentalization of \textit{Xist} RNA during development. Upon XCI, the X chromosome that is chosen to be inactivated undergoes a dynamic reorganization. \textit{Xist} RNA starts to accumulate on one X at onset of XCI and triggers the inactivation of this chromosome [51]. From then on, \textit{Xist} RNA is localized to the chromosome territory of the Xi and induces a heteropycnotic configuration that is cytologically distinctive with Giemsa or DAPI staining. X chromosome sequences may not contribute to the unique nuclear distribution of \textit{Xist} RNA since DNase treatment does not abolish the \textit{Xist} RNA localization [350]. A matrix protein, scaffold attachment factor A, is enriched on the Xi through its RNA binding activity and DNase treatment has no influence on its nuclear architecture either [350]. Thus, \textit{Xist} RNA may interact with nuclear matrix to form a stable compartment. This nuclear compartment is an impenetrable structure to transcriptional machinery, depleted of RNA Polymerase II and other transcription factors, and is also associated with the silencing of repetitive sequences on the X [99, 350]. Upon XCI, some genes shift from outside to inside of this nuclear structure and are subsequently inactivated, while \textit{Jarid1c} displayed a more exterior location [96]. Therefore, \textit{Xist} RNA likely forms a repressive nuclear compartment around the Xi that initiates and/or
maintains the silent status of X-inactivated genes [96]. Genes shifting into the Xist RNA territory are inactivated, whereas those located at the periphery or outside of this area escape XCI, perhaps enabling interaction with transcription machinery. In our FISH studies, both endogenous and transgenic Jarid1c signals were frequently seen at the periphery of the Xist RNA-marked Xi domain (Figure 4.6), consistent with the idea that the peripheral positioning is critical for escape gene expression. Yet more critical analysis, to specifically analyze their location relative to other inactivated genes, will be necessary to confirm this observation.

The nuclear localization of X-linked genes may be controlled by local genomic context and specific regulatory elements as both have been found to correlate with XCI status [220, 227, 228]. However, neither repeat sequence distribution nor boundary elements alone can explain how genes can be expressed from Xi. As discussed previously, none of the repeat elements that have been tested on human and mouse X chromosomes correlate with Xi expression status perfectly [220, 227, 228] (also see section 1.3.2.1). DNA insulators with CTCF binding activities or sequences that interact with nuclear matrix, such as MARs, are not sufficient to establish an active region on the Xi [178, 246]. However, these factors are not mutually exclusive and may coordinately determine the nuclear localization of X-linked genes.

How could this cooperation mediate the transcriptional activity of different regions along the Xi? The whole mammalian genome is believed to be organized into distinct areas by forming functional loop domains [351]. The loop structures topologically separate active regions from surrounding inactive chromatin and form self-contained functional domains [351]. The Jarid1c escape domain defined in our studies has unique sequence composition and is potentially flanked by CTCF sites. These two features may cooperate to determine the peripheral location of Jarid1c by looping out of the Xist RNA territory and allow this gene to escape XCI (Figure 5.1). CTCF may aid this function by tethering chromatin loops to the nuclear matrix or a specific compartment [96, 244]. Other sequences, such as repeat elements, may facilitate exterior localization of the Jarid1c gene. It has been shown that repetitive sequences are able to form higher-order structures on chromosomes, which may influence nucleosome positioning and subsequent heterochromatin formation [352]. More specifically, transcriptionally silenced interspersed repeats, such as L1s, MIRs and LTRs, are the most
Figure 5.1 Model of escape from X inactivation.
internalized sequences within the Xi territory [96, 99]. The Jarid1c region is depleted of these repeats and thus is located outside of the Xi territory. Likewise, Alu-rich sequences may also promote the outer localization since they largely correlate with escape from XCI on human X and at mouse Jarid1c. Thus, X-inactivated and escape genes reside in separate chromatin loops that are interior or exterior of the Xist territory, respectively, and are regulated according to their nuclear localization. It should be noted that all escape genes may not share the same sequence organization or regulatory elements. Other escape domains may have distinct repeat content or different boundary elements, such as MARs [350] and GATA binding sites [230]. Whether or not a specific control element is present may depend on the individual gene and the particular region of the X chromosome in which that gene resides. The X-linked transgenes in many studies were inactivated likely because they integrated into regions that were inactivated, contained mammalian sequences that served to further propagate inactivation, or lacked appropriate sequences that would allow them to escape X inactivation.

