



Gadd45g and Mammalian Testis Determination

A thesis presented for the degree of Doctor of Philosophy (D. Phil)

Wolfson College, University of Oxford

Joel Benjamin May

Mammalian Genetics Unit, MRC Harwell Institute

Department of Physiology, Anatomy and Genetics, University of Oxford

Supervisors: Dr Andy Greenfield (MGU, MRC Harwell)

Dr Helen Christian (DPAG, Oxford)

Trinity term 2016

Word count: 41,031

Gadd45g and mammalian testis determination

Joel May

Wolfson College, Oxford
MRC Mammalian Genetics Unit, MRC Harwell
Department of Physiology, Anatomy and Genetic, University of Oxford

A thesis presented for D. Phil Trinity term, 2016

Abstract

Mammalian sex is determined by the presence or absence of *Sry*. This gene, located on the Y chromosome, promotes the development of a testis from the biopotential gonad by initiating a transcriptional programme that results in Sertoli cell formation. SRY is required to reach a specific threshold level during a critical time-period in order to induce the testicular differentiation pathway. We have previously shown that XY mouse embryos lacking Growth arrest and DNA damage inducible 45 gamma (*Gadd45g*) exhibit gonadal sex reversal due to a delay in *Sry* expression.

Efforts to study the role of GADD45γ in sex determination have been stymied by the lack of a working antibody. As an alternative, I have produced a BAC transgene produces a similar expression profile to endogenous *Gadd45g*. This BAC was also shown to rescue the sex reversal found in B6-Y^{POS} gonads, though further research is required to understand the molecular mechanisms through which this occurs. As a result, this BAC was chosen for modification using BAC recombineering to contain the sequence for a poly-His/FLAG epitope tag. Though this edited BAC was generated successfully, no transgenic lines have yet been produced.

Members of the *Gadd45* family have previously been associated with active DNA demethylation. As such, I have conducted whole genome bisulphite sequencing on DNA from XY wild-type, XX wild-type and XY *Gadd45g*$^{-/-}$ embryos at the 18 tail somite stage, where *Sry* expression is at its peak. *Gadd45g* and *Sry* are only expressed in the somatic cells of the gonads, which have been purified through fluorescence-activated flow cytometry using the transgenic line *Sf1-EGFP*. Due to the low number of cells that are purified from the embryonic gonad at this stage, sequencing libraries were generated using post-bisulphite adaptor tagging (PBAT). Analysis of the acquired sequences revealed that there are no significant differences between the methylomes of XY wild-type, XX wild-type and XY *Gadd45g*$^{-/-}$ mutant gonadal somatic cells. It therefore appears that GADD45γ does not have a role in establishing the methylome at this particular stage of sex determination.

Acknowledgments

I'd first and foremost like to thank my supervisor Dr Andy Greenfield for all of the support and guidance throughout my D. Phil. Thank you for pushing me and encouraging me when I needed it. I am also grateful to you for proofreading this thesis and advising me through the writing process. I would also like to thank the members of the Disorder of Sex Development lab for all of their help over the years. To Nick Warr for laying the groundwork for the BAC transgenesis project and developing the *Gadd45gΔ* WMISH probe. Data courtesy of him is found in figure 1.4, figure 3.4 and figure 3.6. To Gwenn Carre for having the patience of a saint, with both me and the various technical issues we ran into with cell sorting and bisulphite conversion. I thank Pam Siggers too, for providing guidance and a great sense of humour when it was needed. Huge thanks go to Abi Harris, Catherine Livermore and Madeline Pope for helping with dissection, providing cake and being there for tea breaks. I also thank the members of the Norris lab for their help and advice on developmental biology matter. I would also like to thank the rest of the students at MRC Harwell for providing a great atmosphere to work in over the years.

I am extremely thankful to the MRC Harwell Core Services for their assistance provided throughout my D. Phil. In particular, my sincerest thanks go to Michelle Simon and the bioinformatics team for helping me process my WGBS data. I thank the team for providing the tools to aid in aligning the obtained sequences to the reference genome and conducting the genome wide search for differentially methylated regions. A huge thanks goes to the staff of the Mary Lyon Centre, in particular Sara Wells, Jackie Harrison, Lee Kent, Alison Gallop and Anju Paudyal, for providing extensive knowledge and taking care of my mice. I also thank Jeremy Sanderson for his help with imaging, and thanks to the Transgenics, Histology and GEMS teams. A huge thank you goes to Gavin Kelsey and Jiahoa Huang for inviting me to their lab to aid in the PBAT protocol.

Thank you to my family, my parents for supporting me and inspiring me to get where I am today. Thank you to my sister and all of my friends. I finally thank my amazing wife Sarah, who has been my rock through my time in Oxford. All of your support has been invaluable and I hope to repay it in kind as you finish your thesis.

I would also like to thank Wolfson College for providing me with a home away from home. Finally I thank the Medical Research Council for funding this research for my D. Phil.

Table of Contents

List of Figures x

List of Tables xii

Abbreviations xiii

Chapter 1. Introduction 1

1.1 Overview 2

1.2 Mammalian sex determination 4

1.2.1 The formation of the gonadal ridge 4

1.2.2 The gonad as a bipotential organ 11

1.2.3 *Sry* and the establishment of the testis 15

1.2.4 The regulation of *Sry* expression 22

1.2.5 The requirement of mitogen-activated protein kinase signalling and *Gadd45g* in mammalian sex determination 28

1.2.6 Epigenetic regulation in sex determination 30

1.3 Clinical relevance of Disorders of sex development (DSD) 33

1.4 The mouse as a model of human DSD 35

1.5 *Gadd45* members and roles 37

1.5.1 The anti-tumour role of Gadd45 members 38

1.5.2 *Gadd45g* expression profile 39

1.5.3 GADD45 proteins in development 40

1.5.4 GADD45 proteins as mediators of DNA repair 41

1.6 DNA methylation and epigenetic regulation 43

1.6.1 Regulation of gene expression by epigenetic factors 43

1.6.2 Active DNA methylation 44

1.6.3 Active DNA demethylation 46

1.6.4 The role of GADD45 proteins in active DNA demethylation 49

1.7 Main aims: summary 52

Chapter 2. Materials and Methods 55

2.1 General methods and solutions 56

2.2 Mouse lines and genotyping 56

2.2.1 Chromosomal sexing 57

2.2.2 *Sf1-EGFP* 57

2.2.3 *Gadd45g* 58

2.2.4 *Gadd45g* BAC Transgenic line 58

2.3 Tissue Collection 60

2.3.1 Embryo Collection 60

2.4 Tissue Processing 61

2.4.1 Fluorescence-activated cells sorting (FACS) 61

2.4.2 Wholemount in situ hybridisation 61

2.4.3 RNA extraction 62

2.4.4 Histology 62

2.5 BAC recombineering 62
2.5.1 Generation of rpsL/neo flanked with homology arms 62
2.5.2 Transformation of the pRed/ET plasmid 63
2.5.3 RedET mediated recombineering 64
2.6 Fluorescence Activated Cell Sorting (FACS) 65
2.7 CpG Methylation 66
2.7.1 Bisulphite conversion 66
2.7.2 Bisulphite sequencing 67
2.7.3 Multiple Displacement Amplification (MDA) 68
2.7.4 Post Bisulphite Adaptor Tagging (PBAT) 68
2.8 Cloning 70
2.8.1 TA cloning 70
2.8.2 Transformation 71
2.8.3 DNA preparation from plasmid DNA 72
2.9 Whole Genome Bisulphite Sequencing 72
2.10 Quantitive real-time polymerase chain reaction 72
2.10.1 RNA extraction and cDNA synthesis 73
2.10.2 Plate setup and qRT-PCR 73
2.10.3 Analysis of qRT-PCR results 74
Equation B: Relative expression (R) 74
Equation C: Standard error of the mean 74
2.11 Droplet Digital PCR 75

2.12 Wholemount in situ hybridisation 75

2.12.1 Riboprobe synthesis 75

2.12.2 Wholemount in situ hybridisation (WMISH) 76

2.13 Statistical power calculation 78

Chapter 3: Characterisation of a transgenic mouse line overexpressing *Gadd45g* 79

3.1 Introduction 80

3.1.1 The requirement for *Gadd45g* during mammalian sex determination 80

3.1.2 B6.Y^{DOM} models of *Sry*-dependent gonadal sex reversal 81

3.1.3 Expression studies using transgenic mice 82

3.1.4 Generation of a *Gadd45g* BAC transgene 84

3.1.5 *E15-Gadd45g* rescues the *Gadd45g*$^{+/-}$ Y^{AKR} sex reversal phenotype 85

3.2 Results 90

3.2.1 *E15-Gadd45g* induces overexpression of *Gadd45g* but does not increase *Sry* beyond normal C57Bl/6J levels 90

3.2.2 E15-Gadd45g specific expression shares a spatial expression profile with endogenous Gadd45g, but not similar temporal profile 92

3.2.3 E15-Gadd45g partially rescues the sex reversal of B6.Y^{POS} 93

3.3 Conclusions and Discussion 99

3.3.1 E15-Gadd45g induces overexpression of *Gadd45g* in the gonads 100

3.3.2 *E15-Gadd45g* shares a similar spatial expression pattern to endogenous *Gadd45g* but not a similar temporal profile 102

3.3.3 E15-Gadd45g rescues testis formation in B6.Y^{POS} mice 103

Chapter 4: Introduction of an epitope tag into the GADD45γ coding sequence using BAC recombineering 105

4.1 Introduction 106

4.1.1 Background information 106

4.1.2 Aims and hypotheses 109

4.2 Results 110

4.2.1 Generation of a BAC containing a modified *Gadd45g* gene with a C-terminal His/FLAG epitope tag sequence 110

4.2.2 Generation of transgenic mice using modified BAC lines 119

4.3 Conclusions and Discussion 121

4.3.1 Aims revisited 121

4.3.2 Discussion 121

Chapter 5: Investigating methods of whole genome bisulphite sequencing using gonadal somatic cells 128

5.1 Introduction 129

5.1.1 The major cell types in the developing gonad 129

5.1.2 Fluorescence-activated cell sorting using *Sf1-EGFP* 130

5.1.3 Epigenetic studies of gonadal cell lineages 132

5.1.4 Genome-wide DNA methylation analysis methods 133

5.1.5 Aims and Hypotheses 134

5.2 Results 135

5.2.1 Enrichment of gonadal somatic cells using transgenic *Sf1-EGFP* mice 135

5.2.2 Improving the DNA yield from sorted gonadal somatic cells 140

5.3 Conclusions and Discussion 145

5.3.1 Aims revisited 145

5.3.2 Sf1-EGFP can be used to enrich for somatic gonadal cells expressing *Gadd45g* and *Sry* 145

5.3.3 Multiple Displacement Amplification can be used to study DNA methylation from low amounts of DNA 148

Chapter 6: The methylomes of gonadal somatic cells at the sex-determining stage and the role of *Gadd45g* in its establishment 150

6.1 Introduction 151

6.1.1 Post-bisulphite adaptor tagging (PBAT) 151

6.1.2 Analysis of whole genome bisulphite sequencing data 153

6.1.3 Differentially Methylated Region (DMR) Detection 154

6.1.4 Aims and Hypotheses 155

6.2 Results 155

6.2.1 Generation of WGBS libraries from gonadal somatic cells using PBAT 155

6.2.2 Analysis of WGBS and alignment 158

6.2.3 Dynamics and global trends of DNA methylation in mouse embryonic gonadal somatic cells 162

6.2.4 The methylation state around known sex determination genes 163

6.3 Conclusions and Discussion 166

6.3.1 Aims revisited 166

6.3.2 No detectable differences in the methylomes of XY, XX or XY *Gadd45g*$^{-/-}$ somatic gonadal cells at 18 ts 166

6.3.3 *Sox9* and *Foxl2* are hypomethylated across their gene bodies 173

Chapter 7: General Discussion 175

7.1 Overview 176

7.2 Key Findings 177

7.2.1 *E15-Gadd45g* is a functional overexpression construct 177

7.2.2 *E15-Gadd45g* is amenable to modification through Red/Et mediated recombineering 178

7.2.3 The methylomes of 11.5 dpc (18 ts) somatic cells from XX, XY and XY *Gadd45g*$^{-/-}$ gonads show no major differences 179

7.3 Future Work 181

7.3.1 A role for *Gadd45g* in active DNA demethylation remains theoretical 181

7.3.2 Establishing an epigenetic profile of gonadal somatic cells during sex determination 184

7.3.3 Closing remarks 185

References 187

Appendices 205

List of Figures

1.1 Diagram showing the location of the urogenital ridge at 10.5dpc

1.2 Specification of cells and the formation of the gonadal ridge

1.3 The cell types within the developing gonads

1.4 S*ry* expression timing and levels are critical for testis determination

1.5 Molecular pathways of mammalian sex determination

1.6 Known regulators of *Sry* expression and their actions

1.7 The reactions governing methylation and demethylation of cytosine

3.1 Location of the sequences encoded in the *E15-Gadd45g* and *K5-Gadd45g* BACs

3.2 Generation of transgenic mice using BACs

3.3 Testing the functionality of *Gadd45g* BACs by crossing onto the B6.Y^{AKR} strain

3.4 Overexpression of *E15-Gadd45g* rescues B6.Y^{AKR} *Gadd45g*$^{+/-}$ sex reversal

3.5 *E15-Gadd45g* induces a 5.2 fold increase in *Gadd45g* and does not increase expression levels of *Sry* expression at 18 ts

3.6 Expression of the probe *Gadd45gΔ* and the sense strand probe produced in conjunction.

3.7 Expression of *Gadd45g* from the *E15-Gadd45g* BAC is overexpressed at 10.5 dpc

3.8 Expression of *Gadd45g* from the *E15-Gadd45g* BAC is limited to the body of gonad at 11.5dpc

3.9 *Gadd45g* expression from E15-Gadd45g is not found in the coelomic epithelium

3.10 Expression of endogenous *Gadd45g* has ceased at 13.5 dpc but continues from the *E15-Gadd45g* BAC

3.11 *E15-Gadd45g* rescues testis development in B6.Y^{POS}

4.1 Overview of Red mediated recombination

4.2 Overview of Red mediated recombination

4.3 Integration of the rpsl-neo selection cassette

4.4 Generation of 500bp homology arm flanked *His/FLAG* sequence

4.5 Confirmation of the integration of the *His/FLAG* sequence prior to the translational stop codon of *Gadd45g*

4.6 Sequencing confirmation of integration of 500bp homology arms and the *His/FLAG* sequence

4.7 *ScaI* digest of unmodified *E15-Gadd45g*, *E15-rpsl-neo* and *R15-Gadd45g-His/FLAG* have no other major recombination events

4.8 Genotyping of transgenic founders for *E15-Gadd45g-His/FLAG*

5.1 Expression of *Sf1-EGFP* in the developing gonad

5.2 Gating used to sort *Sf1-EGFP* positive and negative cells

5.3 Quantitive Real-time PCR analysis of sorted cells

5.4 Bisulphite sequencing data of the proximal *Sry* promoter in FACS sorted cells

5.5: Principles of Multiple Displacement Amplification

5.6 Methylation status of the *Sry* promoter with or without MDA treatment

6.1 Post-bisulphite adaptor tagging (PBAT) main principles

6.2 Bioanalyser gel images from generated PBAT libraries

6.3: A schematic computational pipeline of PBAT Bisulphite Sequencing data

6.4: Read coverage and average read depth of sequence alignments

6.5: Examples of Sex Determining genes and methylation in the surrounding area

Appendix I: Tail somite staging of mouse embryos

List of Tables

2.1 List of standard reagent solutions

2.2: Primer sequences for general genotyping of mouse colonies

2.3 Primers and sequences used in BAC recombineering

2.4 Primers and sequences used in targeted bisulphite sequencing

2.5 Primers and sequences used in Post-bisulphite adaptor tagging

2.6 Primers and sequences used to genotype TA cloning products

2.7 Primers and sequences used in qRT-PCR

2.8 Primers and sequences used in ddPCR

4.1 Statistics on the generation of transgenic embryos from modified version of the *E15-Gadd45g* BAC

6.1: Concentration of PBAT libraries according to KAPA qPCR

6.2: Summary of Pipeline results and CpG methylation percentages

6.3. Number of differentially methylated cytosines detected

6.4. Hypo- & hyper-differentially methylated bases between XY & XX, with a methylation difference >25% in the XY mice and a q-value = 0.05

Abbreviations

AID	Activation induced deaminase
AMH	Anti-Müllerian hormone
AP site	Apyramidic/apurinic site
BAC	Bacterial artificial chromosome
BC-DNA	Bisulphite-converted DNA
BER	Base excision repair
Bp	Base pairs
CGD	Complete gonadal dysgenesis
CRISPR	Clustered regularly interspaced palindromic repeats
ChIP	Chromatin immunoprecipitation
ddPCR	Digital droplet- polymerase chain reaction
DHH	Desert Hedgehog
DMC	Differentially methylated cytosine
DMR	Differentially methylated region
DMRT1/*Dmrt1*	Double-sex and Mab-3 related transcription factor 1
DNA	Deoxyribonucleic acid
DNMT	*De novo* methyltransferase
DOT1L/*Dot1l*	Disruptor of telomeric silencing 1-like
DPC	Days post coitum
DPO	Days post ovulation
DSB	Double-strand break
DSD	Disorder of sexual development
dsDNA	Double-stranded DNA
EGFP	Enhanced-GFP
EMT	Epithelial-mesenchymal transition
FACS	Fluorescence-activated cell sorting
FDR	False discovery rate
FGF	Fibroblast growth factor
FGFR	Fibroblast growth factor receptor
FOG	Friend of GATA
FOXL2/*Foxl2*	Forkhead box L2
FSC	Forward scattered light
GADD45γ/*Gadd45g*	Growth arrest and DNA damage inducible protein 45 gamma
GD	Gonadal dysgenesis
gDNA	Genomic DNA
H3K4	Histone 3 lysine 4
H3K9	Histone 3 lysine 9
HISS	Heat inactivated sheep serum
HMG	High mobility group
HPRT1/*Hprt1*	Hypoxanthine phosphoribosyltransferase 1
IGFR1/*Igfr1*	Insulin-like growth factor receptor 1
INSR/*Insr*	Insulin receptor
JMJD1A/*Jmjd1a*	Jumonji domain –containing 1A
JNK	c-JUN amino terminal kinase
Kb	kilobases
LB	Luria Bertani broth
LHX9/*Lhx9*	Lim homeobox protein 9
MAPK	Mitogen activated protein kinase
MAP2K	MAP kinase kinase

MAP3K	MAP kinase kinase kinase
MBP	Methyl-binding protein
MDA	Multiple displacement amplification
miRNA	Micro RNA
NER	Nucleotide excision repair
NEO/*Neo*	Neomycin resistance gene
ORF	Open reading frame
PBAT	Post-bisulphite adaptor tagging
PCR	Polymerase chain reaction
PGC	Primordial germ cells
PGD2	Prostaglandin D2
PK	Proteinase K
PMC	Peritubular Myoid cell
PRC	Polycomb repressive complex
qPCR	Quantitive PCR
qRT-PCR	Quantitive real-time reverse-transcriptase PCR
RNAseq	RNA sequencing
RRBS	Reduced representation bisulphite sequencing
RSPO1/*Rspo1*	R-spondin 1
SSC	Side scattered light
ssDNA	single-stranded DNA
SF1/*Sf1*	Steroidogenic factor 1
siRNA	Small-interfering RNA
SIX	Sine occulus homeobox
SOX9/*Sox9*	SRY box protein 9
SRY/*Sry*	Sex-determining region of the Y-chromosome
TBE	Tris/borate/EDTA buffer
TBS	Tris-buffered saline
TBST	Tris-buffered saline with tween
TDG	Thymine deglycosylase
TES	Testis specific enhancer of *Sox9*
TESCO	TES core element
ts	Tail somites
UGR	Uro-genital ridge
UTR	Untranslated region
WGA	Whole genome amplification
WGBS	Whole genome bisulphite sequencing
WMISH	Wholemount *in situ* hybridisation
WNT4/*Wnt4*	Wingless-type MMTV integration site family type 4
WT	wild-type
WT1/*Wt1*	Wilm's Tumour suppressor 1
XP	Xeroderma pigmentosum

Chapter 1. Introduction

1.1 Overview

Sexually reproducing species require individuals to develop as male or female. This process of sexual differentiation can be initiated by various means. In some species of reptiles and fish, environmental stimuli such as temperature control which sex an animal will develop towards (Warner and Shine, 2008). In mammals the process is controlled by the sex chromosomes and therefore is effectively initiated at fertilisation. Zygotes will commonly inherit either a single copy each of the X and Y chromosomes or two copies of the X chromosome, establishing the chromosomal sex as XY or XX respectively. Later in development, the process of primary sex determination takes place, in which the embryonic gonad commits to either the testicular or ovarian fate, according to the chromosomal sex of the embryo. Usually, XY embryos will develop testes whilst XX embryos will develop ovaries, reviewed in (Wilhelm et al., 2007). The establishment of gonadal sex then drives subsequent development of the internal and external genitalia towards a male or female fate through the production of organ specific androgens (Wilhelm and Koopman, 2006).

The embryonic gonad is initially bipotential, meaning it has the capability to develop as either a testis or ovary irrespective of chromosomal sex. The primary instigator of mammalian sex determination is the Y-linked gene *SRY*. The expression of *SRY* will drive the gonad toward the testicular fate whilst its absence will result in the development of ovarian tissue. In mice *Sry* expression in the gonad is first detected at 10.5dpc and reaches peak expression at 11.5dpc, after which it begins to fall and by 12.5dpc is absent from the gonad (Bullejos and Koopman, 2001). The expression of *Sry* leads to the establishment of testicular fate and the inhibition of ovarian fate (Sinclair et al., 1990; Koopman et al., 1991). The only known direct target of SRY, *Sox9*, encodes a transcription factor which induces a feedback loop maintaining its own expression as well as the maintenance of the testicular pathway (Sekido and Lovell-Badge, 2008). SOX9 also inhibits components of the ovarian pathway and maintains the pathway after *Sry* expression has ceased (Kashimada and Koopman, 2010).

In human embryos, the gonads begin to form from 33 days post-ovulation (dpo). *SRY* is first detected in XY gonads between 41 – 44 dpo and reaches peak expression levels at 44 dpo (Hanley et al., 2000). Mutations in various genes involved in these pathways can cause a discrepancy between the chromosomal and gonadal sex during development - a phenomenon known as gonadal sex reversal. This can manifest as a variety of human disorders which are collectively known as disorders of sex development (DSDs). These can include congenital adrenal hypoplasia; cryptorchidism and 46, XY gonadal dysgenesis (Hughes, 2008). It has been estimated that only approximately 20 - 30% of cases of 46, XY gonadal dysgenesis are given a specific molecular diagnosis (Hughes et al., 2006). It is therefore imperative to better understand the network of genes that regulate expression of *SRY*, functionally downstream targets and other components of the testis determination pathway in both humans and mice. The mouse is an excellent model of human sex determination: it has a similar developmental biology and many of the genes known to function in human sex determination, such as SRY and SOX9, also do so in mice.

The expression of mouse *Sry* is required to reach a specific threshold by a specific time point to ensure proper testis determination (Hiramatsu et al., 2009). A failure to do so can lead to a range of phenotypes in mice, from the formation of ovotestis to complete gonadal and phenotypic sex reversal. It has previously been shown that the gene *Growth arrest and DNA damage 45 Gamma (Gadd45g)* is required for the timely expression of *Sry* within the murine sex determination system (Gierl et al., 2012; Warr et al., 2012). The research presented within this thesis attempts to further understand the role *Gadd45g* plays in murine sex determination, with a focus on the role that *Gadd45g* potentially plays in active DNA demethylation (Barreto et al., 2007) and the production of transgenic *Gadd45g* mouse lines aimed at improving our ability to detect GADD45γ *in vivo*. This

introductory chapter will outline the current understanding of murine testis determination, the known roles of *Gadd45g* and epigenetic regulation during sex determination.

1.2 Mammalian sex determination

1.2.1 The formation of the gonadal ridge

The origins of the gonads lie in the formation of the urogenital ridge (UGR). This structure is derived primarily from the intermediate mesoderm in the early embryo and runs along the length of the embryo either side of the midline and dorsal aorta, forming at around 9.0dpc (Fig 1.1). The UGR can be further divided into three sections from anterior to posterior: the pronephros, mesonephros and metanephros. The pronephros later degrades in mammals; the mesonephros acts as a single nephron during embryogenesis which is superseded by the formation of the kidney from the metanephric mesenchyme (Stark et al. 1994). Along the UGR runs the mesonephric (or Wolffian) duct. Later, a parallel duct, the paramesonephric (or Müllerian) duct, forms from migratory cells originating from the coelomic epithelium, which separates the UGR from the rest of the body cavity (Swain and Lovell-Badge, 1999).

Around 9.5 dpc, cells within the coelomic epithelium proliferate and undergo epithelial-to-mesenchymal transition (EMT) (Tanaka and Nishinakamura, 2014). This newly designated mesenchyme then protrudes away from the mesonephroic mesenchyme leading to the formation of the gonadal ridge. Several transcription factors are required for the proper formation of the gonadal ridge, including empty spiracle homeobox 2 (*Emx2*)(Miyamoto et al., 1997) GATA-binding protein 4 (*Gata4*)(Hu et al., 2013), Lim homeobox 9 (*Lhx9*)(Birk et al., 2000), steroidogenic factor 1 (*Sf1*: also known as *Nr5a1*)(Achermann et al., 1999) , Wilms tumour 1 (*Wt1*) (Miyamoto et al., 1997),

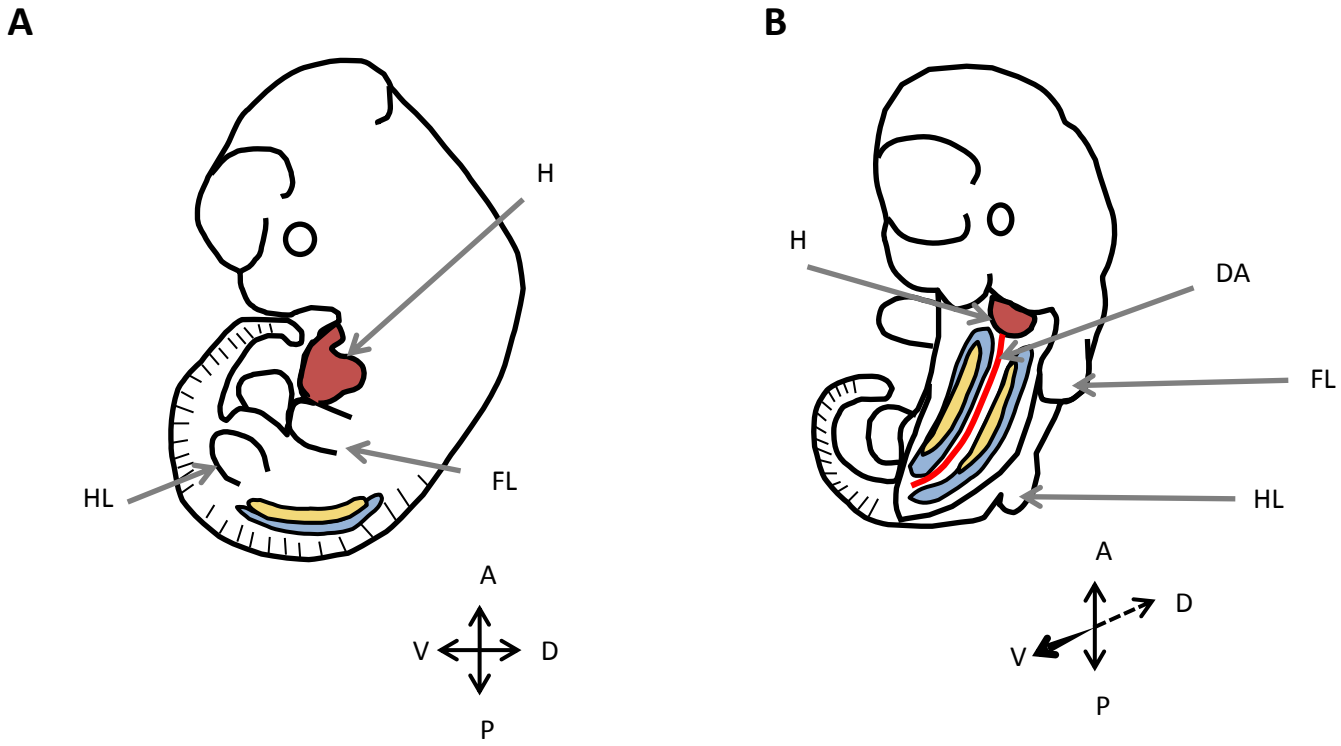


Figure 1.1 Diagram showing the location of the urogenital ridge at 10.5dpc

A) The urogenital ridges are derived from the intermediate mesoderm along the dorsal part of the body cavity. The coelomic epithelium of the mesonephros, light blue, begins to thicken from 9.5dpc, forming the gonadal ridge, yellow.

B) The gonadal ridges form on the ventromedial surface of the mesonephroi, which lie either side of the dorsal aorta (DA). Heart (H), Forelimb (FL), hindlimb(HL)

chromobox 2 (*Cbx2*: also known as *M33*)(Katoh-Fukui et al., 1998), insulin receptor (*Insr*) and insulin-like growth factor type I receptor (*Igfr1*)(Pitetti et al., 2013). Whilst many of these genes are essential for the proper formation of the gonadal ridge, they also appear to be required for the timely expression of *Sry* and other factors involved in sex determination (outlined in Fig 1.2).

SF1 is a member of the orphan nuclear receptor family and acts as a transcription factor containing two zinc finger domains. It is required for the regulation of steroidogenic enzymes as well as the early stages of gonad formation. Mice lacking *Sf1* die by postnatal day 8 from adrenocortical insufficiency. Homozygous animals lacked adrenal glands and gonads, and developed externally as females (Luo et al., 1994). Expression of *Sf1* within the coelomic epithelial layer of both XX and XY embryos at 9.5dpc, marks the initial cell population which will develop into the somatic cells of the gonadal ridge. Several factors are involved in the initial expression and maintenance of *Sf1*, which is thought to be required for the continued survival of cells within the developing gonad (Tanaka and Nishinakamura, 2014). Later, *Sf1* contributes towards the timely expression of *Sry*, which will be discussed in a later chapter.

GATA4 is a transcription factor that is required for the early development of multiple organs including the heart, liver and pancreas. Expression of *Gata4* precedes the initial thickening of the coelomic epithelium at 9.5dpc, progressing in an anterior to posterior direction. Inducible knockouts of *Gata4* fail to induce the coelomic epithelium thickening and do not express *Lhx9* or *Sf1*. Due to the lack of expression of other genes required for the formation of the gonadal ridge and its early expression within the mesonephros, *Gata4* is thought to be the initiating factor for the development of the gonadal ridge (Hu et al., 2013).

First identified at a locus associated with Wilm's tumour, *Wt1* encodes a four zinc fingered protein containing proline and glutamine rich regions (Call et al., 1990). It has since been shown to have an essential role in the development of the urogenital system. Mice deficient in *Wt1* fail to develop a gonadal ridge by 12.5dpc and the metanephoric cells, which form the embryonic kidney, undergo early apoptosis leading to the death of the embryo (Kreidberg et al., 1993). Human patients carrying mutations in *WT1* show a variety of conditions including Denys-Drash syndrome, Frasier syndrome and mesothelioma (Englert, 1998) (Pelletier et al., 1991). Patients of Denys-Drash syndrome, caused by premature truncations and missense mutations within the zinc finger domains, show urogenital abnormalities. Frasier syndrome patients, who carry mutations within intron 9, show XY gonadal dysgenesis. These mutations appear to affect the recognition of an alternative splice donor site and cause 46,XY gonadal dysgenesis. This splice site controls the insertion or deletion of three amino acid residues between zinc fingers 3 and 4, designated as the +KTS and –KTS isoforms respectively. Mouse models in which the +KTS isoform has been ablated, show similar phenotypes to Frasier syndrome patients including XY sex reversal. WT1 can act as both a transcriptional repressor or activator, though these roles are thought to be due to the high number of isoforms that the gene produces through alternative transcription start sites, alternative splice sites and RNA editing (Hammes et al., 2001)

Lhx9 is a member of the LIM homeobox domain gene family of transcription factors (Hobert and Wesphal, 2000). It is first expressed in the medial surface of the gonadal ridge at 9.5dpc and later in the coelomic mesothelial lining and the nearby mesenchyme at 11.5dpc. Both XX and XY mice lacking *Lhx9* develop as infertile females, exhibit atrophied uteri, vaginas and oviducts and no oestrus cycle (Birk et al., 2000). Despite high levels of *Lhx9* expression in the embryonic nervous system and limbs, no phenotypes were found in these regions. At 11.5dpc, the gonads of *Lhx9*$^{-/-}$ mice appear normal, by

12.0dpc they appeared smaller compared to wild-type littermates, and by 13.5dpc appeared to lack gonads. BrdU staining experiments at 11.5dpc show that the large amounts of cell proliferation, expected at this time within the gonad, are absent in *Lhx9*$^{-/-}$ gonads, a possible reason for the failure for the gonadal ridge to survive. Tissue from *Lhx9*$^{-/-}$ gonads showed decreased levels of expression of *Sf1* but not *Wt1* suggesting that it lies upstream of *Sf1* but downstream of *Wt1* in the gonadal ridge formation pathway (Birk et al., 2000) (Fig 1.2).

Gata4 is often associated with members of the friends of GATA (FOG) family. One such factor, Fog2, has been shown to be required for proper expression of *Sry*, but does not appear to affect gonad size (Tevosian et al., 2002), indicating it may not have a role in gonadal ridge development. However, double knockouts of the two transcription factors *Sine oculis homeobox* (*Six*) *1* and *Six4*, which upregulate *Fog2*, show XY male-to-female sex reversal and reduced gonad size. The latter phenotype is likely to result from a downregulation of *Sf1*, and a failure of the gonadal primordium to grow. Despite the presence of an *Sry* transgene, which rescued the sex reversal phenotype, the gonads still appeared smaller than wild-type gonads, indicating a role for *Six1* and *Six4* in the growth of the developing gonad independent of testis determination (Fujimoto et al., 2013).

Insulin signalling pathways, and the related insulin like growth factors, are imperative for a number of physiological processes including metabolism, differentiation, cell survival and proliferation. Signals are mediated to the target cells through the actions of the insulin receptor (*Insr*) and insulin-related growth factor type 1 receptor (*Igf1r*). Double knockouts of both receptors cause male to female sex reversal in mice, as well as slower rates of embryonic growth (Louvi et al., 1997). At 16.5dpc XY double KO gonads were histologically indistinguishable from XX gonads and resembled ovaries though both sexes were smaller than the wild-type counterparts. Low levels of *Sf1*, *Wt1* and *Lhx9*

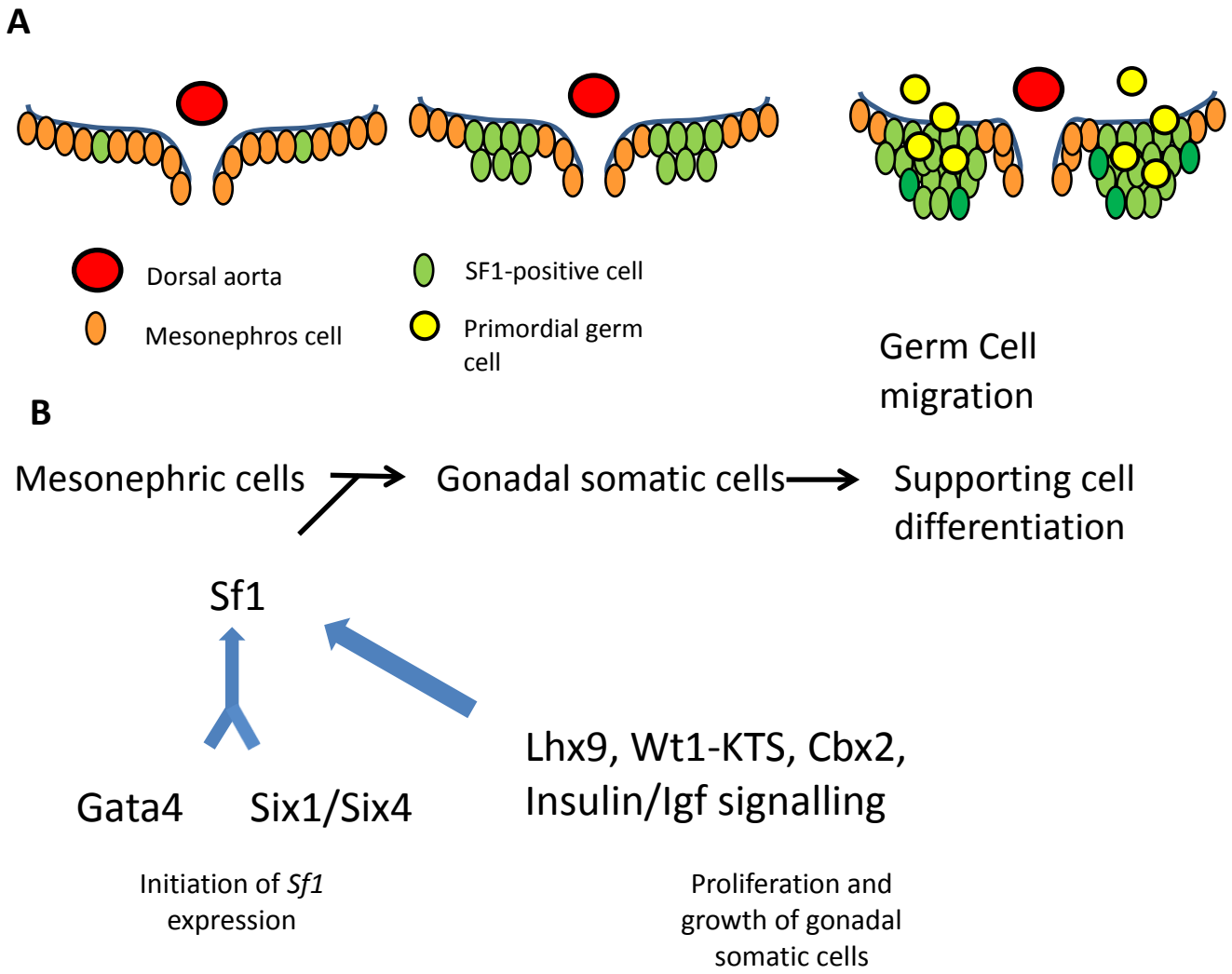


Figure 1.2 Specification of cells and the formation of the gonadal ridge

A) Cells within the coelomic epithelium of the mesonephros begin to express *Sf1* at 9.5dpc. As the cells begin to proliferate, they extend away from the mesonephros. Migrating PGCs, enter the gonadal ridge from 10.5dpc.

B) The expression of *Sf1* is initiated by GATA4 and SIX1/SIX4 at 9.5dpc. This expression, along with the gonadal somatic cell identity is maintained by LHX9, WT1 (-KTS), CBX2 and insulin/IGF signalling acting through INSR1 and IGFR2

expression were detected in both XX and XY gonads prior to sex determination, which showed approximately 39% fewer somatic gonadal progenitor cells at 11.5dpc (Pitetti et al., 2013). This lack of proliferation can lead to fewer cells which will later express *Sry* (Kashimada and Koopman, 2010). The sex reversal phenotype therefore appears to result from the impaired growth of the primordial gonad, similar to the phenotype of knockouts of other factors involved in this process. This study also revealed that the double KO of *Insr* and *Igf1r* caused downregulation of members of the mitogen activated protein kinase pathway and associated regulators, including *Map3k4* and *Gadd45g* (Pitetti et al., 2013). These two genes have a critical role in sex determination which will be elaborated later.

CBX2 is a member of the polycomb group protein family. First found as a paralog of *Drosophila* polycomb genes, *Cbx2* forms a part of the polycomb-repressive complex 1 (PRC1), which controls epigenetic gene regulation through the recognition of methylated histone marks and methylation of histones. In humans, the *CBX2* gene has been found to be essential for sex development. A phenotypically female patient with 46,XY complete gonadal dysgenesis (CGD) had ovaries showing a normal tissue structure and thought to be a result of compound heterozygous mutations in *CBX2* (Biason-Lauber et al., 2009). Mice lacking both copies of *Cbx2* had a higher than 50% mortality rate before weaning and showed numerous skeletal defects. The XY mice showed an external female phenotype, but examination of the internal gonads revealed a variety of phenotypes, such as ovaries with follicles, indistinct gonads and a mix of testicular and ovarian tissue. XX gonads developed as smaller ovaries or were completely absent. At 11.5dpc mutant gonads were morphologically indistinguishable from wild-type. However at 13.5dpc, gonads in both sexes were noticeably smaller and appeared to be a similar size to those at 11.5dpc. (Katoh-Fukui et al., 1998). Microarray expression profiling analyses on gonads from 11.5dpc XY $Cbx2^{+/+}$, XX $Cbx2^{+/+}$, XY $Cbx2^{-/-}$ and XX $Cbx2^{-/-}$ revealed the downregulation of many genes encoding known sex differentiation transcription factors

including *Dax-1*, *Dmrt1*, *Gata4* and *Lhx9* in *Cbx2*$^{-/-}$ in mice of both sexes. It was subsequently found that a low number of cells were expressing the markers for gonadal ridge growth at 10.5dpc and 11.0dpc and that proliferation of the mesenchymal cells was reduced in the *Cbx2*$^{-/-}$ in both sexes. (Katoh-Fukui et al., 2012). Chromatin immuno-precipitation (ChIP) experiments conducted on Y-1 cells, a line derived from mouse adrenal gland, showed that CBX2 directly bound to the *Sf1* locus (Katoh-Fukui et al., 2005). While not required for the initial growth of the gonadal ridge, it appears that *Cbx2* is essential for the continued growth of the gonad out from the mesonephros.

The gonadal ridge develops in a similar manner in both XX and XY gonads. It is not until a later stage that differences start to emerge. Until then the gonad has the capacity to develop towards either of the testicular or ovarian fates and hence is described as being in a bipotential state.

1.2.2 The gonad as a bipotential organ

Before sex determination, the gonads of XX and XY gonads are poised in a state of readiness to differentiate into either a testis or ovary. Signalling cascades involving Wingless-type MMTV integration site family member 4 (*Wnt4*) and Fibroblast growth factor 9 (*Fgf9*), genes which are later expressed exclusively in ovarian or testicular tissue respectively, appear to maintain a balance in the bipotential gonad until sex determination takes place (Kim and Capel, 2006). As the gonadal ridge forms, several different cell lineages begin to arise from the initial mix of coelomic epithelial cells and transitioned mesenchymal cells (Fig 1.3). Each of these cell types retains the ability to differentiate towards two different fates depending on the surrounding environmental signals.

Among these cell types are the supporting cell precursors, which will differentiate into Sertoli cells in a testis and granulosa cells in an ovary. These cell types are the first to differentiate and send signals to the surrounding cells to differentiate towards their sex-specific fate. Later in development,

supporting cells will surround the germ cells, nurturing their development as either spermatogonia or oocytes. Another of the somatic cell populations comprises the precursors of steroidogenic cells, which will later produce the hormones which establish the secondary sex characteristics in the rest of the body. In the testis, these cells will differentiate into Leydig cells, which will later produce the precursor molecules for testosterone; whilst in ovaries, theca cells are formed, which will produce oestrogen precursors (Wilhelm and Koopman, 2006). A population of connective cells also reside in the gonad. In both sexes they assist in the structural formation of their respective organ. In the testis the connective cells will form peritubular myoid cells, which surround the Sertoli cells and contribute towards the formation of the testis cords and basal lamina (Swain and Lovell-Badge, 1999). It is still not clear, however, if the somatic cell lineages are already distinct from each other at the bipotential stage of gonad development, or if they diverge from somatic precursors due to signals from the differentiating supporting cells (Kim and Capel, 2006).

As the gonadal primordium grows away from the mesonephros, primordial germs cells (PGCs) migrate into the gonad. PGCs are first specified during gastrulation in the posterior of the primitive streak in a subset of epiblast cells expressing PR domain zinc finger protein 1 (*Prdm1*) at 6.5dpc. By 7.5dpc the PGCs have begun to migrate towards the definitive endoderm and at 8.0dpc they have relocated to the hindgut, where they undergo global demethylation thus removing the majority of inherited epigenetic marks (Guibert et al., 2012). By 11.5dpc most of the PGCs have migrated along the gut mesentery and into the gonadal ridge, where they remain and later differentiate into either spermatogonia or oocytes (Molyneaux et al., 2001). This process is dependent on the surrounding cellular environment and is not influenced by the chromosomal content of the individual germ cells. Germ cells are not required for the development of the testis, or the early development of the ovary. However, the presence of germ cells is necessary for the proper formation of ovarian follicle at

perinatal stages and postnatal maintenance of the ovarian structure (Hashimoto et al., 1990). These defects are thought to be due to the absence of germ cells rather than defects in ovarian somatic cells (Maatouk et al., 2012).

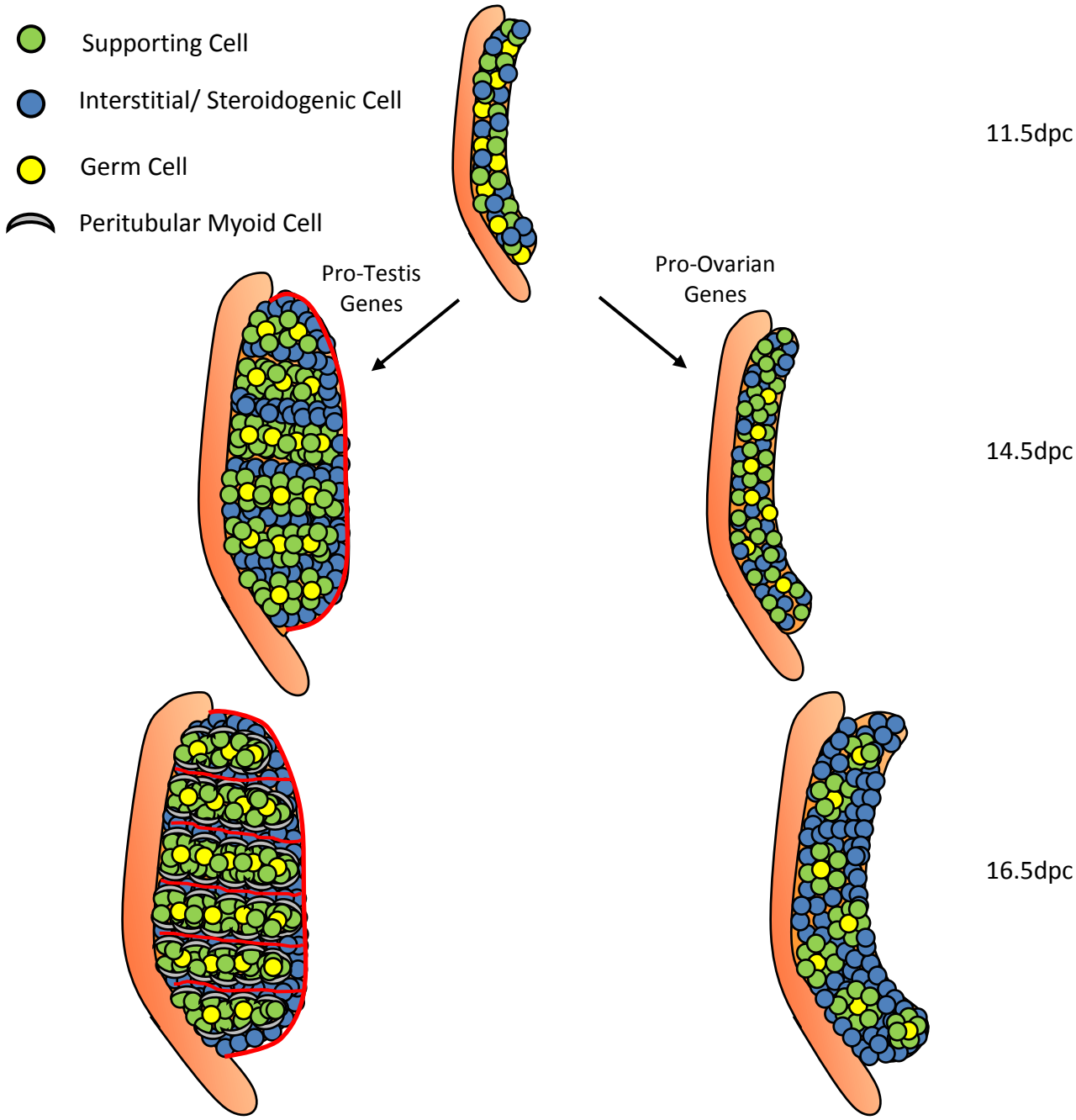


Figure 1.5 The cell types within the developing gonads
Diagram of various cell types within the developing gonad at different stages. Testes and ovaries share a similar morphological structure at 11.5dpc. By 12.5dpc, the supporting cells of the testis have coalesced around the germ cells and begin to form testis cords. At 16.5 the testis cords are surrounded by peritubular myoid cells. In the ovaries the supporting granulosa cells begin to coalesce around the germ cells, which will develop as follicles.

1.2.3 *Sry* and the establishment of the testis

The presence of a dominant testis-determining factor on the Y chromosome was first posited in the 1950s after a series of experiments investigating the phenotypes caused by different aneuploidies of sex chromosomes (Welshons and Russell, 1959). In the 1990s, a series of mapping experiments investigating sequences common to both XY and XX males had limited the testis determining factor (*TDF*) to a 35 kilobase (kb) region of the Y chromosome (Sinclair et al., 1990). It was within this region that a testis-specific transcript was found, which was designated Sex-determining region Y (*SRY*) (Sinclair et al., 1990). A 14kb DNA fragment containing *Sry* was shortly after found to be sufficient to induce the development of testes in XX mice (Koopman et al., 1991). Together, these results established *SRY* as the master sex determination factor in humans, mice and other subsequently studied mammals, including marsupials (Harry et al., 1995; Parma et al., 1999; Warr and Greenfield 2012).

The process of sex determination in mice begins around 10.5dpc (11 ts) with the onset of *Sry* expression in XY embryonic gonadal somatic cells, which initiates the testicular pathway and promotes the expression of pro-testis genes. The absence of *Sry* in XX individuals, and the subsequent lack of pro-testis expression, leads to the upregulation of the ovarian pathway. A series of various genes present in the gonads at this time lead to segments of the population of somatic cells within the XY gonad begin expressing *Sry*, which acts in a cell autonomous manner to induce the differentiation of precursor Sertoli cells. This expression marks the first divergence between XX and XY gonads. As sex determination continues, pre-Sertoli cells will continue to play a key signalling and organisational role in the formation of the testis (Kashimada and Koopman, 2010).

Sry is the founding member of the SRY-related HMG box (SOX) protein family and encodes a nuclear high-mobility group (HMG) domain-containing transcription factor, which allows SRY and related proteins to bind the sequence (A/T)ACAA(T/A) and induce a 65 - 80° bend in the DNA. The HMG domain is flanked by nuclear localisation sequences and when compared to the rest of the protein shares a much higher rate of conservation between species. The importance of this region can also be demonstrated by the high incidence of mutations found in human DSD patients relative to other regions of *SRY* (Kashimada and Koopman, 2010). The expression pattern of *Sry* in mice shows a wave like profile originating in the centre of the gonadal ridge and extending in tandem to the poles. By 11.5dpc (18 ts), expression is observed throughout the entire gonad. Expression of *Sry* then begins to wane and by 12.5dpc (27 ts) has ceased in the newly formed testis. (Bullejos and Koopman, 2001). To induce Sertoli cell differentiation, *Sry* is required to reach a specific threshold level of expression within a specific time window (Hiramatsu et al., 2009). In addition to cellular threshold of Sry expression, it is also important to recruit a certain minimum percentage of cells in the gonad to the Sertoli fate, to ensure proper testis development (Fig 1.4). In XX and XY chimeric mice containing approximately 35 – 40% XY cells, testes are formed. However embryos containing less than 10% of XY somatic cells form ovaries instead (Palmer and Burgoyne, 1991).

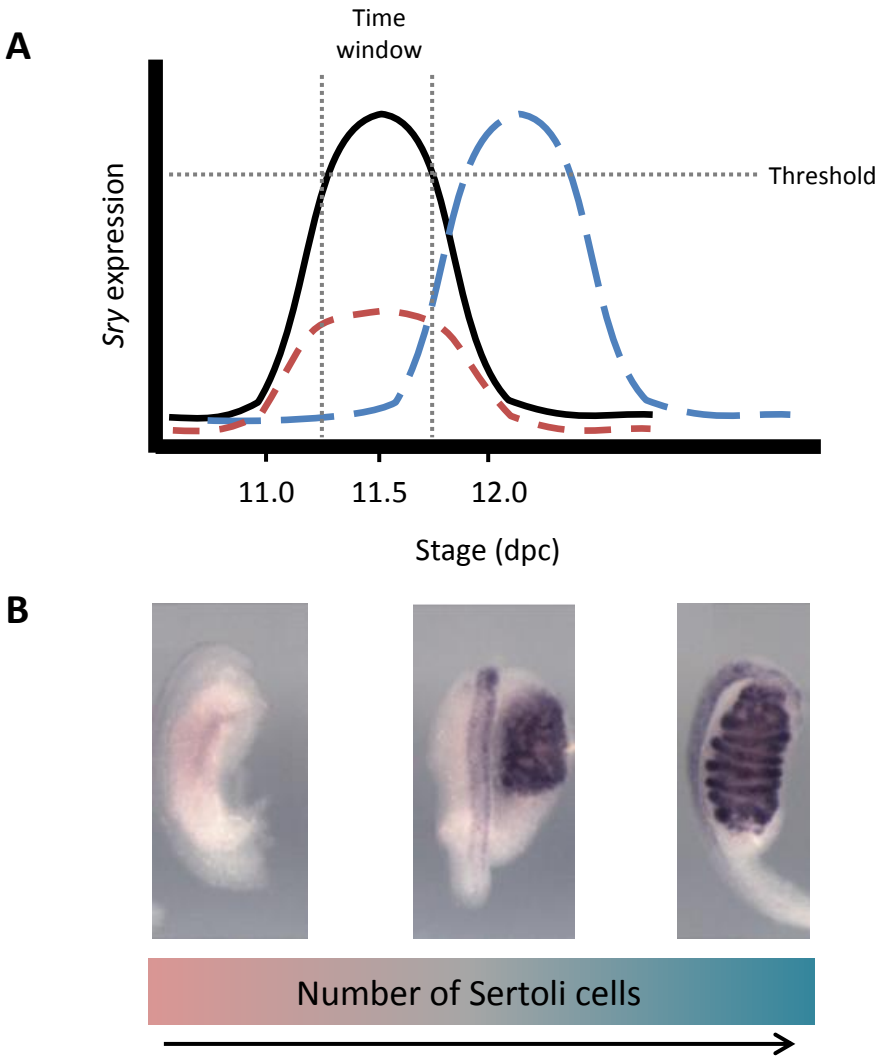


Figure 1.4 *Sry* expression timing and levels are critical for testis determination

A) *Sry* expression must reach a specific threshold level within a specific time window in order to induce Sertoli cell differentiation and testis development. Failure to reach the threshold (red dashed line), or a delay in *Sry* expression (blue dashed line), can result in insufficient numbers of Sertoli cells to drive normal testis development (Adapted from (Kashimada and Koopman, 2010)

B) *In situ* hybridisations of the Sertoli cell marker Sox9. The number of Sertoli cells within a gonad determines the sexual fate of the organ. Lower levels of Sertoli cells can impede the formation of testes (right), leading to the development of ovotestes (centre) or ovaries (left). WMISH images courtesy of Dr Nick Warr.

The SOX family is comprised of 20 members in both mice and humans that typically act as cell differentiation switches, playing important roles in many aspects of organogenesis. The transcription factor *Sox9* in mice is located on chromosome 11. SOX9 has previously been implicated in the differentiation of mesenchymal cells to chrondrocytes; mutations within *SOX9* are linked to cases of campolemic dysplasia (CD). Patients often also exhibit XY gonadal dysgenesis, indicating a role for *SOX9* in the testis determining pathway (Lefebvre and Crombrugghe, 1998). SOX9 is found at low levels in the somatic cells of both XX and XY gonads during the bipotential stage and is absent from the nucleus (Sekido and Lovell-Badge, 2008). However, soon after expression of *Sry*, it is strongly upregulated in XY gonads and downregulated in the XX gonads. XY mice lacking *Sry* lose expression of *Sox9* and transgenic *Sry* XX gonads show increased *Sox9* expression. This huge upregulation is initiated by the binding of SRY, in association with SF1, to a 3.2kb region 13 – 10kb 5' to the *Sox9* transcription start site termed the testis-specific enhancer of *Sox9* (TES). This region was later narrowed down to a 1.4kb conserved region termed the TES core element (TESCO) (Sekido and Lovell-Badge, 2008). In a similar fashion to *Sry*, *Sox9* has been shown to be essential for testis development (Barrionuevo et al., 2006) and can induce testicular tissue in XX individuals (Vidal et al., 2001). *Sox9* also shows a similar expression pattern to *Sry*, beginning in the centre of the gonad and moving out to the poles in a dynamic wave. While the expression of *Sry* falls and is undetectable in the gonads at 12.5dpc, expression of *Sox9* remains consistent, suggesting that it usurps *Sry* as the major driver of testis formation through the differentiation of pre-Sertoli cells and acts as a key factor in the continued identity of Sertoli cells (Kashimada and Koopman, 2010).

