7-16-2015

Insight into the Function of X-linked Lymphocyte Regulated 3 (XLR3) and the Mechanism Regulating its Imprinted Expression

Robert James Foley
University of Connecticut - Storrs, RFoley99@gmail.com

Insight into the Function of *X-linked Lymphocyte Regulated 3* (XLR3) and the Mechanism Regulating its Imprinted Expression

Robert J. Foley

University of Connecticut
2015

The Synaptonemal Complex Protein 3 (SYCP3), a key component of the synaptonemal complex in mammalian meiosis, has many closely related X-linked homologs on the X chromosome of which the functions are unclear. In rodentia, these genes comprise the *Xlr* superfamily of proteins and studies elucidating spatial and temporal profiles of these homologs reveal primarily testis specific function. The following study provides evidence for murine testis and ovary function of a newly discovered *Xlr* superfamily protein, X-linked Lymphocyte Regulated 3 (XLR3). In addition to showing the immunolocalization of the XLR3 protein at the condensed X and Y chromosomes throughout Meiotic Sex Chromosome Inactivation (MSCI), preliminary experiments behind the mechanism controlling expression of the imprinted *Xlr3b/4b/4c* copies reveals a differentially methylated region (DMR) in the adjacent single copy *Factor 8 associated gene A (F8a)* consistent with differential binding of the master regulator of chromatin conformation factor CTCF.

Additionally, this work also aims to provide the groundwork for deletion of XLR3 from meiosis to assess its significance in meiotic drive theory, sex chromosome conflict, and infertility. Two similar *Xlr* superfamily genes, SLX and SLY, whose absence causes near sterility and litter gender bias, have been shown to control gene expression emanating from the sex chromosomes specifically during meiosis. Uncovering the function of these rapidly evolving superfamily proteins in the transmission of the sex chromosomes will add to the growing

evidence for a mammalian sex chromosome conflict model of inheritance involving additional layers of gene regulation.

Insight into the Function of *X-linked Lymphocyte Regulated 3* (XLR3) and the Mechanism Regulating its Imprinted Expression

Robert J. Foley

B.S., University of Delaware 2009

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

At the

University of Connecticut

2015

Approval Page

Doctor of Philosophy Dissertation

Insight into the Function of *X-linked Lymphocyte Regulated 3* (XLR3) and the Mechanism Regulating its Imprinted Expression

Presented by

Robert J. Foley, B.S.

Major Advisor_____
Dr. Michael J. O'Neill

Associate Advisor_____
Dr. Rachel O'Neill

Associate Advisor_____
Dr. Judy Brown

University of Connecticut

2015

Acknowledgements

I would like to thank my family for their loving support throughout my college and graduate school career. I would also like to thank my grandmother and great aunt for their enthusiasm in my endeavors as well as monetary support in tough times.

Thank you, Mike, for allowing me to study with freedom and independence and always being there for critical guidance and insight. My ability to think like a scientist is all thanks to you.

I also want to thank my friends for being my life outside of work.

Thanks to Barbara, Rachel, Judy, Andrew, Charlie, Leighton and all others who aided me in asking relevant questions during the past 6 years. Also thank you to our collaborators at The Jackson Laboratory; Dr. Laura Reinholdt and Anne Czechanski, for generating X-monosomic mESCs and an *Xlr3* shRNA mouse model.

Craig Obergfell, thank you for keeping the labs running smoothly and constantly keeping up with the order list.

Thanks to Seth Kasowitz for teaching me molecular biology and providing me with a unique perspective towards lab work and life in general.

Thanks to the other M. O'Neill lab graduate students; Michael Murphy, Sohaib Qureshi, Glenn Milton, Inga Nesbitt, Nate Jue, Alicia Lui, and Natali Naveh for being patient and informative during our long lab meetings and lab discussions.

Thank you to my undergraduate students; Amy Coffey, Andrew Sadowski, Emily Clancy, Katelyn DeNegre, and Drew Weyman for keeping lab life fun and being flexible with my experiments.

Table of Contents

TABLES AND FIGURES

Chapter 1: Introduction

Overview

Meiosis is a unifying process among sexual reproducing organisms involving the maturation of progenitor germ cells into haploid gametes for reproduction. Meiosis is divided into two parts; Meiosis I marks a dynamic transformation in germ cell DNA where homologous chromosomes replicate, condense, pair, synapse, recombine and segregate, and Meiosis II which provides a second and necessary segregation of the diploid genome into haploid gametes.[1]

One of the most important and best characterized aspects of meiosis occurs during the pairing, synapse, and recombining of DNA (maternal and paternal genomes) to ultimately contribute to variation seen within species. The effort of organizing the DNA is accomplished through a tightly coordinated proteinaceous complex known as the synaptonemal complex (SC) which aligns homologous chromosomes along an axis in preparation for recombination.[2]

Necessary to this scaffold is the synaptonemal complex protein 3 (SYCP3) which is an autosomally located gene that has evolved many X-linked duplicons on the mammalian X.[3] SYCP3 duplicons comprise a gene family that is defined by the presence of the COR1 domain, and are termed *X-linked Lymphocyte Regulated (Xlr)* superfamily genes. Recent studies in mouse reveal that many of these genes play a role in meiosis and the majority have testis-specific expression profiles (SLY, SLX, SLXL1, XLR6/SLX2); however, their functions and relation to SYCP3 remains somewhat unclear.[4–7] Another cluster of X-linked SYCP3 duplicons (*Xlr3, Xlr4,*

and Xlr5) located at XA7.2 were shown to have imprinted expression in neonatal mouse brain.[8] This study aimed to determine protein functionality of the multicopy gene, *Xlr3*, and elucidate its mechanism of imprinting through three specific aims:

1) Determine whether the *Xlr3* gene produces a functional protein and elucidate the XLR3 spatial and temporal profile in mouse through immune-fluorescent labeling.

2) Generate an *Xlr3* knockdown mouse model to determine whether XLR3, like SYCP3, is necessary for meiosis.

3) Identify and characterize the imprinting control region (ICR) regulating imprinted expression at *Xlr3b/4b/4c*.

The following body of work provides evidence for the existence of XLR3 as a highly regulated meiotic protein, which localizes to the inactive XY body in meiotic spermatocytes and to nucleolus organizer regions (NORs) in oocytes. This work also suggests that a hypothetical function of XLR3 may involve, through structural means, global or local activation or repression of the sex chromosomes, implicating XLR3 in meiotic drive theory and sex chromosome conflict. Evidence is also presented of a differentially methylated region (DMR) in a gene adjacent to the imprinted *Xlr* paralogs, termed *F8a*, which displays differential methylation and CTCF occupancy between maternal and paternal alleles that may constitute the Imprinting Control Region (ICR).

Germ Cells in Mammals

In contrast to *Drosophila* and *C. elegans*, where the preformed germplasm gives rise to germ cells, the laboratory mouse, *Mus musculus*, germ cells form through induction by which any cell in the blastocyst can become part of the germ line. At embryonic day 6.25 (E6.25) mouse embryos express the earliest known marker to indicate germ cell development, PR containing domain 1 *(Prdm1)*, which can be detected in only about six cells in the posterior epiblast.[1] WNT3 signaling induces competency of these cells to bone morphogenic protein 4 (BMP4) creating a germ cell "precursor" state. The subsequently formed primordial germ cells (PGCs) must initially travel through the gut of the developing embryo while replicating and resisting differentiating factors from the surrounding region until they reach the developing gonad. Germ cells accomplish stability throughout this migration with high expression of the pluripotency factors *Octamer binding protein 4 (Oct4)*, *Nanog Homeobox (Nanog)*, and *SRY-box containing gene 2 (Sox2)*.

Development of the PGCs into oocytes or spermatogonia is dependent on cues from the gonads, which initially develop as female by default. In the early female gonad (E12) retinoic acid and *Deleted in azoospermia-like (Dazl)* signal *Stimulated by retinoic acid 8 (Stra8)* to initiate meiosis causing the PGCs to develop into oocytes.[9] The male gonad inhibits the effects of retinoic acid with CYP26b1 which is expressed from fetal sertoli cells, and instead male differentiation factors such as *Nanos2* and *Dnmt3l* are expressed. This sexually dimorphic event thrusts female mouse embryos into meiosis at E13 while male mice initiate meiosis much later when CYP26b1 is repressed around 6 days post-partum (dpp). It should be noted that S*tra8*

induction and, ultimately, male/female germ cell commitment, is thought to depend on the ratio of retinoic acid and male differentiating *Fibroblast growth factor 9 (Fgf9).*[10,11]

Meiosis (General)

Meiosis allows for genetic variation in offspring by recombining DNA while accounting for the proper chromosomal ploidy needed for sexual reproduction.[1] In 6 day old male mice, the precursor germ cell state, termed “Type A” spermatogonia, replicate to form a constant supply of Type A spermatogonia as well as “Type B” spermatogonia.[12] Type B spermatogonia develop into primary spermatocytes and undergo chromosomal replication increasing the genome size from 2n to 4n. Chromosomes in the primary spermatocyte then dynamically condense during the first stage of prophase I, allowing them to begin pairing with homologs. Once paired, non-sister chromatids of homologous chromosomes form chiasmata, or attachment points, and begin homologous recombination between maternal and paternal genomes.[2] Chromosomes in prophase I align in a highly coordinated fashion by an essential proteinaceous complex termed the Synaptonemal Complex (SC).[2,13–15] The role of this structure is to mediate chromosomal crossover events during prophase I. After crossover events occur chromosomes remain attached by chiasmata and "relax" allowing a slight increase in transcription. In females, meiosis arrests in prophase I after crossover events until cued to continue division after birth. The last substage of prophase I is indicated by nuclear membrane disintegration along with meiotic spindle formation. Reduction through metaphase I, anaphase I, and telophase I results in breaking of chiasmata and the newly diploid genome is further

reduced to a haploid genome in meiosis II, where reduction occurs similarly to mitosis. The resulting haploid germ cells mature into gametes as ova or mature sperm.

Meiotic Prophase Events and the Synaptonemal Complex

When primary spermatocytes initiate meiosis, DNA replicates during interphase I in order to prepare homologous chromosomes for recombination during prophase I. The initiator protein SPO11, an ancient topoisomerase-like enzyme, induces double-stranded breaks (DSBs) globally which has been shown to be necessary for downstream recombination events to occur. *Spo11*$^{-/-}$ mice arrest at prophase due to a lack in accumulation of DSBs and DNA repair protein RAD51, ultimately abolishing the synapsis of homologs.[16,17] In anticipation of induced DSBs via SPO11, the synaptonemal complex (SC) begins to form as a result of an ataxia telangiectasia and Rad3 related (ATR) phosphorylation cascade event.

Figure 1 displays the complete scaffold of the synaptonemal complex and core elements involved. During the first stage of prophase I, leptotene or *leptonema ("Thin Threads")*, the "Lateral elements (LEs)" are the first SC proteins to form between each parental chromosome and are comprised of SYCP2 and SYCP3 fiber polymers.[2,18] These two proteins have been shown to self-assemble homo- and hetero-dimers with each other to create short fibers *in-vitro*, likely through use of C-terminal coiled coil motifs.[19,20] Chromatin at this stage dynamically condenses down and associates with the LEs in a loop structure that is not yet well understood. SYCP2/SYCP3 loss of function studies have shown meiosis to fail at this stage due to improper assembly of the lateral elements and impairment of chromatin synapsis.[21,22]

Zygonema or zygotene *("Paired Threads")*, the second stage of prophase, marks assembly of the "transverse filaments (TF)" and the "central elements (CE)" which allows the LEs to come together in a zipper-like fashion known as "synapsis." Chromosomes are therefore dynamically aligned with homologs along an axis poised for recombination.

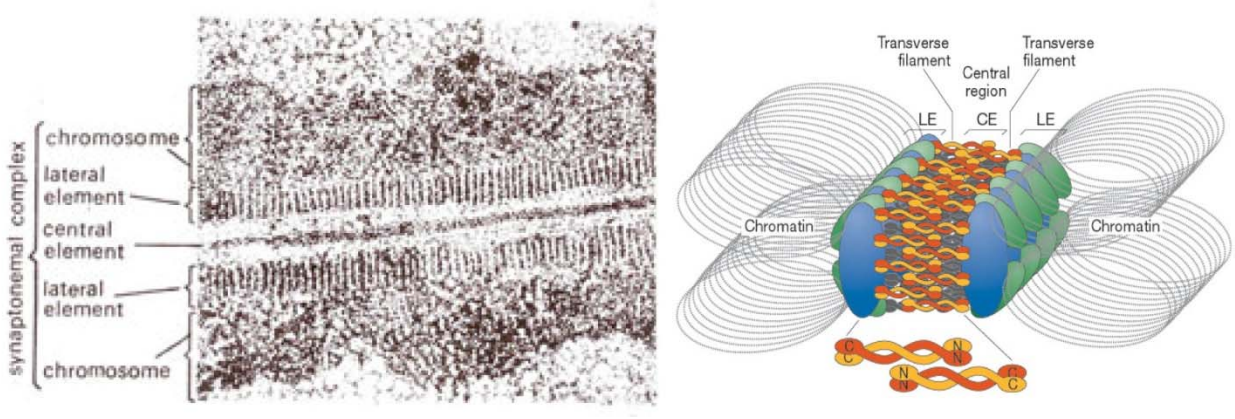


Figure 1. EM and illustration provided by M. Westergaard and D. Von Wettstein of how the synaptonemal complex is structured into three regions relative to chromatin.[23]

Pachynema or pachytene *("Thick Threads")*, marks the longest stage in meiosis lasting three days in mice. At this stage, synapsis of homologous chromosomes through the SC is complete leaving a stable intermediate structure allowing for recombination events to occur (see Figure 2 below). Recombination proteins RAD51 and DMC1 aggregate at ~300 specific regions at the LEs to produce early recombination nodules (ENs) visible through immuno-staining.[2,24] These sites are apparent even before complete SC synapsis, however, they are not thought to be undergoing recombination early in pachytene. The ENs shift to an intermediate state known as transition nodules (TNs) indicated by recruitment of stabilizing and DNA repair proteins RPA, BLM, MSH4 and MSH5. While these proteins mediate a connection between both LEs,

recombination won't occur until MLH1 is recruited. MLH1 is paramount to class I crossover events which only occurs in ~25 of the initial 300 early nodules. In this sense, recombination crossover events occur about once or twice per SC - these mature crossovers are known as late recombination nodules (RN). The TNs that are not destined to crossover are dislocated from each other and associate back with the rest of the SC as non-crossover events. It should be noted that some sites of recombination occur even when MLH1 has not been recruited, and these are referred as to class II crossover events.

Additionally during pachytene in males, the X and Y chromosomes are transcriptionally inactivated and packaged into a separate nuclear compartment known as the “XY body” or "sex-body." This phenomenon is known as meiotic sex chromosome inactivation (MSCI) and is part of a more general mechanism known as meiotic silencing of unsynapsed chromatin (MSUC).[25] The X and Y synapse exclusively at the pseudoautosomal regions (PARs) but lack of homology elsewhere activates MSUC. This is thought to occur to prevent aneuploidies which may result due to the cells’ failure to recognize the unsynapsed XY during meiosis, but additional theories for existence of this mechanism stem back to Fischer (1931) and Rice (1984).[25] They predicted that many of the genes on the sex chromosomes are sexually antagonistic and, due to the hemizygous nature of the X chromosome in males, genes are immediately selected for if they are advantageous to male reproduction. In fact, a disproportionate amount of genes with roles in spermatogenesis are found on the murine X and the Y chromsomes.[26,27] It is thought that, if allowed to express during meiosis, these genes may distort the inheritance of the sex chromosomes and thus a global silencing of both chromosomes is key.[28]

Perhaps the most important factor necessary to development of the XY body is the histone variant H2AX which becomes phosphorylated upon induction of global DSBs by SPO11.[25] This variant is also activated in response to somatic cell DSBs and once phosphorylated it is termed γH2AX. In meiosis, H2AX is activated in two waves, first localizing in response to SPO11 induced DSBs globally in the nucleus and secondly to chromatin of the X and Y during MSCI. In $H2AX^{-/-}$ male mice spermatogenesis fails in pachytene due to failed MSCI, a necessity for proper spermatogenesis.[29] The exact function of this histone modification has yet to be elucidated but it is seen at DNA at class I crossover events and is known to recruit a mediator of DNA damage checkpoint protein MDC1.[30] H2AX is phosphorylated by ATR (ataxia telangiectasia and Rad3 related), a PI3-like kinase similar to ATM (ataxia telangietasia mutant).[16] ATR is surprisingly recruited by BRCA1 and is seen localizing with BRCA1 at the XY body until dephosphorylation of γH2AX in early diplotene. It is therefore no surprise that BRCA1 deficient mutant mice fail to phosphorylate H2AX resulting in meiotic arrest.[25,31,32] Epigenetically speaking, γH2AX, ATR, and BRCA1 are necessary for numerous downstream silencing marks on the sex chromosomes such as H2A ubiquitylation, H3K9me2, and deacetylation of H3 and H4.[25]

While a large amount of evidence supports that the genome in prophase I of meiosis is active and the X and Y are actively silenced, Page et al. (2012) conversely proposed that transcription is low genome-wide and is then activated in the early stages of prophase I except at the inactive X and Y.[33]

The three previously discussed stages of prophase I mark formation and primary function of the SC, which won't serve any post-meiotic function. Below are SYCP3 fluorescently labelled spermatocytes showing formation and later degradation of the SC.

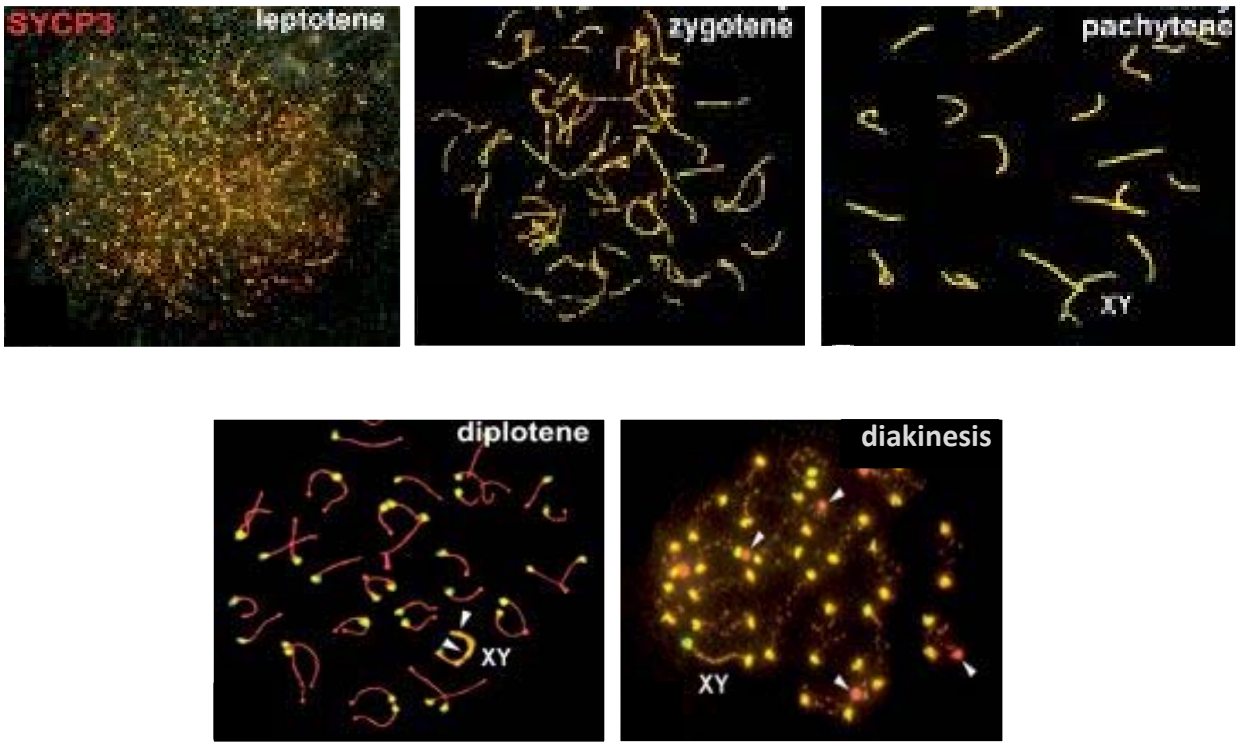


Figure 2. The Synaptonemal Complex via SYCP3 Staining. Fluorescently labelled SYCP3 allows for staging of the SC in primary spermatocytes. Leptotene, Zygotene, and Pachytene represent formation of the SC while Diplotene and Diakinesis mark degradation of the SC. Figure adapted from SC staining by Herr'an, Y. *et al.*[34]

Following completion of pachytene the SC begins to degrade allowing for desynapsis of homologous chromosomes to prepare for segregation. In contrast to male mice, females undergo these initial stages of prophase I during gestation at embryonic day 13 (E13) and arrest in meiosis at diplotene around birth (~E20/P0), also termed *dictyate ("Two Threads")*. This finite amount of halted oocytes will selectively resume meiosis starting within the first month of post-natal mouse maturity. Conversely, male mice will continue meiosis after diplotene and enter

the final stage of prophase I known as diakinesis *("Moving Through")*. At this stage, chromosomes condense further and await segregation via metaphase I. Male mice don't begin meiosis until ~6dpp, but continually cycle this process after the first wave is complete at 18.5dpp.[35] Interestingly, the synaptonemal complex shows a dimorphism between male and female mice, having far less serious phenotypes in SC protein deficient oocytes than spermatocytes.[18]

To date seven basic components of the mammalian SC have been identified. The lateral elements are composed of SYCP2 and SYCP3, the transverse filaments (TF) are exclusively composed of SYCP1, and the central elements are composed of SYCE1, SYCE2, SYCE3 and Tex12.[2] Additional protein components, however, are needed to adhere sister chromatids from S phase through segregation. The proteins responsible to ensure this fidelity are known as Cohesin Complex (CC) proteins and have a structural maintenance of chromosome (SMC) domain which acts in organizing DNA.[36] This complex acts as a clamp around chromatin and involves two core subunits, SMC1 and SMC3, which interact with each other to make up the main "hinge" structure of the complex. The ATPase heads of these proteins are then connected through SCC1, a bridge kleisin protein, to form a stable clamp.[37] This SCC1 component is cleaved during metaphase to allow chromosomal segregation to occur.[34] In addition to mitosis cohesin factors SMC1α, SMC3, and RAD21, there have been four meiosis specific cohesin subunits identified; REC8, RAD21L, SMC1β, and STAG3[18,34]. The cohesin complex has additionally been shown to be necessary for proper SYCP2 and SYCP3 localization and is thought to interact with the lateral elements.[18]

More recently, a group of meiotic "surveillance" factors, termed HORMA (Hop1, Rev7, and Mad2) proteins, were discovered in mouse. These proteins were initially identified in yeast (Hop1) and are thought to act as a meiotic pachytene checkpoint system to ensure the fidelity of chromosomal ploidy. Hop1 is additionally responsible for ensuring that homologs recombine rather than sister chromatids. Fukuda et al. identified HORMAD1 in mice, which immunolocalizes to the SC at points which have yet to synapse.[38–40] This signal is void once synapsis is complete, but re-localizes at the beginning of diplotene to regions of desynapsis. Deficiency in another HORMA protein, HORMAD2, results in lack of ATR accumulation and results in defects related to homolog pairing.[41] It is therefore now clear that HORMA domain proteins monitor and localize to unsynapsed regions during pachytene to prevent mistakes that result in aneuploidies.

The Synaptonemal Complex Protein 3 (SYCP3)

The synaptonemal complex protein 3 (SYCP3) is a 254aa protein in mouse that works by polymerizing to itself and SYCP2 to form the lateral elements of the SC at the beginning of the leptotene stage of meiotic prophase I. In yeast, Hop1 and Red1, serve as lateral element proteins in yeast; however, no structural similarities exist in comparison to mammalian SYCP2/3.[42] SYCP3 is highly conserved in Mammalia, containing a central α-helix domain along with two flanking non-helical domains.[19,20] The α-helix contains many coiled coil motifs, which are thought to promote homotypic interactions. Coiled-coil domain proteins are characterized by rigid and fibrous structures that interact with intermediate filaments (IF) creating a physical cellular network. In fact, SYCP3 looks very similar to IFs whose coiled coil domains functions

through dimerization while two flanking non-helical domains contribute to lateral and longitudinal assembly.[20] While knockout studies are currently underway for these specific domains, male meiocytes deficient in SYCP3 arrest at zygotene/pachytene due to lack of a SC. Interestingly, female meiocytes deficient in SYCP3 were still able to form core units of the SC and remarkably remained somewhat fertile further exemplifying the sexual dimorphism of the SC.[18,22]

The *X-linked Lymphocyte Regulated (Xlr)* Superfamily

SYCP3 is located on chromosome 10 in mouse, but has generated many X-linked duplicons on the mammalian X. In rodents, these duplicons are termed *X-linked lymphocyte regulated (Xlr)* superfamily genes due to the foundation member, *Xlr*, having high expression in B- and T-lymphocytes.[43,44] It was first thought that the majority of these copies were pseudogenes; however, it is now clear that many duplicons code for functional proteins, of which a disproportionate amount have roles in spermatogenesis.[26,44]

The foundational member, *X-linked lymphocyte regulated (Xlr)* was initially discovered in 1985 when Cohen et al. isolated a dozen highly expressed cDNAs from B and T lymphocytes to determine whether any hybridize to the X chromosome. Since many lymphocyte maturation defects map to genes on the X chromosome in other mammals, Cohen proposed that the one cDNA he saw hybridizing to the X chromosome, *X-linked Lymphocyte Regulated* (*Xlr)*, may have a role in immunodeficiency disorders.[45–47] It was not until 2002 that the XLR protein was shown to have a role in mouse oocytes leaving its high expression in B and T lymphocytes a mystery. Since the initial discovery of *Xlr*, hundreds of copies of very similar looking genes have been

identified on the X and Y chromosomes.[44] On the murine X these genes cluster to regions known as *Xlr1 and Xlr2* based on their cytogenetic location.[3] While it is not known if all these *Xlr* superfamily genes code a functional protein, a handful have recently been elucidated in mouse.

Transcription of two multicopy *Xlr* superfamily genes, *Sycp3-like X-linked (slx) and Slx-like1 (slxl1)*, become upregulated in male post-meiotic germ cells and code for proteins which localize to compartments in the cytoplasm of round spermatids.[3,5,35] The Y chromosome counterpart, *Sycp3-like Y-linked (Sly)*, which is upregulated around the same time, also codes a protein that is predominately cytoplasmic in post-meiotic germ cells, but some of which is imported into the nucleus and localizes to the X and Y chromosomes (see Figure 3 below).[48] Abolishing SLY from spermatogenesis reveals de-repression of the X and Y chromosomes at specific genes including *Slx and Slxl1*.[4] In fact, SLX and SLXL1 localize to the X and Y in post-meiotic spermatids in the absence of SLY. This observation is suggestive of antagonistic functions for the SLX and SLY proteins and provides a genetic basis for post-meiotic X and Y chromosomal intergenomic conflict that may point to a general function of all *Xlr* superfamily proteins. This conflict hypothesis is bolstered by the fact that *Sly* deficient mice generate litters with sex ratios favoring females, while *Slx/Slxl1* deficient mice generate litters with sex ratios favoring males. Remarkably, deficiencies in both genes recovers a 1:1 litter sex ratios.[28] Ultimately, it is thought that *Slx* and *Slxl1* are distorters of X chromosome inheritance, acting as selfish elements to increase X chromosome inheritance. *Sly* is thought to be a repressor of distortion, which is exemplified by its repressive role on the *Slx*, *Slxl1*, and the X and Y in general (discussed further in Chapter 3).

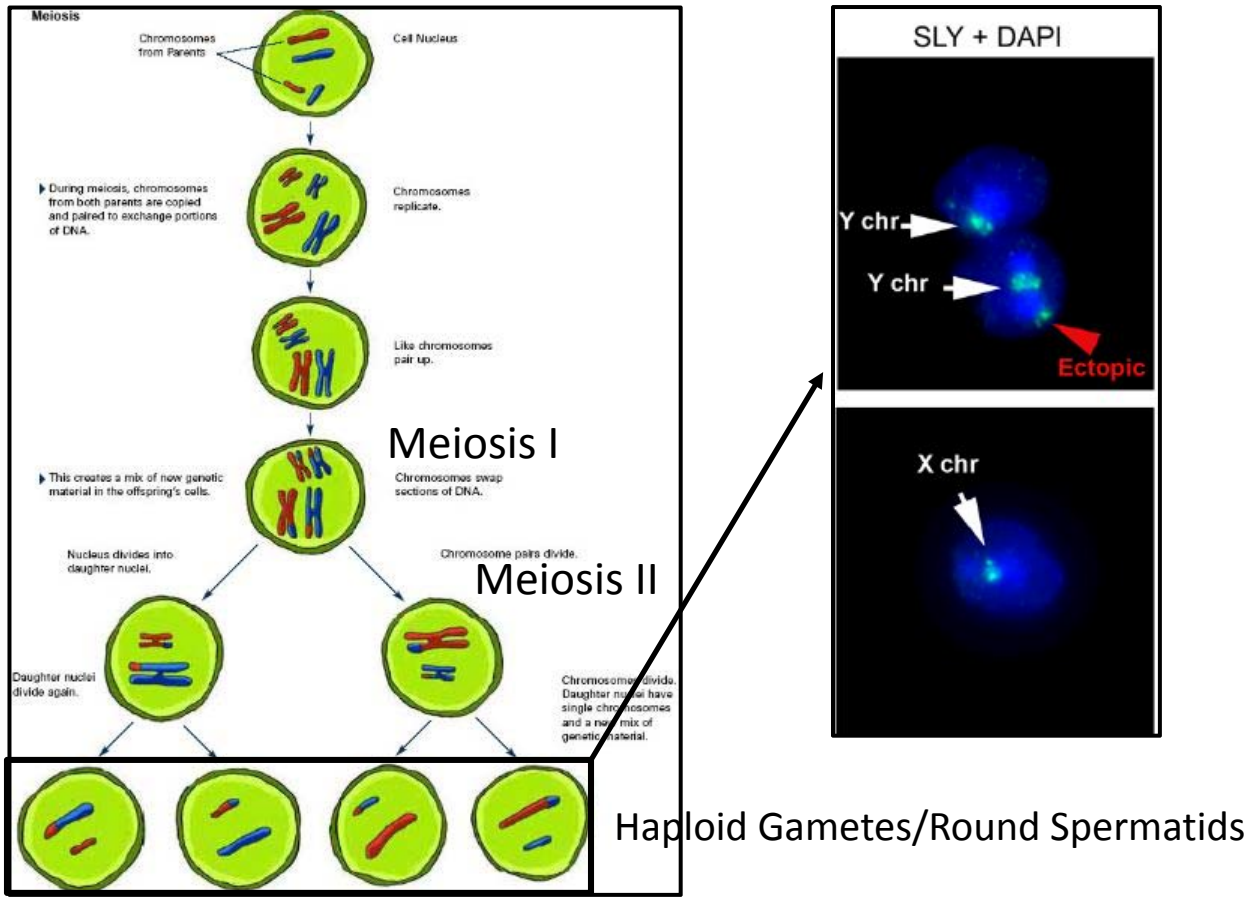


Figure 3. SLY Localization in Post-Meiotic Round Spermatids. Displays the localization of the Xlr Superfamily protein SLY to the X chromosome in X-bearing round spermatids and to the Y chromosome in Y-bearing round spermatids. Figure adapted from NIH's image and video gallery. http://images.nigms.nih.gov/index.cfm?event=viewDetail&imageID=2546

The most recent *Xlr* superfamily protein elucidated, *Xlr6/Slx2* (single copy), has been shown to localize separate from *Slx/Sly* to the inactive XY body in pachytene of male meiosis, the significance of which is not known.[6,7] From a global view, the spatial and temporal profiles of these *Xlr* superfamily genes are very much like their ancestor, SYCP3, in that they are up-regulated in the male germ line almost exclusively; however, each appears to have a slightly different local meiotic profile.

The *X-linked Lymphocyte Regulated (Xlr) Locus*

Other than the XLR1 and XLR2 cytogenetic loci, an additional cluster of *Xlr* superfamily genes were discovered at XA7.2 termed, *Xlr3, Xlr4,* and *Xlr5*. This locus is comprised of fifteen *Xlr* superfamily genes that share extremely high sequence similarity to one another due to the fact that they are the result of *Xlr3*, *Xlr4*, and *Xlr5* duplication events (Figure 4). These duplications have created paralogs which are arranged into triads (a, b, c, d, e), each one having a single copy of *Xlr3*, *Xlr4*, and *Xlr5* (d and e triads not shown).

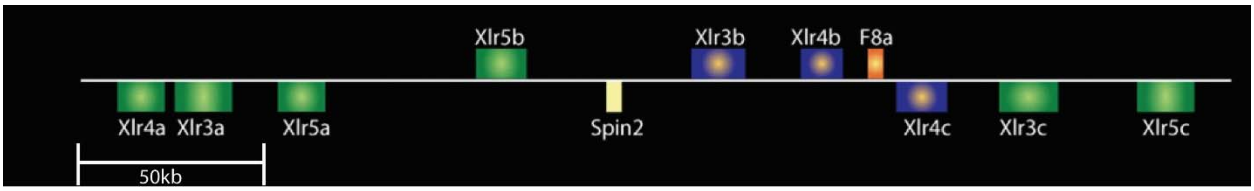


Figure 4. Illustration of the *X-linked Lymphocyte Regulated* (*Xlr)* locus at XA7.2 (mm10).

