



Loss- and gain-of-function analyses of the role of MAP kinase signalling in testis determination

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

St Cross College, University of Oxford

Madeleine Jane Pope

Mammalian Genetics Unit, MRC Harwell

Department of Physiology, Anatomy and Genetics, University of Oxford

Supervisors: Dr Andy Greenfield (MGU, MRC Harwell)

Dr Helen Christian (DPAG, Oxford)

Trinity term 2015

Word count: 45,102

Loss- and gain-of-function analyses of the role of MAP kinase signalling in testis determination

Madeleine Pope

St Cross College, Oxford

MRC Mammalian Genetics Unit, MRC Harwell

Department of Physiology, Anatomy and Genetics, University of Oxford

A thesis presented for D. Phil, Trinity term, 2015

Abstract

Sexual differentiation is the process by which an organism develops as either male or female. In mammals, this process is governed chromosomally, principally by the presence or absence of the Y chromosome. In the mouse, sex determination occurs at 11.5 days *post coitum* (dpc), when expression of *Sry* in the XY gonad kick-starts testis determination. SRY drives the subsequent expression of genes that both promote testis differentiation and inhibit ovarian differentiation, leading to specification of the testis. In cases where sex determination is disrupted, disorders of sexual development (DSD) can occur. Many genes have been described that are required for testis determination. However, we still know relatively little about how *Sry* itself is regulated.

MAP kinase (MAPK) signalling is involved in testis determination in mice and humans. MAP3K4 is required for testis determination in the mouse: loss of MAP3K4 is associated with male-to-female sex reversal in embryos and insufficient expression of *Sry* at 11.5 dpc. Likewise, loss of function of a protein known to interact with MAP3K4, GADD45 γ , also leads to sex reversal. In humans, mutations in the gene encoding the related kinase, MAP3K1, have been identified as a key cause of DSD. Thus, MAPK signalling appears to play a key role in testis determination both in humans and mice.

This thesis describes experiments designed to further elucidate the role of MAPK-related signalling in testis determination. Firstly, an attempt was made to generate mice bearing a human sex-reversing mutation in *Map3k1*, however this was not successful. Additionally, RNA-seq was employed to investigate transcriptional targets of GADD45 γ . Finally, the role of MAP3K4 as a dose-dependent modifier of sensitivity to sex reversal in mice is described, utilising a *Map3k4* transgene to rescue sex reversal in B6 mice harbouring a Y chromosome from the *Mus poschiavinus* strain. These mice undergo sex reversal in the absence of the *Map3k4* transgene, and also exhibit delayed *Sry* expression. By utilising this genetic tool, this thesis investigates the mechanism of sex reversal in these mice, with particular focus on the role of *Map3k4* in *Sry* expression.

Acknowledgments

First and foremost, thank you to my supervisor, Dr Andy Greenfield, for your guidance and support throughout my D.Phil. I am extremely grateful for your advice on matters science-related and otherwise, for always challenging me and for proofreading this thesis. My sincere thanks also go to all the members of the Sexual Development group who have helped me along the way. In particular, thanks to Nick Warr for masterminding the Y^{POS} project that forms the bulk of this thesis. Thanks also for teaching me so many good habits. Huge thanks also go to Gwenn Carre, for your patience in teaching me qRT-PCR, Bisulphite sequencing and cell sorting. I thank Pam Siggers for her guidance and sense of humour when it was needed. Huge thanks also to Catherine Livermore, Joel May and Abi Harris for your help with my many dissections and for providing cake in times of need. Many thanks also go to the rest of the Development lab, particularly Dominic Norris and Becca Walker. I'd also like to thank all the students at MRC Harwell for making the unit a thoroughly enjoyable place to spend four years.

I am very grateful to the MRC Harwell Core Services for their assistance throughout my D. Phil. In particular, my sincere gratitude goes to all the staff of the Mary Lyon Centre. In particular, thank you to Sara Wells, Lee Kent and Jackie Harrison, for your incredible knowledge and excellent care of all our mice. None of the work within this thesis would be possible without your continual support. Thanks also to Jeremy Sanderson and Kevin Glover for help with imaging, and to the Histology team for your hard work cutting sections. Also thank you to the Proteomics, Transgenics and Necropsy teams.

Thank you to my wonderful family, particularly my parents, for supporting me and helping me to be where I am today. Thank you to my brother and sister, to my grandparents and to my friends, especially Helen Oakes. Finally, I would like to thank my partner Gavin, whose support in the decade that we've spent together has been incredible. Without you I would not have had the courage to return to university as a mature student or to take on a D. Phil. Thank you for putting your own career on hold for seven years while I pursued my ambitions.

I also thank the Genetics Society and St Cross College, both of whom provided much-needed financial support that enabled me to present this work at the Seventh International VSD Symposium in Hawaii.

Finally, I thank the Medical Research Council for funding this research and my D.Phil.

Table of Contents

Table of Contents	iii
List of Figures	viii
List of Tables	x
Table 2.6: Identification of targeted <i>Map3k1</i> recombinants by restriction digest	x
Abbreviations	xi
Tail somite-staging of mouse embryos	xiii
Chapter 1. Introduction	1
1.1. Overview	2
1.2. Mammalian sex determination	3
1.2.1. Formation of the bipotential gonad	3
1.3. <i>Sry</i> : the genetic switch in sex determination	8
1.4. After <i>Sry</i> : canalising the male pathway	11
1.5. Understanding <i>Sry</i>	15
1.5.1. Structure and function of <i>SRY</i>	15
1.5.2. Expression of <i>Sry</i>	17
1.5.3. <i>Sry</i> regulation	20
1.6. MAPKs and the control of gene expression	31
1.6.1. MAPK signalling modules	31
1.6.2. MAPK interactions with chromatin	34
1.7. Disorders of sexual development (DSD)	35
1.8. The mouse as a model of DSD	38
1.9. Aims of this thesis	39
Chapter 2. Materials and Methods	40
2.1. General materials and methods	41
2.2. Mouse Lines and genotyping	42
2.2.1. Chromosomal sexing	42
2.2.2. <i>Gadd45g</i>	42
2.2.3. B6.Y ^{DOM} Lines	43
2.2.4. B6.Y ^{POS} - <i>Map3k4</i> Transgenic line	43
2.2.5. <i>Map3k4</i> knockout	44
2.2.6. <i>Sf1-EGFP</i>	44
2.2.7. 110 Region	44
2.3. Tissue collection and processing	46

2.3.1.	Embryo collection.....	46
2.3.2.	WMISH.....	46
2.3.3.	Quantitative real time PCR (qRT-PCR).....	46
2.3.4.	Western Blots.....	47
2.3.5.	Immunohistochemistry (IHC)	47
2.3.6.	Fluorescence-associated cell sorting (FACS)	47
2.3.7.	Necropsy and histology	48
2.4.	Cell biology	49
2.4.1.	Fluorescence-activated cell sorting (FACS)	49
2.5.	DNA methods	49
2.5.1.	DNA extraction for NGS.....	49
2.5.2.	DNA extraction for genotyping assays	50
2.5.3.	DNA extraction from ES cells in 96-well plates	50
2.5.4.	DNA extraction from FACS-sorted cells.....	51
2.5.5.	Polymerase chain reaction	51
2.5.6.	Gel electrophoresis	51
2.5.7.	Purification of PCR products	52
2.5.8.	Cloning.....	52
2.5.9.	DNA sequencing	54
2.6.	<i>Map3k1</i> gene targeting.....	60
2.6.1.	L189R construct design	60
2.6.2.	Vector preparation and transfection	61
2.6.3.	Strategy for screening ES cells.....	61
	Table 2.6: Identification of targeted <i>Map3k1</i> recombinants by restriction digest.....	62
2.6.4.	ES Cell screen.....	62
2.6.5.	Restriction digests of ES cell gDNA in 96 well plates.....	62
2.6.6.	Southern blotting	62
2.6.7.	Southern Hybridisation	63
2.7.	RNA Methods	64
2.7.1.	3' UTR riboprobe design.....	64
2.7.2.	Riboprobe synthesis	65
2.7.3.	Wholemound <i>in situ</i> Hybridisation (WMISH).....	66
2.7.4.	RNA extractions.....	67
2.7.5.	Reverse transcription	68
2.7.6.	qRT-PCR.....	68

2.7.7.	RNA Sequencing	71
2.8.	Protein methods.....	72
2.8.1.	Western blots	72
2.8.2.	Immunohistochemistry	75
Chapter 3.	Generation of mice bearing a human sex-reversing mutation in <i>Map3k1</i>	77
3.1.	Introduction	78
3.1.1.	<i>MAP3K1</i> mutations in human DSD.....	79
3.1.2.	Targeted mutagenesis in the mouse.....	82
3.1.3.	Aims.....	86
3.1.4.	The L189R construct.....	86
3.2.	Results	90
3.2.1.	Identification of targeted ES cell clones.....	90
3.3.	Discussion.....	93
3.3.1.	Summary	93
3.3.2.	<i>MAP3K1</i> mutations are a common cause of human DSD	93
3.3.3.	Future prospects	94
3.3.3.1.	CRISPR-Cas: a revolutionary tool for mouse genetics	94
3.3.4.	Future experiments.....	95
3.3.5.	Final remarks.....	97
Chapter 4.	Identifying transcriptional targets of GADD45 γ -mediated signalling in the developing mouse gonad by RNA-seq	98
4.1.	Introduction	99
4.1.1.	<i>Gadd45g</i> in the mouse gonad	99
4.1.2.	RNA-Sequencing.....	102
4.2.	Results	106
4.2.1.	Proof-of –principle	107
4.2.2.	Differential expression: XY <i>Gadd45g</i> ^{+/+} vs XY <i>Gadd45g</i> ^{-/-}	108
4.2.3.	Further analysis of downregulated transcripts	112
4.2.4.	Further analysis of <i>Pnliprp2</i>	117
4.3.	Discussion.....	118
Chapter 5.	The role of <i>Map3k4</i> in B6.Y ^{POS} sex reversal	125
5.1.	Introduction	126
5.1.1.	Genetic background affects sensitivity to XY sex reversal	126
5.1.2.	B6.Y ^{DOM} : a tool for studying <i>Sry</i> regulation	127
5.1.3.	<i>Map3k4</i> and <i>T</i> -associated sex reversal (<i>Tas</i>).....	129