After the initial upregulation, a series of positive feedback loops is established to maintain *Sox9* expression. SOX9 can replace SRY and interact with SF1 to bind to TESCO and promote its own expression (Sekido and Lovell-Badge, 2008). Experiments using gonads from chimaeric XX-XY mice

showed that the vast majority of Sertoli cells observed were XY despite a similar number of cells of each karyotype in gonads. However, up to 10% of the Sertoli cells were found to be XX, suggesting another mechanism of recruitment to the Sertoli cell fate in addition to the cell-autonomous action of *Sry* and *Sox9* (Palmer and Burgoyne, 1991). Such recruitment is thought to be the result of paracrine signalling mediated by fibroblast growth factor 9 (FGF9) (Kim et al., 2006) and prostaglandin D2 (PGD2) (Moniot et al., 2009).

At 11.5dpc, *Fgf9* is found throughout the gonads of both sexes in mice. However by 12.5dpc, its expression is exclusive to XY gonads within cells in the coelomic epithelium and the developing testis cords. *Fgf9*$^{-/-}$ mice show normal expression of *Sry* and *Sox9* expression at 11.5dpc, but *Sox9* expression was not seen in XY gonads at 12.5dpc and these failed to express Sertoli cell markers such as Desert hedgehog (*Dhh*) (Kim et al., 2006). Similar phenotypes are also observed in mice carrying conditional knockouts or loss of function mutations in *Fgf receptor 2* (*Fgfr2*) (Kim et al., 2007; Siggers et al., 2014), identifying it as the likely target receptor of FGF9 in the gonads. These results show the requirement for FGF signalling in the maintenance of *Sox9* expression but not its initial upregulation. Fgf9 signalling also appears to have a role in the repression of the pro-ovarian WNT4 signalling pathway. These two signalling pathways, which were previously balanced in the early bipotential gonad, act at the sex-determining stage in a mutually antagonistic manner to promote the testicular or ovarian fate (Kim et al., 2006). This mutual antagonism of the testis- and ovary-determining pathways is thought to canalize gonad development towards one of the two desired developmental fates (Carré and Greenfield, 2016).

At the same time as FGF signalling, SOX9 upregulates the expression of lipocalin-type prostaglandin D2 synthase (*L-Pgds* or *Ptdgs*) which converts the hormone-like lipid prostaglandin H2 into

prostaglandin D2 (PGD2). PGD2 is expressed in a testis-specific manner and has been shown to upregulate testis-specific factors in explanted XX gonadal cells. Once sex determination occurs, PGD2 signalling stimulates the nuclear localisation of SOX9, thus increasing the expression of *L-Pgds* and further translocation of SOX9 to the nucleus. This feedback loop acts independently of FGF9 signalling, and appears to act as a backup to maintain *Sox9* expression (Moniot et al., 2009).

As pre-Sertoli cells differentiate, they coordinate a series of morphological and cellular changes throughout the gonad. From 11.5 to 12.5dpc, Sertoli cells begin to condense around the primordial germ cells (PGCs) and form visible cords, which will eventually form the seminiferous tubules found in the adult testis. The PGCs then enter a state of mitotic arrest and do not commence meiosis until after birth (Warr and Greenfield, 2012). Sertoli cells will still aggregate in the absence of germ cells or *in vitro* in the presence of an extracellular matrix (Combes et al., 2009). Somatic cells then undergo increased levels of proliferation, extending the gonad away from the mesonephros (Wilhelm and Koopman, 2006). During this time, endothelial cells from the mesonephros migrate into the gonad. It is thought that these cells help partition the clustered Sertoli and germ cells into cords across the gonad (Combes et al., 2009). At this time, Sertoli cells begin to express anti-Müllerian hormone (*Amh*). This leads to the regression of the Müllerian ducts within the mesonephros, the precursors to the fallopian tubes and uterus. As a result of testicular androgens the Wolffian ducts do not regress, but instead form the epididymis, vas deferens and seminal vesical (Wilhelm and Koopman, 2006).

The testis cords become coated in a layer of smooth muscle-like peritubular myoid cells (PMCs) at 13.5dpc. Originally hypothesised to have developed from the migratory mesonephric cells, it is now thought that PMCs originate from the interstitial cell population or cells from the coelomic epithelium which have undergone epithelial-to-mesenchymal transformation (Svingen and Koopman, 2013).

Between the growing cords lie the interstitial cells, which include a population of precursor cells that express Patched 1 (*Ptch1*). This receptor is activated by Desert hedgehog (*Dhh*) secreted by the Sertoli cells, triggering differentiation of fetal Leydig cells (Barsoum and Yao, 2010). Though these cells will be supplemented by adult Leydig cells derived from a progenitor cell type maintained by *Notch* signalling, the fetal Leydig cells begin the process of masculinisation of the rest of the embryo, such as promoting Wolffian duct development and effecting testicular descent (Welsh et al., 2009), through the production of testosterone (Svingen and Koopman, 2013).

The lack of *Sry* expression in XX gonads enables a pro-ovarian genetic program to predominate, as the lack of pro-testis genes does not lead to the subsequent inhibition of this pathway. The continued upregulation of the canonical WNT signalling pathway, initiated by R-spondin 1 (RSPO1) and WNT4, leads to the stabilisation of the intracellular signal transducer β-catenin (CTNNB1) (Chassot et al., 2014). Together with the transcription factor forkhead box L2 (FOXL2), these factors induce the differentiation of supporting cells towards the granulosa cell fate, as well as inhibit the differentiation of Sertoli cells (Uhlenhaut et al., 2009). As the ovarian pathway is established, the PGCs within the gonad enter meiosis. This process is triggered by retinoic acid signalling at 13.5dpc. Meiotic germ cells then begin to express the gene *Stra8*, which can be used as a marker of ovarian meiotic germ cells (Koubova et al., 2006). The molecular genetic pathways in sex determination are outlined in Fig 1.5.

By 13.5dpc the cells of the gonad have committed to the either the testicular or ovarian pathway. However these cells retain a certain amount of plasticity and require the maintenance of directive signals throughout life. In males doublesex and mab-3 related transcription factor 1 (*Dmrt1*) is expressed by pre-meiotic germ cells and Sertoli cells. Post-natal XY gonadal knockouts of *Dmrt1* show trans-differentiation of Sertoli cells into granulosa like cells (Matson et al., 2011). An opposite

phenotype (female-to-male) is seen in XX post-natal gonadal knockouts of *Foxl2* (Uhlenhaut et al., 2009), indicating that these two factors act to maintain the sexual fate of the supporting cell lineage.

It has since been shown that *Dmrt1* maintains the testicular fate by blocking the activation of female specific genes by retinoic acid signalling, a necessary component in the proper maturation of germ cells in both sexes (Minkina et al., 2014). As a result, it appears that sex determination is a life-long process requiring continued active signals to maintain the testicular or ovarian fate.

1.2.4 The regulation of *Sry* expression

Sry expression occurs in a highly regulated manner showing a dynamic expression pattern (Bullejos and Koopman, 2001; Wilhelm and Koopman, 2006). In humans, *SRY* expression commences in the gonad between 41 and 44 days post ovulation (dpo) and peaks at 44dpo. Unlike in mice however, low levels of *SRY* transcript are detectable in the gonads throughout embryonic development (Hanley et al., 2000). Other mammalian species also show differences in the timing of *Sry* downregulation (Harry et al., 1995; Parma et al., 1999), suggesting that the time-point at which *SRY* expression ceases is not vital to the proper development of the testis. The timely onset of *Sry* expression is critical, however, as evidenced by the phenomenon of B6.Y^{DOM} sex reversal. Mice with autosomes and an X chromosome derived from the C57BL/6J (B6) background that carry the Y chromosome from various *Mus domesticus* sub-species (such as *Mus poschiavinus*) show a range of phenotypes including delayed testis development, the formation of ovotestes or full XY male-to-female sex reversal. The HMG domain and other areas of the SRY protein have been shown to have retained high levels of homology between the different sub-species involved, indicating that they are likely to be functionally equivalent. Instead differing degrees of delay in *Sry* expression result in a delay or failure of sufficient *Sry* expression to induce the testis fate throughout the entire gonad (Eicher et al., 1996; Nagamine et al., 1999).

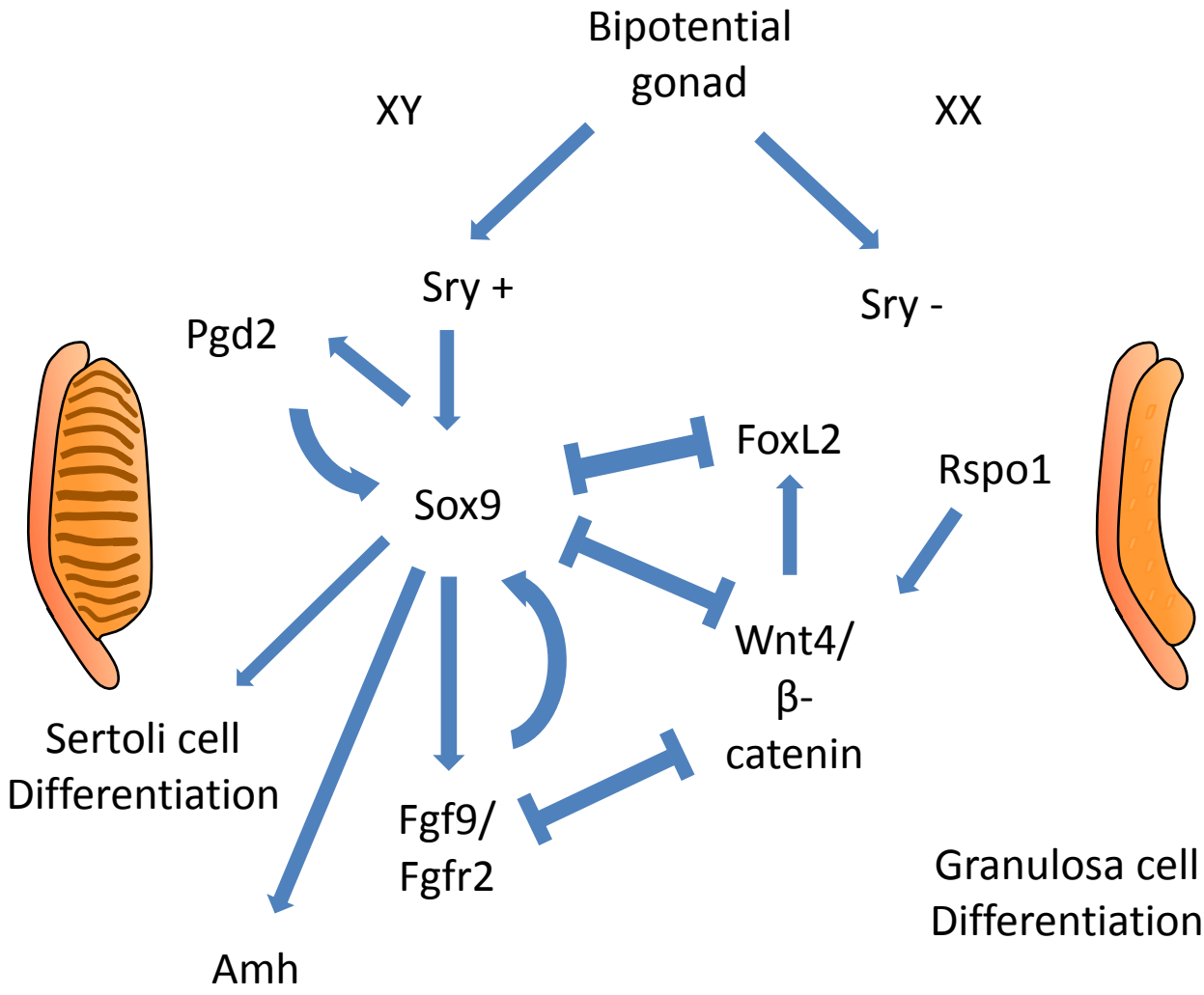


Figure 1.5 Molecular pathways of mammalian sex determination
The onset of *Sry* expression at around 10.5dpc in XY gonads leads to the upregulation of *Sox9*. SOX9 upregulates FGF9/FGFR2 signalling and the expression of *Pgd2*. Both of these pathways then maintain *Sox9* expression through a positive feedback loop. SOX9 also inhibits the actions of the pro-ovarian molecule FOXL2 and members of the WNT signalling pathway. FGF9 signalling also contributes to the repression of *Wnt4*. Continued SOX9 activity induces the differentiation of supporting cells to pre-Sertoli cells. These cells then begin to produce AMH , which induces morphological changes in the rest of the urogenital region. In the absence of *Sry* expression, pro-ovarian genes, such as *Rspo1*, *Wnt4* and *Foxl2* are not downregulated and lead to the differentiation of supporting cells towards the granulosa cell fate. Pro-testis genes, such as *Sox9* and *Fgf9* are downregulated to avoid disruption to the ovarian pathway.

To date a number of different factors have been linked to the timely upregulation of *Sry* (Warr and Greenfield, 2012). Often identified through mouse models exhibiting XY sex reversal, these proteins are involved in a range of different pathways, some of which converge on regions surrounding the *Sry* gene (outlined in Fig 1.6). However the precise mechanism by which such molecules regulate *Sry* expression has not been established.

The Y chromosome consists of a large number of repeated sequences and a small number of genes which inconsistent expression levels and location on the chromosome between mammalian species (Graves, 1994). However, comparisons of the genomic sequences upstream of *Sry* in humans, cows, pig, goat and mice reveal four highly conserved regions containing putative binding sites (Ross et al., 2008).

Amongst these is the target sequence for the GATA protein family: a member of which, *Gata4*, was previously discussed in the formation of the gonadal ridge, a common theme between many of the known regulators of *Sry*. Gata4 is the sole member of the GATA family found in the somatic cells of the gonads of both sexes at 11.5dpc. At 13.5dpc, its expression becomes sexually dimorphic. In XX gonads, *Gata4* is downregulated in all gonadal cells. In XY gonads, *Gata4* expression is downregulated in the interstitial cells but upregulated in the Sertoli cells, where it continues to be expressed in the postnatal testis (Tanaka and Nishinakamura, 2014). Inducible knockouts of *Gata4* at 10.5dpc, but not later times, result in XY sex reversal (Manuylov et al. 2011). ChIP experiments have also shown that GATA4 binds to the mouse *Sry* promoter in two locations (Gierl et al., 2012).

Post-translational modifications of transcription factors can affect their ability to bind DNA as well as their sub-cellular localisation. Gata4 has been shown to only bind to the *Sry* promoter when it is

phosphorylated (Gierl et al., 2012). This phosphorylation is thought to be carried out by p38 MAPK, consistant with this GATA4 activating role is the sex reversal observed in the double knockouts of *p38α* and *p38β* MAPK XY gonads (Warr et al., 2012). Components of the phosphorylation cascade upstream of *p38* MAPK also appear to be involved in sex determination. Of particular importance to this thesis are Mitogen-activated protein kinase kinase kinase 4 (*Map3K4*) and Growth arrest and DNA damage 45 gamma (*Gadd45g*) (Gierl et al., 2012). These components will be discussed in more detail in a later section.

Multiple co-factors are required for normal *in vivo* function of GATA factors, particularly members of the FOG (friend of GATA) family. The requirement for FOG2 for *Sry* expression has been demonstrated through knockout experiments and the ablation of the interacting domain on Gata4 ($Gata4^{ki}$). $Fog2^{-/-}$ XY gonads produced insufficient levels of expression of *Sry* for Sertoli cell differentiation and later failed to initiate the expression of genes involved in steroidogenic functions, such as *Dhh* and *Amh*, indicating that this co-factor has further downstream applications during gonad development (Tevosian et al., 2002). The upregulation of *Fog2* is in turn controlled by the transcription factors SIX1 and SIX4 (Fujimoto et al., 2013).

Wt1, another gene previously discussed in the formation of the gonadal ridge, is also required for the activation of *Sry.* Both the +KTS and –KTS isoforms of WT1 have been shown to bind to the TCC repeat elements found in *Sry* sequence and interact with the GATA4 protein. Both isoforms are capable of activating the *Sry* promoter, and to a greater extent than GATA4 shown using *in vitro* models. This interaction is mitigated by the binding of WT1 to a proximal binding site within the *Sry* promoter. Specific *Wt1(-KTS)* null XY mice still express *Sox9* and *Amh* at 12.5dpc. However, decreased gonad size is observed in both sexes, likely due to its role in establishing the gonadal ridge (Hammes

et al., 2001). In contrast, the +KTS isoform of Wt1 appears to be essential for *Sry* expression in a cell-autonomous manner. *Wt1(+KTS)* null mice show a reduced level of *Sry* mRNA, and decreased levels of cell proliferation at 11.5dpc. This isoform also has a high affinity for RNA and aids in the processing of mRNA containing viral or cellular constitutive transport elements (CTE) (Bor et al., 2006). As a result, it may also have a role in stabilising *Sry* mRNA (Bradford et al., 2009). This suggests that WT1 is required to act in concert with GATA4 to increase *Sry* expression to the threshold level required for Sertoli cell differentiation (Miyamoto et al., 2008).

Though its role is unknown, mutations in *SF1* have been shown to cause 46,XY complete gonadal dysgenesis (Achermann et al., 1999). SF1 was found to bind to the porcine *Sry* promoter at two potential binding sites. A proximal site lies 260bp 5' to the *Sry* transcription start site, while a distal site is found 1369bp upstream. Though orthologues of these regions have been found in the sheep *Sry* promoter, none were found in the sequence upstream of human or mouse *Sry*. An SF1 binding site was found 315bp 5' to the human *SRY* transcriptional start site, though this appears to repress *SRY* activity if SF1 has been phosphorylated (Pilon et al., 2003). It is difficult to verify the action of *Sf1* on *Sry* in mice as these specific sites do not have homologues in the mouse *Sry* promoter. Instead, many short sequences more closely resembling a short SF1 binding motif are found throughout the mouse *Sry* promoter. It is therefore assumed that SF1 binds to the *Sry* promoter, but in an unknown location (Larney et al., 2014).

As in the formation of the gonadal ridge, Cbx2 appears to have a role in the regulation of *Sry* expression. Human mutations in *CBX2* have been found in patients exhibiting 46,XY complete gonadal dysgenesis, an observation mirrored in mutant mice. It is still debatable, however, whether this

phenotype is the result of a lack of *Sry* expression or a failure to differentiate a sufficient number of gonadal somatic cells which express *Sry* at an earlier stage (Larney et al., 2014).

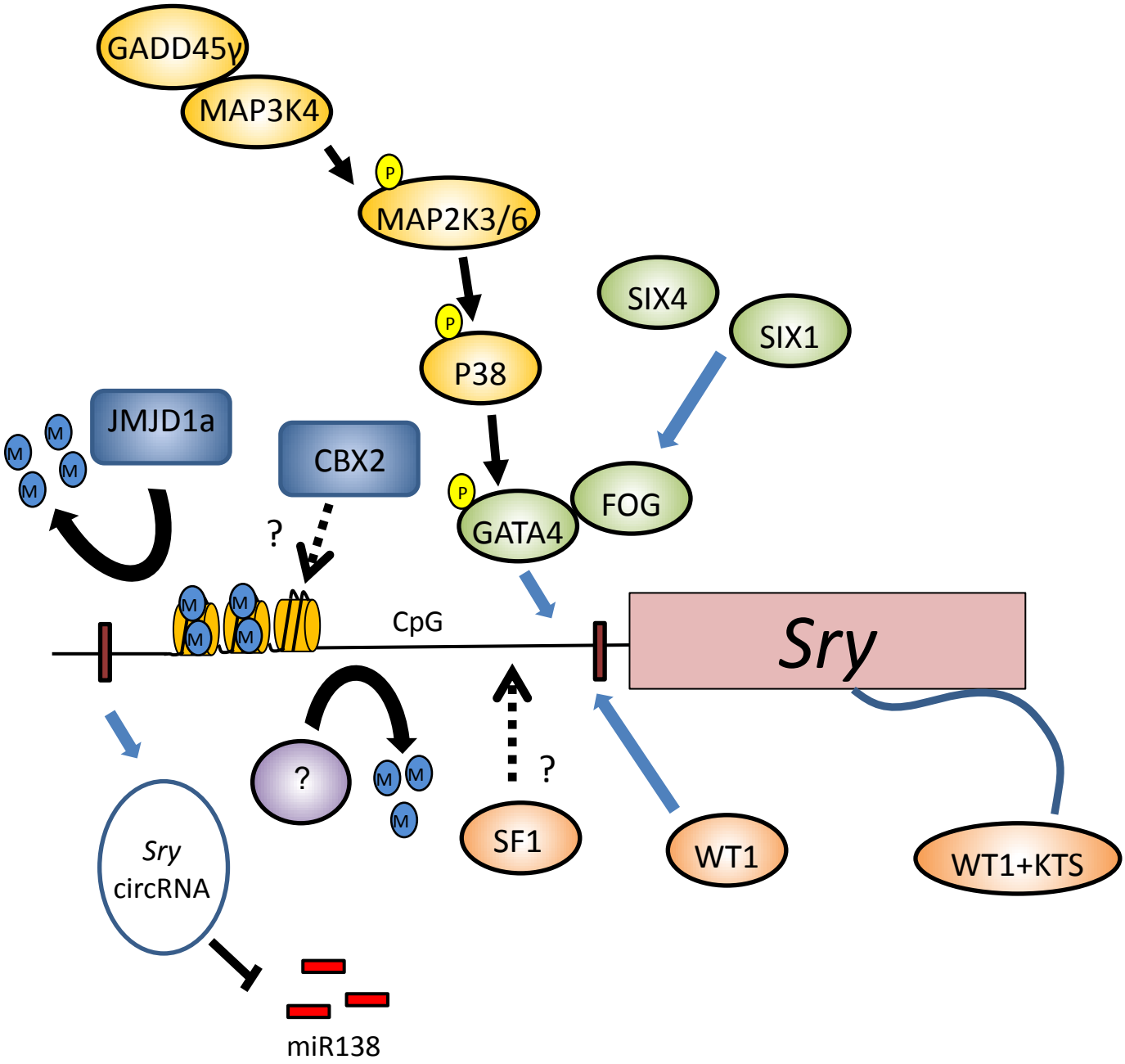


Figure 1.6 Known regulators of *Sry* expression and their actions
Diagram of the known pathways and molecules implicated in the timely expression of *Sry*. Blue arrows indicate transcriptional activation of the target gene, solid black arrows a protein interaction; dashed arrows indicate where a direct link has not yet been reported. Yellow circles represent phosphate groups, blue circles represent methyl groups.

Mouse *Sry* can be transcribed as a 1.1kb or a 5kb linear RNA molecule. The latter RNA includes regions from the large inverted repeat sequence which flank the gene. The open reading frame is flanked by a splice donor and a splice acceptor site. Together these sequences facilitate the formation of circular *Sry* RNA containing the ORF but a new stop codon. This alternative transcript is found in adult male mouse germ cells as well as embryonic gonads at 12.5dpc (Dolci et al., 1997). *In vitro* experiments have shown that the *Sry* circRNA has the potential to act as a sponge towards micro RNA-138 (miR-138). The mature product of miR-138 is a 22nt product used by DICER and the RNAi machinery to post-transcriptionally regulate target genes. Degradation of the protein coding *Sry* transcript is therefore downregulated by the expression of the *Sry* circRNA (Hansen et al., 2013).

1.2.5 The requirement of mitogen-activated protein kinase signalling and *Gadd45g* in mammalian sex determination

Cellular signalling using mitogen activated protein kinases (MAPKs) is a common method of converting extracellular signals to intracellular responses. MAPKs exist in a hierarchically organised phosphor-relay cascade. The uppermost components of this phosphorylation cascade phosphorylate specific conserved residues on their target MAPK proteins. Once activated, these proteins phosphorylate the conserved tyrosine and threonine residues on other MAP kinases, continuing the cascade.

Mammals contain four distinct MAPK families: extracellular signal-related kinase (ERK)-1/2, Jun amino-terminal kinases (JNK 1 – 3), p38 MAPK proteins (p38α/β/γ/δ) and ERK5. Each one is activated by a specific MAP2K: ERK-1/2 by MAP2K1/2; JNK by MAP2K4/7; p38s by MAP2K3/6 and MAP2K5 for ERK5. However each MAP2K can be activated by various MAP3K proteins (Chang and Karin, 2001). Numerous extracellular signals have been implicated in the activation of the various MAP3K proteins through *in vitro* studies. These include cellular stress, growth factors signals and cytokines. However

other studies have shown that a more complex relationship exists *in vivo* and that the relationships begin signalling cascades can vary between cell types (Chang and Karin, 2001).

The role of MAPK in mouse sex determination was first established by identification of the boygirl (*byg*) mutant in a forward genetic screen. This ENU-induced point mutation was mapped to nucleotide 1144 of the *Map3K4* open reading frame, causing an A→T transition, replacing an arginine with a premature stop codon. The resulting protein is truncated at amino acid 382, and lacks the S_TKc kinase domain required for proper function. This resulted in a delay in *Sry* expression, and embryonic sex reversal (Bogani et al., 2009). Subsequent studies found that this process begins a phosphorylation cascade through MAP2K3 and MAP2K6 to p38-MapK which phosphorylates GATA4, allowing it to bind to the *Sry* promoter. *Map2k3*$^{+/-}$ *Map2K6*$^{-/-}$ mice showed limited amount of ovarian tissue developing within the XY gonad. When crossed onto the sensitive B6.Y^{AKR} strain, this phenotype was elevated to full XY sex reversal (Warr et al., 2016). Lower levels of phospho-p38 MAPK and phospho-GATA4 were detected in gonadal tissue from *Map3k4*$^{-/-}$ mice (Warr et al., 2012).

Other members of the MAPK family may have a role in mammalian sex determination. Familial and non-familial cases of human 46, XY DSD (including 46, XY CGD) were found to carry mutations in the *MAP3K1* gene (Pearlman et al., 2010). *Map3K1* is expressed in murine gonads at 11.5dpc, the stage of sex determination. However, mice lacking the kinase domain of *Map3k1* do not show any signs of sex reversal but did exhibit increased gonadal length. *Map3k1* mutants also showed increased embryonic lethality on C57BL/6J. Though the XY individuals who survived to adulthood were fertile, XX mice showed an incompletely penetrant imperforated vagina phenotype. The differences between the human and the mouse phenotypes may be due to the human conditions not reflecting a loss of function *MAP3K1* mutation (Pearlman et al., 2010; Warr et al., 2011).

Numerous factors have been implicated in the activation of the MAP3K4 signalling cascade, among these is a member of the GADD45 family, GADD45γ. In T cells, GADD45γ activates the p38-MAPK and JNK pathways (Takekawa and Saito, 1998). GADD45γ interacts with the auto-inhibitory region on MAP3K4, which allows MAP3K4 to auto-phosphorylate and activate itself (Miyake et al., 2007). XY mice on the C57BL/6J background lacking *Gadd45g* were found to exhibit XY gonadal sex reversal due to a delay in *Sry* expression, similar to the *Map3k4* knockout. This was demonstrated to be associated with a deficiency in the levels of phosphorylated p38-MAPK and GATA4 (Gierl et al., 2012; Warr et al., 2012). The other members of the GADD45 family, GADD45α and β, do not appear to have a role in sex determination (Johnen et al., 2013).

Much of the work on mammalian sex determination has focused on transcriptional networks (Warr and Greenfield, 2012) and the transcriptomes of the various cell types involved (Jameson et al., 2012;). The next section will instead look at the role epigenetics has been shown to play in this process.

1.2.6 Epigenetic regulation in sex determination

Epigenetics is a term first used to explain the process of terminal differentiation of cells from common precursors. In the modern sense, it refers to the non-base modifications to DNA and post-translational modifications to the surrounding histones which affect the accessibility of the DNA to the transcriptional machinery and thus differences in gene expression between cell types.

Post-translational modifications to histone tails can condense or relieve the chromatin around genes, thus allowing access to the transcriptional machinery. The histone demethylase Jumonji domain-containing 1a (*Jmjd1a*) has been shown to be required for normal *Sry* expression; knockout mutants

show diminished *Sry* expression at the critical time point for testis determination. Somatic cells from 11.5dpc knockout XY gonads showed increased levels of di-methylated lysine 9 on histone 3 (H3K9me2) at the *Sry* promoter, an epigenetic mark associated with transcription repression, compared to wild-type counterparts. The knockout cells also showed lower levels of methylation on lysine 4 of histone 3 (H3K4me2/3), a mark associated with transcriptionally active areas. Combined with the low levels of *Sry* expression, these results suggest that *Jmjd1a* mediates the demethylation of H3K92me, opening the chromatin around the *Sry* locus to facilitate transcription factor access and further open chromatin marks such as H3K4me (Kuroki et al., 2013).

Cbx2 is another known regulator of *Sry* and has a role in upregulating cell proliferation in the gonadal ridge. Though the role of Cbx2 is unclear in *Sry* regulation, its contribution to the polycomb repressive complex 1 (PRC1), responsible for the reorganisation of chromatin, raises the possibility it has a potential role in the epigenetic regulation of *Sry*. CBX2 has been shown to recognise methylated histone residues through its chromodomain whilst ChIP experiments in ES cells on the PRC1 highlight many of the genes involved in sex determination as potential targets (Katoh-Fukui et al., 2012).

DNA methylation is another well studied epigenetic mechanism. The most well characterised target for DNA methylation are cytosine residues immediately 5' to a guanine residue, a CpG site. These CpG dinucleotides are found through-out the mammalian genome. Often the addition of a methyl group to the carbon at the 5 position on the cytosine base is associated with negative gene regulation. Differing cell types show differential methylation at the CpG sites of differentially expressed genes. Typically, CpG sites are hypomethylated in actively expressed genes and hypermethylated in the regulatory regions of inactive genes. CpG islands are regions of the genome containing a much higher frequency of CpG dinucleotides relative to the rest of the genome. Half of the genes in the human

genome contain a CpG island over their promoter region and first exon, for example the promoter region of human *Sox9* contains a CpG island (Aleman et al., 2008). Murine *Foxl2* also contains a CpG island covering its proximal promoter and 5' untranslated region. This region was found to be hypomethylated in tissue samples from adult ovaries compared to testis, liver and heart (Tran et al., 2013).

Two clusters of CpG sites reside within the *Sry* promoter. These sites become hypomethylated specifically in the somatic cells of the XY gonad at the time of *Sry* expression. The first region (region I) contains 4 CpG sites and lies 3kb upstream of the *Sry* transcriptional start site, potentially in the promoter for the circular *Sry* transcript. At the earliest stages of embryogenesis, this region is hypermethylated. However, this methylation is lost by 11.5dpc at the peak of *Sry* expression. The second region (region II) contains seven CpG sites and lies in a more proximal position to the linear *Sry* promoter. Similarly to region I this site becomes hypomethylated at 11.5dpc as *Sry* expression is initiated. Unlike region I however, region II becomes hypermethylated again by 13.5dpc coinciding with the decrease in *Sry* expression (Nishino et al., 2004).

Later reports describe methylation at a non-CpG site on the internal cytosine at the sequence 5'-CCTGG-3' in region II. The methylation status of this site varies across differing mouse strains. In the ICR-J, ICR-A, Bal/c, DBA-2 and C3H strains; methylation was observed with concurrent hypomethylation in the CpG sites of the region. Methylation was not seen in the B6 or 129 strains. Interestingly, methylation at this site appeared to boost the activity of the *Sry* promoter *in vitro*. The strain specific differences at this particular site could be part of the underlying differences in the regulation of *Sry* between these strains, and therefore their differing susceptibility to XY sex reversal (Ninshino et al., 2011).

The discovery of the sex reversal phenotype in *Gadd45g*$^{-/-}$ mice provided the intriguing possibility of an essential factor controlling DNA methylation at the *Sry* locus. Previously member of the *Gadd45* family have been implicated in active DNA demethylation (Kurtz and Ray-Coquard, 2012; Niehrs and Schafer, 2012). However, *Gadd45g*$^{-/-}$ mice did not show increased levels of methylation around the *Sry* locus at 11.5dpc (Gierl et al., 2012; Warr et al., 2012).

Currently there are no other known tissue-dependent, differentially methylated regions (DMRs) associated with genes active in somatic cells of the early embryonic gonad between 11.0-12.0 dpc. Finding such DMRs would provide insight into the regulation of many of the genes involved in sex determination and may also find new genes in the pathway. The search for such elements and their relationship to *Gadd45g* will be a major component of this thesis and will be discussed more in chapters 3 and 4.

1.3 Clinical relevance of Disorders of sex development (DSD)

Disorders of sex development (DSD) are defined as congenital conditions in which the development of chromosomal, gonadal or phenotypic sex is atypical (Hughes, 2008). Such disorders can arise through disruptions to the sex determination process, as described in section 1.2, or during the differentiation of the internal and external genitalia, a process controlled by the secretion of hormones from the gonads. Patients are at an increased risk of gonadal tumours (Looijenga et al., 2007), low fertility (Bashamboo and McElreavey, 2015) and potentially lifelong impacts on their psychology and social development (Hiort et al., 2014).

Rates of testicular cancer have shown an increased rate of incidence since the last quarter of the 20th century (Bray et al., 2006; Walsh et al., 2006). It has been proposed that this increase is due to testicular dysgenesis syndrome, which links various clinical observations such as cryptorchidism, hypospadias, sub-level fertility and gonadal cancers to the same underlying disruptions to development of the foetal testes (Wohlfahrt-Veje et al., 2009). These disruptions can be linked to environmental factors, such as exposure to endocrine disrupting chemicals (EDCs), as well as genetic factors (Giwercman and Giwercman, 2011). Such developmental disruptions may also be occurring in ovarian development (Louis et al., 2011). As a result, it is of great importance that our understanding of the genetic pathways controlling the process of sex determination is further developed. This will not only aid in our understanding of the various gonadal cancers but can also provide diagnoses for DSD patients.

DSDs can be categorised into three groups: sex chromosome DSDs, 46 XY DSDs and 46 XX DSDs. Sex chromosome DSDs arise as a result of aneuploidy of the sex chromosomes. These include Turner's syndrome (45 XO) and Klinefelter syndrome (47, XXY). Children with Turner's or Klinefelter syndrome do not present genital differences in early life, unlike patients with 45, XO/46, XY mosaicism, which will often exhibit atypical genitalia (Achermann et al., 2015). 46, XX DSDs, such as congenital adrenal hyperplasia, can result from excessive levels of foetal androgen exposure, resulting in ambiguous genitalia tending towards the "male" side of the spectrum (Kousta et al., 2010). 46, XY DSDs can present with a large range of phenotypes. They are typically the result of impaired production of testosterone, the lack of conversion of testosterone to dihydrotestosterone or abnormalities of the androgen receptor and present ambiguous external genitalia or a tendency towards the female end of the spectrum. Examples of the latter include failure of testicular descent (cryptorchidism) and abnormalities of the position of the urethral opening (hypospadias) (Mendonca et al., 2009). 46, XY

DSDs may also be the result of disruptions to the endocrine system, In both 46, XX and 46, XY DSDs the gonadal tissue can resolve during sex determination as a complete ovary or testis or an ovotestis, an organ containing a mix of tissue types from the two organs. Gonadal dysgenesis (GD) encompasses disorders in which patients exhibit external genitalia distinct to that normally associated with their chromosomal content. In 46,XY GD and complete gonadal dysgenesis (CGD) this typically involves to the formation of streak gonads, underdeveloped gonads comprised mostly of fibrous tissue (Bashamboo and McElreavey, 2015). The formation of a testis or ovotestis in 46,XX individuals is known as 46,XX testicular or ovotesticular DSD. Approximately 1 in 3 cases of this are caused by the translocation of *SRY* to the X chromosome (Verkauskas et al., 2007). Other cases have been reported to be caused by mutations in *RSPO1, DAX1* and *WNT4* (Domenice et al., 2004; Parma et al., 2006).

46, XY CGD and GD are characterised by complete or partial abnormal testis development, respectively. CGD is characterised by female internal and external genitalia, a normal or tall stature and the presence of bilateral dysgenic gonads. A variety of different genetic causes have been reported for 46, XY CDG; mutations or deletions of *SRY* affect 20% of patients, the majority of which are found within the HMG domain. Mutations leading to under-expression of *WT1*, *SF1*, *SOX9* and *DMRT1* as well as over-expression of various ovarian genes such as *DAX1* and *WNT4* have been reported (Mendonca et al., 2009). These cases have helped identify components of human sex determination and underline the dosage-sensitive nature of the gonadal differentiation pathway. However, a formal genetic diagnosis cannot be made in around 80% of 46, XY DSD cases (Hughes et al., 2006), underlining the importance of identifying novel genes involved in the sex determination process.

1.4 The mouse as a model of human DSD

The mammalian sex determination pathway is highly conserved across mammalian species, though less so across phyla. SRY and SOX9 function in both human and mouse testis development, as evidenced by phenotypes caused by their disruption; many of the other components of the sex differentiation pathway are also implicated in both species, including CBX2, FGFR2, DMRT1, WNT4, DAX1 and SF1(Wilhelm and Koopman, 2006).

The mouse has been an invaluable tool in confirming the identity of sex determining genes first implicated through the study of human DSD. Moreover is has allowed the experimental investigation of functional mechanisms. The ability to manipulate the mouse genome to engineer point mutations, global knockouts and conditional knockouts has allowed the identification of novel sex-determining genes and their characterisation (Hammes et al., 2001; Bogani et al., 2009; Warr et al., 2012; Hu et al., 2013).

Genetic background can also play a major role in the penetrance of DSD phenotypes. Parents can carry mutations which do not disrupt the sex determination pathway or affect their fertility. However when these mutations are inherited DSDs can develop in children. This suggests the existence of modifier genes in the parent and offspring. The importance of genetic background for sex determination is well established in mice. The effects of mutations on some inbred strains which lead to full sex reversal can be muted on other backgrounds (Johnen et al., 2013). For example, the C57BL/6J (B6) background is particularly sensitive to disruptions to testis development, which means that sex reversal caused by gene mutation is a more common outcome on this background when compared to others. Expression profiling has revealed that this is a result of higher levels of expression of ovarian genes during gonad development (Munger et al., 2009). B6 can be further

sensitised through the introduction of Y chromosomes from the various strains of *Mus domesticus* (Nagamine et al., 1999). As a result, such distinct genetic backgrounds can be used to dissect the underlying gene networks controlling testis development, even when genes are performing relatively minor roles.

Transgenic overexpression constructs have also been used to investigate genetic pathways in testis determination, especially in gain of function experiments (Warr et al., 2014). They can also be genetically modified to produce transgenic line containing reporter genes used for a variety of purposes (Beverdam and Koopman, 2006). Such constructs can allow the isolation of the various cell types in the developing gonads through fluorescence-activated cell sorting (FACS). By investigating these distinct cell populations, global expression profiles can be generated or epigenetic profiling performed, that specify the genetic or epigenetic landscape in each cell type. (Jameson et al., 2012; Kobayashi et al., 2013). Investigating the specific molecular nature of each cell can illuminate the cell-specific roles required for proper sex determination to take place, without 'noise' arising from irrelevant cell-types.

Mouse models have been used to show the requirement for the gene *Gadd45g* (Gierl et al., 2012; Warr et al., 2012). Though a mechanism through which GADD45γ activates a MAP kinase cascade targeting GATA4 for phosphorylation has been proposed, GADD45γ has been implemented in many cellular roles. These roles will be elaborated on in the following section.

1.5 *Gadd45* members and roles

The Growth arrest and DNA Damage (GADD) 45 protein family consists of three members; GADD45α, GADD45β (MYD118) and GADD45γ (CR6). This family of small, approximately 18kDa, acidic proteins

are evolutionarily conserved across species, with 93 – 95% identity between the amino acid sequences between the human and mouse orthologues, and share 50 – 60% homology between members. The crystal structure of GADD45γ reveals a 3 layered structure comprising of 2 α-helices (α2 and α3) followed by 5 β-sheets at its core, and a further 3 α-helices on the opposite side of the sheets (α1, α4 and α5). The β-sheets show the most homology between GADD45 proteins with the outer helices α1, α4 and α5 appearing more variable (Schrag et al., 2008).

First identified through *in vitro* screens for genes induced by stress or differentiation cues in Chinese hamster ovary (CHO) cells, *Gadd45a* was found to be rapidly induced by UV radiation (Zhan et al., 1994). *Gadd45b* was first reported from a screen investigating the effects of differentiation agents on M1 myeloblastic leukaemia cells. Later it was observed to share high levels of homology and similar charge characteristics with *GADD45a* (Selvakumaran et al., 1994). *GADD45g* was first found as result of Interleukin 2 stimulation on T lymphocytes, resulting in clonal expansion (Beadling et al., 1993). First designated as Cytokine response factor 7 (*CR7*), it was later renamed to *GADD45g* after noting its similar, but not identical, roles to *GADD45a* and *GADD45b (Zhang et al., 1999)*.

1.5.1 The anti-tumour role of Gadd45 members

Since the discovery of the *Gadd45* genes, the GADD45 proteins have been shown to interact with various components of different cellular processes. The promoter region of *GADD45a* contains a binding site for p53, the best studied tumour suppressor gene also known as “the guardian of the genome”. A variety of different cytotoxic stresses can induce the activation of p53, which sets up various checkpoints in the cell cycle, which can lead to cell cycle arrest or apoptosis (Kastan et al., 1992). GADD45 proteins lack any obvious enzymatic or DNA binding activity but instead act as recruitment proteins for various activators. Examples include p21, a p53 target implicated in cell cycle arrest; Proliferating cell nuclear antigen (PCNA), a nuclear protein which plays a central role in DNA

repair; and the Cdc2/CyclinB1 complex, a protein which mediates the transition from G2 to M phase, inhibited by the presence of GADD45 members (Liebermann and Hoffman, 2008). In many cases these interactions require the dimerization of GADD45 proteins through their α-helices α2 and α3, which contain the most highly conserved residues between the members (Schrag et al., 2008). Dimerization of GADD45γ was shown to be essential for the inhibition of proliferation of HepG2 cells (Schrag et al., 2008).

Due to their role in stress signalling, it was hypothesised that the GADD45 proteins could also have a role in the prevention of cancers. Since then GADD45 proteins have been shown to exhibit anti-tumour activity both *in vitro* (Ju et al., 2009) and *in vivo* (Hollander et al., 1999). Lower levels of *GADD45G* have been found in a number of tumours including pituitary adenomas, hepatocellular carcinomas, lymphomas, prostate carcinomas and nasopharyngeal carcinomas (Yang et al., 2009). Rather than a result of direct mutation or inactivation to the *GADD45G* gene, this appears to be the result of hypermethylation at the proximal *GADD45G* promoter (Guo et al., 2013). This region spans 355bp and contains 33 CpG sites, showing increased levels of methylation in various carcinoma and lymphoma cell lines (Ying et al., 2005). GADD45γ has also been implicated in the control of apoptosis, this process can be negatively regulated by the micro RNA miR-383 binding to its 3' UTR (Zhao et al., 2014).

1.5.2 *Gadd45g* expression profile

In adult mice, expression of *Gadd45g* is found predominantly in the skeletal muscles, kidneys and liver (Zhang et al., 1999). A comprehensive series of experiments has revealed the early expression pattern of the three *Gadd45* genes during mouse development. Quantitative PCR (qPCR) experiments on whole mouse embryos from 6.5dpc to 10.5dpc reveal differential and dynamic profiles for each of

the *Gadd45* genes. *Gadd45g* expression was found to increase continuously from 6.5dpc till 10.5dpc. Comparisons between the three *Gadd45* genes at 10.5dpc found that *Gadd45g* was expressed at a higher rate than the other two genes at this stage (Hollander et al., 1999).

At 6.5dpc, *Gadd45g* expression was found in the ectoplacental cone and some cells of the extraembryonic-ectoderm. At 8.5dpc, expression was observed in the presomatic mesoderm and within some of the neural precursor cells found in the neural fold. *Gadd45g* expression was then strongly seen in the central nervous system in 9.5dpc and 10.5dpc embryos. This included regions such as: the neural tube, the dorsal midbrain and the cranial and dorsal root ganglia (Kaufmann et al., 2011).

1.5.3 GADD45 proteins in development

Though each of the members of the *Gadd45* family show differential expression patterns during development, expression is observed in many cells while they are undergoing differentiation, suggesting that they may have a role in various aspects of development (Kaufmann et al., 2011).

Null mutants for *Gadd45a* do not show any major developmental problems and grow to adulthood without any obvious differences compared to wild-type littermates. When *Gadd45a* nulls were mated however, 50% of females failed to deliver their first litter and subsequently died as a result (Hollander et al., 1999). Double morpholino mediated knockdowns of *Gadd45a* and *Gadd45g* in NF stage 41 *Xenopus* embryos exhibited gastrulation defects, head defects and reduced pigmentation. At the earlier NF stage 18, double knockdowns showed a lack of various neuronal differentiation markers and an upregulation of pluripotency markers, suggesting *Gadd45a* and *Gadd45g* act redundantly to regulate the exit from pluripotency for neuronal precursor cells. Milder phenotypes were seen in

Gadd45b knockdowns, coinciding with the detectable but lower levels of expression of *Gadd45b* compared to *Gadd435a* and *Gadd45g* at this time (Kaufmann and Niehrs, 2011).

Gadd45 genes are thought to play a role in the differentiation of haematopoietic stem cells, which form the various cell lineages involved in the innate and adaptive immune systems. *In vitro* studies show that bone marrow cells lacking either *Gadd45a* or *Gadd45b* show reduced numbers of macrophages or granulocytes upon the induction of differentiation (Gupta et al., 2006). In contrast *Gadd45g* did not appear to have a role in the differentiation of myeloid cells (Hoffmeyer et al., 2001). However type 1 T-helper (Th1) cells from *Gadd45g* deficient mice showed lower levels of IFN-γ production due to impaired p38-MAPK activity and were less prone to apoptosis (Lu et al., 2001). GADD45γ was required to induce the differentiation of long-term repopulating haematopoietic stem cells via MAP3K4 and p38-MAPK (Thalheimer et al., 2014).

1.5.4 GADD45 proteins as mediators of DNA repair

DNA repair is a critical survival mechanism for cells. Damage which arises from alkylation, oxidative damage or deamination of DNA bases can lead to non-canonical base pairings. If not corrected these mismatches can lead to the accumulations of point mutations throughout the genome. Such instances of damage are usually repaired by the base excision repair pathway (BER). In mammals this process is mediated by four core components: a DNA glycosylase, an AP endonuclease or AP DNA lyase, a DNA polymerase and a DNA ligase. The damaged base is recognised and cleaved by a base specific DNA glycosidase at the N-glycosidic bond leaving an apyramid or apurinic site (AP site). The DNA backbone is then cleaved at the AP site by a DNA AP endonuclease or DNA AP lyase. A DNA polymerase fills in the gap generated by the nick with the correct nucleotide. The repair is then

completed by a DNA ligase which catalyses the phosphodiester bonds along the backbone (Robertson et al., 2009).

Mouse embryonic fibroblasts lacking *Gadd45a* have been shown to have an impaired ability for BER, due to the inability to remove AP sites from the genome. A significantly lower level of nuclear AP Endonuclease 1 (APE1) was observed in *Gadd45a*$^{-/-}$ cells, suggesting a role for GADD45 proteins in the recruitment of components of the BER pathway (Jung et al., 2007).

UV radiation can lead to a variety of helix-distorting modifications to DNA including the formation of dimers between two adjacent thymidine residues or adjacent pyramidine residues. In order to repair these bulky helix-distorting lesions, cells employ components of the nucleotide excision repair (NER) pathway. A small single-stranded segment of the damaged DNA, typically 12 -13 nucleotides, is removed by dual incisions around the lesion. The gap is then filled using the intact strand as a template. The NER process in mammals involves up to 30 proteins to recognise and repair the damaged site. Mutations in the various genes involved have been linked to various human conditions including xeroderma pigmentosum (XP), Cockayne syndrome and trichothiodystrophy (Boer and Hoeijmakers, 2000). UV radiation has been shown to upregulate the expression of *Gadd45* genes (Hildesheim et al., 2002; Ying et al., 2005) and the failure of cells lacking *Gadd45a* to repair DNA lesions suggest a role in mediating the NER response. In mice, *Gadd45a* and the NER gene *Xpc* have been shown to genetically interact in the formation of lung tumours (Hollander et al., 2005). All three members of the GADD45 family have been shown to interact with PCNA (Azam et al., 2001), itself a critical component of the NER pathway (Essers et al., 2005). These results suggest that the *Gadd45* gene play a crucial role in coordinating the NER response as part of their role in responding the DNA damage.

1.6 DNA methylation and epigenetic regulation

This thesis will investigate the potential role that *Gadd45g* may have in active DNA demethylation (Niehrs and Schafer, 2012) and whether this may be the cause of the observed sex reversal seen in XY *Gadd45*$^{-/-}$ mice. In order to investigate this role it is important to survey the current understanding of DNA methylation, how it is established and removed, and its significance to the fields of epigenetics and development.

1.6.1 Regulation of gene expression by epigenetic factors

Epigenetic regulation has been demonstrated to play an important role in the regulation of *Sry* expression (Nishino et al., 2004; Kuroki et al., 2013); however little else is known about the epigenetic landscape of the genome in cellular lineages of the gonads at the time of sex determination. Since its inception, the field of epigenetics has been linked to developmental biology. Phenomena such as terminal cell differentiation, genomic imprinting and X chromosome inactivation have been demonstrated to depend on epigenetic modifications. The effects of epigenetic modifications are often reversible and can be influenced by environmental signals, allowing cells to retain a level of plasticity after differentiation (Holliday, 2014).

As well as DNA methylation, which will be expanded on in a later section, modifications to the tails of core histone molecules affect the expression of nearby genes. The flexible tails of histones are highly basic and can be modified by a large range of post-translation modifications. These include acetylation, methylation, phosphorylation and ubiquitination. Each distinct modification is associated with specific transcriptional states of surrounding genes. The histone code hypothesis proposes that specific patterns of modifications act as a bar code to recruit chromatin modifying complexes to a site, bringing about the associated chromatin state (Cosgrove and Wolberger, 2005).

Many methods of RNA-mediated gene expression are regarded as epigenetic events. Distinct cells may both transcribe the same gene, but alternative splicing of the resulting mRNA can leads to the production of distinct proteins. The chromatin state around specific exons can affect how a transcript is spliced by modifying the speed at which RNA Pol II transcribes. By slowing the rate of transcription at specific exons due to a more densely packed chromatin structure, more time is allowed for the alternative splicing machinery to be recruited to the emerging transcript (Robson-Dixon and Garcia-Blanco, 2004). Specific marks on histone tails may also aid in the recruitment of the alternative splicing machinery. Long non-coding RNAs can also regulate splicing by acting as a sponge for splicing competent serine-arginine rich proteins. By sequestering these proteins to the speckles in the nucleus, exons are included in a number of genes (Luco and Misteli, 2011).

Micro RNAs (miRNAs) and small interfering RNAs (siRNAs) are another method of post-transcriptional regulation. Typically 20 – 25nt long, these RNAs control gene expression through the recruitment of the RNA-induced silencing complex (RISC). Complementary mRNA sequences are then targeted for degradation, usually at the 5' and 3' untranslated regions (UTRs) (Liu et al., 2004). Intronic regions downstream of alternatively spliced exons may also be a target of siRNAs. This method of regulation instead induces the formation of localised heterochromatin around the exon affecting its inclusion in the transcript (Luco and Misteli, 2011). Epigenetics is now seen as the biochemical manifestation of the differences between various cells types, with each cell seemingly exhibiting its own unique epigenome.

1.6.2 Active DNA methylation

The majority of CpG dinucleotides in animal genomes are methylated bimodally, targeting the cytosine residue on both DNA strands. Those located within CpG islands surrounding active regions of DNA are typically the exception, and are unmethylated. DNA methylation was first proposed as a molecular method for gene expression control in 1975 (Riggs, 1975). The enzymatic addition of a

methyl group to the carbon at the 5 position on cytosine bases in the CpG context was suggested to affect the expression of surrounding genes as well as provide a method for daughter cells to inherit gene expression patterns. This enzyme would either be sequence specific or recruited to target sequences by various protein complexes (Holliday, 2014). DNA methyltransferase (DNMT) 3a and 3b were found to be essential for de novo methylation of cytosine residues and are required for normal mammalian development. Mice lacking *Dnmt3a* appeared normal at birth but failed to thrive and died at 4 weeks of age. *Dnmt3b*$^{-/-}$ mice died in utero and showed slow growth and neural tube defects. Double knockouts lacked somites and failed to show embryonic turning. Embryos from *Dnmt3a*$^{-/-}$*Dnmt3b*$^{-/-}$ mice were smaller at 8.5 and 9.5dpc, and had died by 11.5dpc (Okano et al., 1999). Endogenous C-type retroviral DNA sequences found throughout the genome showed lower levels of methylation in *Dnmt3a*$^{-/-}$, *Dnmt3b*$^{-/-}$ and double knockout mice compared to wild-type blastocysts, indicating that the Dnmt3s are required for de novo methylation, which takes place after implantation (Okano et al., 1999). Dnmt3a also targets cytosines in the non-CpG context for methylation (Ramsahoye et al., 2000).

An elusive question in the field of epigenetics is whether DNA methylation patterns or modifications to histone tails are the first modifications to be established. Crosstalk between the two phenomena has been noted at various time points. As stated earlier, the repressive histone mark unmethylated H3K4 is required for DNMT3L to recruit DNMT3A for *de novo* methylation (Jia et al., 2007). Trimethylation of H3K9 (H3K9me3) by Suv39h1 HMTase recruits DMNT3B to pericentromic repeats, but not at other repeat sequences. The Lymphoid specific helicase (LSH) is a member of the SNF2-helicase family of chromatin remodelling proteins. It is required for the methylation of specific imprinting regions in embryonic development as well as recruitment of DNMT1, DSNM3B and histone deacetylases to a target region. LSH also appears to be involved in the recruitment of the polycomb

repressive complexes, factors closely associated with methylated regions of DNA (Gopalakrishnan et al., 2008).

However, DNA methylation has also been shown to have an effect on the histone modifying machinery. As DNA is replicated, it is likely that the surrounding chromatin is disrupted. The DNA methylation pattern can then be used to re-establish the chromatin state. Proteins which bind-specifically to methylated DNA such as MECP2 and the Methyl-binding domain (MBD) family have the potential to recruit histone deacetylases to regions of highly methylated DNA. The pro-expression histone mark H3K4me, also appears to be inhibited in the presence of methylated DNA (Cedar and Bergman, 2009).

1.6.3 Active DNA demethylation

As cells differentiate, genes which were once repressed become expressed. Sites such as these show a loss of methyl groups from nearby CpG sites. Demethylation can take place in an active or passive fashion. Passive demethylation refers to the progressive loss of methylation due to the absence of the methylation maintenance machinery. The loss of DNMT1 or UHRFf1 leads to a 50% reduction in methylation every cell division. However, large-scale demethylation is observed in the paternal pro-nucleus of the zygote and non-dividing migratory PGCs, indicating that another active pathway is present in mammalian cells (Schomacher, 2013). In human cells, the promoter of the estrogen-receptor target *pS2* undergoes cyclical methylation and demethylation within a 100 minute cycle when treated with estrogen. DNA methylation and demethylation can therefore be seen to be a reversible and dynamic process (Kangaspeska et al., 2008).

While the enzymes responsible for DNA methylation are relatively well characterised, those controlling active DNA demethylation are numerous and not as well understood. As the carbon-

carbon bond between the methyl group and the rest of the cytosine is difficult to break thermodynamically, it is unlikely that the demethylation reaction directly removes the methyl group from 5mC. Instead a series of intermediate reactions are used to remove the methyl group from the cytosine (Fig 1.7).