5.3 X-LINKED TRANSGENE SYSTEM

Besides a use in characterizing the mouse Jarid1c locus, our transgenic ES cell lines also establish a transgene system to investigate the Xi regulation of any gene or sequences of interest. The single-copy BAC insert in lines B079, B202 and B259 is flanked by one loxP and one lox511 site. Because different lox sites do not recombine with each other, the BAC insert can be replaced by any sequence flanked by the same lox sites through a cre-mediated double crossover event, named recombinase mediated cassette exchange (RMCE) (Figure 5.2) [349]. Thus, we have created a series of randomly tagged sites on the mouse X chromosome. These sites are permissive for transgene expression from active X chromosomes and thus transgenes are less likely subjected to position effects. Although these sites are inactivated upon XCI, transgenes can either be inactivated or expressed. As the precise location of these sites has been determined, local sequence features have been preliminarily investigated and can be analyzed more thoroughly. Hence, sequences of interest can be rapidly inserted into well-characterized locations on the X, and Xi regulation of other genes or function of candidate regulatory sequences can be quickly tested at multiple locations with known XCI status. These ES cell lines can also be used for mouse transgenesis to study gene function and regulation by the transgene approach.
Figure 5.2 Principle of recombinase mediated cassette exchange (RMCE). A DNA fragment flanked by heterospecific lox sites (loxP and lox511) to prevent self excision is previously inserted into a chromosomal locus. Cre recombinase recognizes the lox sites at the tagged locus and those in targeting vectors and catalyzes recombination between the same types of lox sites. This cre-mediated recombination results in exchange of the sequences at the tagged genomic locus with the sequences in the targeting constructs. Suitable selection strategies permit the recovery of the desired exchange event.
Chromosome locus

\[ \text{lox511} \rightarrow \text{neo} \rightarrow \text{loxP} \]

targeting vector

\[ \text{bsr} \]

** neo: neomycin resistance gene  
bsr: blasticidin resistance gene **
5.4 BROAD APPLICATION OF THE METHODOLOGY

The two approaches that we utilized to assess transgene XCI status are valuable for gene expression analysis. The allelic expression assay, Q-SNaPshot, can not only determine Xi expression patterns, but also quantify Xi expression levels [25]. In fact, due to the ability to measure allelic ratios, this assay can be extended to analyze allelic expression of any gene at the RNA level and allele presence at the DNA level. This assay can directly determine the XCI status of X-linked genes when uniform cell populations that inactivate the same X chromosome are tested, like the non-randomly inactivated cell line B119. Our studies also demonstrated that mixed populations of differentiated and undifferentiated cells were effective to assess Xi expression status. In non-transgenic cells, endogenous Tspyl2 and Jarid1c showed allelic expression ratios consistent with the XCI status established in previous studies, with calculated Xi expression levels fairly close to reported levels (Figure 4.3 and 4.4) [174, 196]. This approach was successfully applied to two additional endogenous genes, the X-inactivated gene Crsp2 and escape gene Ddx3x (data not shown). When transgenic Tspyl2 and Jarid1c were assessed by comparing transgenic and non-transgenic cells, this approach yielded results that were consistent with the sequential RNA and DNA FISH method (Figure 4.3 and 4.4). Therefore, this approach is useful for future Xi expression studies.