5.1.4.	Mouse strain nomenclature	133
5.1.5.	Aims	133
5.2.	Results	134
5.2.1.	Analysis of adult B6.Y ^{POS} Tg mice.....	134
5.2.2.	<i>Sry</i> expression.....	138
5.2.3.	<i>Sox9</i> expression.....	141
5.2.4.	Immunohistochemical analysis of SRY and SOX9 expression	144
5.2.5.	Effect of loss of one copy of <i>Map3k4</i> on the gonadal phenotype of B6.Y ^{POS} embryos	149
5.3.	Discussion.....	152
5.3.1.	B6.Y ^{POS} Tg mice develop as normal males.....	152
5.3.2.	Overexpression of <i>Map3k4</i> leads to enhanced <i>Sry</i> expression in B6.Y ^{POS} embryos	152
5.3.3.	<i>Sox9</i> is modestly elevated in B6.Y ^{POS} Tg embryos	153
5.3.4.	Losing a copy of <i>Map3k4</i> exacerbates B6.Y ^{POS} sex reversal	154
5.3.5.	Variation and the analysis of B6.Y ^{POS} gene expression.....	154
5.3.6.	Summary and future directions	155
Chapter 6.	Molecular studies of B6.Y ^{POS} sex reversal.....	157
6.1.	Introduction	158
6.1.1.	Questions arising from Chapter 5	158
6.1.2.	Approaches to understanding <i>Sry</i> regulation	158
6.1.3.	Varying severity of B6.Y ^{POS} sex reversal	163
6.1.4.	Aims.....	165
6.2.	Results	166
6.2.1.	Analysis of p38 and GATA4 phosphorylation.....	166
6.2.2.	Expression of known regulators of <i>Sry</i> in B6.Y ^{POS} gonads	170
6.2.3.	Sequencing <i>Sry</i> ^{POS}	172
		177
6.2.4.	Analysis of CpG methylation in B6.Y ^{POS} gonadal somatic cells.....	186
6.2.5.	Assaying for presence of the 110 region in Harwell B6.Y ^{POS} mice.....	191
6.3.	Discussion.....	194
6.3.1.	p38 and GATA4 phosphorylation are not affected by overexpressing <i>Map3k4</i>	194
6.3.2.	<i>Wt1</i> and <i>Sf1</i> are not differentially expressed in B6.Y ^{POS} gonads, but GATA4 is elevated	195
6.3.3.	There are modest sequence differences between the <i>Sry</i> ^{B6} and <i>Sry</i> ^{POS} 5' flanking regions	196

6.3.4.	A preliminary study of <i>Sry</i> ^{POS} methylation	197
6.3.5.	The Harwell B6.Y ^{POS} colony does not possess the protective 110 region	198
Chapter 7.	General Discussion	200
7.1.	Overview	201
7.2.	Key findings	201
7.2.1.	<i>Pnliprp2</i> is expressed in the 11.5 dpc gonad and is downregulated in gonads lacking <i>Gadd45g</i>	201
7.2.2.	<i>Map3k4</i> modifies sensitivity to B6.Y ^{POS} sex reversal <i>via</i> an effect on <i>Sry</i> expression	202
7.2.3.	Delayed testis determination in B6.Y ^{POS} Tg embryos	203
7.2.4.	No clear mechanism for delayed <i>Sry</i> expression in B6.Y ^{POS} embryos	204
7.2.5.	POS-specific <i>Sry</i> polymorphisms	204
7.3.	Future work	205
7.3.1.	Further RNA-seq experiments.....	205
7.3.2.	The role of MAPKs in Y ^{POS} and beyond	206
7.3.3.	Closing remarks	208
	References.....	209
	Appendix 1: <i>Sry</i> ^{POS} sequence reads	221
	Appendix 2: Alignment of <i>Sry</i> ^{POS} and <i>Sry</i> ^{TIR} open reading frames	233
	Appendix 3: Annotated sequence alignment of <i>Sry</i> ^{B6} and <i>Sry</i> ^{POS}	234
	Appendix 4: <i>SRY</i> ^{AKR} sequence reads	239

List of Figures

Figure I: Tail somite-staging of mouse embryos	xiii
Figure 1.1 Schematic of the location of the urogenital ridge at 10.5 dpc	5
Figure 1.2: Cellular and molecular pathways of mammalian sex determination	14
Figure 1.3: SRY protein structure in humans and mice	17
Figure 1.4 Dynamics of <i>Sry</i> and <i>Sox9</i> expression	19
Figure 1.5 Pathways regulating <i>Sry</i>	22
Figure 1.6 <i>Sry</i> expression levels and timing are key in testis determination	24
Figure 1.7 Tissue specific, differentially methylated regions (T-DMRs) of <i>Sry</i>	28
Figure 1.8 MAPK signalling pathways	33
Figure 3.1 Collection of EGFP-positive and EGFP-negative cells by FACS.	49
Figure 3.1 Human sex-reversing <i>MAP3K1</i> mutations	80
Figure 3.2 Generation of mice carrying a targeted mutation derived in ES cells	85
Figure 3.3 Design of the L189R targeting construct	88
Figure 3.3. Southern blot screening strategy for the L189R targeting construct	89
Figure 3.5 Screening ES cells for L189R homologous recombinants	92
Figure 4.1 Potential roles for GADD45 γ in testis determination	101
Figure 4.2 cDNA library preparation and sequencing for RNA-seq	104
Fig 4.3 RNA-seq proof-of-principle	107
Figure 4.4 qRT-PCR analysis of top 9 downregulated genes from RNA-seq analysis	115
Figure 4.5 Analysis of <i>Pnliprp2</i> expression by WMISH	118
Figure 5.1 Overexpression of <i>Map3k4</i> rescues B6.Y ^{POS} sex reversal	132
Figure 5.2 Analysis of adult B6.Y ^{POS} Tg animals	136
Figure 5.3 Delayed <i>Sry</i> expression in B6-Y ^{POS} embryos is ameliorated by overexpressing <i>Map3k4</i>	140
Fig 5.4 Delayed <i>Sox9</i> expression in B6.Y ^{POS} embryos is ameliorated by overexpressing <i>Map3k4</i>	143
Figure 5.4 SRY expression in B6.Y ^{POS} Tg gonads	146
Figure 5.5 SOX9 expression in B6.Y ^{POS} Tg gonads	148
Figure 5.6 Loss of a copy of <i>Map3k4</i> exacerbates B6.Y ^{POS} sex reversal	151
Figure 6.1 Western blot analysis of p38 and GATA4	168
Figure 6.2 Expression of known regulators of <i>Sry</i> in B6, B6.Y ^{POS} and B6.Y ^{POS} Tg gonads	171
Figure 6.3 Strategy for sequencing the flanking regions of <i>Sry</i>	174

Figure 6.4 Summary of sequence polymorphisms in <i>Sry</i> ^{POS} flanking regions	177
Figure 6.5 Alignment of <i>Sry</i> ^{B6} and <i>Sry</i> ^{POS} and <i>Sry</i> ^{AKR} CA repeat regions	180
Figure 6.6 Alignment of sequences from <i>Sry</i> ^{B6} , <i>Sry</i> ^{AKR} and <i>Sry</i> ^{POS} containing CpG dinucleotides of the <i>Sry</i> T-DMR	182
Figure 6.7 Alignment of <i>Y</i> ^{AKR} , <i>Y</i> ^{B6} and <i>Y</i> ^{POS} fragments containing polymorphisms 5' R5 and 5' R6	185
Figure 6.8 qRT-PCR analysis of sorted cells	188
Figure 6.9 Analysis of methylation of the <i>Sry</i> T-DMR in B6. <i>Y</i> ^{POS} gonadal somatic cells	190
Figure 6.10 Genotyping for the 110 region in Harwell B6. <i>Y</i> ^{POS} mice	192