In 2005, Raefski and O'Neill showed imprinting of some of these genes in neonatal mouse whole brain and fibroblast (C57BL/6J).[8] The imprinted paralogs (shown in blue in Figure 4) cluster within the b and c triads and display maternal specific expression. Interestingly, a biallelically expressed gene *Factor 8-associated gene A (F8a)* sits in between the imprinted paralogs. Studies aimed at elucidating the imprinting mechanism controlling these three genes were carried out in Dr. Michael O'Neill's laboratory at the University of Connecticut by Dr. Michael Murphy, Dr. Seth Kasowitz, Dr. Ben Carone, and Dr. Sohaib Qureshi. These previous studies revealed co-transcriptional stalling of RNA polymerase II exclusively on the paternal allele of the imprinted paralogs. Further experiments aimed at revealing differential enrichment of histone modifications showed allelic differences in H3K36me3, a mark for elongating RNA PolII. This is likely a downstream consequence of the controlling elements regulating this locus.

Further experiments focusing on gene regulation at this locus are discussed in Chapter 5, “Allele Specific Differences at the Imprinted *Xlr3/4/5* Locus”.

Chapter 2: Characterization of the X-linked Lymphocyte Regulated 3 (XLR3) Protein

The existence of protein products and their potential function is currently not known for any of the Xlr3/4/5 genes, including the imprinted paralogs (*Xlr3b, Xlr4b,* and *Xlr4c*). In the "Epigenetics" chapter of this thesis it is revealed that certain autosomal loci are imprinted to maintain protein dosages necessary for proper embryo developmental. In the case of reciprocally imprinted *Igf2/H19* on chromosome 7 in mouse, the expression of a single allele of the potent mitogenic factor *insulin-like growth factor 2 (Igf2)* is absolutely necessary for proper embryo development and therefore the reason for its imprinted expression can be easily understood in light of its protein function. In order to gain an understanding of why the *Xlr3b* copy is imprinted, this chapter aims at identifying its protein product and shedding light on its function. Additionally, elucidating XLR3 protein function may give insight into the underlying evolutionary cause of the massive Xlr Superfamily expansion on the mammalian X chromosome.

XLR3 Has Similarities to SYCP3

Due to high sequence similarities, Xlr3a/b/c were assayed together throughout this study and henceforth will be referred to as Xlr3. The Xlr3 amino acid sequence is 226 amino acids in length harboring a COR1 domain which extends nearly the entire length of the protein categorizing it as a member of the Xlr Superfamily along with other meiotic proteins XLR, SLX/XMR, SLX2/XLR6, SLY and SYCP3. In fact, the XLR3 COR1 peptide has 32% sequence identity

with SYCP3 and at least 12% identity with all other known superfamily members (Figure 5). The exact function of the COR1 domain is not known, though it is thought to interact with other COR1 domain containing proteins. Additionally, the N-terminus of XLR3 contains a canonical monopartite nuclear localization signal (KRKR), which is not well conserved among other superfamily proteins.

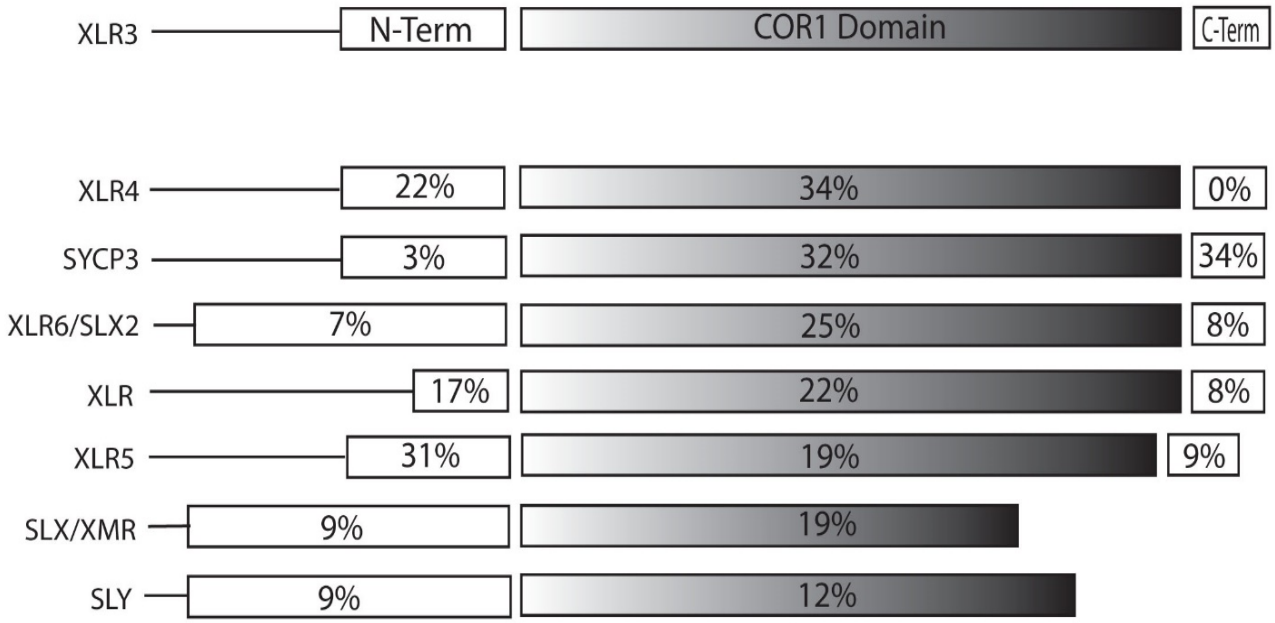
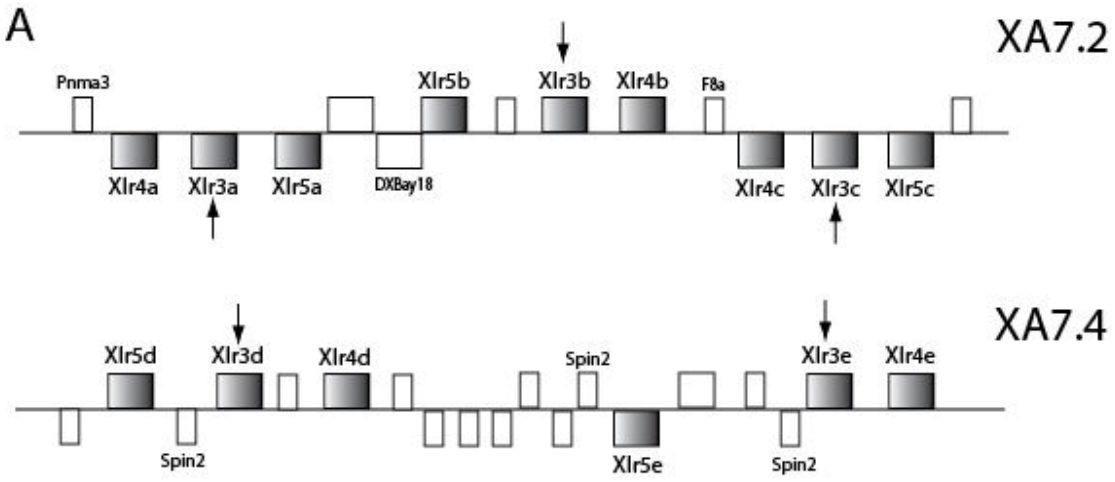


Figure 5. Protein Alignment of the Xlr Superfamily. Amino acid identity from full length Xlr Superfamily proteins broken up into N-terminus, Cor1, and C-terminus domains relative to XLR3.

Xlr3 **Becomes Up-Regulated During Spermatogenesis**

Five *Xlr3* paralogs are found across 2 genetic loci on the mouse X-chromosome (*Xlr3a, Xlr3b, Xlr3c, Xlr3d* and *Xlr3e*) each holding >97% mRNA sequence homology to each other (Figure 6A and B). The latter two paralogs are silent pseudogenes however Xlr3a/b/c retain an open reading frame and all potentially code for the same functional protein.



B

mRNA	Xlr3a	Xlr3b	Xlr3c
Xlr3a	--	0.977	0.974
Xlr3b	0.977	--	0.99
Xlr3c	0.974	0.99	--

Protein	XLR3A	XLR3B	XLR3C
XLR3A	--	0.938	0.938
XLR3B	0.938	--	0.991
XLR3C	0.938	0.991	--

Figure 6. Paralogous nature of *Xlr3.* **A.)** Genomic coordinates and relative positioning of *Xlr3/a/b/c/d/e* paralogs. **B.)** Nucleotide and protein sequence identity matrix displaying high homology among the Xlr3 paralogs.

Xlr3 transcripts were analyzed for relative expression in various tissues via qRT-PCR using primers common to all 3 active paralogs. Consistent with the expression status of other *Xlr* superfamily genes, *Xlr3*, shows high relative expression in mature adult testis (Figure 7A). In mice, the maturation of germ cell subtypes coincide closely with specific time points during prepubertal mouse development. In order to characterize *Xlr3* expression in mature testis, transcripts were additionally analyzed at time points during postnatal male gonad development. Testis sample collections from 6dpp through 18.5dpp encompass the first wave of spermatogenesis during which all *Xlr* superfamily genes become up regulated at various

points. *Xlr3* shows a 4-fold up regulation from 6 days post-partum (dpp) to 18.5dpp, encompassing maturation of the first wave of mature spermatids (Figure 7B). Basal expression of *Xlr3* is apparent in the pre-meiotic stages, however, from 7.5dpp to 10.5dpp expression is doubled indicating potential pre-leptotene/leptotene protein function (~10dpp). This expression status matches more closely to *SYCP3* and *XLR6*/*SLX2* as opposed to SLX and SLY which are up-regulated post-meiotically. To further elucidate this meiotic function an antibody was generated against the Xlr3 amino acid sequence.

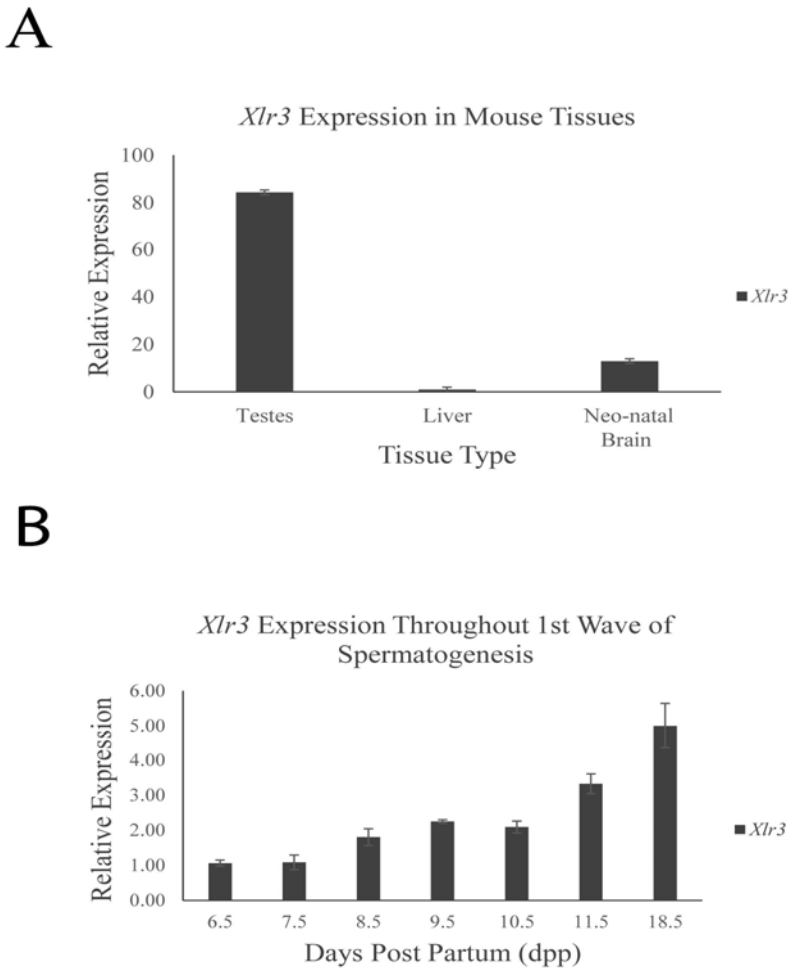


Figure 7. Real-time quantitative RTPCR analysis of *Xlr3*. **A.)** Relative expression of *Xlr3* in various tissues, normalized to *Actb*. **B.)** Relative expression of *Xlr3* throughout the first wave of mature mouse spermatogenesis. Error bars display 95% confidence intervals.

XLR3 is a Nuclear Protein First Present at Leptotene

A custom antibody was raised against the first 18 amino acids of the N-terminus of XLR3A/B/C peptide sequences, which holds no homology to any other Xlr superfamily protein. The anti-XLR3 antibody was used to probe western blots of cytoplasmic and nuclear fractions from adult mouse tissues. A single band consistent with the putative protein size (26.1kDa) is detected exclusively in the nuclear fraction of the testis sample confirming translation of these paralogs (Figure 8A). Next, post-natal testis samples were collected, fractionated, and protein was extracted from time points encompassing the first wave of spermatogenesis (Figure 8B). Western blots probed with anti-XLR3 reveal a band first present in the nuclear fraction at 10.5dpp concurrent with maturation of spermatogonia into pre-leptotene/leptotene stage spermatocytes (meiosis - prophase I). Additional probing with anti-γH2AX, a control for double stranded breaks during leptotene, appears in large amounts at 10.5dpp. In order to visual the localization of XLR3 at leptotene and stages after, immunohistochemistry was conducted on mature adult spermatocytes.

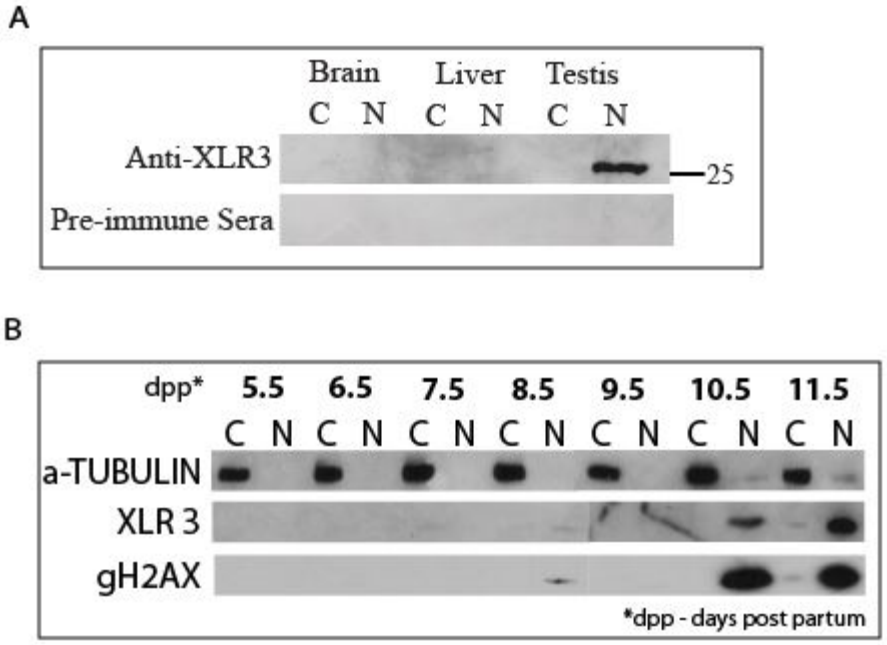


Figure 8. Western blot analysis of XLR3 protein. **A.)** Displays nuclear and cytoplasmic extracts collected from various tissues and probed with anti-XLR3. Positive signal is seen in the nuclear fraction of testis relative to negative control pre-immune sera. **B.)** Displays nuclear and cytoplasmic testis extracts collected early in the first wave of mouse spermatogenesis probed with anti-XLR3. Alpha-tubulin is a control for proper cytoplasm separation while gH2AX confirms proper nuclear separation as well as the beginning of leptotene I of meiosis I.

XLR3 Associates with the XY body During Meiosis I

Fixed meiotic spermatocyte spreads were probed with anti-SYCP3 and anti-XLR3 and coupled with immunofluorescence in order to visually assess XLR3 protein localization at all stages of germ cell maturation (Figure 9). Anti-XLR3 immunolocalizes diffusely at leptotene agreeing with the mRNA expression increase and presence of a protein band first seen at 10.5dpp on the western. A key characteristic of male meiosis is the transcriptional inactivation of the sex chromosomes (XY body) in a separate nuclear compartment throughout the next two stages; zygotene and pachytene, termed Meiotic Sex Chromosome Inactivation (MSCI). XLR3 signal is next detected at the inactivated XY body first at zygotene and lasting throughout

pachytene, confirmed through γH2AX staining, also a marker for XY body formation. In fact, this XLR3 signal is seen closely mirroring SYCP3 at the inactive XY indicating a potential interaction with the Synaptonemal Complex (SC).

A more in-depth look at early, mid, and late pachytene spermatocytes (Figure 10) shows early diffuse XLR3 signal at the XY body becoming more intimately localized with SYCP3 until it is finally seen in late pachytene exclusively at the Synaptonemal Complex of the XY body. This specific localization is new to Xlr superfamily proteins and suggests XLR3 has a transient role in SC formation at the inactive XY.

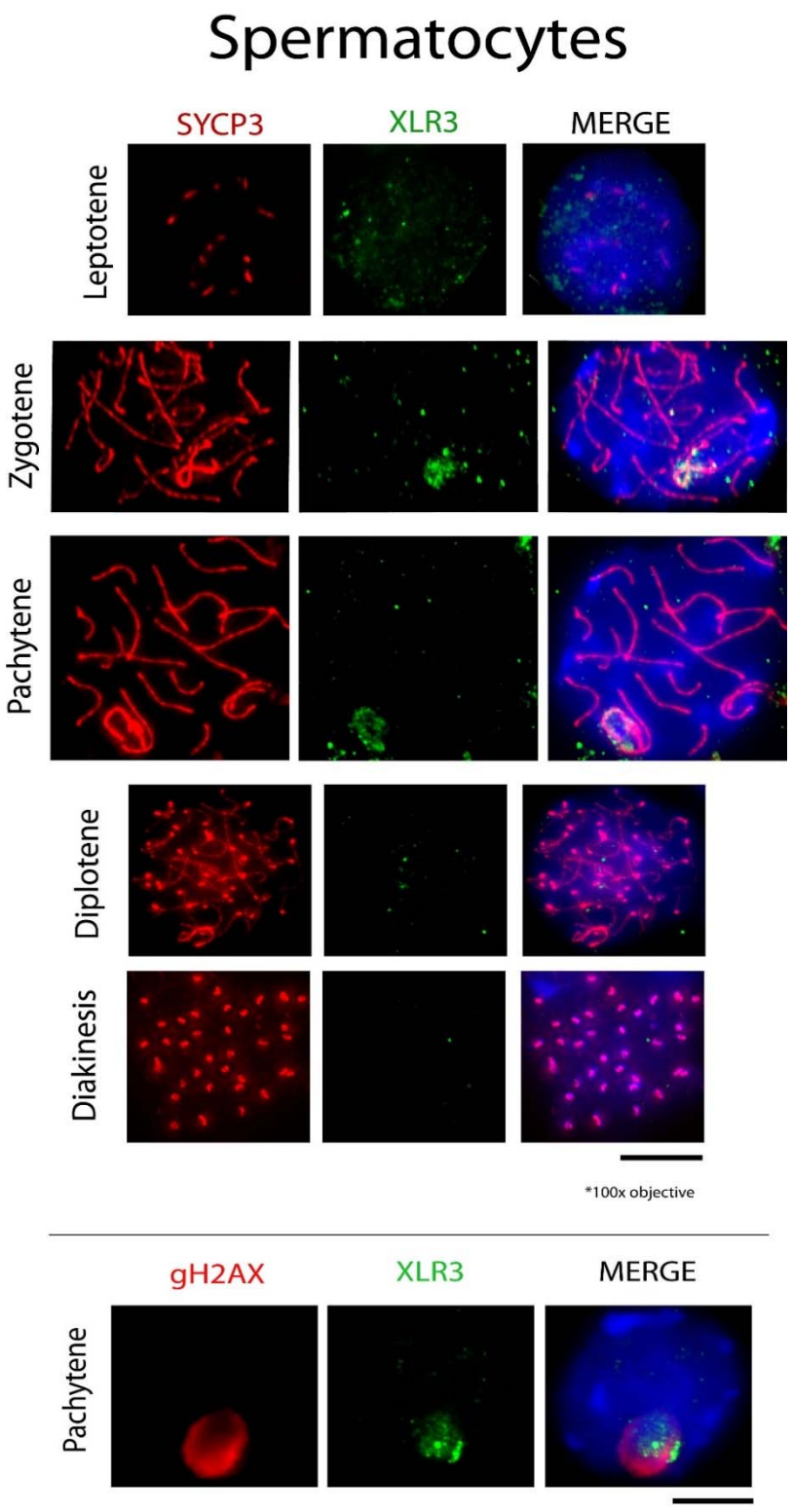


Figure 9. Immmunohistochemical Analysis of XLR3 on Meiotic Spermatocyte Spreads. Meiosis I stages are labelled relative to SYCP3 staining. XLR3 signal appears in green and DAPI stains DNA blue. The bottom panel shows gH2AX stained in red confirming XY body localization of XLR3.

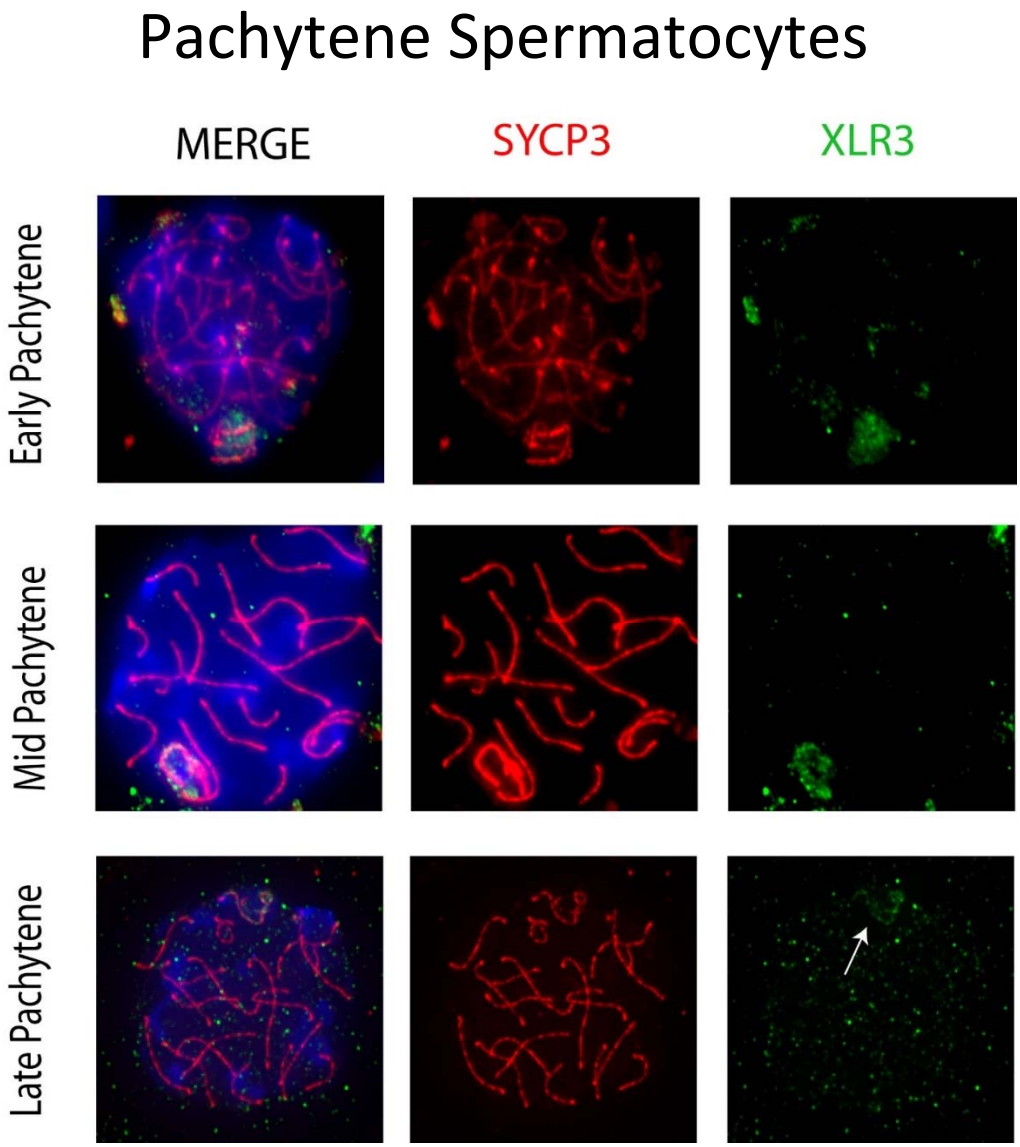


Figure 10. Immunohistochemical Staining of SYCP3 and XLR3 on Pachytene Stage Spermatocytes. A refined look at sex-body localization of XLR3 at early, mid, and late pachytene. Note: XLR3 staining in late pachytene localizes with the lateral elements at the inactive XY.

XLR3 Relocalizes to Post-Meiotic Sex Chromatin (PMSC)

To determine whether XLR3 plays a role after MSCI, immunofluorescence was conducted on similar spermatocyte spreads containing post-meiotic spermatids. After meiosis I/II the X and Y chromosomes are independent of each other, one per haploid round spermatid,

and additionally there is a slight de-repression of the X and Y chromosomes which is termed Post-Meiotic Sex Chromatin (PMSC). The X and Y chromosomes at this stage tend to localize adjacent to the nucleolus which is consistent with XLR3 immunostaining (Figure 11). This result is also consistent with persistent *Xlr3* mRNA expression through to the end of the first wave of spermatogenesis at 18.5dpp. Co-staining with X and Y chromosome paints and anti-XLR3 on round spermatids reveals co-localization of the two signals which is later absent in mature spermatids. As discussed earlier, two other Xlr superfamily proteins, SLX and SLY, have the same post-meiotic localization to PMSC and have additionally been implicated in helping to repress certain genes on these chromosomes. It is thought that SLX and SLY may be part of an inter-genomic conflict mechanism at this stage, a phenomenon that XLR3 itself may have a role in.

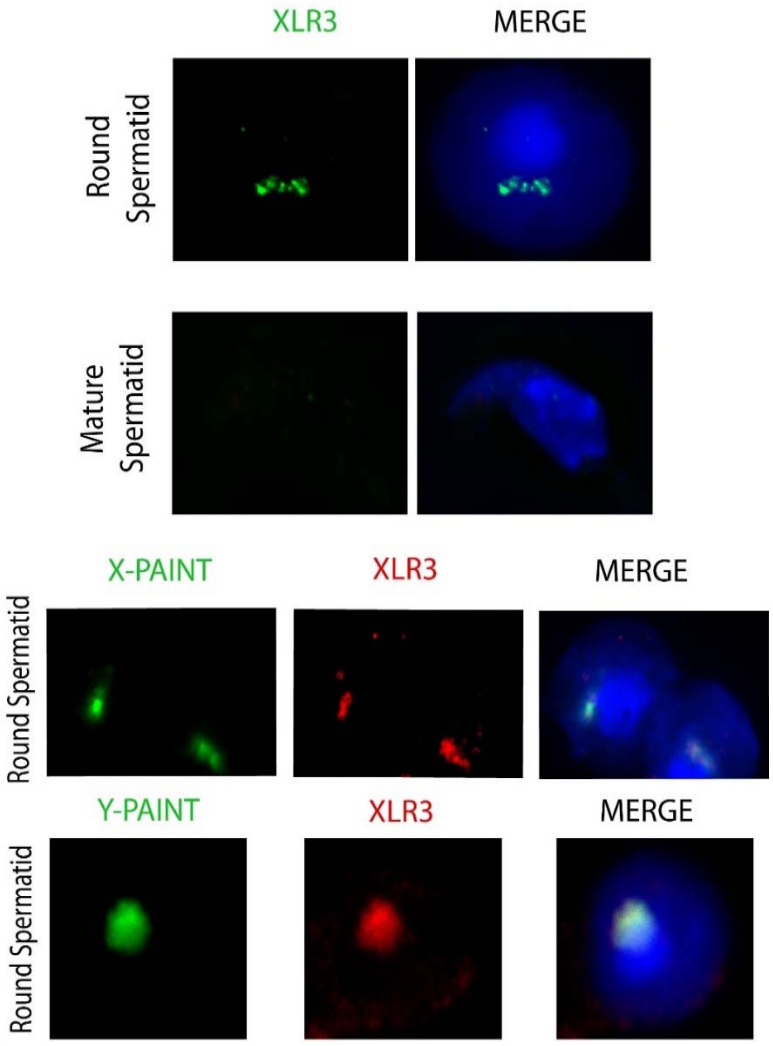


Figure 11. Immunohistochemical Staining in Post-Meiotic Spermatids. XLR3 localization in round spermatids and mature spermatids. X and Y whole chromosome paints are both labelled green in the bottom panels - XLR3 is labelled in red.

XLR3 Localizes to Nucleolus Organizer Regions (NORs) in Oocytes

Some but not all Xlr superfamily proteins are expressed during oogenesis. The localization of XLR3 to the sex chromosomes during MSCI and PMSC during male meiosis prompts the investigation of the protein in female meiosis, which lacks these two mechanisms. In order to determine whether XLR3 has a function in female meiosis, meiotic oocyte spreads were made from embryonic ovaries representing all stages of oocyte maturation up until

dictyate. Oocyte spreads were probed with anti-SYCP3 and anti-XLR3 revealing a dimorphic SC compared to male spermatocytes. Regions of aggregated SYCP3 are seen from leptotene through dictyate and are believed to be Nucleolus Organizer Regions (NORs). XLR3 signal is consistently detected at these aggregate foci even at dictyate after the SC has served its purpose (Figure 12). NORs were identified through co-staining with anti-Fibrillarin. The implications of XLR3 localization to these aggregates is not known.

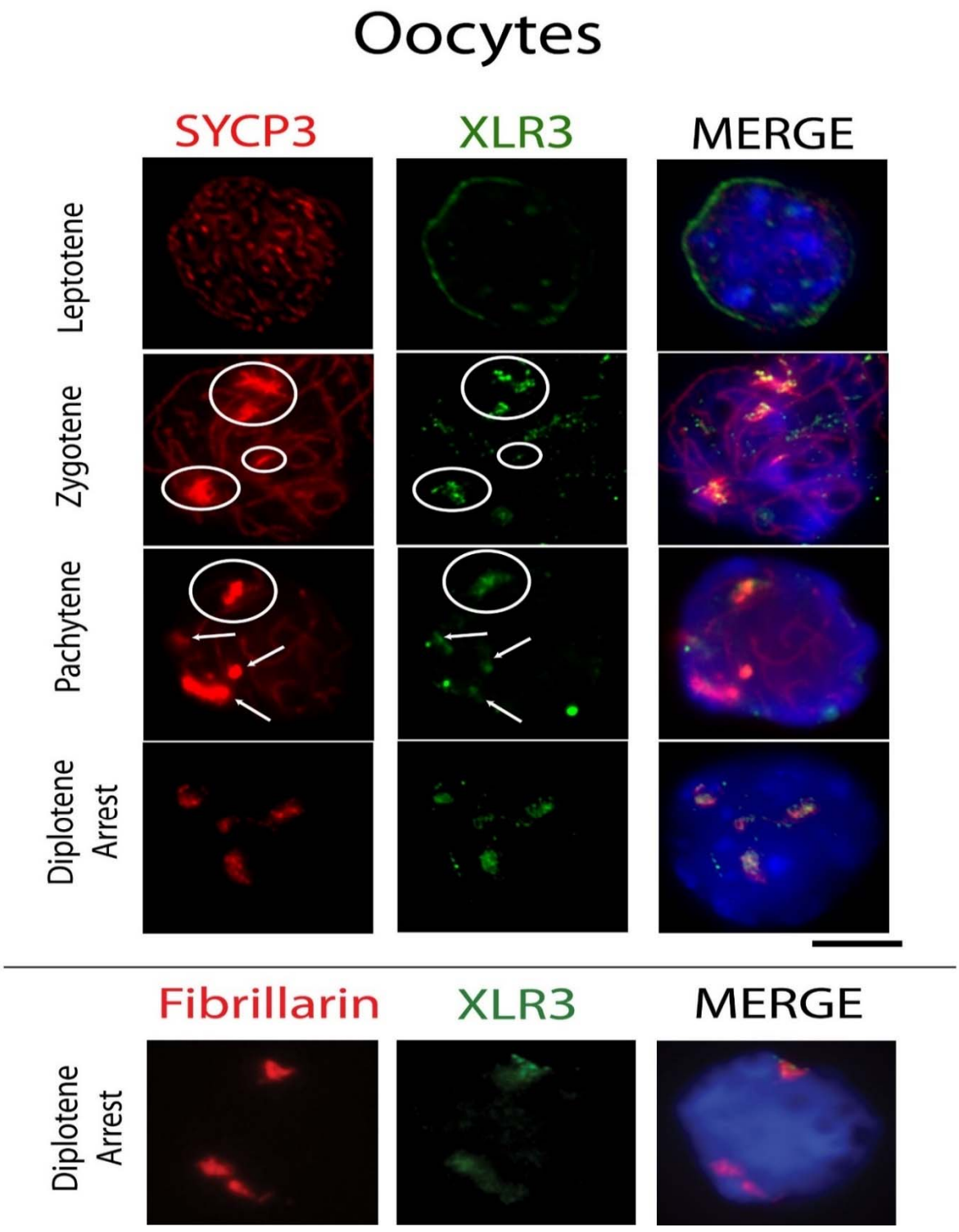


Figure 12. Immunohistochemical Analysis of XLR3 in Meiotic Oocytes. XLR3 appears to localize to aggregate areas of intensely staining SYCP3. The bottom panel displays XLR3 localization with Fibrillarin, a marker for NORs.