The Ten-eleven translocation (Tet) family of enzymes have the ability to oxidise the methyl group on 5mC in a progressive fashion. Human TET1 has been shown to oxidise 5mC to 5-hydoxymethyl cytosine (5hmC). Subsequently it was shown that all of the Tet members, Tet1-3, have this ability. At this stage many proteins which act on 5mC can no longer recognise 5hmC, potentially blocking DNA silencing constructs (Wu and Zhang, 2011). Iterative oxidation of 5hmC can generate 5-formyl cytosine (5fC) and 5-carboxylcytosine (5caC). These bases may contribute to passive DNA demethylation, through a lack of recognition by the methylation maintenance machinery, but can also contribute to active demethylation as well. By generating a 5fC or 5caC base, the base can be targeted by thymine-DNA glycosylase (TDG) generating an AP site which can be replaced with a non-methylated cytosine through BER (Kohli and Zhang, 2013).

Repair based demethylation can also be achieved in a Tet-independent manner. 5mC sites can spontaneously deaminate, generating a thymine, at a rate of 5.8×10^{-13} per second (Shen et al., 1994). As a result a G-T pair mismatch is generated, leading to the recruitment of the BER machinery. The AID/APOBEC family of enzymes can more directly catalyse this reaction *in vivo* in higher vertebrates. Activation induced deaminase (AID) is essential for proper immune function. By targeting cytosine residues in the variable segment of the rearranged antibody gene, AID can diversify the pool of available antibodies. APOBEC3 confers protection against viruses, by inhibiting reverse transcription. Over expression of AID APOBEC2a or APOBEC2b in zebrafish increased rates of

demethylation (Rai et al., 2008). During induced pluripotency in human cells, siRNA induced knockdown of AID resulted in remethylation around the promoters of *OCT4* and *NANOG*. PGCs

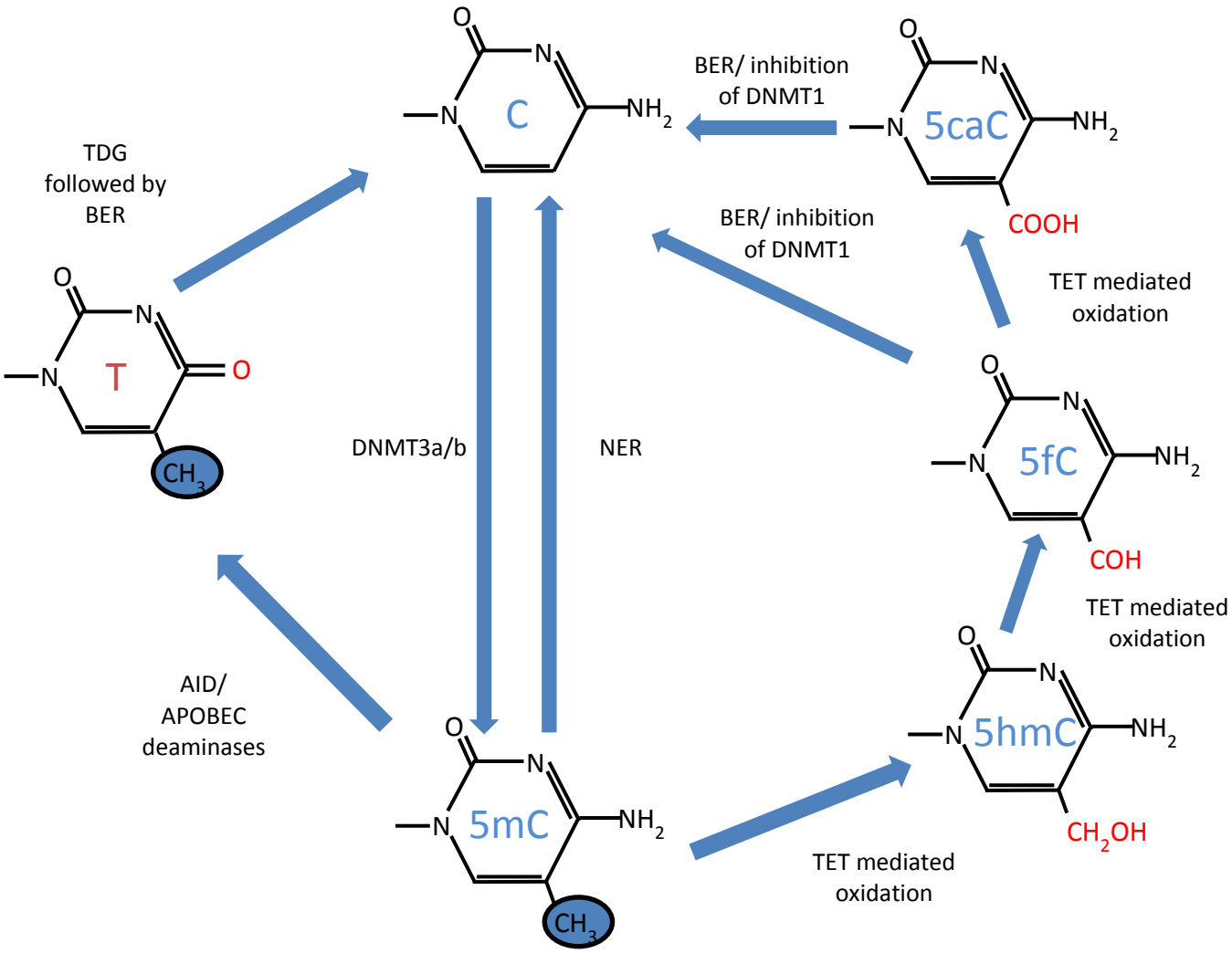


Figure 1.7 The reactions governing methylation and demethylation of cytosine
Model of the various reaction and intermediates formed during methylation and active demethylation of cytosine

lacking AID also showed a decreased but not incomplete level of global demethylation (Schomacher, 2013). However, *in vitro* experiments in human cell lines showed that only APOBEC3A, but not AID or APOBEC3G showed efficient deamination of 5mC (Wijesinghe and Bhagwat, 2012).

Components of the NER pathway have also been implicated in active DNA demethylation. Demethylation of the *oct4* promoter in *Xenopus* embryos was shown to require the 3'-nuclease of NER, *xpg*. In humans, XPG is required for the demethylation of the *RARβ2* promoter (Schomacher, 2013). In contrast to BER acting on specific CpG sites, NER appears to act on larger patches of DNA. NER may therefore act as a more dynamic method of active DNA demethylation.

1.6.4 The role of GADD45 proteins in active DNA demethylation

Since active DNA demethylation is regulated at specific loci, it is imperative to find the non-enzymatic proteins coordinating this process in order to understand its control. With respect to the role of the GADD45 family the majority of research so far has been carried out on GADD45α, with a growing number of studies investigating GADD45β. However the high level of homology between the members of the family means that GADD45γ may also carry out this role. GADD45α was first identified as a potential mediator of active DNA demethylation through its reactivation of methylation-silenced reporter plasmids (Barreto et al., 2007). Similar observations have been obtained for GADD45β and GADD45γ (Niehrs and Schafer, 2012). Global levels of methylation appear normal in *Gadd45a*$^{-/-}$ mice and they do not show increased levels of methylation at 11 normally hypomethylated sites (Engel et al., 2009). Other studies have not been able to replicate the original experiment that suggested a role for GADD45α in active demethylation using bisulphite sequencing (Jin et al., 2008). This does not preclude the possibility that GADD45 proteins target specific genes and their nearby CpG sites (Engel et al., 2014). It is also a possibility that in certain cell lines the GADD45 proteins act redundantly with each other. In zebrafish, the combined knockdown of four of

the six GADD45 isoforms greatly reduced DNA demethylation compared to individual knockouts (Rai et al., 2008).

Many examples of GADD45-induced demethylation have been found throughout different model organisms, often induced by differentiation or cellular stress responses. Both *Gadd45a* and *Gadd45b* are induced in epidermal progenitor cells by calcium signalling. This then leads to the demethylation of the *S100P* promoter (Sen et al., 2010). Adult neurogenesis in mice requires GADD45β and appears to be responsible for the demethylation of key factors including *Bdnf IX* and *Fgf1* (Ma et al., 2009; Guo et al., 2011).

As GADD45 proteins lack any obvious enzymatic activity, their role in mediating BER and NER has been suggested to be the mechanism through which they regulate DNA demethylation. The key targets for BER mediated demethylation appear to be a deaminase and a glycosylase to generate and remove a GT mismatch. In zebrafish, GADD45 proteins have been shown to interact with the deaminases AID and Apobec2a, as well as the deglycosylase Methyl binding protein (Mbd) 4 (Rai et al., 2008). In human cells, the deglycosylase TDG was found to interact with GADD45α (Schomacher, 2013). TDG has also been shown to mediate the removal of oxidative derivatives of 5mC. As GADD45α has been shown to physically and functionally interact with TET1, this suggests that GADD45 proteins can mediate both TET-dependent and independent demethylation (Kienhofer et al., 2015). The overlap in expression of Tet1 and GADD45β in the brain suggests this may not be an exclusive property of GADD45α (Ma et al., 2009; Guo et al., 2011)

Components of the NER pathway are also essential for GADD45α mediated DNA demethylation. During *Oct4* demethylation in *Xenopus* oocytes, BrdU is incorporated into the genome, suggesting

stretches of DNA had been synthesised and repaired (Barreto et al., 2007). A similar result was found in arrested mammalian cells at the ribosomal DNA locus (Schmitz et al., 2009).

GADD45 mediated demethylation does not reduce global levels of methylation and is often found to be targeted to certain regions within genes (Schafer et al., 2013). One method of targeting the demethylation machinery to specific regions of the genome may lie in the interaction between GADD45α and inhibitor of growth1 (human ING1b and mouse Ing1). ING1b and GADD45α interact in HEK293T cells, and are required for demethylation of silenced reporters. Similarly to GADD45 proteins, ING1b has been implicated in stress responses and cell cycle arrest. Within the protein are interaction domains for PCNA and a PHD finger domain which can bindH3K4 tri-methylation, a mark associated with transcriptional activation. ING1b may therefore guide GADD45α and the rest of the DNA demethylation machinery to the DNA located near H3K4me3 containing histones. Similar results were found in MEFs, revealing the requirement for the GADD45α-Ing1 interaction for the targeted demethylation of the *Mageb2* promoter upon UV irradiation (Schafer et al., 2013). Analysis of cells lacking *Gadd45a*, *Ing1* or both revealed an extra 100 potential target genes activated by this complex. Surprisingly, an equal number of gene were found to be upregulated in double knockouts, but this may be due to the other processes mediated by GADD45α or Ing1 (Schafer et al., 2013).

Intriguingly components of the NER pathway are often found at the promoters of genes exhibiting low levels of transcription. There, they promote the reorganisation of the surrounding chromatin, which shows decreased levels of DNA methylation and increased levels of positive histone marks (Le May et al., 2010). TAF12, a member of the TAT-box binding protein associated factor family, is found in Pol I and Pol II initiation complexes. It appears to be a factor in recruiting GADD45α to ribosomal DNA loci, in order to induce demethylation. GADD45α was found at the promoter regions of active

rDNA genes, which were found to be hypermethylation in the absence of *Gadd45α*. Depletion of TAF12 caused hypermethylation and prevented GADD45α and the Pol I complex formation at the rDNA loci. The specificity of GADD45α recruitment to specific promoter regions may be explained by its interaction with TAF12 (Schmitz et al., 2009). While GADD45 proteins have been shown to play a role in active demethylation in many contexts, how this process is regulated to target specific genes remains mostly unclear.

1.7 Main aims: summary

Gadd45g has been identified as a key gene in testis determination, acting to ensure the timely expression of *Sry*. The mechanism through which this process is mediated is still unknown, although the activation of MAPK signalling seems to be one important element (Gierl et al., 2012; Warr et al., 2012). As transgenic overexpression constructs have proven to be an invaluable tool in ascertaining the role of other genes in sex determination (Hiramatsu et al., 2009; Warr et al., 2014), a similar construct containing the sequence for *Gadd45g* will allow us to perform a variety of genetic tests. Such a tool could also prove valuable to other fields in which *Gadd45g* has been shown to have a role (Yang et al., 2009; Thalheimer et al., 2014). I aim to therefore produce a functional *Gadd45g* overexpression transgenic mouse line, which produces a significantly increased level of expression while still restricted to the same tissues as endogenous *Gadd45g* expression. I also aim to investigate whether over-expression of *Gadd45g* can rescue models of sex reversal, highlighting its importance in testis determination further.

Nothing is known about the loci that GADD45γ is recruited to during testis development and whether these correspond to gene promoters which will become transcriptionally active, or to the other proteins required to interact with GADD45γ. Finding such interacting partners and domains may not only provide answers to the role of *Gadd45g* in testis determination, but may also potentially lead to the identification of new candidate genes in the testis determining pathway. Such proteomic experiments require a well characterised and accurate antibody. No such antibody has been produced and instead the interacting partners of GADD45γ are known through the use of tagged constructs in *in vitro* experiments (Hansen et al., 2013; Schafer et al., 2013). If an *in vivo* construct can be successfully produced, then it may be amendable to genetic modification. This would allow the construction of an epitope-tagged version of the GADD45γ protein. I therefore aim to generate such a construct using BAC recombineering to add a combined His/FLAG tag sequence to *Gadd45g* and generate a transgenic mouse line containing this modified BAC.

Though members of the GADD45 protein family have been implicated in active DNA demethylation (Niehrs and Schafer, 2012), it seems unlikely that *Gadd45g* is aiding the active demethylation of the proximal *Sry* promoter (Gierl et al., 2012; Warr et al., 2012). This does not preclude the possibility of other loci showing differential methylation due to the absence of *Gadd45g*. Rather than test individual loci, I instead aim to take a global search for differentially methylated regions between genomes of XY wild-types, XX wild-types and XY *Gadd45g*$^{-/-}$ in the developing gonad. *Gadd45g* appears to be expressed only within the somatic cells of the gonad (Warr et al., 2012). As a result, it will be necessary to enrich for the somatic cells of the gonad before examining the state of DNA methylation. To achieve this I aim to sort cells from gonads of wild-type and *Gadd45g*$^{-/-}$ mouse embryos at 11.5dpc (18 ts), the highest point of *Sry* expression (Bullejos and Koopman, 2001), using the fluorescent transgene *Sf1-EGFP* and fluorescence activated cell sorting (FACS). With these cells I

then aim to produce bisulphite-converted sequencing libraries for each genotype. This will provide single-base resolution sequences, allowing for analysis of the methylome of gonadal somatic cells from wild-type XY and XX animals and the *Gadd45g* deficient XY animals to observe if any differentially methylated regions are present at this critical time point in sex determination.

Chapter 2. Materials and Methods

This chapter describes the materials and methods utilised in this thesis. Variations employed for specific experiments are described in their relevant chapters

2.1 General methods and solutions

Solutions were prepared according to Sambrook et al 1989 (Sambrook and Russell, 1989). Thermocycling reactions were performed using a G-Storm GS4 Thermocycler or a BioRad Tetrad2. All chemicals were obtained from Sigma-Aldrich unless stated.

Table 2.1 List of standard reagent solutions

Reagent	Composition
Luria Bertani (LB) broth	10 g/l Bactotryptone, 5 g/l Bactoyeast extract, 5 g/l NaCl, pH 7 at 25 ° C
PBAT Cell Lysis buffer	10mM Tris, 0.5% SDS
Streptavidin Binding and Washing (B&W) buffer (2x)	10mM Tris-HCl (pH 7.5), 1mM EDTA, 2M NaCl
PBT	PBS +0.1% Tween 20
NTMT	100mM NaCl, 100mM Tris-HCl (pH 9.5), 50mM $MgCl_2$, 0.1% Tween 20, in water
PBST	PBS + 0.1% Tween 20
TBE (10x)	1 M Tris base, 1 M boric acid, 0.02 M EDTA disodium salt in water. Diluted to 1 x for use in electrophoresis
Tris buffered saline (TBS)	50mM Tris-Cl pH 7.5, 150mM NaCl
TBST	50mM Tris-Cl pH 7.5, 150mM NaCl, 0.1% Tween 20
SSC (20x)	3m NaCl, 300mM sodium citrate, in water
WMISH pre-hybridisation solution	50% formamide, 5 x SSC, 0.1% Tween 20, 50µg/ml heparin in water
WMISH hybridisation solution	Prehybridisation solution + 100µg/ ml yeast RNA, 100µg/ml salmon sperm DNA
WMISH wash solution (20ml)	10ml formamide, 4ml 20 x SSC pH 4.5, 2ml 10% SDS, 4 ml water

2.2 Mouse lines and genotyping

Mouse work was conducted under the authority of the Home Office, Licence number PPL 30/2877. All animals were maintained in accordance with the Animals (Scientific Procedures) Act 1986 (ASPA) amended 2012 and in accordance with the ARRIVE guidelines. Animals were housed in IVC 1284L

cages at a controlled temperature and humidity in a controlled light-dark cycle, with darkness beginning at 7pm and ending at 7am.

All mice used in experiments were maintained on the C57BL/6J (B6) background, due to their increased sensitivity to disruption to testis determination. Primer sequences are available in Table 2.2. Unless stated otherwise reactions were prepared by mixing 10ng of extracted genomic DNA with 10µl Maxima Hot Start PCR Master Mix (2X) (ThermoFisher), 1µl 10µM of each relavant primer (given in table 2.2) and made up to 20µl with RNase free water.

2.2.1 Chromosomal sexing

Chromosomal sexing was performed using a PCR assay amplifying a region of Ube1X and Ube1Y1, present on both the X and Y chromosome respectively, though a larger band is amplified from the Y chromosomal copy (Warr et al., 2009). Reactions were performed with the following assay: 95 ˚C for 5 minutes; 30 cycles of 30 seconds at 95 ˚C, 30 seconds at 54 ˚C, 30 seconds at 72 ˚C; followed by 5 minutes at 72 ˚C. Products were then run on a 2% agarose gel. A single band indicated the DNA was from an XX animal while 2 bands indicates an XY animal.

2.2.2 *Sf1-EGFP*

The *Sf1-EGFP* reporter line was a gift donated from J. Bowles (University of Queensland, AU). Adult males *Sf1-EGFP* were crossed with female *Gadd45g*$^{+/-}$ to generate *Gadd45g*$^{+/-}$ studs carrying the *Sf1-EGFP* transgene. These were then mated with female *Gadd45g*$^{+/-}$ mice to generate *Sf1-EGFP* positive *Gadd45g*$^{+/+}$ and *Gadd45g*$^{-/-}$ embryos of both sexes for FACS sorting. A presence or absence PCR assay targeting *EGFP* was used to maintain the transgene within the colony. Reactions were performed with the following assay: 95 ˚C for 5 minutes; 30 cycles of 30 seconds at 95 ˚C, 30 seconds at 58 ˚C,

30 seconds at 72˚C; followed by 5 minutes at 72˚C. Products were then run on a 2% agarose gel. A single band indicated the DNA was from a transgenic animal.

2.2.3 *Gadd45g*

The *Gadd45g* knockout line was a gift donated by P. Trainor (Stowers Institute for Medical Research, Kansas, USA) previously described in (Hoffmeyer et al., 2001). In this line, most of the coding region contained in exons 1 to 4 has been replaced by a Neo selection cassette. The colony was maintained on a heterozygous background and genotyped using a multiplex assay consisting of a common forward primer and 2 alternative reverse primers. One primer anneals to the endogenous sequence whilst the other to the selection cassette and produce bands of different sizes. Reactions were performed with the following assay: 95˚C for 5 minutes; 30 cycles of 30 seconds at 95˚C, 30 seconds at 60˚C, 30 seconds at 72˚C; followed by 5 minutes at 72˚C. Products were then run on a 2% agarose gel. Two bands indicated that the DNA from from a heterozygous animal, a single band at 150bp indicated a homozygous knockout animal and a single band at 225bp indicated a wild-type animal.

2.2.4 *Gadd45g* BAC Transgenic line

A BAC transgenic line containing the coding region and surrounding sequence of C57BL/6J *Gadd45g* was generated by the MRC Harwell Transgenics team by pro-nuclear injections on 1-cell stage embryos. The BAC, known as RCPI-23-194E15 (E15), consists of the B6 sequence from chromosome 13 bases 51787574 to 52025876, approximately 238 kb, and the BAC backbone pBeloBAC11 (genome build GRCm38.p5). Animals were genotyped for the presence of the BAC using the chloramphenicol resistance gene, *Cmr*, contained on the pBeloBAC11 backbone. Reactions were treated with the

following assay: 95˚C for 5 minutes; 30 cycles of 30 seconds at 95˚C, 30 seconds at 55˚C, 30 seconds at 72˚C; followed by 5 minutes at 72˚C. Products were then run on a 2% agarose gel. A single band indicated the DNA was from an XX animal while 2 bands indicates an XY animal. The presence of a 200bp band indicated the DNA was from a transgenic animal.

In tissues from animals carrying the *E15-Gadd45g* transgene but also containing at least one copy of the *Gadd45g* knockout allele, the endogenous *Gadd45g* sequence was genotyped by quantitive PCR targeted towards the Neo selection cassette (TaqMan). In this assay, a probe is generated towards the target DNA sequence, in this case the sequence of the neomycin gene cassette (*Neo*) used to disrupt the expression of *Gadd45g*. The probe is labelled with the FAM fluorophore at the 5' end and a quencher molecule at the 3' end. While normally the fluorophore is excited by the laser and emits light at 520nm, the presence of the quencher eliminates this emission signal. Primers are designed to target the region bound by the probe for amplification. As Taq polymerase amplifies the target region, the annealed probe is degraded by its 5' to 3' exonuclease activity, releasing the fluorophore from the quencher and allowing the emitted light to be detected. Each round of amplification releases more of the fluorophore, which correlates to the amount of degraded probe and therefore the number of copies of the target region in the sample. The number of cycles taken to reach a specific threshold level of fluorescence is then measured, the cycle threshold (Ct). Samples which produce a lower Ct correspond to a higher copy number of the neo gene, as fewer cycles are required to reach the threshold value. The calculated copy number corresponds to the number of copies of the *Gadd45g* null allele within each sample.

Cycle threshold values were normalised across samples with the CT of *Dot1l*, run in conjunction with the Neo assay. Each sample was run in duplicate alongside several *Neo* heterozygous controls.

Relative copy numbers were then calculated using the CopyCaller software (Applied Biosystems), with n=1 equalling an endogenous heterozygote.

Table 2.2: Primer sequences for general genotyping of mouse colonies, all sequences are in 5' to 3' order.

Gene Name	Forward Primer	Reverse Primer
Chromosomal Sexing	TGGATGGTGTGGCCAATG	CACCTGCACGTTGCCCTT
EGFP	CCTACGGCGTGCAGTGCTT	CGGGCGAGCTGCACGCTGCCGTCCTC
Gadd45g –wt	GCCTAGGCCTTCCTGTGCTA	TGGTCTGATCGACTTGGTGA
Gadd45g –Mut	GCCTAGGCCTTCCTGTGCTA	CTTGCAAAACCACACTGCTC
Neomycin resistance	GGTGGAGAGGCTATTCGGC	GAACACGGCGGCATCAG
Neo probe	TGGGCACAACAGACAATCGGCTG	N/A
Dot1l	GCCCCAGCACGACCATT	TAGTTGGCATCCTTATGCTTCATC
Dot1l probe	CCAGCTCTCAAGTCG	N/A

2.3 Tissue Collection

2.3.1 Embryo Collection

Timed mates were set up to procure embryos of various ages. Male and female mice would be housed together over night. If a vaginal plug was found the following morning, noon that day was considered to be 0.5dpc. Pregnant females were culled by cervical dislocation (Schedule One) according to Home Office regulations. The abdomen was opened and the uterine horns were collected. Embryos were removed from the uterus and immediately placed in ice cold PBS in accordance with Home Office regulations. Embryos aged 14.5 dpc or older were immediately culled by decapitation. Dissection of embryos was carried out in cold PBS using a Leica M80 dissection microscope. Embryos aged 10.5 – 12.5 dpc were staged by counting tail somites (ts), outlined in appendix i.

2.4 Tissue Processing

2.4.1 Fluorescence-activated cells sorting (FACS)

Embryos for cell sorting were removed from the dissected embryos and placed into PBS in individual wells of 24 well tissue culture plates. A small piece of tissue from the neck of each embryo was then used for genotyping and immediately processed. Once genotyping was complete, UGRs were sorted into relevant experimental groups. The pooled gonads were sub-dissected, removing the mesonephros and gonad from the dorsal aorta and other parts of the urogenital ridge and placed in individual 1.5ml tubes with minimal PBS and kept on ice. Gonads were then resuspended in 250µl of 0.25% trypsin in phenol-red free DMEM and incubated for 5 mins at 37˚C. 25µl of FBS was added to each sample to stop the reaction and mixed by tapping. 500µl of PBS was then added followed by 10 minutes of centrifugation at 300 x g at 4˚C. The supernatant was removed and samples were resuspended in 500µl of fresh PBS. Gonads were mechanically dissociated by successively pipetting with a P1000, P200,and a 23Ga needle using a 1ml syringe. The suspended cells were then passed through a 30µm filter and washed with 250µl PBS to remove residual cells in the filter.

2.4.2 Wholemount in situ hybridisation

Gonads were harvested from 10.5 dpc to 14.5 dpc embryos. Embryos aged 12.5 dpc and younger were staged by counting the number of tail somites (ts) posterior to the hind limbs. Urogenital ridges (UGRs) in 11.5 dpc to 12.5 dpc, or the paired gonad and mesonephros from 13.5 dpc to 14.5 dpc, were dissected and fixed in 4% paraformaldehyde in PBS overnight. The next day, tissue was dehydrated through a series of methanol washes. Tissue was washed for 5 minutes in each solution twice in PBS, 25% methanol in PBS, 50% methanol in PBS, 75% methanol in PBS and twice in 100% methanol. Tissue was then stored in 100% methanol at -20˚C.

2.4.3 RNA extraction

RNA was extracted from samples using an AllPrep DNA/RNA micro kit (Qiagen) according to the supplied instructions. Frozen samples were resuspended in

2.4.4 Histology

Tissues required for histological sectioning were fixed in 4% PFA in PBS overnight and transferred to PBS and processed into paraffin wax by the MRC Harwell Histology team. Tissues were embedded in a transverse orientation and 4µm sections were and placed onto glass slides before eosin staining by the MRC Harwell histology team.

2.5 BAC recombineering

BAC recombineering was performed using the Counter-Selection BAC modification Kit (Gene Bridges). All antibiotic concentrations are standardised as the following across this protocol:

- Chloramphenicol 30µg/ml
- Streptomycin 50µg/ml
- Tetracycline 3µg/ml
- Kanamycin 50µg/ml

2.5.1 Generation of rpsL/neo flanked with homology arms

Primers were designed carrying the 50 nucleotides directly upstream of and including the *Gadd45g* translational start site or directly upstream of the translational stop codon, 5' of the sequence:

5'-GGCCTGGTGATGATGGCGGCGGGATCG-3'

The 50 nucleotides immediately downstream of the *Gadd45g* translational start site or the translational end site (stop codon included) were transformed into the reverse compliment orientation and placed 5' of the sequence:

5’-TCAGAAGAACTCGTCAAFAAGGCG-3’

(See Table2.3 for complete primer sequences)

Primer pairs for the translational start site or the end site were then used with the supplied *rpsL/neo* template in a PCR reaction to generate the homology arm flanked *rpsL/neo* selection cassette in a 50µl reaction: 39.5µl ddH_2O. Five reactions were then run on a 1% agarose gel and size checked for the correct PCR product, the 1320bp selection cassette plus the 2 x 50bp homology arms. Products were then cleaned using a PCR purification kit (Qiagen).

2.5.2 Transformation of the pRed/ET plasmid

E. coli cultures containing the BAC to be modified were first checked for sensitivity to streptomycin, a requirement for this protocol. Cultures were inoculated into LB containing chloramphenicol or chloramphenicol and streptomycin and grown overnight at 37˚C.

Cultures which successfully grew were then inoculated into 1.4ml LB with chloramphenicol and cultured at 37˚C with 250 rpm shaking for 2 hours. Cultures were then centrifuged at 13,000 x g for 30 sec at 2˚C. The supernatant was removed and the pellet resuspended in 1ml chilled ddH_2O. The centrifugation was repeated and the cells were resuspended in 30µl ddH_2O. 1µl of the supplied *pRed/ET* was then mixed with the cells and transferred to a 1mm chilled electroporation cuvette. Cells were electroporated at 1350 V and a 5ms pulse. Cells were resuspended in 1ml LB and placed in a 1.5ml Eppendorf tube. Cells were incubated at 30˚C for 70 mins with 250 rpm shaking before plating on LB agar containing chloramphenicol and tetracycline. Plates were then grown overnight at 30˚C. The following day colonies were picked and grown in 5ml LB containing chloramphenicol and tetracycline at 30˚C the next night to confirm the presence of *pRed/ET*.

2.5.3 RedET mediated recombineering

For the first round of recombineering, fresh bacterial cultures would be prepared for the following day. 30µl of fresh culture were inoculated into 1.4ml fresh LB with tetracycline and chloramphenicol and grown at 30˚ for 2 hours. These cells were then supplied with 50µl 10% L-arabinose in ddH_2O and grown at 37˚C for 1 hour. Once the Red recombinases had been induced, cultures were centrifuged for 30 seconds at 13,000 x g at 4˚C. The supernatant was removed and cells were resuspended in 1ml chilled ddH_2O. Cells were centrifuged again but this time resuspended in 30µl ddH_2O along with 1.5µl of 100ng/µl *Gadd45g-N-flanked rpsL/neo*. Cells were electroporated at 1350V in a 1mm cuvette for 5ms, and allowed to recover in 1ml LB at 30˚C with shaking for 1 hour. Cultures were then spread onto LB agar plates containing chloramphenicol, kanamycin and tetracycline overnight at 30˚C.

For the second round: 30µl fresh culture was inoculated into 1.4ml of LB containing chloramphenicol, tetracycline and kanamycin. Cells were grown for 2 hours and the recombinases induced as before. After, cells were spun and prepared for electroporation with the homology flanked FLAG/His tag sequence. Cells were then allowed to recover in 1ml LB at 37˚C with shaking before spreading onto LB agar plates containing chloramphenicol and streptomycin. The following day, colonies were then genotyped with the previously used assay surrounding the homologous regions.

Table 2.3 Primers and sequences used in BAC recombineering, all sequences are in 5' to 3' order.

Product	Forward Primer	Reverse Primer
50 bp *Gadd45g* translation start homology flanked Rpsl-Neo	CGGGTTCCCTCCGCACTCTTTTGGATAACTTGCTGTTCGTGGATCGCACAGGCCTGGTGATGATGGCGGGATCG	GCTGTGCTTTCCGGAACTGTATCCTGGCCACGGACTTCTTCCAGAGTCATTCAGAAGAACTCGTCAAGAAGG
50 bp *Gadd45g* translation stop homology flanked Rpsl-Neo	GAGGAGAGCCGCAGCTTCAACGACTGGGTGCCCAGCATCACCCTTCCCGAGGGCCTGGTGATGATGGCGGGATCG	CGCTAGAGTGTCACCAAGTCGATCAGACCAAGGTCCCTGCCAGGCTGTCATCAGAAGAACTCGTCAAGAAGG
50 bp Gadd45g translation stop homology flanked His/FLAG	GAGGAGAGCCGCAGCTTCAACGACTGGGTGCCCAGCATCACCCTTCCCGAGCATCATCACCATCACCACATGGACTACCAAGGAC	CGCTAGAGTGTCACCAAGTCGATCAGACCAAGGTCCCTGCCAGGCTGTCACTTGTCATCGTCGTCCTTGTAGTCCATCTGTGGTG
Gadd45g translation start genotyping	CGCCGAGCCTGTCTATAAAA	ACCAGACAAGGCCAACTCTC
Gadd45g translation stop genotyping	TGCCTTGGAGAAGCTCAGTT	ACATCGCCTGCCAGTCTC
5' 500 bp Gadd45g translation stop homology arm	TCTACGAGTCCGCCAAAGTC	CACCCAGTCGTTGAAGCTG
3' 500 bp Gadd45g translation stop homology arm	TGGTCTGATCGACTTGGTGA	GCCGTGAAACATCCTTTACC

2.6 Fluorescence Activated Cell Sorting (FACS)

Dissociated cell were sorted using an S3e cell sorter (Biorad). Cells were vortexed briefly to ensure they remained in suspension before being loaded into the sorter. Cells are passed through a fluid stream and vibrated to generate a stream of droplets. Droplets were then exposed to a 488nm laser. Sensors measuring the amount of forward scattered light (FSC) and side scattered light (SSC) were used to measure the size and granularity of particles within the droplets. A sensor which could detect light between 510-540nm was used to detect fluorescence from the EGFP within cells. Gates were set up to sort both the EGFP-positive and EGFP-negative cells and were collected into 1.5ml tubes

containing 100µl of PBS. Samples were then spun at 4˚C for 10 minutes at 300 x g. The excess PBS was removed and cells were stored at -80˚C.

2.7 CpG Methylation

2.7.1 Bisulphite conversion

DNA was bisulphite-converted using either the EZ DNA methylation Direct or EZ DNA methylation Lightning Kits (Zymo). CT conversion reagent is prepared with the addition of 900µl water, 300µl M-dilution buffer and 50µl M-Disolving buffer to a tube of CT conversion reagent powder. The solution is mixed with frequent vortexing for 10 minutes at room temperature. M-Wash buffer was also prepared with the addition of 24ml of 100% ethanol to the 6ml of concentrated M-Wash buffer.

Extracted DNA in a volume of 20µl was mixed with 130µl of the prepared CT conversion reagent and mixed by pipetting. Samples were then incubated in the thermal cycler at 98˚C for 10 minutes, 64˚C for 2.5 hours and stored at 4˚C. For each sample, 600µl of M-Binding buffer was added to Zymo-Spin^{tm} columns followed by the converted DNA. The cap was closed and the column inverted several times before centrifugation at 20,000 x g for 30 seconds at room temperature. After discarding the flow through, 100µl of prepared M-Wash buffer is added to the column and spun at 20,000 x g for 30 seconds. The flow through is discarded again and 200µl of M-Desulphanation buffer is added to each sample and left to stand at room temperature for 20 minutes. After the incubation, samples are then spun at 20,000 x g for 30 seconds and the flow through discarded. The columns are then washed twice with 200µl M-Wash buffer before elution with 11.5µl RNase free water. Samples are then quantified using a Nanodrop ND-8000 using the RNA setting, which more closely resembles converted DNA.

2.7.2 Bisulphite sequencing

CpG methylation at individual sites was assessed by bisulphite-sequencing. DNA was extracted from cells and bisulphite-converted using the EZ Methylation Direct Kit (Zymo). After conversion, bisulphite-converted DNA (BC-DNA) was used as a template in a nested PCR reaction. In the first round 10ng of BC-DNA was used as a template and mixed with 10µl Maxima Hot Start PCR Master Mix (2X) (ThermoFisher), 1µl 10µM of the primers BC-Sry Com and BC-Sry r1 (given in table 2.4) and made up to 20µl with RNase free water. Reactions were then run through the following program on the thermocycler: 1 minute, followed by 33 cycles of 94 °C for 30 seconds, 60 °C for 30 seconds, and 72 °C for 1 minute; followed by a final extension phase of 5 minutes at 72 ° C. For the second round, 1µl of the first round reaction was used as a template in a reaction with 10µl Maxima Hot Start PCR Master Mix (2X) (ThermoFisher), 1µl 10µM of the primers BC-Sry Com and BC-Sry r2 (given in table 2.4) and made up to 20µl with RNase free water. 5µl of each reaction were run on the same PCR program as the first round, and then run on a 2% agarose gel to ensure that no other products had been generated. Remaining PCR products were then purified using a PCR Purification Kit (Qiagen) according to the supplied instructions.

The product was then cloned into the cloning vector pGemT (Promega), PCR products were mixed with 5µl of the supplied ligation buffer, 1µl 50ng pGemT, PCR product in a 3:1 ratio to the molarity of pGemT, 1µl T4 ligase and made up to 10µl with RNase free water. Reactions were then stored at 4 ° C overnight. The next day, plasmids were transformed into DH5α subcloning efficiency cells (ThermoFischer Scientific). 50µl cells were allowed to thaw on ice and then mixed with 1µl of the ligated plasmid and incubated on ice for 30 minutes. Samples were then heat shocked at 42 ° C for 40 seconds and allowed to recover on ice for 2 minutes, then incubated at 37 ° C in a shaking incubator for 1 hour. Cells were then spread on agar plates containing LB media and 100µg/ml ampicillin, and

incubated over night at 37 ° C. The next morning individual colonies would be picked, grown in 5ml LB media the following night at 37 ° C with shaking at 250 rpm. The plasmids from each clone would then be extracted and sent for Sanger sequencing using the M13 reverse primer (GATC Biotech). The obtained sequence reads were analysed using the BiQ analyser tool (Max Planck institute, Germany), which aligns and automatically detects CpG sites calculating their methylation status according to the bisulphite-converted sequences. Lollipop diagrams were then generated using the BiQ analyser's diagram function.

Table 2.4 Primers and sequences used in targeted bisulphite sequencing, all sequences are in 5' to 3' order.

Primer Name	Primer sequence
BC-Sry Com	AAGGAAATGATAGTTAGTTTTGATT
BC-Sry r1	ACCAAAATATACTTATAACAAAAATTTTAA
BC-Sry r2	TCAAACAAAACCCCTCAAATTTATAAA

2.7.3 Multiple Displacement Amplification (MDA)

Bisulphite-converted DNA was amplified with the EpiTect Whole Bisulfitome kit (Qiagen). 50ng of Bisulphite-converted DNA in 10µl RNase free water was incubated with 29µl EpiTect WBA Reaction Buffer and 1µl REPLI-g Midi DNA polymerase at 28 ° C for 8 hours followed by incubation at 95 ° C for 5 minutes. Amplified DNA was then purified with AmPure XP beads and stored at -20 ° C.

2.7.4 Post Bisulphite Adaptor Tagging (PBAT)

Frozen aliquots of sorted somatic gonadal cells were used to generate sequencing libraries by Post-Bisulphite Adaptor tagging at the Babraham Institute, Cambridge. 5µl of approximately 20,000 cells was resuspended in 6µl of 10mM Tris, 12µl PBAT Cell lysis buffer, 1µl Proteinase K and 1µl 0.5pg/µl λ phage DNA. This solution was incubated for 60 minutes at 37 ° C. Bisulphite conversion was then performed using an Imprint DNA Modification Kit (Sigma) according to the manufacturer's

instructions, however converted DNA was purified using the columns and supplied reagents from the EZ DNA Methylation Direct kit (Zymo). The DNA was then eluted using 20µl of 10mM Tris. 19.5µl of the eluted DNA is then mixed with 1µl 10µM dNTPs, 1µl 10µM PBAT-9N-F primer and 2.5µl 10 x NEB buffer 2 before being incubated at 65˚C for 2 minutes and cooled to 4˚C. 1µl of Klenow exo (NEB) was then added to the mix and briefly vortexed before incubating at 4˚C for 5 minutes, rising to 37˚C at a rate of 4˚C per minute and held at 37˚C for 90 minutes before storing at 4˚C. 1µl of Exonuclease I (NEB) was then added to the samples and incubated at 37˚C for 60 minutes, followed by storage at 4˚C overnight.

The next day 24µl of 10mM Tris is added to each reaction followed by 40µl of room temperature Agencourt AmPure XP beads. Each sample was then incubated at room temperature for 5 minutes to allow the DNA to bind to the beads, followed by 5 minutes at room temperature on a magnetic rack. The supernatant was removed and beads were washed twice in fresh 80% ethanol then allowed to air dry for 7 minutes. The dried beads were resuspended in 50µl 10mM Tris. During this time, 10µl per sample fresh Dynabeads M-280 Streptavidin were prepared on a magnetic rack, washed twice with 1ml of 1x PBAT B&W buffer and resuspended in 50µl per sample 2x PBAT B&W buffer. The Streptavadin bead mix was then added to each reaction and rotated at 30 rpm for 20mins at room temperature. The tubes are then returned to the magnet and washed twice with 100µl of 0.1N NaOH, followed by two washes of 10mM Tris. Beads are resuspended in 2µl 10µM dNTPs, 2µ 10µM PBAT-9n-R primer, 5µl 10 x NEB buffer 2 and 39µl water. Solutions were then incubated at 95˚C for 1 minute and cooled to 4˚C before the addition of 2µl Klenow fragment. The temperature then raised to 37˚C at a rate of 4˚C per minute and incubated at 37˚C for 90 minutes, every 30 minutes the reaction was paused and the solutions briefly vortexed to resuspended the streptavidin beads. Samples were then cooled to 4˚C.

Samples were placed on the magnetic rack until the beads had adhered to the side of the tube and the supernatant removed. Beads were then washed twice in 50µl 10mM Tris before being resuspended in 10µl HF Phusion buffer, 1µl Phusion DNA polymerase (both NEB), 1µl 10µM dNTPs, 1µl 10µM PE1.0 primer, 1µl 10µM barcode containing reverse library primer (see table 2.2) and 36µl water. Reactions were then incubated at 95˚C for 30 seconds followed by 10 cycles of 95˚C for 10 seconds, 65˚C for 30 seconds and 72˚C for 30 seconds. Samples were further incubated for 5 minutes at 72˚C before cooled to 4˚C.

Libraries were then purified by placing the reaction on the magnetic rack. The supernatant was retained and then purified using the AmPure beads as described earlier. Libraries were resuspended in 15µl 10mM Tris and stored at -20˚C.

Table 2.5 Primers and sequences used in Post-bisulphite adaptor tagging

Primer Name	**Primer sequence**
PBAT-9N-F	Btn-CTACACGACGCTCTTCCGATCTNNNNNNNNN
PBAT-9N-R	TGCTGAACCGCTCTTCCGATCTNNNNNNNNN
PE 1.0	Phos-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
PE 2.0	ACACTCTTTCCCTACACGACGCTCTTCCGATCT

2.8 Cloning

2.8.1 TA cloning

Double-stranded DNA products flanked by overhanging 3’ adenine bases, a by-product of Taq based PCR amplification, were cloned into pGEM-T vector (Promega). This linearised vector contains 3’ T overhangs to ensure proper ligation, flanked by T6 and SP6 sites for transcription as well as M13

bacteriophage sites used for genotyping. The PCR products and the pGEM-T vector, in a 3:1 molar ratio, were incubated with 1µl T4 ligase and 5µl 2x ligase buffer at 4̊C overnight. As the vector is 3kb and 50ng is required for the ligation reaction, the following formula was used to calculate the volume of PCR product required:

$$\frac{ng\ of\ vector \times kb\ of\ insert}{kb\ of\ vector} \times insert : vector\ molar\ ration = ng\ of\ insert$$

The following day 2µl of each reaction was used for transformations.

Table 2.6 Primers and sequences used genotyping TA cloning products

Gene Name	Forward Primer	Reverse Primer
M13	GTAAAACGACGGCCAGT	GGAAACAGCTATGACCATG

2.8.2 Transformation

Plasmids were transformed into high efficiency JM109 competent cells (Promega) following the manufacturer's instructions. 50µl of competent cells were added to 2µl of ligated plasmid and incubated on ice for 20 minutes. Cells were then heat shocked at 42 ˚C for 40 seconds followed by 2 minutes on ice to recover. 950µl of SOC media was then added to each reaction and incubated at 37 ˚C for 90 minutes in a shaking incubator at 250 rpm. Volumes between 50 - 100µl of culture was spread onto LB agar plates containing 100µg/ml ampicillin and incubated at 37 ˚C overnight.

2.8.3 DNA preparation from plasmid DNA

Colonies were picked from LB agar plates the following day and inoculated into 5ml LB media with 100µg/ml ampicillin. Colonies were genotyped using the M13 bacteriophage sites flanking the insertion site on the pGEM-T. Colonies showing a band equal to the size of the insert +200bp were incubated overnight at 37 ° C with shaking at 250 rpm. The next day plasmid DNA was harvested from the culture using the QIAprep Spin Miniprep Kit (QIAGEN) according to the manufacturer's instructions. The concentration of the eluted DNA was calculated using a Nanodrop ND-8000 (Thermo Scientific) by measuring the absorbance at 260nm.

2.9 Whole Genome Bisulphite Sequencing

PBAT libraries were sequenced at the Babraham Institute (Cambridge, UK). Library quality and concentration was assessed by a High Sensitivity DNA chip (Agilent) run on an Agilent 2100 Bioanalyser, and by qPCR using a KAPA Library quantification kit (KAPA Biosystems).

Sequencing data was quality control checked before alignment to the B6 reference genome MM10 and an *in silico* generated fully converted C → T genome by the MRC Harwell Bioinformatics team.

2.10 Quantitive real-time polymerase chain reaction

Quantitive real-time polymerase chain reaction (qRT-PCR) was performed using the Fast SYBR® Green assay. The SYBR® Green dye is incorporated into double stranded DNA during Taq polymerase based amplification, allowing the quantification of cDNA. Each time the dye is incorporated into a new double stranded DNA molecule, it emits energy detected as fluorescence at 521nm. As the number of DNA molecules increases with each round of amplification the fluorescence increases in intensity, and correlates to the amount of cDNA used as the template. The number of cycles taken to reach a

specific threshold level of fluorescence is then measured, the cycle threshold (Ct). Samples which produce a lower Ct correspond to a more abundant amount of starting cDNA, as fewer cycles are required to reach the threshold value.

2.10.1 RNA extraction and cDNA synthesis

Total RNA was extracted from embryonic gonads stored at -80˚C using an RNA/DNA micro prep Kit (Qiagen) according to the manufacturer's instructions. Lysis buffer containing 1% β-mercaptoethanol is added to frozen tissue and lysed by vortexing briefly. The lysate is then loaded into a "gDNA-eliminator" column, which will bind DNA but not RNA, and centrifuged. The RNA is then precipitated with 70% ethanol and loaded into a second Rneasy column and centrifuged. The bound RNA is washed several times to remove excess salts and then eluted in 16µl RNase free water run through the column. RNA concentrations were then calculated by measuring the absorbance at 260nm using a Nanodrop ND-8000.

200 – 220ng of RNA was used to generate cDNA using a High Capacity Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's instructions. Reactions were performed in a thermocycler under the following program. 10 minutes at 25˚C, 120 minutes at 37˚C, 5 minutes at 85˚C. After synthesis, cDNA was diluted 1:6 in H_2O.

2.10.2 Plate setup and qRT-PCR

Plates were prepared by first making a master mix for each reaction. 1 x Fast SYBR® Green Master Mix containing, in an optimised buffer: AmpliTaq® fast DNA polymerase, SYBR® Green dye, dNTPs and Uracil DNA glycosylase (Invitrogen) was mixed with 0.2µM of each primer specific for the desired reaction. 15µl of master mix was then added to each well of a Microamp fast optical reaction plate (Applied Biosystems), along with 5µl of cDNA (approximately 9ng). Each assay was performed in

duplicate. Real-Time thermocycling and fluorescence detection was performed using the 7500-Fast real time PCR system (Applied Biosystems).

Table 2.7 Primers and sequences used in qRT-PCR

Gene Name	Forward Primer	Reverse Primer
Hprt1	CATTATGCCGAGGATTTGGAA	CACACAGAGGGCCACAATGT
Sry	TTATGGTGTGGTCCCGTGGT	GGCCTTTTTTCGGCTTCTGT
Sox9	AAGAAAGACCACCCCGATTACA	CAGCGCCTTGAAGATAGCATT
Gadd45g	GGAGACCTGCATTGCATCCTCATTT	ACTCGGGAAGGGTGATGCTGG
Sf1	TTCGTCTGTCTCAAGTTCCTCATC	CCTTTACGAGGCTGTGGTTGTT
Rspo1	GCTCCAACTCTCGGAGACAC	GTCACTGTGCCCAGGTAGGT
Oct4	CTCAGCCTTAAGAACATGTG	TTCTCTTGTCTACCTCCCTT

2.10.3 Analysis of qRT-PCR results

After the completion of the reaction, the Ct value was calculated using the 7500 Fast system software (ThermoFisher). All of the following data analysis was performed using Excel 2010 (Microsoft). Relative expression levels were calculated using the comparative Ct method, which provides the ratio (R) of expression of the target gene in comparison to the endogenous control gene *Hprt1* (equation b) (Schmittgen & Livak 2008).

Equation B: Relative expression (R)

$$R = \frac{2^{-Ct_{target}}}{2^{-Ct_{endogenous}}}$$

The mean R of all of the samples of a particular genotype and age range (n= 3 - 4 pairs of sub-dissected gonads) was used to produce a plot of relative expression. The standard deviation (Stdev) of each sample set was calculated and then used to estimate error through the calculation of the Standard error of the mean (STEM) (equation c).

Equation C: Standard error of the mean

$$STEM = \frac{Stdev_{samples}}{\sqrt{number\ of\ samples}}$$

2.11 Droplet Digital PCR

Droplet digital PCR (ddPCR) was carried out on the Bio-Rad QX200 ddPCR system (Bio-Rad). Single chromosome targets were run in parallel with *Dot1L* as a reference for 2 copies per cell. Reactions were made to 20 µl containing 1 µl crude DNA lysate, 1x ddPCR Supermix for probes (Bio-Rad), 225 nM of each primer (two primers per assay for *Cmr* and *Dot1L*) and 50 nM of each probe (*Dot1L* VIC®-labelled probe and *Cmr* FAM™-labelled). These reaction mixes were loaded into the Bio-Rad QX200 AutoDG and droplets generated as per the manufacturer's instructions. Post droplet generation, the oil/reagent emulsion was transferred to a 96 well semi-skirted plate (Eppendorf). Samples were then amplified on a Bio-Rad C1000 Touch thermocycler (95°C for 10 min, followed by 40 cycles of 94°C for 30 s and 58°C for 60 s, with a final elongation step of 98 °C for 10 min). The plate containing the droplet amplicons was subsequently loaded into the QX200 Droplet Reader. Standard reagents and consumables supplied by Bio-Rad were used, including cartridges and gaskets, droplet generation oil and droplet reader oil.

Table 2.8 Primers and sequences used in ddPCR

Primer Name	Sequence
Dot1l forward primer	GCCCCAGCACGACCATT
Dot1l reverse primer	TAGTTGGCATCCTTATGCTTCATC
Dot1l probe	VIC-CCAGCTCTCAAGTCG
Cmr forward primer	GCCCGCCTGATGAATGCT
Cmr reverse primer	CGGTGTAACAAGGGTGAACACTA
Cmr probe	FAM-TCCGGAGTTCCGTATGGCAATGAA

2.12 Wholemount in situ hybridisation

2.12.1 Riboprobe synthesis

Plasmid containing the desired riboprobe sequence is first linearised by restriction digest targeting the 5' end of the sense strand. 1mg of plasmid in 20µl of water was added to 3µl 10 x buffer, 2µl restriction enzyme (all reagents from NEB) and made to 50µl total volume with water. Digests were

incubated at 37˚C for 2 hours. To ensure that complete digestion had taken place, a 5µl sample of cut plasmid was run on a 1% agarose/ TBE gel alongside the uncut plasmid. The linearised DNA was precipitated using 0.1 x volume of 3M sodium acetate (pH 4.5) and 2.5 x volume of 100% ethanol. The solution was then incubated on dry ice for 30 minutes before centrifugation at 3,000 x g at 4˚C or 30 minutes. Supernatant was removed and replaced with 200µl 70% ethanol. The solution was centrifuged again at 20,000 x g at 4˚c for 10 minutes. Ethanol was removed and the pellet allowed to air dry for 5 minutes before resuspension in 20µl water.

To synthesise the riboprobe, 3µl of linearised probe was added to 2µl 10 x transcription buffer, 2µl DIG RNA labelling mix, 1µl 40 U/µl RNase inhibitor, 2µl 20 U/µl RNA polymerase (SP6, T7 or T3) (all reagents from Roche) and made up to 20µl with RNase free water. The reaction was incubated at 37˚C for 2 hours, followed by the addition of 2µl 20 U/µl DNase I and incubation at 37˚ for a further 20 minutes. The reaction was halted and RNA was precipitated by the addition of 2µl 200mM EDTA, 2.5µl 4M lithium chloride and 75µl 70% ethanol. Reactions were then incubated at -20˚C overnight. The following day reactions were centrifuged at 20,000 x g at 4˚C for 30 minutes. The supernatant was removed and the pellet washed in 200µl ice cold 70% ethanol and centrifuged for another 2 minutes. The ethanol was then removed and the pellet resuspended in 100µl RNase free water. Riboprobes were stored at -20˚C until required.

2.12.2 Wholemount in situ hybridisation (WMISH)

WMISH was performed in 12-well culture dishes. Netwell (Fisher) inserts were used to move tissue from well to well. Each step was conducted at room temperature unless stated otherwise. Tissue which had previously been dehydrated by the method described in section 2.4.2 was rehydrated through successive washes in 75%, 50%, 25% methanol in PBS, with 5 minutes for each wash. Tissue was then washed in PBST 3 times for 5 minutes, followed by bleaching in 6% hydrogen peroxide in

PBST for 40 minutes and 3 more 5 minute washes in PBST. Samples were the digested in 10µg/ml proteinase K (Roche) for 20 minutes, followed by 5 minutes of incubation in 2mg/ml glycine to halt the reaction. Three more 5 minute washes in PBST were then performed, followed by post-fixation in 0.2% glutaraldehyde in 4% paraformaldehyde (PFA) in PBST for 20 minutes and two further 5 minute washes in PBST. Tissues were then incubated in pre-hybridisation solution at 70˚C for at least 1 hour. The required volume of RNA probe was denatured at 70˚C for 10 minutes, followed by cooling then added to the hybridisation solution at a concentration of 5µl per ml hybridisation solution. Tissue was then transferred into the probe containing hybridisation and incubated at 70˚C overnight.

The next day, tissue was washed in WMISH wash solution, twice in succession at 70˚C followed by two washes at 65˚C. Tissues were then returned to room temperature and washed 3 times in TBST followed by pre-blocking in TBST containing 10% heat inactivated sheep serum (HISS) for at least 2.5 hours. Meanwhile, embryo powder was resuspended in 1 ml TBST and heat inactivated for 30 minutes at 70˚C. The solution was then centrifuged briefly and the TBST removed. Embryo powder was then resuspended in 1% HISS in TBST. 1µl per ml of reaction anti-DIG-alkaline phosphatase conjugated Fab fragments (Roche) was then added and the solution mixed on a rotor for at least 1 hour at 4˚C. The solution was then centrifuged at 20,000 x g at 4˚C for 5 minutes, the antibody containing supernatant removed and transferred into wells containing 1% HISS in TBST. Samples were then added and incubated overnight at 4˚C.

The next day samples were washed in TBST 3 x 5 minutes followed by 5 x 1 hour all at room temperature and incubated at 4˚C overnight in TBST. The following day, samples were first washed in NTMT for 3 x 10 minutes and incubated in alkaline phosphatase stain (20µl NBT/BCIP (Roche) in NTMT) in the dark until sufficient colour development had occurred. The reaction was stopped by

replacing the solution with PBST and incubated for up to 3 days in the dark to remove excess staining. Samples were then post fixed in 4% PFA for 1 hour, and then stored in 0.4% PFA in PBST at 4 ˚C in the dark. Samples were images on a Leica MZ16 microscope with an attached digital camera, on a 1% agarose dishes flooded with PBS.

2.13 Statistical power calculation

In order to determine the number of required samples for any experiment it is vital to calculate statistical power. Statistical power refers to the probability of not committing a type II error, the null hypothesis is rejected when it is true, when using the supplied number of samples. This was calculated using the online tools available at http://powerandsamplesize.com/ (HyLown Consulting LLC, 2013). For qRT-PCR, mean values were input into the 2 mean: 2-sample equivilance calculator to determine statistical power of the number of samples which had been run. For bisulphite sequencing data, total percentages of methylation across the whole region of each treatment were used in the 2 proportions: 2-sample, 2-sided equality calculator, using the total value of all CpG sites within the sequenced regions as the number of samples. t-tests and Fisher's exact tests were performed on Microsoft Excel.

Chapter 3: Characterisation of a transgenic mouse line overexpressing *Gadd45g*

3.1 Introduction

This chapter will explore the various experiments conducted to investigate potential *in vivo Gadd45g* overexpression constructs. Two bacterial artificial chromosomes (BACs) were investigated. I aimed to investigate the relative level of overexpression, and determine whether the expression of *Gadd45g* from the BAC in transgenic animals shares a similar spatiotemporal profile to that of endogenous *Gadd45g*. I also aimed to establish if an overexpression of *Gadd45g* can rescue the phenotype of known sex reversal models such as B6.Y^{DOM}.

3.1.1 The requirement for *Gadd45g* during mammalian sex determination

In order to properly induce the testicular fate, *Sry* is required to be expressed to a specific threshold value during a specific time window (Hiramatsu et al., 2009). Multiple factors are required to ensure this timely and sufficient expression. Among these is the gene G*add45g*. In the absence of *Gadd45g* XY mice sex reverse on the C57BL/6 background (Gierl et al., 2012; Warr et al., 2012). On a mixed 129/Svlmj-C57BL/6 background, XY sex reversal is observed at a lower penetrance. XY mice can also develop with an intersex phenotype (Johnen et al., 2013). XY sex reversal in *Gadd45g*$^{-/-}$ mice is caused by a delay in *Sry* expression (Warr et al., 2012). This is hypothesised to be the result of a failure to activate the MAP3K4/p38-MAPK phosphorylation cascade, which phosphorylates GATA4 allowing it to bind to and activate the *Sry* promoter (Gierl et al., 2012).