List of Tables

Table 2.1 Standard solutions	41
Table 2.2 Genotyping PCR primers and assay conditions	45
Table 2.3 PCR assays used to amplify <i>Sry</i> for sequencing	55
Table 2.4 Primers sequences for amplification of <i>Sry</i> ORF and flanking regions	56
Table 2.5 Primers used for <i>Sry</i> sequencing	57
Table 2.6: Identification of targeted <i>Map3k1</i> recombinants by restriction digest	62
Table 2.7. Primers and PCR assay conditions for synthesis of <i>Pnliprp2</i> riboprobe	64
Table 2.8. Conditions for RNA probe synthesis	65
Table 2.9 Sequences of qRT-PCR primers	69
Table 2.10 qRT-PCR assay efficiency	70
Table 2.11 Antibodies used for Immunoblotting	74
Table 2.12 Antibodies used for immunohistochemistry	75
Table 4.1 RNA seq sample quantity and quality	106
Table 4.2 Differentially expressed genes in WT XX vs XY samples	108
Table 4.3 Top 44 differentially expressed genes ranked by <i>p</i> value	109
Table 4.4 Top 10 down-regulated transcripts in XY ^{-/-} gonads	111
Table 5.1 Summary of <i>p</i> values from qRT-PCR profiling of <i>Sry</i> expression	139
Table 5.2 Summary of <i>p</i> values from qRT-PCR profiling of <i>Sox9</i> expression	142
Table 5.3 Phenotypes of 14.5 dpc gonads from B6.Y ^{POS} <i>Map3k4</i> ^{+/+} and B6.Y ^{POS} <i>Map3k4</i> ^{-/-} gonads	150
Table 6.1 <i>Sry</i> ^{B6} - <i>Sry</i> ^{POS} polymorphisms	178
Table 6.2 Predicted transcription factor binding	184
Table 6.3 Assays to determine presence of the 110 region	191

Abbreviations

AMH	Anti-Müllerian hormone
BAC	Bacterial artificial chromosome
bp	base pairs
BSA	Bovine serum albumin
CBX2	chromobox homolog 2
CD	Campomelic dysplasia
cDNA	Complementary DNA
CGD	Complete gonadal dysgenesis
ChIP	Chromatin immunoprecipitation
CpG	Cytosine-phosphate-guanine (cytosine 5' to guanine)
Ct	Cycle threshold
DIG	Digoxigenin
DMRT1	Double-sex and Mab-3 related transcription factor 1
DNA	Deoxyribonucleic acid
dpc	Days <i>post coitum</i>
DSD	Disorder of sexual development
EGFP	Enhanced-GFP
ENU	<i>N</i> -ethyl <i>N</i> -nitrosurea
ERK	Extracellular signal-related kinase
FGF	Fibroblast growth factor
FGF9	Fibroblast growth factor 9
FGFR	Fibroblast growth factor receptor
FGFR2	FGF receptor 2
FOG	Friend of GATA
FOXL2	Forkhead box L2
FSH	Follicle stimulating hormone
GADD45g	Growth arrest DNA damage inducible protein 45 gamma
GATA4	GATA-binding protein 4
GD	Gonadal dysgenesis
gDNA	Genomic DNA
H&E	Haematoxylin and Eosin
H3K4	Histone 3 lysine 4
H3K9	Histone 3 lysine 9
HISS	Heat-inactivated sheep serum
HMG	High mobility group
HPRT1	hypoxanthine phosphoribosyltransferase 1
IHC	immunohistochemistry
JMJD1A	Jumonji domain-containing 1A
JNK	c-Jun amino terminal kinase
LB	Luria Bertani broth
LHX9	Lim homeobox protein 9
MAP	Mitogen activated protein
MAP2K	MAP kinase kinase
MAP3K	MAP kinase kinase kinase
MAPK	MAP kinase

MeOH	Methanol
miRNA	Micro RNA
mRNA	Messenger RNA
NaCl	Sodium chloride
ORF	Open reading frame
PBS	Phosphate-buffered saline
PDGS	Prostaglandin D2 synthase
PFA	paraformaldehyde
PGC	Primordial germ cell
PGD2	Prostaglandin D2
PK	Proteinase K
qRT-PCR	Quantitative real-time reverse-transcriptase PCR
RNA-seq	RNA sequencing
RSPO1	R-spondin 1
RT	Room temperature
SD	Standard deviation
SEM	Standard error of the mean
Seq	Sequencing
SF1	Steroidogenic factor 1
SIX	sine oculis-related homeobox
SNP	Single nucleotide polymorphism
SOX9	SRY box protein 9
SRY	Sex-determining region of the Y chromosome
SSC	Saline-sodium citrate buffer
STRA8	Stimulated by retinoic acid 8
TAD	Transcription activator domain
TBE	Tris/borate/EDTA buffer
TBS	Tris-buffered saline
TBST	Tris-buffered saline with tween
T-DMR	Tissue-specific differentially methylated region
TE	Tris-EDTA
TESCO	Testis-specific enhancer of <i>Sox9</i> core
ts	Tail somites
TSS	Transcription start site
UGR	Urogenital ridge
UTR	Untranslated region
WMISH	Wholemount <i>in situ</i> hybridisation
WNT4	Wingless-type MMTV integration site family 4
WT	Wild-type
WT1	Wilms tumour suppressor 1

Tail somite-staging of mouse embryos

Throughout this thesis, timing of the development of the gonad is discussed in relation to the age of the embryo, given in tail somites. This page is intended to orientate the reader and provide a frame of reference for the staging of the embryos used in this work.

Somites are segmented regions of mesoderm that arise along the dorsal aspect of the embryo, either side of the neural tube. The timing of somites is closely coordinated and one somite emerges approximately every two hours (Tam, 1981). Due to this, precise staging of embryos can be carried out by counting somites. In embryos aged 11.0 dpc to 12.5 dpc, tail somites are counted. These are the somites posterior to the hindlimb (Fig I). Table I gives a guideline for the relationship between tail somites and embryo age in days *poist coitum* (dpc).

Table I: Tail somite stages of mouse embryos

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

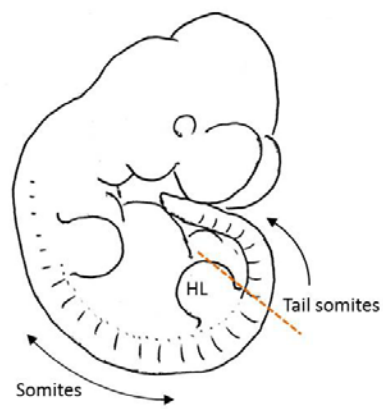


Figure I: Tail somite-staging of mouse embryos
Somites form along the dorsum of the embryo, either side of the neural tube. Tail somites (ts) are counted from the posterior limit of the hindlimb (HL) to the end of the tail.

Chapter 1. Introduction

1.1. Overview

Sexual differentiation is the process by which an organism develops as either male or female and is essential for sexual reproduction in mammals. Different organisms utilise varying mechanisms of sex determination. Some species, such as many reptiles and fish, depend upon the environment for sex determination, whereas in mammals, such as humans and mice, the process is governed chromosomally. In mammals, the process begins at fertilisation, with the establishment of the chromosomal sex of an individual: commonly either XY or XX. Primary sex determination is the process by which the embryonic gonad commits to either the testicular or ovarian fate. Gonadal fate is established by chromosomal sex: XY embryos usually develop testes whilst XX embryos develop ovaries. Sex determination in the gonad drives the subsequent processes of sexual differentiation that result in the development of male or female reproductive systems, including internal and external genitalia (Swain and Lovell-Badge, 1999).

The developing gonadal primordium is initially bipotential: irrespective of its chromosome constitution it retains the capacity to develop as either a testis or ovary. Its subsequent differentiation is principally governed by the presence or absence of the dominant male determinant, *SRY*, encoded by the Y chromosome. In the mouse, sex determination occurs at 11.5 days *post coitum* (dpc) and expression of *Sry* during this critical period is necessary for testis determination to occur (Sekido and Lovell-Badge, 2009). *SRY* drives the expression of genes that both promote testis differentiation and inhibit ovarian differentiation, thus tipping the balance of two mutually-antagonistic genetic pathways towards a testicular fate (Brennan and Capel, 2004). In cases where sex determination is disrupted, discrepancies can occur between

chromosomal and gonadal sex. This phenomenon is broadly referred to as sex reversal in mice, but is manifested as a variety of human disorders that will be discussed later in this chapter.

Many genes have been linked with the control of *Sry* expression. Of particular interest within this thesis is the mitogen activated protein (MAP) kinase pathway, which has been shown to be important for testis determination in both mice and humans (Bogani et al., 2009, Pearlman et al., 2010, Warr et al., 2012, Warr et al., 2014). The research described herein attempts to further elucidate the role of the MAP kinase pathway in testis determination in the mouse, with particular focus upon how it is implicated in the regulation of *Sry* expression. A description of our current understanding of murine testis determination, as well as the function and regulation of *Sry*, will be given in this introductory chapter.

1.2. Mammalian sex determination

1.2.1. Formation of the bipotential gonad

Sex determination in mammals is initiated after the formation of the urogenital ridge. This structure arises from intermediate mesoderm from around 9.0 dpc in the mouse embryo, forming a pair of mounds running in an antero-posterior orientation along the body wall, either side of the neural tube and dorsal aorta. The gonad forms as a thickening of the coelomic epithelium on the ventromedial surface of the mesonephros, a structure that gives rise to the mesonephric (Wolffian) duct. The bilateral gonads and mesonephroi comprise the urogenital ridge, along with two other structures: the pronephros, which has no known function and regresses shortly after it forms, and the metanephros, which gives rise to the kidneys (Swain and Lovell-Badge, 1999) (Fig 1.1).