Chapter 3: Generation and Characterization of XLR3 Deficient Mice

Transmission Distorters

A transmission/segregation distorter is a genetic element that exhibits meiotic drive, the ability of a genetic element to transmit inter-generationally more frequently than the expected Mendelian ratio of 50%. Genic meiotic drive relies on close linkage of elements increasing in size through genetic hitchhiking of other genetic elements. Typically, transmission distorter elements are referred to as "ultraselfish" DNA meaning that they offer no selective advantage to the organism, but at the same time have no deleterious effects either.[49] These genetic elements are typically rich in chromosomal rearrangements as to decrease recombination events for preservation of the region. These regions are referred to as "haplotypes" and have the ability to cause large scale chromosomal meiotic drive events such is the case with chromosome Ab10 in maize.[50] In this case, a seemingly functionless repeat sequence on chromosome 10 acts as a "knob" or heterochromatic region similar to a neocentromere. This feature causes the chromosome to precociously move along the spindle poles and end up in the upper or lower egg allowing for >50% chance of being inherited.[51–53]

Meiotic drive is seen in nature in many systems including spore-killer (Sk) in fungi, segregation distorter (Sd) in *Drosophila melanogaster*, and t-haplotypes in mouse. The latter case is the result of a variant form of chromosome 17 discovered ~70 years ago which maintains a haplotype of inverted repeat sequences in tight linkage. This haplotype spans 20-40Mb and is observed in about 10-20% of the *Mus domesticus* and *Mus musculus* lineages.[50] Interestingly, a heterozygotic mouse carrying the t/+ alleles will preferentially pass on the sperm with a "t" allele presumably due to induced disruption of the wild type (+) sperm carrier.

This system requires the existence of a T complex responder *(Tcr),* a region within the t-haplotype that is required for preferential transmission and is found on both the t and + allele.[54–56] The Tcr in this case is a chimeric kinase protein capable of signaling flagellar development which was discovered after finally being cloned in 2000 by John Schimenti and co-workers.[57] This flagellar development protein causes defects for the *trans* (+) allele but has no effect on the *cis* allele. The *cis* allele t-haplotype contains T complex distorters *(Tcd)* which are thought to work in causing the defect in *trans*.[58] To date, 3 *Tcd* regions have been identified, but how they cause sperm defects by manipulating the Tcr in trans is unclear.[56]

Drosophila melanogaster has perhaps the best studied segregation distorter mechanism first discovered in 1956 by Sandler and Hiraizumi.[59] In this system a responder (*Rsp*) element exists on the second chromosome of some *Drosophila* alleles with a certain amount of copies of a 240-bp repeat sequence. *Rsp* sensitive (Rsp^s) alleles are composed of repeats arrays in the range of 700-2,500, while insensitive alleles (Rsp^i) have fewer repeats in the responder region (100-200). Alleles containing the segregation distorter (*Sd*) locus always carry the Rsp^i while the other allele (Sd^+) is subject to defects induced by the *Sd* allele. *Sd* codes for an enzymatically wild-type RanGAP protein, which for an unknown reason, mis-localizes in the cell of drive sensitive spermatids (Rsp^s). This effect occurs at Meiosis I and results in the death of Rsp^s sperm, which are unable to transition chromosome conformation to the lysine/arginine-rich histone state.[50,60] Many enhancers, modifiers and suppressors of the Sd allele have been identified, indicating that regulation of segregation distortion is more complicated than initially thought.

Theory predicts that the effects of distorters will eventually be diminished by evolution of unlinked suppressors restoring Mendelian ratios.[28,61] In hemizygous systems such as heterogametic sexes, sex chromosomes can evolve elements distorting their inheritance as well as suppressors of distorters, ultimately leading to genetic arms races characteristic of intragenomic conflict. This is best exemplified by S*tellate*, an X-linked locus in flies, which has a varying amount of repeats that directly distorts the inheritance of the Y chromosome. In low copy number, *Stellate*, a protein homologous to the beta unit of casein kinase II, forms linear crystals leading to a severe reduction in male fertility. In high copy number, *Stellate*, forms "star-like" crystal structures in spermatids resulting in completely sterile males due to non-disjunction in meiosis.[62–64] Y-chromosomes carrying the *crystal (cry)* or *Suppressor of Stellate* [Su(Ste)] allele are able to suppress the distortion of *Stellate,* correcting the transmission of the Y chromosome to 1:1. This suppressor effect in turn has led to an arms race where the X chromosome is selected to accumulate more copies of *Stellate* while the Y-chromosome evolves more copies of Su(Ste) to suppress *Stellate*.

Intragenomic Conflict in Mice

In 2005 Burgoyne et al. published evidence for intragenomic conflict between X- and Y-linked genes based on mice with various size Y-chromosome long arm (MSYq) deletions. These deletions resulted in an up regulation of 23 transcripts, 15 of which emanate from the X and Y chromosomes.[65] Phenotypic changes in these mice included teratospermia becoming worse with greater deletions and ultimately leading to infertility in some cases.[24] In order to further narrow-down Y-linked genes responsible for this effect, Cocquet and Burgoyne, generated a

transgenic mouse ubiquitously expressing a short-hairpin RNA targeting the abundantly expressed multicopy *Sycp3-like Y-linked (Sly)* gene with hundreds of deleted copies in the Y long-arm deletion studies (Figure 13).[4,48] Their scheme utilizes the strong RNA PolIII promoter "U6" in order to drive high expression of the hairpin targeting transcripts from over 300 *Sly* duplicates on the Y chromosome.[66]

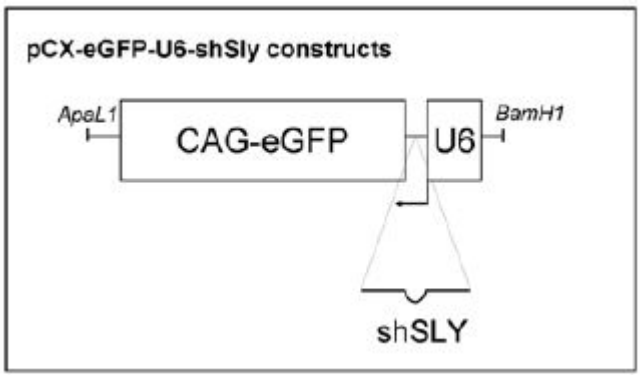


Figure 13. Short Hairpin Targeting Scheme for Knockdown of *Sly*.

The ubiquitous expression of this short hairpin resulted in no off target effects and seemed to only manifest a phenotype in testis. This phenotype matched very similarly to the Y chromosome deletion studies having abnormal mature spermatid heads and significant gene up regulation from the X and Y chromosomes. In fact, 86 genes on the X chromosome were significantly up regulated along with 27 genes on the Y chromosome while only 1 gene showed down regulation.[28] Under wild-type conditions these genes appear to be repressed by *Sly* through recruitment of heterochromatic marks (H3K9me3, CBX1)[4]. The genes under *Sly* repression include *Gmd1l, Ssxb, Rhox, Speer, Slx, Slxl1, H2al1*, and *Cypt* on the X and *Ssty1/2, Asty, Orly* on the Y. The majority of these genes are found in multiple copies and have post-meiotic functions in testis.[67]

Two of these repressed X-linked multicopy genes, *Sycp3-like X-linked (Slx)* and *Slx-like 1 (Slxl1),* show the same pattern of copy number amplification as *Sly* over the evolution of the mouse sex chromosomes. Like Sly these genes also show predominantly cytoplasmic localization in post-meiotic spermatids.[68] While conflict between *Sly* and its counterparts *Slx/Slxl1* is suggested by Sly's transcriptional repression of these X-linked genes, other evidence suggests strong antagonism exists. Immunostaining of SLY and SLX/SLXL1 reveals competitive localization to the sex chromosomes in round spermatids where they likely perform antagonistic roles. Even more evidence for conflict was seen when litters from the *Sly* deficient mice had litter sex ratios biased towards females while *Slx/Slxl1* deficient mice had litters biased toward males. This reveals a clear distortion effect for *Slx/*Slxl1 to increase the inheritance of the X chromosome while *Sly* tries to silence this affect. Remarkably, mice having deficiencies in both *Sly and Slx/Slxl1* (presumably removing the distorter and repressor activity completely), recovered a 1:1 sex ratio implicating these antagonistic genes in a pathway that regulates the sex chromosome expression and inheritance. With cytoplasmic and nuclear localizations of both these proteins it is unclear where exactly the conflict is occurring, however, the downstream manifestations are very apparent.

Manifestation of transmission distorters are apparent just prior to allele inheritance. Many of the phenotypic errors caused by distorters manifest during or after meiosis I resulting in gamete dysfunction i.e; death of spores in (Sk) *Neurospora* caused by first meiotic division segregation failure, improper flagellar development in t-haplotype drive-sensitive spermatids, segregation distorter (Sd) in *Drosophila* causing RanGAP protein mislocalization in drive sensitive spermatids, and *Stellate* in Drosophila which causes high levels of non-disjunction and

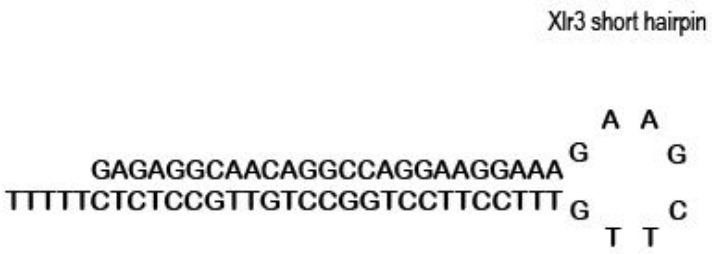
distorted sperm genotype ratios.[50,63] Even the recently proposed *Sly vs. Slx/Slxl1* conflict appears to be occurring throughout spermiogenesis at the sex chromosomes in mice.[4,5,48] XLR3, sharing ~10% identity with SLX and SLY, also immunolocalizes to the sex chromosomes in round spermatids, potentially implicating XLR3 in the same X and Y conflict. *Xlr3* is a compelling candidate to study in the context of sexual antagonism because one of three *Xlr3* paralogs is paternally imprinted, suggesting additional parent of origin regulation of this gene.[8,69] Additionally, bioinformatic QTL analysis by Good, Dean, and Nachmann shows an 8.44Mb region directly adjacent to the *Xlr3* locus implicated in hybrid sterility, suggesting that loss-of-function studies may reveal phenotypic abnormalities or sterility.[68]

Proof of Principle and Construction Scheme for *Xlr3* Knockdown *in vivo*

In the present study, we created an *Xlr3* deficient transgenic mouse in hopes to narrow biological functionality of this multicopy gene as well as to observe evidence for a role in distortion of the X and Y. As seen in chapter 2, XLR3 transiently accumulates during zygotene and pachytene in meiosis I, immunolocalizing with SYCP3 exclusively at the XY body. Consequently, a deficiency in XLR3 may phenotypically manifest as early as *zygonema*. Female mice with XLR3 deficiencies may also manifest a phenotype based on XLR3 immunostaining in oocytes. To engineer such a deficiency we chose a shRNA knockdown approaching targeting all active copies of Xlr3. The short hairpin design shown below was manually generated based on guidelines from Invivogen (http://www.invivogen.com/review-sirna-shrna-design) for targeting of the mature RNA from all of the *Xlr3a,b,c,d* and *e*, copies. In order to first test siRNA biogenesis of the target sequence in cell culture, the sequence was cloned downstream of a U6

promoter in an expression vector also containing a GFP selectable marker (U6-Xlr3sh-eGFP plasmid seen in Appendix). An empty GFP vector was used as a control for the neonatal mouse fibroblast transfections. Figure 14A displays the target hairpin which resulted in a ~50% reduction in *Xlr3* transcript in fibroblasts due to *Xlr3* siRNA biogenesis (Figure 14B).

A.)



B.)

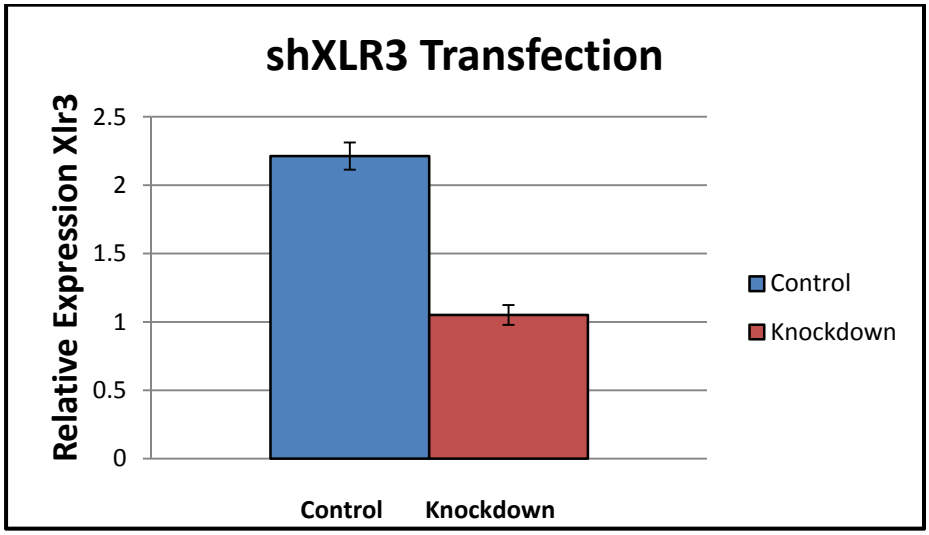


Figure 14. *Xlr3* Short-Hairpin Targeting Data. **A.)** Illustration of the *Xlr3* siRNA in hairpin form. **B.)** Quantitative RT-PCR analysis of transcription levels of *Xlr3* in control vs test RNAi knockdown. Error bars represent confidence at 95% interval.

Full knockdown of the *Xlr3* transcripts were not achieved likely due to incomplete transfection of cells (efficiency - 50%) ascertained through GFP fluorescence. FACs selection was not used because the decrease in transcript is apparent without the separation of non-transfected cells.

Preliminary Characterization of *Xlr3* Deficient Mice

With confirmation of the target sequence's ability to reduce *Xlr3* transcripts, a mammalian targeting vector was created to generate the knockdown system. The *Rosa26* locus on *Mus* chromosome 6 is commonly targeted for generation of high ubiquitous expression of transgenes.[70–72] *Rosa26* is typically targeted with high efficiency in embryonic stem cells, reaching up to 40% recombination efficiencies in many cases. Current evidence suggests XLR3 has a role confined to testis/ovary; however, to avoid the possibility of embryonic lethality or infertile chimeric founders from loss of this protein, a conditional knock down strategy was implemented. This is commonly achieved through insertion of a "floxed" STOP cassette in front of the transgene (Figure 15), allowing for control of lox-p site recombination in a tissue specific manner. The Jackson Laboratories house many tissue specific Cre expressing strains including the following pre-meiotic expressing testis promoter strains: vasa/ddx4-cre and str8-cre (http://jaxmice.jax.org/strain/006954.html).

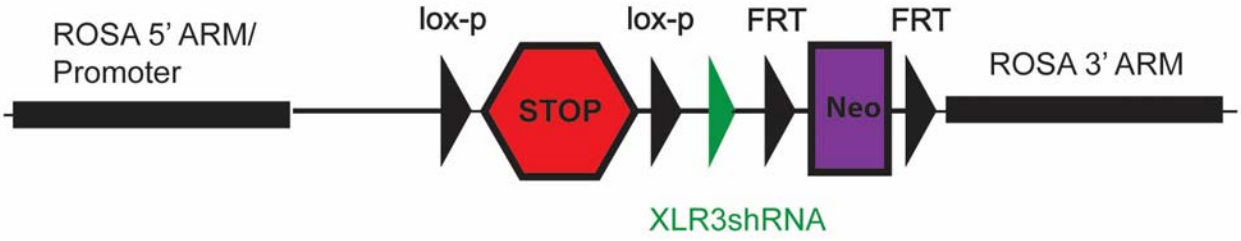
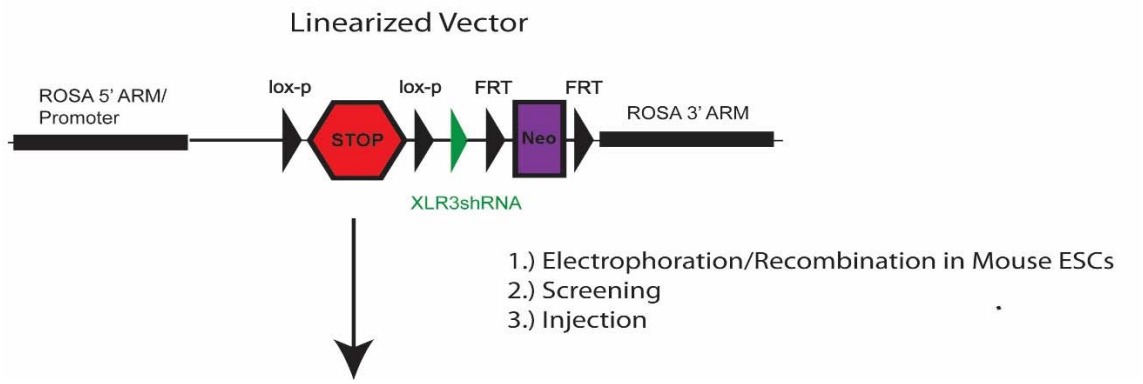


Figure 15 – Illustration of the In-vivo XLR3 shRNA Targeting Construct. The XLR3 siRNA preceded by a floxed STOP cassette is contained within the 5' and 3' ROSA locus targeting arms (not drawn to scale). Neomycin resistance for selection in mammalian cells is removable via Flippase.

The steps already taken to generate this transgenic model are subsequently discussed; however, breeding schemes are currently still on-going to generate the knock down model. The XLR3shRNA linearized construct was electroporated into C57BL/6J mESCs due to the fact that

most of the information known about the *Xlr* locus comes from the C57 derived X-chromosome. The neomycin cassette provides G418 resistance for positive selection and diphtheria toxin A (not shown) provides negative selection during mESC culturing. mESCs which survived selection were screened for presence of a recombination event, expanded, and injected into a recipient blastocyst by our collaborators at The Jackson Laboratory. With transgenic mESC giving rise to black fur and recipient mESC giving rise to white fur, the resulting mice are referred to as "chimeras" for having a mixed coat color. One would hope that the transgenic mESCs contribute to the germline of these chimeric mice and therefore allow for generation of a fully transgenic founder line when bred. The founder line is currently being bred, and initially, these founders will be crossed with mice ubiquitously expressing flippase (FLP) recombinase for prudent removal of the Neomycin selection cassette in all offspring. The resultant mice can then be crossed with Cre-recombinase expressing mice under a testis/time specific promoter. Initially, reduced fertility will be scored via litter numbers and the sex ratios of those litters, however, histology and SC integrity at the XY body will need to be examined via SYCP3 immunostaining. The breeding scheme is illustrated below in Figure 16.



Chimera

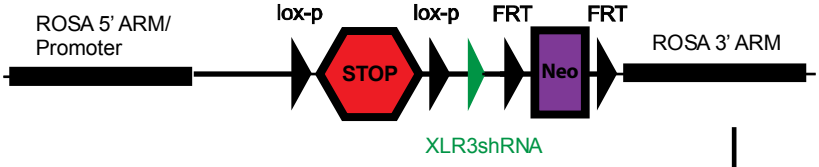
X

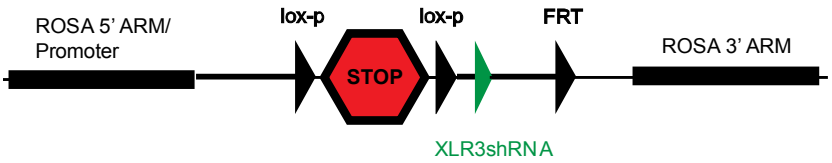
C57

Founder line

X

Flippase (FLP)





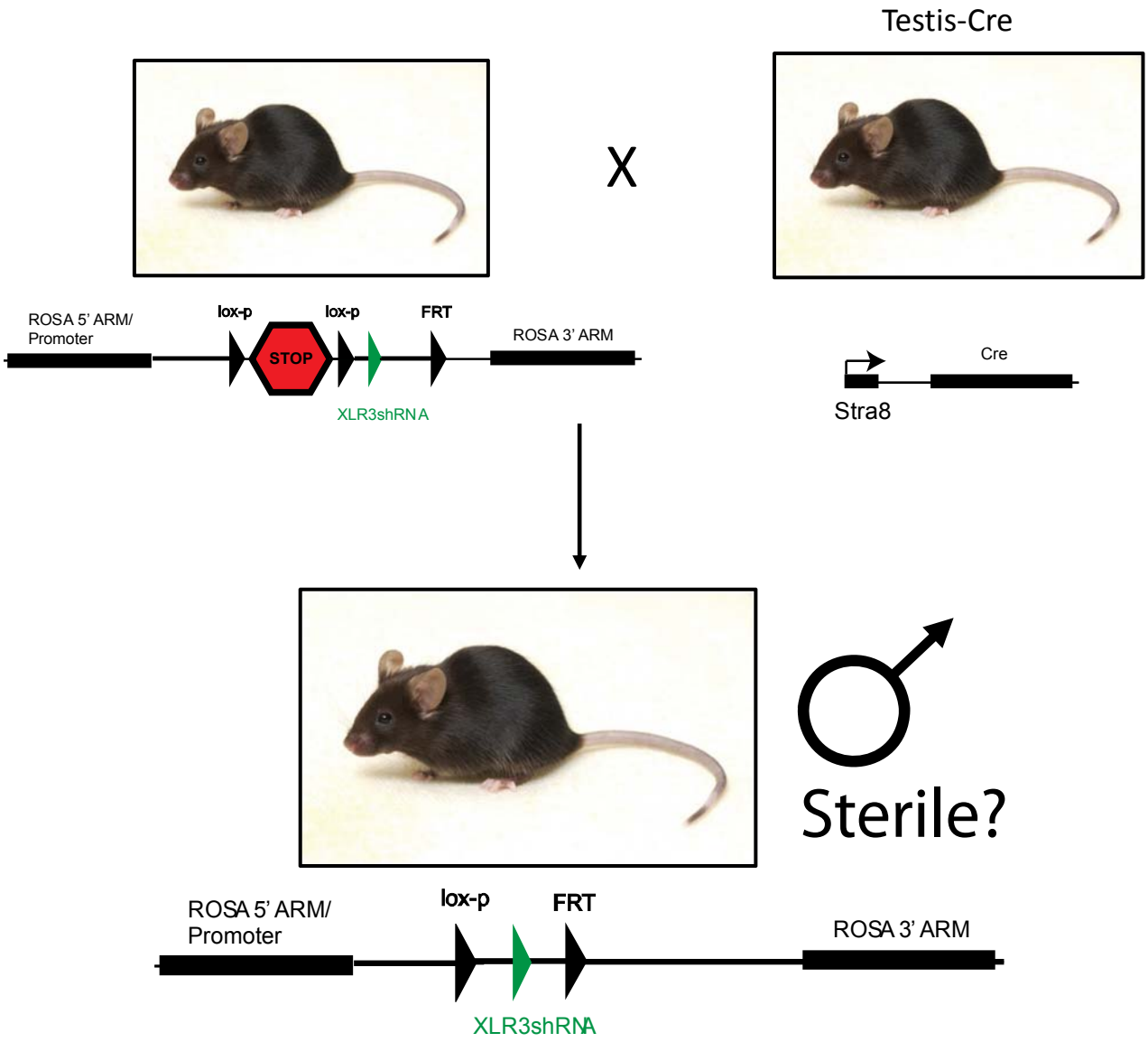


Figure 16 - Illustration of Breeding Scheme for XLR3 Knockdown Mouse. In this scheme, the linearized construct was electroporated into mESC for generation of chimeric mice were bred for germ line transmission. Founders will be crossed to Flippase expressing mice and resulting litters will be crossed to Testis-Cre expressing mice for removal of the STOP cassette and activation of the XLR3 shRNA (not drawn to scale).

Chapter 4: Epigenetics

Epigenetics and Genomic Imprinting

In 1942 C. H. Waddington proposed the term "epigenotype" to refer to the causal mechanisms occurring in the void between genotypes and phenotypes.[73] This definition suggests a general mechanism which could and does have many components and complexities. The term "epigenetics" was soon (1958) refined by David L. Nanney who claimed that aside from genetic components owing to cellular outcome, "auxiliary mechanisms" control the expression of the genetic material.[74] Therefore, epigenetics describes heritable traits that are independent of DNA sequence. Today's understanding of the term "epigenetics" is paramount to understanding how genomes are poised for developmental processes and environmental perturbations as was hinted so many years ago. Epigenetic mechanisms are known to manifest in many ways, showing commonalities between species, and allowing insight into the evolutionary necessity of these marks. Genomic imprinting and X chromosome inactivation (XCI) are the two most studied epigenetic mechanisms in mammals, the molecular basis of which is applied to the studies conducted in this thesis.

Genomic imprinting is characterized by mono-allelic gene expression based on parent of origin effects. Upon inheritance, maternal and paternal genomes retain specific "marks" or modifications that don't change the nucleotide sequence but rather predispose it to activation or silencing in the offspring. An evolutionary theory for this parental allelic predisposition, termed "Parent-Offspring Conflict" was realized in its earliest form by R.L. Trivers (1974), stating that "...parent and offspring are expected to disagree over how long the period of parental

investment should last, over the amount of parental investment that should be given, and over the altruistic and egoistic tendencies of the offspring as these tendencies affect other relatives."[75] However, in addition to conflicting psychological tendencies toward offspring, maternal and paternal contribution is also unequal at the epigenetic level as was shown by Takagi and Cooper in the early 1970's when revealing preferential inactivation of the paternal X chromosome in extra embryonic tissues in mouse and kanagroo.[76–78]

In 1989, David Haig and Mark Westoby built upon Triver's realization, proposing a biological relevance for what they termed "parent-specific gene expression" (PSGE) as it relates to provisioning sibling resources. They state that one sibling can only be half related to the next sibling of a different father, and therefore has an interest to benefit itself more than the next sibling. This benefit manifests as the offspring re-allocating maternal resources toward itself during development. In this sense, the paternal genomic contribution should select for alleles which exploit extra maternal resource, while this reallocation of maternal resources selects for maternal alleles that avoid a drain of resources in hopes of future reproductive success. Interestingly, Haig correctly predicts "…that there will exist a class of loci for which paternally derived alleles are considerably more strongly expressed than maternally derived alleles. These loci are predicted to encode proteins responsible for acquiring resources from the mother for the offspring, and the PSGE in question should be found in the offspring tissue that acquires resources."[79] In 1991, evidence for the first gene subject to PSGE was discovered, termed Insulin-like growth factor II (*Igf2*), an extremely potent mitogenic factor expressed during fetal development.

In their study, Dechiara et al, bred mice with a mutated form of *Igf2* resulting in "growth deficiencies" in progeny when passed through the paternal lineage. Passing the allele through the maternal lineage revealed normal progeny and nuclease protection assays in the same study would reveal paternal specific expression in embryos while the maternal alleles remained silent.[80] *IGF2* was additionally shown to be imprinted in humans in a study exploiting a 3' UTR *ApaI* polymorphism to show paternal specific expression in human placentae.[81] In mouse, *Igf2* is located on distal mouse chromosome 7 in a 1Mb region of 12 imprinted genes, including the non-coding RNA *H19* which is ~80Kb downstream from *Igf2*.[82,83] *Igf2/H19* are reciprocally imprinted loci which fall under control of the same mechanism, relying on a differentially methylated region (DMR) located 2Kb upstream of *H19* identified by Tremblay and Bartolomei.[84] This region harbors at least 12 CpG dinucleotides that display hypermethylation on the paternal allele coinciding with paternal specific repression of *H19,* and maternal allele hypomethylation coinciding with maternal specific expression of *H19*.

The mechanistic link between *Igf2* and H19 didn't become apparent until 1995 when Leighton and colleagues disrupted an enhancer 3' of *H19* affecting both *H19* and *Igf2* expression.[85] Additional work done 3 years later by Thorvaldsen et al. deleted the differential methylated region (DMR) 5' of *H19,* resulting in loss of imprinted expression of *H19* and *Igf2,* thereby confirming the first imprinting control region (ICR).[86] In 2000, Szabo and colleagues identified 4 repeats of a 21 bp motif in the previously identified ICR and were able to show a significant footprint of a zinc-finger DNA binding protein on the mouse maternal allele which was absent on the paternal allele.[87] In the same month Bell et al. would show this zinc-finger

protein, termed CCCTC binding factor (CTCF), to bind DNA at the ICR in a methylation dependent manner.[88,89]

Over the span of many decades the ideas of Waddington, Trivers, and Haig would be realized, culminating in the mechanism seen in Figure 17 below. On the maternal mouse chromosome 7, an imprinting control region (ICR) ~2kb upstream of the *H19* transcriptional start site is hypomethylated allowing for the insulator protein CTCF to bind. This results in the preferential association of downstream enhancers with the *H19* promoter, activating it. Lack of enhancer activity at the maternal *Igf2* promoter results in silencing. On the paternal allele, the ICR is preferentially hypermethylated abolishing CTCF occupancy at this region and inducing a long-range chromatin interaction (~80kb) of the 3' enhancers with the *Igf2* promoter, activating it. Paternal *H19* is silenced due to lack of enhancer activity at the hypermethylated promoter. It should be noted that in addition to the differential methylation seen at the ICR upstream of H19, there exists 2 other DMR's, "IGF2 DMR0" and "IGF2 DMR2", which are located in intron 2 and the 3' UTR of *Igf2,* respectively. These regions are normally paternally methylated and are not considered part of the ICR; however, studies involving growth impairment reveal hypomethylation of these alternate DMRs coinciding with abnormal *Igf2* gene expression.[90]

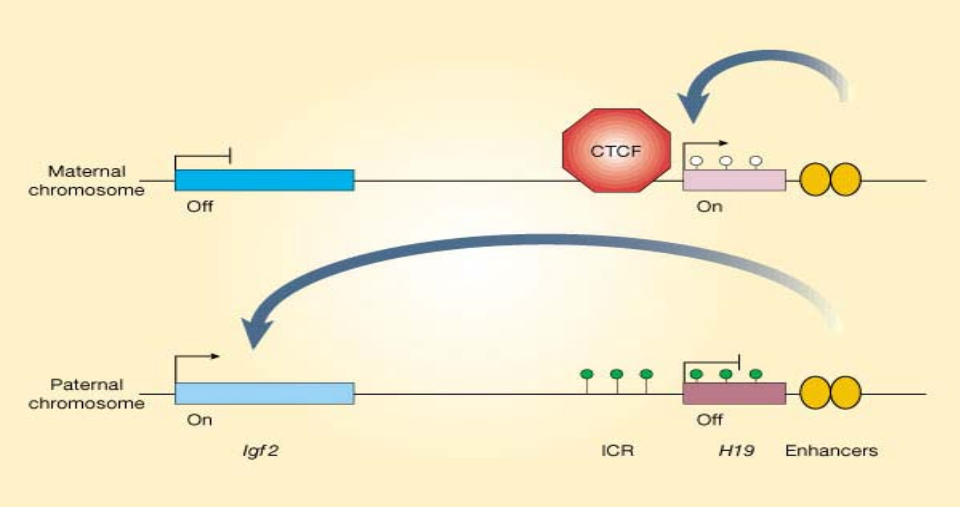


Figure 17. Mechanistic View of the Imprinted *Igf2/H19* Locus in Mouse. This locus is regulated by a differentially methylated region upstream of *H19*.[91]

Paternal allelic expression of *Igf2* aims to increase nutrient exploitation for growth at the provider's expense, and is therefore repressed by the maternal lineage. Also implicated in this conflict and antagonistic to *Igf2* is the cation-independent mannose-6-phosphate receptor, *Igf2r*, which acts as a receptor for Igf2 in mouse and humans. Interestingly, *Igf2r* is also imprinted displaying maternal-specific expression after embryo implantation.[92] Mutations in this receptor results in embryonic over-growth as well as increased extra cellular levels of Igf2 suggesting it controls the growth signaling of Igf2.[93–95] With the maternal allele active, Ifg2r acts in regulating high expression of paternal Igf2, while the paternal lineage predisposes this receptor to be inactive with the purpose of preventing Igf2 regulation. In this sense, maternal and paternal alleles must contribute unequally to the progeny in order to keep a balance in growth factors/receptors for creation of viable offspring. One would hypothesize, if the parental genomes contribute equally, then two paternal or two maternal genomes could be combined in a parthenogenetic manner for successful generation of an offspring. In fact this is

not the case with mammals as was demonstrated by zygotic nuclear transplantation experiments from which Surani and Barton concluded that the genomes must be marked such that different maternal and paternal contributions are necessary for viable offspring.[96–98]

To date, there have been ~160 imprinted genes identified between mouse and human, the vast majority of which fall into clusters on autosomes (www.mousebook.org). Many of the well-studied imprinted loci share other commonalities such as the presence of reciprocally imprinted long non-coding RNAs, which are currently being intensely studied in the field (i.e. *Kcnq1ot1*, *Air,* and *Xist/Tsix)*.[99] But, above all else, the hallmark of epigenetic regulation at these loci involves the presence of differentially methylated regions. As discussed earlier, the DMR upstream of H19 controls imprinting of the cluster, however, control of imprinted expression on the same chromosome at *Kcnq1* is determined by a separate DMR/ICR. This DMR regulates silencing of a cryptic promoter for a long non-coding RNA (*Kcnq1ot1*) on the maternal allele but results in activation of *Kcnq1* on the paternal allele.[100] It should be noted that imprinted clusters are much more complicated than originally thought, as is the case with *Dlk*/*Gtl2,* which has 3 DMRs, two of which control imprints in different ways when deleted and result in many differences in activating and repressive histone modifications.[101] On top of this, *Dlk*/*Gtl2* harbors many snoRNAs which are thought to act through RNAi biogenesis to regulate *Dlk* in *trans.*[99]