The range of functions of members of the GADD45 family, as shown in section 1.4, suggests that the other roles of GADD45γ may be important in mammalian sex determination. Despite the potential role of GADD45γ in active DNA demethylation, its absence has been shown not to affect the methylation profile of the proximal *Sry* promoter (Gierl et al., 2012; Warr et al., 2012). The absence of an effect on *Sry* does not rule out the possibility that other areas of the genome may require the

active DNA demethylating activity of GADD45γ to ensure proper testis determination. It is also unclear whether its pro-MAPK signalling function (Bogani et al., 2009) is is an aspect of its potential role in DNA demethylation. GADD45γ and DNA methylation will be explored more in chapters 5 and 6.

3.1.2 B6.Y^{DOM} models of *Sry*-dependent gonadal sex reversal

Despite the stringent requirements of *Sry* expression to induce testis development, inbred strains of mice are able to properly develop testes. Distinct strains however, show different levels of penetrance of sex reversal when harbouring mutations in known sex determining genes. This phenomenon can be exploited to understand the interaction between the testis and ovary determining pathways (Nikolova et al., 2008; Munger et al., 2009; Gierl et al., 2012). The C57BL/6J (B6) strain appears particularly sensitive to sex reversal, as mutations which disrupt testis determination are found to increase the severity of sex reversal phenotypes once backcrossed to this particular strain.

One much studied model of sex reversal occurs when the Y chromosome of sub-species of *Mus domesticus* is introgressed onto the B6 background. After successive backcrosses to B6, the Y chromosome from the *M. domesticus* is the only part of the genome which is retained. The offspring produced, known as B6.Y^{DOM}, may show disruption to testis determination to varying degrees depending on the *M. domesticus* subspecies. The severity of the disruption is classified into three levels. B6 mice carrying the Y chromosome from the FVB/N strain, do not cause sex reversal and constitute the first class. The second class, such as Y chromosome for the AKR/J strain, contain ovarian tissue in their fetal gonads during gonad development but resolve to form testis in adulthood. The final classification, which includes the B6.Y^{POS} strain, is associated with sex reversal.

The B6.Y^{POS} strain is derived from a wild subspecies from Val Poschiavo (Switzerland). Approximately 75% of B6.Y^{POS} will develop as females, the remaining 25% as intersex (Bullejos and Koopman, 2005).

Other strains, such as DBA/2J and BALB/c, have been crossed to generate Y^{POS} derivatives. However, these strains do not show sex reversal, highlighting the increased sensitivity of C57BL/6 to disruptions to testis determination (Eicher et al., 1996; Eicher and Washburn, 2001). Studies of gonadal development in B6.Y^{DOM} strains suggest that sex reversal is caused by a delay in the expression of *Sry* relative to C57BL/6J (B6.Y^{B6}). (Bullejos and Koopman, 2005).

B6.Y^{DOM} lines can be used to increase the phenotypic effect of mutations which have a limited effect on testis determination in B6. Double knockouts of *Map2k3* and *Map2k6* are embryonic lethal at 11.5dpc and so cannot be investigated for their effect on sex determination without a complex double conditional knockout approach. *Map2k3*$^{+/-}$, *Map2k6*$^{-/-}$ embryos show limited amounts of ovarian tissue in 14.5dpc XY gonads. When crossed onto the B6.Y^{AKR} strain, however, full XY gonadal sex reversal is seen (Warr et al., 2016). Similarly, XY *Gadd45g*$^{+/-}$ develop normal testis on B6. However, on B6.Y^{AKR} full and partial XY sex reversal is observed in *Gadd45*$^{+/-}$ gonads. This phenotype was rescued with the expression of a BAC transgene overexpressing *Map3k4*, indicating that it and *Gadd45g* genetically interact to ensure the timing of *Sry* expression (Warr et al., 2012). This same *Map3k4* transgene was used to rescue the *T*-associated sex reversal (Tas) model of sex reversal (Warr et al., 2014).

3.1.3 Expression studies using transgenic mice

Transgene expression in mice is an important tool in establishing the role of genes in sex determination. Transgenic expression of *Sry* and *Sox9* in XX embryos revealed their role in promoting testis determination (Koopman et al., 1991; Vidal et al., 2001). One approach to studying the place

occupied by *Gadd45g* in the testis-determinating gene network is studying the regulatory regions which promote its expression. By comparing the expression of various transgenic constructs, insight can be gained about the surrounding regulatory regions and their requirements for expression in specific tissue types. Such studies have shown the presence of binding sites at different positions for ASCL1 (Huang et al., 2010); OCT and NF-Y(Campanero et al., 2008); and MicroRNA-383, a negative regulator of Gadd45g expression. (Zhao et al., 2014).

An *in vivo* Gadd45g-d4Venus construct, using the 1767bp 5' of the Gadd45g transcriptional start-site, was sufficient to induce expression of the D4Venus-GFP throughout the developing central nervous system (Kawaue et al., 2014). However no expression was reported in the developing gonad (Kawaue pers. comm.).

Though the majority of regulatory sequences are located close to the transcription start-site of their respective genes, some enhancer elements can be found many kilobases 5' or 3' of the coding sequence. One such element required for sex determination is the TESCO enhancer of *Sox9*, which is located 14kb 5' of *Sox9*. Other regions controlling the expression of this particular gene have been mapped to regions 1Mb upstream of Sox9 (Sekido and Lovell-Badge, 2008). One way to identify regulatory sequences is to use an overexpression construct containing large amounts of the upstream and downstream sequence.

Bacterial artificial chromosomes (BACs) are large regions of chromosomal sequence cloned into vectors based on the bacterial fertility-plasmid (F-plasmid). These cloned regions can encapsulate up to 500kb of transgenic DNA, and are therefore more likely to contain distal regulatory elements. BACs are therefore more likely to reproduce similar spatial-temporal expression patterns to endogenous

genes, while expressed in a manageable copy-number dependent manner (Van Keuren et al., 2009). The previously described *Map3k4* transgene was developed through BAC transgenesis. Producing a similar transgenic line overexpressing *Gadd45g* would provide a tool to further investigate *Gadd45g* activity in testis determination. Moreover, a functional BAC, if identified, could be used to introduce modifications into the *Gadd45g* gene, such as an epitope tag to aid protein detection or a fluorescent reporter to aid imaging.

3.1.4 Generation of a *Gadd45g* BAC transgene

Previously in the lab, two BACs containing the *Gadd45g* coding region were chosen to develop transgenic mouse lines (Fig 3.1). RPCI-23-97K5 (*K5-Gadd45g*) and RPCI-23-194E15 (*E15-Gadd45g*) cover bases 51683090 – 51881327 and 51787574 – 52025877 on chromosome 13, respectively (genome build GRCm38.p5). These particular BACs were chosen due to their large coverage of areas 5' or 3' of the *Gadd45g* locus. *K5-Gadd45g* covers 198.2kb, including 163.9kb upstream of *Gadd45g* and 32.8kb downstream. This area also encompasses the coding region for the genes *Sema4d* and an unknown protein coding transcript. *E15-Gadd45g* encompasses a 238.3kb region covering 59.1kb 5' of *Gadd45g* and 177.4kb 3' of the gene. A suspected non-protein coding transcript lies at the 3' end of this region. However, *Sema4d* is not included in the sequence of this BAC.

BACs were acquired (Life Technologies) and the *Gadd45g* exons were sequenced to ensure the entirety of the gene was present and matched the reported sequence. Once confirmed, BACs were extracted from the cells and micro-injected into 1-cell stage embryos by the transgenics facility at MRC Harwell (Gardiner and Teboul, 2009) (Fig 3.2). After numerous injection sessions founders for both *K5-Gadd45g* and *E15-Gadd45g* were generated (Table 3.3). Founders were tested for the presence of the chloramphenicol resistance gene *Cmr* found on the BAC. Founders were then mated with B6 mice and the resulting offspring were tested for the presence of the transgene. Two founders

were generated using the *K5-Gadd45g* construct which were not found to transmit. Five transgenic *E15-Gadd45g* founders were produced, of these lines 1, 2, 3 and 5 transmitted (Nick Warr pers. comm.).

3.1.5 *E15-Gadd45g* rescues the *Gadd45g*$^{+/-}$ Y^{AKR} sex reversal phenotype

In order to test if the transmitted *E15-Gadd45g* was functional in the context of sex determination, the transgene was crossed onto B6.Y^{AKR} *Gadd45g*$^{+/-}$ mice, in an attempt to rescue the sex reversal phenotype. This series of crosses was chosen over rescue of *Gadd45g*$^{-/-}$ on the C57BL/6 background as the genotyping assays used at the time could not distinguish between heterozygous and homozygous *Gadd45g* knockouts in the presence of the transgene, which would amplify in addition to any endogenous wild-type sequence (Fig 3.3). Transgenic lines were first crossed with B6.Y^{AKR} males to produce *E15-Gadd45g-B6-Y*^{AKR} offspring. Studs from this line were then crossed with *Gadd45g*$^{+/-}$ females to generate B6.Y^{AKR} *Gadd45g*$^{+/-}$ pups, some of which would carry the *E15-Gadd45g* transgene. As it had previously been shown that B6.Y^{AKR} *Gadd45g*$^{+/-}$ embryos develop ovaries (Warr et al., 2012), it was hypothesised that sufficient functional expression of *Gadd45g* from a transgene would rescue the sex reversal phenotype. B6.Y^{AKR} *Gadd45g*$^{+-/}$ gonads carrying *E15-Gadd45g* produced from the fifth founder line, were found to express *Sox9* in the testis cords throughout the gonad and lacked expression of the meiotic germ cell marker *Stra8*, indicating a rescue of the sex reversal phenotype (Fig 3.4). The usual limited amounts of *Stra8* expression and disruption of the testis cords can be seen at the poles of B6.Y^{AKR} control gonads (Fig 3.4 B and G). This control phenotype was also rescued with the presence of the transgene (Fig 3.4 D, E, I and J).

In light of these data, I aimed to test more hypotheses concerning the *E15-Gadd45g* transgene. Would it be able to rescue the sex reversal phenotype in *Gadd45g*$^{-/-}$ B6 mice and more severe Y^{DOM} sex reversal phenotypes such as Y^{POS}? How much of an increase in expression was produced from the BAC and is expression found within the same tissues and timeframe as endogenous *Gadd45g* within the gonad? These are important experiments that would allow me to determine whether this BAC has the regulatory elements that drive *Gadd45g* expression with its normal spatiotemporal profile and whether it was, therefore, suitable for being manipulated in order to introduce alterations into *Gadd45g* for subsequent applications.

Hypotheses:

A *Gadd45g* BAC will induce an overexpression of *Gadd45g* within a transgenic animal.

Transgenic *Gadd45g* expression will be limited to the same spatiotemporal regions as endogenous *Gadd45g*.

Overexpression of *Gadd45g* will rescue models of sex reversal such as B6.Y^{DOM}.

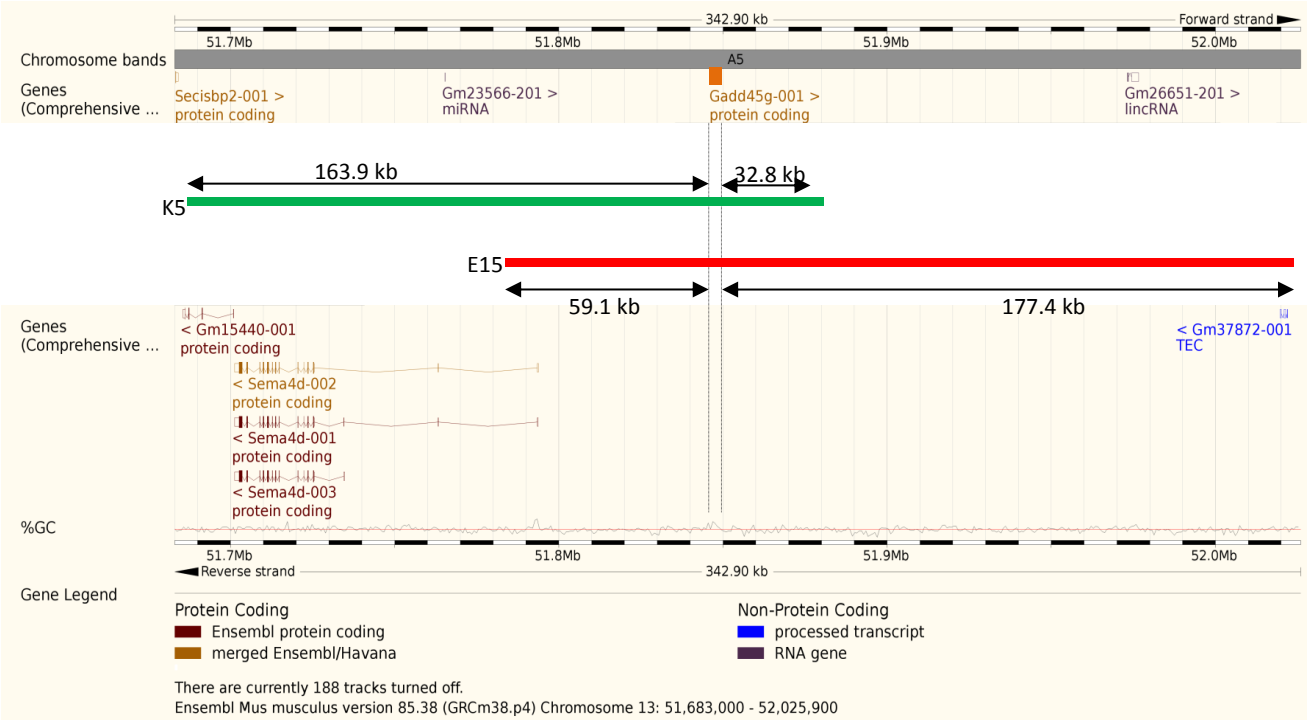


Figure 3.1 Location of the sequences encoded in the *E15-Gadd45g* and *K5-Gadd45g* BACs
Edited from Ensembl genome browser. The diagram shows the region of chromosome 13 at the *Gadd45g* locus (orange box) along with the regions encoded by the *E15-Gadd45g* and *K5-Gadd45g* BACs, highlighted in red and green respectively. The respective distances 5' and 3' of the *Gadd45g* gene covered by the BACs are shown with arrows in black.

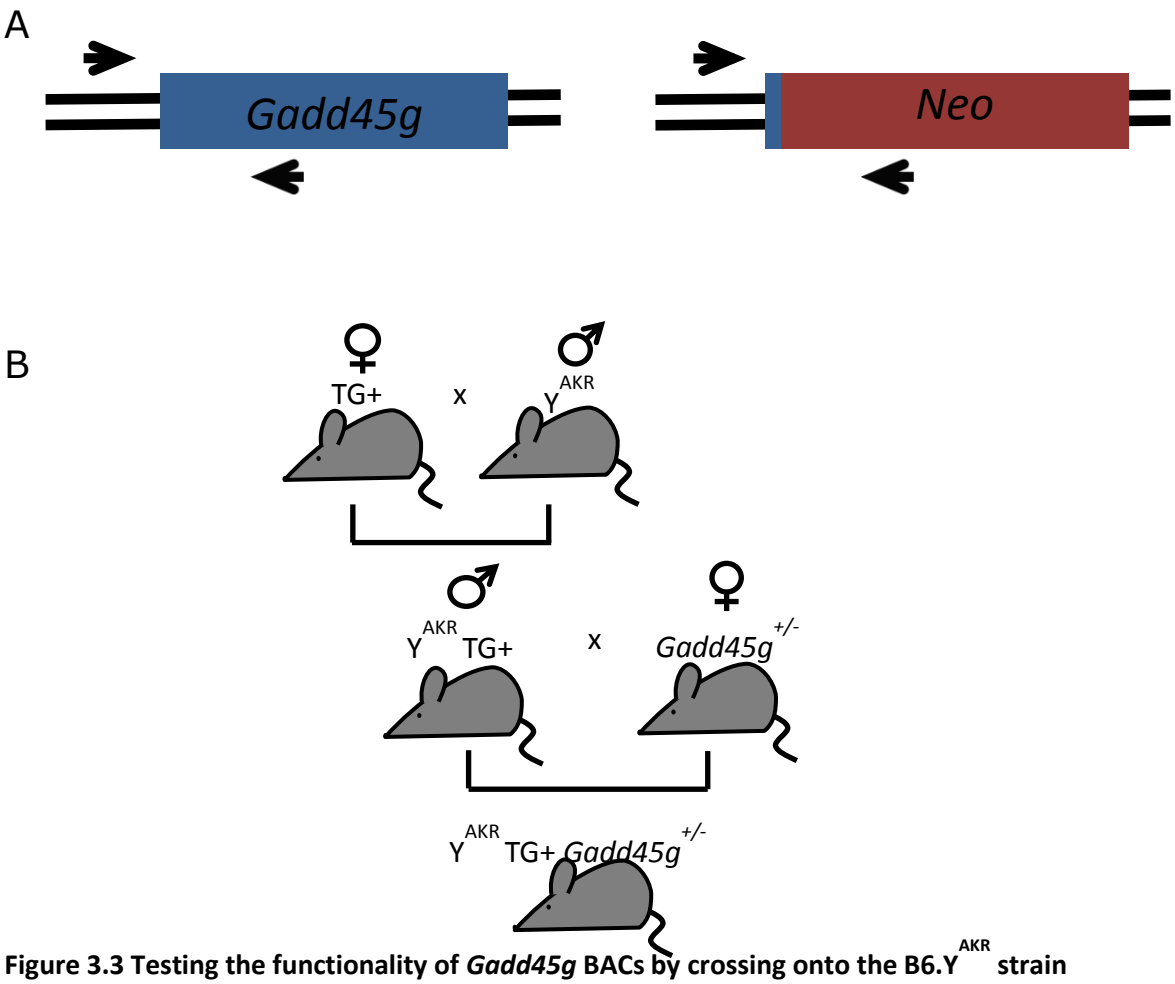


Figure 3.3 Testing the functionality of *Gadd45g* BACs by crossing onto the B6.Y^{AKR} strain
A) The *Gadd45g* genotyping assay used in the lab uses 3 primers: a common forward primer; a reverse primer targeting the *Gadd45g* sequence; and a reverse primer targeting the *Neo* selection cassette used in the knockout allele. Amplicons from the wild-type allele produce a 217bp band, while the knockout allele produced a 120bp band. The presence of the *Gadd45g* transgene will produce large amounts of the wild-type band. This prevents distinguishing between transgenic animals heterozygous and homozygous knockout for *Gadd45g.* **B**) The functionality of the transgene is instead tested on the B6. Y^{AKR} strain, in which *Gadd45g*$^{+/-}$ induces *XY* sex reversal. Transgenic females are first crossed to B6.Y^{AKR} males. Male offspring carrying the transgene are then crossed to *Gadd45g*$^{+/-}$ females, producing B6.Y^{AKR} *Gadd45g*$^{+/-}$ carrying the transgene.

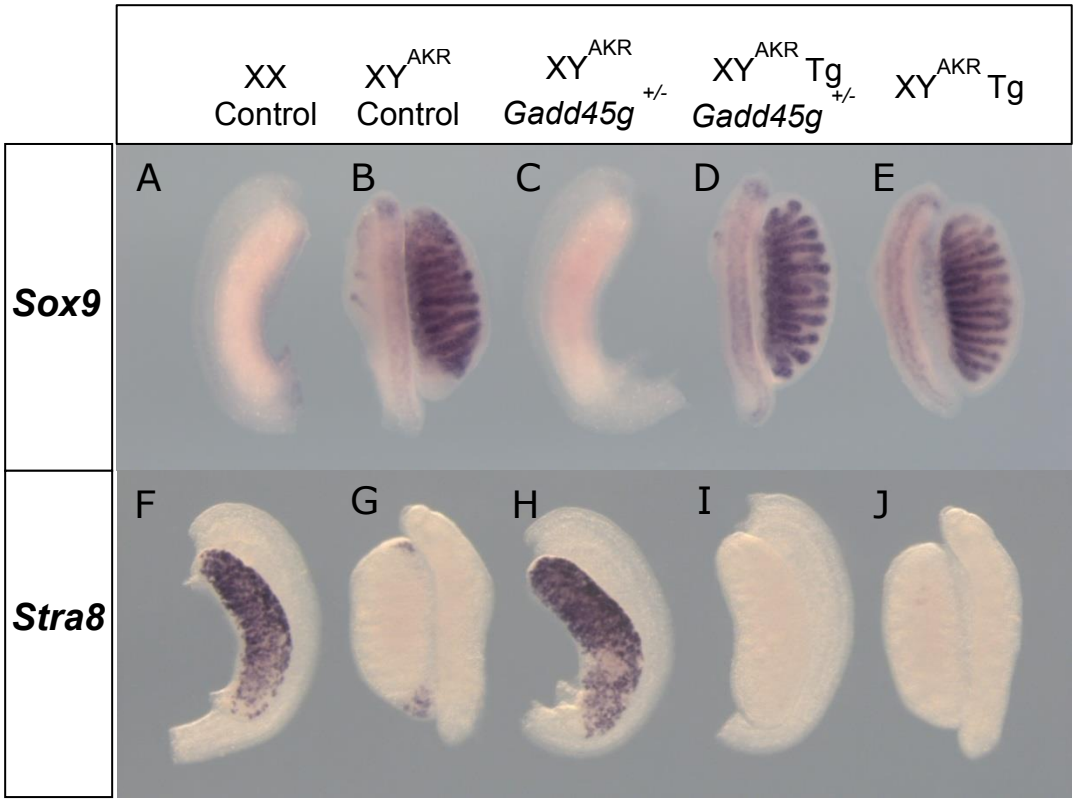


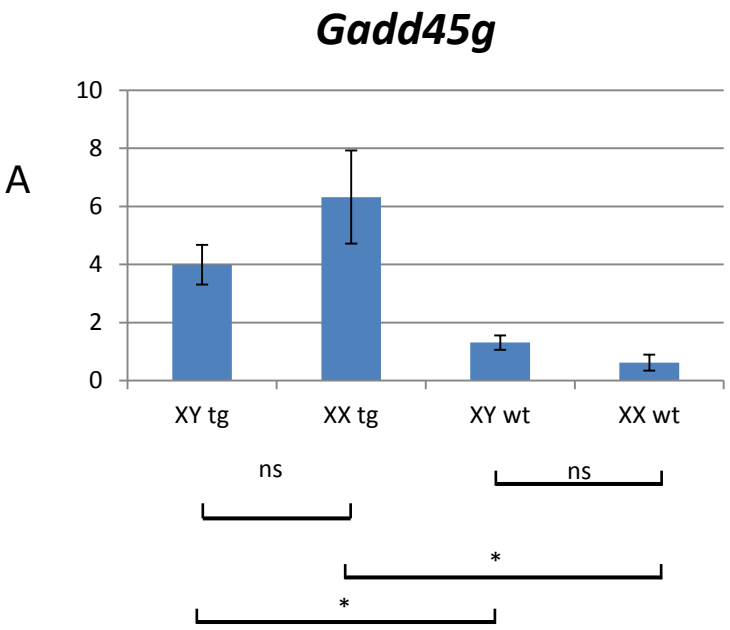
Figure 3.4 Overexpression of *Gadd45g* rescues B6.Y^{AKR} *Gadd45g*$^{+/-}$ sex reversal
WMISH was performed on 14.5 dpc gonads from the strains indicated in the top panels. Expression of the Sertoli cell marker *Sox9* (**A-E**) and the meiotic germ cell marker *Stra8 (***F–J**) were used to mark testicular or ovarian tissue respectively. In XY C57BL/6 gonads *Sox9* is expressed throughout the testis cords, whereas XX C57BL/6 gonads do not show any *Sox9* expression (**A**). Expression of *Stra8* is found throughout wild-type XX B6 gonads (**F**) but not in testicular tissue. B6.Y^{AKR} gonads show expression of *Sox9* but does not extend to the poles of the gonad (**B**)**,** where *Stra8* expression is found (**G**)**.** Testis cord structure is also disrupted at the poles. When one copy of *Gadd45g* is deleted on the B6.Y^{AKR} background, XY sex reversal is induced, *Sox9* expression is absent and *Stra8* expression is present (**C** & **H**). In the presence of the *E15-Gadd45g* transgene the sex reversal is rescued. *Sox9* expression is restored and *Stra8* expression is not observed (**D & I**). The structure of the testis cord is also restored at the poles of the gonad, indicating that the B6.Y^{AKR} disruption phenotype is also rescued. Data courtesy of Dr Nick Warr.

3.2 Results

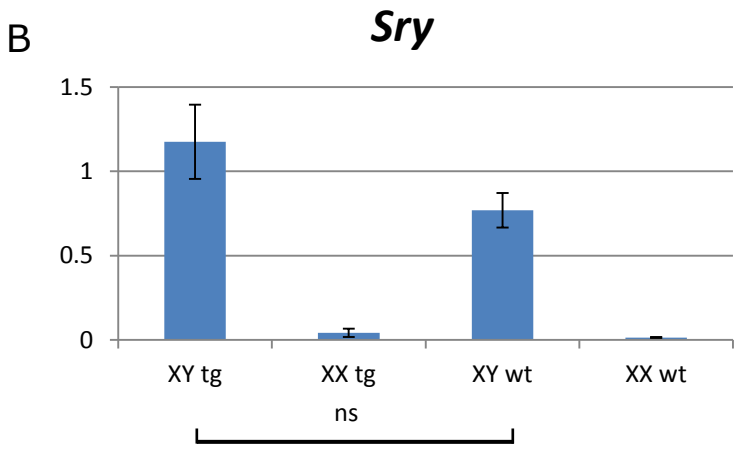
3.2.1 *E15-Gadd45g* induces overexpression of *Gadd45g* but does not increase *Sry* beyond normal C57Bl/6J levels

Gadd45g expression within the developing gonads has a similar spatiotemporal profile to *Sry*, though precedes the onset of *Sry* expression. The peak of *Gadd45g* expression occurs at 18 ts, when it is found throughout the gonadal ridge (Warr et al., 2012). The amount of *Gadd45g* overexpression contributed by the *E15-Gadd45g* was determined using quantitive real time PCR (qRT-PCR). Gonads were collected from 11.5dpc (18 ts) wild-type B6 XY and XX embryos, and XY and XX embryos carrying the *E15-Gadd45g* BAC transgene. Gonads were dissected away from the mesonephros and stored at -80°C until needed. RNA was extracted from 4 pairs of gonads of each genotype and treated with reverse transcriptase to generate cDNA. cDNAs were then used as the template in a series of qRT-PCR reactions targeting *Gadd45g, Sry* and *Hprt1*. Cycle threshold values were collected and used to generate relative expression levels between the four genotypes according to section 2.9.3 (Fig 3.5).

Wild-type XY and XX gonads did not show significantly different levels of *Gadd45g* expression (2-tailed t-test, p=0.132735), as expected. When compared to their wild-type counterparts, transgenic XX and XY gonads showed a 10.2 fold and 2.6 fold increase in *Gadd45g* expression, respectively. However, statistical analysis showed that there was no significant difference in *Gadd45g* expression between transgenic XX and XY gonads (2-tailed t-test, p=0.298614), or between wild-type XX and XY gonads. When averaged across all transgenic and all wild-type individuals, the transgene produced a 5.2 fold increase in *Gadd45g* expression. GADD45γ is required for *Sry* expression at this time; however, there was no significant difference in *Sry* expression between XY wild-type and XY transgenic gonads (2-tailed t-test, p=0.144744).



P-value	Gadd45g	Statistical power
xy tg v xx tg	0.298614	0.6088
xy tg v xy wt	0.014788	0.7284
xx tg v xx wt	0.030762	0.7095
xy wt v xx wt	0.132735	0.4338



P-value	*Sry*
xy tg v xy wt	0.144744

Figure 3.5 *E15-Gadd45g* induces an approximate 5.2 fold increase in *Gadd45g* and does not increase expression levels of *Sry* expression at 18 ts on C57BL/6J

Quantitive real-time PCR on 11.5 dpc (18 ts) reveals an average of 5.2-fold increase in *Gadd45g* expression in gonads from embryos carrying the *E15-Gadd45g* transgene (**A**). No significant difference was found in expression between XY and XX transgenic gonads (2-tailed T-test, P values and statistical powers in accompanying table). **B**) Despite increased levels of *Gadd45g* expression, transgenic XY gonads did not show a significant difference in *Sry* expression at this stage compared to wild-type XY gonads (2-tailed T-test, P =0.144744, statistical power =0.2587).

3.2.2 E15-Gadd45g specific expression shares a spatial expression profile with endogenous Gadd45g, but not similar temporal profile

The increased expression of *Gadd45g* from the transgene may have led to expression in cells which do not normally express *Gadd45g*. To confirm that transgene specific expression of *Gadd45g* shared a similar spatial-temporal profile to that of endogenous *Gadd45g*, mice lacking the endogenous *Gadd45g* sequence but carrying the transgene were generated. Male carriers of *E15-Gadd45g* were crossed with *Gadd45g*$^{+/-}$ females. *Gadd45g*$^{+/-}$ males carrying the transgene were then crossed with *Gadd45g*$^{+/-}$ females to produce transgenic mice lacking endogenous *Gadd45g* in a 1:8 ratio. Due to the necessity of finding transgenic embryos lacking endogenous Gadd45g a new genotyping assay was designed. Mice were genotyped for the endogenous *Gadd45g* locus by qPCR targeting the *Neo* gene found in the knockout allele to distinguish between heterozygous and homozygous *Gadd45g* knockouts carrying the transgene.

At 10.5 dpc *Gadd45g* expression is first seen at the 9 ts stage in the centre of the gonadal ridge of both XY and XX C57BL/6 embryos. This expression pattern then extends towards the poles at 10 ts. By the 11 ts stage, expression is found throughout the gonad. Expression of *Gadd45g* peaks at 18 ts (11.5 dpc) where it is limited to the body of the gonad and not found in the mesonephros or coelomic epithelium. After 20 ts, Gadd45g expression begins to decline and by 13.5 dpc, expression cannot be detected (Warr et al., 2012).

A series of wholemount *in situ* hybridisations (WMISH) was performed on gonads from wild-type, *Gadd45g*$^{-/-}$ and transgenic *Gadd45g*$^{-/-}$ XX and XY animals. 3 – 4 UGRs or gonad pairs were used per stage and a representative gonad was selected for photography A probe designed to target between exon 2 and 3, regions deleted in the knockout allele (Hoffmeyer et al., 2001), was used to

differentiate between wild-type *Gadd45g* and any short transcription products from the knockout allele (Fig 3.6).

WMISH using this strategy using gonads from 9 – 10 ts stage embryos showed expression of *Gadd45g* in B6 gonads but had not reached the poles of the gonad. Conversely, expression in gonads carrying the transgene showed expression throughout the gonad and at a much higher intensity (Fig 3.7).

WMISH using 17 – 18 ts gonads revealed both XY and XX B6 gonads show full expression of *Gadd45g* in the gonad, but not in the mesonephros. XY and XX gonads expressing *Gadd45g* from the transgene only showed a higher intensity of expression and limited expression in parts of the gonad-mesonephros boundary (Fig 3.8). Transverse sections of 4µm were cut and stained with eosin to check for expression within the coelomic epithelium (Fig 3.9). In transgenic XY and XX gonads no expression was seen within the coelomic epithelium.

By 13.5 dpc B6 gonads no longer express *Gadd45g*. However expression from the transgene persists at this stage. In XY gonads expression is limited to the testis cords. The morphology of the XY transgenic *Gadd45g*$^{-/-}$ gonads is robustly testicular, indicating that *E15-Gadd45g* rescues the sex reversal phenotype of *Gadd45g*$^{-/-}$. Expression in XX gonads was found throughout the ovary but not in the ducts (Fig 3.10).

3.2.3 E15-Gadd45g partially rescues the sex reversal of B6.Y^{POS}

As *E15-Gadd45g* has previously been shown to rescue sex reversal in B6.Y^{AKR} *Gadd45*$^{+/-}$ and improve testis formation in B6.Y^{AKR} embryos, the potential of *E15-Gadd45g* to rescue other B6.Y^{DOM} derivatives was investigated. The most severe sex reversal appears in B6.Y^{POS} mice. Transgenic females were

crossed with fertile B6.Y^{POS} males. The resulting male offspring were genotyped for the presence of the transgene and then crossed to B6 females to generate embryos at 14.5 dpc.



Figure 3.6 Expression of the probe *Gadd45g-del* and the complementary sense strand probe.
WMISH was performed on wild tpye 11.5dpc gonads from 18 ts embryos using the *Gadd45g-del* probe, which targets exons 2 and 3 of *Gadd45g*, which are replaced by a *Neo* selection gene in the knockout allele. The corresponding sense strand probe was run alongside as a negative control.

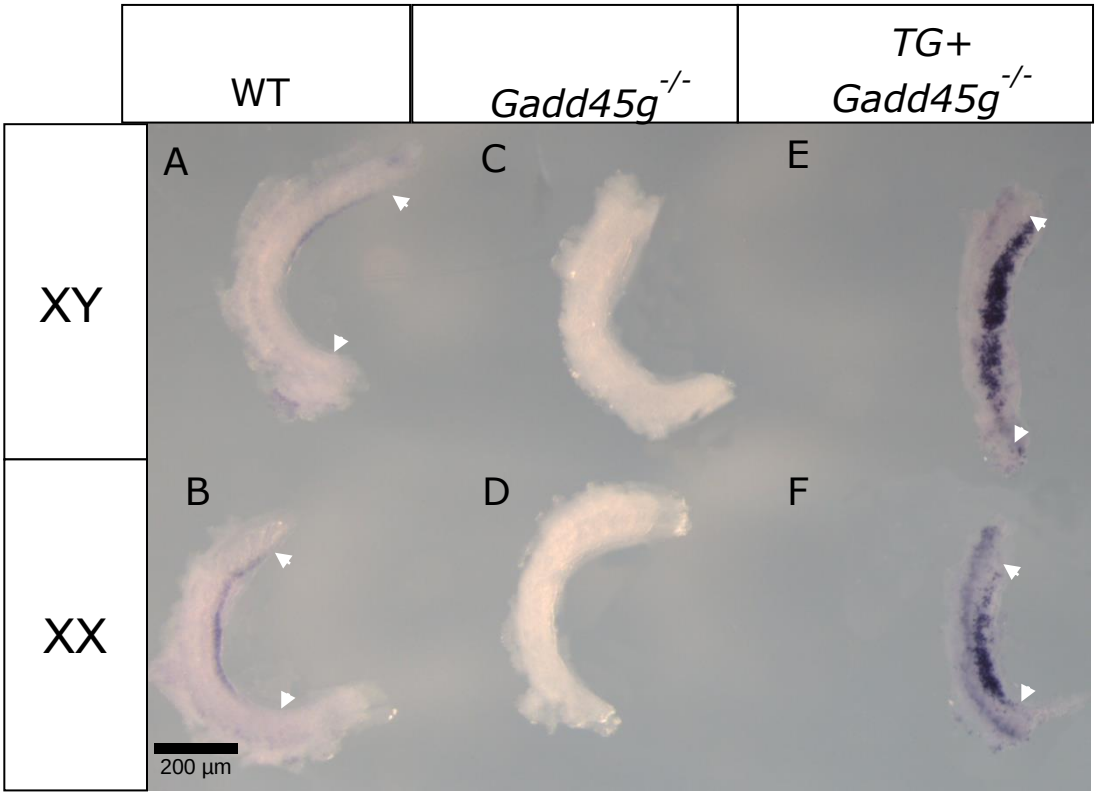


Figure 3.7 Expression of *Gadd45g* from the *E15-Gadd45g* BAC is overexpressed at 10.5 dpc

WMISH was performed on 10.5dpc gonads from wild-type, *Gadd45g*$^{-/-}$ and *Gadd45g*$^{-/-}$ transgenic XY and XX embryos. Expression of *Gadd45g* was examined using a probe which targets the exon 2/3 boundary, a region missing in the knockout allele. At 10.5dpc (9 – 10 ts), *Gadd45g* is found in the gonads of both XY (A) and XX (B) embryos, but expression has not reach the poles. This expression is completely absent in the knockout embryos (C & D). Transgenic expression of *Gadd45g* from the *E15-Gadd45g* BAC (**E** & **F**) was found throughout the gonad and stain appeared at a higher intensity in all observed gonads. White arrows indicate the poles of the gonad.

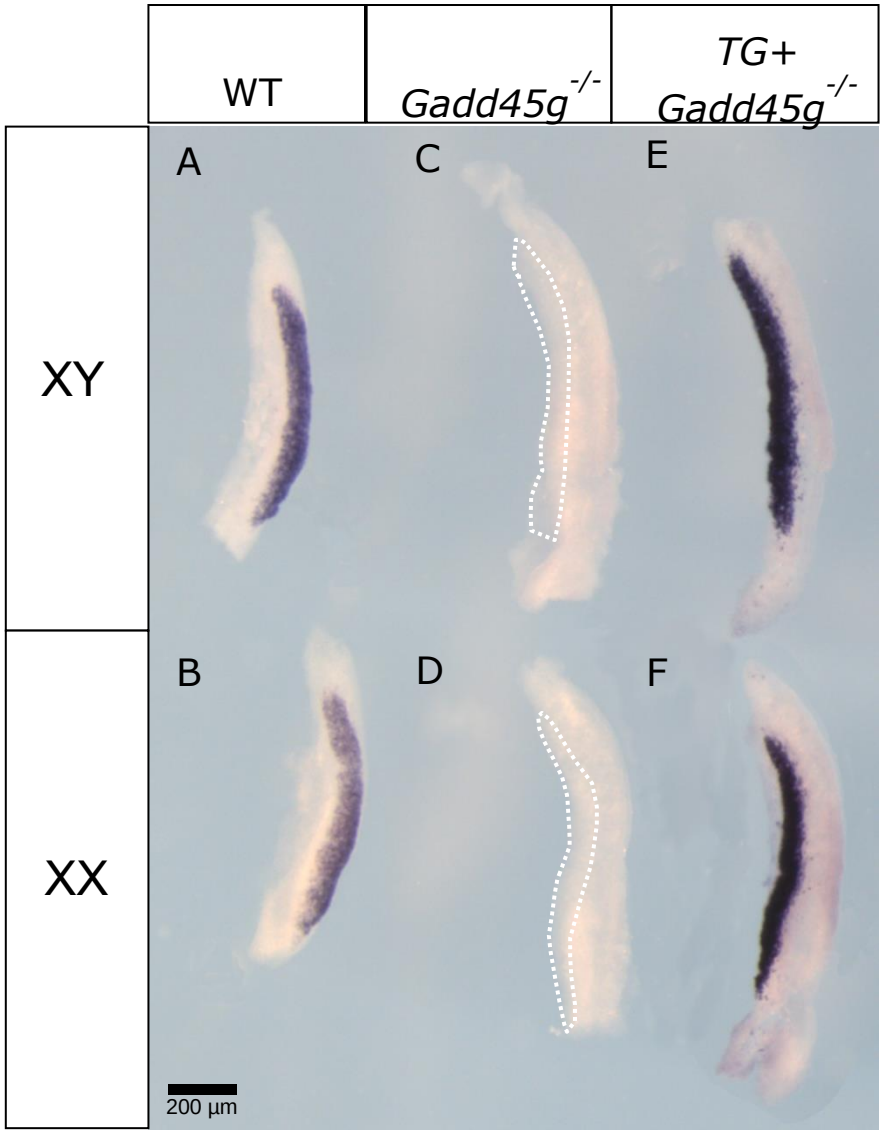


Figure 3.8 Expression of *Gadd45g* from the *E15-Gadd45g* BAC is limited to the body of gonad at 11.5dpc.
WMISH was performed on 11.5dpc gonads from wild-type, *Gadd45g*$^{-/-}$ and *Gadd45g*$^{-/-}$ transgenic XY and XX embryos. At 11.5dpc, *Gadd45g* is found in the gonads of both XY (A) and XX (B) embryos. This expression is completely absent in the knockout embryos (C & D). Transgenic expression of *Gadd45g* from the *E15-Gadd45g* BAC was found to be limited to similar tissues to the endogenous expression, though at a much higher intensity. Limited amounts of expression seen at the gonad mesonephros boundary, but appears to mimic the spatial expression profile of endogenous *Gadd45g* at this stage.

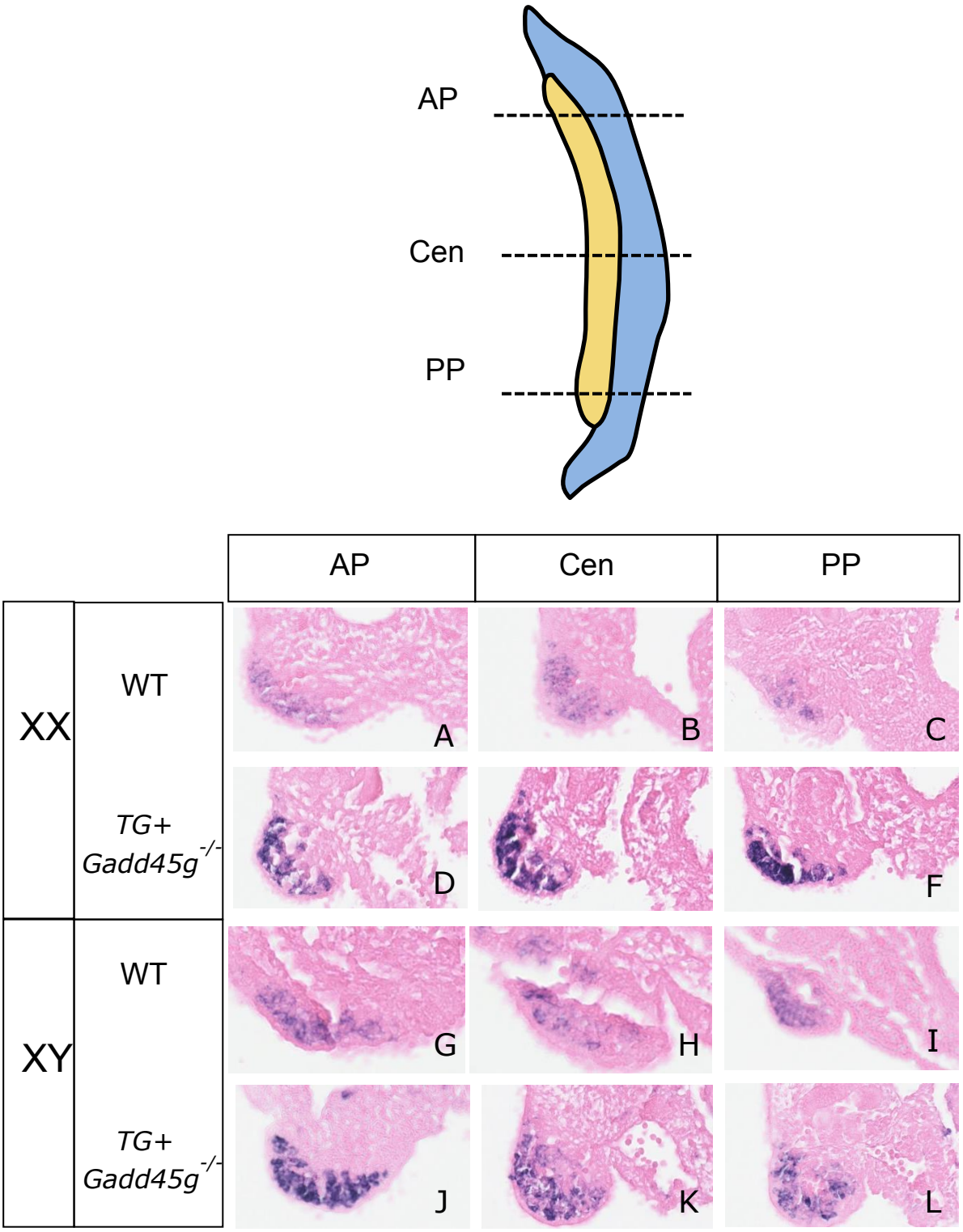


Figure 3.9 *Gadd45g* expression from E15-Gadd45g is not found in the coelomic epithelium
WMISH of UGRs from from wild-type, *Gadd45g*$^{-/-}$ and *Gadd45g*$^{-/-}$ transgenic XY and XX embryos were sectioned and stained with eosin. Sections from the anterior pole (**AP**), centre (**Cen**) and posterior pole (**PP**) of UGRs were photographed. *Gadd45g* expression is detected in the main body of the gonad in wild-type XX and XY embryos (**A** - **C** and **G** - **I**). Expression from the *E15-Gadd45g* BAC was found at a higher intensity in the gonad and not found in the coelomic epithelium in both XX and XY gonads (**D – F** and **J – L**).

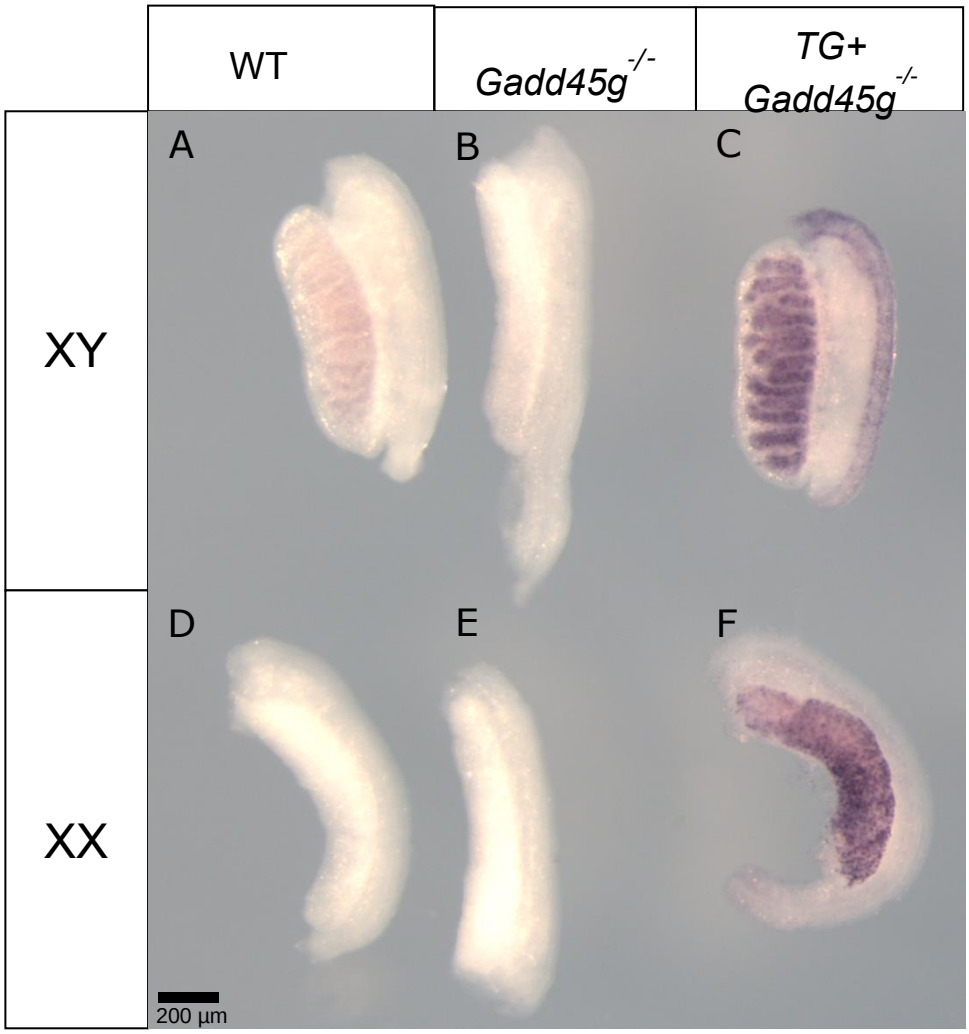


Figure 3.10 Expression of endogenous *Gadd45g* has ceased at 13.5 dpc but continues from the *E15-Gadd45g* BAC. WMISH was performed on 13.5dpc gonads from wild-type, *Gadd45g*-/- and *Gadd45g*-/- transgenic XY and XX embryos. At 13.5dpc, *Gadd45g* expression has subsided in C57BL/6 gonads of both XY and XX embryos (**A** & **D**). In the absence of *Gadd45g*, XY gonads develop as ovaries and show a similar morphology to XX gonads, which show ovarian development despite the lack of *Gadd45g* (**B** & **E**). The *E15-Gadd45g* transgene is still active during this period (**C** & **F**). The transgene is sufficient to rescue the sex reversal phenotype of *Gadd45g*$^{-/-}$. XY embryos produce gonads with a similar morphology to wild-type testis and testis cords are present (**C**). Expression of transgenic *Gadd45g* is found in the cords of XY gonads, whereas expression is found throughout the XX gonad.

Embryos were dissected and the morphology of their gonads was examined. B6.Y^{POS} gonads can show a range of sex reversal phenotypes at 14.5 dpc, from Xy ovary development to the formation of ovotestes. The majority of transgenic B6.Y^{POS} gonads appeared as testes at 14.5 dpc, with a number developing at least one ovotestis consisting of predominantly testicular tissue (3/7). This result was confirmed with WMISH using a probe targeting *Sox9* and *Stra8* (Fig 3.11). Limited *Sox9* expression was found in B6.Y^{POS} gonads and testis cords were not visible. Some B6.Y^{POS} gonads carrying the transgene showed similar *Sox9* expression levels as wild-type C57BL/6 gonads. Gonads which had ovotesticular morphology did not show *Sox9* specific staining at the poles (Fig 3.11E). *Stra8* expression was observed throughout XX gonads and at a slightly lower intensity in B6.Y^{POS} gonads. Though absent in XY^{B6} gonads, limited *Stra8* expression was observed in the poles of transgenic B6.Y^{POS} gonads. Though the extent of rescue shows some variability, a largely testicular phenotype has been restored to the B6.Y^{POS} gonads carrying *E15-Gadd45g*. These results suggest that the *E15-Gadd45g* BAC has rescued the sex reversal phenotype of B6.Y^{POS}.

3.3 Conclusions and Discussion

The beginning of this chapter outlined the aims of finding what level of over-expression was induced in the transgenic gonad; whether this transgenic expression shared a similar spatiotemporal profile to endogenous *Gadd45g* and whether the overexpression could rescue the sex reversal phenotype of B6.Y^{DOM}. The expression and spatiotemporal profile of *Gadd45g* was restricted to a few stages of development and specifically within the gonad. The *E15-Gadd45g* transgene was shown to successfully rescue the sex reversal found in B6.Y^{POS}. However, other models of sex reversal can be investigated to find the mechanisms through which *Gadd45g* overexpression rescues this phenomenon

3.3.1 E15-Gadd45g induces overexpression of *Gadd45g* in the gonads

Increased *Gadd45g* expression was detected at 18 ts within both XX and XY gonads. However, this approximately 5-fold increase in mRNA levels did not lead to increased levels of *Sry* expression (Fig 3.5). While t-tests suggested a significant difference between the expression levels of Gadd45g between transgenic and wildtype gonads, analysis of the statistical power shows that there is a relatively high risk of type 2 errors being committed with the number of samples used (statistical power =0.7284 and 0.7095 for comparisons between the XY and XX transgenics versus wild-types respectively) (HyLown Consulting LLC, 2013). As a result further qRT-PCR analysis will be required to properly calculate the level of overexpression. This suggests that *Sry* expression at this time in B6 XY gonads is maximal. As both *Gadd45g* and *Sry* show a dynamic expression profile which peaks at 18 ts, investigating the levels of expression over a time course would reveal any other potential effects of *Gadd45g* overexpression on *Sry*. Such effects could include prolonging the window in which *Sry* is expressed or a change in the relative level of expression of *Gadd45g*. However, in this context fertile adult XX and XY were readily identified, suggesting no significant disruption to gonad function by the transgene.

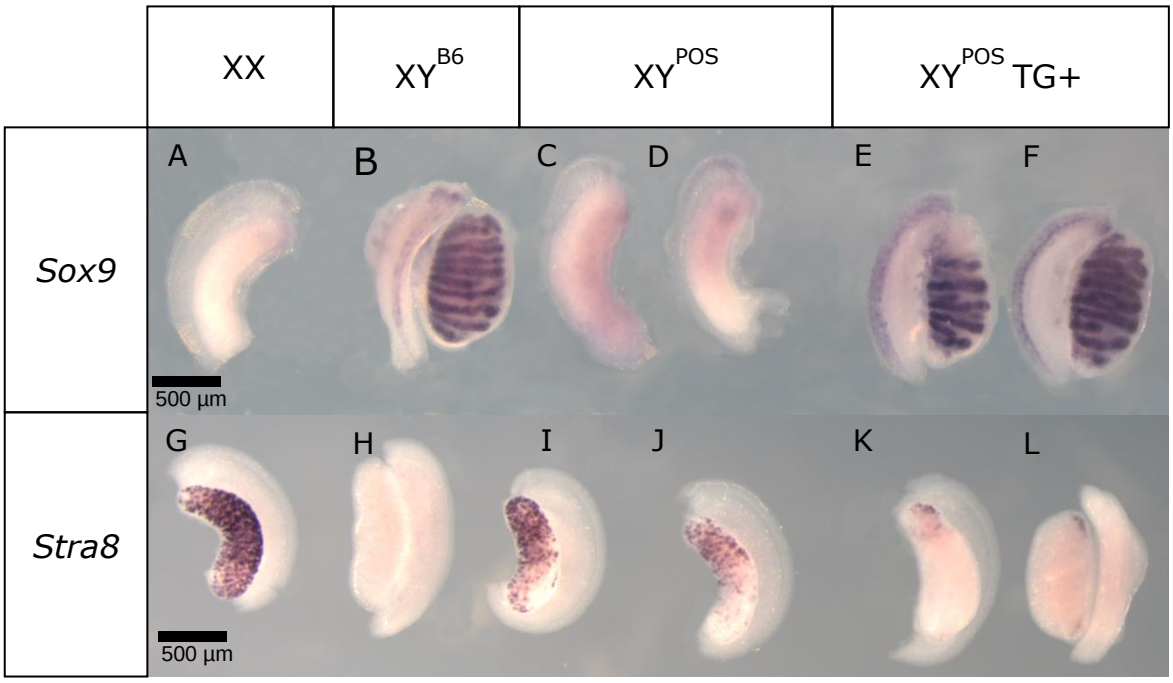


Figure 3.11 *E15-Gadd45g* rescues testis development in B6.Y^{POS}.
Gonads dissected from the indicated strains were used in WMISH using probes targeting *Sox9* (**A** – **F**) and *Stra8* (**G** - **L**). *Sox9* expression is observed in testicular tissues, which is nearly absent in B6.Y^{POS} gonads. This expression is restored in gonads containing *E15-Gadd45g*. Expression of *Stra8* is seen in meiotic germ cells, which are present in ovarian tissue at this time. *Stra8* is seen in B6.Y^{POS} gonads, but is limited to small segments of the gonadal poles in gonads carrying *E15-Gadd45g.*

3.3.2 *E15-Gadd45g* shares a similar spatial expression pattern to endogenous *Gadd45g* but not a similar temporal profile

In wild-type B6 gonads *Gadd45g* expression is first detected at 10.5 dpc at the 9 ts stage in the centre of the gonad. By 11 ts expression has reached the poles of the gonads, where it continues until approximately 27 ts, when expression begins to fall in the centre of the gonad. By 13.5 dpc, *Gadd45g* expression can no longer be detected. Throughout this period *Gadd45g* expression is limited to the main body of the gonad and not found in the mesonephros or the coelomic epithelium (Warr et al., 2012). *E15-Gadd45g* was found to produce higher levels of expression at 10.5 dpc (9 – 10 ts) compared to both XX and XY non-transgenic C57BL/6 gonads. Expression from the *E15-Gadd45g* BAC was found to be expressed only within the main body of the gonad at 11.5 dpc (18 ts) though also at much higher levels. Expression from the BAC continued 13.5 dpc. Further investigation is needed to determine when expression from the BAC begins and ceases. This will require further WMISH experiments at other time points earlier and later than those presented here.

Interestingly the expression of *Gadd45g* from the BAC is specific to the testis cords of 13.5 dpc XY transgenic gonads. As it has previously been established that *Gadd45g* is not expressed in germ cells (Warr et al., 2012), then the expression within the testis cords may indicate that GADD45γ is exclusive to Sertoli cells. If this expression pattern extends to the supporting cell lineage in XX gonads and before the onset of *Sry* expression, then *E15-Gadd45g* could prove to be a useful tool for generating reporter constructs for the sex determination field. This idea is further elaborated on in chapter 4.