The proliferation of coelomic epithelial cells gives rise to at least two lineages of bipotential somatic precursors. The first, the supporting cell lineage, is absolutely required for the initiation of sex determination. It is within these cells that the expression of *Sry* occurs. The supporting cell progenitors differentiate into either Sertoli cells in the testis, or granulosa cells in the ovary, and play a key role in supporting germ cells as they develop. A second lineage, comprising steroidogenic cell precursors, will differentiate either into Leydig cells of the testis or theca cells of the ovary. A third somatic lineage, the connective tissue, includes stromal cells, endothelial cells and peritubular myoid cells, which are important for sex-specific organisational structure of the testis or ovary following sex determination. Finally, primordial germ cells (PGCs) originate outside of the UGR in a region of extraembryonic mesoderm posterior to the primitive streak. They form at 6.25 dpc but do not arrive in the gonad until between 10.5 and 11.5 dpc, prior to which they migrate through the gut mesentery towards the genital ridge, to populate the gonad. PGCs later respond to sexually dimorphic signals from the somatic cells, which drive their differentiation into either oogonia or spermatogonia (Swain and Lovell-Badge, 1999).

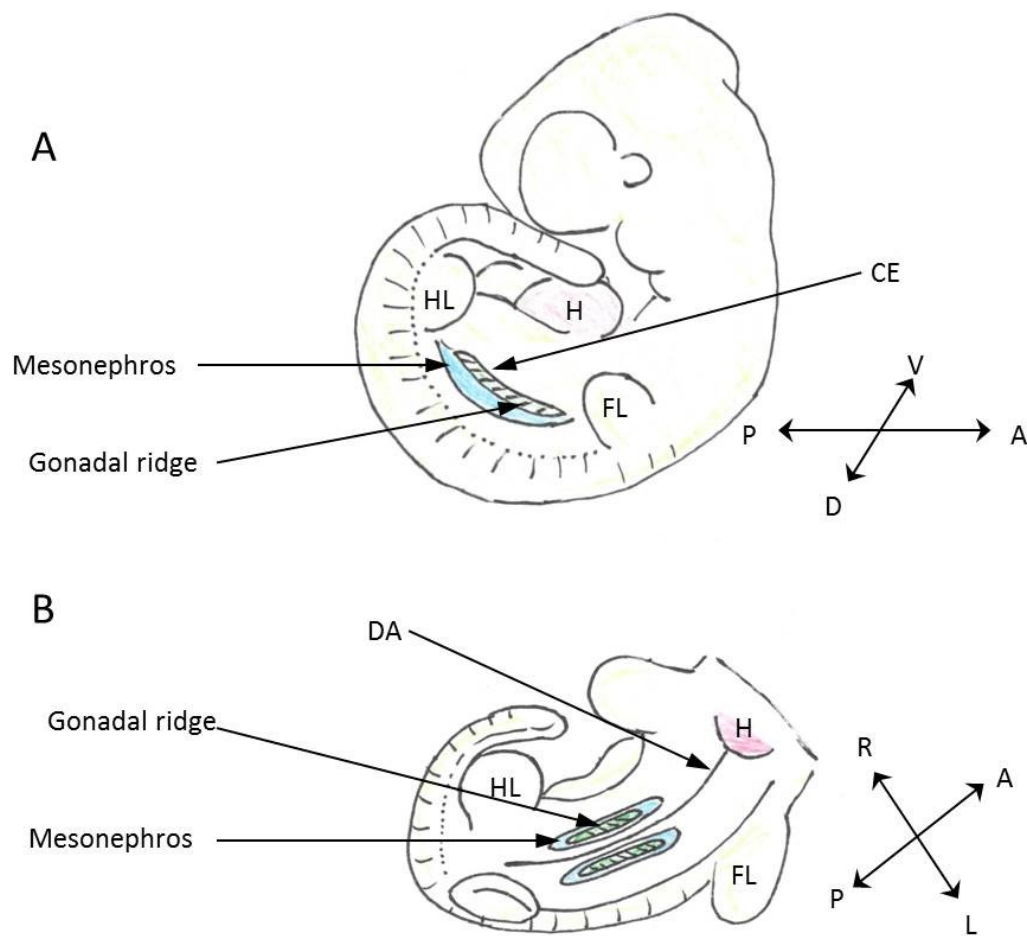


Figure 1.1 Schematic of the location of the urogenital ridge at 10.5 dpc

(A) The urogenital ridges are located in the intermediate mesoderm running along the dorsal body wall. The gonadal ridge (hatched area) arises from a thickening of the coelomic epithelium (CE) on the ventromedial surface of the mesonephros (blue).

(B) The gonadal ridge is located on the ventromedial surfaces of the bilateral mesonephroi, either side of the dorsal aorta (DA). FL, forelimb; H, heart; HL, hindlimb; A↔P, anteroposterior axis; D↔V, dorso-ventral axis; L↔R, left-right axis.

Characterisation of gene expression in the somatic cells of the bipotential gonad has identified several genes required for establishment of the population of bipotential somatic cells in the gonad. These include *Wt1* (Wilms tumour suppressor 1) (Hammes et al., 2001), *Lhx9* (Lim homeobox protein 9) (Birk et al., 2000), *Emx2* (empty spiracles homeobox 2) (Kusaka et al., 2010), *Sf1* (steroidogenic factor 1) (Luo et al., 1994), *Cbx2* (chromobox 2; also known as *M33*) (Katoh-Fukui et al., 2012, Katoh-Fukui et al., 1998), *Gata4* (Hu et al., 2013) and *Six1/Six4* (sine oculis-related homeobox 1/4) (Fujimoto et al., 2013). Of particular relevance to this thesis are *Sf1*, *Wt1* and *Gata4*, which will be discussed in more detail below.

SF1, also known as NR5A1 (nuclear receptor subfamily 5, group A, member 1) is an orphan nuclear receptor that is required for the early stages of gonad development, and more widely for steroidogenesis. Mice lacking *Sf1* fail to develop adrenal glands and gonads and exhibit severe deficiency in steroid hormones such as corticosterone (Luo et al., 1994). In the gonad, *Sf1* is expressed exclusively in the somatic cells and is thought to regulate their survival and proliferation. Gonads of embryos with a targeted disruption of *Sf1* initially form gonads, but they degenerate between 11.0 and 11.5 dpc, due to increased apoptosis (Swain and Lovell-Badge, 1999). Additionally, *Sf1* also regulates expression of *Sry*, a role that will be discussed in the upcoming sections.

Wt1 is expressed from 9.0 dpc in the urogenital ridge of the mouse embryo (Armstrong et al., 1993) and is required for the formation of the bipotential gonad and kidneys. Mice lacking *Wt1* die early in embryonic development and fail to form gonads or kidneys (Kreidberg et al., 1993). Additionally, gonads of *Wt1* knockout embryos fail to express *Sf1* (Wilhelm and Englert, 2002). *Wt1* encodes at least 24 splice isoforms, which all contain 4 zinc fingers at the C terminal domain. Within this domain, the

presence or absence of a three amino acid sequence, KTS (lysine, threonine, serine), confers distinct functions. These splice variants are accordingly designated +KTS or –KTS. Targeted disruption of the –KTS variant leads to regression of both XX and XY gonads by 12.5 dpc, whereas embryos lacking the +KTS variant specifically exhibit male-to-female gonadal sex reversal (Hammes et al., 2001). In early gonad development, the -KTS variant binds and activates the *Sf1* promoter, and is therefore thought to promote the survival and proliferation of the early gonadal primordium (Wilhelm and Englert, 2002). In contrast, the +KTS variant is thought to play a male-specific role in the regulation of *Sry* expression (Bradford et al., 2009).

Gata4 encodes a member of the GATA family of zinc-finger transcription factors that recognise the consensus motif 5'AGATAG3' (Arceci et al., 1993). *Gata4* expression has recently been shown to precede the thickening of the coelomic epithelium, and it is thought to be required for the earliest stages of gonad formation (Hu et al., 2013). Embryos lacking *Gata4* die between 7.0 – 9.5 dpc due to failure of ventral morphogenesis and heart tube formation (Kuo et al., 1997, Molkentin et al., 1997), but conditional ablation of *Gata4* at 8.5 dpc has demonstrated that it is required for thickening of the coelomic epithelium (Hu et al., 2013). These embryos also fail to express early gonadal markers *Lhx9* and *Sf1*. GATA proteins require interaction with zinc finger cofactors of the FOG (friend of GATA) family for their normal function. GATA4 associates with FOG2 (also known as ZFPM2), however, knockout of *Fog2* is not associated with disruption to early gonad formation, but instead results in XY sex reversal and reduced expression of *Sry* (Tevosian et al., 2002). A *Gata4* mutant harbouring a knock-in allele that abrogates the protein's interaction with FOG2 shows a similar phenotype to the *Fog2* knockout, suggesting that GATA4/FOG2 interaction is

required for normal expression of *Sry* (Tevosian et al., 2002). Together, these reports suggest that GATA4 has distinct functions at different stages of gonad development, both requiring FOG2 interaction and also independent of it.

1.3. *Sry*: the genetic switch in sex determination

Until approximately 11.2 dpc, the bipotential gonad is poised in a state of readiness for genetic cues driving it towards a male or female fate. At this point, the transcriptomes of XX and XY gonads are indistinguishable (Nef et al., 2005, Munger et al., 2013). In XY gonads, the expression of *Sry* from 10.75 dpc initiates the testis-determining pathway and drives the expression of pro-testis genes, enabling the gonad to develop as male. In the XX gonad, the absence of *Sry* means that the testis-determining pathway is not initiated and the ovary-determining pathway instead dominates, enabling the gonad to develop a female identity. The genetic processes governing development of the bipotential gonad as either male or female are mutually antagonistic, with genetic cues acting to both promote one pathway whilst simultaneously inhibiting the other. In the XY gonad, it is the expression of *Sry* that initially tips this balance towards the male fate, and for this reason *Sry* has been denoted the “master switch” in testis determination (Kashimada and Koopman, 2010).