Sequential RNA and DNA FISH is a sensitive method to detect transcription from the site of a gene in individual cells and therefore gives definitive results for transcription activity. This single-cell analysis is crucial for analyzing genes that may have different Xi expression patterns between cells, such as mouse Jarid1c ([174, 209], also see section 1.5). Additionally, in transgenic cell lines, endogenous expression provides internal controls for transgene activity. Endogenous Xa alleles are a positive control for gene transcription and hybridization conditions. Transgene Xa expression is important for evaluating gene silencing that may occur during ES cell differentiation but unrelated to XCI. Transgene XCI status can be determined by comparing to endogenous Xi alleles. Therefore, sequential RNA and DNA FISH is an excellent approach to assess gene expression.

5.5 FUTURE DIRECTIONS

From these and other studies, it is clear that regulation on the Xi is complex. It is likely that control of Xi genes occurs at one or multiple levels including genomic sequences,
boundary elements, nuclear localization, and even gene-specific promoters and enhancers. Our transgene studies provide strong evidence for an important conclusion that escape from XCI is an intrinsic feature of mouse Jarid1c. Consequently, we have defined an autonomous escape domain that is instructive for further dissection of key regulators of Xi expression. The tractable transgene systems described here are powerful to directly address local or long-range factors influencing Xi expression as such has broad applications for understanding gene regulation on the Xi. Here, I describe several research projects that follow my thesis work and will generate important results in XCI studies.

5.5.1 Key regulatory sequences at the Jarid1c locus

We have established that the two 175 kb and 186 kb BACs or, more likely, the 112 kb overlap region is sufficient for escape from XCI. However, they are still relatively large and important regulatory sequences remain elusive. In order to narrow the region essential for escape and eventually identify key cis-regulatory sequences, we need to generate smaller constructs to test the hypothesis that these constructs still include all sequences necessary for escape.

Strategy I: In our transgene studies, we identified twelve X-linked BAC transgenes and PCR analysis showed that some of them were partially deleted (Figure 3.6). These truncated transgenes are convenient and useful tools for our deletion analysis. The breakpoints of the deletions can be defined by the Q-SNaPshot assay and the SNPs present throughout the BAC genomic insert as discussed in section 3.3.2. Transgene XCI status can be determined by sequential RNA and DNA FISH or the Q-SNaPshot assay that are described in my thesis. If Jarid1c in the truncated BAC DNA is X-inactivated, the sequences included are not sufficient for escape. Deleted regions contain important sequences and long-range chromosome effects influence XCI status. If any truncated transgenes are able to escape XCI, they may still contain all sequences necessary for escape. This possibility needs to be validated at other chromosomal location. For instance, similar truncations can be generated from original BACs by recombineering, targeted into characterized integration sites by RMCE and tested whether it can escape XCI. We can further narrow down the Jarid1c escape domain only if the Jarid1c transgene always escapes XCI. To understand the role of adjacent sequences, we will identify the integration sites of truncated transgenes, determine the XCI status at these loci and
compare levels of repetitive elements, particularly for transgenes that can be X-inactivated. If candidate sequences are important for inactivation, we expect high frequency at the integration sites of inactive transgenes. Otherwise, their frequency will not correlate with transgene XCI status. Similarly, if any sequences are important for escape, they should be present or enriched specifically at the integration sites that allow escape from XCI.

Strategy II: Intact BAC DNA can be modified by recombineering to generate a series of deletions (such as 20, 40 kb, etc. at either end). Each modified BAC can be knocked into the tagged integration sites to test transgene XCI status. Same as the expectations in strategy I, deleting important sequences will cause the \textit{Jarid1c} transgene to be X-inactivated at one or more loci. As discussed in section 4.3.2.4, a reasonable assumption regarding the \textit{Jarid1c} escape domain (also tested in these experiments) is that it extends at most to the flanking X-inactivated genes. By this definition, this domain is \(\sim 70\) kb including 8 kb upstream and 21 kb downstream of the \textit{Jarid1c} genomic region. This fragment can be produced by modifying the original BACs using recombineering and inserted onto the X chromosome to test whether the 70 kb region is sufficient for escape.