X-chromosome Inactivation

A crucial epigenetic mechanism exists in eutherian mammals during X-chromosome inactivation (XCI), the random transcriptional repression of one X chromosome in all female

cells. Muller saw such a mechanism as necessitated by the unequal inheritance of the sex chromosomes between the homo/heterogametic sexes in order to effect dosage compensation, manifest as silencing one X in females to match the hemizygous nature of the male X.[102,103] Additionally, according to Susumu Ohno (1967), in order to meet expression of the duplicate autosomes, the remaining active X must compensate by doubling expression of its 2,000+ genes.[104,105] Recently, Ohno's hypothesis has been challenged by new studies utilizing RNA-seq and microarray technologies to cast doubt on the hypothesized 1:1 X:AA expression ratio.[105–109]

Visually, the inactivated X chromosome was discovered by Barr and Bertram who observed dense chromocenters exclusively in female neurons indicative of an inactive X chromosome, later termed "Barr Body."[110] Mary Lyon also showed evidence for this mechanism in 1961 when observing the variegated coat colors in female mice due to their X chromosome mosaicism.[111] The genes subject to silencing during XCI remain so in future cell lineages, a clear example of epigenetic inheritance. XCI is controlled by the X-inactivation center (Xic) where the gene *X-inactive specific transcript* (*Xist)* is located and transcribes a long non-coding RNA (15kb) which "coats" the X chromosome destined for inactivation.[112,113] Ultimately, *Xist* is thought to be regulated by a multitude of other non-coding RNAs in the region (*Xite, RepA, Jpx, and Tsix*), and its spread along the X undergoing inactivation may occur by one of two proposed models. With evidence of only one control center (*Xic)* it could be that each gene region has its own recruiting capabilities for silencing; however, more popular is the idea that *Xist* loads at the *Xic* and is able to propagate outward through YY1 tethering to “spaced booster elements” to aid in global silencing.[114]

The stable repression of the inactivated X is ensured through epigenetic means such as the heterochromatic marks, H3K9 dimethylation, H4K20me, H2AK119ub1, H3K27me3, and incorporation of the histone variant macroH2A.1, while active marks are removed via H3K9 hypoacetylation and H3K4 hypomethylation.[115–118] Conversely, the X chromosome destined to remain active transcribes *Tsix*, which is an antisense transcript to *Xist*, that prevents initiation of silencing on the future active X.[119,120] Again, it should be noted that an imprinted form of X-inactivation occurs early in development (2-cell stage) of metatherian and eutherians in which the paternal X chromosome is exclusively inactivated. The paternal X remains inactive in all tissues in metatherians, however in eutherian blastocyst stage embryos all cells in the Inner Cell Mass (ICM) reactivate to prepare for random XCI, while cells associated with the trophectoderm keep their paternal X inactive.[76,121] XCI is therefore perhaps the most dynamic and large scale epigenetic mechanism known, involving differences in all major players; DNA methylation, non-coding RNA, variable histone marks, and conformation factors. It should be noted that *Xist* has not been located in marsupials but rather an analogous long non-coding RNA termed *Rsx* seems to play a role in inactivation of the paternal X.[122]

Roles for CTCF and YY1 in Epigenetic Gene Regulation

Gene expression follows a general mechanism of transcription factor binding to a regulatory sequence (RNA Pol II and TFII family members) followed by nearly simultaneous recruitment of activating or repressive nucleosome remodeling complexes and histone modifying enzymes. In addition to this orchestration of positive and negative factors, eukaryotes have evolved machinery connecting distant enhancers and promoters to achieve

proper transcription in many cases. The hallmark protein of such interactions, CCCTC-binding factor (CTCF), is an 11 zinc-finger DNA binding protein first discovered for its repressive regulation of the *c-myc* oncogene by binding a consensus sequence in the promoter region.[123,124] CTCF is highly conserved from fruit flies to humans and the central zinc finger domain retains 100% identity between birds and humans.[125] Interestingly, CTCF has a very loose consensus binding sequence attributed to the fact that it uses different combinations of the 11 zinc fingers to bind a variety of regions.[126] In mammals CTCF has ~55,000-65,000 binding sites of which ~5,000 are extremely conserved between different species and tissues.[127] About 50% of total mammalian sites appear to be used in tissue specific instances making it important in laying the unique topography of many cell lineages. Additionally, 50% of total mammalian CTCF sites fall near promoters and within gene bodies bolstering a role in gene expression mechanisms.[128–131]

While it was initially discovered as a repressive transcription factor, CTCF was soon also appropriately defined as an "insulator" for its roles in suppressing the spreading of heterochromatin. CTCF's function is best exemplified by circular chromosome conformation capture (4C) experiments suggesting extensive inter- and intra-chromosomal interactions at the *H19* DMR, dependent on CTCF occupancy, which is ultimately dependent on DNA methylation.[132] Additional work showed that enhancers downstream of *H19* interact with its promoter in the presence of CTCF, but form a larger chromatin loop to the *Igf2* promoter in the absence of CTCF at the ICR.[89,133–135] CTCF therefore lays down a topography for which DNA looping is determined. This topography is exemplified at the mouse *homeobox A* cluster *(Hoxa)*, where two chromatin loops form around a CTCF binding site which ultimately splits the cluster

into two distinct domains. The *Hoxa1-7* loop is enriched for the activating histone modification H3K4me3, while the *Hox9-13* loop is enriched for the repressive mark H3K27me3. This allows for specific Hox gene expression in a tissue and time specific manner. Knockdown of CTCF in this system reveals a spread of H3K27me3 across the entire locus resulting in silencing of all *Hox* genes in the cluster.[136]

In a study targeting the long range interaction landscape of gene promoters through 5C, Dekker and colleagues found that only 7% of chromatin interactions were with directly proximal genes and, contrary to repressive activity, CTCF is significantly enriched at distal interactions along with many active enhancer histone modifications. In the same study 79% of distal interactions showed a lack of CTCF blockage and rather an enrichment of enhancer activating marks further supporting its ability to tether distant enhancers to their promoters in a positive manner.[137]

The “looping” which is governed by CTCF can be extremely complicated as seen in the example below (Figure 18).

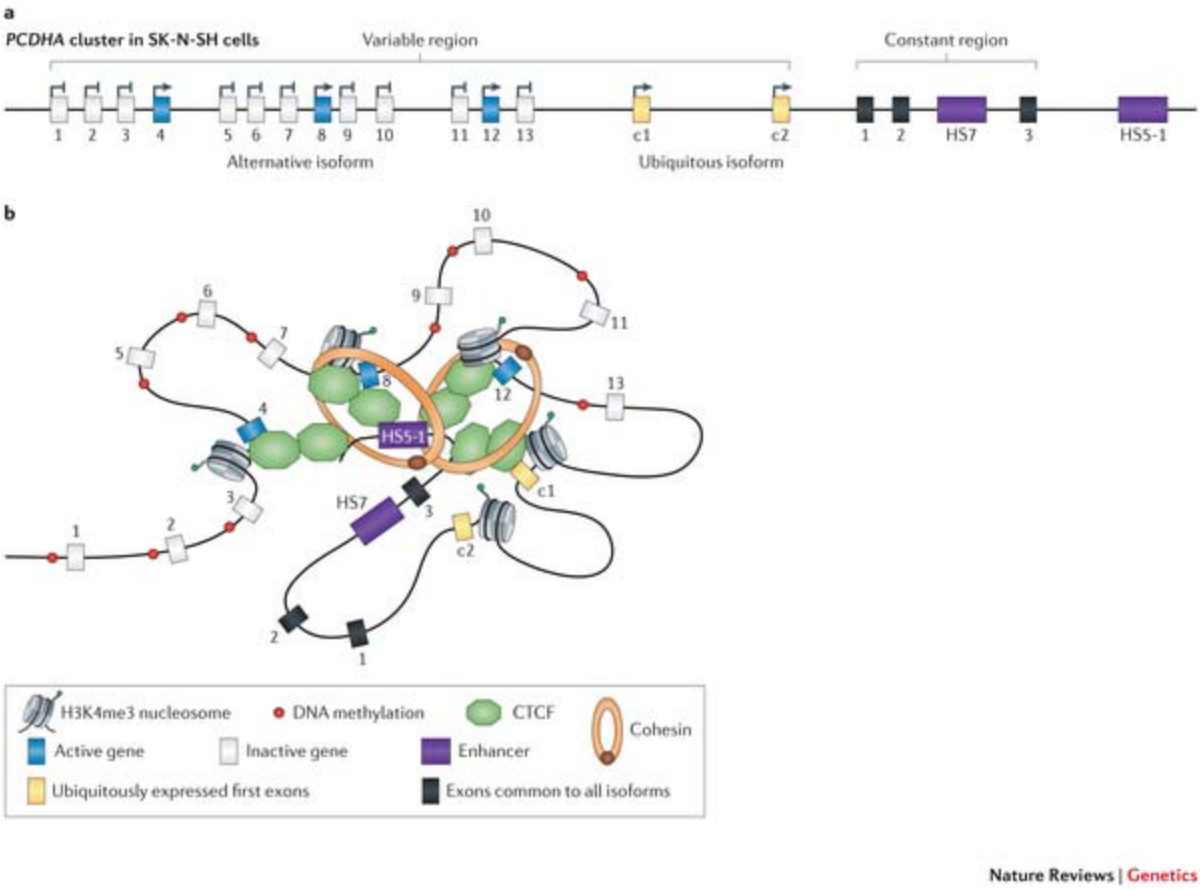


Figure 18. Mechanistic View of the Protocadherin (PCDH) Genes. Generation of varying 5' exons is accomplished through looping coordinated by Cohesin, CTCF, and enhancers – Figure courtesy of Ong and Corces.[131]

In this model, the protocadherin (PCDH) genes, which are responsible for neuronal diversity, have a very complex mechanism for promoter and alternatively spliced exon choice.[131] The many different forms of PCDH have incorporation of variable 5' exons followed by a common domain seen in all isoforms. The HS5-1 enhancer seen above, is tethered to certain promoters via cohesion and CTCF. Each of the first 13 exons has its own promoter and in this example exons 4, 8, and 12 are poised for expression and incorporation into a variable form of PCDH.[138–140]

These mechanisms utilize other transcription factors including PR domain zinc-finger protein 5 (PRDM5), zinc-finger protein 143 (ZNF-143), chromodomain helicase DNA-binding protein 8 (CHD8), PARP1, MYC-associated zinc-finger protein (MAZ), JUND, nucleophosmin and perhaps most importantly – Cohesin.[141] 50-80% of CTCF binding motifs are occupied by cohesin whose SA2 subunit has been shown to interact with the C-terminal domain of CTCF.[135,142] Cohesin is ultimately responsible for the physical structuring of DNA around CTCF sites and is indispensable for proper *Igf2/H19* looping and imprinted expression.[135] More recent studies involving CTCF and cohesin utilized a technique termed "Hi-C" which captures all interacting chromatin regions through protein crosslinks and sequences this pool for a global view of the genomes' "interactome." Interestingly, only 15% of CTCF binding sites fell in the border regions of Topologically Associated Domains or TADs while the other 85% of CTCF sites fell inside the TADs. TADs are defined by regions of genes that show a large amount of interaction with each other and not with other TADs. With the majority of CTCF falling within these regions it seems likely that CTCF's major role isn't to form TAD borders but is instead to form intra-TAD interactions.[143–146] This role is exemplified through the many examples listed earlier.

Another major player in these tethering mechanisms is Yin Yang 1 (YY1), a 4 zinc-finger protein known to interact with TBP, p300, c-myc, and HDAC2. Additionally, YY1 harbors a "REPO" domain, which recruits polycomb group proteins (PcG) resulting in trimethylation of H3K27. YY1 was initially discovered when Kassis et al. cloned the *Drosophila* Pleiohomeotic (PHO) gene sequence, and noticed 2 domains with homology to mammalian YY1. PHO itself is PcG protein and is thought to anchor PcG proteins to DNA, a function consistent to YY1, which behaves like a PcG protein *in vivo*.[147,148] YY1, like CTCF, functions very broadly in the regulation

of gene expression, having been implicated in X-chromosome inactivation, imprinting, oncogenesis, viral gene expression, and embryonic development.[149,150]

CTCF and YY1 work together in one of the best-studied looping mechanisms involving the B-cell immunoglobulin (Ig) and T-cell receptor (Tcr) loci which must be conformationally poised correctly in order for proper immunoglobulin formation. These loci are characterized by a variable region (V), a diversity region (D), and a joining region (J), which can create an enormously large diversity of receptors during lymphocyte maturation and development. These loci span very large regions (~3Mb) and become rearranged through use of two recombinases, termed Recombinase-Activating Gene 1 and 2 (RAG1/RAG2).[131,151,152] In Figure 19, the immunoglobulin heavy chain complex *(Igh)* locus uses a combination of enhancer sequences, RAG sequences, CTCF and YY1 sites for correct recombination between the many gene segments.

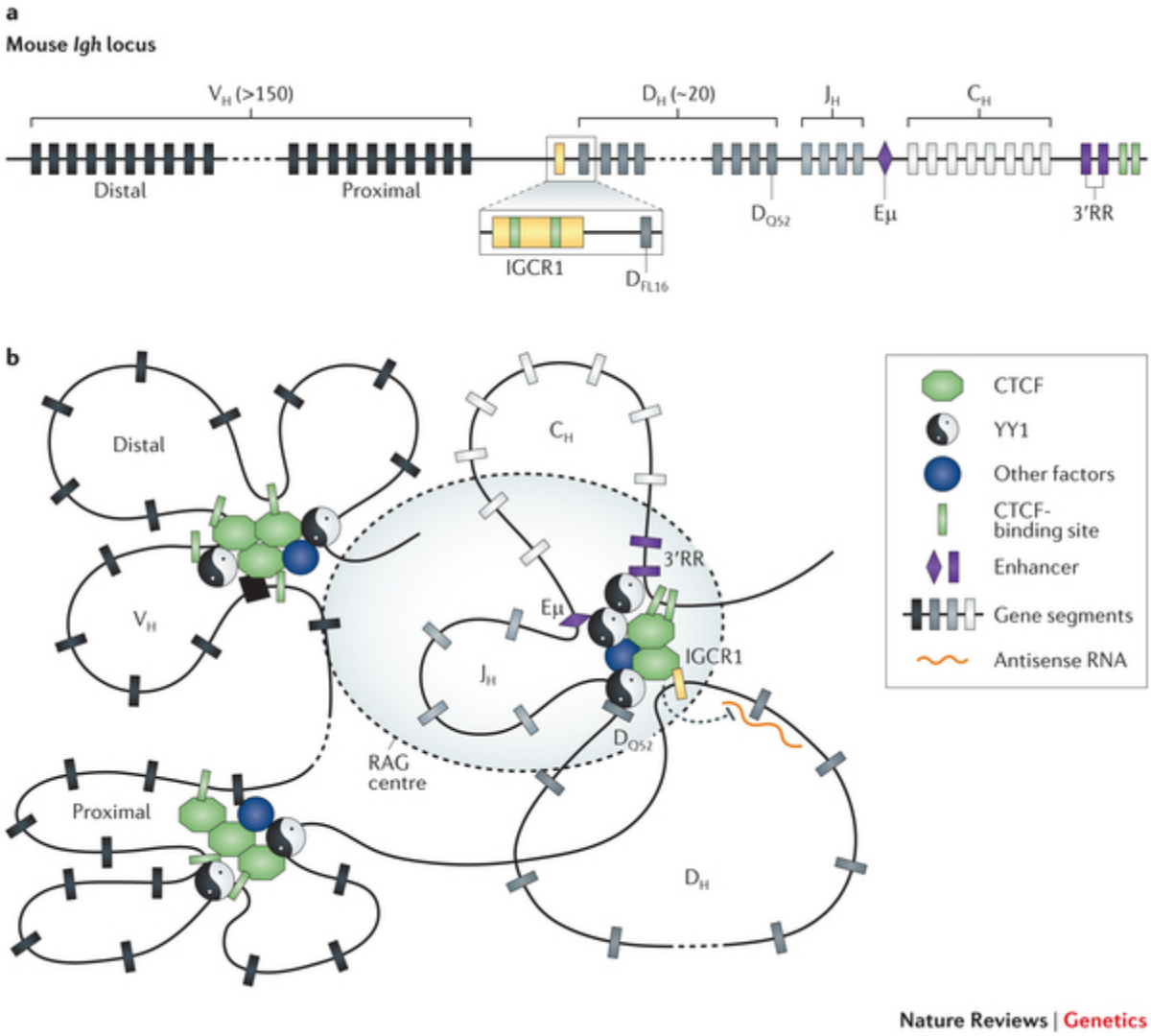


Figure 19. Mechanistic View of the B-cell Immunoglobulin and T-cell Receptor (Tcr) Locus. **A.)** Gene segment representation of the V_H, D_H, J_H, and C_H regions. **B.)** Illustration of the complicated looping mechanisms involving CTCF, YY1, and enhancers that direct reorganization of the *Igh* locus. Figure courtesy of Ong and Corces.[131]

Issues with the formation of these critical loops can result in immunodeficiencies and often lead to lymphoid malignancies.[153] This thesis aims to look at CTCF involvement in regulation of an imprinted locus, but it should be mentioned that CTCF, cohesin, and YY1 also play roles in mitosis during replication as well as chromosomal condensation making them critical in setting up landscapes throughout the entirety of the cell cycle.

Methylation and ICRs

Cytosine residues adjacent to a guanosine residue (CpG) can be actively targeted by DNA methyltransferase proteins in order to create a modified nucleoside termed 5-methylcytosine. These marks are commonly found in regions of high CpG content known as CpG islands where the CG residues exist in an observed-to-expected ratio greater than 60%.[154] This is generally known to be a chromatin silencing mark resulting in tighter chromatin compaction. 70-80% of total CpGs in mammals are methylated, while, conversely, 70% of CpGs found at promoters remain hypomethylated, presumably to allow transcription to occur.[155,156] *De-novo* methylation, such as allele specific methylation that occurs in imprinting, is set up by the DNA methyltransferases, *Dnmt3a, Dnmt3b* and *Dnmt3l*.[157] *De-novo* methylation at imprinted loci is set up in the germ line and remains in place throughout global zygotic demethylation of the paternal genome, followed later by the maternal genome.[158] Deleting *Dnmt3b* and *Dnmt3l* results in embryonic lethality, while deletion of *Dnmt3a* leads to post-natal lethality, further exemplifying the necessity of these proteins to set up important *de-novo* epigenetic marks for developing embryos. A fourth DNA methyltransferase, *Dnmt1*, is known to maintain DNA methylation marks throughout the cell cycle, having the ability only to fully methylate hemi-methylated DNA.[159–161]

Numerous methyl-binding domain proteins work in further recruitment of DNA modifying and chromatin effector proteins. One such methyl-binding protein, *Methyl CpG Binding Protein 2* (*MeCP2*), localizes to CpG methylation globally. In general, *MeCP2* binds to methylated DNA to block or enhance DNA binding of activators or repressors. *MeCP2* is located on the X chromosome in mouse and humans, and its various mutated forms underlie

neuropathological disorders, including Rett Syndrome (RTT).[162] RTT manifests in early development in humans (6-18 months) where phenotypes such as microcephaly, general growth retardation, ataxia and seizures are seen.[163] More recently, studies in mice show *MeCP2* to be necessary for recruitment of *alpha thalassemia/mental retardation syndrome X-linked (Atrx)* chromatin remodeler encoded by a gene that when mutated causes Alpha Thalassemia X-linked Intellectual Disability Syndrome (ATR-X), a severe intellectual disability disorder.[164] ATRX contains a SWI/SNF domain allowing it to remodel chromatin in heterochromatic regions and appears to be enriched at GC-rich regions of the genome, repeat sequences and telomeres. In fact, MeCP2 targets ATRX to DNA secondary structural regions known as 'G-quadruplexes' (G4s) which are defined by 4 consecutive duplicates or triplets of guanosines that interact through Hoogsteen pairing.[164,165] This type of pairing occurs between guanosines when the DNA is opened up at the replication fork, or during transcription, causing a replication fork or transcriptional stall.[166] This affect has been shown to occur in an allele-specific manner at 8.5% of the genome including imprinted genes and regions of large deletions. ATRX is targeted to these regions, especially at telomeres to counter G4 formation to ensure telomere structural stability. In fact, deficiencies in ATRX induce telomere dysfunction, potentially through abnormal lengthening of telomeres.[167,168] ATRX is thought to cull such affects through recruitment of H3.3, a histone variant associated with open chromatin and active transcription.[169] Interestingly and significant to this thesis, ATRX was recently found to be recruited to the promoter of the mouse X-linked imprinted gene *Rhox5*, contributing to regulation of its expression.[170] Additionally, MeCP2/ATRX binds the maternal allele ICR of imprinted *H19*, loss of which prevents post-natal repression of this non-coding RNA.[171] In fact,

loss of *ATRX* was also shown to derepress many imprinted genes post-natally, including Gtl2/Dlk1. In their study, Kernohan et al, revealed allele-specific binding of MeCP2/ATRX at ICRs for many imprinted loci, loss of which caused a change in nucleosome occupancy at CTCF consensus binding sites common to ICRs resulting in abolishment of long-range chromatin interactions coincident with changes in gene expression.[164]

While MeCP2 seems to function broadly by binding methylated CpGs globally, another zinc-finger DNA binding protein, ZFP57, is shown to bind and protect maternal methylation exclusively at ICRs and its 6 base pair consensus sequence (TGCCGC) is found at all known ICRs except one. ZFP57 is thought to set up methylation memory at ICRs specifically on the maternal allele, and loss of ZFP57 results in embryonic lethality.[172,173]

Epigenetics in Cancer and Disease

Global DNA hypomethylation is a key characteristic of tumorigenesis with the exception of certain sites with specific CpG hypermethylation.[174] The hypomethylation of many different genetic elements, including repetitive sequences and retrotransposons, leads to DNA instability which can take the form of increased expression of growth factor genes such as *T-Tas*, *MAPSIN*, *MAGE* (melanoma-associated antigen), and even *IGF2*.[175,176] In a seminal study by Ehrlich in 2003, a patient with immunodeficiency, centromeric region instability and facial anomalies syndrome (ICF) was revealed to carry a mutant form of DNMT3B, which resulted in global hypomethylation, linking hypomethylation to instability.[177,178] Loci that are selectively hypermethylated tend to fall under the category of "tumor suppressor genes" (*p16, MLH1, and BRCA1*).[179] As discussed earlier, BRCA1 and MLH1 are crucial factors involved in the DNA repair

pathway in somatic and germ cells, and are therefore essential to proper cell development. How these genes are targeted for hypermethylation isn't well understood currently.

Aberrant DNA methylation is only partially explanatory in tumorigenesis. For a multitude of cancers, high expression of histone deacetylation proteins (HDACs) is consistent with loss of the activating histone modifications, H4K16ac/H4K20me3, resulting in repression of tumor suppressor genes.[180,181] Additionally, the histone methyltransferase, EZH2, is over-expressed in prostate and breast cancers, and is responsible for altering the methylation of H3K27 and H3K9 at aberrantly expressing genes.[182] EZH2 is also responsible for the reversible silencing of key developmental genes by trimethylating H3K27. In cancer cells, this trimethylation mark was recently shown to be a precursor for *de novo* DNA methylation, potentially explaining at least a portion of aberrant methylation.[183] It should be noted that miRNAs, which are involved in transcription factor regulation, cell proliferation and apoptosis, also become deregulated in cancer, acting as both oncogenes and tumor suppressors. Another facet of chromatin architecture relatively unexplored in cancer is nucleosome occupancy, connecting fine-scale physical gene structure to tumor formation.[178]

With such power over gene expression, it is no surprise that epigenetic mis-regulation can result in disease. As mentioned earlier, parentally established allelic marks from both parents are essential for viable development of mammalian offspring. Angelmann Syndrome (AS) is caused by inheritance of both chromosome 15s from the paternal lineage or maternal deletion of a key region on chr15.[184] Phenotypic manifestations include more frequent laughter and smiling, abrupt limb movements, and general mental retardation. The locus responsible resides at 15q11-13 which harbors multiple imprinted genes with paternal specific expression

(*Snurf/Snrpn, Znf127, Nectin, Mkrn3, and Magel2)*.[185–187] Conversely, when the progeny inherits both chromosome 15s from the maternal lineage, or a paternal chr15 regional deletion, a similar disorder known as Prader-Willi Syndrome (PWS) results. PWS is characterized by hypotonia and hypogonadism early on and docile behavior coupled with excessive eating after infancy. These phenotypes are consistent with the idea that disruptions to imprinted genes generally lead to growth abnormalities making them good candidates for gene therapy.

Improper genomic imprinting has also been hypothesized to contribute to Autism Spectrum Disorder (ASD). ASD is a complex developmental disability manifest as impairments in communication, learning, and forming relationships. Interestingly, 14.7 per 1,000 8-year old children were shown to have some form of ASD, with a sex bias affecting males much more commonly than females (4.3 to 1).[188] Fetal testosterone is being intensely looked at to explain this bias, however, to date, no male-specific factor has been identified to explain its prevalence, and instead a new hypothesis has begun to emerge termed the "Female Protective Effect" (FPE). FPE states that unknown female-specific factors are able to somehow prevent reaching the threshold of ASD in a background of non-sex-specific predisposing influences.[189] As mentioned previously, fetal testosterone has been suggested many times as being the male-specific vulnerability factor, potentially causing an "extreme male brain effect" which manifests as ASD, but evidence for this theory has so far been lacking.[190]

Equally as relevant is the idea that the significant sex bias in autism could be due to an imprinted locus on the X chromosome, which will be phenotypically much more prominent in males who are hemizygous for the X-chromosome. This idea emerged from a study conducted by Dr. David Skuse, a British psychiatrist studying social cognitive defects in Turner Syndrome

patients. Turner syndrome is characterized by X chromosome monosomy in females resulting in phenotypes similar to ASD; i.e. - behavioral deficiencies such as hyperactivity, learning impairments, and poor social interactions.[191,192]

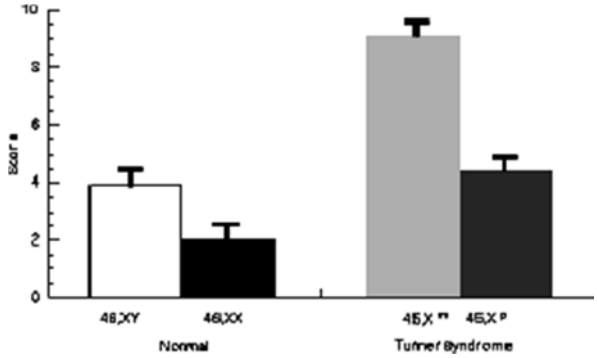


Figure 20. Turner Syndrome Cognitive Study. Relative scores from cognitive tests conducted by Dr. David Skuse on X monosomic (Turner Syndrome) females. Higher scores coincide with poorer cognitive ability. Figure courtesy of Skuse et al.[191]

In his study Dr. Skuse subjected 45,X^m and 45,X^p female patients to social cognitive tests and revealed a significant difference in performance based on the parental origin of the X chromosome (Figure 20). X^m carrying females scored higher and therefore worse compared to X^p carrying females. He later hypothesized that this could be due to an X-linked imprinted gene locus or loci. As discussed earlier, in 2005 Raefski and O'Neill discovered the first X-linked imprinted genes in mouse, *X-linked Lymphocyte Regulated (Xlr) 3b/4b/4c,* work that was inspired by the observations of Skuse and colleagues.[8]

The *Xlr3b, Xlr4b, Xlr4c* Imprinted Locus

Like many autosomally imprinted genes, *Xlr3b, Xlr4b and Xlr4c* are present in a cluster displaying maternal specific expression (Figure 21A and B). Raefski and O'Neill utilized X monosomic female mice with either the maternal X (39, Xm) or the paternal X (39, Xp) to show

significant allelic expression differences in *Xlr3b/4b/4c*. Reciprocal crosses of two inbred mouse strains, C57/C3H, provided proof of the imprint in normal females, thereby providing the first candidate genetic locus for Skuse's theory.

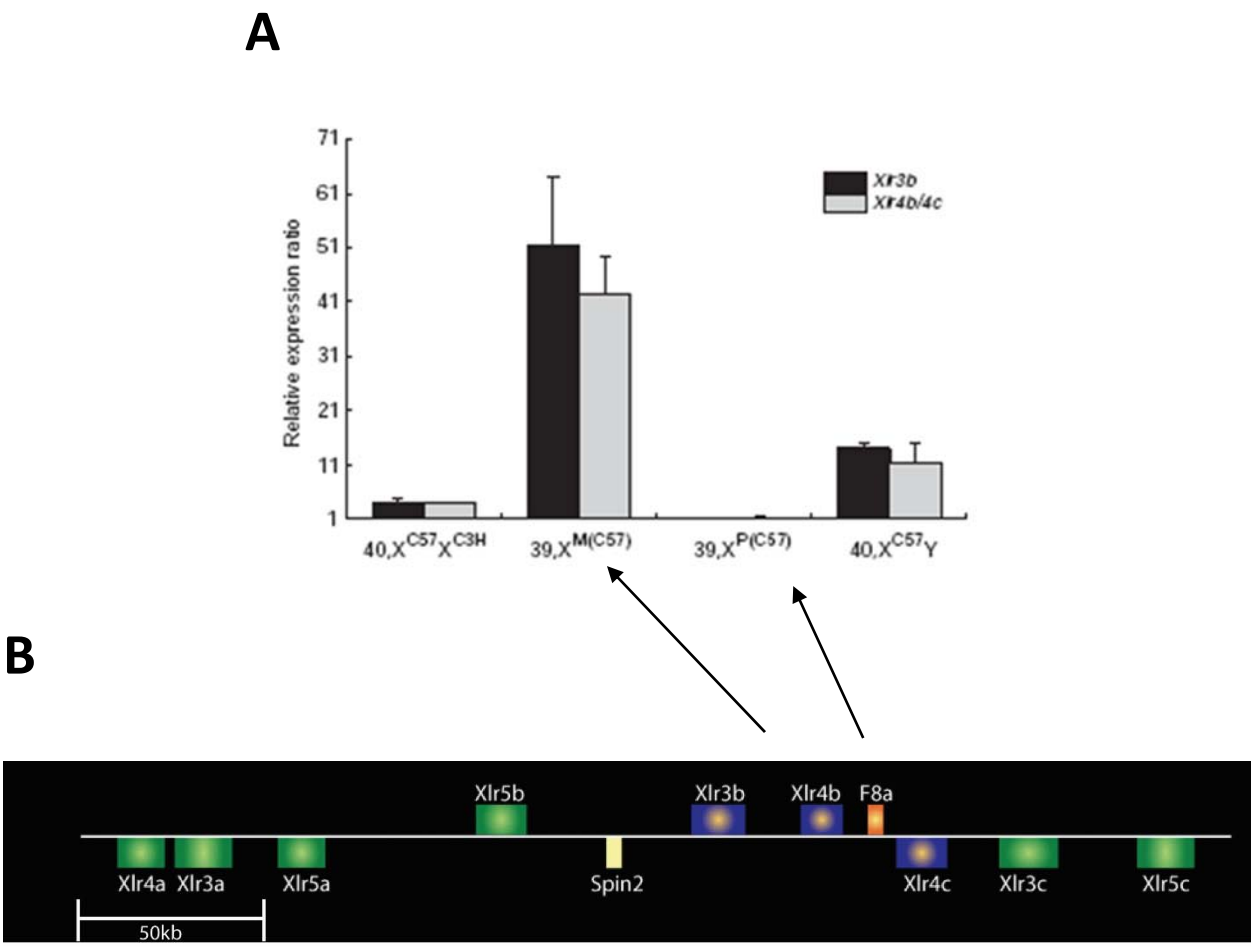


Figure 21. Discovery of an Imprinted Cluster of Genes at the *Xlr* Locus. **A.)** Allele/Paralog-specific imprinting assay on 39,Xm and 39,Xp neonatal mouse fibroblast. **B.)** Illustration of the *X-linked Lymphocyte Regulated* (*Xlr)* locus at XA7.2 (mm10) - imprinted paralogs are seen in blue. Figures courtesy of Raefski and O'Neill.[8]

This locus (XA7.2) harbors members of the *Xlr* superfamily of genes, which exist in hundreds of copies on the murine X and Y chromosomes. As discussed in Chapter 1, the primary function of this gene family appears to be involvement in repression of gene expression from the X and Y

chromosomes during spermatogenesis. Although relatively high expression of X*lr3/4/5* transcripts in brain, fibroblast, and lymphocytes suggested other functions, the XLR3 protein has only been detected in testis and oocytes. Interestingly, an antibody raised against the Xlr4 N-terminus reveals a band of the correct size on a western predominantly in the cytoplasm of neonatal whole brain and fibroblast (Figure 22). This is counter to the expression status of all other XLR superfamily proteins but is consistent with the high mRNA expression of *Xlr4* in brain and fibroblast.

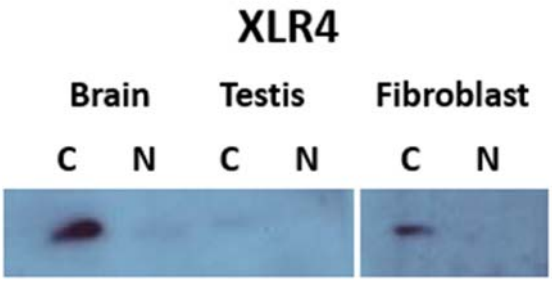


Figure 22. Western Blot probed with anti-XLR4. Cytoplasmic and nuclear fractioned tissue was probed with a custom anti-XLR4 antibody revealing a prominent band in mouse brain and fibroblast.