The existence of a dominant testis-determining Y-encoded factor was discovered through an elegant pairing of human and mouse genetics that spanned some 30 years. The first major discoveries took place in the late 1950s, with studies of sex chromosome aneuploidy indicating that the Y chromosome carries a male-determining factor (Ford et al., 1959, Jacobs and Strong, 1959). Much later, in 1990, a comprehensive search for conserved sequences between Y chromosome translocations from human XX male patients led to the discovery of the human *SRY*

gene (Sinclair et al., 1990). Around the same time, the mouse sex-determining region was also identified in studies of XY female mice carrying a Y chromosome deletion (Gubbay et al., 1990). Shortly after this, a 14 kb genomic fragment containing mouse *Sry* was shown to be sufficient for male development of chromosomally female mice, confirming that it is the genetic switch driving male sex determination (Koopman et al., 1991). Together, these studies demonstrated that *Sry* was the key Y-linked testis-determining gene in humans and mice, and its importance has since been demonstrated in almost all mammalian species studied (Sekido and Lovell-Badge, 2009). Crucially, in the mouse, the spatiotemporal expression profile of *Sry* is tightly controlled: it is only expressed from 10.75 dpc to 12.5 dpc. It exhibits a wave-like pattern of expression, initiating in the centre of the gonad and spreading to the poles, and reaches a peak at 11.5 dpc (Bullejos and Koopman, 2001). This tightly coordinated pattern of expression is critical for SRY function and shall be discussed later in this chapter.

Sry is a short, single exon transcript that encodes the founding member of the SOX (SRY-related HMG box) family of transcription factors, comprising 20 members (Kamachi and Kondoh, 2013). *Sry* is thought to have evolved from the X-linked gene *Sox3*. *Sox3* is the most similar to *Sry* of all the SOX proteins, sharing some 90% similarity across the HMG domain. In support of the hypothesis that *Sry* arose from this X-linked counterpart, chromosomal rearrangements involving the regulatory regions of *Sox3* have been reported in several cases of XX sex-reversal in humans (Lim et al., 2000, Moalem et al., 2012, Haines et al., 2015). These studies implicate ectopic expression of *Sox3* in the developing gonadal ridge as a putative cause of sex reversal. In further support of this, Sutton *et al* found that ectopic expression of *Sox3* in the

mouse genital ridge causes XX sex reversal. In their report, the authors found that integration of a *Sox3* transgene upstream of the gonadally expressed gene *Aldh1a1* led to its expression in the bipotential gonad, and subsequent upregulation of *Sox9* in XX gonads, leading to sex reversal (Sutton et al., 2011). This suggests that by acquiring the ability to express in the bipotential gonad, the proto-Y allele of *Sox3* may have emerged as the mammalian testis-determining factor and led to the divergence of the sex chromosomes into the X and Y.

Another member of the SOX family, *Sox9*, is SRY's most compelling target, based on a variety of experimental data (Sekido and Lovell-Badge, 2008), although other targets have also been reported (Li et al., 2014). SOX9 is a vital transcription factor required for development of many tissues, in particular in the development of the skeleton and chondrogenesis (Akiyama, 2008). In humans, mutations in *SOX9* are associated with campomelic dysplasia (CD), a skeletal dysmorphology syndrome. Often, patients with CD also exhibit XY sex reversal, and on this basis it was hypothesised that *SOX9* might be involved in sex determination (Foster, 1996). Studies in mice in which deletion of *Sox9* caused XY sex reversal confirmed its requirement for testis determination (Chaboissier et al., 2004, Barrionuevo et al., 2006), while gain-of-function analyses also demonstrated that *Sox9* is sufficient to drive testis development in XX mouse embryos (Vidal et al., 2001).

Sox9 is expressed at low levels in the somatic cells of both XX and XY gonads before *Sry* is expressed and this is thought to be mediated by SF1 at this early stage (Sekido and Lovell-Badge, 2009). However, at 11.5 dpc, it is strongly upregulated specifically in XY gonads. This upregulation of *Sox9* requires synergistic binding of SRY and SF1 to an enhancer called the testis-specific enhancer of *Sox9* core element, or TESCO (Sekido

and Lovell-Badge, 2008), and occurs shortly after *Sry* transcript reaches its peak at 11.5 dpc. Unlike *Sry*, which diminishes after this stage, *Sox9* expression persists and becomes self-regulating, maintaining expression by binding to its own enhancer and also via positive feedback with downstream pro-testis genes, discussed below.

Sox9 upregulation by SRY promotes the testis-specific program of gene expression in the gonad, driving Sertoli cell differentiation and, thereby, other cell-fate decisions that lead to differentiation and proliferation of the principal cell types of the testis, as well as their arrangement to form characteristic testicular structures, the testis cords.

1.4. After *Sry*: canalising the male pathway

Following SRY activation of *Sox9* expression, the commitment of the somatic precursors to the male fate must be maintained and propagated. SOX9 is essential for this process and maintains its own expression via two means: firstly, it is able to autoregulate its own expression by directly binding to its own TESCO enhancer, achieved through its HMG domain (Sekido and Lovell-Badge, 2008). Secondly, SOX9 activates expression of fibroblast growth factor 9 (FGF9), a paracrine growth factor, which in turn maintains *Sox9* expression by a positive feedback loop (Kim et al., 2006). FGF9 is required for the maintenance of *Sox9* expression after 11.5 dpc and for the proliferation of Sertoli cells (Kim et al., 2007). Mice lacking FGF9 or its receptor FGFR2 exhibit reduced *Sox9* expression and XY sex reversal (Colvin et al., 2001, Bagheri-Fam et al., 2008, Siggers et al., 2014). In addition to maintaining *Sox9* expression, FGF9 also antagonises WNT4 signalling, which is required for maintaining female fate (Kim et al., 2006). Indeed, analysis of *Fgf9/Wnt4* double knockouts suggests that FGF9's role is primarily inhibitory (Jameson et al., 2012a). To further reinforce propagation of the male signal, *Sox9* also facilitates paracrine recruitment of neighbouring cells to the

Sertoli cell lineage. This is achieved by prostaglandin D2 (PGD2), which is synthesised by the enzyme prostaglandin D2 synthase (PGDS). The gene encoding PGDS, *Ptdgs*, is positively regulated by SOX9, and PGDS is able to induce expression of *Sox9* and subsequent Sertoli cell differentiation (Wilhelm et al., 2005, Wilhelm et al., 2007, Moniot et al., 2009). Thus, maintenance of *Sox9* expression in the XY gonad canalises the process of gonad development, driving the pro-testis pathway whilst simultaneously inhibiting the ovarian pathway.

By 12.5 dpc male fate is well established and male-specific differentiation and proliferation take place. Tightly coordinated cellular processes, involving the introgression of endothelial cells from the mesonephros, guide this process, and lead to the formation of the characteristic vasculature and cords of the testis (Brennan et al., 2003, Cool et al., 2011, Combes et al., 2009). At 12.5 dpc it is possible to distinguish male and female gonads by their morphology, as the early testis cords can be observed, as well as the prominent coelomic vessel, which runs along the ventral surface of the coelomic epithelium. The cords, consisting of Sertoli cells surrounding the germ cells, are encapsulated by peritubular myoid cells (Wilhelm and Koopman, 2006). At this stage, germ cells enter mitotic arrest and do not commence meiosis until after birth, when they differentiate into spermatogonia (Kocer et al., 2009). Foetal Leydig cells, also arising from SF1-positive somatic precursors, are found in the interstitium between testis cords from 12.5 dpc (Brennan and Capel, 2004).

In the XX gonad, absence of *Sry* expression and subsequent expression of pro-testis genes enables an ovary-specific genetic program to proceed. This involves expression of genes encoding members of the canonical WNT signalling pathway, including WNT4, β -catenin (CTNNB1) and R-spondin 1 (RSPO1), which promotes *Wnt4* and *Ctnnb1*

expression. The transcription factor forkhead box L2 (FOXL2) is also a key ovary determinant and is required for maintenance of ovarian identity (Uhlenhaut et al., 2009). These genes act in concert to drive granulosa cell differentiation and inhibit Sertoli cell differentiation, with no single gene playing the role of a master ovary-determining gene whose absence results in XX sex reversal. The first cellular event that distinguishes the specified female gonad is the entry of germ cells into meiosis. This occurs at 13.5 dpc and is triggered by retinoic acid (Bowles et al., 2006, Koubova et al., 2006). Ovarian germ cells can be identified by their expression of *Stra8*, a gene stimulated by retinoic acid (Koubova et al., 2006, Koubova et al., 2014). The molecular and cellular processes leading to sex determination are summarised in Fig 1.2.