\textbf{5.5.2 CTCF binding sites at the 5’ end of \textit{Jarid1c}}

CTCF sites have been proposed to regulate escape genes, potentially by establishing active chromatin domains [220]. It is also possible that CTCF proteins promote \textit{Jarid1c} Xi expression by gene-specific mechanisms since they can function as transcription activators [243]. As all \textit{Jarid1c} transgenes tested escape XCI and contain two CTCF sites at the promoter and 5’ UTR, it is important to test the hypothesis that these sites are necessary for escape from XCI. CTCF is involved in global transcription regulation and is essential for cell viability [243] such that knockout is lethal. This protein has been also shown to regulate several important components of XCI, such as \textit{Xist} and \textit{Tsix} [103, 239]. Therefore, conditional knockout or knockdown of CTCF will affect many housekeeping pathways including the XCI process and are not useful to understand whether CTCF plays any role in \textit{Jarid1c} Xi expression. We can mutate these CTCF sites in BAC DNA using recombineering by replacing these sequences with synthesized oligonucleotides that have 6-10 nucleotides differences and can abolish CTCF binding activities [220]. Modified BACs can be inserted back into the integration sites by \textit{cre}-mediated RMCE to test the XCI status of transgenic \textit{Jarid1c}. As these
sites are located within the promoter region, the normal function of modified promoters to start transcription needs to be confirmed. Alternatively, a “promoter swap” approach can be used to replace the mouse promoter in BACs with the human syntenic region. The human JARID1C promoter is absence of CTCF sites while the sequences and function are highly conserved [220]. If these CTCF sites are important for escape, we expect to see the Jarid1c transgene is X-inactivated in one or more mutated BAC transgenes. If, in contrast, mutated BACs are still able to be expressed from the Xi at different loci, the CTCF sites are not necessary for escape.
REFERENCES


218. Luoh, S.W., et al., *CpG islands in human ZFX and ZFY and mouse Zfx genes:...


317. Yant, S.R., et al., *High-resolution genome-wide mapping of transposon integration in*
Curriculum Vitae
Nan Li

ACADEMIC ADDRESS:
Department of Biochemistry and Molecular Biology
Penn State University, College of Medicine
500 University Drive
Hershey, PA 17033, USA
Telephone: 717-531-8154
Email: nzl103@psu.edu

HOME ADDRESS:
1262 Peggy Drive
Hummelstown, PA 17036, USA
Telephone: 717-566-4838

CITIZENSHIP:
People's Republic of China

EDUCATION
2001-present Pennsylvania State University College of Medicine, Hershey, PA, USA
Intercollege Graduate Program in Genetics (Laura Carrel, mentor)
Ph.D. in Genetics, Expected August 2009
1998-2001 China Pharmaceutical University, Nanjing, China
M.S. in Biopharmaceutical Sciences (Wutong Wu, Shigan Lin, mentors).
1994-1998 China Pharmaceutical University and Nanjing University, Nanjing, China.
B.S. in Pharmaceutical Biotechnology

HONORS
Karl H. Beyer, Jr., M.D., Ph.D. Scholarship, Pennsylvania State University, 2002
Scholarship for Outstanding Undergraduate Students, China Pharmaceutical University, 1995-1998

PUBLICATIONS

CONFERENCES
2003,2005,2007 Poster, Symposium in Molecular Biology, Penn State University, University Park, PA
2003-2006 Poster, Symposium of Intercollege Program in Genetics, Penn State University, PA
2003-2006 Poster, Graduate Student Research Forum, Penn State College of Medicine, PA

LEADERSHIP
Jun. 2007 Organizing committee, 12th Annual conference of Chinese Biopharmaceutical Association, Rockville, MD
May 2007 PSU coordinator, Annual conference of Sino-American Pharmaceutical Association, Philadelphia, PA