An *in vivo Gadd45g* overexpression construct could have multiple applications across other fields which study *Gadd45g*. Experiments conducted *in vitro* showing the effects of *Gadd45g*

overexpression on differentiation (Kaufmann and Niehrs, 2011; Thalheimer et al., 2014), tumorigenesis (Zerbini and Libermann, 2005), DNA repair (Rai et al., 2008) and active DNA demethylation (Niehrs and Schafer, 2012) could be elaborated upon *in vivo* using transgenic mice. Further WMISH analyses on *Gadd45g*$^{-/-}$ transgenic embryos will be required to establish non-gonadal sites of expression.

3.3.3 E15-Gadd45g rescues testis formation in B6.Y^{POS} mice

At the time of writing, three B6.Y^{POS} animals carrying the *E15-Gadd45g* have been allowed to develop to adulthood. These individuals have developed as phenotypic males and have successfully mated and produced embryos, indicating they are fertile. Gonads from some transgenic B6.Y^{POS} embryos were observed to contain ovarian tissue. It is therefore imperative to examine the phenotypic state of the adults as ovotestes may resolve to form testis in adults, as is the case with B6.Y^{AKR} animals (Nagamine et al., 1987).

The molecular mechanisms underlying this phenomenon will need to be investigated further. *Sry* expression is delayed in B6.Y^{POS} animals and is the likely cause of the sex reversal phenotype (Bullejos and Koopman, 2005). Previously, a *Map3k4* BAC was shown to restore *Sry* expression to normal levels within the critical time window and rescue T-associated sex reversal (Warr et al., 2014). This same BAC was later shown to also rescue B6.Y^{POS} (Pope, 2015). Given the interaction between *Gadd45g* and *Map3k4* (Warr and Greenfield, 2012), it is likely that a similar restoration of the C57BL/6 *Sry* expression profile to B6.Y^{POS} gonads is occurring. However, this will require confirmation through experimentation using WMISH and qrtPCR. Further studies will be required to establish the mechanism by which MAP3K4 and GADD45γ overexpression impacts on *Sry* expression. Unfortunately, time restraints prevented such experiments from being performed.

As *E15-Gadd45g* has been shown to rescue Y^{AKR} and Y^{POS} sex reversal it may have potential in rescuing other sex reversal phenotypes. One key experiment to conduct would be investigating the effects of *E15-Gadd45g* on *Igfr1*/*Insr* double knockout mice, which show XY sex reversal. Previous experiments have shown decreased levels of *Gadd45g* expression in the gonads of *Igfr1*/*Insr* double knockouts and it has been speculated that an overexpression of *Gadd45g* may be able to rescue the sex reversal phenotype (Pitetti et al., 2013).

Chapter 4: Introduction of an epitope tag into the GADD45γ coding sequence using BAC recombineering

4.1 Introduction

4.1.1 Background information

Gadd45g is expressed in the bipotential gonads of both XX and XY mice before the initiation of *Sry* expression. Though knowledge of its cell-specific expression is still not precisely clear, it is not found in the coelomic epithelium and does not show an overlap with the primordial germ cell (PGC) marker *Oct4* (Warr et al., 2012), suggesting it is expressed in gonadal somatic cells. Closer examination of cellular and subcellular distribution of GADD45γ protein has been stymied by the absence of a suitable working antibody that can detect endogenous protein.

As GADD45 proteins have no DNA-binding capability themselves, they are suspected to act as scaffolding proteins, facilitating the interactions between key enzymes and DNA (Niehrs and Schafer, 2012). Experiments which have detected interacting partners of the various GADD45 proteins have been conducted *in vitro*, using overexpression plasmids encoding tagged versions of GADD45 proteins (Azam et al., 2001; Schafer et al., 2013; Kienhofer et al., 2015). Investigating the *in vivo* interactions between GADD45γ and its various partner proteins in the developing gonad would illuminate its role in sex determination. Such experiments, however, require an antibody with a strong affinity towards its target protein. Previous efforts by this lab, and others, to generate an antibody towards endogenous GADD45γ have not produced a sufficiently effective antibody.

In order to further elucidate the function of GADD45γ in the developing testis, it would be beneficial to generate transgenic lines expressing epitope-tagged versions of the GADD45γ polypeptide or carrying fluorescent-reporters controlled by the *Gadd45g* regulatory sequences. Modification of BAC sequences can be achieved through homologous recombination mediated genetic engineering,

termed BAC recombineering (Copeland et al., 2001). As the E15 BAC transgenic line both rescues the phenotype of *Gadd45g* deficient gonads and mimics expression of endogenous *Gadd45g* in the gonad, it likely carries all of the regulatory sequences required for gonadal *Gadd45g* expression, making it an ideal candidate for generating a transgenic construct.

Recombineering utilises the Red/ET recombination system of the lambda bacteriophage to successfully modify large DNA molecules in *E. coli* cells (Wang et al., 2009). Linear constructs flanked by 42 nucleotides of homologous sequence can be transformed into circular DNA through the use of the Rac prophage encoded proteins RecE and RecT (Zhang et al., 1998) In *E. coli*, homologous recombination is mediated by the ATPase RecA. However, proteins from lambda phage can act independently of RecA to successfully initiate double-strand breaks (DSBs) and repairs (Muyrers et al., 2000). Linear double-stranded DNA (dsDNA) is unstable in *E. coli* due to the presence of RecBCD, a ATP-dependent linear dsDNA exonuclease which targets double strand breaks and recruits RecA for DNA repair (Copeland et al., 2001). The rate of recombination was further improved by replacing *recE* and *recT* with their orthologues from the lambda phage *redα* and *redβ*, and by including the RecBCD inhibitor *redγ.*

RecE and Redα (λ exonuclease) are 5’ – 3’ exonucleases, which in the event of a DSB generates a 3’ single-stranded overhang from dsDNA. RecT or Redβ then binds to and coats the single-stranded DNA generated by the respective exonuclease. This protein nucleic acid complex then associates with homologous regions of DNA. A DSB can then be repaired through the pairing of complementary protein-DNA complexes. In the event of homologous recombination, the protein DNA complex invades an unbroken homologous region generating a D-loop. This then acts as a priming event for the replication machinery to substitute in the recombination target (Fig 4.1). Redγ is a short

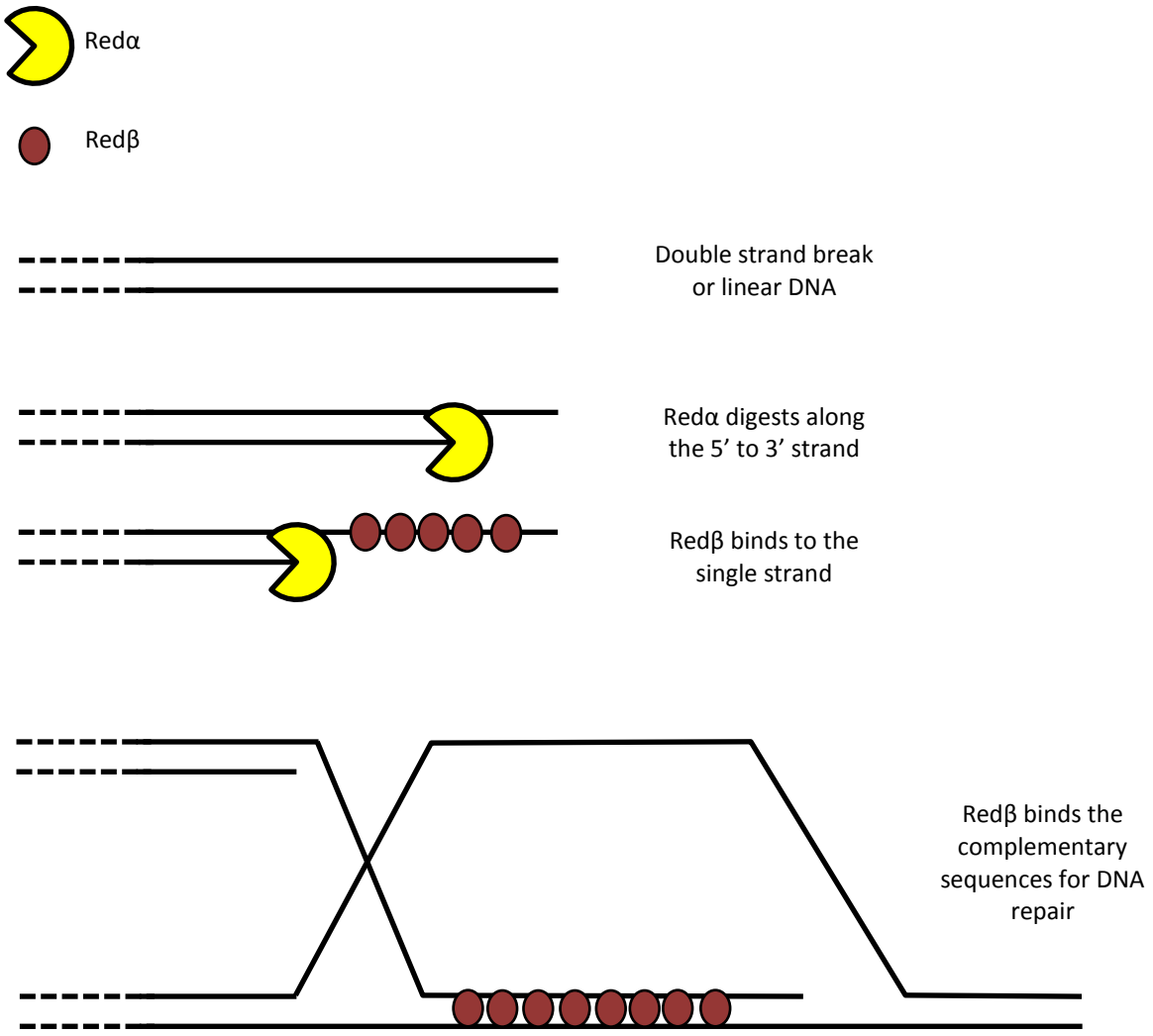


Figure 4.1 Overview of Red mediated recombination
Redα has a 5’ to 3’ dsDNA exonuclease activity, which generated 3’ overhangs on linear dsDNA. The overhang is subsequently bound by Redβ which then binds to the complementary sequence found in the target construct. The D-loop formed by these process then acts as a target for the DNA repair machinery to integrate the linear construct.

polypeptide, which when dimerised with itself inhibits the activity of the *E. coli* RecBCD complex, protecting the linear DNA targeting cassette from degradation (Zhang et al., 1998).

In this particular study the plasmid pRed/ET (GeneBridges) was electroporated into the *E. coli* containing the *E15-Gadd45g* BAC. pRed/ET contains the λ phage operon *redγβα* expressed under the control of the inducible pBAD promoter, which confers tetracycline resistance. The gene *araC*, encoded on the reverse strand, is used to regulate the pBAD promoter both positively and negatively. The AraC protein binds to L-arabinose leading to the upregulation of pBAD. In the absence of L-arabinose, AraC instead binds to the pBAD promoter, inhibiting expression (Schleif, 2010). Precise and limited expression of the Red proteins is required to prevent both the toxic effects from continuous expression of *redλ* as well as unwanted recombination events. The pRed/ET plasmid is a derivative of pSC101. RepA, a protein encoded on pSC101, is required for the proper replication and partitioning of the plasmid during replication. However, RepA is temperature sensitive and cells are required to be grown at 30 ˚C to maintain pSC101 or its derivatives. Growth at 37 - 42 ˚C will lead to an 80% loss of the plasmid after four generations of bacterial growth (Miller et al., 1995). Sustained growth at 37 ˚C and lack of tetracycline selection will lead to the progressive loss of pRed/ET, leaving the host cell as it was before the transformation.

4.1.2 Aims and hypotheses

Given the lack of success in generating a strong *Gadd45g* antibody, both within the Greenfield group and other laboratories, I aimed to produce an edited version of the *E15-Gadd45g* transgene containing the sequence for an epitope tag using RedET mediated recombineering. Using this new construct, I hope to generate a transgenic mouse line which will allow me to detect GADD45γ protein in embryonic tissues, and allow me to confirm the cellular populations to which *Gadd45g* is expressed within the gonad. If detection by the epitope tag antibody is strong enough, I aim to

attempt proteomic experiments such as immunoprecipitation and chromosome immunoprecipitation to find the interacting partners and target sequences for GADD45γ within the developing gonad.

If RedET mediated recombineering proves to be successful, then other types of transgene such as fluorescent reporters or inducible Cre exonucleases can be considered to be edited into the E15-*Gadd45g* BAC. This will allow me to further try and elucidate the mechanism of action for GADD45γ during sex determination.

4.2 Results

4.2.1 Generation of a BAC containing a modified *Gadd45g* gene with a C-terminal His/FLAG epitope tag sequence

Epitope tags are small sequence motifs, meaning that the corresponding DNA sequences are typically >100bp, and are non-selectable by antibiotic means. Recombineering was therefore achieved through a two-step process using the *rspL*$^{+}$/*neo* selection method, see section 2.5 for methodologies. In the first round, a selection cassette is inserted into the target region of the gene. Cells that have successfully integrated the selection cassette are selected for using the respective antibiotics. This is followed by a second round of recombinase induction and the replacement of the selection cassette with the epitope tag sequence (Fig 4.2). This selection is facilitated by the removal of the *rspL*$^{+}$ sequence, which encodes the bacterial S12 ribosomal sub-unit, the target of streptomycin. In many strains of *E. coli*, mutations in the chromosomal copy of the gene confer streptomycin resistance. However, with the addition of *rspL*$^{+}$ this resistance is lost (Wang et al., 2009).

The supplied *rpsL*/*neo* selection cassette sequence was flanked by 50bp of homology surrounding the translational stop codon of the *Gadd45g* open reading frame (ORF).

Bacterial cultures containing the *E15-Gadd45g* BAC were first tested for streptomycin resistance. Fresh cultures of *E15-Gadd45g* containing *E. coli* were grown in overnight with chloramphenicol and streptomycin. The next day, colonies were picked and genotyped with primers flanking the homologous sequence (Fig4.3). In the unmodified BAC, this assay produced a single 200bp band, which would increase to 1.5kb with the addition of the selection cassette. Colonies which produced a single 1.5kb band were grown overnight in chloramphenicol and kanamycin and prepared the following day. Each colony was tested for streptomycin sensitivity by overnight growth on LB two agar plates containing chloramphenicol and kanamycin or chloramphenicol and streptomycin. One colony was found to successfully integrate the *rpsL*/*neo* cassette in the correct position. This colony also showed the highest sensitivity to streptomycin. Designated *pRed/ET-E15-Gadd45g-C-rpsL*/*neo*, this colony was then prepared for the second round of recombineering (Fig 4.3).

A combination epitope tag containing the sequences for a combination poly-His/FLAG tag was selected to be recombined into the *E15-Gadd45g* BAC. These particular epitopes were chosen because there are corresponding highly specific and commonly used antibodies that have previously been used in numerous biochemical applications, including immunoprecipitation experiments (Einhauer and Jungbauer, 2001; Nishi et al., 2002; Ohkouchi et al., 2013). The sequence for the His-FLAG tag flanked by the 50bp homology arms was generated by two 90bp overlapping primers G45gC FLAG F and G45g C FLAG R.

A fresh overnight culture of *pRed/ET-E15-Gadd45g-C-rpsL*/*neo* was made and used in the second round of recombineering as describe in section 2.5.2. The next day colonies were tested using primers flanking the homology arm sequence. Multiple 96 colony screens were carried out, however out of 564 colonies picked; only one had integrated the epitope sequence and was later found to

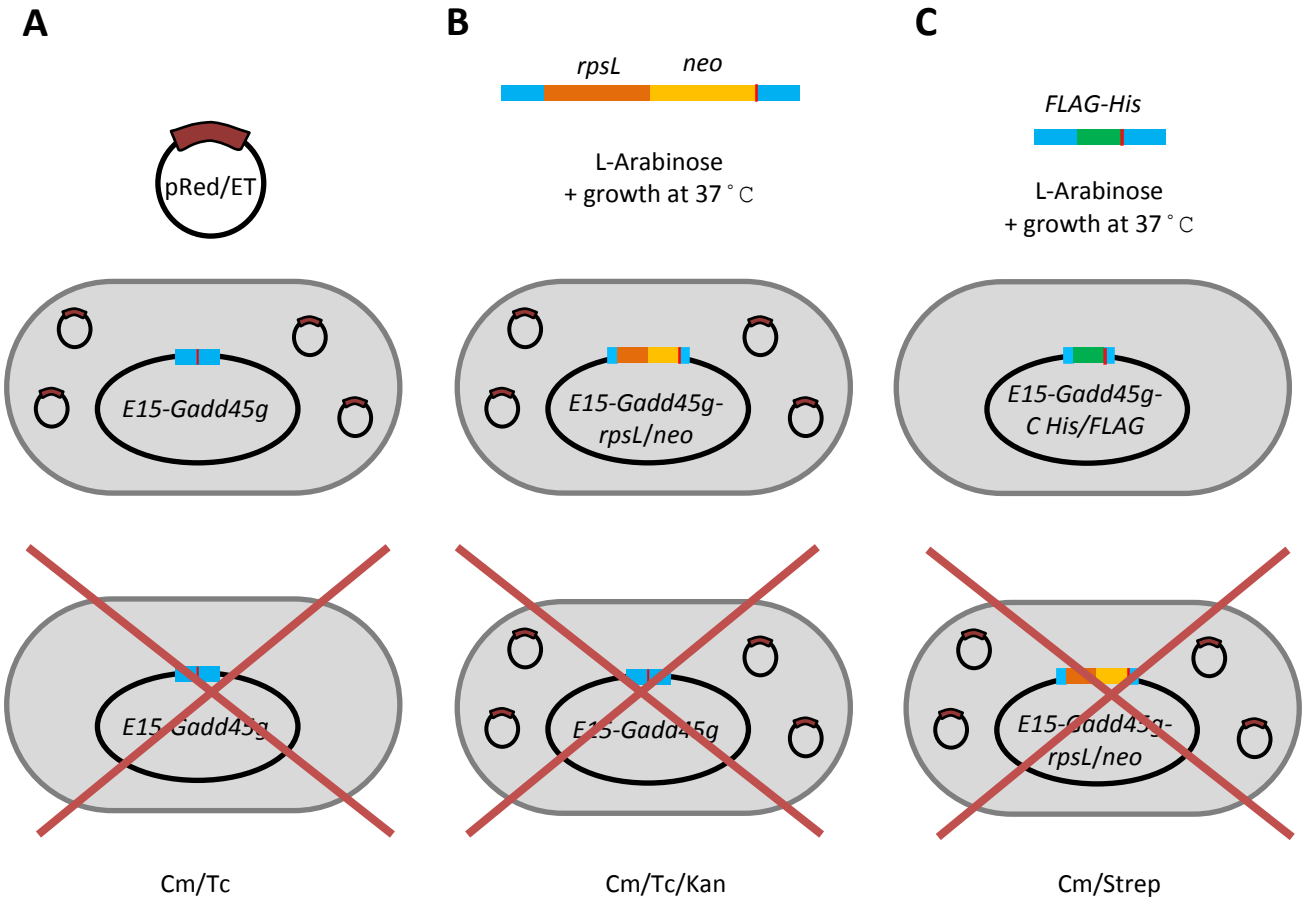


Figure 4.2 Overview of Red mediated recombination

Strategy for integration of *His/FLAG* tag sequence into the coding region of *Gadd45g* within the *E15-Gadd45g* BAC. The plasmid pRED/ET is transformed into the *E. Coli* cells carrying the *E15-Gadd45g* BAC. Selection is then carried out with plates containing chloramphenicol and 3µg/ml tetracycline. **B**) After selection, the RED recombinases are induced. Cells were then transformed with linear *rpsL-neo* flanked by homology arms targeting the region around the *Gadd45g* translational stop codon. Integration for the selection cassette was selected for with kanamycin, with the chloramphenicol and tetracycline as before. **C**) The selection cassette is then replaced by the sequence for the *His/FLAG* tag by inducing the RED recombinases followed transformation with the linear *His/FLAG* sequence flanked by the same homology arms as before. Cells lacking the selection cassette are then selected for with streptomycin and chloramphenicol.

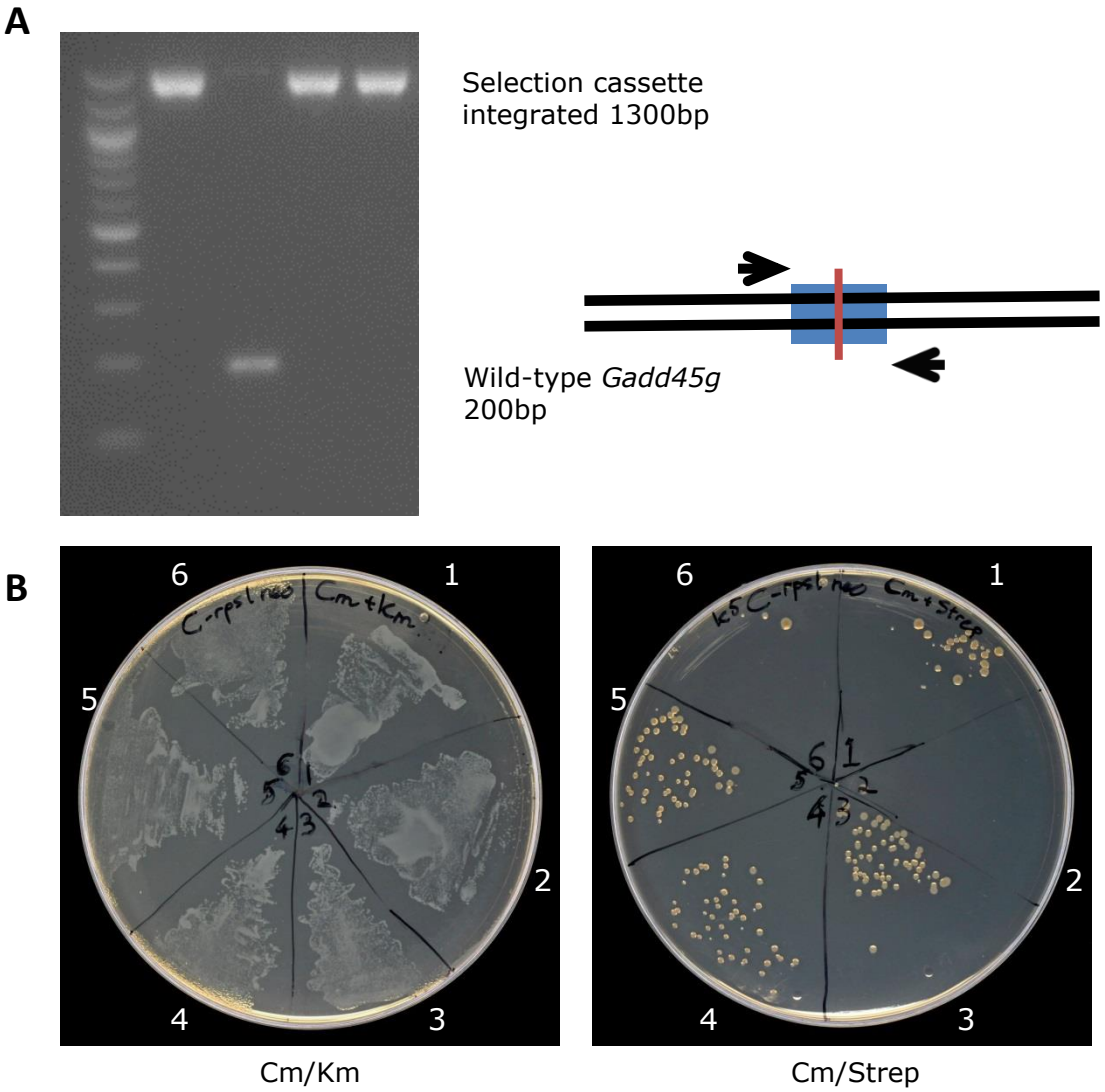


Figure 4.3 Integration of the rpsl-neo selection cassette
A) Colonies from the Cm/Tc/Km selection process are screened with a PCR assay amplifying a region outside of the homology arms. A single band at 1.3kb shows the integration of the *rpsL-neo* selection cassette in the correct location. B) Colonies are then screened with antibiotic resistance. With the integration of the selection cassette, the cells gain a sensitivity to streptomycin. In this particular screen colony 2 shows resistance to Cm/Km but not to Cm/Strep. Colony 2 was then taken forward for use in the second step of BAC recombineering.

have excised other regions of the BAC in the process. As a result, it was decided that the homology arms would be extended to 500bp, which has been shown to improve the efficiency or Red/ET mediated recombination (Westenberg et al., 2010). The *His/FLAG* sequence flanked by 500bp homology arms was generated by overlap PCR, using the original 50bp homology arm flanked sequence and 500bp up- and down-stream from the *Gadd45g* translational stop codon (Fig 4.4). As before, fresh cultures of *pRed/ET-E15-Gadd45g-C-rpsL/neo* were made and the *Red* recombinases were induced with L-arabinose and growth at 37°C. Cells were then transformed with the 500bp flanked *His/FLAG* construct and plated onto LB agar plates containing chloramphenicol and streptomycin.

The following day, colonies were picked and genotyped with a new assay which amplified from within the expanded homology regions and outside (Fig 4.5). Colonies showing a single band of 650bp were then grown in 5ml LB with chloramphenicol and streptomycin overnight. 1µl of culture was then used as a template with the original genotyping primers, which targeted around the 50bp homology arms. The smaller amplicon could then be compared to the amplicon from the unmodified BAC. A 45bp increase in size was observed, indicating the sequence for the *His/FLAG* tag had integrated. Cultures were then prepared and sent for Sanger sequencing. Sequencing results revealed that one colony, *E15 C-FLAG* 6A, had successfully integrated the His/FLAG sequence immediately 5' of the *Gadd45g* stop codon. Additional bases had neither been added nor lost at the border of the homologous regions (Fig 4.5 and 4.6).

The expression of the *Red* genes may have led to other recombination events occurring, compromising the integrity of the BAC and potentially removing or rearranging key regulatory elements. To establish that no such events had occurred, a series of restriction digest fingerprints

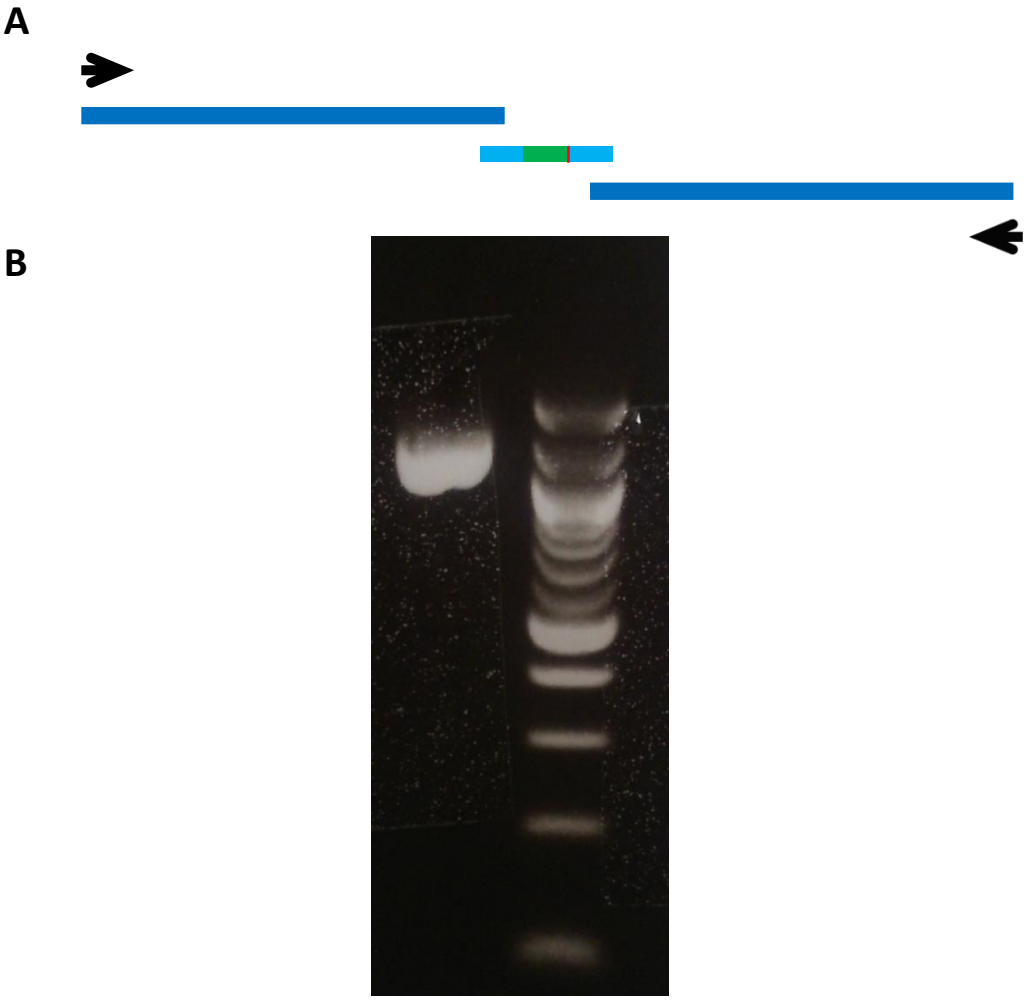


Figure 4.4 Generation of 500bp homology arm flanked *His/FLAG* sequence
A) 450bp regions either side of the 50bp homology arms were amplified by PCR. The forward primer used to generate the 5' homology arm and the reverse primer from the 3' homology arm were then used in an overlap PCR to produce the 500bp homology arm flanked version of the *His/FLAG* sequence. **B**) Products were run on a 2% agarose gel and found to be the correct size of 1kb.

were produced from the unmodified BAC, the *rpsL/neo* containing BAC and *E15 C*-His/FLAG 6A. An *in silico* digest on the *E15* BAC sequence was performed using restriction mapper (Blaiklock, 2009) to find a suitable restriction enzyme. Digestion with *ScaI* was found to produce 38 fragments ranging from 18kb to 55bp, with the site of modification within a 4.3kb band. 15µg of BAC DNA from *E15*, *E15-C-rpsL/neo* and *E15 C-His/FLAG* were digested overnight with *ScaI*. Samples were then run on a 0.8% ethidium bromide gel for 17 hours at 40v. The resulting gel produced almost identical fingerprints from each of the BACs, indicating that no other major recombination events had occurred (Fig 4.7). A 4.3kb band present in the unmodified and epitope sequence modified BAC was not present in the *rpsL/neo* containing BAC. Instead an extra 5.6kb band was visible in the fingerprint from *E15-C-rpsL/neo.* Given that the *rpsL/neo* cassette is 1.3kb long, this extra band likely contains the *Gadd45g* translational stop codon with the 5' inserted *rpsL/neo* cassette. After the second round of recombineering and the replacement of the selection cassette with the epitope tag sequence, this site produced a similar sized fragment as the unmodified BAC. Having confirmed the integrity of the BAC and the successful modification of the BAC sequence to produce a C-terminal His/FLAG tagged version of GADD45γ, the BAC was prepared for microinjection.

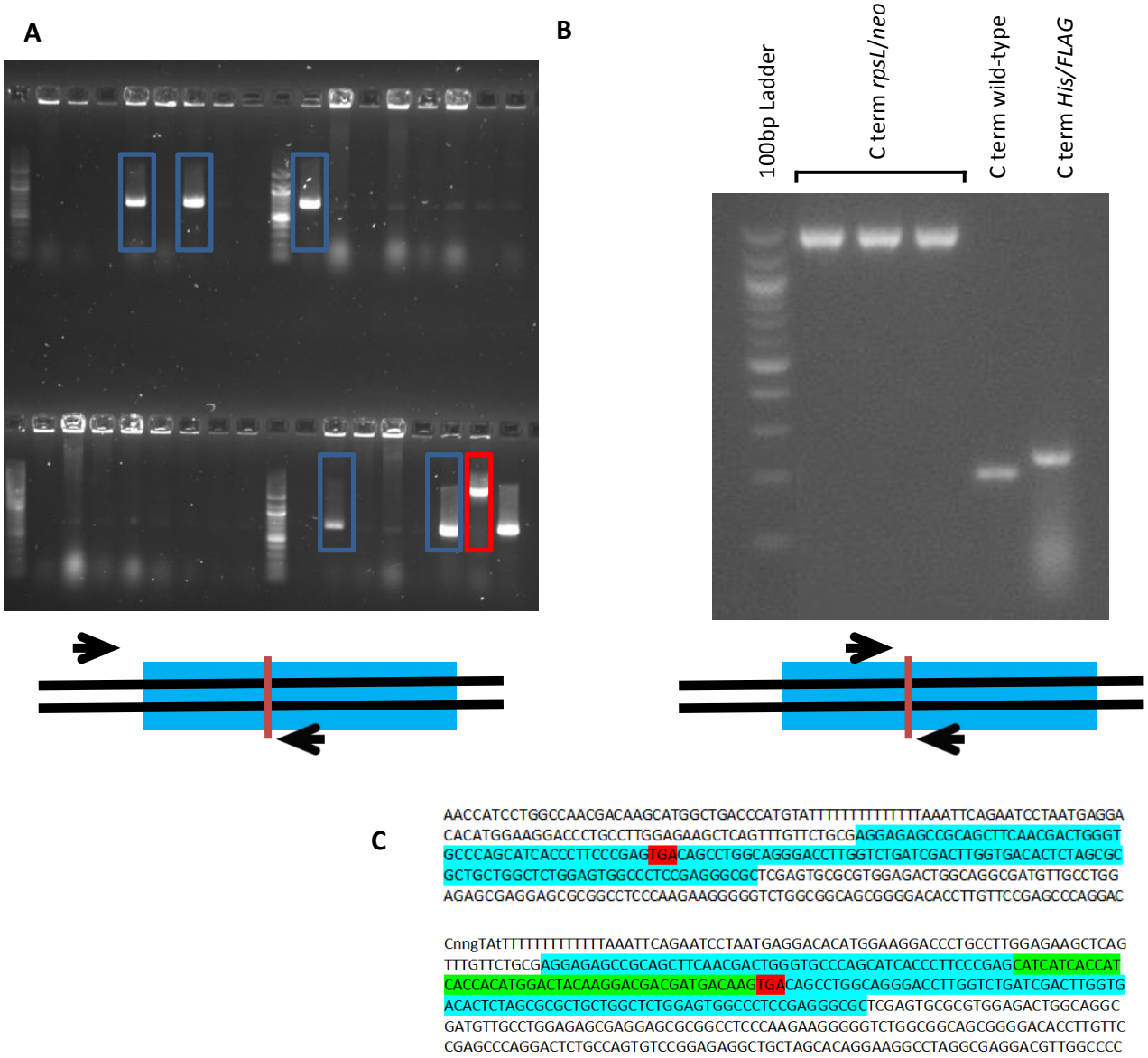


Figure 4.5 Confirmation of the integration of the *His/FLAG* sequence prior to the translational stop codon of *Gadd45g*

A) Integration of the His/FLAG sequence was confirmed with a PCR assay with one primer annealing to an area outside of the homology arms. With the removal of the *rpsL-neo* cassette, the resulting band decreases in size from 1.8kb (band in red rectangle) to 600bp (bands in blue rectangle). Colonies which produced a single band of 600bp were taken for further analysis, as this . **B**) A second PCR assay targeting the region around the 50bp homology arms revealed the integration of the His/FLAG sequence. The integrated *His/FLAG* sequence is 45bp larger than the endogenous *Gadd45g* sequence, which can be observed with the products of this assay. **C**) Sequencing around the 50bp homology arms confirms the presence of the *His/FLAG* sequence.

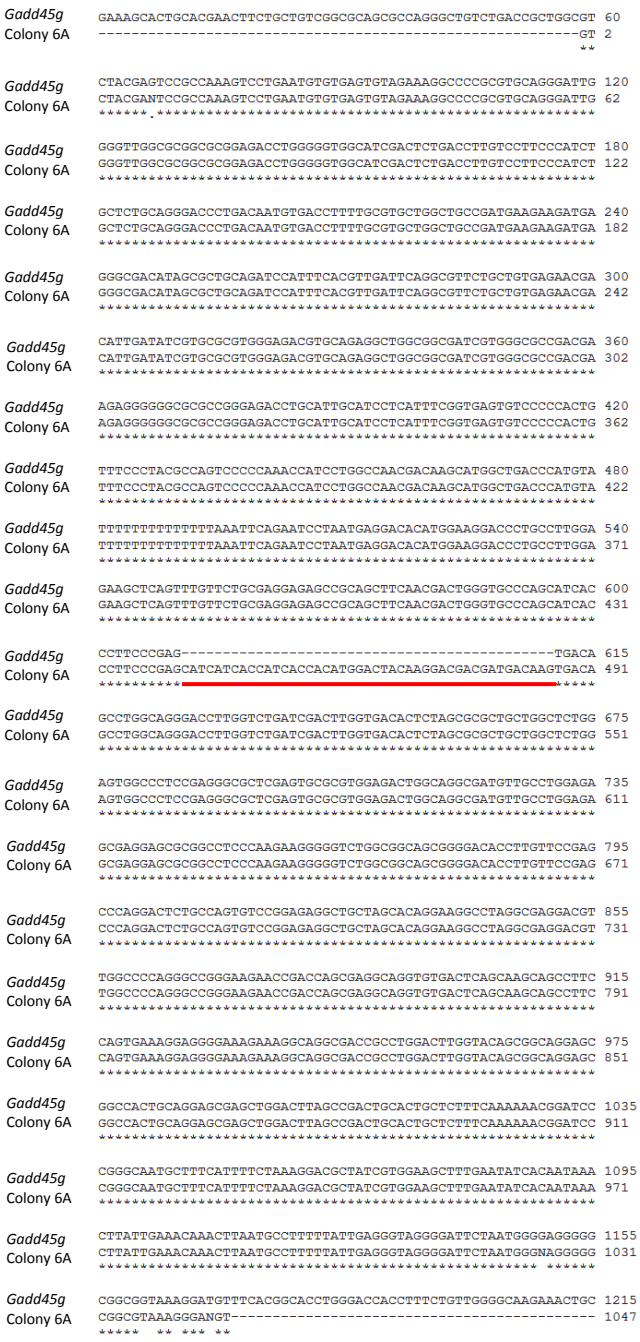


Figure 4.6 Sequencing confirmation of integration of 500bp homology arms and the *His/FLAG* sequence

Alignment of the *His/FLAG* containing colony 6A with the endogenous *Gadd45g* sequence. Sequences appear to align 500bp either side of the site of modification. The 45bp consisting of the *His/FLAG* sequence, underlined in red.

4.2.2 Generation of transgenic mice using modified BAC lines

The bacterial line containing the modified BAC was sent to the MRC Harwell transgenics team for microinjection. 500ml cultures of *E15 C-His/FLAG* were made and prepared. Purified BACs were then microinjected into 1-cell C57BL/6J (B6) embryos (Gardiner and Teboul, 2009). Across 5 sessions, 347 embryos were injected, leading to the birth of 46 pups. Once they had reached weaning stage, pups were genotyped for the presence of the chloramphenicol resistance gene on the pBeloBAC11 backbone. A single male transgenic founder (3213.1h) was found to be *Cmr* positive. This founder was then mated with several B6 females to test if the transgene would be transmitted to the next generation.

From this mating, 9 litters were sired which produced 62 pups that survived to weaning. These pups were then genotyped for presence of the *Cmr* gene. However, all were negative. These results were confirmed with qPCR targeting the *Cmr* gene normalized against *Dot1l*. Due to the lack of transgenic offspring the original founder was then subjected to droplet digital PCR (ddPCR), to reassess his transgenic status. In this process, the components of a qPCR reaction are suspended in an oil emulsion, generating approximately 20,000 individual reactions. The droplets are placed in a well and the PCR reaction is run to the amplification end-point (40 cycles). After the reaction is complete, droplets are processed into a stream and run single file passed a fluorescence detector. Droplets which contain the probe specific template will show fluorescence due to the release and cleavage of the probe. The fraction of positive droplets to negative is then used to calculate the concentration of the target sequence. A Poisson distribution is then applied to extend the range of the concentration to account for varying levels of template across droplets. The reaction targeting the *Cmr* gene is run in conjunction with an assay for *Dot1l.* As *Dot1L* is an autosomal gene, there will usually be two

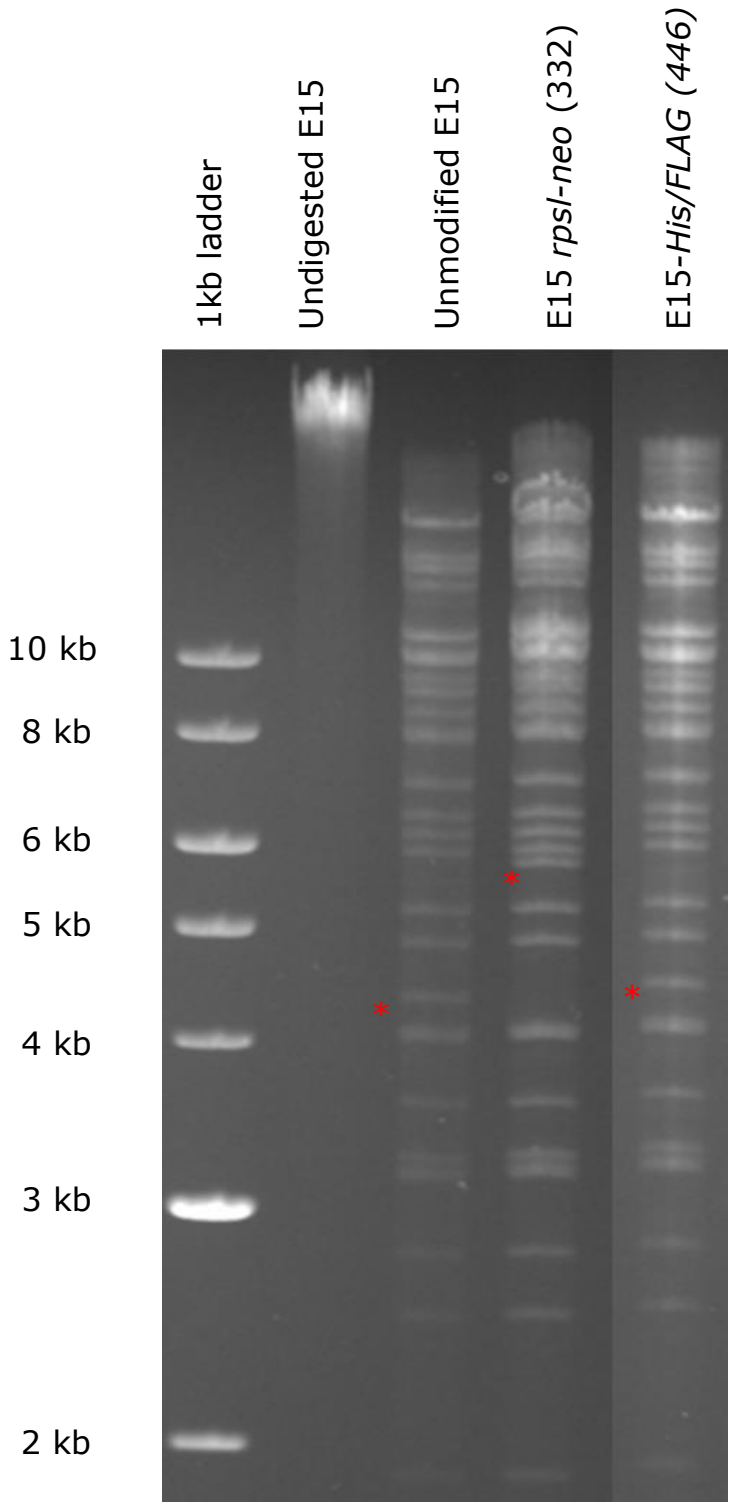


Figure 4.7 *ScaI* digest of unmodified *E15-Gadd45g*, *E15-rpsl-neo* and *R15-Gadd45g-His/FLAG* have no other major recombination events

Samples of the unmodified *E15-Gadd45g*, *E15-Gadd45g-rpsl-neo* and the modiifed *E15-Gadd45g-His/FLAG* BACs were digested overnight with *ScaI*. Digests were then run on a 0.8% agarose gel. The resulting fingerprints show only a single band difference, highlighted with a red asterisk. According to an *in silico* digest a 4.3kb band contains the site targeted for modification. With the addition of the 1.3kb selection cassette, this band increases to 5.6kb. In the second round of recombineering, the selection cassette is replaced by a 45bp sequence for the His/FLAG epitope. The lack of a 5.3kb band and the reappearance of a 4.3kb band in the modified BAC fingerprint suggests this is the site of modification and that no other recombination events have occurred.

copies per cell. By normalizing the concentration of *Cmr* against that of *Dot1l*, a highly accurate estimate of the number of copies per cell of a target gene can be generated (Hindson et al., 2011).

Results of the ddPCR on an ear biopsy from 3113.1h showed that it had a *Cmr* copy number per cell of 0.064 (Fig 4.8). This result was significantly higher than a wild-type litter mate, which produced a copy number of 0.0151 per cell. However when coupled with the lack of transgenic offspring, these data are consistent with the founder being a mosaic; and it is unlikely that the BAC had integrated into its germ line. Sperm was collected from 3213.1h and stored in the MRC Harwell FESA facility, should there be a potential use in the future for this line.

4.3 Conclusions and Discussion

4.3.1 Aims revisited

In this series of experiments I aimed to produce a transgenic mouse line carrying a modified version of the *Gadd45g* coding sequence of the *E15-Gadd45g* to incorporate the sequences for a combination *His*/*FLAG* epitope tag. While I have successfully managed to produce the modified BAC, a transgenic line had not been produced at the time of writing. As a result my other aims for this project could not be met due to the lack of a transgenic line.

4.3.2 Discussion

As shown previously in this thesis, *E15-Gadd45g* is a functional BAC capable of rescuing the gonadal sex reversal phenotype of *Gadd45g*$^{-/-}$, *Gadd45g*$^{+/-}$ Y^{AKR} and B6.Y^{Pos}. It is therefore likely to contain the regulatory regions required for the expression of *Gadd45g* in the developing testis. BACs are amenable to genetic modification, opening the possibility of using E15 as the basis for transgenic

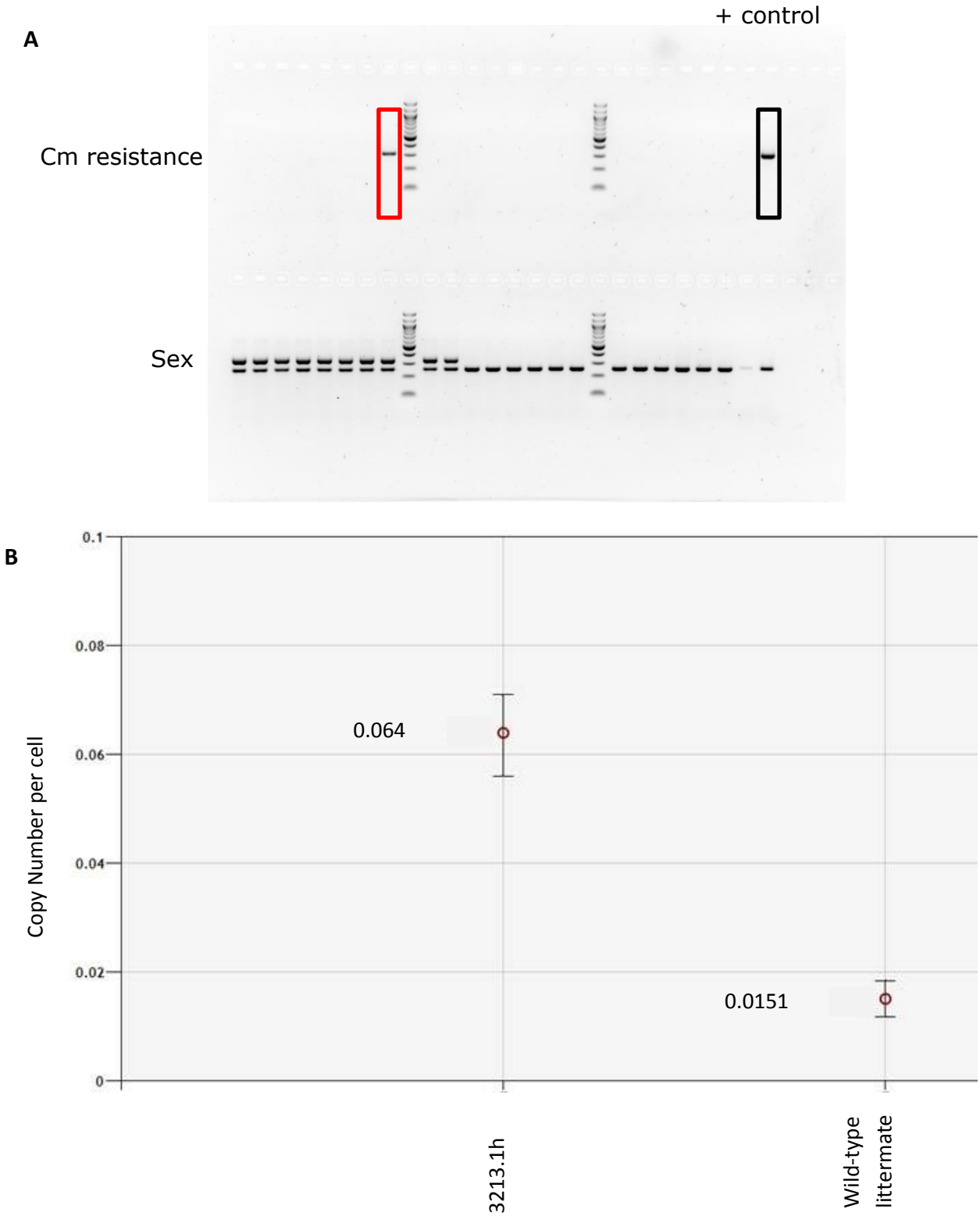


Figure 4.8 Genotyping of transgenic founders for *E15-Gadd45g-His/FLAG*

Ear biopsies of transgenic founders were genotyped for the presence of the *Cmr* gene present on the backbone of the *E15-Gadd45-His/FLAG* BAC. One potential transgenic founder was found, 3213.1h. A series of matings were set up between 3213.1h and C57Bl/6 females. However no transgenic offspring were produced. **B**) To reassess the transgenic status of the founder, droplet digital PCR was performed on a fresh ear biopsy. Results revealed that while 3213.1h showed significantly higher copy numbers of the transgene compared to wild-type littermates, the copy number only averaged 0.064 copies per droplet. Therefore it is unlikely that the transgene had entered the germline.

mouse line harbouring *Gadd45g* controlled constructs including tags or reporters. The addition of the sequence for a combination His/ FLAG epitope tag to the *Gadd45g* sequence within the BAC could provide an alternative to an antibody to the endogenous protein. This would allow us to investigate the interacting partners, chromatin recruitment and cellular location of GADD45γ in testis determination.

Using Red/Et mediated recombineering, I have successfully modified the *E15-Gadd45g* BAC to contain the sequence for the combination *His/FLAG* tag prior to the translational stop codon of *Gadd45g*. The modification is in the correct location; it has not deleted or added additional base pairs at the modification sites; and it does not show signs of any other major recombination events having occurred. This BAC was then micro-injected into 1-cell stage embryos to generate a transgenic line similar to that established by the unmodified BAC. Unfortunately this has not been the case, as a line containing the modified BAC has yet to be generated.

This is perhaps a surprising outcome given the number of sessions and embryos injected are almost equivalent to those which produced the unmodified BAC transgenic line (table 4.8). There does however appear to be an increased rate at which embryos injected with the modified BAC have lysed 24 hours after the injection compared to the unmodified BAC. In turn, this lead to a decreased number of embryos transferred into pseudo-pregnant females and around half of the number of pups being born. There are numerous possibilities for this increased lysis rate. BAC size and concentration has been shown to affect transgenesis rates. While E15 is relatively large for a BAC at 238kb, this is within the 61kb to 303kb range at which transgenic efficiency does not appear to be affected by size (Van Keuren et al., 2009). The concentration of an injected BAC is inversely proportional to birth rates but also correlates with the rate of transgenesis. BAC DNA preparations

were often reported to be highly viscous during injection. This increased viscosity can impede the injection, and may also be indicative of higher concentrations of DNA than the expected 1-2ng/µl. This may be the cause of the increased lysis of injected embryos observed in the *E15-Gadd45g-His* line.

Table 4.8 Statistics on the generation of transgenic embryos from the *E15-Gadd45g* BAC

Construct	No. sessions	Injected embryos	Lysed after injection	Transferred	Pups born	Transgenics
G45g-His	5	374	273	101	46	1
G45g-mCherry	1	40	16	24	2	0
G45g-194E15	6	331	158	237	104	4

Number of injection sessions carried out; the total number of injected embryos; the number of lysed embryos 10 hours after injection; the number of embryos transferred into pseudo-pregnant females; number of pups born and the number of transgenic founder generated for the unmodified *E15-Gadd45g* BAC (G45g-194E15), recombineered *E15-Gadd45g-His/FLAG* (G45g-His) and *E15-Gadd45g-mCherry* (G45g-mCherry).

Another possible explanation for the low numbers of transgenic animals could be the injected construct is deleterious to early development. This may be the result of a gene present on the BAC which cannot be tolerated at higher expression levels. Given the role of GADD45γ in cell cycle arrest, an overexpression of such genes could cause early death in embryos by impeding their early growth. However, the successful generation of the unmodified *E15-Gadd45g* line with a similar number of injected embryos suggests that it is unlikely to be any sequences located on the BAC. The modification to the *E15-Gadd45g* BAC was the insertion of a 45bp sequence encoding a *His/FLAG* tag immediately prior to the *Gadd45g* translational stop codon. This modification was shown to be in-frame for the *Gadd45g* reading frame and did not result in the removal or insertion of extra bases at

the edge of the homologous regions. Given the small size of this modification, it is also unlikely to be responsible for the increased lysis of injected embryos.

Despite the lack of transgenic animals at the time of writing, I have shown that *E15-Gadd45g* can be modified to contain transgenic constructs under the control of the *Gadd45g* regulatory region. In conjunction with the development of the *E15-Gadd45g-His/FLAG* construct, a modified *E15-Gadd45g* containing the coding sequence for the fluorescent reporter mCherry was generated in the lab (N. Warr unpublished data). The mCherry coding sequence followed by a stop codon and a poly-A signal was inserted at the *Gadd45g* translational start site within *E15-Gadd45g*. The correct integration and lack of other recombination events was confirmed in a similar manner to *E15-Gadd45g-His/FLAG*, using both Sanger sequencing and restriction digest fingerprinting. So far no transgenic animals have been produced containing *E15-Gadd45g-mCherry*, however this is likely due to only a single series of injections having been performed so far (table 4.8). So far 40 embryos have been injected, leading to the birth of only 2 pups, neither of which carried the transgene. *E15-Gadd45g* has been shown to be expressed in the embryonic gonads at the time of endogenous *Gadd45g* expression. Therefore, the *E15-Gadd45g-mCherry* construct will likely show a similar expression pattern. This would produce a fluorescent reporter line found in both sexes, controlled by a gene known to be critical for *Sry* expression; making it an invaluable tool for the sexual development field and research into the function of GADD45γ. Other such modifications may also be useful, including the introduction of a Cre recombinase gene.

The recent rise of the Clustered regularly interspaced short palindromic repeats(CRISPR)/Cas9 system of DNA editing presents a powerful alternative method of generating transgenic mouse lines (Yang et al., 2013). This technique could be applied to *Gadd45g* to produce either an epitope-tagged

or fluorescent version of the GADD45γ protein under the control of the endogenous promoter. BAC transgenic animals can show variability between different lines due to the position where the DNA has integrated as well as copy number variations between lines (Van Keuren et al., 2009). Targeted endonuclease systems such as CRISPR, Transcription activator-like effector nucleases (TALENs) or modified zinc finger nuclease (ZFNs) can produce much more efficient and reproducible transgenic lines (Beil et al., 2012; Auer et al., 2014). As the transgene is under the control of an endogenous promoter, it is more likely to reflect naturally occurring expression patterns. However, BAC transgenes do possess some advantages over genome editing approaches. One hypothesis for the lack of a GADD45γ antibody that is capable of detecting endogenous protein is that due to its role in cellular stress responses and growth arrest and thus the requirement for dynamic expression and repression, GADD45γ has a very high turnover rate. This may result in low levels of available protein for antibody detection. Though *in situ* hybridisation experiments have shown plentiful *Gadd45g* expression levels in the gonad during sex determination (Warr et al., 2012), the *Gadd45g* mRNA has been shown to have a half-life of less than 1 hour (Sharova et al., 2009). This could also lead to low levels of the protein for detection. By overexpressing *Gadd45g* with an epitope tag using BAC transgenesis, this potential problem might be overcome. Whilst data reported here indicate that BAC transgenesis is not an efficient process, and requires excellent microinjection skills, neither is the introduction of large constructs or subtle edits using CRISPR-Cas9. Most labs use CRISPR in ES cells for introducing larger constructs, such as a fluorescent reporter, into a gene of interest, rather than one-cell injection; and subtle edits are often associated with unwanted non-homologous end-joining events that introduce indels (Lydia Teboul, MLC pers. comm.). Perhaps most significantly, alteration of an endogenous gene may result in disrupted function and thus haploinsufficiency. For example, the *Wt1*CreERT line that is employed in our lab for inducible gene targeting in gonadal somatic cells lacks one functional copy of *Wt1*, and this alone generates a subtle sex determination phenotype that

must be taken into account when using this line (Gwenn Carre, pers. comm.). This is not a problem encountered when using BAC transgenes.