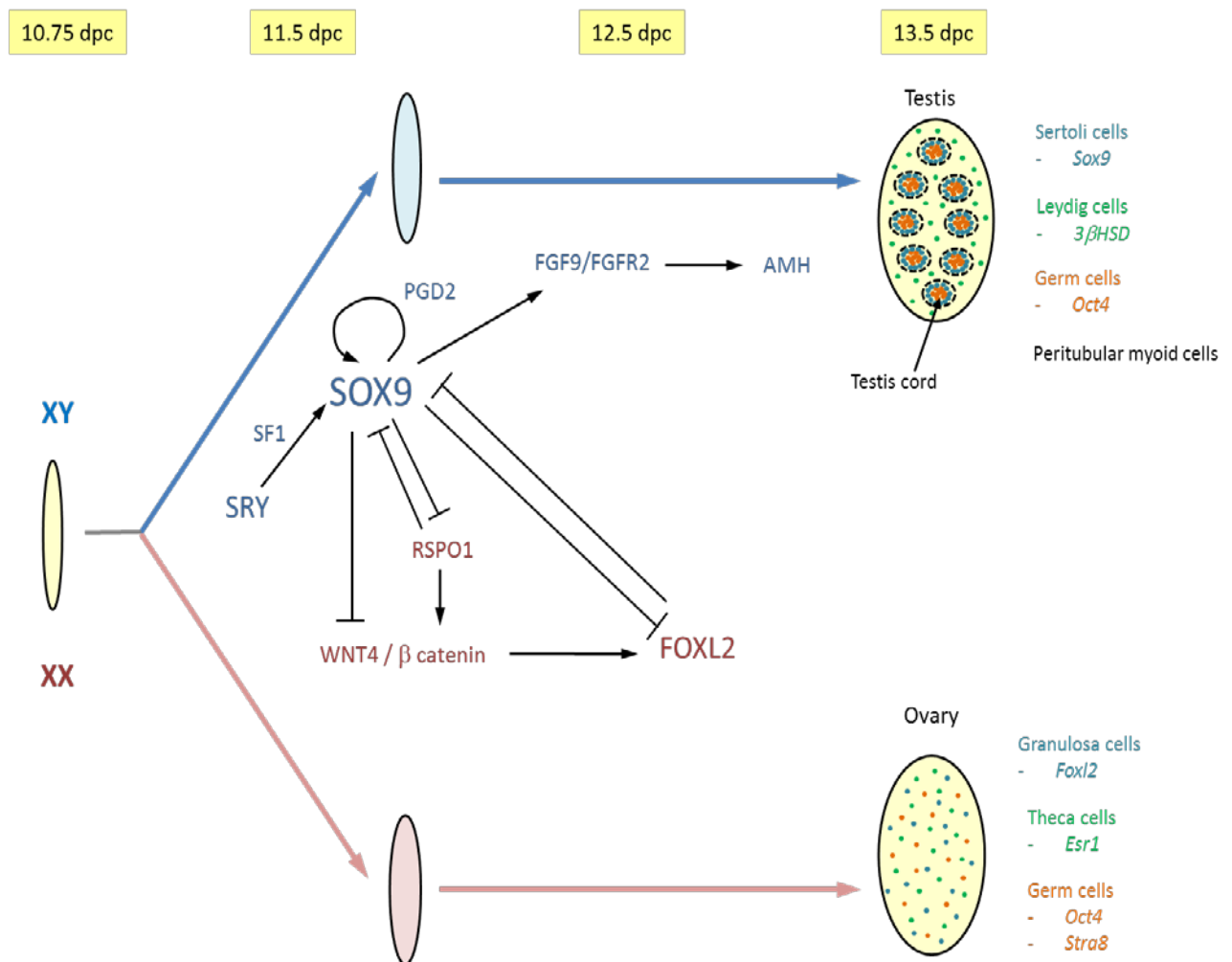


Figure 1.2: Cellular and molecular pathways of mammalian sex determination

At 10.75 dpc, *Sry* is expressed in the XY gonad and initiates the pro-testis pathway by upregulating *Sox9*. *Sox9* expression is maintained in a positive feedback loop. SOX9 upregulates FGF9/FGFR2 and inhibits expression of WNT pathway members and *Foxl2*. By 13.5 dpc, the specified testis contains Sertoli cells (blue), Foetal Leydig cells (green) and germ cells (orange) arranged in testis cords encapsulated by peritubular myoid cells (black dashed line). In the XX gonad, *Sry* is not expressed so pro-ovarian genes dominate, leading to development of the ovary. By 13.5 dpc the specified ovary contains Granulosa cells (blue), Theca cells (green) and meiotic germ cells (orange). Markers for identification of cell lineages in the 13.5 dpc gonad are given in the appropriate colour. Pro testis genes are shown in blue whereas pro-ovarian genes are shown in pink.

Following commitment of gonadal somatic cells to their respective male and female programs, their fate remains plastic, and maintenance of directive signals is required throughout life. Ovarian fate is maintained postnatally by FOXL2, which directly represses *Sox9* expression by binding to the TESCO enhancer. Loss of FOXL2 in adult life leads to transdifferentiation of granulosa cells into Sertoli cells, and the expression of testosterone by Leydig-like cells (Uhlenhaut et al., 2009). Conversely, loss of the transcription factor DMRT1 (doublesex and mab-3 related transcription factor 1), which is not required for embryonic testis determination (Raymond et al., 2000), causes male-to-female transdifferentiation of somatic lineages when deleted in the post-natal testis (Matson et al., 2011). Thus, there is a continued requirement for active signals to maintain testicular or ovarian fate.

1.5. Understanding *Sry*

1.5.1. Structure and function of SRY

The SOX family of transcription factors is characterised by the high mobility group (HMG) DNA-binding domain. Intriguingly, this appears to be the only domain highly conserved between SRY proteins of different species (Kashimada and Koopman, 2010). The 79 amino acid HMG domain of SRY binds the motif (A/T)ACCA(T/A) and induces bending of DNA by 60-85° (Harley et al., 1994), a property that is essential for SRY function. Whilst, so far, the HMG domain is the only evolutionarily conserved part of SRY, there is evidence of species-specific functions of other regions of murine SRY, which will be discussed below. It could be argued that, since it is the only evolutionarily conserved domain, it is perhaps the only one required for SRY to function. In support of this, XX mice transgenically expressing either human or goat *SRY* develop as male (Lovell-Badge et al., 2002, Pannetier et al., 2006). However, contrary to this hypothesis, human *SRY* must be of full-length to bind DNA *in vitro* (Dubin and Ostrer,

1994), so it is possible that non-HMG domains of the protein may be conserved in their function, but perhaps not in sequence.

Mouse SRY is unusual. Whereas SRY proteins of other mammals contain an N-terminal domain of 30-60 amino acids, in mouse this consists of only two (Gubbay et al., 1990). Additionally, mouse SRY contains a long C-terminal region comprising a bridge domain and a long polyglutamine (polyQ) tract, encoded by tandem CAG repeats (Fig 1.3). The size of this region varies between strains, consisting of 8 (*Mus domesticus*) to 21 (*Mus musculus*) blocks of 2 to 13 glutamine residues, interspersed by histidine-rich spacer sequences (Coward et al., 1994, Albrecht and Eicher, 1997). Absence of the polyQ domain has been linked with failure to cause male development in XX transgenic embryos (Bowles et al., 1999). Initially it was unclear whether it was required for the activity of SRY or for stability of the protein, and it was suggested that ablation of this domain led to premature degradation. More recently it has been demonstrated that both may be the case. The polyQ domains of both *M. musculus* and *M. domesticus* are able to transactivate a TESCO reporter *in vitro* and are required for SRY to upregulate *Sox9* expression (Chen et al., 2013c, Zhao et al., 2014). In addition, the size of the C-terminal domain, rather than the sequence, was found to be crucial for stability of *Sry* protein. Ablation of the polyQ domain *in vitro* resulted in degradation via the proteasomal pathway, but its replacement with mCherry rescued it from degradation (Zhao et al., 2014). Together these data suggest that the polyQ tract acts both as a protein stabilisation domain and as a transcriptional activator domain (TAD).

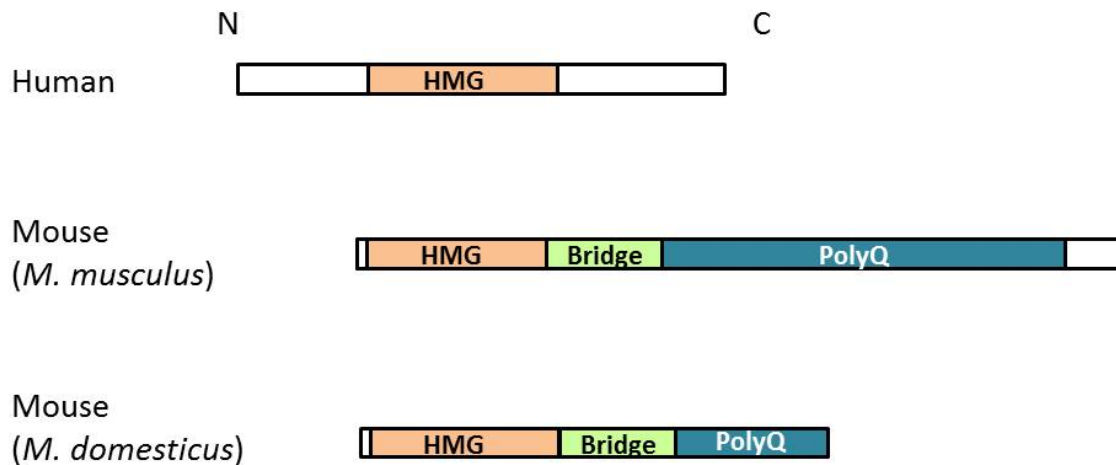


Figure 1.3: SRY protein structure in humans and mice

Human *SRY* consists of N- and C-terminal domains (white) and the highly-conserved HMG domain (orange). Mouse *SRY* shares the HMG domain but possesses a bridge domain (green) and long glutamine-rich domain (PolyQ domain, blue). In *M. musculus* *SRY*, this domain is considerably longer than that of *M. domesticus* strains.