Two other imprinted genes have been found on the mouse X-chromosome; *ferritin, heavy polypeptide-like* 17 *(Fthl17)*, a maternally silent gene which becomes paternally expressed starting at the 2-cell stage of embryogenesis, and *Rhox5/Pem*, also a maternally silent gene which is paternally expressed up until the morula stage when the imprint flips and maternal expression predominates.[193–195] Both genes have high expression in testis/oocytes and appear to have roles in gonad development.

Investigation of the imprinting mechanism involved in paternal silencing of *Xlr3b/4b/4c* was initially carried out by Dr. Michael Murphy, Dr. Sohaib Qureshi, Dr. Seth Kasowitz, and Dr.

Ben Carone. Since parental-allelic differences in DNA methylation are the global epigenetic mark of all well-studied imprinted loci, the interrogation of the Xlr3b/4b/4c imprinting mechanism began with a search for differential methylation of CpG residues. CpG islands were bioinformatically identified at XA7.2 and subject to sodium bisulfite treatment followed by Sanger sequencing. Figure 23 below shows the methylation status of CpG dinucleotides at each island across the locus. Surprisingly, both parental copies are hypermethylated for each of these loci, despite active transcription of only the maternal copies. No significant allelic differential methylation is seen except at *F8a* in the normal XX genotype. F8a was included as a control since it resides in the middle of the imprinted Xlr cluster, but is not itself imprinted. F8a shows the expected 50% methylation attributable to XCI in normal females, but hypomethylation in monosomic females.

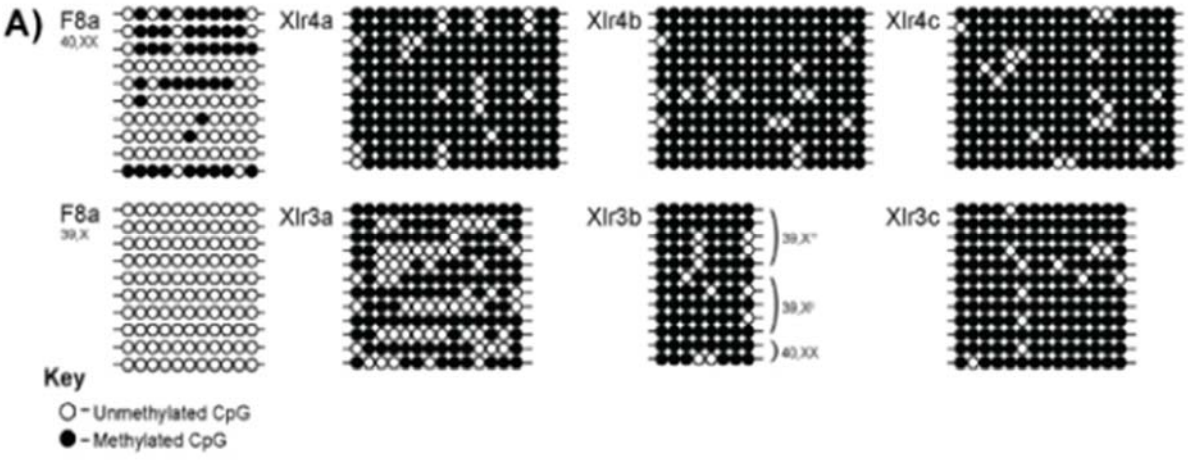


Figure 23. Sodium Bisulfite Sequencing of CpG Islands at XA7.2. Blocks of clones represent various CpG islands across the Xlr locus. Methylated CpGs are indicated by filled in "lollipops" while unmethylated CpGs are blank. Results courtesy of Dr. Seth Kasowitz and Dr. Ben Carone

Methylated Binding Domain protein immunoprecipitation hybridized to tiling array (MeDIP-chip) experiments performed by Dr. Michael Murphy did not reveal a DMR within 2 Megabases

spanning the Xlr locus (data not shown). While investigating the processing and disposition of the *Xlr3b* mRNA in the cell nucleus, Dr. Seth Kasowitz discovered that while the processed transcript showed maternal dominant expression, the 5' end of the primary transcript was biallelically transcribed (Figure 24). Maternal dominant expression isn't detected in the primary transcript until at least intron 7.

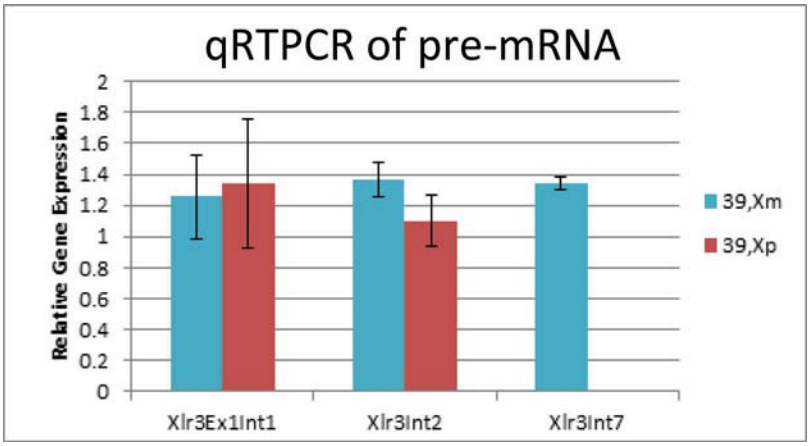


Figure 24. Quantitative RT-PCR Analysis of *Xlr3b* Primary Transcript. qRTPCR of the *Xlr3b* primary transcript assayed at various positions across the gene revealing lack of a transcript on Xp neonatal brain at Intron 7. Figure courtesy of Dr. Seth Kasowitz.

These findings suggested that the imprinted expression of *Xlr3b* may be manifest at transcript elongation or processing. To investigate this possibility, Dr. Sohaib Qureshi performed Chromatin Immunoprecipitation (ChIP) to characterize the chromatin architecture and the profile of RNA Polymerase II occupancy on the two parental alleles of *Xlr3b*. Equal expression of the 5' end of the paternal transcript suggested that transcription may be initiating but failing to complete elongation or processing on the paternal allele. Figure 25A below shows enrichment of total RNA Polymerase II across the *Xlr3b* gene in X-monosomic tissues. There appears to be an increase of RNA Pol II on the paternal allele at the Intron 3/Exon 4 boundary while more maternal RNA PolII successfully makes it to the end of the gene. This data is corroborated by

results from an immunoprecipitation using an antibody specific to “elongating” RNA Pol II, which shows a clear increase in actively transcribing Pol II at Intron 3/Exon 4 while, again, more maternal RNA PolII makes it to the end of the gene (Figure 25B). This data suggests an RNA Pol II stall on the paternal allele of *Xlr3b* due to an unknown mechanism.

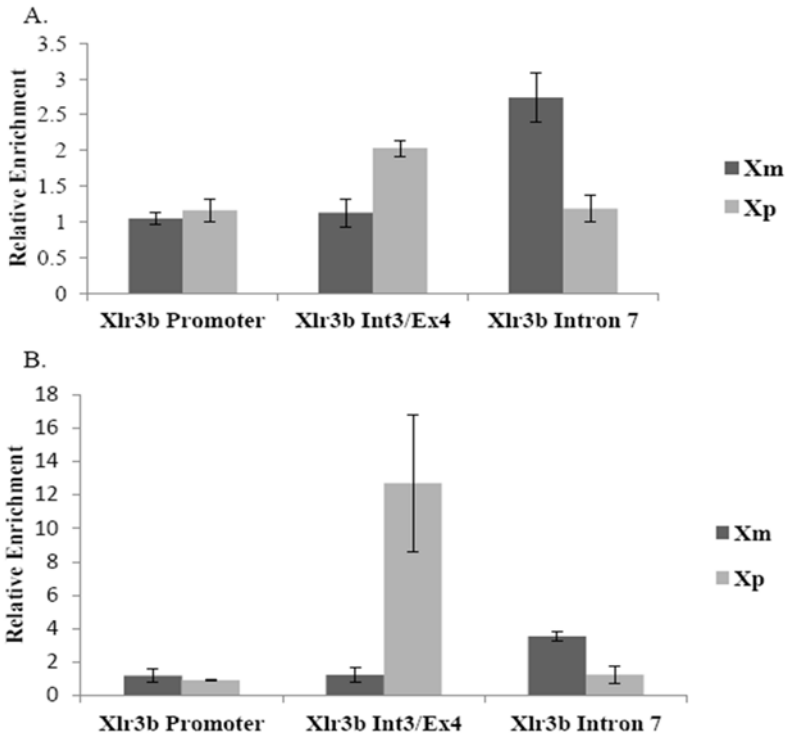


Figure 25. RNA Polymerase II Enrichment across Xlr3b. **A.)** Quantitative RT-PCR of RNA PolII (4h8) ChIP enrichment across Xlr3b. **B.)** Quantitative RT-PCR of RNA PolII (Ser2) ChIP enrichment at Xlr3b. Both histograms indicate a buildup of RNA PolII in the middle of the gene at intron3/exon4. Figure courtesy of Dr. Sohaib Qureshi.

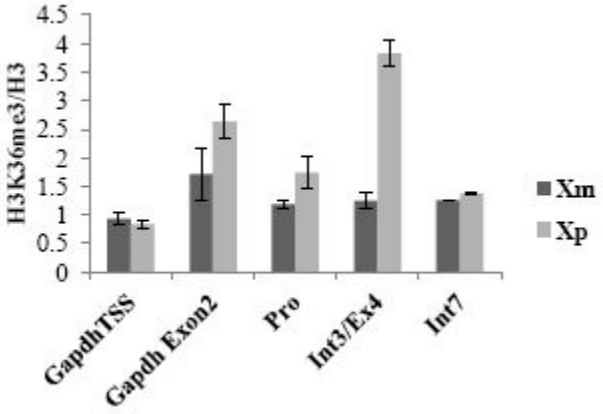


Figure 26. Real-Time PCR of H3K36me3 ChIP Enrichment at *Xlr3b*. Graph shows the paternal allele is 3 to 4 fold more enriched for H3K36me3 than the maternal allele at intron3/exon4 (Int3/Ex4). Enrichment at all regions is relative to total histone H3 ChIP and normalized to the *Gapdh* TSS. Figure courtesy of Sohaib Qureshi.

The only other chromatin difference observed between the maternal and paternal alleles of *Xlr3b* is an increase of H3K36me3 on the paternal allele at the Intron3/Exon 4 boundary (Figure 26) while H3K4me3, H3k9me2, and H3K27me3 remained equally enriched at all regions (data not shown). This mark could be a result of an increase of stalled RNA PolII on the paternal allele, which in turn increases the recruitment of SETD2, the H3K36 trimethylase enzyme that works co-transcriptionally with RNA PolII. While these results show how the imprint is manifesting during transcription it doesn't make clear the cause of the RNA Pol II stall. The next chapter shows that a closer look into the proximal gene, *Factor 8 associated gene (F8a),* reveals regulatory elements consistent with other imprinted loci and leads to a model of *Xlr3/4* imprinting.

Chapter 5: Allele Specific Differences at the Imprinted *Xlr3/4/5* Locus

Factor 8 associated gene A (F8a)

Internal to the imprinted *Xlr* paralogs exists a biallelically expressed gene, *Factor 8 associated gene A (F8a)* (Figure 27). This gene has one exon (~2.5Kb) and in humans it is located inside intron 22 of coagulation *Factor 8 (F8)*. While *F8* is involved in blood clotting, *F8a* has been shown to code for a 40kDa protein termed HAP40, or Huntingtin-associated protein 40, which co-purifies with the protein product involved in Huntington's disease termed Huntingtin (HTT).[196] It is suggested that *F8a* is necessary for proper function of HTT. As seen below *F8a* is single copy in the mouse genome but has 3 copies (F8a1, F8a2, F8a3) in the human genome, one residing in intron 22 of *F8* (Figure 28).

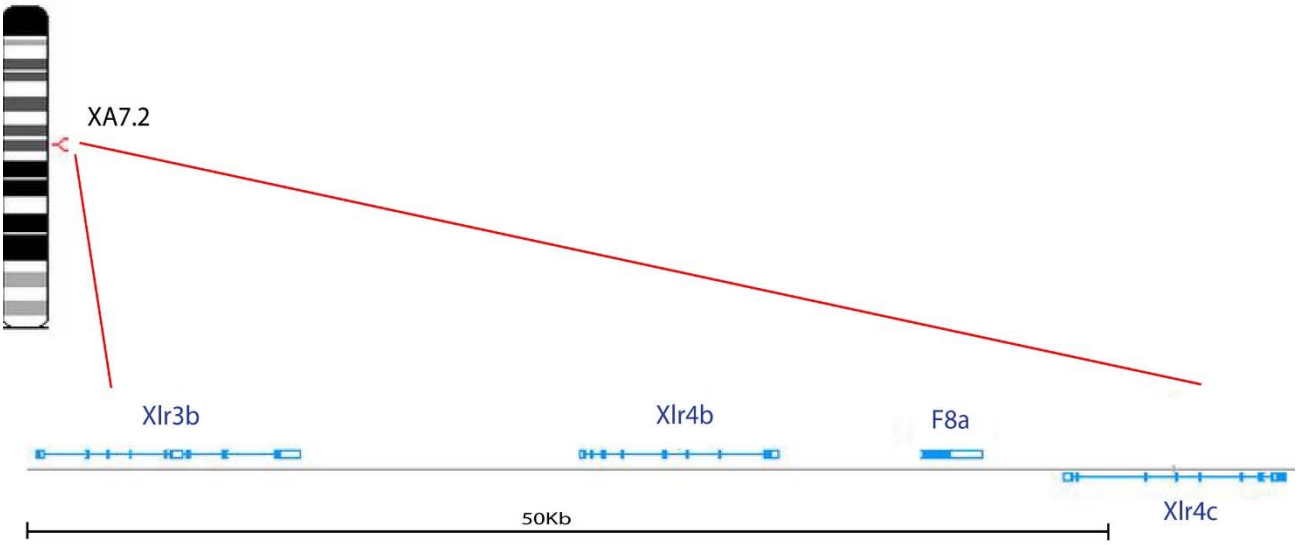


Figure 27. Cytogenetic Location of *F8a* in Mouse. Genomic representation of single copy F8a and adjacent imprinted genes.

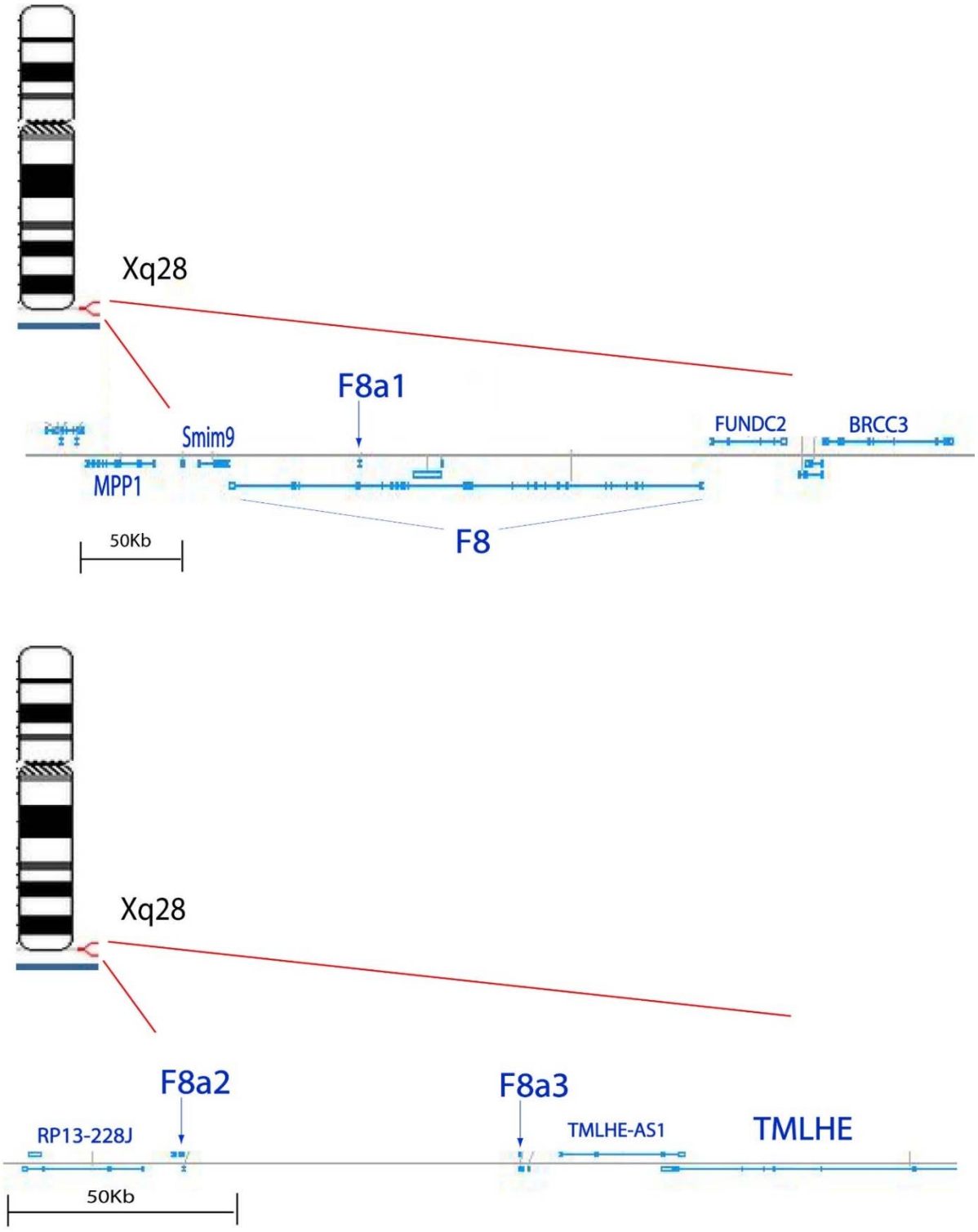


Figure 28. Cytogenetic Location of *F8a* in Human. F8a is multi-copy in humans (F8a1, F8a2, and F8a3) all found near the telomere.

Interest in relating *F8a* to the *Xlr3b/4b/4c* imprinting mechanism came when ENCODE released ChIP-seq tracks of transcription factor binding locations revealing prominent enrichment of the insulating protein CTCF in the gene body of *F8a* in several tissues and cell lines from mouse. As discussed earlier, CTCF has been implicated in the regulation of gene expression and in particular imprinted expression at the *Igf2/H19* locus. CTCF directs chromatin loop formation to promote enhancer/promoter interactions and acts in insulating key regions. Interestingly, p300, a mark for active enhancers, is seen flanking the main CTCF peak in F8a, as shown in Figure 29 below.

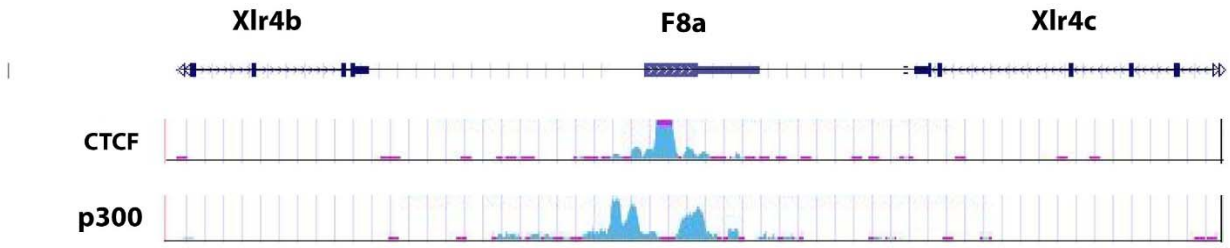


Figure 29. Pile-up of CTCF and p300 Tracks from ENCODE. CTCF insulator and p300 enhancer marks at F8a - figure adapted from ENCODE.

Internal to *F8a,* where the CTCF peak resides is a large CpG island which extends for the first half of the F8a coding region. Figure 30 below illustrates the location of this CpG island within F8a and displays the location of consensus binding sequences for CTCF, YY1, and ZFP57.

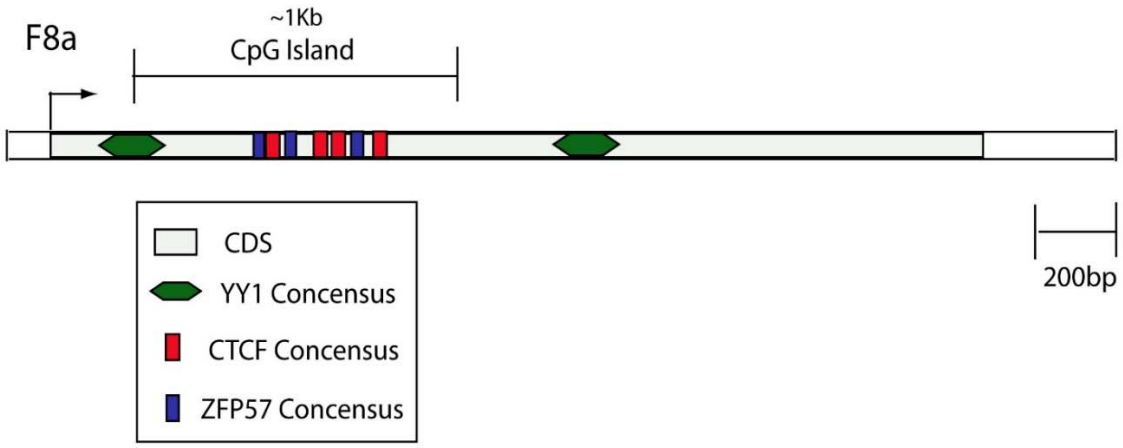


Figure 30. Illustration of Consensus Sequence Locations in F8a. Indicators of potential regulation and looping exist at the 5' end of F8a.

The numerous indicators of looping interactions sitting in the middle of the imprinted *Xlr* paralogs, compelled a series of experiments to determine if these chromatin features may be involved in the imprinted expression of the neighboring paralogs. Since DNA methylation is the differentiating hallmark of ICR's, this was the first epigenetic mark to be investigated in F8a. In order to do so, mouse models were used to produce X-monosomic mice with either the maternal or paternal X chromosome (Figure 31). In this breeding scheme, germ cells carrying the In(x) or Paf alleles will undergo non-disjunction during meiosis resulting in the passing of a single wildtype C57 X chromosome.

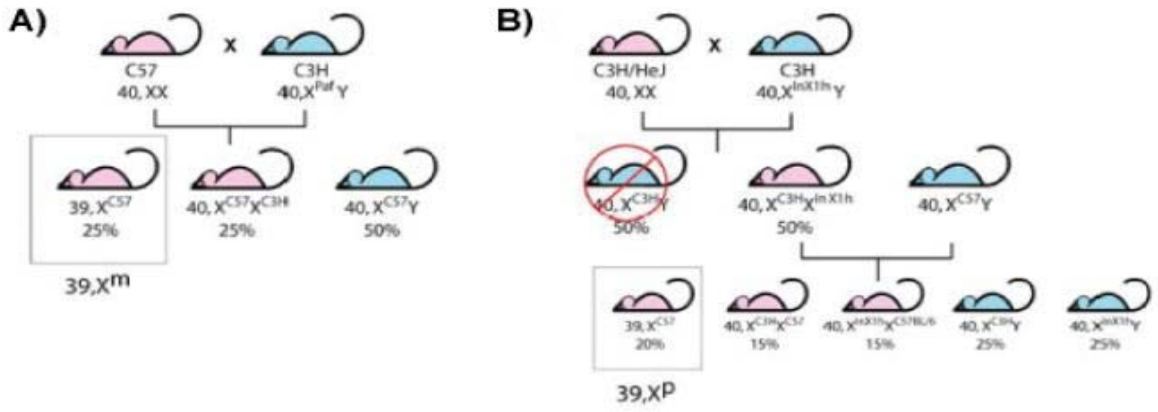


Figure 31. Breeding Schemes for the Generation of X-monosomic Mice. Courtesy of Dr. Michael O'Neill.

These mice are viable and only show slight behavioral differences when compared to wild-type C57 mice.[197] Previous work in this mouse model by Raefski and O'Neill show *Xlr3b/4b/4c* to be imprinted in neonatal whole brain and neonatal fibroblast making them good target tissues for identification of a DMR.[8]

Differential Methylation at F8a

In a simple assay to test for a DMR, Xm and Xp neonatal whole brain DNA was digested with a methylation sensitive restriction enzyme (HpaII) and subject to PCR amplification. Existence of a PCR band suggests a methylation block of HpaII cutting while absence of a band represents complete cutting and therefore complete hypomethylation. In Figure 32 below, a clear difference can be seen at F8a, which has no methylation in Xp but at least some methylation in Xm.

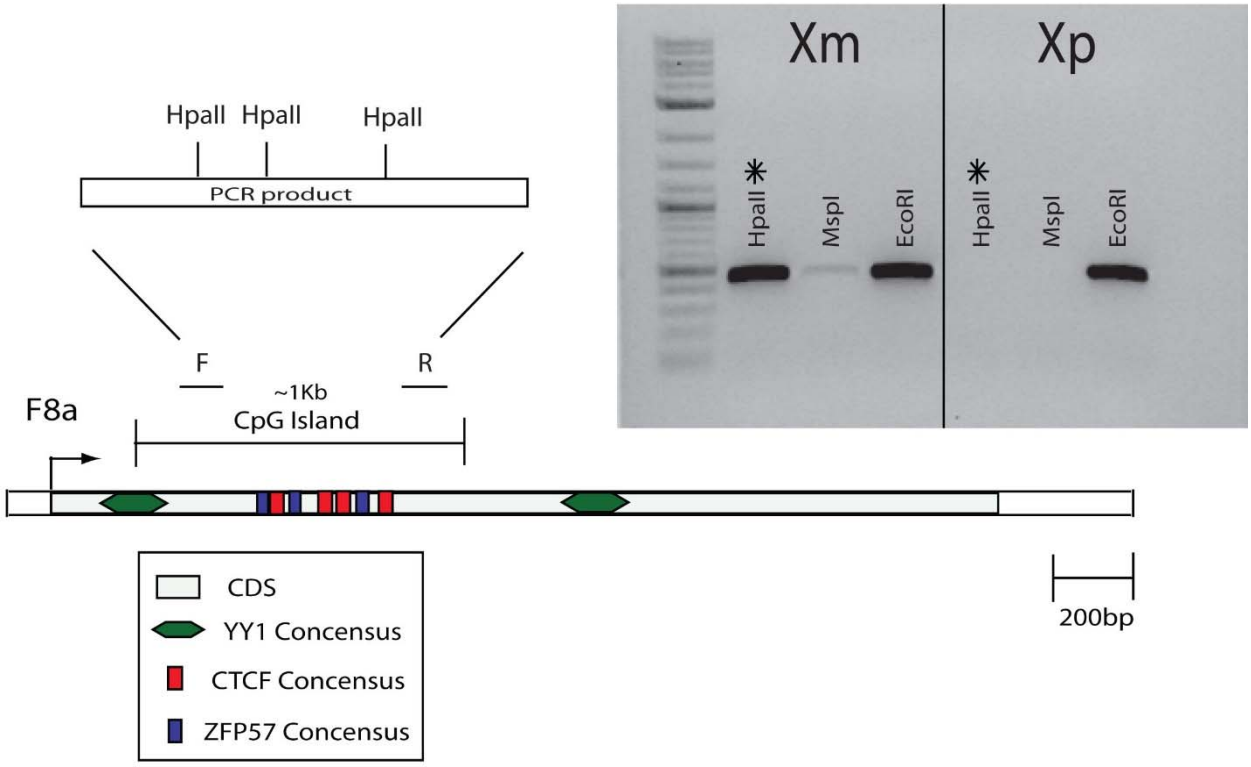


Figure 32. HpaII Methylation Assay at F8a. "F" and "R" mark the primers used to amplify the region containing 3 HpaII sites in neonatal whole brain. Successful cutting of one of these sites would be indicative of no band and therefore hypomethylation. Conversely, a positive band would indicate incomplete cutting and therefore hypermethylation.

In order to quantitate the difference in DNA methylation, the assay above was carried out using qRT-PCR. Ct values from the HpaII amplifications can be normalized to EcoRI CT values (which shouldn't digest the DNA at all) to determine percentage of methylation. Results from this normalization are shown in the bar graph below suggesting ~20% of the time F8a is methylated at this region on the maternal allele while paternal inheritance retains <1% methylation (Figure 33).

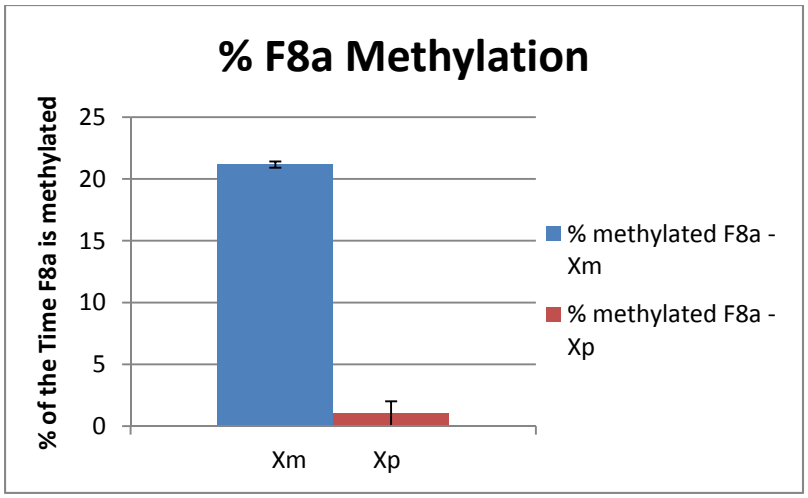


Figure 33. Quantitative RT-PCR Methylation Assay at F8a. Neonatal whole brain Xm and Xp genomic DNA was cut with HpaII or EcoRI and primers were used to amplify a region containing 3 HpaII and 0 EcoRI sites in F8a. HpaII Ct cycles (test) were normalized to the EcoRI Ct cycles (control) to determine % efficiency of cutting. Error bars represent 95% confidence.

To further assess and vet the extent of methylation differences seen between the Xm and Xp alleles, DNA from neonatal mouse whole brain and neonatal fibroblast were subject to sodium bisulfite sequencing. The results are illustrated below in "lollipop" format where filled in circles represent methylated CpGs and blank circles are unmethylated CpGs. The region assayed spans the ZFP57 and CTCF sites shown in Figure 32 on the previous page.

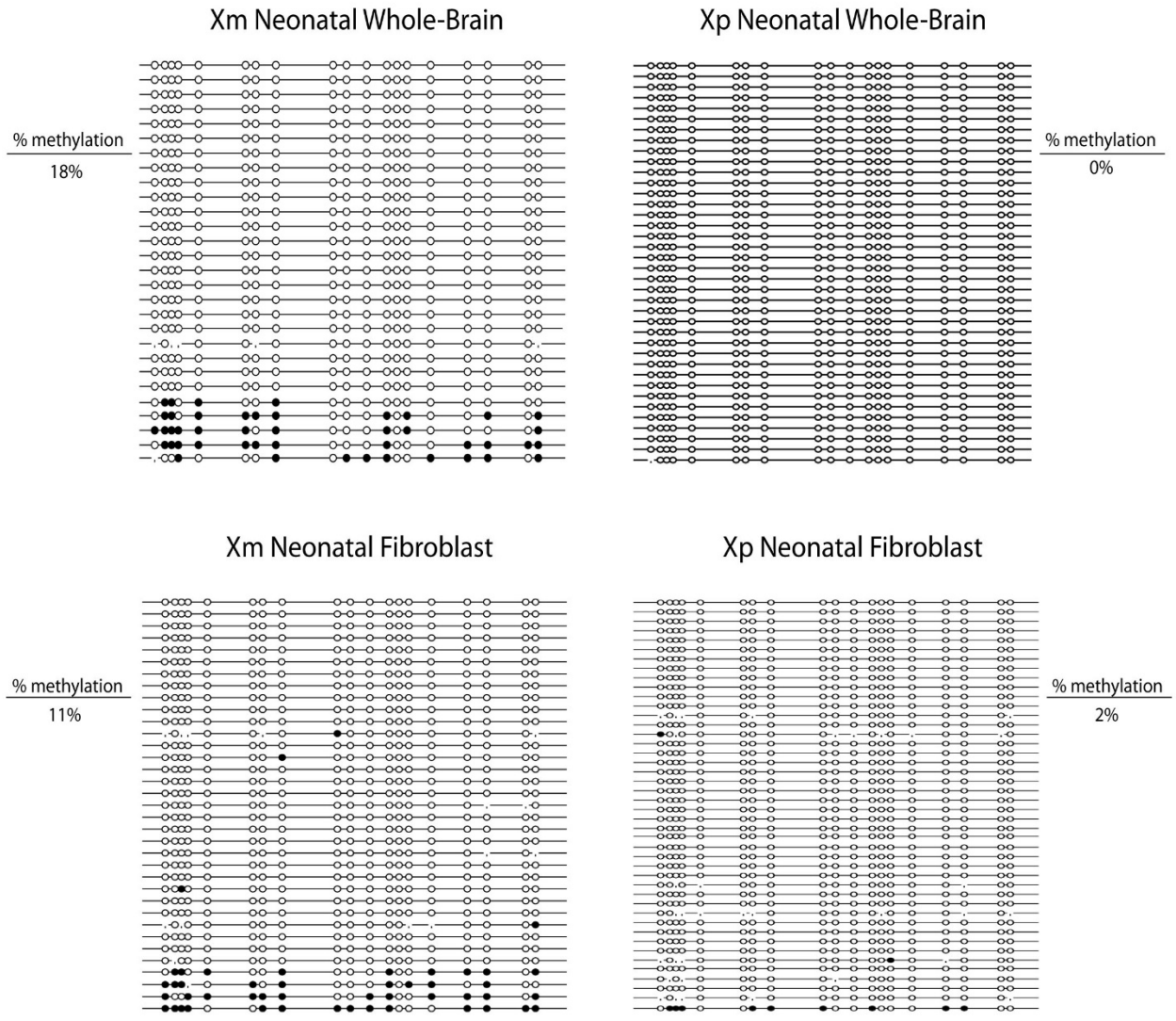


Figure 34. Bisulfite Sequencing of Xm and Xp Neonatal Brain and Fibroblast. CpG "lollipop" diagram showing percentage of clones that were methylated in monosomic Xm and Xp neonatal whole brain and fibroblast. Filled in circles represent methylated CpG dinucleotides while blank circles are unmethylated CpGs.