Chapter 5: Investigating methods of whole genome bisulphite sequencing using gonadal somatic cells

5.1 Introduction

5.1.1 The major cell types in the developing gonad

Mammalian gonads contain a number of different cells types. The first of these to show molecular divergence in XY and XX embryos are the supporting cell lineage. In XY gonads the expression of *Sry* within this lineage induces the supporting cells to differentiate into pre-Sertoli cells, which further produce signals to the surrounding cells to direct them to their respective testis specific lineages. In XX gonads, the absence of *Sry* expression leads to the upregulation of canonical WNT/beta-catenin signalling. This antagonises the expression of *Sox9* and *Fgf9* (Kim et al., 2006), ensuring that gonadal cells adopt ovarian fates. The other major population of gonadal somatic cells are the steroidogenic cells. After sex determination has taken place within the gonad, the hormones secreted by these cells will control how sex differentiation occurs throughout the rest of the body. In the testis, the secretion of DHH from pre-Sertoli cells causes steroidogenic cells to differentiate into foetal-Leydig cells (FLCs) (Barsoum and Yao, 2010). These cells produce androgens that will masculinise non-gonadal structures. In ovaries, the differentiation of the supporting cells towards the granulosa lineage induces the surrounding steroidogenic precursor cells to differentiate towards the theca cell fate. In a similar fashion to FLCs, the theca cells produce androgens that are converted to oestrogen by granulosa cells. The primordial germ cells (PGCs) migrate into the gonad from the hind gut from 9.5 – 11.5dpc. The fate of PGCs is decided by the surrounding supporting cell types. Pre-Sertoli cells coalesce around PGCs resulting in mitotic arrest. Conversely, retinoic acid signals induce meiosis in PGCs in ovarian tissue (Warr and Greenfield, 2012). Other lineages within the testis include the endothelial cells and the peritubular myoid cells, which both aid the formation of testis cords (Combes et al., 2009; Svingen and Koopman, 2013).

5.1.2 Fluorescence-activated cell sorting using *Sf1-EGFP*

Each of the cell types found within the developing gonad shows differential expression of growing number of genes from around 11.2 dpc (Jameson et al., 2012). In such studies, which aim to study the genetic networks controlling lineage-specific commitment, it is imperative that the different cell populations are enriched in order to reduce transcriptional 'noise' from the other cell types. This is most often achieved through the use of fluorescence-activated cell sorting (FACS). A particular method of flow cytometry, FACS allows for the purification of cell populations according to the presence of various fluorescent markers. These markers can be either the result of the expression of transgenic fluorescent proteins, chemical dyes or fluorescent antibodies targeting cell surface antigens (Herzenberg et al., 1979; Weston and Parish, 1990; Beverdam and Koopman, 2006). Cells are dissociated into a single cell suspension and forced into a stream by hydrodynamic processing. The stream is passed through a nozzle vibrating at high speeds to generate droplets. Droplets are then passed through a laser and detectors, before encountering a charger which can add a charge to droplets exhibiting the desired signal profile generated by the laser and detectors. Charged droplets are then directed into collection tubes by electromagnets.

In this particular study, the transgenic line *Sf1-EGFP* was used to mark gonadal somatic cells. This transgene is expressed in the somatic cells of both XX and XY gonadal ridges from 10.5dpc and continues until at least 16.5dpc (Fig 5.1)(Beverdam and Koopman, 2006). This line has been extensively used throughout the sex development field to purify somatic cells from the gonad. This has provided much insight into the transcriptomes of the somatic cells of the gonads (Nef et al., 2005; Beverdam and Koopman, 2006; Jameson et al., 2012).

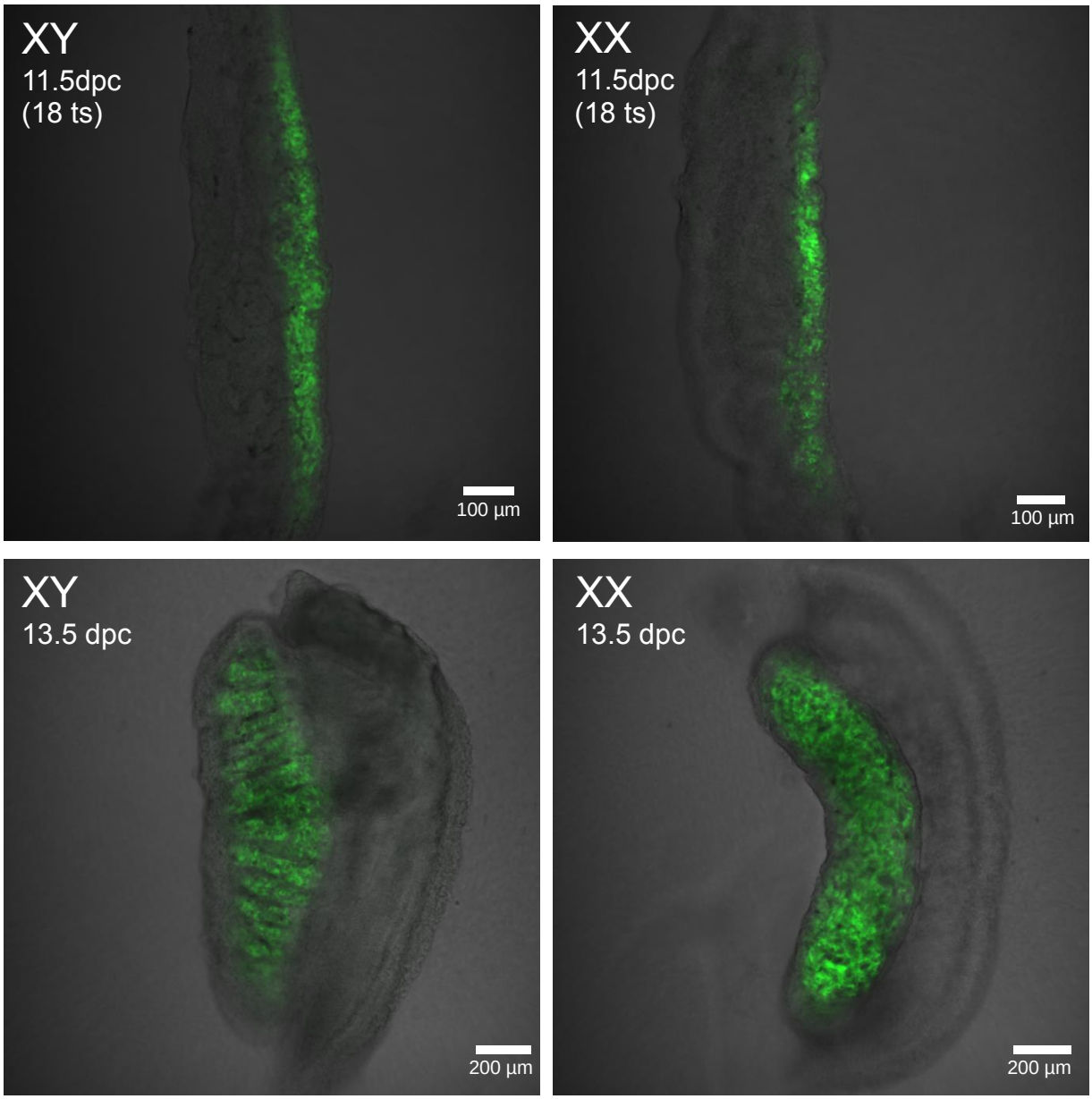


Figure 5.1: Expression of *Sf1-eGFP* in the developing gonad
Fluorescence microscopy of gonads from embryos carrying the *Sf1-eGFP* transgene show fluorescence throughout the gonads of both XY and XX embryos within the somatic cell lineage. At 11.5 dpc, XY and XX gonads are morphologically similar. By 13.5 dpc in XY gonads, the testis cords have formed. The Sertoli cells within the cords fluoresce more brightly than the interstitial cells at this period. In the ovary, the various somatic cell lineages are still relatively homogenous throughout the gonad.

5.1.3 Epigenetic studies of gonadal cell lineages

While there has been extensive research into the differences between the transcriptomes of the differentiating cell of the gonad, there are few studies investigating any role that epigenetics plays in establishing these observed transcriptional profiles. The study of chromatin and DNA chemical marks in a genome-wide fashion – so-called epigenomics – is of much interest in fundamental biological and clinical research. Major international colaborative projects such as the encyclopaedia of DNA elements (ENCODE) have established the epigenomic profile of numerous cell lines and tissues, including testicular tissue from adult mice (8 and 10 weeks of age) (Yue et al., 2014).However, these tissues are unlikely to reflect the epigenetic state that is present as the time of sex determination. Nor is there any indication that cells have been sorted to enrich for the supporting (Sertoli) cells implicated in sex determination.

Some studies have taken a more targeted approach. A series of CpG sites in the *Sry* promoter were revealed to show dynamic methylation in relation to the expression of *Sry* (Nishino et al., 2004; Nishino et al., 2011). A recent study looked specifically at the methylation status of the promoters of *Sox9* and *Oct4* in the testis at various ages (Pamnani et al., 2016). Methylation of specific CpG sites appeared to correlate with the level of expression of these genes in their respective tissues. Curiously though, data are only shown for the adult ovary, preventing us from drawing any conclusions about the requirement for methylation of *Sox9* in sex determination. No indication is given as to whether somatic and germ cells were purified from each other. This step is necessary given the expression of *Oct4* is limited to PGCs, which are undergoing extensive demethylation at the time of sex determination (Yoshimizu et al., 1999; Seisenberger et al., 2012). Little else is known about genome-

wide DNA methylation profiles (the methylome) in gonadal somatic cells and how it relates to the control of sex determining genes.

5.1.4 Genome-wide DNA methylation analysis methods

Due to its implied role in active DNA demethylation (Niehrs and Schafer, 2012), it was hypothesised that the delay in *Sry* expression found in *Gadd45g*$^{-/-}$ mice could be due to a failure to remove the methyl groups from the CpG dinucleotides at the *Sry* promoter. However, the *Sry* promoter was shown to be unmethylated in *Gadd45g*$^{-/-}$ gonadal somatic cells at the 15/16 tail somite stage (Gierl et al., 2012; Warr et al., 2012). This does not exclude the possibility that GADD45γ is affecting the methylation levels of other loci required for sex determination. To find such loci, it was decided that an approach that examined DNA methylation throughout the genome would be taken.

Various methods exist to investigate methylation at a global level throughout a population of cells. Methyl-DNA immunoprecipitation (MeDIP) is a technique which enriches areas of the genome with high levels of methylated CpGs for sequencing. Sheared DNA is treated with an antibody raised against 5mC. Immunoprecipitation is then used to pull down the 5mC containing DNA, which is then sequenced (Jacinto et al., 2008). Though this method can be used to find major DNA methylation differences between two cell types, individual regulatory CpGs or those in low CpG region cannot be readily detected. Methylation sensitive restriction endonucleases allow another method of investigating methylation across a genome. The most commonly used enzymes are *HpaII* and *SmaI*. Each of these enzymes has a closely related protein that can target the same sequence but is not sensitive to the presence of a methyl group on the CpG site; *MspI* for *HpaII* and *XmaI* for *SmaI*. Samples can be treated with a methyl-sensitive enzyme and then sequenced using next generation sequencing techniques. Though this approach can cover large swaths of the genome and give

information on specific CpG sites, analysis is limited to sequences which can be targeted by the restriction enzyme.

The gold standard for base-level methylome detection is whole genome bisulphite sequencing (WGBS). Denatured DNA treated with sodium bisulphite leads to the deamination of unmethylated cytosine producing a uracil base. Methylation of the 5 position carbon on cytosine protects the base from the deamination, maintaining the base as a cytosine. In subsequent PCR amplification, the modified uracil base will be read as a thymine and amplified accordingly. These amplified reads can then be sequenced. Reads from bisulphite-converted DNA can be compared against a reference genome, which are converted *in silico* so that all cytosines are converted to thymines. This C to T conversion allows methylation to be observed at a base-level resolution on any known DNA sequence. This process can be applied in a targeted manner towards a particular region of the genome or globally using next generation sequencing (Laird, 2010). It is this approach that I describe in this thesis for the investigation of the methylome of gonadal somatic cells from XY wild-types, XX wild-type and XY *Gadd45g*$^{-/-}$ mice. Not only would such base-resolution methylome analysis illuminate the role of GADD45γ in mammalian sex determination, but it may also uncover new genes or regulatory elements involved in this process.

5.1.5 Aims and Hypotheses

I aim to sort cells from *Sf1-EGFP* gonads in order to enrich for the somatic cells of the gonad, in which sex determination takes place and *Gadd45g* is expressed. This will be achieved through fluorescence activated cell sorting, which I will need to ensure works in my hands by demonstrating that the sorted cells show a similar expression and epigenetic profile to somatic gonadal cells. I will also aim to produce a library suitable for whole genome bisulphite sequencing from these cells.

5.2 Results

5.2.1 Enrichment of gonadal somatic cells using transgenic *Sf1-EGFP* mice

As *Gadd45g* expression is not found in PGCs (Warr et al., 2012), the somatic component of the gonads needs to be purified to minimise noise from cell types not contributing to sex determination. As a result the *Sf1-EGFP* line was imported from the Koopman lab, University of Queensland, Australia. To test if somatic cells could successfully be sorted from transgenic gonads a series of experiments using the *Sf1-EGFP* line was conducted. Studs from this line were first crossed to C57BL/6J (B6) females to generate transgenic embryos. The urogenital ridges from 11.5dpc (17 - 18 ts) embryos, the age at which *Sry* expression is at its highest point, were dissected and genotyped. Fluorescence from the S*f1-EGFP* transgene was detected using fluorescence microscopy. UGRs were then pooled according to their genotype and fluorescence, and then further sub dissected to remove the gonad and mesonephros from the dorsal aorta and surrounding tissue. Gonads were then dissociated in trypsin and resuspended in PBS. Samples from GFP positive embryos were then loaded into the S3e cell sorter (BioRad) and briefly allowed to run through the stream to set the appropriate gating (Fig 5.2). Cells from non-transgenic littermates were then run to ensure the gates marking fluorescent cells were sufficiently stringent to not include any GFP negative cells. This particular gating scheme produced approximately 15,000 – 20,000 EGFP+ cells per pair of gonads. To ensure that gonadal somatic cells had successfully been enriched, qRT-PCR and targeted bisulphite sequencing were performed.

For qRT-PCR, RNA was extracted from GFP positive (EGFP+) cells; GFP negative (EGFP-) cells from S*f1-EGFP* UGRs and cells from sub-dissected B6 UGRs, pairs of gonads and their attached mesonephroi) which had undergone FACS. Each sample contained cells from 3 – 5 pairs of gonads, and was run in

duplicate. Mean relative expression was calculated according to section 2.9.3 (Fig 5.3). Both XY and XX showed higher relative expression of *Sf1* in EGFP+ cells compared to their respective EGFP- cells, which showed negligible levels of expression, as did the non-transgenic gonads. Relative expression of *Gadd45g* was found to be higher in EGFP+ cells compared to EGFP- cells in both sexes as well.

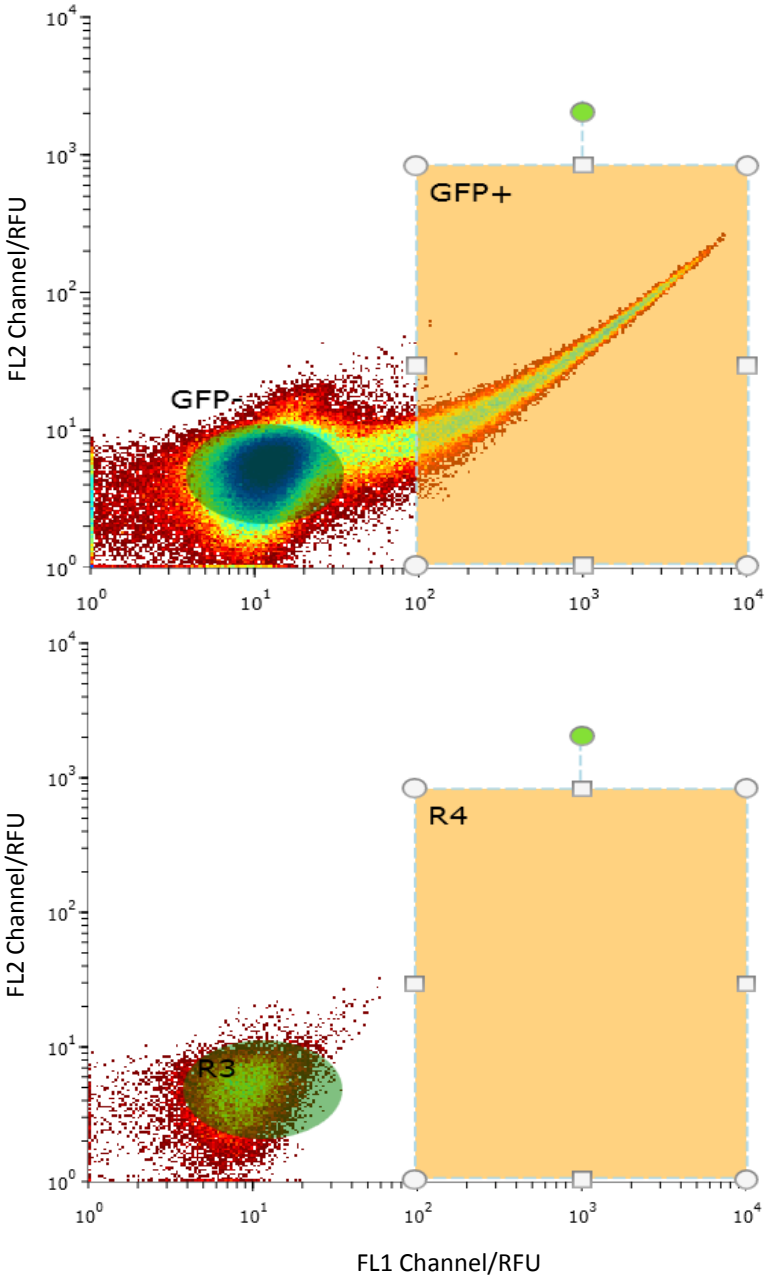
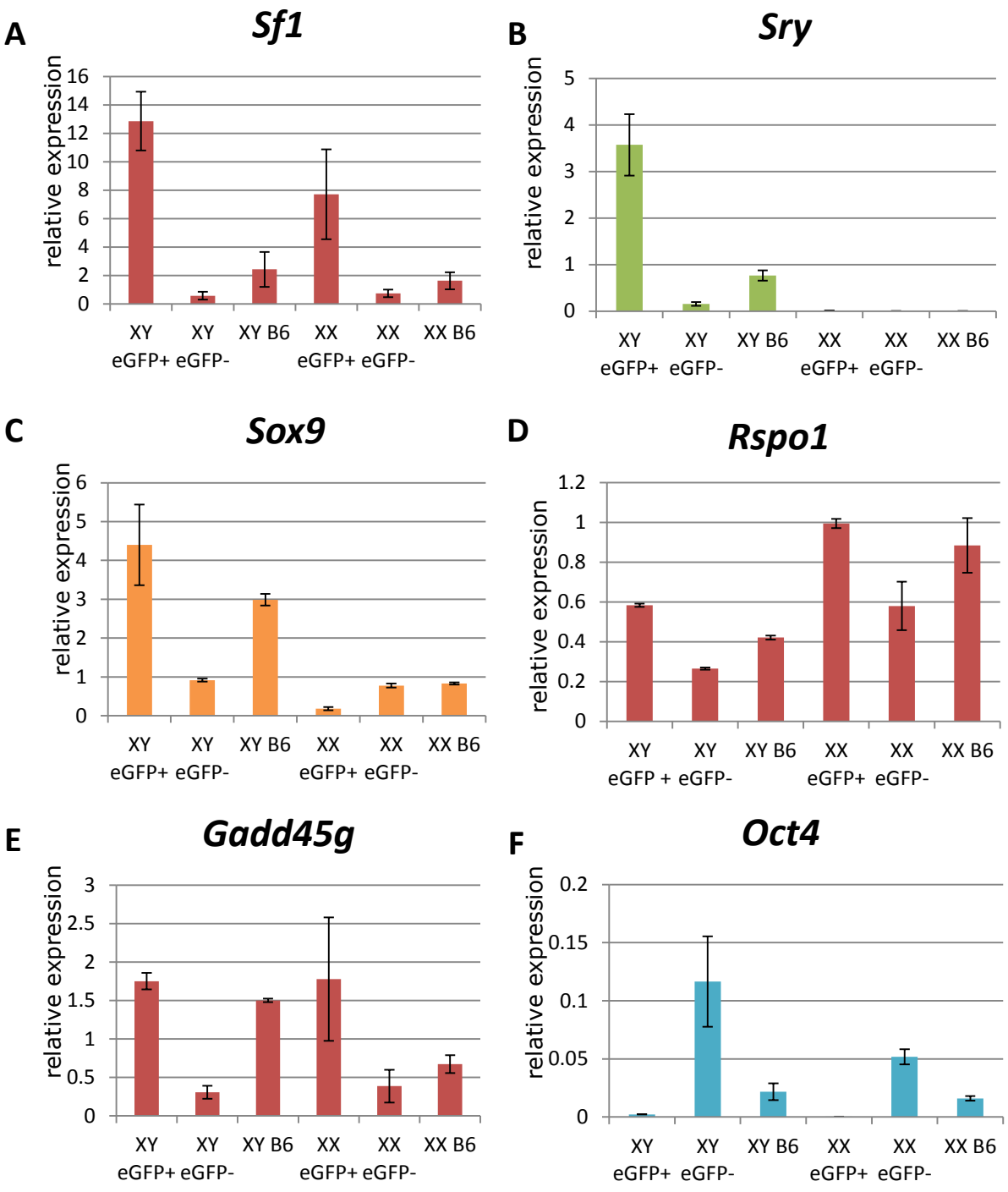


Figure 5.2: Gating used to sort *Sf1-EGFP* positive and negative cells
(**A**) Cells from gonads carrying the *Sf1-EGFP* transgene were sorted using the above gating on an S3e cell sorter (BioRad). FL1 refers to the channel which detects fluorescence between 500 – 530nm, while FL2 detects between 520 – 550nm. RFU refers to relative fluorescence units. **B**) cells from gonads of non-transgenic animals do not produce cells with sufficient levels of fluorescence to fall within the GFP+ gate


A
Sf1
relative expression
16
14
12
10
8
6
4
2
0
XY eGFP+
XY eGFP-
XY B6
XX eGFP+
XX eGFP-
XX B6
B
Sry
relative expression
5
4
3
2
1
0
XY eGFP+
XY eGFP-
XY B6
XX eGFP+
XX eGFP-
XX B6
C
Sox9
relative expression
6
5
4
3
2
1
0
XY eGFP+
XY eGFP-
XY B6
XX eGFP+
XX eGFP-
XX B6
D
Rspo1
relative expression
1.2
1
0.8
0.6
0.4
0.2
0
XY eGFP +
XY eGFP-
XY B6
XX eGFP+
XX eGFP-
XX B6
E
Gadd45g
relative expression
3
2.5
2
1.5
1
0.5
0
XY eGFP+
XY eGFP-
XY B6
XX eGFP+
XX eGFP-
XX B6
F
Oct4
relative expression
0.2
0.15
0.1
0.05
0
XY eGFP+
XY eGFP-
XY B6
XX eGFP+
XX eGFP-
XX B6

G

T-test P values	*Sf1*	*Sry*	*Sox9*	*Rspo1*	*Gadd45g*	*Oct4*
XY EGFP+ v XY EGFP-	0.000527	0.000883	0.004579	2.82E-05	0.010421	0.007711
XY EGFP+ v XY B6	0.001674	0.001959	0.086397	0.000131	0.466817	0.01074
XY EGFP- v B6	0.074269	0.000876	2.2E-05	5.21E-05	0.01317	0.015519
XX EGFP+ v XX EGFP-	0.021376	0.605434	0.000104	0.004255	0.229192	0.000169
XX EGFP+ v XX B6	0.034331	0.894486	2.25E-05	0.003905	0.044107	0.000172
XX EGFP- v XX B6	0.039217	0.000363	6.37E-05	0.000119	0.477785	0.019113
XY EGFP+ v XX EGFP+	0.07249	0.000745	0.002242	0.003101	0.004308	7.05E-05
XY EGFP- v XX EGFP-	0.488984	0.002832	0.01939	0.021525	0.317304	0.051998
XY B6 v XX B6	0.411036	0.000302	1.76E-05	9.76E-06	0.008616	0.287512

H

stat power	*Sf1*	*Sry*	*Sox9*	*Rspo1*	*Gadd45g*	*Oct4*
XY EGFP+ v XY EGFP-	0.9993	0.9943	0.8298	0.9999	0.9999	0.7205
XY EGFP+ v XY B6	0.9921	0.9581	0.2193	0.9999	0.5246	0.6404
XY EGFP- v B6	0.9994	0.9973	0.9999	0.9999	0.9999	0.5603
XX EGFP+ v XX EGFP-	0.9999	0.0651	0.9999	0.8433	0.3246	0.9999
XX EGFP+ v XX B6	0.9999	0.0519	0.9999	0.8581	0.2233	0.9999
XX EGFP- v XX B6	0.8344	0.057	0.1475	0.0508	0.2139	0.9973
XY EGFP+ v XX EGFP+	0.2923	0.9968	0.9413	0.9634	0.0504	0.9999
XY EGFP- v XX EGFP-	0.0832	0.9127	0.6476	0.5015	0.0624	0.3034
XY B6 v XX B6	0.2178	0.9999	0.9999	0.9999	0.9999	0.1055

Figure 5.3: Quantitive Real-time PCR analysis of sorted cells
qRT-PCR was performed on cDNA synthesised from RNA extracted from FACS-sorted cells (eGFP+ and eGFP-) or from the gonad and mesonephros of wild-type mice (B6). **A**) *Sf1* was enriched in eGFP positive populations in both XY and XX samples. Negligible expression was found in eGFP negative cells. **B**)*Sry* was enriched in eGFP positive XY cell compared to wild-type cells. *Sry* was not detected in XY eGFP negative cells or XX cells. **C**) A similar pattern was seen in *Sox9* expression, though limited ammounts of expression were found in eGFP-cells from XY and XX gonads. *Sox9* was not detected in XX eGFP positive cells. **D**) *Rspo1* was found at higher levels in all XX lineages compared to XY lineages. Expression was also found to be enriched in eGFP positive cells compared to eGFP negative cells. **E**) *Gadd45g* was found to be enriched in eGFP positive cells compared to eGFP negative cells in both XY and XX cells. **F**) The germ cell marker *Oct4* was not found in eGFP positive cells and enriched in eGFP negative cells. Two tailed t-test p values are given in the table (**G**)**,** those that are >0.05 and therefore considered non-significant are highlighted in red. **H**) Statistical power of the above t-tests is also provided.

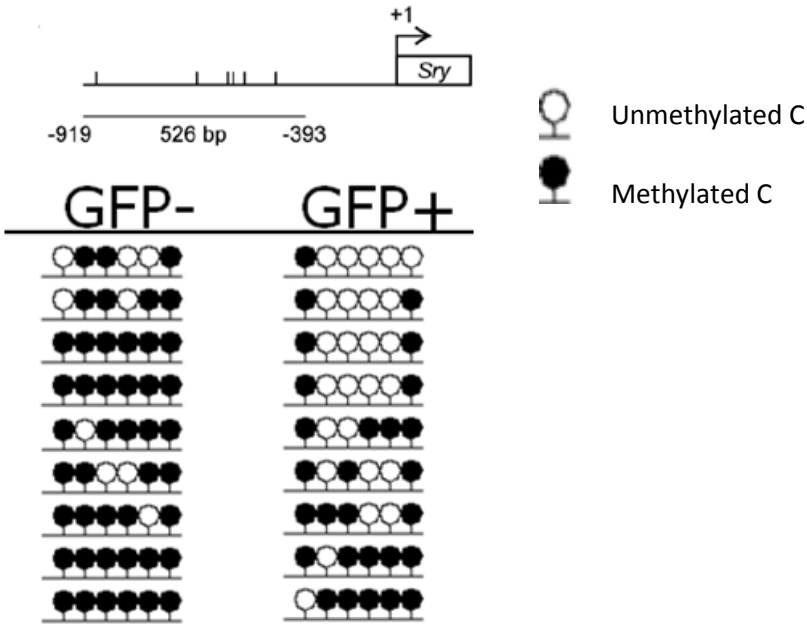
Expression of *Sry* and *Sox9* were also found to be at higher levels in the EGFP+ cells of XY gonads compared to both the EGFP- and non-transgenic gonads. XX gonads acted as negative controls for the expression of *Sry. Sox9* expression was significantly lower in XX EGFP+ cells compared to XY EGFP+ and non-transgenic cells. Expression of the germ cell marker *Oct4* was found at negligible levels in EGFP+ cells of both sexes, and found to be enriched in the EGFP- cells compared to non-transgenic.

Targeted bisulphite sequencing was used to investigate the methylation profile of the *Sry* proximal promoter in sorted cells from 18 ts XY gonads. DNA extracted from these cells was bisulphite treated, amplified by PCR using primers targeting the bisulphite-converted *Sry* region and products were cloned into the pGem-T vector as described in section 2.6.1 and 2.6.2. A Fisher's exact test revealed that the distribution of methylated and non-methylated CpGs was significantly different between the two populations (p= 0.000864) (Bock, 2012), sample size of 54 CpGs gave a statistical power of 96.03% (HyLown Consulting LLC, 2013). Hypomethylation of the *Sry* proximal promoter was found in EGFP+ cells, whilst CpGs were comparatively hypermethylated in the promoter from EGFP- cells (Fig 5.4). Together these expression and methylation data suggest that the gating used to sort cells had successfully enriched for gonadal somatic cells.

5.2.2 Improving the DNA yield from sorted gonadal somatic cells

One of the limiting factors of WGBS is the requirement for relatively large amounts of DNA, typically up to 5µg (Miura et al., 2012). WGBS libraries are generated in a similar method to other next generation sequencing libraries. DNA is sheared through either sonication or restriction digests, the ends are repaired and an adenine base added at the 3' end. Adaptors are then added, which in WGBS are fully methylated at cytosine bases to prevent conversion in the subsequent bisulphite treatment. Adaptor tagged sequences are then enriched for with four rounds of PCR (Lister et al., 2009; Kulis et al., 2012). The requirement for large amounts of DNA comes from the numerous steps required for

the proper fragmentation, end repair, A-tailing and adaptor ligation, which results in loss of material at each stage. Bisulphite treatment can degrade DNA however, resulting in reduced levels of library template (Miura et al., 2012).



	GFP-	GFP-
Me-C	45	28
non Me	9	26

Fisher's exact test
P value = 0.000864

Figure 5.4 Bisulphite sequencing data of the proximal *Sry* promoter in FACS sorted cells
Lollipop diagrams representing the methyaltion status of 6 CpGs within the proximal *Sry* promoter of *sf1-eGFP* positive and negative cells sorted by FACS. Fisher's exact test shows that there is a significant difference between the two samples in the distribution of the number of methylated and non-methylated cytosines. Hypomethylation is observed in eGFP+ cells compared to eGFP- cells, suggesting that gonadal somatic cells have been sorted to this population

Through successive cell sorts it was found that each pair of gonads yielded approximately 15,000 – 20,000 EGFP+ cells. A series of DNA extraction across different numbers of cells revealed that 20,000 cells provided approximately 80ng of DNA (Table 5.1). In order to achieve the required 5µg, approximately 62 embryos of each genotype would be needed. Given that the chance of any given embryo produced by this cross being either XX *Sf1-EGFP +*, *Gadd45g*$^{+/+}$; XY *Sf1-EGFP +*, *Gadd45g*$^{+/+}$ or XY, *Sf1-EGFP*, *Gadd45g*$^{-/-}$ is 3/16, this would require approximately 330 stage-matched embryos. This has yet to be achieved, even in three years of breeding and sorting cells.

Numerous methods have been devised to generate methylome data from low amounts of input DNA. Tagmentation-based whole genome bisulphite sequencing (T-WGBS) is a method which ensures the more efficient ligation of adaptor sequences to DNA fragments. In conventional WGBS the numerous reactions needed for the steps between DNA fragmentation and adaptor ligation reduce the proportion of DNA fragments successfully adaptor ligated. T-WGBS uses a Tn5 transposase which targets a 19bp target sequence. This sequence is amended to the 3' end of a methylated adaptor sequence which is then used to shear and add the adaptor sequence in one step through the action of the transposase (Adey and Shendure, 2012; Wang et al., 2013).

Reduced representation bisulphite sequencing (RRBS) is a method which utilises restriction enzymes which are insensitive to the methylation status of the CpG site contained within their target sequence. Digestion of genomic DNA with such an enzyme allows for the selection of fragments with a relatively high density of CpG sites, usually located around promoters and CpG islands (Meissner et al., 2005; Gu et al., 2011). As a result the amount of sequencing and the required tissue is greatly reduced, at the expense of genome coverage.

Whole genome amplification (WGA) aims to amplify the initial low amounts of DNA to a useable amount. This technique has been applied to bisulphite-converted DNA to generate WGBS libraries Early methods of WGA utilised PCR such as primer extension preamplification (PEP) and degenerate-oligonucleotide-primed PCR (DOP) (Lasken and Egholm, 2003). However, these methods can preferentially amplify certain areas of the genome (Hosono et al., 2003). Multiple displacement amplification (MDA) minimises the differences in amplification of loci to >6-fold compared to up to 10^3 –fold differences found in PEP (Lasken and Egholm, 2003). MDA utilises the DNA polymerase from bacteriophage ϕ29, which amplifies DNA stably at 30 ° C (Fig 5.5). The ϕ29 polymerase binds to DNA at a priming event, initiated by random hexamers. It proceeds in a 5' to 3' direction and will displace other priming events it encounters due to its high binding affinity to the template strand. These displaced strands of DNA can then be primed again, and act as another template strand. This reaction continues until all of the available oligonucleotides or primers have been used up, generating up to 80µg of DNA from a 100µl reaction (Lasken and Egholm, 2003). MDA based WGA has previously been used in samples as low as 2000 cells to generate libraries for WGBS (Kobayuki et al., 2012). Due to its previous use and the relative ease compared to other methods, MDA was selected as a potential method of increasing the amount of DNA available to generate a WGBS library from the sorted gonadal cells.

As the only known differentially methylated region in gonadal somatic cells is the *Sry* promoter, this region was investigated to ensure that MDA was compatible with bisulphite-converted DNA (bcDNA) in my hands. DNA was extracted from the XY EGFP+ and EGFP- cells and bisulphite-converted using an EZ DNA methylation Gold kit (Zymo). 50ng of converted DNA from each of the cell types was then treated with the Epitect Whole Biusulfitome kit (Qiagen). The following day MDA treated samples

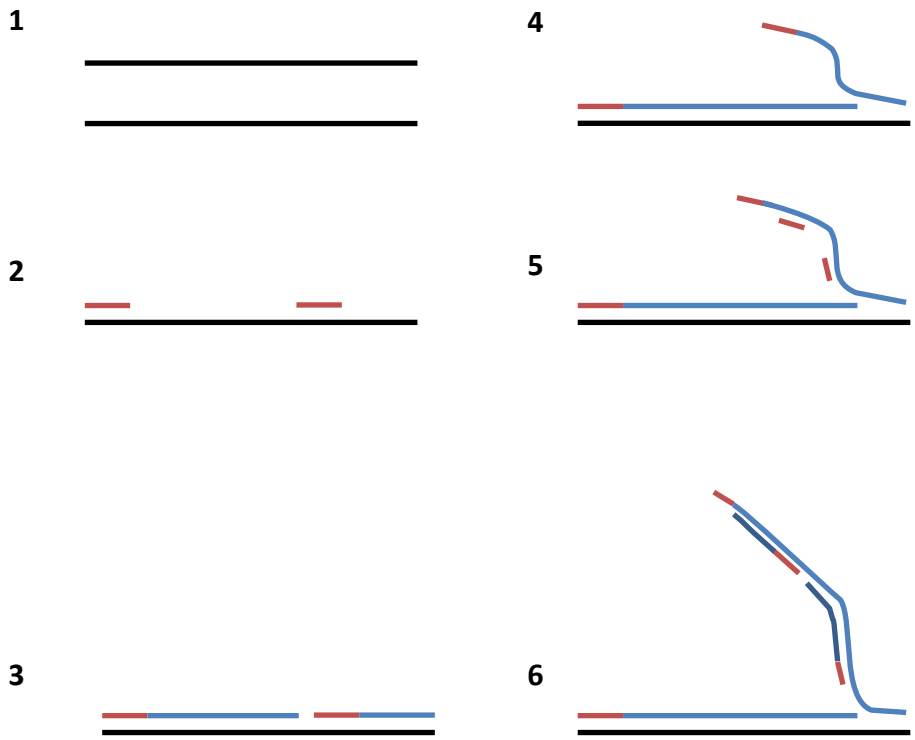


Figure 5.5: Principles of Multiple Displacement Amplification
1) DNA is denatured by incubation at 95 ˚C for 5 minutes. **2**) Random hexamers prime along the denatured DNA. **3**) The ϕ29 polymerase elongates the DNA in a 5' to 3' direction from each priming event. **4**) Once another priming event is encountered the primer and coding strand are displaced. **5**) The displaced strand is then primed again and (**6**) becomes a template strand.

were quantified, revealing yields ranging from 2.3 – 4.4µg. The MDA reaction can be increased in scale by increasing the volume of the reaction(Lasken and Egholm, 2003), and as a result can be increased to generate the 5µg required for WGBS. MD- treated samples were then diluted and 10ng of DNA was used as a template, along with the non-MDA treated samples, in a PCR reaction targeting the proximal promoter of *Sry*. Results showed that the MDA and non-MDA treated samples from the EGFP+ cells showed similar levels of methylation (Fisher's exact test, P = 0.090633), as did those from the EGFP- cells (Fisher's exact test, P= 0.847842)(Fig 5.6). As before, DNA from EGFP+ cells appeared to be less methylated compared to those from EGFP- cells regardless of whether samples were amplified with MDA or not (non-MDA treated GFP+ versus GFP- (Fisher's exact test, P=0.000323); MDA treated GFP+ versus GFP- (Fisher's exact test, p= 0.000001)).

5.3 Conclusions and Discussion

5.3.1 Aims revisited

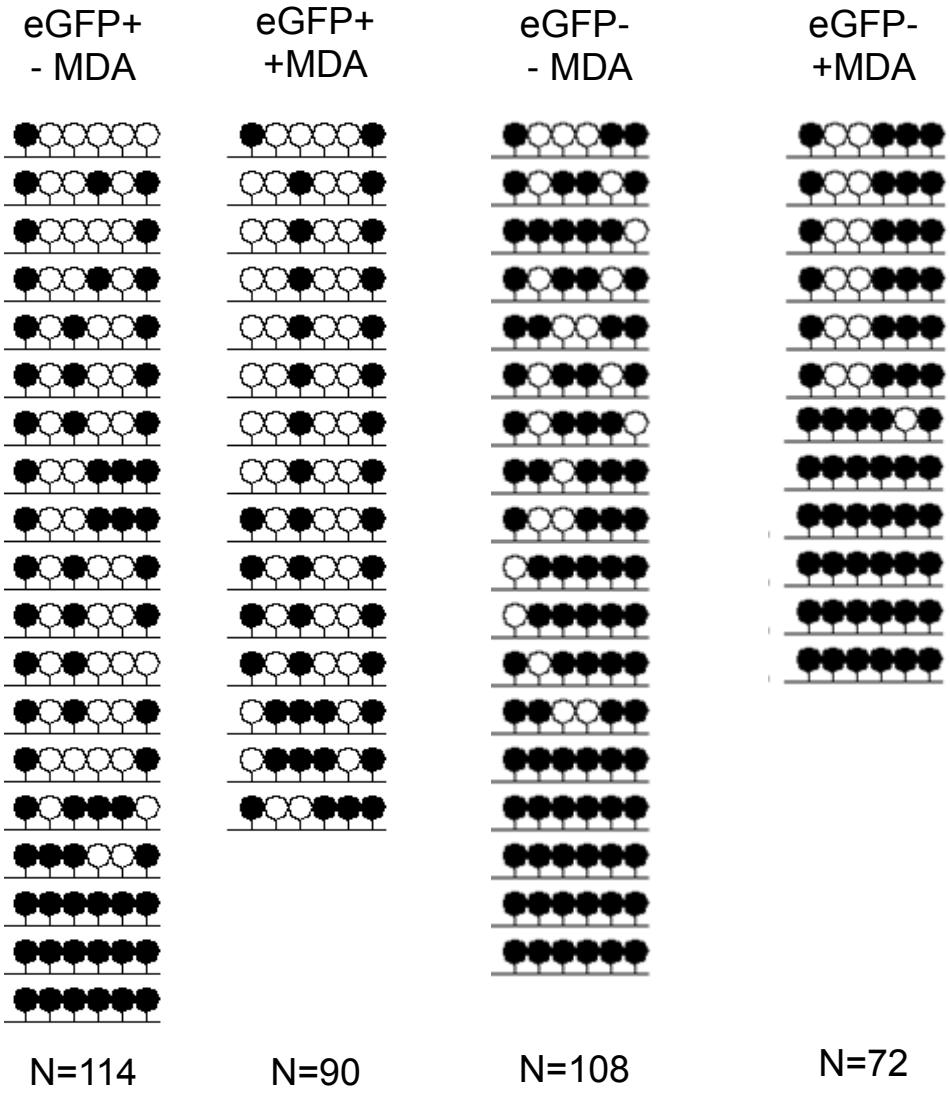
In this chapter I aimed to ensure that I could enrich for somatic gonadal cells using FACS on gonads from embryos carrying the *Sf1-EGFP* transgene. I also aimed to investigate potential methods of generating a whole genome bisulphite sequencing library using DNA from these cells.

5.3.2 Sf1-EGFP can be used to enrich for somatic gonadal cells expressing *Gadd45g* and *Sry*

When studying epigenomic phenomena, the enrichment of different cell types is important to ensure accuracy of the obtained data. This was achieved using the *Sf1-EGFP* transgenic line and FACS. Cells were sorted from embryos carrying the *Sf1-EGFP* transgene, into EGFP+ and EGFP- fractions. Cells from the EGFP+ fraction were shown to be enriched for gonadal somatic cell markers such as *Sry*,

Sox9 and *Sf1*. These cells also showed negligible levels of the germ cell marker *Oct4*. Crucially these cells also showed increased levels of *Gadd45g* expression, suggesting that *Gadd45g* is expressed at higher levels in the somatic component of the gonad. In XY gonads, the EGFP+ cells were also shown to have a similar DNA methylation profile to gonadal somatic cells at the *Sry* promoter. The six CpG sites within the *Sry* promoter were hypomethylated compared to those from EGFP- cells. Together, these results suggest that the sorted EGFP+ cells are a population of gonadal somatic cells.

A



B

Fishers exact P values		statistical power
GFP+ MDA- v GFP+ MDA+	0.090633	0.773
GFP- MDA- v GFP- MDA+	0.847842	0.4364
GFP+ MDA- v GFP- MDA-	0.000323	0.9631
GFP+ MDA+ v GFP- MDA+	0.000001	0.9997

Figure 5.6: Methylation status of the *Sry* promoter with or without MDA treatment
A)Lollipop diagrams showing the methylation state of the six CpG sites within the *Sry* proximal promoter of eGFP positive (eGFP+) and negative cells (-eGFP) with MDA treatment (+MDA) and without MDA treatment (-MDA). Black circles represent methylated CpGs, white circles unmethylated CpG sites. **B**) Fisher's exact T-test value comparing the percentage of methylated cytosines within the *Sry* promoter region between the four treatments, along with the statistical power calculation of the number of samples used. N in the above statistical test refers to the number of CpGs tested per treatment.

5.3.3 Multiple Displacement Amplification can be used to study DNA methylation from low amounts of DNA

Each pair of gonads carrying the *Sf1-EGFP* transgene yields 15,000 – 20,000 EGFP+ cells. In turn this number of cells would produce approximately 80ng of DNA. Whole genome bisulphite sequencing typically requires 5µg of DNA. As this would require an excessive number of embryos, multiple displacement amplification (MDA) was investigated as a potential method of increasing bisulphite treated DNA. MDA did not appear to have a major effect on the *Sry* promoter. Hypomethylation of the six CpGs was observed in EGFP+ cells from *Sf1-EGFP* carrying embryos, compared to EGFP- cells. MDA generated samples from EGFP+ and EGFP- cells also largely reflected this pattern. MDA treated samples also yielded 3 - 5µg of DNA from an input of 50ng of bisulphite-converted DNA (bcDNA). Together these results suggest that MDA can be used to generate sufficient volumes for whole genome bisulphite libraries.

However, it should be noted that while the MDA samples reflected the hypomethylation of EGFP+ cells compared to EGFP- cells, the generated methylation patterns appear less complex in MDA samples. Only three different patterns of methylation were observed in all of the reads of the MDA EGFP- samples, compared to eleven in the non-MDA samples. This was also observed in the EGFP+ cells though to a lesser extent. Five different patterns were seen in the MDA samples compared to 10 in the non MDA samples. This is a phenomenon which has been noticed before with WGA-based methods of generating WGBS libraries (Bundo et al., 2012). The amplification of samples can preferentially amplify specific reads, which then amplify further in other rounds of amplification. As a result, methylation calls of specific CpG sites trend towards a bimodal distribution. As many sites within the genome show a range of methylation percentages, this bias can lead to the false identification of differentially methylated regions. This effect becomes more extreme with lower

levels of input DNA (Bundo et al., 2012). Other estimates of the accuracy of MDA samples showed a high correlation of methylation percentage between specific CpG sites. However these sites are known to show methylation levels of >80% or <20% (Mill et al., 2006). In this thesis 50ng of DNA was used to test the MDA protocol. Given that each pair of gonads typically yields 80ng of DNA, this increased input, along with pooling of gonads of a similar genotype, could help minimise the biasing effect.

Whilst MDA can be utilised to generate WGBS libraries, many studies in recent years have utilised post-bisulphite adaptor tagging (PBAT) as an alternative method of generating WGBS libraries from small amounts of tissue. The next chapter describes the use of this method to generate WGBS libraries.

Chapter 6: The methylomes of gonadal somatic cells at the sex-determining stage and the role of *Gadd45g* in its establishment

6.1 Introduction

6.1.1 Post-bisulphite adaptor tagging (PBAT)

Though amplification-based methods can be used to generate WGBS libraries from small amounts of DNA (Mill et al., 2006), many published studies instead have employed the post-bisulphite adaptor tagging (PBAT) protocol (Kobayuki et al., 2012; Kobayashi et al., 2013; Sirane et al., 2013; Smallwood et al., 2014) (outlined in Fig 6.1). PBAT has been shown to produce single-base resolution libraries without the requirement for large amounts of amplification, thus removing the biases that can occur. Unlike traditional WGBS library preparation, sample DNA is bisulphite-converted in the first step of PBAT. Adaptors are then added to the converted DNA preventing the loss of available template through bisulphite-induced degradation (Miura et al., 2012).

Following bisulphite treatment, converted DNA is fragmented and becomes single stranded. Converted DNA is then used as a template in a random primer extension reaction. 5' biotinylated primers containing the adaptor sequence and 3' random hexamers are used to generate a complementary strand to the converted DNA. Primer dimers are removed and the copied strand (first-strand) DNA is purified using streptavidin beads followed by alkaline denaturation. The immobilised first-strand is then used as a template with a second primer containing the other adaptor sequence and random hexamers at its 3' end. The resulting double stranded-DNA is then eluted from the beads (Miura et al., 2012). At this stage, if sufficient amounts of DNA for a sequencing library have not been produced, then limited cycles of amplification can be performed using primers targeting the adaptor sequences. Products are then size-selected and purified for next generation sequencing.

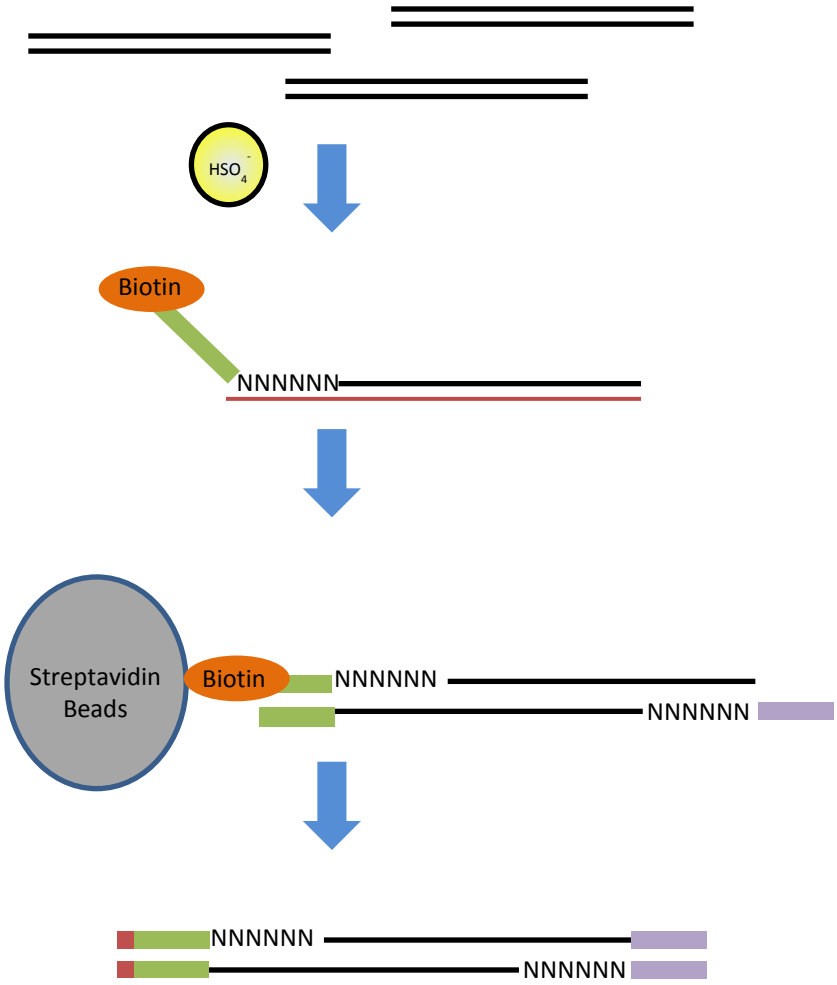


Figure 6.1 Post-bisulphite adaptor tagging (PBAT) main principles
Outline showing the main progress through the PBAT protocol. DNA is bisulphite converted, which also causes the fragmentation of DNA. Random primers containing the 5' adaptor tagged with biotin are used to generate a strand of DNA from the template converted DNA (in red), resulting in first strand synthesis. The synthesised first strands are purified with streptavadin beads and then used as a template to produce a second DNA strand. This process uses a primer containing a random hexamer and the other adaptor sequence. The adaptor tagged DNA is then purified and can be amplified for a limited number of PCR cycles if needed.

As PBAT has been adopted as something of an industry leader, and even used to generate methylome data from single cells (Smallwood et al., 2014), it was decided that this approach would be taken to generate bisulphite-converted libraries from EGFP+ cells of XY, XX and XY *Gadd45g*$^{-/-}$ embryos. Three sorted-cell samples of each genotype were taken to the Kelsey lab at the Babraham Institute (Cambridge, UK), where libraries are regularly made through PBAT using small numbers of cells. I followed established Kelsey lab protocols in order to generate PBAT libraries (see section 2.7.4).

6.1.2 Analysis of whole genome bisulphite sequencing data

Though the PBAT protocol significantly increases sequencing yield, analysis of sequence data from WGBS libraries following traditional bisulphite treatment and PBAT pose a substantial bioinformatic challenge. Working with the bioinformatics team at MRC Harwell, a whole-genome bisulphite sequencing pipeline was developed. Analyses of the large amount of sequence data are ongoing.

The pipeline consists of 3 parts: i) data processing and quality control; ii) statistical methods for detecting differential methylation; and iii) the annotation of differentially methylated regions (DMRs) alongside the interpretation of the DMRs in a biological context. In this thesis, I describe the preliminary analyses of sequence data from the three genotypes and restrict the discussion primarily to methylation in the CpG context. Bisulphite-treated DNA converts cystosine residues to uracil (U), which are copied as thymines during the strand synthesis stages, but leaves the 5-methylcytosine residues intact (C). By aligning the converted DNA sequence reads to the reference genome it is possible to determine the percentage of Cs and Ts aligning to any given cytosine location in the genome. Therefore, the percentage methylation refers to the number of C bases aligning to a given cystosine location in the genome. Sequencing aligners typically have two approaches for aligning bisulfite treated short-reads. Wildcard aligners replace all Cs in the reference genome to a wildcard letter Y. Any Cs and Ts within the treated reads can then align to any Ys within the reference. This

increases the sequencing coverage at the expense of specificity. This method will introduce a methylation percentage bias towards higher methylation levels. By contrast, three-letter aligners replace all Cs to Ts in the bisulphite reads and work on a reduced genomic alphabet of As, Gs and Ts. This way mapping reads to the genome is simplified and faster than the former approach; however, a lower read coverage is achieved due to the large percentage of reads discarded because of ambiguous alignment positions. For the initial analysis of my data we chose Bismark, a three-letter aligner (Krueger and Andrews, 2011) due to the speed and well-validated results (Angermueller et al., 2016; Hanna et al., 2016).

6.1.3 Differentially Methylated Region (DMR) Detection

Global trends of DNA methylation can be simply achieved by calculating the percentage of Cs at any given cytosine. However, to reduce false positives other factors must be accounted for, such as the sequence quality scores at the loci alongside the likelihood that the C is methylated; therefore many methods have been developed to statistically model the methylated and unmethylated allele distributions and filter the ambiguous methylated sites. After global analysis of the methylated levels the next step is to identify differentially methylated regions (DMRs) between case and control groups in this instance XX, XY and *Gadd45g*$^{-/-}$, making 6 individual comparisons. Typically, DMRs in vertebrates range from 100s to 1000s of bases but there are instances of individual C methylated CpGs that resulted in changes to gene expression and disease (Nile et al., 2008). The majority of methylation callers produce a percentage methylation status at every differentially methylated cytosine (DMC) and DMRs are produced from these results. In most cases a simple t-test, Fischer's exact test or Wilcoxon rank-sum test is used to determine methylation levels between two cases. Spurious methylation levels can be detected due to the presence of SNPs being called as a methylated base, misplaced alignment reads or poor quality sequence. Therefore other methods include modelling the methylation levels to determine significant methylation changes including

DMRFinder and IMA (Bock, 2012). Recently, more sophisticated algorithms have been developed to take into account specific sequence features including the relationship between CpGs and their neighbouring sequence context. Like most next generation sequencing technologies, the read coverage and depth is most influential when applying confidence to a sequence statistic. Here we use MethylKit to detect differentially methylated cytosines (Akalin et al., 2012).

6.1.4 Aims and Hypotheses

As PBAT has become a leading method of producing whole genome bisulphite sequencing libraries I aim to produce libraries suitable for sequencing from sorted cells of *Sf1-EGFP* gonads from wild-type XY and XX and XY *Gadd45g*$^{-/-}$ embryos.

After these libraries are generated, they will be sent for next generation sequencing, analysis of which will allow me to deduce if there are differences between the methylomes of XY and XX gonadal somatic cells at 18 ts, the height of *Sry* expression. I also aim to investigate if the absence of *Gadd45g* causes any differences in the methylome of XY gonadal somatic cells.