It has been suggested that human *SRY*, and that of other mammals, requires a protein partner for transcriptional activation of its target, because human *SRY* has no intrinsic transactivation activity *in vitro* (Dubin and Ostrer, 1994) and *SRY* proteins of most mammals lack domains that share homology with other known TADs. Therefore, it has been postulated that the acquisition of the TAD-functioning polyQ domain in mouse *SRY* may have abrogated the need for a partner protein for transcriptional activation.

1.5.2. Expression of *Sry*

In the mouse, *Sry* expression is transient and follows a distinct spatiotemporal pattern (Fig 1.4). *Sry* is initiated in the somatic cells in the central region of the gonad at around 10.75 dpc, and spreads in a wave-like pattern towards the poles until *Sry* transcript is detectable throughout the entire length of the gonad, reaching a peak at 11.5dpc. After this peak, it begins to be downregulated in a pattern mirroring its onset, with the last remaining expression being in the caudal gonadal pole, and becoming

undetectable by 12.5dpc (Bullejos and Koopman, 2001). Fundamental experiments using epitope tagging of SRY have estimated that, on a cellular level, the protein itself is only present for around 6 hours (Sekido et al., 2004). Thus, in this short time, it must be able to sufficiently upregulate the expression of *Sox9* via TESCO to facilitate its autoregulation and lead to commitment of the gonadal somatic cells to the Sertoli cell fate.

The characteristic wave-like pattern of *Sry* transcript expression is also mirrored in the expression of SRY protein (Correa et al., 2012). This dynamic pattern of expression over space and time means that the cells in the centre of the gonad are the first to be exposed to SRY, and therefore the first to begin upregulating *Sox9*. The cells at the poles, particularly the caudal pole, express *Sry* last, and therefore are the last to upregulate *Sox9*. As a result, upregulation of *Sox9* expression after 11.5 dpc follows a wave-like pattern reflective of *Sry*, but instead of being downregulated is maintained at high levels after SRY becomes undetectable.

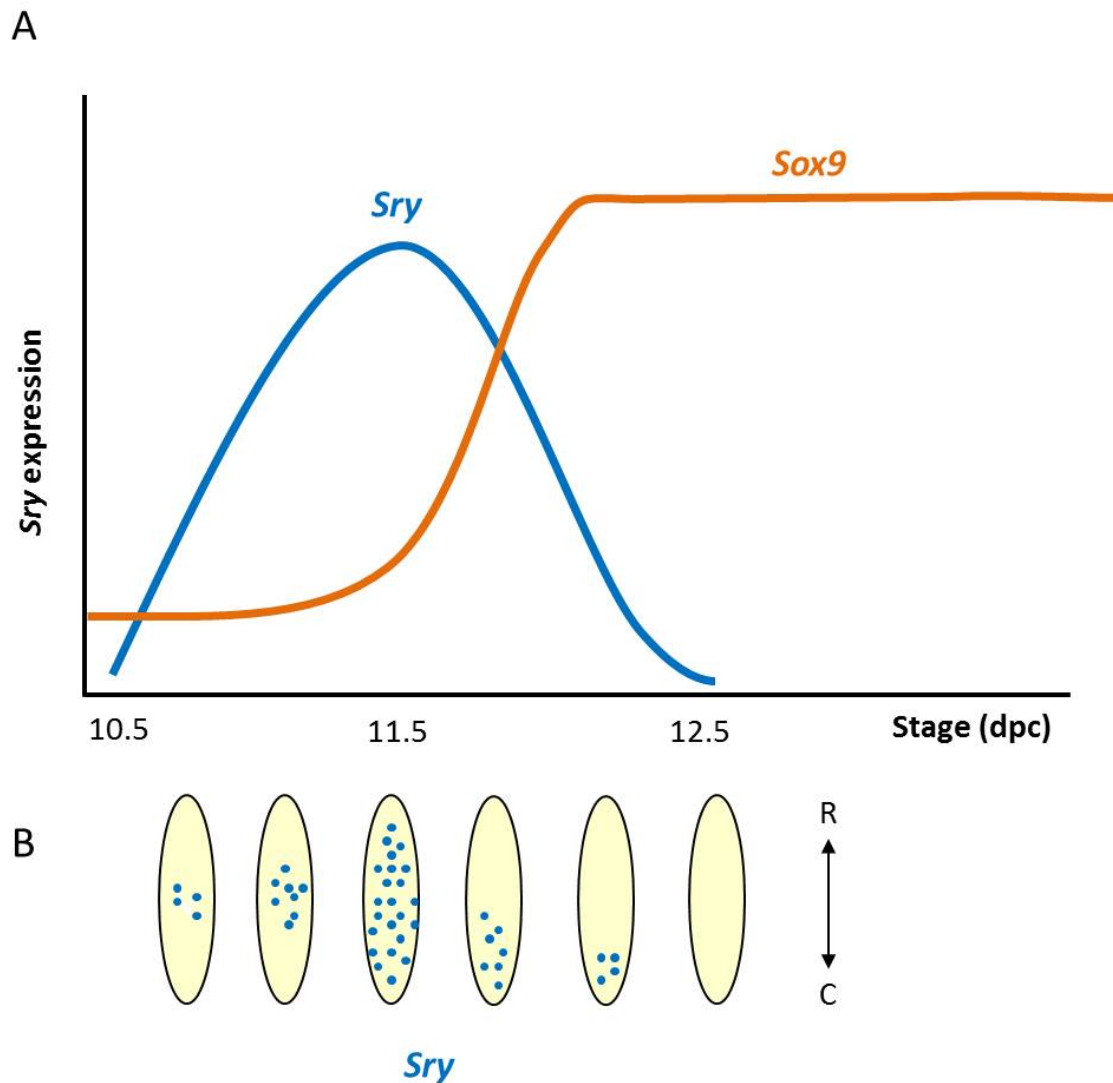


Figure 1.4 Dynamics of *Sry* and *Sox9* expression

- (A) *Sry* (blue) is detectable from 10.75 dpc in the mouse gonad, after which it increases and peaks at 11.5 dpc. Expression then declines and has diminished by 12.5 dpc. *Sox9* (orange) is strongly upregulated after the peak of *Sry* expression and is maintained after *Sry* expression has declined.
- (B) Schematic of *Sry* expression in the gonad (ovals) following the timings of the graph axis above. *Sry* expression (blue dots) is initiated in somatic cells in the central region of the gonad. It spreads along the length of the gonad and by 11.5 dpc is expressed throughout its entire length. As it declines, *Sry* becomes restricted to the caudal pole of the gonad. R↔C rostro-caudal axis

1.5.3. *Sry* regulation

1.5.4. Genes required for *Sry* expression

Due to its characteristic, transient expression pattern, much focus has been placed upon how *Sry* is regulated. Appropriate *Sry* expression is crucial for the ability of its protein product to effectively upregulate *Sox9*, and insufficient *Sry* expression at 11.5 dpc is associated with disrupted testis determination that can range from the presence of ovarian tissue in embryonic gonads, to complete XY sex reversal. Therefore, the regulation of *Sry* expression is of paramount importance to its ability to act as a molecular switch.

Many genes have been identified that, when disrupted, are associated with reduced *Sry* expression, and are summarised in Fig 1.5. The list of such genes is continually growing, and at the time of writing over 10 genes have been implicated in *Sry* regulation (Carre and Greenfield, 2014). This paints a complex picture of the control of *Sry* expression, and is perhaps reflective of the key importance of testis determination to enabling the development of two sexes. Multiple *Sry*-regulatory pathways may have evolved to ensure that there is a robust genetic network underpinning the expression of this key testis-determining gene. Some of these genes are discussed below.

The +KTS splice variant of WT1 has been shown to be required for appropriate *Sry* expression. Studies of mice lacking this variant demonstrated that loss of WT1 (+KTS) is associated with a 25% reduction in levels of *Sry* compared with wild-type (WT) (Hammes et al., 2001). A later study found that this is the result of a combination of reduced *Sry* expression per cell, as well as a reduced number of *Sry*-expressing cells (Bradford et al., 2009). This suggested that it has a role both in the regulation of expression of *Sry* itself, as well as an earlier role in the proliferation of somatic cells of

the gonad, described above. WT1 has also been shown to transactivate *Sry* promoter constructs from mouse, human and pig *in vitro* and also binds the promoter of human *SRY* (Miyamoto et al., 2008, Hossain and Saunders, 2001).

One very recently identified regulator of *Sry* expression is the histone demethylase JMJD1A (Jumonji domain-containing 1A). This enzyme demethylates lysine 9 of histone 3 (H3K9), and is the first gene for which a direct epigenetic link with *Sry* expression has been described. Its role in the epigenetic control of *Sry* expression will be described in a later section of this introduction, but, in brief, mice lacking *Jmjd1a* undergo XY sex reversal characterised by severely reduced *Sry* expression (Kuroki et al., 2013).

Another gene implicated in epigenetic regulation of *Sry* is *Cbx2*. CBX2 (chromobox homolog 2, also known as M33), is a member of the polycomb group of proteins, which were originally described as repressors of *Hox* genes in *Drosophila* (Beuchle et al., 2001). Within this group, the CBX subset of proteins is involved in epigenetic gene regulation through the recognition of methylated histones. Mouse embryos lacking *Cbx2* have hypoplastic gonads and exhibit XY sex reversal, associated with reduced expression of *Sry* (Katoh-Fukui et al., 1998, Katoh-Fukui et al., 2012). Sex reversal in XY knockout gonads was rescued by transgenic overexpression of *Sry* or *Sox9*, although the gonads were still smaller in size than WT. This suggested that CBX2 has a role in both the regulation of *Sry* expression and, independently, in cell proliferation within the gonad (Katoh-Fukui et al., 2012). Whilst the mechanism underlying the reduced *Sry* expression in *Cbx2* mutant gonads has not yet been elucidated, these studies provide an intriguing potential link between the epigenome and *Sry*.