Immediately it is obvious that F8a is hypomethylated on both alleles consistent with qRT-PCR results. The paternal allele appears to be completely hypomethylated in neonatal whole brain; however, the maternal allele retains a fraction of methylation (18%) consistent with the qRT-PCR results (~20%). This general trend is also seen in neonatal fibroblasts, however, in both cases it is apparent that the difference in methylation is not extensive. In light of

parent/offspring conflict theory and the fact that many imprinted genes function during embryogenesis, it is possible that the imprint is being relieved by the time of birth. The assay above may be indicative of a relaxing methylation/imprint considering samples were taken at birth (P0). Earlier stages may reveal much more maternal methylation, while the paternal allele likely remains hypomethylated.

Differential CTCF Occupancy at F8a

These results raise the question - how is the DMR working to cause imprinted expression at *Xlr3b/4b/4c*? Methylation can be a large factor in regulating insulator binding as is the case with CTCF which can't bind methylated CpGs.[198] Interestingly, at least two groups have suggested that CTCF binding is actually the cause of differential methylation as opposed to being the consequence.[199,200] To look for differential enrichment of CTCF at the DMR in F8a, chromatin immunoprecipitation (ChIP) using antibodies to CTCF were carried out and analyzed for presence of F8a. The results summarized in the bar graph below show significant enrichment of CTCF on the paternal allele of F8a consistent with complete hypomethylation for CTCF binding (Figure 35). The maternal allele, however, shows reduced CTCF binding across the two regions assayed, presumably because of increased methylation of F8a on this allele.

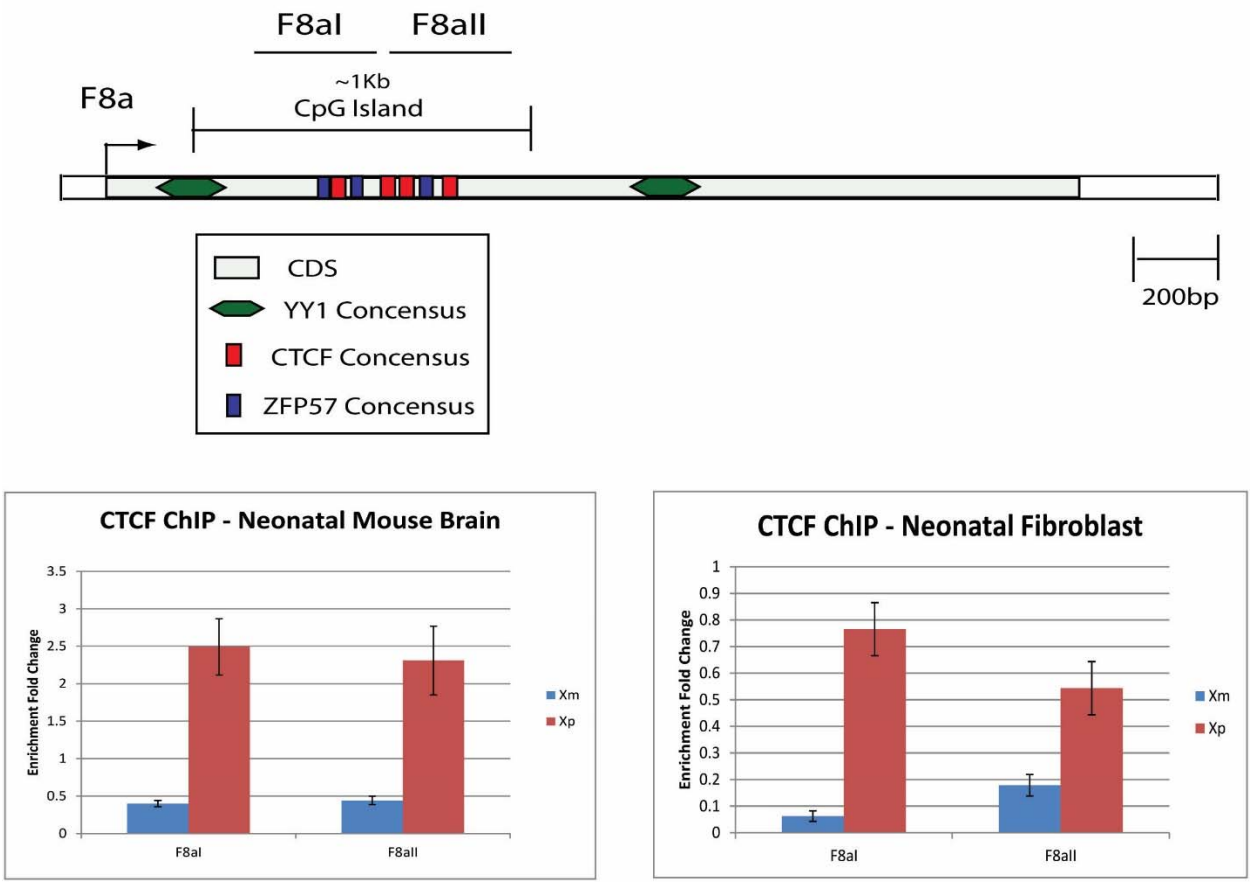


Figure 35. Enrichment of CTCF at F8a. ChIP of CTCF at primer locations F8aI and F8aII (locations depicted in illustration above graph) showing significant enrichment in Xp when normalized to CTCF bound H19.

The differential CTCF bound chromatin along with differential methylation fits a traditional model of methylation dependent binding of CTCF as seen from the Igf2/H19 imprinting mechanism. This region in F8a has many attributes consistent with loop forming interactions that could be involved in silencing genes in the same neighborhood as F8a, namely *Xlr3b/4b/4c*. In this respect, F8a remains CTCF bound on the paternally imprinted allele while the maternal allele allows for at least some methylation, hindering CTCF, and ultimately bringing about activation of maternal *Xlr3b/4b/4c*. Additional ChIP experiments at the same regions within F8a

show higher enrichment of the Methyl-CpG binding protein, MeCP2, in Xm neonatal whole brain (Figure 36). This result might be expected due to the fact that the maternal allele shows an increase in methylation (20%) relative to the paternal allele (<1%) and MeCP2 is known to bind methylated regions genome-wide. In the previous chapter it was discussed that MeCP2 can recruit ATRX which is able to help deposit H3.3 within certain areas of G-quadruplex formation. If G-quadruplex formation in *Xlr3b/4b/4c* are the cause of an RNA Polymerase II stall, then it stands to reason that MeCP2 may be recruiting ATRX to deposit H3.3 in *Xlr3b/4b/4c* to relieve G-quadruplex formation allowing maternal Pol II read-through. It would be interesting to reveal recruitment of ATRX to the locus in an allele specific manner – this would explain maternal allele activation of the surrounding imprinted Xlrs.

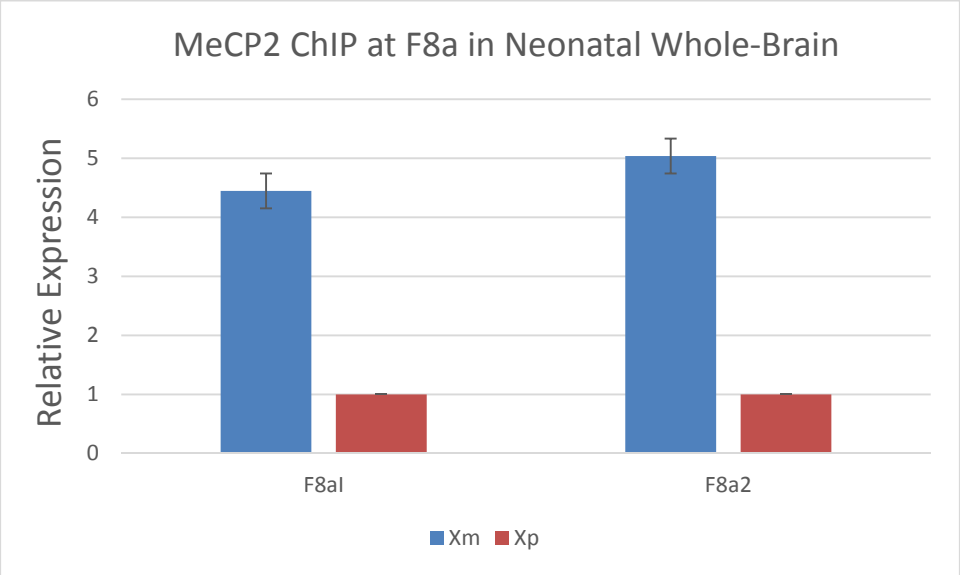
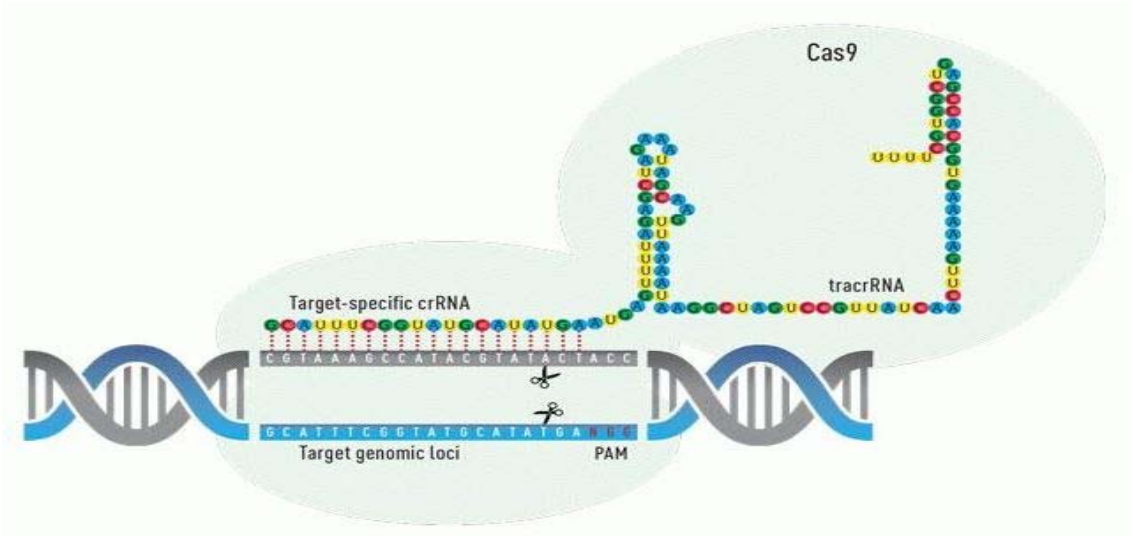


Figure 36. Enrichment of MeCP2 at F8a. ChIP of MeCP2 at primer locations F8aI and F8aII showing significant enrichment in Xm neonatal whole brain relative to Xp when normalized to MeCP2 bound H19.

Deletion of the Putative ICR in F8a

In order to conclusively determine whether the CpG island in F8a is an imprinting control region, a deletion strategy was pursued using CRISPR-Cas9 technology. In this scheme two plasmids are used, each with a unique "guide RNA" that is homologous to a region of the genome. These guide RNAs are cloned 5' of a "tracr RNA" which is recognized by the Cas9 protein (Figure 37A). The Cas9 protein is therefore targeted to the tracr and guide RNAs and cuts any proximal PAM sequence (NGG) making a double stranded break. In our case, 2 cuts were made using 2 different guide RNAs that flank the CpG island in F8a (plasmids seen in Appendix: *Cas9+OFP+*). To correct the two double strand breaks, the cell will activate the non-homologous end joining (NHEJ) DNA repair pathway causing a deletion of everything between the two cut sites (Figure 37B).

A



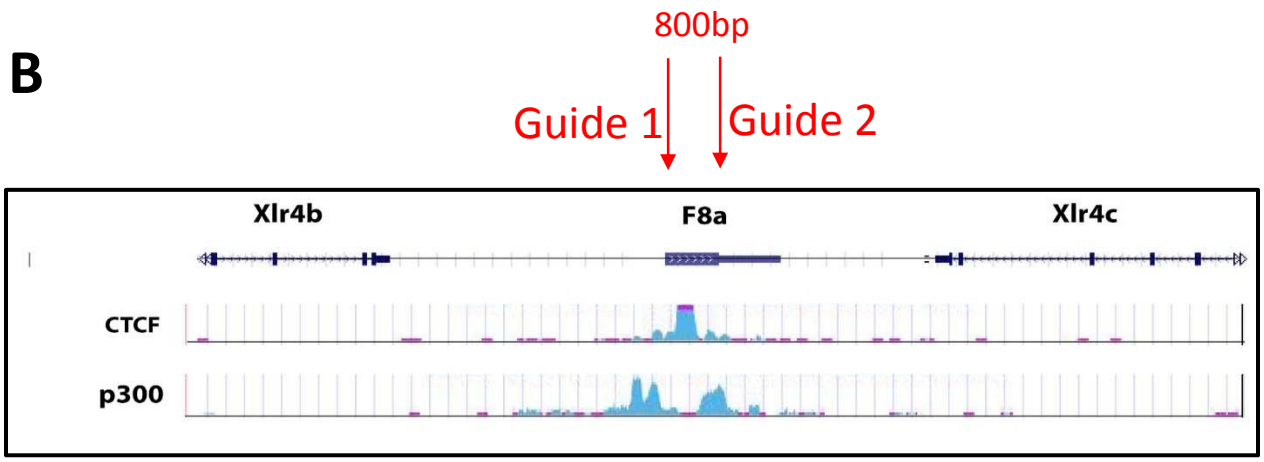
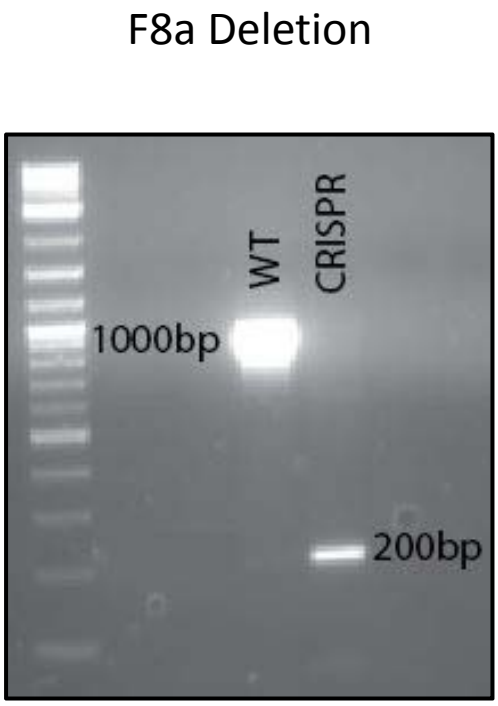


Figure 37. CRISPR/Cas9 Scheme Targeting F8a. **A.)** Depiction of CRISPR/Cas9 mechanism for targeted deletions. **B.)** Location of targeted guide RNAs for deletion of the CpG island in F8a.

Neonatal male (XY) fibroblasts were cultured for transfection of the CRISPR deletion vectors in order to remove the CpG island in F8a to assess its effect on *Xlr3b* expression status. Male mice have the Xm and hence have active copies of *Xlr3b/4b/*4c. Deletion of F8a would be expected to either silence *Xlr3b* or upregulate it if F8a acts as a local ICR (discussed further in Synthesis chapter). Figure 38A shows successful deletion of F8a in CRISPR transfected cells compared to WT conditions. The high deletion efficiency was achieved through FACS sorting of the cells expressing Orange Fluorescent Protein originating from the CRISPR/Cas9 plasmids. Cells were allowed to grow for 12 days post-FACS sort and then analyzed for differences in *Xlr3b* expression relative to control cells. Unfortunately, significant expression differences are not apparent in this deletion assessment (Figure 38B). No significant change was seen when F8a was deleted in Xm fibroblasts (Figure 38C) or Xp fibroblasts (data not shown). There are a multitude of reasons an expression change wasn't achieved and I remain hopeful that this region is not only a DMR, but also an ICR. Perhaps the imprinting mechanism at F8a is set up early in embryogenesis, and once established, deletion of F8a can no longer have an effect on

the imprinted paralogs. To combat this, an F8a knockout mouse should be generated and can be bred into the X-monosomic breeding scheme to generate F8a deletions in Xm and Xp mice. This would allow analysis of the DMR's effects throughout embryogenesis. Also, with methylation of maternal F8a tracking with active *Xlr3b/4b/4c,* a more appropriate experiment may be to methylate F8a via methyltransferase targeting to look for activation of the paralogs around it.

A



B

C

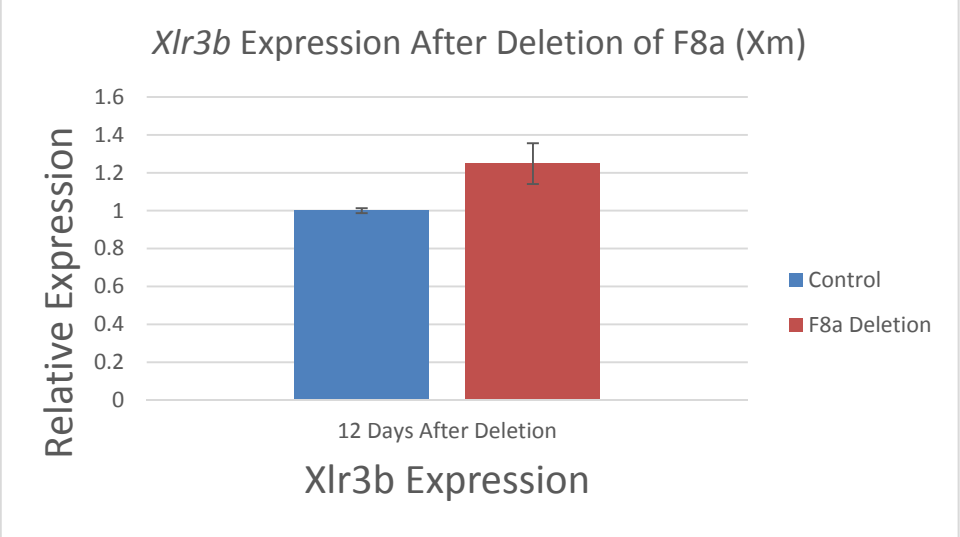


Figure 38. CRISPR/Cas9 Deletion in XY Fibroblasts **A.)** Gel image showing successfully 800bp deletion of the CpG island in F8a compared to control cells. **B.)** Displays a lack of significant expression difference in *Xlr3b* across control and F8a deleted XY fibroblast cells. **C.)** Displays a lack of significant expression difference in *Xlr3b* across control and F8a deleted Xm fibroblast cells.

Chapter 6: Synthesis and Future Directions

The initial aim of this dissertation was to elucidate the protein functionality of genes within the *Xlr* locus with the purpose of revealing a general reason for their existence and massive expansion in rodentia. Concurrently, efforts were made to determine how, mechanistically, the *Xlrs* are subject to parent specific gene expression as was discovered by Dr. Adam Raefski in 2005.[8] The wealth of knowledge on genes displaying parental specific behavior (genomic imprinting) implicates many in parent-offspring conflict, a theory in which gametes predispose certain developmental genes to be silent or expressed with the intention of increasing individual fitness in the offspring. *Insulin-like growth factor II* (*Igf2*) is a hallmark example of this type of regulation, relying on one silent and one expressed allele for proper dosage of this growth hormone in offspring.[81]

What about the *Xlrs*? One third of the active copies in the *Xlr* locus are paternally silenced (*Xlr3b, Xlr4b, and Xlr4c*) but remain maternally expressed. Additionally, these paralogs are members of a larger group of similar looking genes which have duplicated to form hundreds of copies on both the X and Y chromosomes in rodentia, begging the question – What developmental role are these multi-copy genes playing that relies on so many copies and additional levels of transcriptional regulation?

Dr. Paul Burgoyne and others have shown at least 4 of these *Xlr* superfamily proteins (SLX, SLXL1, SLY, XLR6/SLX2) to have high expression throughout spermatogenesis/spermiogenesis similar to their autosomal ancestor, SYCP3.[3,35,44] Extensive examination into the potential mode of function reveals close linkage of these proteins to

expression emanating from the sex chromosomes throughout meiosis in male mice, even potentially acting in an inter-genomic battle for inheritance; X vs Y.[4,5,28] Conflict mechanisms have already been shown to exist in mammals as seen with t-haplotypes in mice and could explain the large build-up of these antagonistic genes analogous to an X vs Y "arms-race."[57] It makes sense that proteins involved in such a mechanism act throughout meiosis, a critical time in which inheritance to the next generation is eminent.

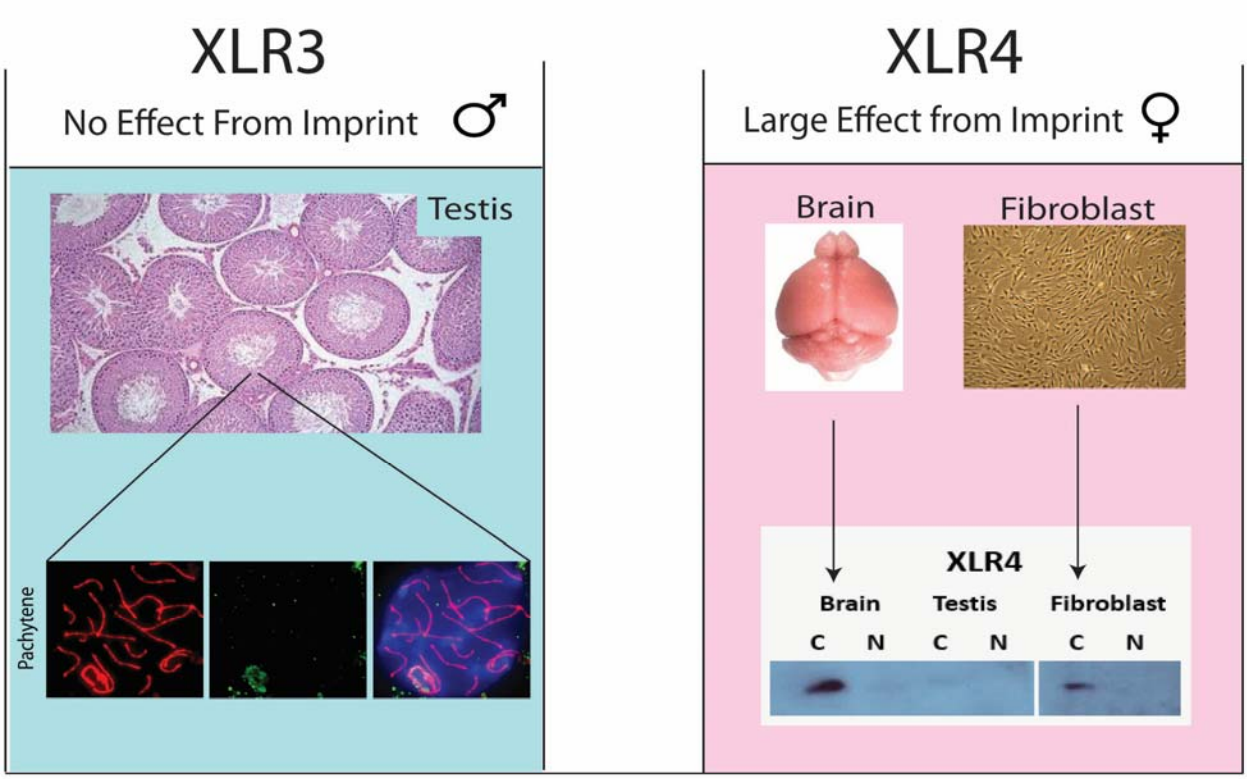
Dr. Michael O'Neill's lab attempted to determine protein functionality of the imprinted *Xlr3b* paralog by designing an antibody initially to be used for probing lysates from brain and fibroblast, the two tissues in which the imprint was known to manifest. It was therefore somewhat surprising not to reveal a band in these tissues but rather see a prominent band in nuclear testis lysate. However, taking into account the expression profile of other *Xlr* superfamily proteins, it soon became clear that XLR3 was also a meiotically expressed protein. Investigation into localization revealed XLR3 associates closely with the synaptonemal complex exclusively at the inactive X and Y chromosomes in pachytene spermatocytes. Additional work would show close association of XLR3 with the repressed X and Y chromosomes in post-meiotic round spermatids adding it to the growing list of X and Y linked genes with similar behavior. To confirm its function, an *Xlr3* shRNA knockdown transgenic mouse model was produced to eliminate XLR3 from meiosis in order to determine the exact effect it will have on gene expression emanating from the sex chromosomes. It is ultimately proposed that XLR3 is controlling gene expression from the X and Y chromosomes globally during meiosis, perhaps through interaction with the synaptonemal complex.

This protein expression profile seems to further convolute the active imprinting of *Xlr3b*. The problem at hand is that the tissue in which XLR3 functions, testis, is genotypically X_mY meaning no imprint can manifest – so why imprint a protein that won't feel the effects? Female tissue would be directly affected by an imprint, and in fact, it was initially thought that X-linked imprinted genes couldn't exist due to absent expression entirely in half of female cells - a consequence of X-chromosome inactivation (XCI). Interestingly, XLR3 protein was shown in this study to also localize in meiotic oocytes raising the theory that with 1 copy silent and 5 copies active, perhaps dosage of the protein is the key reason for imprint. Duplication of multiple paralogs is one way of ensuring "back-up" copies in the event one copy gains deleterious mutations or becomes imprinted - further bolstering a dosage necessity.

I propose taking another perspective on the matter - *Xlr4b and Xlr4c* are also both imprinted paternally. This represents one third of the active Xlr4 copies in females and two thirds of the active copies in males, which most likely results in a severe down regulation of XLR4 protein. In this study, a custom XLR4 antibody was generated revealing prominent bands in the cytoplasm of neonatal whole brain and fibroblast tissues, again, the two tissues that manifest the imprint. With imprinted expression of *Xlr4b/c* and translation of XLR4 protein both occurring in brain and fibroblast it seems likely that the targeting of the imprint is intended for the Xlr4 paralogs. This would indicate that the imprinting mechanism works within a genomic "neighborhood" which may accidently silence 1 of the *Xlr3* copies along with the 2 intended *Xlr4* copies. This may not affect XLR3 protein which has 5 more copies that can express to keep this protein available to the oocyte and has no effect on the 3 active copies in spermatocytes. Figure 39 below summarizes this conundrum.

While the imprint may not have directly evolved to regulate XLR3 protein, the results of the XLR3 protein analysis remains significant. Although there is no direct human ortholog to XLR3, there does exist a distantly homologous locus sharing ~8% identity to XLR3, termed FAM9A, FAM9B, and FAM9C. These proteins have not been studied extensively, however, a quote from the only paper that exists on these human homologs reveals many similarities between XLR3 and FAM9. It states, "The deduced amino acid sequence of the three novel proteins [FAM9A/B/C] shows homology to SYCP3, a component of the synaptonemal complex located along the paired chromosomes during meiosis. *FAM9A*, *FAM9B*, and *FAM9C* are expressed exclusively in testis; their proteins are located in the nucleus, and FAM9A localizes to the nucleolus."[201] Suffice it to say there are many similarities between this localization of FAM9A/B/C in humans and XLR3A/B/C in mice.

Very recently, the existence of an XLR5 protein product was found and it appears to be upregulated during prophase I of meiosis as well. It is seen localizing with the synaptonemal complex from leptotene through pachytene.[202]



Tissue	Xlr Superfamily Protein Involved
Testis	SLY, SLX, SLXL1, XLR5, XLR6/SLX2, XLR3, SYCP3
Ovary	XLR3, XLR, XLR6/SLX2, SYCP3
Brain/ Fibroblast	XLR4

Figure 39. XLR3/XLR4 Conundrum. Depiction of the effect imprinting has on the protein coding genes at the *Xlr* locus. XLR3 is unaffected by the imprint in males where it localizes to the inactive XY chromosome and only slightly affected in oocytes where 5 duplicate copies exist. A potentially large effect would manifest in tissues where the imprint is severe and where a protein is translated (i.e. XLR4 in brain and fibroblast).

In keeping with the theme of a "neighborhood" imprint it is possible that the mechanism silencing *Xlr3b/4b/4c* is acting regionally at XA7.2 rather than targeting individual genes. Imprinted genes in general are known to occur in clusters and are typically subject to the same mode of silencing or activation mechanisms as a whole. This theory breaks down slightly when considering the bi-allelically expressed *Factor 8-associated A* (*F8a*) which sits in the middle of the imprinted *Xlr* paralogs, however, it does make *F8a* an interesting candidate for being the imprinting controlling region (ICR) at this locus. Interest in F8a came from ChIP-seq data on ENCODE showing a relatively large build-up of CTCF at the 5' end of *F8a* coincident with a CpG island. This peak is dramatic because it is the only one seen in a 200kb span across the locus. CTCF is a master regulator protein of chromatin conformation as well as an insulator protein implicated in a long list of looping mechanisms some of which are involved in genomic imprinting.[143] This new target inspired an in-depth search within *F8a* for a differentially methylated region (DMR), which would force differential binding of CTCF and with it the foundational evidence for a local ICR.

Dr. Ben Carone initially searched for allele-specific differential methylation in a locus-wide CpG island analysis revealing no DMR but instead extreme hypermethylation of all islands except the hypomethylated one inside F8a. Our methylation study re-analyzed F8a for differential methylation through use of HpaII methyl sensitive restriction enzyme digests/PCR revealing a clear allelic difference. This difference was quantified through using qRT-PCR suggesting the region gained methylation in Xm brain and fibroblast at a rate of ~20% while Xp tissues revealed a 0-2% gain in methylation. Bisulfite sequencing of these two tissues would reveal consistent percentages of methylation and with it a bonafide DMR. While this is

admittedly not a strong methylation difference we wanted to determine whether it could affect allelic enrichment of CTCF binding at this region. Allele specific ChIP of CTCF was carried out in the same tissues resulting in a 5-fold enrichment of CTCF occupancy on the paternal allele of *F8a* vs the maternal allele. This fits with the theory that 20% methylation of *F8a* on the maternal allele is "blocking" CTCF from binding while bringing about a higher occupancy on the hypomethylated paternal allele. The retained methylation on the maternal allele is proposed to be necessary to activate expression of neighboring genes while complete paternal hypomethylation results in no expression of neighboring genes. It should be noted that while CTCF appears to have a reverse correlation with methylation, it is still likely secondary and not causing silencing but rather recruited to ensure silencing.

Consensus sequences for CTCF as well as YY1 (tethering protein) are found within this CpG island in F8a bolstering a potential looping or insulating mechanism to regulate the Xlrs. Additionally, three ZFP57 consensus sites, a maternal mark preserving methylation at all known ICRs, is seen within the *F8a* CpG island.[172,203] We proposed to delete this F8a region via CRISPR/Cas9 technology in hopes of activating paternal *Xlr3b/4b/4c* or silence the maternal *Xlr3b/4b/4c*. Unfortunately, deletion of this region in neonatal fibroblast cells revealed no change in gene expression of the surrounding *Xlrs*. It seems a more appropriate experiment would be to extensively methylated F8a on the paternal allele due to the fact that methylation of F8a correlates with activation of the surrounding *Xlrs* as seen from the methylated maternal allele of *F8a* and the expressed *Xlrs* around it. This can potentially be accomplished through TAL or CRISPR targeting of a methyltransferase specifically to F8a. The mechanisms shown in Figure 40 provides a potential mode for silencing the paternal allele through CTCF binding and

activation of the maternal allele through methylation retention. In this model, 20% of Xm derived cells retain methylation (likely through ZFP57 protection) of the CpG island in *F8a* during development which hinder CTCF binding and abolishes CTCF insulator activity. The methylated CpGs in *F8a* are therefore able to recruit a methyl-binding domain protein MeCP2. MeCP2 has recently been shown to recruit ATRX a DNA helicase with chromatin remodeling abilities.[171] ATRX is known to aid in deposition of the histone variant H3.3 at telemetric sequences to allow for stabilization of regions with high G-quadruplex complexes.[169] G-quadruplexes result when 4 di-guanosine residues associate with each other in a complex when the DNA is opened up for RNA PolII transcription.[204] H3.3 deposition prevents these complexes from forming presumably allowing PolII to successfully transit the region. Therefore the methylated maternal *F8a* allele recruits MeCP2 which in turn recruits ATRX which is able to deposit H3.3 at points of G-quadruplex formation in *Xlr3b/4b/4c* allowing RNA PolII to read through to the end of the gene. All cells in Xp derived tissues retain hypomethylation of *F8a* and therefore CTCF insulates the region to ensure no methylation can occur. With blockage of methylation on the paternal allele of *F8a*, factors such as MeCP2 and ATRX can't be recruited and G-quadruplex formation in the *Xlrs* may be causing an RNA PolII stall in the absence of H3.3 deposition. It should be noted that Xm tissues also display hypomethylation of *F8a* 80% of the time which may reflect the imprint only manifesting in a certain cell type or possibly F8a is acting as a metastable epi-allele, meaning methylation is dependent on environmental effects on the cell.[205] Additionally, metastable epi-alleles gain methylation marks prior to gastrulation and not in the sperm and egg which fits with publications showing no significant methylation of *F8a* in gametes but increased (~10-40%) methylation from blastocyst onwards.[206]

Figure 41 below was generated from the following website (http://bioinformatics.ramapo.edu/QGRS/analyze.php) which predicts G-quadruplex formation in genes. Intron 6 in *Xlr3b/4b/4c* is a strong candidate for G4 formation as seen by the intensity of the peaks in that region. It should be noted that while *Xlr3b* only retains about ~35% homology to *Xlr4b/4c*, this particular region in Intron 6 is 99% identical at the nucleotide level between the paralogs, and therefore acts as a common region for G4 formation in all imprinted *Xlrs*. *F8a* may be able to relieve this affect in the 3 adjacent paralogs due to its proximity to *Xlr3b/4b/4c*. This region is also consistent with where the RNA PolII stall may be occurring. Data from Dr. Sohaib Qureshi and Dr. Seth Kasowitz reveals a potential RNA PolII stall somewhere between Exon 3 and Intron 7 of *Xlr3b* making this Intron 6 region a strong candidate. One can't ignore the strong build-up of G4 formation at the 5' end of *Xlr3b/4b/4c* genes. This is in fact a common feature seen at the origins of most genes; however, according to Bartholdy et al there is no correlation between G4 forming sequences at promoters and initiation efficiency.[166]

Interestingly, *F8a* is well conserved throughout eutherian mammals as well as marsupials even existing in 3 copies on the human X chromosome. While the protein product of *F8a* (HAP40) is involved in binding the Huntingtin protein (Huntington's disease) it is also possible that the CpG island in F8a is acting as a control region for surrounding genes in human as well.[196] One of these *F8a* copies sits directly adjacent to *Trimethyllysine Hydroxylase Epsilon* (*TMLHE*) a gene implicated in autism spectrum disorder (ASD). *TMLHE* acts in the first steps of carnitine biosynthesis, which is essential for fatty acid transport into the mitochondria. It stands to reason that if *F8a* is able to control surrounding gene expression profiles, then perhaps mis-

regulation of *TMLHE* can lead to increased oxidative stress in the brain at critical time points during development, ultimately caused by F8a methylation status.[207–209]

In conclusion, *F8a* is a strong candidate for being the imprinting control region for *Xlr3b, Xlr4b, and Xlr4c*. It should be noted, however, that this DMR does not act like a typical DMR in that it doesn't display complete allelic hypermethylation vs hypomethylation. Instead, F8a appears to retain a certain percentage of maternal methylation at F8a on a cell by cell basis, potentially based on environmental cues.