6.2 Results

6.2.1 Generation of WGBS libraries from gonadal somatic cells using PBAT

Three PBAT libraries of each genotype were generated using gonadal cells sorted from *Sf1-EGFP* embryos. Each library used the cells collected from a single embryo. Following the protocol outlined in section 2.7.4, the quality of each library was checked using a High sensitivity DNA chip on a 2100 Bioanalyzer (both Agilent). Successfully generated libraries showed a range of DNA fragments from 200 – 2000 bp, with a peak at 500 bp (Fig 6.2). Concentrations for the libraries were then calculated

using the KAP library quantification kit (KAPA Biosciences) (Table 6.1). Only libraries showing a concentration higher than 500pM on either the Bioanalyser traces or the KAPA qPCR assay were selected. Three libraries from XY wild-type and XY *Gadd45g*$^{-/-}$ and two libraries from XX wild-type were then submitted for next generation sequencing using an Illumina Hiseq 2500 at the Babraham Institute. All samples were barcoded and split across 6 different lanes. Two XY wild-type samples, JM2 and JM3, had two sequencing cycles run due to the low sequencing yield obtained in the first run.

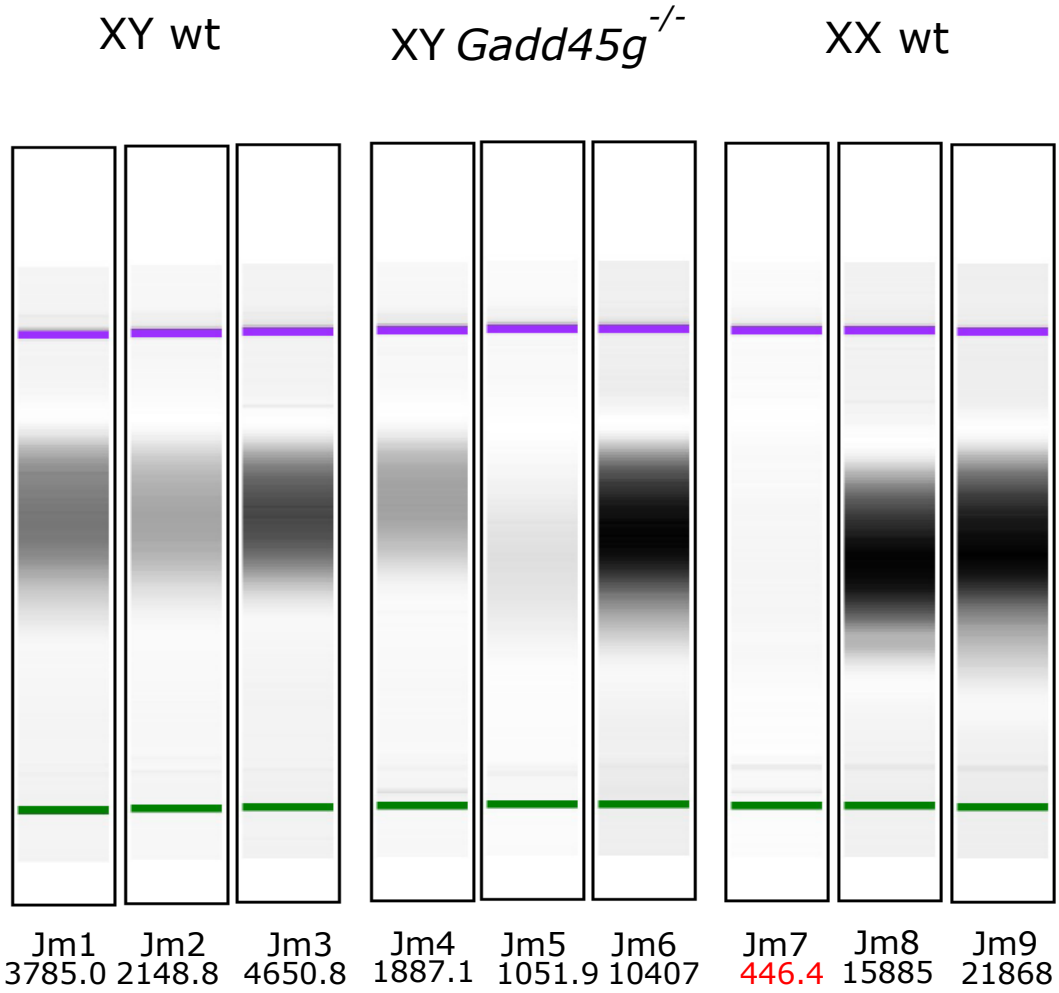


Figure 6.2: Bioanalyser gel images from generated PBAT libraries.
Gel images from bioanalyser from the indicated PBAT libraries. Below are the names assigned to each library along with their calculated concentration in pM. Library Jm7 (highlighted in red) did not produce a visible band and showed a low concentration. Though Jm5 also did not produce visible bands, a sufficient concentration was detected.

Table 6.1: Concentration of PBAT libraries according to KAPA qPCR.

Sample	pM	nM	Average fragment size (bp) - based on Bioanalyser	Size adjusted concentration (nM)	Sample info
1	9218.51	9.22	612	6.81	JM1_xy_wt_PBAT
2	4275.72	4.28	593	3.26	JM2_xy_wt_PBAT
3	43.86	0.04	590	0.03	JM3_xy_wt_PBAT
4	4213.59	4.21	648	2.94	JM4_xy_Gadd45gKO_PBAT
5	2279.08	2.28	441	2.34	JM5_xy_Gadd45gKO_PBAT
6	29534.51	29.53	522	25.57	JM6_xy_Gadd45gKO_PBAT
7	10.69	0.01	459	0.01	JM7_xx_wt_PBAT
8	47605.14	47.61	449	47.92	JM8_xx_wt_PBAT
9	57057.04	57.06	487	52.96	JM9_xx_wt_PBAT
10	1.09	0.00			Blank

The KAPA library quantification kit system (KAPA Biosystems) was used to confirm that the concentrations of the libraries were suitable for next generation sequencing. The XX wild-type

6.2.2 Analysis of WGBS and alignment

Bioinformatic analysis of the PBAT sequencing was performed using an in-house pipeline, developed to ultimately detect DMRs between the three genotypes (Table 6.2). A detailed description of the pipeline can be found in Figure 6.3. Each library generated 13 to 20 million sequence reads. The ends of the reads were assessed for their quality using FastQC, then trimmed to remove the adaptor sequences and aligned to a Burrows-Wheeler transformed C to T converted mouse genome. Sequence alignment and mapping was conducted three times. The paired-end illumina reads were first mapped onto the genome and the unmapped reads from each pair were collected as two single-end files. The single-end reads were each aligned and mapped separately onto the mouse reference genome. The three different alignments (BAM files) were merged into one single BAM and sorted for further analysis. The unmapped reads were subsequently re-mapped to the genome in single-end mode. The alignment achieved an overall average read-depth of either 6 or 7x for all genotypes (Fig 6.3). The low read-depth could be explained by read contamination, PCR bias and unaligned reads. The number of unaligned reads was high, ranging from 30-50%. However, aligning the paired-end reads in single-end mode more than doubled the mapping efficiency and all alignments were in unique mapping mode. Therefore, the aligned reads could not map anywhere else in the genome. In addition, a number of reports indicate that bisulphite-converted sequences will typically only align to ~40% of the genome and for most samples the pipeline achieved higher than this (Miura et al., 2012). Therefore, I was confident that the alignments reflect methylation levels in the different genotypes. The data includes the alignments (BAM format), coverage and methylation (BedGraph format) and differentially methylated sites (text format).

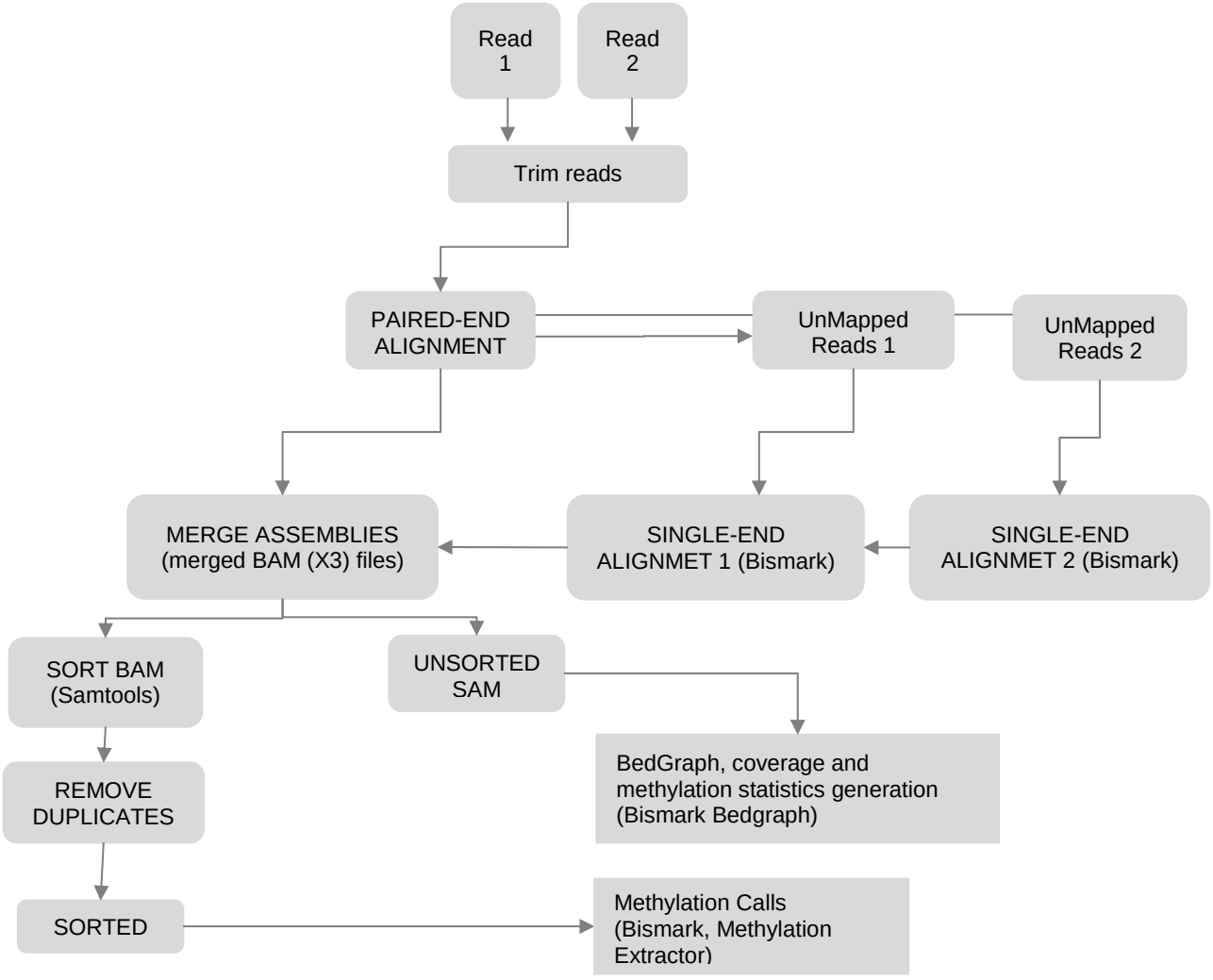
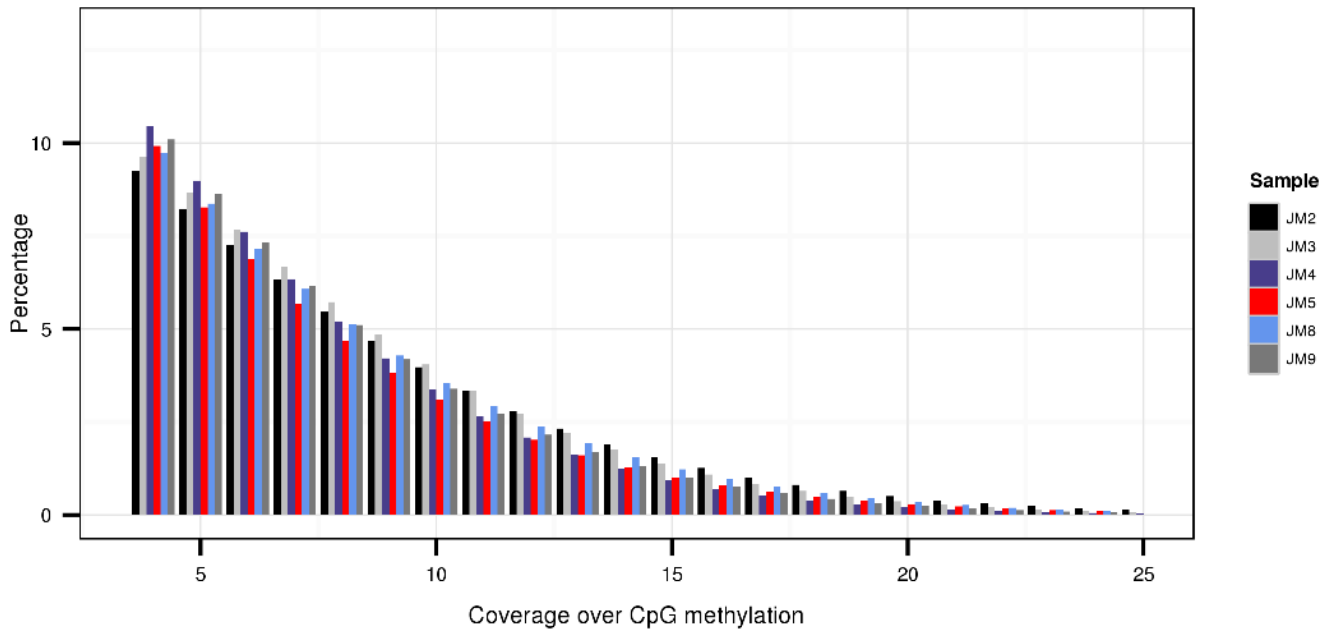


Figure 6.3: A schematic computational pipeline of PBAT Bisulphite Sequencing data
Computational pipeline used to generate alignments from WGBS data. Sequence reads are first trimmed to remove adaptors and low confidence bases, then aligned as paired-end reads to a Burrows-Wheeler converted mouse genome. Unmapped reads are then used in a single-end alignment, before merging the three alignments into a single file. Coverage and read depth are then calculated, followed by statistical tests for DMRs.

Table 6.2: Summary of Pipeline results and CpG methylation percentages

Samples	Alignment	Total number of reads	% Uniquely mapped reads	% C methylated in CpG context	% C methylated in CHG /CHH context
JM1:XY	Paired-end	174926369	31.8	74.4	9.8
	Single-end	119241435	41.6	73.0	14
	Single-end	119241435	34.3	70.3	6.9
JM2:XY	Paired-end	178228400	17.2	72.7	7.8
	Single-end	147523355	60	72.1	11.5
	Single-end	147523355	53.8	70.1	4.2
JM3:XY	Paired-end	196632901	29.4	73	9.5
	Single-end	138776961	44	71.7	12.5
	Single-end	138776961	38.9	68.9	2.8
JM4:XY_GAD	Paired-end	159961248	29.5	72.3	9
	Single-end	112849556	47.8	71.2	11
	Single-end	112849556	39.6	68.5	3.9
JM5:XY_GAD	Paired-end	131888687	39.8	69.8	10.1
	Single-end	79396418	34.4	68.8	12.8
	Single-end	79396418	36.6	65.9	7.1
JM8:XX	Paired-end	168207595	35.5	72.0	5.1
	Single-end	108571170	48.6	71.7	11.7
	Single-end	108571170	23.6	67.6	4.5
JM9:XX	Paired-end	155885765	36.9	71.4	10.7
	Single-end	98289480	37.5	72.2	11.9
	Single-end	98289480	29.9	67.5	8.8

A



B

JM1:XY	JM2:XY	JM3:XY	JM4:XY_GAD	JM5:XY_GAD	JM8:XX	JM9:XX
7	7	6	6	6	6	6

Figure 6.4: Read coverage and average read depth of sequence alignments
A) Depth of coverage of reads across CpG sites from the 5 sequencing libraries. Each coverage map is averaged into the average read depth in (**B**).

6.2.3 Dynamics and global trends of DNA methylation in mouse embryonic gonadal somatic cells

Whole methylome analysis was conducted using two different software packages; Bismark and MethylKit. Bismark was used to calculate the coverage and percentage of methylation and generated BED files for visualization, while MethylKit identified differentially methylated cytosines between the genomes. The number of cytosines analysed ranged from 1.7 to 2.5 billion in all genotypes, where approximately 65-75% were in the CG context. The level of methylated CpG sites was similar in all genotypes (Table 6.2) and between all replicas regardless of genotype (data not shown). Biological replicas within a genotype were more correlated than between genotypes, however the hierarchical clustering did not show distinct genotype groups. These results concur with the similar level of methylation seen between genotypes and suggest the quantity of differentially methylated regions may be low. For all samples, we detected a modest amount of DNA methylation in non-CG contexts (CHG and CHH, where H = A, C, T), comprising around 10% of all methylated cytosines. This supports previous assumptions that mammalian DNA occurs prevalently in the CpG context (Yan et al., 2011).

After the initial analysis of the overall trends in DNA methylation between the genotypes, MethylKit was used to identify differentially methylated cystosines (DMCs). MethylKit outputs the percentage methylation for every base covered by sequence reads. DMCs were only called at sites with at least four reads. MethylKit applies a logistic regression model to compare the number of Cs in two groups at any given cystosine. If the methylation proportion is significantly different between the two groups, the base would be classified as differentially methylated and given a p-value. MethylKit corrects for multiple testing by applying the standard false discovery rate (FDR)-based method (Benjamini-Hochberg) (Hebestreit et al., 2013). The standard identification of a differentially methylated cytosine is typically; C bases that have a q-value (FDR) of 0.01 and percent methylation

difference higher than 25% (Akalin et al., 2012). Table 6.3 shows the number of differentially methylated cytosines for each pairwise comparison of the three different genotypes. For any given comparison, we found there were <5 differentially methylated bases, at the standard q-value cut-off of 0.01, indicating that practically no DMCs were detected between the genotypes. More, but still relatively few DMCs were found at higher q-values, table 6.4 shows the list of DMCs and their location at q=0.05.

6.2.4 The methylation state around known sex determination genes

To investigate the integrity of the sequence alignment and the absence of DMRs at sex-determining genes, the regions surrounding known sex-determining loci was inspected manually. Specific genes showed hypermethylation throughout their gene body, excluding the transcription start-site region, which typically looked hypomethylated compared to the gene body, regardless of genotype. However, a small number of genes seemed highly hypomethylated throughout their entire gene body and corresponding transcription start-site in all three genotypes (Fig 6.5). Upon further inspection we found *Fgf9* was 70% methylated at 100% CpG sites, whilst *Sox9* was 20% methylated at 80% CpG sites. Using manual inspection of the alignments for all genotypes, we observed various other sex-determining genes with the same gene body hypo- or hypermethylation. The genes considered hypermethylated throughout the gene body are *Fgfr2*, *Sf1*, *Fgf9*, *Map3k4*, *Gata4*, *Jmjdla*, *Wt1* and *Wnt4*. The genes which were inspected showing hypomethylation were *Sox9* and *Foxl2*. Naturally, there were other genes we could not classify and the analysis was not exhaustive. A more robust, global analysis is required for future investigations of these datasets.

Table 6.3. Number of differentially methylated cytosines detected

Samples	Hyper & hypomethylation	Difference=25, qvalue=0.05	Difference=25, qvalue=0.1	Difference=25, qvalue=0.5	Difference=25, qvalue=1
XY vs XX	Hyper	2	7	626	141553
	Hypo	0	1	308	109698
XY vs GAD	Hyper	16	20	248	143199
	Hypo	0	0	64	97333
GAD vs XX	Hyper	33	86	4259	539142
	Hypo	5	17	2305	1002630

Number of differentially methylated individual CpG sites detected between each of the genotypes analysed. Lowering the stringency of the statistical test by increasing the q value, raises the number of DMCs but lowers the significance.

Table 6.4. Hypo- & hyper-differentially methylated bases between XY & XX, with a methylation difference >25% in the XY mice and a q-value = 0.05, and the nearest gene to each site

	chr	start	end	strand	pvalue	qvalue	meth.diff	Nearest Gene (distance)*
hypermethylated: distance 25, q-value=0.05	chr15	46708326	46708326	-	5.94296E-09	0.017408906	50.30959752	*4930548G14Rik* (+82.8kb)
	chr15	86621448	86621448	-	7.89599E-09	0.017408906	96.77419355	*Tbc1d22a* (+406,990)
hypermethylated: distance 25, q-value=0.1	chr12	24159970	24159970	+	4.81662E-08	0.053097916	63.63636364	*9030624G23Rik* (-63,003), *Gm16372* (+333,686)
	chr12	109059053	109059053	-	1.39215E-07	0.090226913	73.26388889	*Begain* (+9,164), *Wdr25* (+164,737)
	chr13	60664872	60664872	+	1.63693E-07	0.090226913	62.14285714	*Dapk1* (+62,926), *BC051665* (+121,492)
	chr13	112578307	112578307	-	1.56823E-07	0.090226913	56.8877551	*Il31ra* (+2,355), *Il6st* (+103,300)
	chr14	65766600	65766600	+	4.28656E-08	0.053097916	63.5	*Pbk* (-39,237), *Scara5* (+100,198)
	chr15	46708326	46708326	-	5.94296E-09	0.017408906	50.30959752	*4930548G14Rik* (+82.8kb)
	chr15	86621448	86621448	-	7.89599E-09	0.017408906	96.77419355	*Tbc1d22a* (+406,990)
hypomethylated: distance 25, q-value=0.1	chr14	25411319	25411319	+	7.85621E-08	0.069284785	-100	*Zmiz1* (-47,905), *Rps24* (+920,639)

6.3 Conclusions and Discussion

6.3.1 Aims revisited

This chapter has explored the methylome of gonadal somatic cells through whole genome bisulphite sequencing using libraries generated by post-bisulphite adaptor tagging. Whole genome bisulphite sequencing libraries were successfully produced from cells purified by FACS from XY wild-type, XX wild-type and XY *Gadd45g*$^{-/-}$ 11.5 dpc (18 ts) gonads by using PBAT.

These libraries were then analysed to detect any differences between the methylomes of wild-type XY and XX gonadal somatic cells, as well as investigate if there were observable differences in the methylome of XY gonadal somatic cells in the absence of *Gadd45g*.

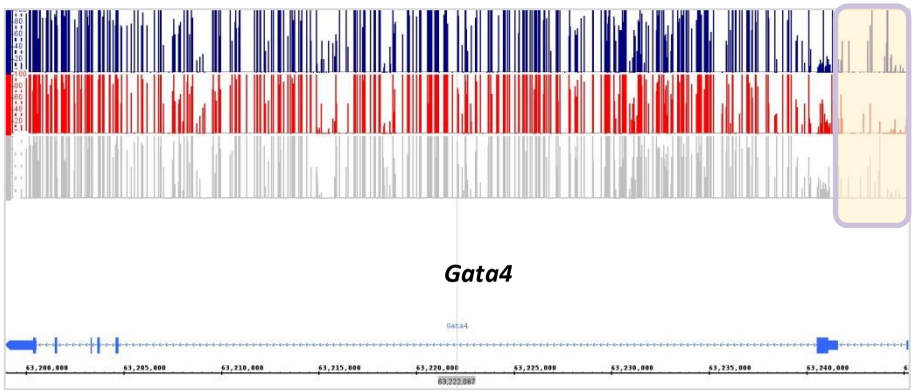
6.3.2 No detectable differences in the methylomes of XY, XX or XY *Gadd45g*$^{-/-}$ somatic gonadal cells at 18 ts

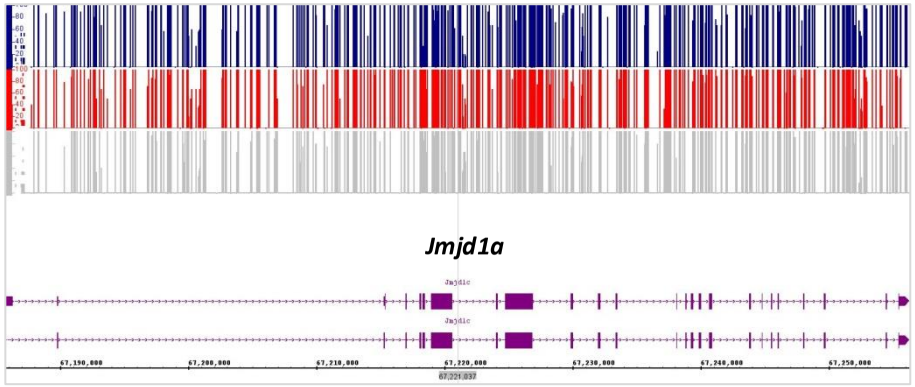
The first signs of *Sry* expression occur in the centre of the gonad at around 10.5 dpc. After 24hours, *Sry* is found throughout the XY gonad and has begun to upregulate *Sox9* expression (Kashimada and Koopman, 2010). Despite the divergence between the XY and XX gonad at this stage in expression of various genes (Nef et al., 2005; Jameson et al., 2012), no major differences were found in the methylation profiles of XY, XX and XY *Gadd45g*$^{-/-}$ gonadal somatic cells.

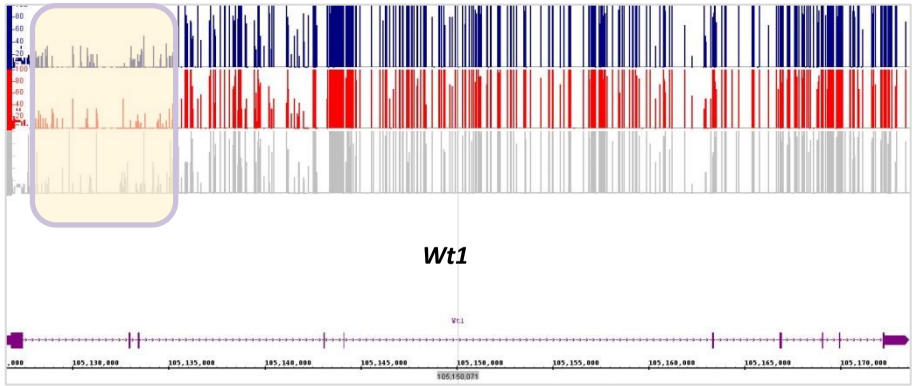
Interestingly, many of the genes involved in sex determination, both pro-testis and pro-ovarian, showed demethylation of their promoters and transcription start sites in all three genotypes. Hypomethylation in these areas is usually associated with nucleosome-depleted regions, areas of chromatin which are not tightly bound. As a result they can more easily be affected by transcription factors, both positively and negatively (Jones, 2012). Given the mutually antagonistic relationship

between the transcription factor networks which promote the Sertoli or granulosa cell fate, it is likely that transcription factor occupancy plays a major role in establishing the identity of cells at this stage. Other epigenetic factors such as chromatin compaction and histone modification may also be differential at this time, but both of these possibilities will require further investigation. Binding of transcription factors to promoter regions is associated with the maintenance of the hypomethylated state by inhibiting the action of the *de novo* methyltransferases (Smith and Meissner, 2013). Methylated areas around transcriptional start-sites are associated with long term gene silencing and often occur in terminally differentiated cells. In mouse embryonic stem cells differentiation is associated with the silencing of pluripotency genes such as *Oct4* and *Nanog*. These genes are first repressed through transcription factor binding, followed by the placement of repressive histone marks and finally DNA hypermethylation around the TSS (Smith and Meissner, 2013). A similar series of events could potentially be occurring in gonadal somatic cells, though at the time of this particular experiment, *de novo* methylation had not taken place at the genes which were to be silenced. A more accurate picture of what state of differentiation the cells are in could be established by investigating various histone marks around sex determining genes at this time through chromatin immunoprecipitation. As morphological differences begin to occur in the gonads, a more stable transcriptome may be established, leading to the long term silencing of genes which are associated with promoting the opposite gonadal fate. Investigating the methylome in these same cell lineages at later stages would help confirm if this is the case.

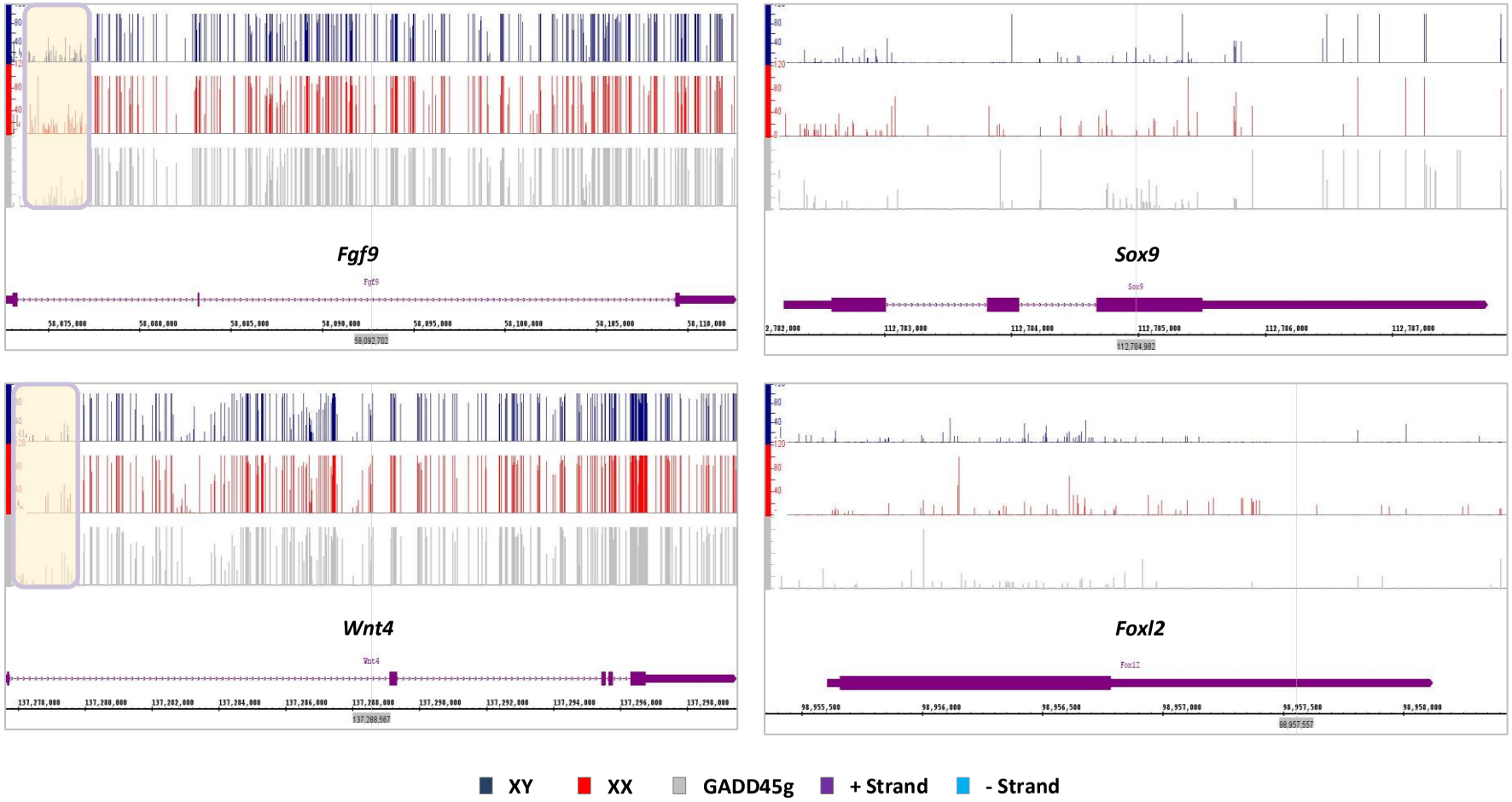
This similarity of the epigenetic states between XY and XX cells at this time could be indicative of the earlier bipotential state. This divergence in differentiation between the many differences in the transcriptome and the lack of differences in the methylome may explain why the time-window in







XY XX GADD45g + Strand - Strand


Fgf9
Sox9
Wnt4
Foxl2
XY
XX
GADD45g
+ Strand
- Strand

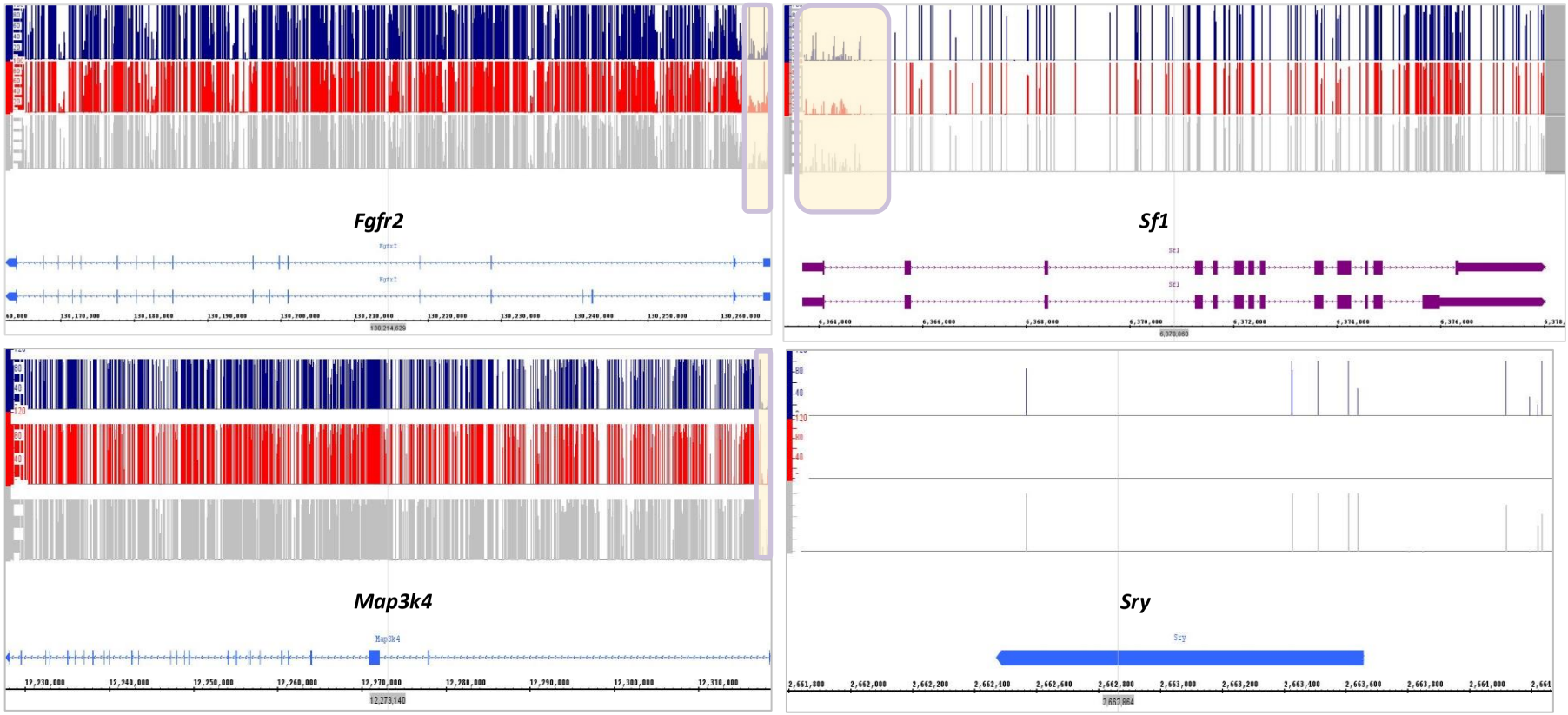


Fig 6.5. Examples of Sex Determining genes and methylation in the surrounding area
Tracks showing the percentage of methylated cytosines at each CpG around known sex determination genes. Areas in boxes correspond to promoter regions. Length of bars represents the % methylation at each CpG site across the gene. While the majority of genes involved in sex determination were found to be hypermethylated across their gene body and relatively hypomethylated around their transcriptional start site; *Sox9* and *Foxl2* showed hypomethylation across their entire coding regions. Reads from XX samples did not align to *Sry*.

which *Sry* must be expressed for proper testis formation is so precise. The absence or failure of SRY and SOX9 to act on epigenetically open states results in the establishment of the ovarian pathway.

Though not seen in the *Gadd45g* knockout, aberrant methylation has been shown to cause mammalian DSDs (Jeong et al., 2016), and may be responsible for the development of gonadal tumours later in life (Kalfa et al., 2008). PGCs located within the gonads develop sex specific differences in DNA methylation at later stage in development (Kobayashi et al., 2013). It is therefore likely that a methylation profile specific for Sertoli or granulosa cells is established at a later stage. A delay in the commitment of the epigenome of gonadal somatic cells towards a testicular or ovarian fate may help explain why delays in *Sry* expression lead to sex reversal (Bullejos and Koopman, 2005; Bogani et al., 2009; Warr et al., 2012). If a cell has committed to an ovarian fate through its epigenome, then it is unlikely that SRY would be able to access its target DNA sequences and activate the testicular pathway. The plasticity retained in the adult gonad, demonstrated by the supporting cell reprogramming associated with the loss of *Foxl2* or *Dmrt1* (Uhlenhaut et al., 2009; Matson et al., 2011), may also be incompatible with the idea of a firmly established methylome. Both of these intriguing possibilities will need to be confirmed by further WGBS at later embryonic stages, such as 14.5 dpc when gonads show distinct morphological differences, or in adult gonadal supporting cells.

The average read-depth achieved from PBAT generated libraries can vary. Libraries which produce up to a 20-fold read depth have been achieved (Miura et al., 2012), though often read-depth is more similar to those described here (Kobayashi et al., 2013; Sirane et al., 2013). Though this read depth is sufficient to determine large scale differences in methylation, such as CpG island hyper/hypomethylation, smaller scale differences in CpG methylation may be more difficult to detect. Non-CpG island methylation is often more dynamic and tissue specific (Jones, 2012), as seen in the

Sry promoter (Nishino et al., 2004). Smaller differences in methylation can be detected through targeted BS-seq. Attempts were made to investigate the methylation status of regions surrounding known sex-determining genes. However, due to time constraints only a very low number of reads was produced. Further targeted investigation will be required to validate the data found by analysing the sequencing of the WGBS libraries.

Sequencing data can also be improved through employing different bioinformatics techniques. Sequence alignment is required to be as accurate and deep as possible to prevent false-positive calls using DMR detection software. In this study, it was observed that aligning reads which had not aligned in a paired-end manner in single-end mode added approximately 30% to the total mapped reads. As a result future alignments will only employ single-end mode of alignment, with the aim of decreasing mis-aligned reads, increasing sequence depth and improving the significance of any DMRs. Alternative methods of effectively detecting DMRs can also be used. In this study the q-value calculates the multiple testing-corrected false discovery rates using every base in the genome. An alternative approach uses a tiling and sliding window method. Rather than the rest of the genome, the local genomic area of a given size will be used to calculate the FDR (Lister et al., 2009), allowing further biological inference to be deduced.

PBAT has previously been used to generate libraries from single cell samples. When multiple libraries are made, which can be achieved within a single plate, sequencing barcodes can allow for the differentiation of which sample is the source of each read (Smallwood et al., 2014; Schwartzman and Tanay, 2015). Through this method, heterogeneity across samples can be measured. As *Sf1-EGFP* can be expressed in precursors to both supporting cells and steroidogenic cells (Gierl et al., 2012; McClelland et al., 2015), this approach can be used to further separate the epigenetic states of these

two cell lineages. In XY gonads, *Sry* expression begins at different times depending on the location of the cell within the gonad (Bullejos and Koopman, 2001). Rather than generate an average of the epigenetic state of the gonad at a particular stage, it may be possible to profile the advancement of epigenetic changes as *Sry* expression continues.

6.3.3 *Sox9* and *Foxl2* are hypomethylated across their gene bodies

Sox9 and *Foxl2* are essential genes for testicular and ovarian determination respectively. Both promote their respective fates as well as inhibiting the other (Kim et al., 2006). Interestingly, WGBS revealed that both genes were extensively hypomethylated across the entirety of their coding region, as opposed to other sex determining genes which were only hypomethylated across their transcriptional start sites. *Sox9* expression is detectable only in wild-type XY gonads at this stage, while *Foxl2* is found in XX gonads (Jameson et al., 2012).

Though a large proportion of CpG islands do not reside in gene bodies, both *Sox9* and *Foxl2* contain CpG islands across the length of their coding regions (Stothard, 2000). The role of gene body methylation has been suggested to correlate with levels of transcription in differentiated cells (Lou et al., 2014), though it has also been argued that methylated DNA can inhibit transcription through inhibition of RNA Pol II elongation (Lorincz et al., 2004). Given that differential transcription of these two critical genes has been observed at this time, there must be controlling molecular mechanisms acting upon them. The nature of this mechanism does not appear to be DNA methylation, which is often the final epigenetic mark established during differentiation (Smith and Meissner, 2013). It is therefore critical that histone modifications around these two genes at this time are analysed using chromatin immunoprecipitation to assess if this could explain the differential expression. The possibility also exists that the only differences at this time are maintained by the differential occupancy of transcription factors, established with the onset of SRY expression.

As the methylomes of all three genotypes do not appear to differ, they may be similar to those found in the early-stage bipotential gonad. Low levels of *Sox9* expression are detected in both XX and XY gonads from 10.5 dpc (Ono and Harley, 2013). Though gonadal *Foxl2* expression is not observed before 11.5 dpc, the transcriptome of bipotential supporting cells has been described as having a "female bias in the underlying program" (Lin and Capel, 2015). Gene body hypomethylation may therefore leave these two genes transcriptionally inactive (or with low activity) but still amenable to dynamic activation or further repression by other transcription factors or epigenetic events. To fully answer this question, however, much more epigenomic data will be required at other stages of sex determination.

Chapter 7: General Discussion

7.1 Overview

This thesis has described experiments aimed at examining the role of GADD45γ in testis determination, in an effort to further understand how it ensures the timely expression of *Sry*. Since the discovery of *Sry*, various genetic pathways have been shown to regulate its expression during a critical time window within the supporting cell lineage. Rather than all acting through a single direct regulatory module, these various pathways appear to impact in different ways on the *Sry* locus to ensure its appropriate expression, although molecular details are often unclear. This thesis has focused on one particular pathway, or possibly pathways, involving GADD45γ.

Data presented provide an evaluation of a *Gadd45g* overexpression transgenic line using the BAC *E15-Gadd45g*, primarily to determine the suitability of the BAC for further functional studies. Analyses include determining the increased levels of expression the BAC has induced as well as the timing and spatial distribution of expression in gonadal tissues in which the transgene is expressed. In addition to its ability to rescue embryonic gonadal sex reversal in the *Gadd45g* knockout, the BAC has also been tested for its ability to rescue the sex reversal phenotype of B6.Y^{POS}. Also presented are efforts to generate a modified version of the BAC including the sequence for a combination His/ FLAG epitope tag using Red/ET recombineering.

This thesis also described methodologies used to generate libraries for whole genome bisulphite sequencing (WGBS) from embryonic gonadal somatic cells, including a successful attempt using post-bisulphite adapter tagging (PBAT). These PBAT libraries were generated in order to examine the methylome of the bipotential gonadal soma at the sex-determining stage (but prior to overt differentiation) and to test the hypothesis that GADD45γ functions in gonad development, at least partly, by establishing gene-specific methylation profiles via its putative role in active DNA

demethylation. Sequencing results are also presented from PBAT from XY wild-type, XX wild-type and XY *Gadd45g*$^{-/-}$ gonadal somatic cells.

7.2 Key Findings

7.2.1 *E15-Gadd45g* is a functional overexpression construct

Two distinct BACs containing the *Gadd45g* coding region were used to generate transgenic overexpression mouse lines, with a view to identifying a BAC containing sequences sufficient for *Gadd45g* functionality. Such a transgenic line could be used for further functional studies of GADD45γ. One particular transgenic line, *E15-Gadd45g*, was shown to rescue the sex reversal phenotype of *Gadd45g*$^{-/-}$, B6.YAKR *Gadd45g*$^{+/-}$ and B6.Y^{POS}. The expression derived exclusively from the transgene was examined across the development of the gonad, using an approach in which the BAC was bred on to a *Gadd45g* -/- background. Through this approach, it was found that *Gadd45g* expressed from the transgene is stronger, begins earlier and lasts longer than expression from the endogenous locus. It was also found that expression from the BAC is sufficient to rescue sex reversal in the homozygous *Gadd45g* knockout. Levels of expression were found to be approximately 5 fold higher than wild-type expression at 18 ts, the peak of endogenous *Gadd45g* expression. Interestingly, levels of *Sry* transcript at 18 ts did not appear to increase significantly compared to non-transgenic gonads. The increased levels and duration of transgenic *Gadd45g* may provide answers to how the BAC rescues sex reversal in various sex reversal lines, but deducing the mechanisms behind this will require further investigation that could not be completed within the time-scale of this thesis.

As well as offering a potent tool for the mammalian sex determination field, other roles of *Gadd45g* mean that other distinct areas of developmental biology, cancer biology and epigenetics may benefit from this *in vivo* overexpression line.

7.2.2 *E15-Gadd45g* is amenable to modification through Red/ET mediated recombineering

Chapter 4 describes the process of generating an epitope tagged version of *E15-Gadd45g*, as previous efforts at Harwell and elsewhere to generate antibodies specific to GADD45γ, which are capable of detecting endogenous protein, had been unsuccessful. The epitope tag sequence was successfully inserted into the *Gadd45g* coding region with no other detectable recombination events. Despite a large number of microinjections into 1-cell stage embryos, a transgenic mouse line containing this modified BAC was not generated. This experiment highlights the difficulties in using BACs as a means of producing transgenic lines. The viscosity of BAC DNA preparations appears to cause difficulties for microinjectors, although this may vary from user to user. While BACs are powerful tools, the low chance of successfully generating founders, especially enough to exclude position-effects upon integration, hinders their application.

An alternative to the use of BAC recombineering to introduce tags into proteins of interest is the tagging of endogenous genes. Small tags are unlikely to disrupt gene function if inserted at the ends of protein coding sequences, but larger tags may well do so. Whilst larger tags, such as GFP or Cre, can be inserted by replacing the endogenous sequences, the tagged allele is likely to be non-functional and mice carrying it are likely to be haploinsufficient, which may interfere with subsequent genetic analyses. Maintaining a functional allele requires the use of an internal ribosome entry site (IRES) after the coding region or the 2A self-cleaving peptide approach (Yang et al., 2013). CRISPR-

Cas9 genome editing is greatly increasing the efficiency of such approaches and it may be possible to avoid the use of ES cells and modify genes by zygote injection.

7.2.3 The methylomes of 11.5 dpc (18 ts) somatic cells from XX, XY and XY *Gadd45g*$^{-/-}$ gonads show no major differences

The developing gonad at 11.5 dpc is a heterogeneous collection of cell-types, including germ cells, endothelial cells, coelomic epithelial cells and precursors of the somatic supporting and steroidogenic lineages (Warr and Greenfield, 2012). This heterogeneity likely makes whole gonad samples unsuitable for epigenomic or transcriptomic profiling. This issue was confronted by using a well-known transgenic reporter line previously used to FACs sort gonadal somatic cells, *Sf1-EGFP* (Nef et al., 2005; Jameson et al., 2012). The enrichment for SF1-postive cells greatly reduces cellular complexity, but does not eliminate it (Gierl et al., 2012; Jameson et al., 2012; McClelland et al., 2015). However, technical issues concerning the use of the cell sorter hindered progress for a long period of time. Once this issue was resolved, significant challenges were encountered in the attempts to produce bisulphite sequencing libraries from small numbers of cells. Multiple displacement amplification could be used to generate the large amounts of bisulphite-converted DNA required for a WGBS library. However, the nature of amplification means that reads of a particular locus can be preferentially amplified. This produces a library in which reads are biased to towards interpretation of a region as hypomethylated or hypermethylated.

In response, libraries were generated using post-bisulphite adaptor tagging (PBAT). Though initially samples from three individuals of the three different genotypes were used to produce the libraries, only two XX wild-type libraries were successfully made. The other libraries were submitted for next generation sequencing.

Following the completion of sequencing, alignments were produced against the MM10 mouse sequence using BisMark and MethylKit. This analysis did not reveal any major differences between the methylomes of XX, XY and XY *Gadd45g*$^{-/-}$ gonadal somatic cells. The only known methylation difference in the gonad at this time is the tissue-specific hypomethylation of the *Sry* promoter. As this is a Y chromosome specific gene, alignment can be difficult due to the large amount of repeat regions within the Y chromosome resulting in a low number of reads around *Sry*. Lack of knowledge regarding other regions which would show a known methylation state prevented the use of such a region as a control against the sequenced libraries. The absence of clearly detectable differentially methylated regions (DMRs) between the sexes/genotypes could be attributed to the absence of such marks. However, the cellular heterogeneity of the SF1-positive cell population might obscure such regions particularly if they are hypermethylated in one cellular subset and hypomethylated in another. Ultimately, cellular heterogeneity might be overcome by the use of single cell approaches, but single-cell methylome analysis would be a formidable technical challenge.

If the absence of DMRs truly reflects the biology of the sex-determining methylome of XX and XY gonads, then we are forced to conclude that, at this stage of gonad development, and in this population of cells, differential gene methylation is not required for the commitment of somatic cells to the Sertoli or granulosa cell fate, and that GADD45γ is also not functioning to establish DMRs. Our search for the molecular basis of such cell fate commitment then switches to other epigenomic events, such as histone post-translational modifications (or marks) and transcription factor occupancy of chromatin at sex-determining loci. In respect of the latter, it is striking how 'plastic' the epigenome of the supporting cell lineage appears to be even in the differentiated adult gonad. Loss of FOXL2 from adult granulosa cells (Uhlenhaut et al., 2009) and DMRT1 from adult Sertoli cells (Matson et al., 2011) results in reprogramming of those cells to the opposite sexual fate and partial sex reversal of

gonadal tissues. These data strongly suggest the absence of layer-upon-layer of epigenomic 'lock-in' in adult supporting cells, perhaps by the recruitment of repressive chromatin complexes and inhibitory conformations, such that the loss of a single transcription factor can have profound effects on the maintenance of cell fate. Presumably, unoccupied loci re-initiate transcriptional activity, resulting in cellular sex reversal. This dominant role played by transcription factor occupancy might begin at the earliest stages of sex determination.

7.3 Future Work

7.3.1 A role for *Gadd45g* in active DNA demethylation remains theoretical

Results in this thesis have shown that the absence of *Gadd45g* does not appear to affect the methylome of gonadal somatic cells at the stage of sex determination. This does not exclude a role for GADD45g in DNA demethylation in other contexts. However, it does raise the question of the mechanisms by which GADD45γ affects testis determination. XX *Gadd45g*$^{-/-}$ mice are fertile and develop normal ovaries (Warr et al., 2012), indicating that despite its expression in both sexes *Gadd45g* is only essential for testis determination. The only known function of GADD45γ in testis determination is to ensure the timely expression of *Sry*, and this may be its only function. It would be possible to test this using an inducible *Sry* transgene (Hiramatsu et al., 2009) to drive early expression of *Sry* in *Gadd45g* -/- gonads in an attempt to rescue gonadal sex reversal. It has not been possible for our lab to source this transgenic line. Reduced phosphorylation of GATA4 in *Gadd45g* and *Map3k4* knockout gonads suggests a mechanism by which GADD45g promotes *Sry* expression, since GATA4 is a known *Sry* regulator (Warr et al., 2012). But it is not clear whether GATA4 is a direct target of MAP3K4-dependent phosphorylation; indeed, it is not even clear if reduced GATA4 phosphorylation is a direct consequence of signalling defects or a consequence of other changes to *Sry* chromatin.

Another potential role for GADD45γ may be the transmission of extracellular signals to ensure the expression of *Sry*. The best candidates for such a signal are *Igf1r* and *Insr*. Double knockouts of these genes result in a down-regulation of *Gadd45g* expression and lead to sex reversal as a result of a delay in *Sry* expression (Pitetti et al., 2013). A link might, therefore, be inferred between insulin signalling and *Gadd45g*. This could be probed by crossing *E15-Gadd45g* onto the double knockout background and testing for rescue of the sex reversal phenotype.

Further understanding of the function of GADD45γ will require a working antibody that can reliably detect endogenous protein with high specificity and sensitivity. This would allow the investigation of interacting partners of GADD45γ, regions of the genome GADD45γ is recruited to and spatiotemporal distribution of GADD45γ protein in the gonad during development. Though various efforts have been made by others and in this thesis, a working GADD45γ antibody is still unavailable. A gene-editing approach using CRISPR/Cas9 can be undertaken to add the sequence for an epitope tag to the endogenous *Gadd45g* locus (Yang et al., 2013). This approach might produce a more accurate pattern of expression compared to a transgenic approach, as outlined in chapter 3, but may not be sufficient if the reason for the difficulty in detecting endogenous GADD45γ is due to a high turn-over rate of the protein and relatively low levels of expression.

Alternatives to antibodies typically employ a common protein scaffold which can be edited through protein engineering to bind to specific target proteins. These include affibodies, based on bacterial protein A and so can be produced within prokaryotic cells; monobodies, based on the fibronectin type III domain (Koide et al., 1998) and Anticalans, based on human lipocalins (Skerra, 2007). These methods rely on a hypervariable loop region to bind to their target protein. Rather than produced

through biological means, large mutagenesis libraries are generated and tested against target proteins.

Though members of the GADD45 protein family have been shown to play a role in active DNA demethylation (Niehrs and Schafer, 2012), there were no apparent differences between the wild-type XY and *Gadd45*$^{-/-}$ XY methylomes at 18 ts. However, expression of *Gadd45g* continues in the gonad after levels of *Sry* have begun to decline (Warr et al., 2012). GADD45γ may therefore still be playing a role in sex determination after *Sry* expression, which would be masked in the *Gadd45g*$^{-/-}$ phenotype due to its role in inducing timely *Sry* expression. Given that *Gadd45g* is expressed in both XX and XY gonads (Warr et al., 2012), this could include a role in active DNA demethylation around genes which will become active following sex determination, during active differentiation. This possibility could perhaps be investigated by generating methylome data at later time points. One other virtue of performing comparisons of methylomes at later time-points is that methylation may become increasingly important, and given that thousands of genes are differentially expressed at such stages (as opposed to the 100s or so reported at 11.5 dpc) (Jameson et al., 2012) it would increase the chances of detecting bona fide DMRs. This would also give confidence that the analysis tools were working optimally for DMR identification. However, since the methylome datasets were generated quite late in the course of this study, there has not been time to perform additional experimentation or, indeed, a variety of alternative bioinformatics analyses. Such work is ongoing in the lab

Given that the timely expression of *Sry* is controlled by GADD45γ-induced activation of MAP3K4 and the MAP kinase cascade which activates GATA4 (Gierl et al., 2012; Warr et al., 2012), it is interesting that MAP3K4 has been shown to activate the acetyl transferase protein CBP in other contexts (Abell et al., 2011). It may be that the role of GADD45γ in affecting the epigenome of gonadal somatic cells

centres not on affecting the methylome, but instead on modifying the histone code through activation of MAP3K4. This suggests yet another reason to further investigate the modifications to histones surrounding sex determining genes using chromatin immunoprecipitation (ChIP) techniques such as ChIP-qPCR and ChIP-seq, both in wild-type and mutant tissue.

7.3.2 Establishing an epigenetic profile of gonadal somatic cells during sex determination

This thesis describes an analysis of the methylome of gonadal somatic cells from XY wild-type, XX wild-type and XY *Gadd45g*$^{-/-}$ gonads at 18 ts. In XY gonads this stage represents the peak of *Sry* expression, by this stage differences in the transcriptome of the XY and XX gonads are observable (Jameson et al., 2012). As the somatic cell methylome does not appear to differ between XY and XX, based on methods and analyses employed here, other epigenomic approaches and comparisons with other epigenomic data-sets will be required for a fuller picture of the epigenomic state of gonadal somatic cells at this time. Large scale projects such as ENCODE have utilised a number of different methods to analyse chromatin states, including DNase hypersensitivity sites analysis (DNase1seq), long-range interactions between chromosomal elements (HiC), locations of various negative and positive histone marks and levels of transcript from various cell lines (RNAseq) (Yue et al., 2014). Approaches such as this generate vast amounts of data, which take expertise and time to analyse and require extensive computing power to detect statistically significant results. Performing them in the context of *in vivo* tissue, dissected from embryos, also requires dealing with the problem of limiting amounts of tissue and cellular heterogeneity, although such methods are being developed. It will also be important to assess the order in which various epigenetic marks are established, meaning that

these kinds of studies will need to be completed over various time points before the significance of each epigenetic mark is deciphered.

As well as 'traditional' epigenetic marks such as DNA methylation and histone modification, emerging techniques which establish the 3-dimensional structure of chromatin will help decipher how enhancers and other trans-acting elements contribute towards differing transcriptional profiles (Belton et al., 2012).