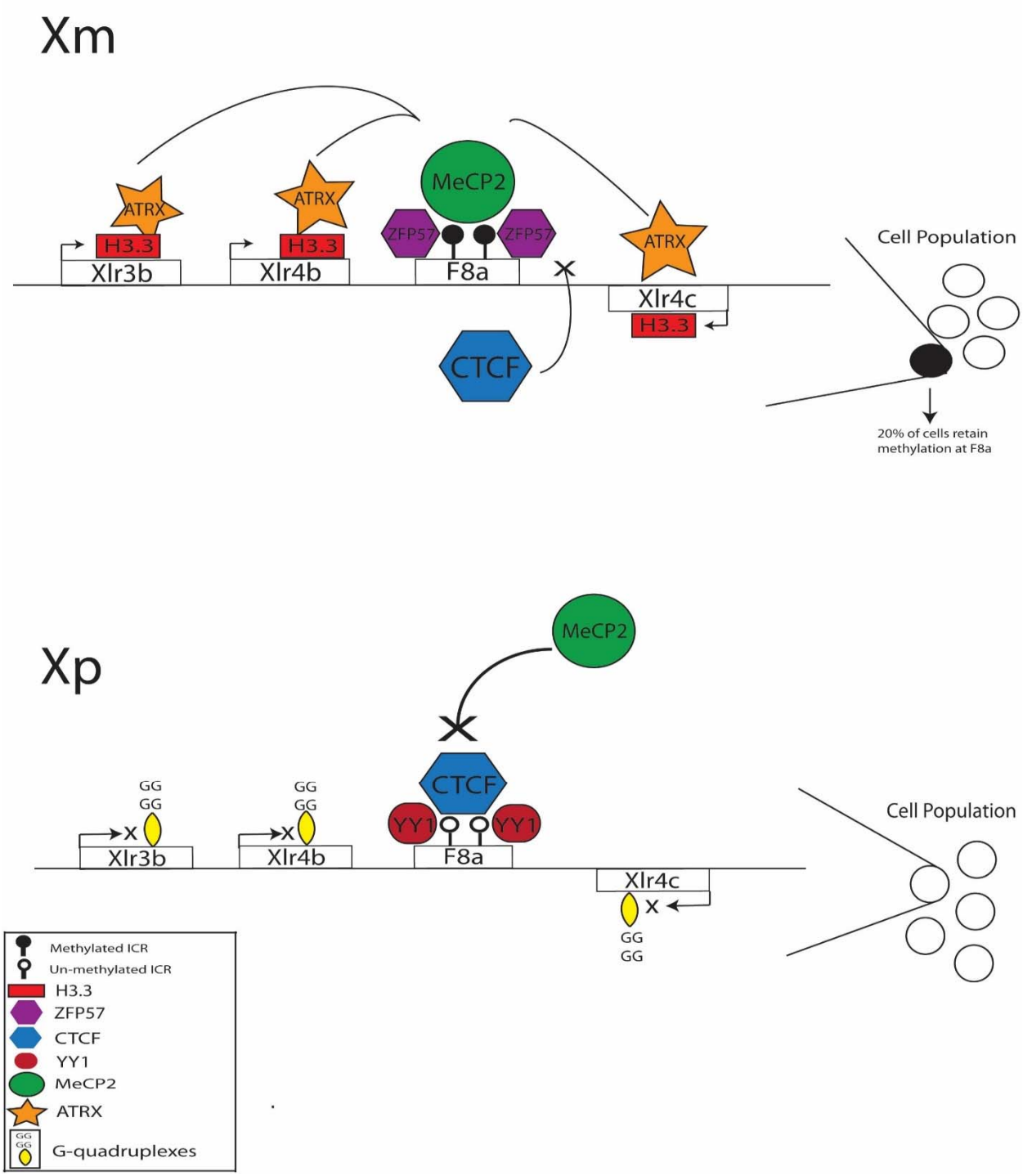


Figure 40. Model of the *Xlr3b/4b/4c* Imprinting Mechanism. In XmO neonatal brain and fibroblast methylation is gained in 20% of cells blocking CTCF binding and allowing recruitment of MeCP2 as well as ATRX to deposit H3.3 to allow PolII to read-through G-quadruplex forming regions. Lack of methylation on the paternal allele causes insulation of CTCF and therefore no mechanism to relieve G-quadruplexes which ultimately causes a PolII stall.

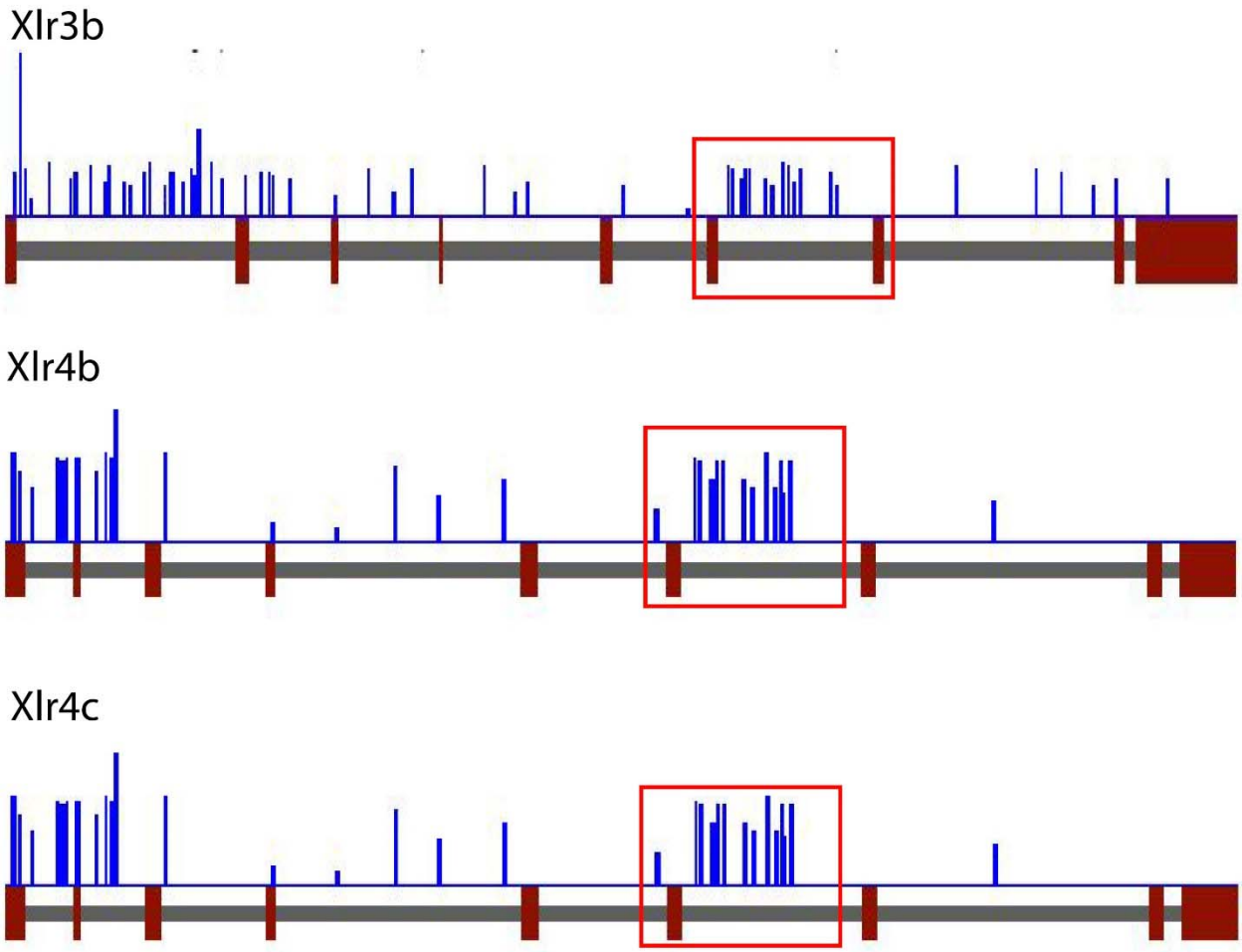


Figure 41. Prediction of G-quadruplexes in *Xlr3b/4b/4c*. Illustration from a g-quadruplex prediction software indicating G4 complexes potentially present in Intron 6 of all imprinted paralogs.

Chapter 7: Materials and Methods, Appendices

Generation of X Monosomic Mice

39, Xm mice were produced by mating C57BL/6J females to C3H/Paf males which generates X(m) monosomic females at a rate of 25% in each litter due to non-disjunction of the Paf X chromosome. 39, Xp mice were produced by crossing C3H/InX1h females to C57BL/6J males which generates X(p) monosomic females at a rate of 20% in each litter also due to non-disjunction in meiosis of the InXh1 chromosome. Genotyping utilizes the DXMit130 PCR primers which are able to distinguish the C57 and C3H X chromosomes, as well as an Smcy primer set to distinguish the males.

Mice and Spermatogenesis Sample Collection

Animal procedures were conducted in accordance with the ethical guidelines and standards of the SSR (The Society for the Study of Reproduction). For sample collection, testis were dissected from C57BL/J6 or C3H/HeJ male mice (Jackson Labs) ranging from 5.5 to 18.5 days old. One teste from each mouse was kept for protein analysis and the other for RNA analysis.

Antibody Production and Western Blot Analysis

A peptide (MSSRKRKATDTAGRHSRM [amino acids 1-18]) was identified to be specific to XLR3A, XLR3B, and having 1 amino acid difference from XLR3C. Synthesis of peptide and immunizations were conducted by Biosynthesis Inc. Two rabbits were immunized followed by five boosters every two weeks as part of the First Class Polyclonal Antibody Package. Total IgG

sera from each mouse was affinity purified on an AminoLink (Thermo) resin column containing immobilized peptide (Biosynthesis Inc.). The antibodies were complemented with pre-immune sera from each rabbit.

Testis lysate was subject to a dounce homogenizer for separation of nuclear and cytoplasmic fractions in Harvest Buffer (10mM HEPES pH7.9, 25mM KCl, 2M Sucrose, 1mM EDTA, 0.5mM Spermidine, 1x Protease Inhibitor, 10% Glycerol). Lysates were then resuspended in 2x SDS sample buffer (2%SDS, 25% Glycerol, 0.06M Tris pH 6.8, 0.01% bromophenol blue, and 5% BME) and denatured for 10 min at 95°C. Ten microliters of sample was electrophoresed through a 12% SDS-polyacrylamide gel at 200V for 50 min. Protein was transferred to Immobilon-P PVDF membrane (millipore) at 15V for 30 min. The membrane was blocked (PBS containing 5% milk powder) for 1h at room temperature and then incubated with primary antibody diluted in blocking solution overnight at 4°C (rabbit anti-XLR3, 1:2000; mouse anti-γH2AX, 1:1000 [ab18311; abcam]; rabbit anti-αTUBULIN, 1:40,000 [ab15246; abcam]). Membranes were then washed 3 three times in PBS containing 0.1% Tween before incubating in the respective secondary antibody peroxidase conjugates for 45 min (anti-rabbit IgG HRP 1:2000 [ab6721; abcam]; anti-mouse IgG HRP 1:100 [32430; Thermo]). After washing the membrane three times (PBS containing 0.1% Tween), signal was detected via chemiluminescence (Western Lightening, Perkin Elmer) on autoradiography film.

Spermatocyte and Oocyte Surface-Spreads

Testis were dissected from three month old males and placed in PBS (MP Biomedicals) for removal of the tunica albuginea. Approximately 25mg of seminiferous tubule was

transferred to 1 mL of cold PBS in a petri dish, chopped up using a blade, and transferred to a round-bottomed tube. Five milliliters of fixative solution (2.6mM sucrose, 2% paraformaldehyde) were added to the cells drop by drop then tubes were swirled to mix. Cells were incubated at room temperature for 5 min and then centrifuged (1,200 rpm, 8 min). The fixative was removed and the cells were resuspended in twenty drops of PBS. Two drops were spread onto Superfrost Plus slides (Fisher) and air dried for 5 min. Cells were then permeabilized by adding 0.5% Triton X-100 onto the slides for 10 min. Slides were washed twice in PBS and stored at -20°C until further use. For oocyte spreads, ovaries were dissected from E15-18 dpc (days post coitum) embryos (~10 ovaries per prep) and prepared using the same protocol as spermatocytes.

Immunofluorescence and Chromosome Paints

Surface spread slides were blocked (5% goat serum in PBS) for 1h at 4°C. Slides were rinsed in PBS (MP Biomedicals) and then incubated in primary antibody overnight at 4°C (rabbit anti-XLR3 1:20; mouse anti-SYCP3 1:200 [sc74569; Santa Cruz]; mouse anti-γH2AX 1:200 [ab18311; abcam]). After three PBS washes, slides were incubated in fluorescent secondary antibodies diluted in PBS for 30 min at 4°C: Alexa Fluor 488-conjugated goat anti-rabbit 1:200 [A11034, Invitrogen]; Alexa Fluor 568-conjugated goat anti-mouse 1:200 [A11004, Invitrogen]. After washing three times in PBS, slides were dried and DAPI (4'6'-diamidino-2-phenylindole, Vector Labs) diluted in mounting media was applied using a cover slip. DAPI was not applied to slides subject to chromosome paints, instead these slides were fixed in 4% paraformaldehyde/0.5% Triton X-100 for 10 min followed by fixation in 3:1 methanol:acetic

acid for 15 min. These slides were aged overnight at room temperature in the dark. Ten microliters of mouse X (XMP D-1420-050; Metasystems) or Y (YMP D-1421-050, Metasystems) chromosome paints labeled with either Texas Red or FITC wavelength emitting fluorophores were applied to the aged slide, and the probe/slide were simultaneously denatured at 75°C for 8-10 minutes before incubating at 37°C in a humid chamber overnight. The slide was then washed in 0.4x SSC at 72°C for 2 min followed by a 30s rinse in room temperature 2x SSC. After drying DAPI was added in mounting media using a cover slip.

Microscopy

All pictures were taken with the 100x objective lens using an Olympus IX 71 (Delta Vision) microscope. This scope is equipped with cubes that excite Texas Red, FITC, and DAPI fluorophores. Pictures were taken at focal lengths in which the target signal was in focus.

RNA Isolation and qRT-PCR

Total RNA from testis was isolated using TRIzol reagent (Invitrogen) as described by the supplier's instructions. One microgram of total RNA was reverse transcribed using Oligo dT and random primers (qScript; Quanta) in a 20-uL reaction as described by the supplier's instructions. 0.5-uL of cDNA template was added to a 15-uL reaction with Sybr Green (Bio-Rad). Reactions were cycled using an iCycler (Bio-Rad) for 40 cycles. Resulting CT values were normalized to the control gene *b-actin* using ΔΔCt method.

Vector Transfection in Cell Culture

The vector used to generate the knockdown, eGFP-shXLR3 (5.3 Kb) and eGFP-empty (5.2Kb), were transfected into WT C57BL/6J neonatal fibroblasts via Invitrogen's Neon Transfection System. The system applies a predetermined electric pulse allowing ~90% transfection efficiency. eGFP was used as a selectable marker for sorting via Fluorescence-Activated Cell Sorting (FACs), followed by RNA extraction of the control and knockdown samples. This RNA was converted into cDNA and qRT-PCR was carried out as discussed previously using B-actin for control expression.

Chromatin Immunopreciptiation

ChIP was conducted on X monosomic (39, Xm/Xp) neonatal female whole brain and fibroblast using the MAGnify Chromatin Immunoprecipitation System from Invitrogen (49-2024). Briefly, samples were fixed in a suspension containing 1% formaldehyde for 10 minutes prior to cell lysis. Chromatin was then sheared down to 100-300bp size range using the Covaris S2 sonicating water bath. Next, protein A/G beads were coupled with 1-3ug of the antibody of choice for 12-16hr (i.e. anti-CTCF - ab70303), a control reaction containing 1ug of rabbit IgG was included. 25ug of the fixed and fragmented chromatin was incubated with this bead/antibody complex at 4°C for 12-16hr. After post-ChIP washes, crosslinks were reversed with Proteinase K at 55°C for 15 min. The remaining DNA was purified again and then eluted off the beads in 30uL of elution buffer and stored at -20°C. Primers were designed for CTCF control and target regions and analyzed through qPCR.

Bisulfite Sequencing

Genomic DNA was extracted from X monosomic mice (39, Xm/Xp) and converted using Zymo's EZ DNA Methylation Kit (D5001). ~500ng of DNA was denatured and incubated with a "CT conversion" reagent for 16 hours at 50°C. After a column clean-up the template can be amplified using primers that will bind CT converted sequence (Zymo - Bisulfite Primer Sequence tool). Primer sequences can be found in the Supplemental tables.

Hpa II Methylation Sensitive Assay

X_m and X_p neonatal whole brain genomic DNA was digested with the methyl sensitive restriction enzyme HpaII and separately EcoRI (control). Primers (F8a1F and F8a1R) flanking the region of interest were used to amplify the digested DNA revealing a band if HpaII didn't cut (methylation). No band results in sufficient cutting of HpaII at 1 or more sites within the amplicon indicating hypomethylation. In order to quantify the level of methylation, the same experiment was run via qRT-PCR and normalized to EcoRI.

CRISPR/Cas9

Target guide RNAs were designed to flank the CpG island in F8a ensuring that the required PAM sequence (NGG) exists downstream for proper Cas9 cutting. The 2 guide RNAs (see Appendix) were cloned into the same type of episomal vector **[Cas9(+) OFP(+)]** (Life Tech) allowing for fluorescent sorting of cells exclusively containing both guide RNAs. 1ug of these two plasmids was transfected into mouse neonatal fibroblasts via Invitrogen's Neon system per manufacturer's protocol. Both vectors carry the Cas9 protein CDS while allowing for high

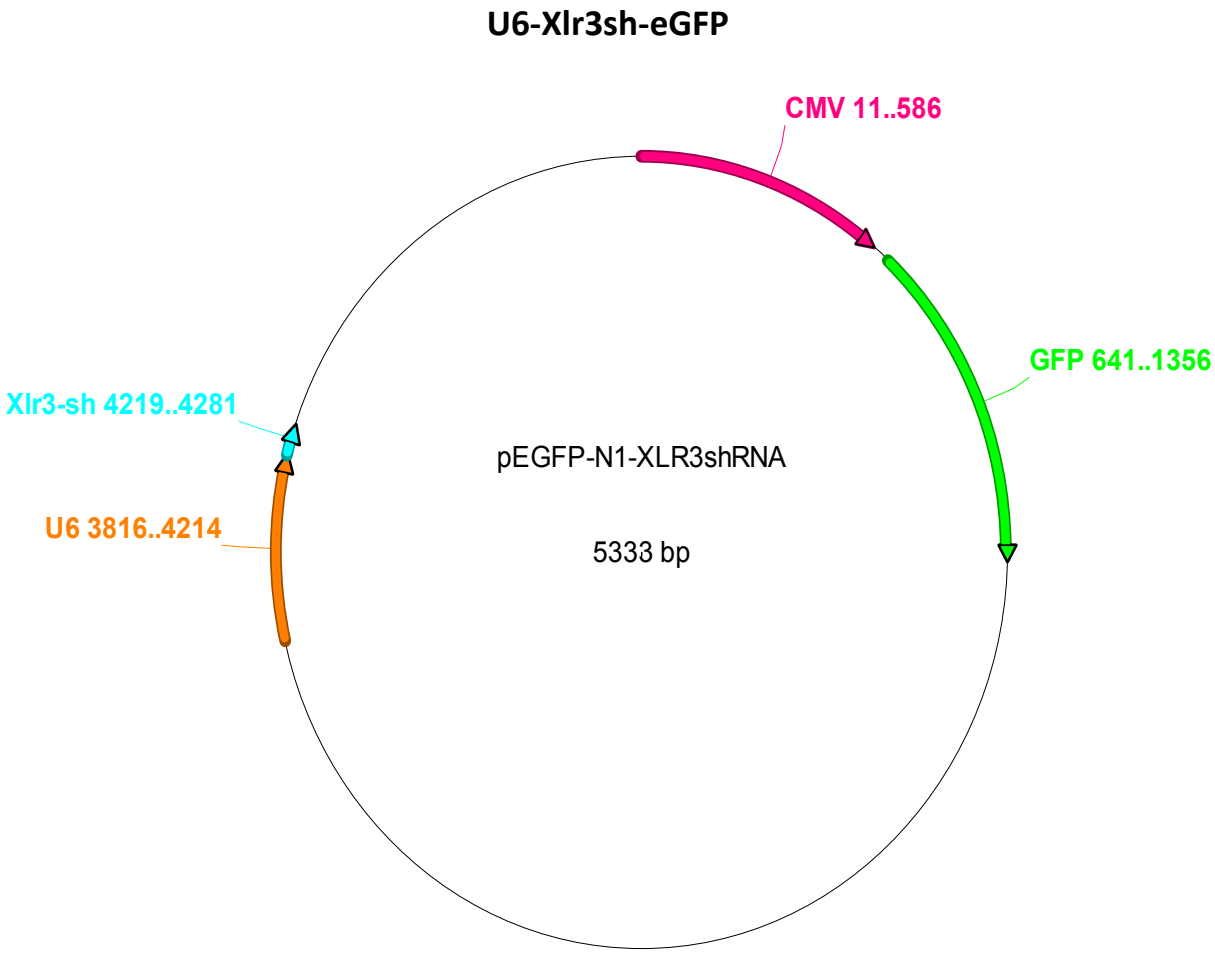
transcription of the respective guide RNAs. After transfection, the cells were fluorescently sorted too confirm a deletion.

Appendix

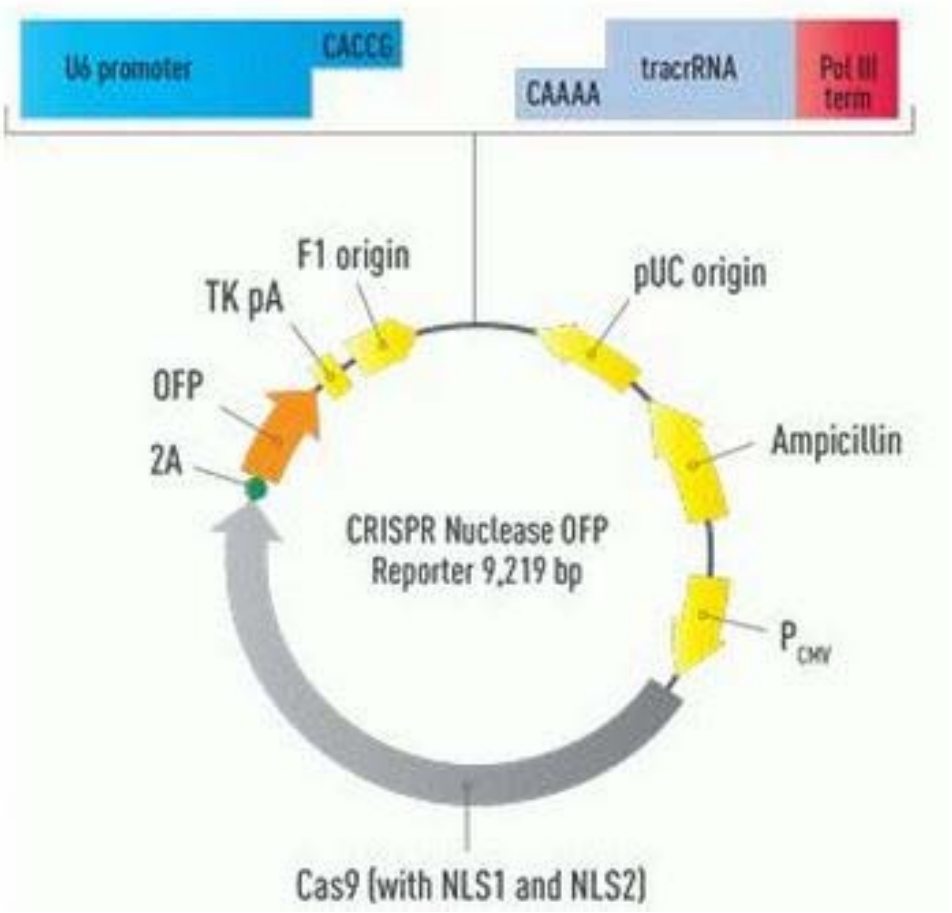
Primer	Sequence (5'-3')
Xlr3abcde F	AAAGGAAGGCCACTGACAC
Xlr3abcde R	TGTTGCCTCTCTGTTCCTGA
Actb F	ACACCCGCCACCAGTTCG
Actb R	CGATGGAGGGGAATACAGCC
MIT130 F	TTCATATCGCCCCAACCTAC
MIT130 R	TATTTTGAAACCTCTGCCATTT
SMCY F	TGAACTGCCTGCTATGCTAC
SMCY R	GCCTCAGATTCCAATGCTC
M13 F	CAGGAAACAGCTATGAC
M13 R	GTAAAACGACGGCCAG
Xlr3b Imp F	AGGCTGCCTTGTGGAGAG
Xlr3b Imp R	TGTCAGTGGCCTTCCTTTTT
F8a F1	CCAGGAGTGCCTGCCTTAT
F8a R1	CTGCAGAGGCTCCCCGTA
F8a F2	CTGCACTGTGCCTGGAACTA
F8a R2	GCGTTGCATGCGTGTAAATA
H19 F	AACCGCCAACAAGAAAGTCTGG
H19 R	GCTTCGGACATTGCTGTGGG
BS F8A F	GGGGAAGYGTTGGTTTTTATAGAGG
BS F8A R	AACAATACACRAAACACCTCACACCTAACTAAAAC
Xlr3sh F	GAGAGGCAACAGGCCAGGAAGGAAAGAAGCTTGTTTCCTTCCTGGCCTGTTGCCTCTCTTTTT
Xlr3sh R	AAAAAGAGAGGCAACAGGCCAGGAAGGAAACAAGCTTCTTTCCTTCCTGGCCTGTTGCCTCTC
F8a 5’ guide	CTTCTTGGCACGCTATCGGC
F8a 3’ guide	GGCTCTGTTGCGCCTTCACC
F8a Deletion F	GTCATAGCTGTTCCCCCTCA
F8a Deletion R	AGTCAGGTGTGAGGTGTCCC

Addgene Plasmids	Description
pEGFP-N1	Expression vector with GFP cassette
Cas9(+)OFP(+)	Contains Cas9, OFP selection and site to clone guide RNA in front of Tracer RNA
ROSA26-PAS	ROSA locus recombination vector with 5' and 3' arms

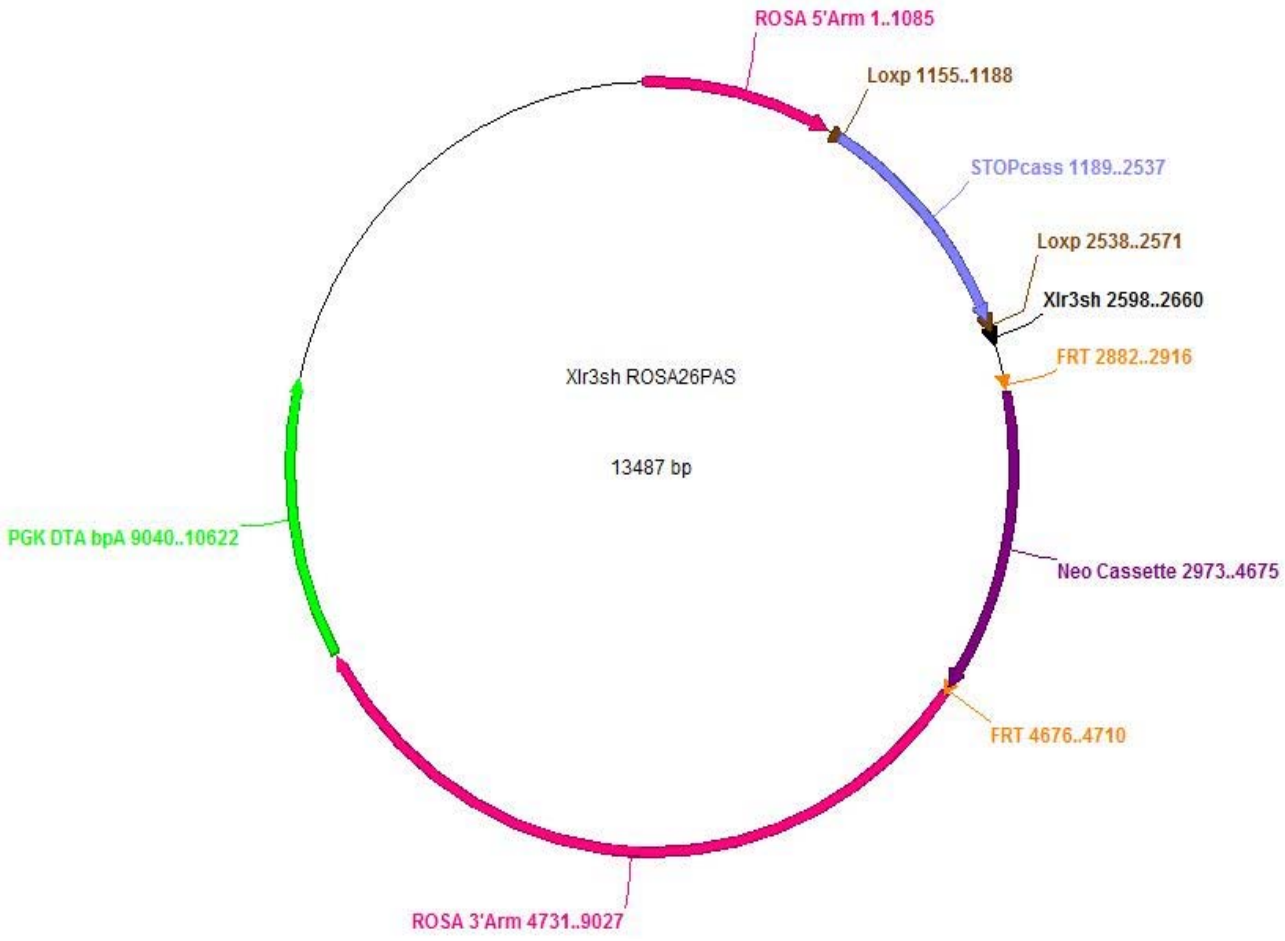
Antibody	Supplier	Cat #
Mouse anti-γH2AX	Abcam	ab18311
Rabbit anti-αTubulin	Abcam	ab15246
Rabbit anti-XLR3	Biosynthesis Inc.	*Peptide:* MSSRKRKATDTAGRHSRM
Rabbit anti-XLR4	GenScript	*Peptide:* MASKIKGRPPKQPKC
Mouse anti-SYCP3	Santa Cruz	sc74569
Rabbit (polyclonal) anti-CTCF	Abcam	ab70303
Rabbit anti-SYCP3	Abcam	ab15092
Rabbit anti-Fibrillarin	Abcam	ab5821
Rabbit anti-MeCP2	Abcam	ab2828



Cas9(+)OFP(+)


U6 promoter
CACCG
CAAAA
tracrRNA
Pol III term
F1 origin
TK pA
pUC origin
OFP
2A
Ampicillin
CRISPR Nuclease OFP
Reporter 9,219 bp
P CMV
Cas9 (with NLS1 and NLS2)

Xlr3sh-ROSA26PAS



Chapter 8: References

1. Bolcun-Filas, E. & Schimenti, J. C. *Genetics of Meiosis and Recombination in Mice*. **298,** (2012).

2. Fraune, J., Schramm, S., Alsheimer, M. & Benavente, R. The mammalian synaptonemal complex: Protein components, assembly and role in meiotic recombination. *Exp. Cell Res.* **318,** 1340–1346 (2012).

3. Escalier, D. Sex-Specific Gene Expression During Meiotic Prophase I: Xlr (X Linked, Lymphocyte Regulated), Not Its Male Homologue Xmr (Xlr Related, Meiosis Regulated), Is Expressed in Mouse Oocytes. *Biol. Reprod.* **67,** 1646–1652 (2002).

4. Cocquet, J. *et al.* The Multicopy Gene Sly Represses the Sex Chromosomes in the Male Mouse Germline after Meiosis. *PLoS Biol* **7,** e1000244 (2009).

5. Cocquet, J. *et al.* Deficiency in the multicopy Sycp3-like X-linked genes Slx and Slxl1 causes major defects in spermatid differentiation. *Mol. Biol. Cell* **21,** 3497–3505 (2010).

6. Tsutsumi, M. *et al.* Characterization of a Novel Mouse Gene Encoding an SYCP3-Like Protein That Relocalizes from the XY Body to the Nucleolus During Prophase of Male Meiosis I. *Biol. Reprod.* **85,** 165 –171 (2011).

7. Shi, Y.-Q. *et al.* SYCP3-like X-linked 2 is expressed in meiotic germ cells and interacts with synaptonemal complex central element protein 2 and histone acetyltransferase TIP60. *Gene* (2013). doi:10.1016/j.gene.2013.06.033

8. Raefski, A. S. & O'Neill, M. J. Identification of a cluster of X-linked imprinted genes in mice. *Nat Genet* **37,** 620–624 (2005).

9. Lesch, B. J. & Page, D. C. Genetics of germ cell development. *Nat. Rev. Genet.* **13,** 781–794 (2012).

10. Western, P. S. *et al.* Mitotic Arrest in Teratoma Susceptible Fetal Male Germ Cells. *PLoS ONE* **6,** e20736 (2011).

11. Bowles, J. *et al.* FGF9 suppresses meiosis and promotes male germ cell fate in mice. *Dev. Cell* **19,** 440–449 (2010).

12. Manku, G. & Culty, M. Mammalian gonocyte and spermatogonia differentiation: recent advances and remaining challenges. *Reproduction* **149,** R139–R157 (2015).

13. Alsheimer, M., Baier, A., Schramm, S., Schütz, W. & Benavente, R. Synaptonemal complex protein SYCP3 exists in two isoforms showing different conservation in mammalian evolution. *Cytogenet. Genome Res.* **128,** 162–168 (2010).

14. Baier, A., Alsheimer, M. & Benavente, R. Synaptonemal complex protein SYCP3: Conserved polymerization properties among vertebrates. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1774,** 595–602 (2007).

15. Botelho, R. J. *et al.* The genomic structure of SYCP3, a meiosis-specific gene encoding a protein of the chromosome core. *Biochim. Biophys. Acta* **1518,** 294–299 (2001).

16. Bellani, M. A. SPO11 is required for sex-body formation, and Spo11 heterozygosity rescues the prophase arrest of Atm-/- spermatocytes. *J. Cell Sci.* **118,** 3233–3245 (2005).

17. Rad51 protein involved in repair and recombination in S. cerevisiae is a RecA-like protein. *Cell* **69,** 457–470 (1992).

18. Kouznetsova, A. SYCP2 and SYCP3 are required for cohesin core integrity at diplotene but not for centromere cohesion at the first meiotic division. *J. Cell Sci.* **118,** 2271–2278 (2005).

19. Baier, A., Alsheimer, M., Volff, J. N. & Benavente, R. Synaptonemal complex protein SYCP3 of the rat: evolutionarily conserved domains and the assembly of higher order structures. *Sex. Dev.* **1,** 161–168 (2007).

20. Yuan, L. *et al.* The synaptonemal complex protein SCP3 can form multistranded, cross-striated fibers in vivo. *J. Cell Biol.* **142,** 331–339 (1998).

21. Yang, F. *et al.* Mouse SYCP2 is required for synaptonemal complex assembly and chromosomal synapsis during male meiosis. *J. Cell Biol.* **173,** 497–507 (2006).

22. Yuan, L. *et al.* The murine SCP3 gene is required for synaptonemal complex assembly, chromosome synapsis, and male fertility. *Mol. Cell* **5,** 73–83 (2000).

23. Structure and function of Synaptonemal complex ~ Biology Exams 4 U. at <http://www.biologyexams4u.com/2012/09/synaptonemal-complex-ultra-structure.html#.VEaHV_nF-T8>

24. Yang, F. & Wang, P. J. The Mammalian synaptonemal complex: a scaffold and beyond. *Genome Dyn.* **5,** 69–80 (2009).

25. Turner, J. M. A. Meiotic sex chromosome inactivation. *Dev. Camb. Engl.* **134,** 1823–1831 (2007).

26. Wang, P. J., McCarrey, J. R., Yang, F. & Page, D. C. An abundance of X-linked genes expressed in spermatogonia. *Nat Genet* **27,** 422–426 (2001).

27. Graves, J. A. M., Gécz, J. & Hameister, H. Evolution of the human X--a smart and sexy chromosome that controls speciation and development. *Cytogenet. Genome Res.* **99,** 141–145 (2002).

28. Cocquet, J. *et al.* A Genetic Basis for a Postmeiotic X Versus Y Chromosome Intragenomic Conflict in the Mouse. *PLoS Genet* **8,** e1002900 (2012).

29. Fernandez-Capetillo, O. *et al.* H2AX is required for chromatin remodeling and inactivation of sex chromosomes in male mouse meiosis. *Dev. Cell* **4,** 497–508 (2003).

30. Stucki, M. *et al.* MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell* **123,** 1213–1226 (2005).

31. Xu, X., Aprelikova, O., Moens, P., Deng, C.-X. & Furth, P. A. Impaired meiotic DNA-damage repair and lack of crossing-over during spermatogenesis in BRCA1 full-length isoform deficient mice. *Dev. Camb. Engl.* **130,** 2001–2012 (2003).

32. Turner, J. M. A. *et al.* BRCA1, histone H2AX phosphorylation, and male meiotic sex chromosome inactivation. *Curr. Biol. CB* **14,** 2135–2142 (2004).

33. Page, J. *et al.* Inactivation or non-reactivation: what accounts better for the silence of sex chromosomes during mammalian male meiosis? *Chromosoma* **121,** 307–326 (2012).

34. Herr'an, Y. *et al.* The cohesin subunit RAD21L functions in meiotic synapsis and exhibits sexual dimorphism in fertility. *EMBO J.* **30,** 3091–3105 (2011).

35. Reynard, L. N. *et al.* Expression analysis of the mouse multi-copy X-linked gene Xlr-related, meiosis-regulated (Xmr), reveals that Xmr encodes a spermatid-expressed cytoplasmic protein, SLX/XMR. *Biol. Reprod.* **77,** 329–335 (2007).

36. Peters, J.-M., Tedeschi, A. & Schmitz, J. The cohesin complex and its roles in chromosome biology. *Genes Dev.* **22,** 3089–3114 (2008).

37. Ishiguro, K., Kim, J., Fujiyama-Nakamura, S., Kato, S. & Watanabe, Y. A new meiosis-specific cohesin complex implicated in the cohesin code for homologous pairing. *EMBO Rep.* **12,** 267–275 (2011).

38. Fukuda, T., Daniel, K., Wojtasz, L., Toth, A. & Höög, C. A novel mammalian HORMA domain-containing protein, HORMAD1, preferentially associates with unsynapsed meiotic chromosomes. *Exp. Cell Res.* **316,** 158–171 (2010).

39. Shin, Y.-H. *et al.* Hormad1 Mutation Disrupts Synaptonemal Complex Formation, Recombination, and Chromosome Segregation in Mammalian Meiosis. *PLoS Genet.* **6,** e1001190 (2010).

40. Daniel, K. *et al.* Meiotic homologue alignment and its quality surveillance are controlled by mouse HORMAD1. *Nat. Cell Biol.* **13,** 599–610 (2011).

41. Wojtasz, L. *et al.* Meiotic DNA double-strand breaks and chromosome asynapsis in mice are monitored by distinct HORMAD2-independent and -dependent mechanisms. *Genes Dev.* **26,** 958–973 (2012).

42. Eichinger, C. S. & Jentsch, S. Synaptonemal complex formation and meiotic checkpoint signaling are linked to the lateral element protein Red1. *Proc. Natl. Acad. Sci.* **107,** 11370–11375 (2010).

43. Bergsagel, P. L., Timblin, C. R., Kozak, C. A. & Kuehl, W. M. Sequence and expression of murine cDNAs encoding Xlr3a and Xlr3b, defining a new X-linked lymphocyte-regulated Xlr gene subfamily. *Gene* **150,** 345–50 (1994).

44. Garchon, H.-J. *et al.* The XLR sequence family: dispersion on the X and Y chromosomes of a large set of closely related sequences, most of which are pseudogenes. *Nucleic Acids Res.* **17,** 9871 –9888 (1989).

45. Cohen, D. I. *et al.* Isolation of a cDNA clone corresponding to an X-linked gene family (XLR) closely linked to the murine immunodeficiency disorder xid. *Nature* **314,** 369–372 (1985).

46. Cohen, D. I., Steinberg, A. D., Paul, W. E. & Davis, M. M. Expression of an X-linked gene family (XLR) in late-stage B cells and its alteration by the xid mutation. *Nature* **314,** 372–374 (1985).

47. Siegel, J. N. *et al.* Sequence analysis and expression of an X-linked, lymphocyte-regulated gene family (XLR). *J. Exp. Med.* **166,** 1702 (1987).

48. Reynard, L. N., Cocquet, J. & Burgoyne, P. S. The Multi-Copy Mouse Gene Sycp3-Like Y-Linked (Sly) Encodes an Abundant Spermatid Protein That Interacts with a Histone Acetyltransferase and an Acrosomal Protein. *Biol. Reprod.* **81,** 250–257 (2009).

49. Werren, J. H. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc. Natl. Acad. Sci. U. S. A.* **108 Suppl 2,** 10863–10870 (2011).

50. Lyttle, T. W. Segregation Distorters. *Annu. Rev. Genet.* **25,** 511–581 (1991).

51. Kanizay, L. B. *et al.* Diversity and abundance of the abnormal chromosome 10 meiotic drive complex in Zea mays. *Heredity* **110,** 570–577 (2013).

52. Kanizay, L. B., Albert, P. S., Birchler, J. A. & Dawe, R. K. Intragenomic conflict between the two major knob repeats of maize. *Genetics* **194,** 81–89 (2013).

53. Buckler, E. S., 4th *et al.* Meiotic drive of chromosomal knobs reshaped the maize genome. *Genetics* **153,** 415–426 (1999).

54. Olds-Clarke, P. Models for male infertility: the t haplotypes. *Rev. Reprod.* **2,** 157–164 (1997).

55. Wallace, L. T. & Erhart, M. A. Recombination within mouse t haplotypes has replaced significant segments of t-specific DNA. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* **19,** 263–271 (2008).

56. Lyon, M. F. Transmission ratio distortion in mouse t-haplotypes is due to multiple distorter genes acting on a responder locus. *Cell* **37,** 621–628 (1984).

57. Schimenti, J. Segregation distortion of mouse t haplotypes: the molecular basis emerges. *Trends Genet.* **16,** 240–243 (2000).

58. Hurst, L. D. & Pomiankowski, A. Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. *Genetics* **128,** 841–858 (1991).

59. Sandler, L., Hiraizumi, Y. & Sandler, I. Meiotic Drive in Natural Populations of Drosophila Melanogaster. I. the Cytogenetic Basis of Segregation-Distortion. *Genetics* **44,** 233–250 (1959).

60. Larracuente, A. M. & Presgraves, D. C. The Selfish Segregation Distorter Gene Complex of Drosophila melanogaster. *Genetics* **192,** 33–53 (2012).

61. Hamilton, W. D. Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science* **156,** 477–488 (1967).

62. Hurst, L. D. Further evidence consistent with Stellate's involvement in meiotic drive. *Genetics* **142,** 641–643 (1996).

63. Belloni, M., Tritto, P., Bozzetti, M. P., Palumbo, G. & Robbins, L. G. Does Stellate cause meiotic drive in Drosophila melanogaster? *Genetics* **161,** 1551–1559 (2002).

64. Stapleton, W., Das, S. & McKee, B. D. A role of the Drosophila homeless gene in repression of Stellate in male meiosis. *Chromosoma* **110,** 228–240 (2001).

65. Ellis, P. J. I. *et al.* Deletions on mouse Yq lead to upregulation of multiple X- and Y-linked transcripts in spermatids. *Hum. Mol. Genet.* **14,** 2705–2715 (2005).

66. Harper, S. Q. *et al.* RNA interference improves motor and neuropathological abnormalities in a Huntington's disease mouse model. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 5820–5825 (2005).

67. Ellis, P. J., Ferguson, L., Clemente, E. J. & Affara, N. A. Bidirectional transcription of a novel chimeric gene mapping to mouse chromosome Yq. *BMC Evol. Biol.* **7,** 171 (2007).

68. Good, J. M., Dean, M. D. & Nachman, M. W. A Complex Genetic Basis to X-Linked Hybrid Male Sterility Between Two Species of House Mice. *Genetics* **179,** 2213–2228 (2008).

69. Davies, W. *et al.* Xlr3b is a new imprinted candidate for X-linked parent-of-origin effects on cognitive function in mice. *Nat. Genet.* **37,** 625–629 (2005).

70. Hippenmeyer, S. *et al.* Genetic mosaic dissection of Lis1 and Ndel1 in neuronal migration. *Neuron* **68,** 695–709 (2010).

71. Premsrirut, P. K. *et al.* A rapid and scalable system for studying gene function in mice using conditional RNA interference. *Cell* **145,** 145–158 (2011).

72. Tasic, B. *et al.* Site-specific integrase-mediated transgenesis in mice via pronuclear injection. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 7902–7907 (2011).

73. Waddington, C. H. The Epigenotype. *Int. J. Epidemiol.* (2011). doi:10.1093/ije/dyr184

74. Nanney, D. L. EPIGENETIC CONTROL SYSTEMS. *Proc. Natl. Acad. Sci. U. S. A.* **44,** 712–717 (1958).

75. TRIVERS, R. L. Parent-Offspring Conflict. *Am. Zool.* **14,** 249 –264 (1974).

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

77. Cooper, D. W., VandeBerg, J. L., Sharman, G. B. & Poole, W. E. Phosphoglycerate kinase polymorphism in kangaroos provides further evidence for paternal X inactivation. *Nature. New Biol.* **230,** 155–157 (1971).

78. VandeBerg, J. L., Cooper, D. W., Sharman, G. B. & Poole, W. E. Studies on metatherian sex chromosomes. IV. X linkage of PGK-A with paternal X inactivation confirmed in erythrocytes of grey kangaroos by pedigree analysis. *Aust. J. Biol. Sci.* **30,** 115–125 (1977).

79. Haig, D. & Westoby, M. Parent-Specific Gene Expression and the Triploid Endosperm. *Am. Nat.* **134,** 147–155 (1989).

80. DeChiara, T. M., Robertson, E. J. & Efstratiadis, A. Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* **64,** 849–859 (1991).

81. Giannoukakis, N., Deal, C., Paquette, J., Goodyer, C. G. & Polychronakos, C. Parental genomic imprinting of the human IGF2 gene. *Nat. Genet.* **4,** 98–101 (1993).

82. Drewell, R. A. *et al.* Deletion of a silencer element disrupts H19 imprinting independently of a DNA methylation epigenetic switch. *Dev. Camb. Engl.* **127,** 3419–3428 (2000).

83. Bartolomei, M. S., Zemel, S. & Tilghman, S. M. Parental imprinting of the mouse H19 gene. *Nature* **351,** 153–155 (1991).

84. Tremblay, K. D., Duran, K. L. & Bartolomei, M. S. A 5’ 2-kilobase-pair region of the imprinted mouse H19 gene exhibits exclusive paternal methylation throughout development. *Mol. Cell. Biol.* **17,** 4322–4329 (1997).

85. Leighton, P. A., Saam, J. R., Ingram, R. S., Stewart, C. L. & Tilghman, S. M. An enhancer deletion affects both H19 and Igf2 expression. *Genes Dev.* **9,** 2079 –2089 (1995).

86. Thorvaldsen, J. L., Duran, K. L. & Bartolomei, M. S. Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. *Genes Dev.* **12,** 3693–3702 (1998).

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

88. Bell, A. C., West, A. G. & Felsenfeld, G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98,** 387–396 (1999).

89. Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405,** 482–485 (2000).

90. Murphy, S. K. *et al.* Frequent IGF2/H19 Domain Epigenetic Alterations and Elevated IGF2 Expression in Epithelial Ovarian Cancer. *Mol. Cancer Res.* **4,** 283–292 (2006).

91. Reik, W. & Murrell, A. Genomic imprinting. Silence across the border. *Nature* **405,** 408–409 (2000).

92. Wutz, A. *et al.* Non-imprinted Igf2r expression decreases growth and rescues the Tme mutation in mice. *Development* **128,** 1881–1887 (2001).

93. Lau, M. M. *et al.* Loss of the imprinted IGF2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality. *Genes Dev.* **8,** 2953–2963 (1994).

94. Wang, Z. Q., Fung, M. R., Barlow, D. P. & Wagner, E. F. Regulation of embryonic growth and lysosomal targeting by the imprinted Igf2/Mpr gene. *Nature* **372,** 464–467 (1994).

95. Ludwig, T. *et al.* Mouse mutants lacking the type 2 IGF receptor (IGF2R) are rescued from perinatal lethality in Igf2 and Igf1r null backgrounds. *Dev. Biol.* **177,** 517–535 (1996).

96. Barton, S. C., Surani, M. A. & Norris, M. L. Role of paternal and maternal genomes in mouse development. *Nature* **311,** 374–376 (1984).

97. Surani, M. A., Barton, S. C. & Norris, M. L. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308,** 548–550 (1984).

98. Surani, M. A. & Barton, S. C. Development of gynogenetic eggs in the mouse: implications for parthenogenetic embryos. *Science* **222,** 1034–1036 (1983).

99. O'Neill, M. J. The influence of non-coding RNAs on allele-specific gene expression in mammals. *Hum. Mol. Genet.* **14 Spec No 1,** R113–120 (2005).

100. Kanduri, C. *et al.* A Differentially Methylated Imprinting Control Region within the Kcnq1 Locus Harbors a Methylation-sensitive Chromatin Insulator. *J. Biol. Chem.* **277,** 18106–18110 (2002).

101. Carr, M. S., Yevtodiyenko, A., Schmidt, C. L. & Schmidt, J. V. Allele-specific histone modifications regulate expression of the Dlk1-Gtl2 imprinted domain. *Genomics* **89,** 280–290 (2007).

102. Reik, W. & Lewis, A. Co-evolution of X-chromosome inactivation and imprinting in mammals. *Nat. Rev. Genet.* **6,** 403–410 (2005).

103. Muller, H. J. in *Harvey Lectures* 165–229 (1947).

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

105. Ohno, S. *Sex Chromosomes and Sex Linked Genes*. (Springer-Verlag, 1967).

106. Gupta, V. *et al.* Global analysis of X-chromosome dosage compensation. *J. Biol.* **5,** 3 (2006).

107. Lin, H. *et al.* Dosage Compensation in the Mouse Balances Up-Regulation and Silencing of X-Linked Genes. *PLoS Biol.* **5,** e326 EP – (2007).

108. Xiong, Y. *et al.* RNA sequencing shows no dosage compensation of the active X-chromosome. *Nat Genet* **advance online publication,** (2010).

109. Jue, N. K. *et al.* Determination of dosage compensation of the mammalian X chromosome by RNA-seq is dependent on analytical approach. *BMC Genomics* **14,** 150 (2013).

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

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

112. Panning, B., Dausman, J. & Jaenisch, R. X Chromosome Inactivation Is Mediated by Xist RNA Stabilization. *Cell* **90,** 907–916 (1997).

113. Sheardown, S. A. *et al.* Stabilization of Xist RNA Mediates Initiation of X Chromosome Inactivation. *Cell* **91,** 99–107 (1997).

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

115. Costanzi, C. & Pehrson, J. R. Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* **393,** 599–601 (1998).

116. Heard, E. *et al.* Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell* **107,** 727–738 (2001).

117. Kohlmaier, A. *et al.* A Chromosomal Memory Triggered by Xist Regulates Histone Methylation in X Inactivation. *PLoS Biol* **2,** e171 (2004).

118. Leeb, M. & Wutz, A. Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. *J. Cell Biol.* **178,** 219–229 (2007).

119. Cohen, D. E. *et al.* The DXPas34 repeat regulates random and imprinted X inactivation. *Dev. Cell* **12,** 57–71 (2007).

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

121. Ropers, H. H., Wolff, G. & Hitzeroth, H. W. Preferential X inactivation in human placenta membranes: is the paternal X inactive in early embryonic development of female mammals? *Hum. Genet.* **43,** 265–273 (1978).

122. Grant, J. *et al.* Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. *Nature* **487,** 254–258 (2012).

123. Filippova, G. N. *et al.* An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol. Cell. Biol.* **16,** 2802–2813 (1996).

124. Merkenschlager, M. & Odom, D. T. CTCF and Cohesin: Linking Gene Regulatory Elements with Their Targets. *Cell* **152,** 1285–1297 (2013).

125. MacPherson, M. J. & Sadowski, P. D. The CTCF insulator protein forms an unusual DNA structure. *BMC Mol. Biol.* **11,** 101 (2010).

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

127. Chen, H., Tian, Y., Shu, W., Bo, X. & Wang, S. Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PloS One* **7,** e41374 (2012).

128. Cuddapah, S. *et al.* Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* **19,** 24–32 (2009).

129. Kim, T. H. *et al.* Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128,** 1231–1245 (2007).

130. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133,** 1106–1117 (2008).

131. Ong, C.-T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* **15,** 234–246 (2014).

132. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* **38,** 1341–1347 (2006).

133. Hark, A. T. & Tilghman, S. M. Chromatin conformation of the H19 epigenetic mark. *Hum. Mol. Genet.* **7,** 1979–1985 (1998).

134. Hark, A. T. *et al.* CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405,** 486–489 (2000).

135. Nativio, R. *et al.* Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. *PLoS Genet.* **5,** e1000739 (2009).

136. Kim, Y. J., Cecchini, K. R. & Kim, T. H. Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 7391–7396 (2011).

137. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489,** 109–113 (2012).

138. Golan-Mashiach, M. *et al.* Identification of CTCF as a master regulator of the clustered protocadherin genes. *Nucleic Acids Res.* **40,** 3378–3391 (2012).

139. Monahan, K. *et al.* Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of protocadherin-α gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 9125–9130 (2012).

140. Guo, Y. *et al.* CTCF/cohesin-mediated DNA looping is required for protocadherin α promoter choice. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 21081–21086 (2012).

141. Zlatanova, J. & Caiafa, P. CTCF and its protein partners: divide and rule? *J. Cell Sci.* **122,** 1275–1284 (2009).

142. Hadjur, S. *et al.* Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* **460,** 410–413 (2009).

143. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 996–1001 (2014).

144. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).

145. Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* **32,** 3119–3129 (2013).

146. Seitan, V. C. *et al.* Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Res.* **23,** 2066–2077 (2013).

147. Brown, J. L., Mucci, D., Whiteley, M., Dirksen, M. L. & Kassis, J. A. The Drosophila Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. *Mol. Cell* **1,** 1057–1064 (1998).

148. Atchison, L., Ghias, A., Wilkinson, F., Bonini, N. & Atchison, M. L. Transcription factor YY1 functions as a PcG protein in vivo. *EMBO J.* **22,** 1347–1358 (2003).

149. Bushmeyer, S., Park, K. & Atchison, M. L. Characterization of functional domains within the multifunctional transcription factor, YY1. *J. Biol. Chem.* **270,** 30213–30220 (1995).

150. Atchison, M. L. Function of YY1 in Long-Distance DNA Interactions. *Front. Immunol.* **5,** 45 (2014).

151. Bossen, C., Mansson, R. & Murre, C. Chromatin Topology and the Regulation of Antigen Receptor Assembly. *Annu. Rev. Immunol.* **30,** 337–356 (2012).

152. Cobb, R. M., Oestreich, K. J., Osipovich, O. A. & Oltz, E. M. in *Advances in Immunology* (ed. Frederick W. Alt) **Volume 91,** 45–109 (Academic Press, 2006).

153. Bassing, C. H., Swat, W. & Alt, F. W. The mechanism and regulation of chromosomal V(D)J recombination. *Cell* **109 Suppl,** S45–55 (2002).

154. Kim, J. K., Samaranayake, M. & Pradhan, S. Epigenetic mechanisms in mammals. *Cell. Mol. Life Sci. CMLS* **66,** 596–612 (2009).

155. Jabbari, K. & Bernardi, G. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene* **333,** 143–149 (2004).

156. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 1412–1417 (2006).

157. Liao, J. *et al.* Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nat. Genet.* **47,** 469–478 (2015).

158. Weaver, J. R., Susiarjo, M. & Bartolomei, M. S. Imprinting and epigenetic changes in the early embryo. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* **20,** 532–543 (2009).

159. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69,** 915–926 (1992).

160. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99,** 247–257 (1999).

161. Bartolomei, M. S. & Tilghman, S. M. Genomic imprinting in mammals. *Annu. Rev. Genet.* **31,** 493–525 (1997).

162. Zimmermann, C. A., Hoffmann, A., Raabe, F. & Spengler, D. Role of Mecp2 in Experience-Dependent Epigenetic Programming. *Genes* **6,** 60–86 (2015).

163. Chahrour, M. & Zoghbi, H. Y. The story of Rett syndrome: from clinic to neurobiology. *Neuron* **56,** 422–437 (2007).

164. Kernohan, K. D., Vernimmen, D., Gloor, G. B. & Bérubé, N. G. Analysis of neonatal brain lacking ATRX or MeCP2 reveals changes in nucleosome density, CTCF binding and chromatin looping. *Nucleic Acids Res.* **42,** 8356–8368 (2014).

165. Simonsson, T. G-quadruplex DNA structures--variations on a theme. *Biol. Chem.* **382,** 621–628 (2001).

166. Bartholdy, B., Mukhopadhyay, R., Lajugie, J., Aladjem, M. I. & Bouhassira, E. E. Allele-specific analysis of DNA replication origins in mammalian cells. *Nat. Commun.* **6,** 7051 (2015).

167. Watson, L. A. *et al.* Atrx deficiency induces telomere dysfunction, endocrine defects, and reduced life span. *J. Clin. Invest.* **123,** 2049–2063 (2013).

168. Law, M. J. *et al.* ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell* **143,** 367–378 (2010).

169. Wong, L. H. *et al.* ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome Res.* **20,** 351–360 (2010).

170. Bagheri-Fam, S. *et al.* Defective survival of proliferating Sertoli cells and androgen receptor function in a mouse model of the ATR-X syndrome. *Hum. Mol. Genet.* **20,** 2213–2224 (2011).

171. Kernohan, K. D. *et al.* ATRX partners with cohesin and MeCP2 and contributes to developmental silencing of imprinted genes in the brain. *Dev. Cell* **18,** 191–202 (2010).

172. Quenneville, S. *et al.* In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol. Cell* **44,** 361–372 (2011).

173. Strogantsev, R. *et al.* Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. *Genome Biol.* **16,** 112 (2015).

174. Feinberg, A. P. & Vogelstein, B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301,** 89–92 (1983).

175. Rainier, S. *et al.* Relaxation of imprinted genes in human cancer. *Nature* **362,** 747–749 (1993).

176. Ogawa, O. *et al.* Relaxation of insulin-like growth factor II gene imprinting implicated in Wilms' tumour. *Nature* **362,** 749–751 (1993).

177. Ehrlich, M. The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease. *Clin. Immunol. Orlando Fla* **109,** 17–28 (2003).

178. Sharma, S., Kelly, T. K. & Jones, P. A. Epigenetics in cancer. *Carcinogenesis* **31,** 27–36 (2010).

179. Baylin, S. B. DNA methylation and gene silencing in cancer. *Nat. Clin. Pract. Oncol.* **2 Suppl 1,** S4–11 (2005).

180. Song, J. *et al.* Increased expression of histone deacetylase 2 is found in human gastric cancer. *APMIS Acta Pathol. Microbiol. Immunol. Scand.* **113,** 264–268 (2005).

181. Halkidou, K. *et al.* Upregulation and nuclear recruitment of HDAC1 in hormone refractory prostate cancer. *The Prostate* **59,** 177–189 (2004).

182. Valk-Lingbeek, M. E., Bruggeman, S. W. M. & van Lohuizen, M. Stem cells and cancer; the polycomb connection. *Cell* **118,** 409–418 (2004).

183. Viré, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439,** 871–874 (2006).

184. Knoll, J. H. M. *et al.* Angelman and Prader-Willi syndromes share a common chromosome 15 deletion but differ in parental origin of the deletion. *Am. J. Med. Genet.* **32,** 285–290 (1989).

185. Nicholls, R. D., Knoll, J. H., Butler, M. G., Karam, S. & Lalande, M. Genetic imprinting suggested by maternal heterodisomy in nondeletion Prader-Willi syndrome. *Nature* **342,** 281–285 (1989).

186. Cattanach, B. M. *et al.* A candidate mouse model for Prader-Willi syndrome which shows an absence of Snrpn expression. *Nat. Genet.* **2,** 270–274 (1992).

187. Leff, S. E. *et al.* Maternal imprinting of the mouse Snrpn gene and conserved linkage homology with the human Prader-Willi syndrome region. *Nat. Genet.* **2,** 259–264 (1992).

188. Newschaffer, C. J. *et al.* The epidemiology of autism spectrum disorders. *Annu. Rev. Public Health* **28,** 235–258 (2007).

189. Werling, D. M. & Geschwind, D. H. Understanding sex bias in autism spectrum disorder. *Proc. Natl. Acad. Sci.* **110,** 4868–4869 (2013).

190. Baron-Cohen, S. The extreme male brain theory of autism. *Trends Cogn. Sci.* **6,** 248–254 (2002).

191. Skuse, D. H. *et al.* Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function. *Nature* **387,** 705–708 (1997).

192. Skuse, D. H. Imprinting, the X-chromosome, and the male brain: explaining sex differences in the liability to autism. *Pediatr. Res.* **47,** 9–16 (2000).

193. Kobayashi, S. *et al.* Comparison of gene expression in male and female mouse blastocysts revealed imprinting of the X-linked gene, Rhox5/Pem, at preimplantation stages. *Curr. Biol. CB* **16,** 166–172 (2006).

194. Kobayashi, S. *et al.* The X-linked imprinted gene family Fthl17 shows predominantly female expression following the two-cell stage in mouse embryos. *Nucleic Acids Res.* **38,** 3672–3681 (2010).

195. Shanker, S., Hu, Z. & Wilkinson, M. F. Epigenetic regulation and downstream targets of the Rhox5 homeobox gene. *Int. J. Androl.* **31,** 462–470 (2008).

196. Peters, M. F. & Ross, C. A. Isolation of a 40-kDa Huntingtin-associated protein. *J. Biol. Chem.* **276,** 3188–3194 (2001).

197. Isles, A. R., Davies, W., Burrmann, D., Burgoyne, P. S. & Wilkinson, L. S. Effects on fear reactivity in XO mice are due to haploinsufficiency of a non-PAR X gene: implications for emotional function in Turner's syndrome. *Hum. Mol. Genet.* **13,** 1849–1855 (2004).

198. De La Rosa-Velázquez, I. A., Rincón-Arano, H., Benítez-Bribiesca, L. & Recillas-Targa, F. Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. *Cancer Res.* **67,** 2577–2585 (2007).

199. Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480,** 490–495 (2011).

200. Feldmann, A. *et al.* Transcription Factor Occupancy Can Mediate Active Turnover of DNA Methylation at Regulatory Regions. *PLoS Genet* **9,** e1003994 (2013).

201. Martinez-Garay, I. *et al.* A new gene family (FAM9) of low-copy repeats in Xp22.3 expressed exclusively in testis: implications for recombinations in this region. *Genomics* **80,** 259–267 (2002).

202. Zhuang, X.-J. *et al.* Identification and Characterization of Xlr5c as a Novel Nuclear Localization Protein in Mouse Germ Cells. *PloS One* **10,** e0130087 (2015).

203. Hanna, C. W. & Kelsey, G. The specification of imprints in mammals. *Heredity* **113,** 176–183 (2014).

204. Zheng, K. *et al.* Co-transcriptional formation of DNA:RNA hybrid G-quadruplex and potential function as constitutional cis element for transcription control. *Nucleic Acids Res.* **41,** 5533–5541 (2013).

205. Harris, R. A., Nagy-Szakal, D. & Kellermayer, R. Human metastable epiallele candidates link to common disorders. *Epigenetics Off. J. DNA Methylation Soc.* **8,** 157–163 (2013).

206. Kobayashi, H. *et al.* Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet.* **8,** e1002440 (2012).

207. Nava, C. *et al.* Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including TMLHE. *Transl. Psychiatry* **2,** e179 (2012).

208. Celestino-Soper, P. B. S. *et al.* Use of array CGH to detect exonic copy number variants throughout the genome in autism families detects a novel deletion in TMLHE. *Hum. Mol. Genet.* **20,** 4360–4370 (2011).

209. Celestino-Soper, P. B. S. *et al.* A common X-linked inborn error of carnitine biosynthesis may be a risk factor for nondysmorphic autism. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 7974–7981 (2012).