Many epigenomic marks are described as correlated with, or associated with, various states of transcription rather than as having a causative role. The DMR upstream of *Sry* is an example (Nishino et al., 2004): it appears to be associated with *Sry* expression, although there is no direct evidence it is required. Using a nuclease deficient Cas9 (or 'dead' Cas9) linked to various histone-modifying enzymes, it is now possible to target epigenetic modifications to specific regions in the genome and investigate their direct effects on expression (Hilton et al., 2015; Thakore et al., 2015). A similar effort could be made to target regions with methyltransferase proteins or components of the demethylation pathway. However, it remains to be seen if components of the DNA methylation or demethylation complexes are compatible with this approach. Moreover, it might be necessary to have precise control over the timing and cellular/molecular specificity of such events. Such approaches are in their infancy.

7.3.3 Closing remarks

In summary, this thesis has described the generation of genetic tools to assist in elucidating the roles that GADD45γ plays in testis determination. This includes an over-expression construct which has been shown to rescue known modals of sex reversal. This construct was then used as the basis for an epitope-tagged version of *Gadd45g* for *in vivo* use, though no transgenic line has yet to be produced.

In addition, it reports on the examination of the potential role GADD45γ plays in DNA demethylation at the time of sex determination. While no obvious candidates for differentially methylated regions governed by *Gadd45g* were found, the striking similarity between the methylomes of XX and XY gonadal somatic cells provides a foundation for further epigenetic studies. Progress in this field will shed light on the testis- and ovary-determining epigenomes and how they are established and maintained, will illuminate the control of *Sry* expression in different genetic models and, perhaps, account for the plasticity of the supporting cell lineage. Finally, such studies will offer mechanistic insights into mammalian sex determination more broadly.

References

Abell, A.N., Jordan, N.V., Huang, W., Prat, A., Midland, A.A., Johnson, N.L., Granger, D.A., Mieczkowski, P.A., Perou, C.M., Gomez, S.M.*, et al.* (2011). MAP3K4/CBP-regulated H2B acetylation controls epithelial-mesenchymal transition in trophoblast stem cells. Cell Stem Cell *8*, 525-537.

Achermann, J.C., Domenice, S., Bachega, T.A., Nishi, M.Y., and Mendonca, B.B. (2015). Disorders of sex development: effect of molecular diagnostics. Nature reviews Endocrinology *11*, 478-488.

Achermann, J.C., Ito, M., Ito, M., hindmarsh, P.C., and Jameson, J.L. (1999). A mutation in the gene encoding steroidogenic factor-1 causes XY sex reversal and adrenal failure in humans. Nature Genetics *22,* 125-126.

Adey, A., and Shendure, J. (2012). Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing. Genome Research *22*, 1139-1143.

Akalin, A., Kormaksson, M., Li, S., Garret-Bakelman, F.E., Figueroa, M.E., Melnick, A., and Mason, C.E. (2012). MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biology *13*, R87.

Aleman, A., Adrien, L., Lopez-Serra, L., Cordon-Cardo, C., Esteller, M., Belbin, T.J., and Sanchez-Carbayo, M. (2008). Identification of DNA hypermethylation of SOX9 in association with bladder cancer progression using CpG microarrays. British Journal of Cancer *98*, 466-473.

Angermueller, C., Clark, S.J., Lee, H.J., Macaulay, I.C., Teng, M.J., Hu, T.X., Krueger, F., Smallwood, S.A., Ponting, C.P., Voet, T.*, et al.* (2016). Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nature Methods *13*, 229-232.

Auer, T.O., Duroure, K., De Cian, A., Concordet, J.P., and Del Bene, F. (2014). Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. Genome Research *24*, 142-153.

Azam, N., Vairapandi, M., Zhang, W., Hoffman, B., and Liebermann, D.A. (2001). Interaction of CR6 (GADD45gamma) with proliferating cell nuclear antigen impedes negative growth control. The Journal of Biological Chemistry *276*, 2766-2774.

Barreto, G., Schafer, A., Marhold, J., Stach, D., Swaminathan, S.K., Handa, V., Doderlein, G., Maltry, N., Wu, W., Lyko, F.*, et al.* (2007). Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. Nature *445*, 671-675.

Barrionuevo, F., Bagheri-Fam, S., Klattig, J., Kist, R., Taketo, M.M., Englert, C., and Scherer, G. (2006). Homozygous inactivation of Sox9 causes complete XY sex reversal in mice. Biol Reprod *74*, 195-201.
Barsoum, I.B., and Yao, H.H. (2010). Fetal Leydig cells: progenitor cell maintenance and differentiation. Journal of Andrology *31*, 11-15.

Bashamboo, A., and McElreavey, K. (2015). Human sex-determination and disorders of sex-development (DSD). Seminars in Cell & Developmental Biology *45*, 77-83.

Beadling, C., Johnson, K.W., and Smith, K.A. (1993). Isolation of interleukin 2 -induced immediate-early genes. Proceedings of the National Academy of Sciences *90*, 2719-2723.

Beil, J., Fairbairn, L., Pelczar, P., and Buch, T. (2012). Is BAC transgenesis obsolete? State of the art in the era of designer nucleases. Journal of Biomedicine & Biotechnology*2012*, 308414.

Belton, J.M., McCord, R.P., Gibcus, J.H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-C: a comprehensive technique to capture the conformation of genomes. Methods *58*, 268-276.

Beverdam, A., and Koopman, P. (2006). Expression profiling of purified mouse gonadal somatic cells during the critical time window of sex determination reveals novel candidate genes for human sexual dysgenesis syndromes. Hum Mol Genet *15*, 417-431.

Biason-Lauber, A., Konrad, D., Meyer, M., DeBeaufort, C., and Schoenle, E.J. (2009). Ovaries and female phenotype in a girl with 46,XY karyotype and mutations in the CBX2 gene. American Journal of Human Genetics *84*, 658-663.

Birk, O.S., Casiano, D.E., Wassif, C.A., Cogliati, T., Zhao, L., Zhao, Y., Grinberg, A., Huang, S., Kreidberg, J.A., Parker, K.L.*, et al.* (2000). The LIM homeobox gene Lhx9 is essential for mouse gonad formation. Nature *403*, 909 - 913.

Blaiklock, P. (2009). RestrictionMapper version 3 (www.restrictionmapper.org: Peter Blaiklock).
Bock, C. (2012). Analysing and interpreting DNA methylation data. Nature Reviews Genetics *13*, 705-719.

Boer, J.d., and Hoeijmakers, J.H.J. (2000). Nucleotide excision repair and human syndromes. Carcinogenesis *21*, 453 - 460.

Bogani, D., Siggers, P., Brixey, R., Warr, N., Beddow, S., Edwards, J., Williams, D., Wilhelm, D., Koopman, P., Flavell, R.A.*, et al.* (2009). Loss of Mitogen-Activated Protein Kinase Kinase Kinase 4 (MAP3K4) Reveals a Requirement for MAPK Signalling in Mouse Sex Determination. PLoS Biol *7,* e1000196.

Bor, Y.C., Swartz, J., Morrison, A., Rekosh, D., Ladomery, M., and Hammarskjold, M.L. (2006). The Wilms' tumor 1 (WT1) gene (+KTS isoform) functions with a CTE to enhance translation from an unspliced RNA with a retained intron. Genes Dev *20*, 1597-1608.

Bradford, S.T., Wilhelm, D., Bandiera, R., Vidal, V., Schedl, A., and Koopman, P. (2009). A cell-autonomous role for WT1 in regulating Sry in vivo. Human Molecular Genetics *18*, 3429-3438.

Bray, F., Richiardi, L., Ekbom, A., Pukkala, E., Cuninkova, M., and Møller, H. (2006). Trends in testicular cancer incidence and mortality in 22 European countries: Continuing increases in incidence and declines in mortality. Epidemiology 188:12*,* 3099-3111

Bullejos, M., and Koopman, P. (2001). Spatially dynamic expression of *Sry* in mouse genital ridges. Developmental Dynamics *221*, 201 - 205.

Bullejos, M., and Koopman, P. (2005). Delayed Sry and Sox9 expression in developing mouse gonads underlies B6-Y(DOM) sex reversal. Dev Biol *278*, 473-481.

Bundo, M., Sunaga, F., Ueda, J., Kasai, K., Kato, T., and Iwamoto, K. (2012). A systematic evaluation of whole genome amplification of bisulfite modified DNA. Clinical Epigenetics *4*.
Call, K.M., Glaser, T., Ito, C.Y., Buckler, A.J., Pelletier, J., Haber, D.A., Rose, E.A., Kral, A., Yeger, H., Lewis, W.H*., et al.* (1990). Isolation and characterisation of a zinc finger polypeptide gene at the human chromosome 11 Wilm's tumour locus. Cell *60*, 509-520.

Campanero, M.R., Herrero, A., and Calvo, V. (2008). The histone deacetylase inhibitor trichostatin A induces GADD45 gamma expression via Oct and NF-Y binding sites. Oncogene *27*, 1263-1272.

Carré, G.-A., and Greenfield, A. (2016). The gonadal supporting cell lineage and mmalian sex determination: the differentiation of Sertoli and Granulosa cells. Molecular Mechanisms of Cell Differentiation in Gonad Development. Springer International Publishing, 46-66.

Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nature reviews Genetics *10*, 295-304.

Chang, L., and Karin, M. (2001). Mammalian MAP Kinase Signalling Cascades. Nature *410*, 37 - 40.

Chassot, A.A., Gillot, I., and Chaboissier, M.C. (2014). R-spondin1, WNT4, and the CTNNB1 signaling pathway: strict control over ovarian differentiation. Reproduction *148*, R97-110.

Combes, A.N., Wilhelm, D., Davidson, T., Dejana, E., Harley, V., Sinclair, A., and Koopman, P. (2009). Endothelial cell migration directs testis cord formation. Developmental Biology *326*, 112-120.

Copeland, N.G., Jenkins, N.A., and Court, D.L. (2001). Recombineering: a poweful new tool for mouse functional genomics. Nature Reviews Genetics *2*, 769 - 779.

Cosgrove, M.S., and Wolberger, C. (2005). How does the histone code work? Biochemistry and cell biology = Biochimie et Biologie Cellulaire *83*, 468-476.

Dolci, S., Grimaldi, P., Germia, R., Pesce, M., and Rossi, P. (1997). Identification of a promoter region generating *Sry* circular transcripts both in germ cells from male adult mice and in male mouse embryonal gonads. Biology of Reproduction *57*, 1128 - 1135.

Domenice, S., Correa, R.V., Costa, E.M.F., Nishi, M.Y., Villain, E., Arnhold, I.J.P., and Nedonca, B.B. (2004). Mutations in the *SRY*, *DAX1, SF1* and *WNT4* genes in Brazillian sex-reversed patients. Brazillian Journal of Mecdical and Biological Research *37*, 145 - 150.

Eicher, E.M., and Washburn, L.L. (2001). Does One Gene Determine Whether a C57BL/6J-YPOS Mouse Will Develop as a Female or as an Hermaphrodite? Journal of Experimental Zoology *290*, 322 - 326.

Eicher, E.M., Washburn, L.L., Schork, N.J., Lee, B.K., Shown, E.P., Xu, X., Dredge, R.D., Pringle, M.J., and Page, D.C. (1996). Sex-determining genes on mouse autosomes identified by linkage analysis of C57BL/6J-YPOS sex reversal. Nature Genetics *14*, 206 - 209.

Einhauer, A., and Jungbauer, A. (2001). The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. Journal of Biochemical and Biophysical Methods *49*, 455 - 465.

Engel, N., Tront, J.S., Erinle, T., Nguyen, N., Latham, K.E, Sapienza, C., Hoffman, B., and Liebermann, D.A. (2009). Conserved DNA methylation in *Gadd45a*$^{-/-}$ mice. Epigenetics *4:2*, 98-99

Engel, N., Tront, J.S., Erinle, T., Nguyen, N., Latham, K.E., Sapienza, C., Hoffman, B., and Liebermann, D.A. (2014). Conserved DNA methylation inGadd45a-/-mice. Epigenetics *4*, 98-99.

Englert, C. (1998). WT1 - more than a transcription factor? . Trends in Biochemical Sciences *23*, 389 - 393.

Essers, J., Theil, A.F., Baldeyron, C., van Cappellen, W.A., Houtsmuller, A.B., Kanaar, R., and Vermeulen, W. (2005). Nuclear dynamics of PCNA in DNA replication and repair. Mol Cell Biol *25*, 9350-9359.

Fujimoto, Y., Tanaka, S.S., Yamaguchi, Y.L., Kobayashi, H., Kuroki, S., Tachibana, M., Shinomura, M., Kanai, Y., Morohashi, K., Kawakami, K*., et al.* (2013). Homeoproteins Six1 and Six4 regulate male sex determination and mouse gonadal development. Dev Cell *26*, 416-430.

Gardiner, W.J.K., and Teboul, L. (2009). Overexpression transgenesis in mouse: pronuclear injection. Methods in molecular biology *561*, 111-126.

Gierl, M.S., Gruhnl, W.H., von Seggem, A., Maltry, N., and Niehrs, C. (2012). GADD45G Functions in Male Sex Determination by Promoting p38 Signaling and Sry Expression. Dev Cell *23*, 1032-1042.

Giwercman, A., and Giwercman Y.L. (2011). Environmental factors and testicular function. Best Practice & Research Clinical Endocrinology & Metabolism *25*, 391-402.

Gopalakrishnan, S., Van Emburgh, B.O., and Robertson, K.D. (2008). DNA methylation in development and human disease. Mutation Research *647*, 30-38.

Graves, J.A.M. (1994). The origins and function of the mammalian Y chromosome and Y-borne genes - an evolving understanding. Bioessays *17*, 311-321.

Gu, H., Smith, Z.D., Bock, C., Boyle, P., Gnirke, A., and Meissner, A. (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nature Protocols *6*, 468-481.

Guibert, S., Forne, T., and Weber, M. (2012). Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome Research *22*, 633-641.

Guo, J.U., Ma, D.K., Mo, H., Ball, M.P., Jang, M.H., Bonaguidi, M.A., Balazer, J.A., Eaves, H.L., Xie, B., Ford, E*., et al.* (2011). Neuronal activity modifies the DNA methylation landscape in the adult brain. Nature Neuroscience *14*, 1345-1351.

Guo, W., Dong, Z., Guo, Y., Chen, Z., Kuang, G., and Yang, Z. (2013). Methylation-mediated repression of GADD45A and GADD45G expression in gastric cardia adenocarcinoma. International Journal of Cancer *133*, 2043-2053.

Gupta, S.K., Gupta, M., Hoffman, B., and Liebermann, D.A. (2006). Hematopoietic cells from gadd45a-deficient and gadd45b-deficient mice exhibit impaired stress responses to acute stimulation with cytokines, myeloablation and inflammation. Oncogene *25*, 5537-5546.

Hammes, A., Guo, J.K., Lutsch, G., Leheste, J.-r., Landrock, D., Ziegler, U., gubler, M.-C., and Schedl, A. (2001). Two splice variants of the Wilm's Tumor 1 gene have distinct functions during sex determination and nephron formation. Cell *106*, 319-329.

Hanley, N.A., Hagan, D.M., Clement-Jones, M., Ball, S.G., Strachan, T., Salas-Cortes, L., McElreavey, K., Lindsay, S., Robson, S., bullen, P.*, et al.* (2000). SRY, SOX9 and DAX1 expression patterns during human sex determination and gonadal development. Mechanisms of Development *91*, 403-407.

Hanna, C.W., Penaherrera, M.S., Saadeh, H., Andrews, S., McFadden, D.E., Kelsey, G., and Robinson, W.P. (2016). Pervasive polymorphic imprinted methylation in the human placenta. Genome Res *26*, 756-767.

Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. Nature *495*, 384-388.

Harry, J.L., Koopman, P., Brennan, F.E., Graves, J.A.M., and Renfree, M.B. (1995). Widespread expression of the testis-determining gene *SRY* in a marsupial. Nature Genetics *11*, 347-349.

Hashimoto, N., Kubokawa, R., Yamazaki, K., Noguchi, M., and Kato, Y. (1990). Germ cell deficiancy causes testis cord differentiation in reconstituted mouse fetal ovaries. The Journal of Experimental Zoology *253*, 61 - 70.

Hebestreit, K., Dugas, M., and Klein, H.U. (2013). Detection of significantly differentially methylated regions in targeted bisulfite sequencing data. Bioinformatics *29*, 1647-1653.

Herzenberg, L.A., Bianchi, D.W., Schröder, J., Cann, H.M., and Iverson, G.M. (1979). Fetal cells in the blood of pregnant women: detection and enrichment by fluorescecne-activted cell sorting. Proceedings of the National Academy of Sciences *76*, 1453 - 1455.

Hildesheim, J., Bulavin, D.V., Anver, M.R., Alvord, W.G., Hollander, M.C., Vardanian, L., and Jr., A.J.f. (2002). Gadd45a protects against UV irradiation-induced skin tumors and promotes apoptosis and stress signalling via MAPK and p53. Cancer Research *62*, 7305 - 7315.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nature Biotechnology *33*, 510-517.

Hindson, B.J., Ness, K.D., Masquelier, D.A., Belgrader, P., Heredia, N.J., Makarewicz, A.J., Bright, I.J., Lucero, M.Y., Hiddessen, A.L., Legler, T.C.*, et al.* (2011). High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Analytical Chemistry *83*, 8604-8610.

Hiort, O., Birnbaum, W., Marshall, L., Wunsch, L., Werner, R., Schroder, T., Dohnert, U., and Holterhus, P.M. (2014). Management of disorders of sex development. Nature reviews Endocrinology *10*, 520-529.

Hiramatsu, R., Matoba, S., Kanai-Azuma, M., Tsunekawa, N., Katoh-Fukui, Y., Kurohmaru, M., Morohashi, K., Wilhelm, D., Koopman, P., and Kanai, Y. (2009). A critical time window of Sry action in gonadal sex determination in mice. Development *136*, 129-138.

Hobert, O., and Westphal, H. (2000). Functions of LIM-homeobox genes. Trends in Genetics *16-2*, 75-83.

Hoffmeyer, A., Piekorz, R., Moriggl, R., and Ihle, J.N. (2001). Gadd45 gamma is dispensable for normal mouse development and T-cell proliferation. Molecular and Cellular Biology *21*, 3137-3143.

Hollander, M.C., Philburn, R.T., Patterson, A.D., Velasco-Miguel, S., Friedberg, E.C., Linnoila, R.I., and Fornace, A.J., Jr. (2005). Deletion of XPC leads to lung tumors in mice and is associated with early events in human lung carcinogenesis. Proceedings of the National Acadamy of Science USA *102*, 13200-13205.

Hollander, M.C., Sheikh, M.S., Bulavin, D.V., Lundgren, K., Augeri-Henmueller, L., Shehee, R., Molinaro, T.A., Kim, K.E., tolosa, E., Ashwell, J.D*., et al.* (1999). Genomic instability in *Gadd45a*-deficient mice. Nature Genetics *23*, 176 - 184.

Holliday, R. (2014). Epigenetics: A Historical Overview. Epigenetics *1*, 76-80.

Hosono, S., Faruqi, A.F., Dean, F.B., Du, Y., Sun, Z., Wu, X., Du, J., Kingsmore, S.F., Egholm, M., and Lasken, R.S. (2003). Unbiased whole-genome amplification directly from clinical samples. Genome Res *13*, 954-964.

Hu, Y.C., Okumura, L.M., and Page, D.C. (2013). Gata4 Is Required for Formation of the Genital Ridge in Mice. Plos Genetics *9,* e1003629.

Huang, H.S., Kubish, G.M., Redmond, T.M., Turner, D.L., Thompson, R.C., Murphy, G.G., and Uhler, M.D. (2010). Direct transcriptional induction of Gadd45gamma by Ascl1 during neuronal differentiation. Molecular and Cellular Neurosciences *44*, 282-296.

Hughes, I.A. (2008). Disorders of sex development: a new definition and classification. Best practice & research Clinical Endocrinology & Metabolism *22*, 119-134.

Hughes, I.A., Houk, C., Ahmed, S.F., Lee, P.A., and Lawson Wilkins Pediatric Endocrine Society/European Society for Paediatric Endocrinology Consensus, G. (2006). Consensus statement on management of intersex disorders. Journal of Pediatric Urology *2*, 148-162.

HyLown Consulting LLC (2013). Power and Sample Size |Free Online Calculators. http://powerandsamplesize.com/Calculators/

Jacinto, F.V., Ballestar, E., and Esteller, M. (2008). Methyl-DNA immunoprecipitation (MeDIP): Hunting down the DNA methylome. BioTechniques *44*, 35 - 43.

Jameson, S.A., Natarajan, A., Cool, J., DeFalco, T., Maatouk, D.M., Mork, L., Munger, S.C., and Capel, B. (2012). Temporal transcriptional profiling of somatic and germ cells reveals biased lineage priming of sexual fate in the fetal mouse gonad. PLoS Genetics *8*, e1002575.

Jeong, Y.H., Lu, H., Park, C.H., Li, M., Luo, H., Kim, J.J., Liu, S., Ko, K.H., Huang, S., Hwang, I.S.*, et al.* (2016). Stochastic anomaly of methylome but persistent SRY hypermethylation in disorder of sex development in canine somatic cell nuclear transfer. Scientific reports *6*, 31088.

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature *449*, 248-251.

Jin, S.G., Guo, C., and Pfeifer, G.P. (2008). GADD45A does not promote DNA demethylation. PLoS Genetics *4*, e1000013.

Johnen, H., Gonzalez-Silva, L., Carramolino, L., Flores, J.M., Torres, M., and Salvador, J.M. (2013). Gadd45g is essential for primary sex determination, male fertility and testis development. PLoS One *8*, e58751.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics *13*, 484-492.

Ju, S., Zhu, Y., Liu, L., Dai, S., Li, C., Chen, E., He, Y., Zhang, X., and Lu, B. (2009). Gadd45b and Gadd45g are important for anti-tumor immune responses. European Journal of Immunology *39*, 3010-3018.

Jung, H.J., Kim, E.H., Mun, J.Y., Park, S., Smith, M.L., Han, S.S., and Seo, Y.R. (2007). Base excision DNA repair defect in Gadd45a-deficient cells. Oncogene *26*, 7517-7525.

Kalfa, N., Fellous, M., Boizet-Bonhoure, B., Patte, C., Duvillard, P., Pienkowski, C., Jaubert, F., Ecochard, A., and Sultan, C. (2008). Aberrant expression of ovary determining gene FOXL2 in the testis and juvenile granulosa cell tumor in children. The Journal of Urology *180*, 1810-1813.

Kangaspeska, S., Stride, B., Metivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F., and Reid, G. (2008). Transient cyclical methylation of promoter DNA. Nature *452*, 112-115.

Kashimada, K., and Koopman, P. (2010). Sry: the master switch in mammalian sex determination. Development *137*, 3921-3930.

Kastan, M.B., Zhan, Q., El-Deiry, W.S., Carrier, F., Jacks, T., Walsh, W.V., Plunkett, B.S., Vogelstein, B., and Jr., A.J.F. (1992). A mammalian cell cycle checkpoint pathway utilizing p53 and *GADD45* is defective in Ataxia-Telaniectasia. Cell *71*, 587 - 597.

Katoh-Fukui, Y., Miyabayashi, K., Komatsu, T., Owaki, A., Baba, T., Shima, Y., Kidokoro, T., Kanai, Y., Schedl, A., Wilhelm, D.*, et al.* (2012). Cbx2, a polycomb group gene, is required for Sry gene expression in mice. Endocrinology *153*, 913-924.

Katoh-Fukui, Y., Owaki, A., Toyama, Y., Kusaka, M., Shinohara, Y., Maekawa, M., Toshimori, K., and Morohashi, K. (2005). Mouse Polycomb M33 is required for splenic vascular and adrenal gland formation through regulating Ad4BP/SF1 expression. Blood *106*, 1612-1620.

Katoh-Fukui, Y., Tsuchiya, R., Shiroishi, T., Nakahara, Y., Hashimoto, N., Noguchi, K., and Hagashinakagawa, T. (1998). Male-to-female sex reversal in M33 mutant mice. Nature *393*, 688 -692.

Kaufmann, L.T., Gierl, M.S., and Niehrs, C. (2011). Gadd45a, Gadd45b and Gadd45g expression during mouse embryonic development. Gene expression patterns *11*, 465-470.

Kaufmann, L.T., and Niehrs, C. (2011). Gadd45a and Gadd45g regulate neural development and exit from pluripotency in Xenopus. Mech Dev *128*, 401-411.

Kawaue, T., Sagou, K., Kiyonari, H., Ota, K., Okamoto, M., Shinoda, T., Kawaguchi, A., and Miyata, T. (2014). Neurogenin2-d4Venus and Gadd45g-d4Venus transgenic mice: visualizing mitotic and migratory behaviors of cells committed to the neuronal lineage in the developing mammalian brain. Development, Growth & Differentiation *56*, 293-304.

Kienhofer, S., Musheev, M.U., Stapf, U., Helm, M., Schomacher, L., Niehrs, C., and Schafer, A. (2015). GADD45a physically and functionally interacts with TET1. Differentiation *90*, 59-68.
Kim, Y., Bingham, N., Sekido, R., Parker, K.L., Lovell-Badge, R., and Capel, B. (2007). Fibroblast growth factor receptor 2 regulates proliferation and Sertoli differentiation during male sex determination. Proceedings of the National Acadamy of Sciences USA *104*, 16558-16563.

Kim, Y., and Capel, B. (2006). Balancing the bipotential gonad between alternative organ fates: a new perspective on an old problem. Developmental dynamics *235*, 2292-2300.

Kim, Y., Kobayashi, A., Sekido, R., DeNapoli, L., Brennan, J., Chaboissier, M.-C., Poulet, F., Behringer, R.R., Lovell-Badge, R., and Capel, B. (2006). *Fgf9* and *Wnt4* Act as antagonistic signals to regulate mammalian sex determination. PLoS Biology *4*, 1000-1009.

Kobayashi, H., Sakurai, T., Miura, F., Imai, M., Mochiduki, K., Yanagisawa, E., Sakashita, A., Wakai, T., Suzuki, Y., Ito, T*., et al.* (2013). High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. Genome Res *23*, 616-627.

Kobayuki, H., Sakurai, T., Imai, M., Takahashi, N., Fukudu, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y*., et al.* (2012). Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish Oocyte-specific heritable marks. Plos Genetics *8,* e1002440.

Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. Nature *502*, 472-479.

Koide, A., Bailey, C.W., Huang, X., and Koide, S. (1998). The fibronectin type III domain as a scaffold for novel binding proteins. Journal of Molecular Biology *284*, 1141 - 1151.

Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P., and Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for *Sry*. Nature *351*, 117-121.

Koubova, J., Menke, D.B., Zhou, Q., Capel, B., Griswold, M.D., and Page, D.C. (2006). Retinoic acid regulates sex-specific timing of meiotic initiation in mice. Proceedings of the National Acadamy of Sciences USA *103*, 2474-2479.

Kousta, E., Papathanasiou, A., and Skordis, N. (2010). Sex determination and disorders of sex development according to the revised nomenclature and classification in 46, XX individuals. Hormones *9(3)*, 218-231.

Kreidberg, J.A., Sariola, H., Loring, J.M., Maeda, M., Pelletier, J., Housman, D., and Jaenisch, R. (1993). WT-1 is required for early kidney development. Cell *74*, 679-691.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571-1572.

Kulis, M., Heath, S., Bibikova, M., Queiros, A.C., Navarro, A., Clot, G., Martinez-Trillos, A., Castellano, G., Brun-Heath, I., Pinyol, M.*, et al.* (2012). Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nature Genetics *44*, 1236-1242.

Kuroki, s., Matoba, S., Akiyoshi, M., Matsumura, Y., Miyachi, H., Mise, N., Abe, K., Ogura, A., Wilhelm, D., Koopman, P.*, et al.* (2013). Epigenetic Regulation of Mouse Sex Determination by the Histone Demethylase Jmjd1a. Science *341*, 1106-1109.

Kurtz, J.E., and Ray-Coquard, I. (2012). PI3 Kinase Inhibitors in the Clinic: An Update. Anticancer Res *32*, 2463-2470.

Laird, P.W. (2010). Principles and challenges of genomewide DNA methylation analysis. Nature Reviews Genetics *11*, 191-203.

Larney, C., Bailey, T.L., and Koopman, P. (2014). Switching on sex: transcriptional regulation of the testis-determining gene Sry. Development *141*, 2195-2205.

Lasken, R.S., and Egholm, M. (2003). Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens. Trends in Biotechnology *21*, 531-535.

Le May, N., Mota-Fernandes, D., Velez-Cruz, R., Iltis, I., Biard, D., and Egly, J.M. (2010). NER factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack. Mol Cell *38*, 54-66.

Lefebvre, V., and Crombrugghe, B.D. (1998). Towards understanding SOX9 function in chrondrocyte differentiation. Matrix Biology *16*, 529 - 540.

Liebermann, D.A., and Hoffman, B. (2008). Gadd45 in stress signaling. Journal of Molecular Signaling *3*, 15.

Lin, Y.-T., and Capel, B. (2015). Cell fate commitment during mammalian sex determination. Current Opinion in Genetics & Development *32*, 144-152.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M.*, et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315-322.

Liu, J., Carmall, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.-J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engiene of mammalian RNAi. Science *305*, 1437 - 1441.

Looijenga, L.H., Hersmus, R., Oosterhuis, J.W., Cools, M., Drop, S.L., and Wolffenbuttel, K.P. (2007). Tumor risk in disorders of sex development (DSD). Best practice & research Clinical Endocrinology & Metabolism *21*, 480-495.

Lorincz, M.C., Dickerson, D.R., Schmitt, M., and Groudine, M. (2004). Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nature Structural & Molecular Biology *11*, 1068-1075.

Lou, S., Lee, H.-M., Qin, H., Li, J.-W., Gao, Z., Chan, L.L., Lam, V.K., So, W.-Y., Wang, Y., Lok, S.*, et al.* (2014). Whole-genome bisulfite sequencing of multiple individuals reveals complimentary roles of promoter and gene body methylation in transcriptional regulation. Genome Biology *15*.

Louvi, A., Accili, D., and Efstratiadis, A. (1997). Growth-Promoting interaction of IGF-II with the insulin receptor during mouse embryonic development. Developmental Biology *189-1*, 33-48.
Lu, B., Yu, H., Chow, C.-w., Li, B., Zheng, W.-p., Davis, R.J., and Flavell, R.A. (2001). GADD45γ mediates the activation of the p38 and JNK MAP kinase pathways and cytokine production iin effector TH1 Cells. Immunity *14*, 583 - 590.

Luco, R.F., and Misteli, T. (2011). More than a splicing code: integrating the role of RNA, chromatin and non-coding RNA in alternative splicing regulation. Current Opinion in Genetics & Development *21*, 366-372.

Luo, X.R., Ikeda, Y.Y., and Parker, K.L. (1994). A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual-differentiation. Cell *77*, 481-490.

Ma, D.K., Jang, M.-H., Guo, J.U., Kitabatake, Y., Chang, M.-L., Pow-anpongkul, N., Flavell, R.A., Lu, B., Ming, G.-l., and Song, H. (2009). Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. Science *323*, 1074 - 1077.

Maatouk, D.M., Mork, L., Hinson, A., Kobayashi, A., McMahon, A.P., and Capel, B. (2012). Germ cells are not required to establish the female pathway in mouse fetal gonads. PLoS One *7*, e47238.

Manuylov, N.L., Zhou, B., Ma, Q., Fox, S.C., Pu, W.T., and Tevosain, S.G. (2011). Conditional ablation of *Gata4* and *Fog2* in mice reveals their distinct roles in mammalian seual differentiation. Developmental Biology *353-2*, 229-241.

Matson, C.K., Murphy, M.W., Sarver, A.L., Griswold, M.D., Bardwell, V.J., and Zarkower, D. (2011). DMRT1 prevents female reprogramming in the postnatal mammalian testis. Nature *476*, 101-104.

McClelland, K.S., Bell, K., Larney, C., Harley, V.R., Sinclair, A.H., Oshlack, A., Koopman, P., and Bowles, J. (2015). Purification and Transcriptomic Analysis of Mouse Fetal Leydig Cells Reveals Candidate Genes for Specification of Gonadal Steroidogenic Cells. Biol Reprod *92*, 145.

Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res *33*, 5868-5877.

Mendonca, B.B., Domenice, S., Arnhold, I.J., and Costa, E.M. (2009). 46,XY disorders of sex development (DSD). Clinical Endocrinology *70*, 173-187.

Mill, J., Yazdanpanah, S., Gückel, E., Ziegler, S., Kaminsky, Z., and Petronis, A. (2006). Whole genome amplification of sodium bisulfite-treated DNA allows the accurate estimate of methylated cytosine density in limited DNA resources. BioTechniques *41*, 603-607.

Miller, C.A., Ingmer, H., and Cohen, S.N. (1995). Boundaries of the pSC101 minimal replicon are conditional. Journal of Bacteriology *177*, 4865 - 4871.

Minkina, A., Matson, Clinton K., Lindeman, Robin E., Ghyselinck, Norbert B., Bardwell, Vivian J., and Zarkower, D. (2014). DMRT1 Protects Male Gonadal Cells from Retinoid-Dependent Sexual Transdifferentiation. Dev Cell *29*, 511-520.

Miura, F., Enomoto, Y., Dairiki, R., and Ito, T. (2012). Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. Nucleic Acids Res *40*, e136.

Miyake, Z., Takekawa, M., Ge, Q., and Saito, H. (2007). Activation of MTK1/MEKK4 by GADD45 through induced N-C dissociation and dimerization-mediated trans autophosphorylation of the MTK1 kinase domain. Mol Cell Biol *27*, 2765-2776.

Miyamoto, N., Yoshida, M., Kuratani, S., Matsuo, I., and Aizawa, S. (1997). Defects of urogenital development in mice lacking *Emx2*. Development *124*, 1653 - 1664.

Miyamoto, Y., Taniguchi, H., Hamel, F., Silversides, D.W., and Viger, R.S. (2008). A GATA4/WT1 cooperation regulates transcription of genes required for mammalian sex determination and differentiation. BMC Molecular Biology *9*, 44.

Molyneaux, K.A., Stallock, J., Schaible, K., and Wylie, C. (2001). Time-lapse analysis of living mouse germ cell migration. Dev Biol *240*, 488-498.

Moniot, B., Declosmenil, F., Barrionuevo, F., Scherer, G., Aritake, K., Malki, S., Marzi, L., Cohen-Solal, A., Georg, I., Klattig, J*., et al.* (2009). The PGD2 pathway, independently of FGF9, amplifies SOX9 activity in Sertoli cells during male sexual differentiation. Development *136*, 1813-1821.

Munger, S.C., Aylor, D.L., Syed, H.A., Magwene, P.M., Threadgill, D.W., and Capel, B. (2009). Elucidation of the transcription network governing mammalian sex determination by exploiting strain-specific susceptibility to sex reversal. Genes Dev *23*, 2521-2536.

Muyrers, J.P.P., Zhang, Y., Buchholz, F., and Stewart, A.F. (2000). RecE/RecT and Redα/Redβ initiate double-stranded break repair by specifically interacting with their respiective partners. Genes & Development *14*, 1971 - 1982.

Nagamine, C.M., Morohashi, K.-i., Carlisle, C., and Chang, D.K. (1999). Sex Reversal Caused by Mus musculus domesticus Y Chromosomes Linked to Variant Expression of the Testis-Determining Gene Sry. Developmental Biology *216*, 182-194.

Nagamine, C.M., Taketo, T., and Koo, G.C. (1987). Studies on the genetics of *tda-1* XY sex reversal in the mouse. Differentiation *33*, 223 - 231.

Nef, S., Schaad, O., Stallings, N.R., Cederroth, C.R., Pitetti, J.L., Schaer, G., Malki, S., Dubois-Dauphin, M., Boizet-Bonhoure, B., Descombes, P*., et al.* (2005). Gene expression during sex determination reveals a robust female genetic program at the onset of ovarian development. Developmental Biology *287*, 361-377.

Niehrs, C., and Schafer, A. (2012). Active DNA demethylation by Gadd45 and DNA repair. Trends in cell biology *22*, 220-227.

Nikolova, G., Sinsheimer, J.S., Eicher, E.M., and Vilain, E. (2008). The chromosome 11 region from strain 129 provides protection from sex reversal in XYPOS mice. Genetics *179*, 419-427.

Nile, C.J., Read, R.C., Akil, M., Duff, G.W., and Wilson, A.G. (2008). Methylation status of a single CpG site in the IL6 promoter is related to IL6 messenger RNA levels and rheumatoid arthritis. Arthritis and Rheumatism *58*, 2686-2693.

Ninshino, K., Hattori, N., Sato, S., Arai, y., Tanaka, S., Nagy, A., and Shiota, K. (2011). Non-CpG Methylation Occurs in the Regulatory Region of the *Sry* Gene. J Reprod Dev *57*, 586-593.

Nishi, T., Shimizu, N., Hiramoto, M., Sato, I., Yamaguchi, Y., Hasegawa, M., Aizawa, S., Tanaka, H., Kataoka, K., Watanabe, H*., et al.* (2002). Spatial redox regulation of a critical cysteine residue of NF-kappa B in vivo. The Journal of Biological Chemistry *277*, 44548-44556.

Nishino, K., Hattori, N., Sato, S., Arai, Y., Tanaka, S., Nagy, A., and Shiota, K. (2011). Non-CpG Methylation Occurs in the Regulatory Region of the Sry Gene. J Reprod Dev *57*, 586-593.

Nishino, K., Hattori, N., Tanaka, S., and Shiota, K. (2004). DNA methylation-mediated control of Sry gene expression in mouse gonadal development. The Journal of Biological Chemistry *279*, 22306-22313.

Ohkouchi, S., Shibata, M., Sasaki, M., Koike, M., Safig, P., Peters, C., Nagata, S., and Uchiyama, Y. (2013). Biogenesis and proteolytic processing of lysosomal DNase II. PLoS One *8*, e59148.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essesntial for De Novo methylation and mammalian development. Cell *99*, 247 - 257.

Ono, M., and Harley, V.R. (2013). Disorders of sex development: new genes, new concepts. Nature Reviews Endocrinology *9*, 79-91.

Palmer, S.J., and Burgoyne, P.S. (1991). *In situ* analysis of fetal, prepuberal and adult XX <-> XY chimaeric mouse testes: Sertoli cells are predominantly, but not exclusively, XY. Development *112*, 265-268.

Pamnani, M., Sinha, P., Singh, A., Nara, S., and Sachan, M. (2016). Methylation of the Sox9 and Oct4 promoters and its correlation with gene expression during testicular development in the laboratory mouse. Genetics and Molecular Biology *39*, 452-458.

Parma, P., Pailhous, E., and Cotinot, C. (1999). Reverse Transcription-Polymerase chain reaction analysis of genes involved in gonadal differentiation in pigs. Biology of Reproduction *61*, 741 -748.

Parma, P., Radi, O., Vidal, V., Chaboissier, M.C., Dellambra, E., Valentini, S., Guerra, L., Schedl, A., and Camerino, G. (2006). R-spondin1 is essential in sex determination, skin differentiation and malignancy. Nature Genetics *38*, 1304-1309.

Pearlman, A., Loke, J., Le Caignec, C., White, S., Chin, L., Friedman, A., Warr, N., Willan, J., Brauer, D., Farmer, C.*, et al.* (2010). Mutations in MAP3K1 cause 46,XY disorders of sex development and implicate a common signal transduction pathway in human testis determination. American Journal of Human Genetics *87*, 898-904.

Pelletier, J., Bruening, W., Kashtan, C.E., Mauer, S.M., Manicel, J.c., Striegel, J.E., Houghton, D.C., Junien, C., Habib, R., Fouser, L.*, et al.* (1991). Germline mutations in the Wilm's tumour supressor gene are associated with abnormal urogenital development in denys-drash syndrome. Cell *67*, 437 - 447.

Pilon, N., Daneau, I., Paradis, V., Hamel, F., Lussier, J.G., Viger, R.S., and Silversides, D.W. (2003). Porcine SRY promoter is a target for steroidogenic factor 1. Biol Reprod *68*, 1098-1106.

Pitetti, J.L., Calvel, P., Romero, Y., Conne, B., Truong, V., Papaioannou, M.D., Schaad, O., Docquier, M., Herrera, P.L., Wilhelm, D.*, et al.* (2013). Insulin and IGF1 receptors are essential for XX and XY gonadal differentiation and adrenal development in mice. PLoS Genetics *9*, e1003160.

Pope, M.J. (2015). Loss-and gain-of-function analyses of the role of MAP kinase signalling in testis determination. In Department of Phsiology, Anatomy and Genetics (University of Oxford), pp. 240.

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell *135*, 1201-1212.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proceedings of the National Academy of Sciences *97*, 5237 - 5242.

Riggs, A.D. (1975). X inactivation, differentiation and DNA methylation. Cytogenetic and Genome Research *14*, 9 - 25.

Robertson, A.B., Klungland, A., Rognes, T., and Leiros, I. (2009). DNA repair in mammalian cells: Base excision repair: the long and short of it. Cellular and Molecular Life Sciences : CMLS *66*, 981-993.

Robson-Dixon, N.D., and Garcia-Blanco, M.A. (2004). MAZ elements alter transcription elongation and silencing of the fibroblast growth factor receptor 2 exon IIIb. The Journal of Biological Chemistry *279*, 29075-29084.

Ross, D.G., Bowles, J., Koopman, P., and Lehnert, S. (2008). New insights into SRY regulation through identification of 5' conserved sequences. BMC Molecular Biology *9*, 85.

Sambrook, J., and Russell, D.W. (1989). Molecular cloning: a laboratory manual. Cold Spring Harbour Press.

Schafer, A., Karaulanov, E., Stapf, U., Doderlein, G., and Niehrs, C. (2013). Ing1 functions in DNA demethylation by directing Gadd45a to H3K4me3. Genes Dev *27*, 261-273.

Schleif, R. (2010). AraC protein, regulation of the l-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action. FEMS Microbiology Reviews *34*, 779-796.

Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C_T method. Nature Protocols *3*, 1101 – 1108.

Schmitz, K.M., Schmitt, N., Hoffmann-Rohrer, U., Schafer, A., Grummt, I., and Mayer, C. (2009). TAF12 recruits Gadd45a and the nucleotide excision repair complex to the promoter of rRNA genes leading to active DNA demethylation. Molecular Cell *33*, 344-353.

Schomacher, L. (2013). Mammalian DNA demethylation: multiple faces and upstream regulation. Epigenetics *8*, 679-684.

Schrag, J.D., Jiralerspong, S., Banville, M., Jaramillo, M.L., and O'Connor-McCourt, M.D. (2008). The crystal structure and dimerization interface of GADD45gamma. Proceedings of the National Acadamy of Science U S A *105*, 6566-6571.

Schwartzman, O., and Tanay, A. (2015). Single-cell epigenomics: techniques and emerging applications. Nature Reviews Genetics *16*, 716-726.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Molecular Cell *48*, 849-862.

Sekido, R., and Lovell-Badge, R. (2008). Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. Nature *453*, 930-934.

Selvakumaran, M., Lin, H.-K., Sjin, R.T.T., Reed, J.C., Liebermann, D.A., and Hoffman, B. (1994). The Novel primary response gene *MyD118* and the Proto-oncogenes *myb*, *myc* and *bcl-2* modulate transforming growth factor β1- induced apoptosis of myeloid leukemia cells. Molecular and Cellular Biology *14*, 2352-2360.

Sen, G.L., Reuter, J.A., Webster, D.E., Zhu, L., and Khavari, P.A. (2010). DNMT1 maintains progenitor function in self-renewing somatic tissue. Nature *463*, 563-567.

Sharova, L.V., Sharov, A.A., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M.S. (2009). Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. DNA Research *16*, 45-58.

Shen, J.-C., III, W.M.R., and Jones, P.A. (1994). The rate of hydrolytic deamination of 5-methylcytosine double-stranded DNA. Nucleic Acids Research *22*, 972 - 976.

Siggers, P., Carre, G.A., Bogani, D., Warr, N., Wells, S., Hilton, H., Esapa, C., Hajihosseini, M.K., and Greenfield, A. (2014). A novel mouse Fgfr2 mutant, hobbyhorse (hob), exhibits complete XY gonadal sex reversal. PLoS One *9*, e100447.

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

Sirane, K., Toh, H., Kobayashi, H., Miura, F., Chiba, H., Ito, T., Kono, T., and Sasaki, H. (2013). Mouse Oocyte methylomes at base resolution reveal genome-wide accumilation of non-CpG methylation and role of DNA methyltransferases. Plos Genetics *9*, e1003439.

Skerra, A. (2007). Alternative non-antibody scaffolds for molecular recognition. Current Opinion in Biotechnology *18*, 295-304.

Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nature methods *11*, 817-820.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nature Reviews Genetics *14*, 204-220.

Stark, K., Vainio, S., Vassileva, G., and McMahon, A.P. (1994). Epithelial transformation of metanephri mesenchyme in the developing kidney regulated by *Wnt4*. Nature *372*, 679-683.

Stothard, P. (2000). The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. In Sequence Manipulation suite (http://www.bioinformatics.org/sms2/cpg_islands.html).

Svingen, T., and Koopman, P. (2013). Building the mammalian testis: origins, differentiation, and assembly of the component cell populations. Genes Dev *27*, 2409-2426.

Swain, A., and Lovell-Badge, R. (1999). Mammalian sex determination: a molecular drama. Genes & Development *13*, 755 - 767.

Takekawa, M., and Saito, H. (1998). A family of stress-inducible GADD45-like proteins mediate activation of the stress-responsive MTK1/MEKK4 MAPKKK. Cell *95*, 521-530.

Tanaka, S.S., and Nishinakamura, R. (2014). Regulation of male sex determination: genital ridge formation and Sry activation in mice. Cellular and molecular life sciences *71*, 4781-4802.

Tevosian, S.G., Albrecht, K.H., Crispino, J.d., Fujiwara, Y., Eicher, E.M., and Orkin, S.H. (2002). Gonadal differentiation, sex determination and normal Sry expression in mice require direct interaction between transcription partners GATA4 and FOG2. Development *129*, 4627 - 4634.

Thakore, P.I., D'Ippolito, A.M., Song, L., Safi, A., Shivakumar, N.K., Kabadi, A.M., Reddy, T.E., Crawford, G.E., and Gersbach, C.A. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nature methods *12*, 1143-1149.

Thalheimer, F.B., Wingert, S., De Giacomo, P., Haetscher, N., Rehage, M., Brill, B., Theis, F.J., Hennighausen, L., Schroeder, T., and Rieger, M.A. (2014). Cytokine-regulated GADD45G induces differentiation and lineage selection in hematopoietic stem cells. Stem Cell Reports *3*, 34-43.

Tran, S., Wang, Y., Lamba, P., Zhou, X., Boehm, U., and Bernard, D.J. (2013). The CpG island in the murine foxl2 proximal promoter is differentially methylated in primary and immortalized cells. PLoS One *8*, e76642.

Uhlenhaut, N.H., Jakob, S., Anlag, K., Eisenberger, T., Sekido, R., Kress, J., Treier, A.-C., Klugmann, C., Klasen, C., Holter, N.I*., et al.* (2009). Somatic Sex Reprogramming of Adult Ovaries to Testes by FOXL2 Ablation. Cell *139*, 1130-1142.

Van Keuren, M.L., Gavrilina, G.B., Filipiak, W.E., Zeidler, M.G., and Saunders, T.L. (2009). Generating transgenic mice from bacterial artificial chromosomes: transgenesis efficiency, integration and expression outcomes. Transgenic Research *18*, 769-785.

Verkauskas, G., Jaubert, F., Lortat-Jacob, S., Malan, V., Thibaud, E., and Nihoul-Fekete, C. (2007). The long-term followup of 33 cases of true hermaphroditism: a 40-year experience with conservative gonadal surgery. The Journal of Urology *177*, 726-731; discussion 731.

Vidal, V.P.I., Chaboissier, M.-C., Rooj, D.G.d., and Schredl, A. (2001). *Sox9* induces testis development in XX transgenic mice. Nature Genetics *28*, 216-217.

Walsh, T.J., Grady, R.W., Porter, M.P., Lin, D.W., and Weiss, N.S. (2006). Incidence of testicular germ cell cancers in U.S. children: SEER program experience 1973 to 2000. Urology *68:2*, 402-405.

Wang, Q., Gu, L., Adey, A., Radlwimmer, B., Wang, W., Hovestadt, V., Bahr, M., Wolf, S., Shendure, J., Eils, R*., et al.* (2013). Tagmentation-based whole-genome bisulfite sequencing. Nature protocols *8*, 2022-2032.

Wang, S.W., Zhao, Y.J., Leiby, M., and Zhu, J.Y. (2009). A New Positive/Negative Selection Scheme for Precise BAC Recombineering. Mol Biotechnol *42*, 110-116.

Warner, D.A., Shine, R. (2008). The adaptive significance of termperature-dependent sex determination in reptiles. Nature *451*, 566-568

Warr, N., Bogani, D., Siggers, P., Brixey, R., Tateossian, H., Dopplapudi, A., Wells, S., Cheeseman, M., Xia, Y., Ostrer, H*., et al.* (2011). Minor abnormalities of testis development in mice lacking the gene encoding the MAPK signalling component, MAP3K1. PLoS One *6*, e19572.

Warr, N., Carre, G.A., Siggers, P., Faleato, J.V., Brixey, R., Pope, M., Bogani, D., Childers, M., Wells, S., Scudamore, C.L*., et al.* (2012). Gadd45 gamma and Map3k4 Interactions Regulate Mouse Testis Determination via p38 MAPK-Mediated Control of Sry Expression. Dev Cell *23*, 1020-1031.

Warr, N., and Greenfield, A. (2012). The molecular and cellular basis of gonadal sex reversal in mice and humans. Wiley Interdisciplinary Reviews Developmental biology *1*, 559-577.

Warr, N., Siggers, P., Bogani, D., Brixey, R., Pastorelli, L., Yates, L., Dean, C.H., Wells, S., Satoh, W., Shimono, A*., et al.* (2009). Sfrp1 and Sfrp2 are required for normal male sexual development in mice. Developmental Biology *326*, 273-284.

Warr, N., Siggers, P., Carre, G.A., Bogani, D., Brixey, R., Akiyoshi, M., Tachibana, M., Teboul, L., Wells, S., Sanderson, J*., et al.* (2014). Transgenic expression of Map3k4 rescues T-associated sex reversal (Tas) in mice. Human Molecular Genetics *23*, 3035-3044.

Warr, N., Siggers, P., Carre, G.A., Wells, S., and Greenfield, A. (2016). Genetic Analyses Reveal Functions for MAP2K3 and MAP2K6 in Mouse Testis Determination. Biol Reprod *94*, 103.

Welsh, M., Sharpe, R.M., Walker, M., Smith, L.B., and Saunders, P.T. (2009). New insights into the role of androgens in wolffian duct stabilization in male and female rodents. Endocrinology *150*, 2472-2480.

Welshons, W.J., and Russell, L.B. (1959). The Y-chromosome as the bearer of male determining factors in the mouse. Proceedings of the National Academy of Sciences *45*, 560 - 566.

Westenberg, M., Soedling, H.M., Mann, D.A., Nicholson, L.J., and Dolphin, C.T. (2010). Counter-selection recombineering of the baculovirus genome: a strategy for seamless modification of repeat-containing BACs. Nucleic Acids Research *38*, e166.

Weston, S.A., and Parish, C.R. (1990). New fluorescent dyes for lymphocyte migration studies Analysis by flow cytometry and fluorescence microscopy. Journal of Immunological Methods *133*, 87 -97.

Wijesinghe, P., and Bhagwat, A.S. (2012). Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. Nucleic Acids Research *40*, 9206-9217.

Wilhelm, D., and Koopman, P. (2006). The makings of maleness: towards an integrated view of male sexual development. Nature Reviews Genetics *7*, 620-631.

Wilhelm, D., Palmer, S., and Koopman, P. (2007). Sex Determination and Gonadal Development in Mammals. Physiological Reviews *87-1*, 1-28

Wohlfahrt-Veje, C., Main, K.M., and Skakkebæk, N.E. (2009). Testicular dysgenesis syndrome: foetal origin of adult reproductive problems. Clinical Endocrinology, 71:4, 459-465.

Wu, H., and Zhang, Y. (2011). Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. Genes Dev *25*, 2436-2452.

Yan, J., Zierath, J.R., and Barres, R. (2011). Evidence for non-CpG methylation in mammals. Experimental Cell Research *317*, 2555-2561.

Yang, H., Wang, H., Shivalila, Chikdu S., Cheng, Albert W., Shi, L., and Jaenisch, R. (2013). One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. Cell *154*, 1370-1379.

Yang, Z., Song, L., and Huang, C. (2009). Gadd45 proteins as Critical signal transducers linking NF-κB to MAPK Cascades. Current Cancer Drug Targets *9*, 915-930.

Ying, J., Srivastava, G., Hsieh, W.S., Gao, Z., Murray, P., Liao, S.K., Ambinder, R., and Tao, Q. (2005). The stress-responsive gene GADD45G is a functional tumor suppressor, with its response to environmental stresses frequently disrupted epigenetically in multiple tumors. Clinical cancer research *11*, 6442-6449.

Yoshimizu, T., Sugiyama, N., Felice, M.D., Yeom, y.I., Ohba, K., Masuko, K., Obinata, M., Abe, K., Schöler, H.R., and Matsui, Y. (1999). Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. Development, Growth and Differentiation *41*, 675 - 684.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D*., et al.* (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355-364.

Zerbini, L.F., and Libermann, T.A. (2005). GADD45 deregulation in cancer: frequently methylated tumor suppressors and potential therapeutic targets. Clinical cancer research *11*, 6409-6413.

Zhan, Q., Lord, K.A., Jr., I.A., Hollander, M.c., Carrier, F., Ron, D., Kohn, K.W., Hoffman, B., Liebermann, D.A., and Jr., A.J.F. (1994). The *Gadd* and *MyD* Gene define a novel set of mammalian genes encoding acidic proteins that synergistically supress cell growth. Molecular and Cellular Biology *14*, 2361-2371.

Zhang, W., Bae, I., Krishnaraju, K., Azam, N., Fan, W., Smith, K., Hoffman, B., and Liebermann, D.A. (1999). *CR6*: A third member of the *MyD118* and *Gadd45* gene family which functions in growth control. Oncogene *18*, 4899 - 4907.

Zhang, Y., Buchholz, F., Muyrers, J.P.P., and Stewart, A.F. (1998). A new logic for DNA engineering using recombination in *Escherichia coli*. Nature Genetics *20*, 123 - 128.

Zhao, L., Gu, H., Chang, J., Wu, J., Wang, D., Chen, S., Yang, X., and Qian, B. (2014). MicroRNA-383 regulates the apoptosis of tumor cells through targeting Gadd45g. PLoS One *9*, e110472.

Appendices

Research Output

Winning talk given at 10th Mammalian Genes, Development and Disease Meeting.

Appendix i: Tail Somite staging of mouse embryos

Throughout this thesis, timing of the development of the gonad and the rest of the embryos is described in tail somites (ts). As a frame of reference, this section summarises the timing of each tail somite stage and its relevant stage of development in days post coitum (dpc). Somites are segmented regions of mesoderm which arise along the dorsal aspect of the embryo. Somite formation is well coordinated and one somite emerges approximately every two hours. As a result, embryos can be precisely staged according to the number of somites. In embryos aged 11.0 to 12.5 dpc, tail somites are counted. These somites are counted as those posterior to the hind limb (Fig I).

A

Stage (dpc)	11.0 dpc						11.5 dpc						12.0 dpc						12.5 dpc					
Tail Somites	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32

B

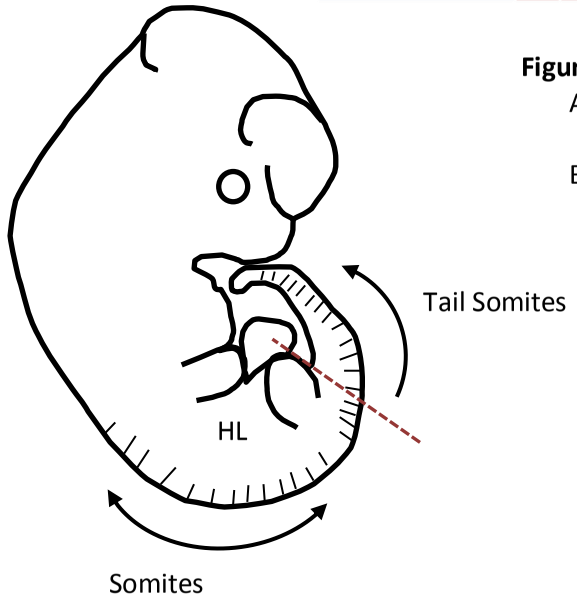


Figure I: Tail somite staging of mouse embryos

A) Table showing the relationship between tail somite stages and age in days post coitum (dpc)

B) Tail somites are counted as those that are posterior to the end of the hind limb to the end of the tail