As described above, GATA4 and appropriate interaction with its partner, FOG2, are required for sufficient *Sry* expression (Tevosian et al., 2002). GATA4 transactivates

promoter constructs of mouse and pig *Sry*, but not human *SRY* (Miyamoto et al., 2008). GATA consensus motifs are located in the sequence proximal to the *Sry* transcription start site (TSS) and, additionally, chromatin immunoprecipitation from 11.5 dpc gonads has demonstrated GATA4 binding to a 100 bp region containing one of these *in vivo* (Gierl et al., 2012). This binding is thought to be dependent upon phosphorylation of GATA4, which may be mediated via MAP3K4/p38 MAPK-mediated signalling (Gierl et al., 2012). Two related genes recently implicated in *Sry* regulation are *Six1* and *Six4*, which encode homeobox transcription factors. Loss of both *Six1* and *Six4* leads to XY sex reversal associated with reduced *Sry* expression. SIX1/SIX4 are thought to regulate *Fog2* expression and thus, indirectly, the expression of *Sry* (Fujimoto et al., 2013).

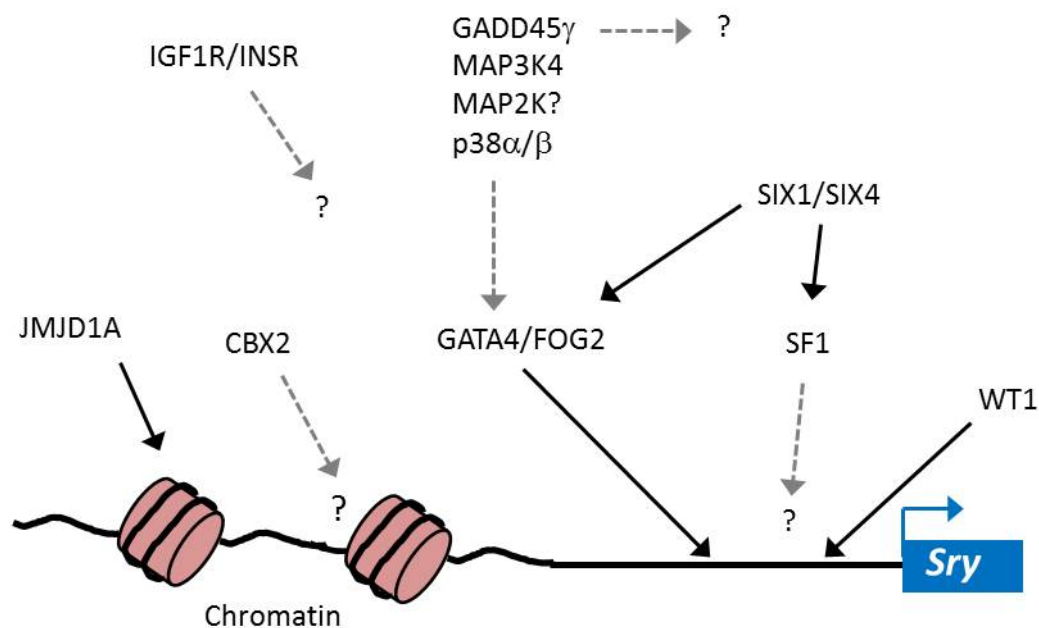


Figure 1.5 Pathways regulating *Sry*

Schematic of pathways and molecules that have been implicated in the regulation of *Sry*. Grey dashed arrows indicate where a direct link has not yet been made.

1.5.5. Sry timing is crucial for testis determination

Comprehensive analyses of *Sry* expression in a variety of mouse mutants has helped develop the concept that not only must *Sry* reach a threshold level, it must also do so within a specific window of time (Fig 1.6). The first observation of the importance of *Sry* timing was made during analyses of B6 mice carrying *Sry* alleles originating from *Mus domesticus* strains. In these mice, an incompatibility between the B6 background and the *domesticus* *Sry* allele results in XY sex reversal, a phenomenon that is known as B6.Y^{DOM} sex reversal, discussed in more detail in Chapter 5. In studies of these mice, it was observed that underlying the observed gonadal sex reversal was delayed expression of *Sry*, and this led to the hypothesis that appropriate timing, as well as levels, of *Sry* is required for testis determination (Nagamine et al., 1999, Bullejos and Koopman, 2005). Later, experiments employing *Sry* under the control of the heat shock protein 70.3 (*Hsp70.3*) promoter facilitated manipulation of the onset of *Sry* expression in transgenic mice. When *Sry* was delayed by 6 hours, the downstream testis-determining pathway was not activated, further confirming the critical requirement for appropriately timed *Sry* expression (Hiramatsu et al., 2009).

It is a reasonable assumption that, if a gene is expressed within a short time-window, any disruption to its timing may affect its ability to regulate downstream targets. For the bipotential gonad, which is under the influence of two mutually antagonistic pathways, this is particularly crucial. For this reason, it is of key importance that the differentiation of Sertoli cells is initiated before pro-ovarian gene expression pathways can attain a dominance that would instead direct the cells towards a granulosa cell fate. Therefore, the requirement for appropriate *Sry* timing may be the result of a need to counteract the ovarian pathway. In support of this, ovotestes are often observed in mice in which *Sry* has been delayed. Ovotestes have a characteristic

appearance at 14.5 dpc, with expression of male markers such as *Sox9* in the central region of the gonad and ovarian markers at the poles. This is reflective of a delay in the receipt of *Sry* signal at the poles of the gonad, which express *Sry* last. In these cells, the ovarian pathway may have already been activated by the time *Sry* reaches them, so they adopt an ovarian fate and the region undergoes characteristic processes of ovarian development (Wilhelm et al., 2009).

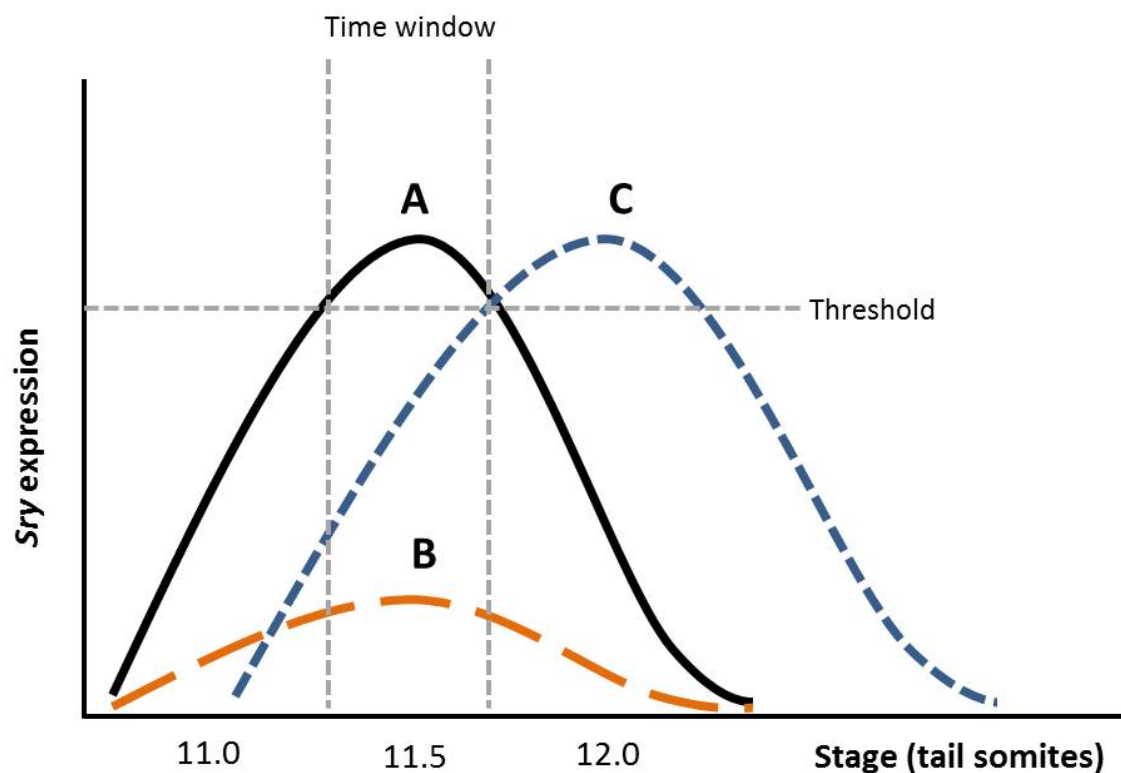


Figure 1.6 *Sry* expression levels and timing are key in testis determination

Normal *Sry* expression (A) must reach a threshold level within the appropriate time (grey dotted lines), to achieve successful testis determination. If *Sry* fails to reach sufficient levels (B, orange), or expression is delayed (C, blue) Sertoli cell differentiation may not be established.

For many of the genes described in the previous section, it is not yet clear whether their disruption is associated simply with reduced levels of *Sry* expression, or with a change in its timing. In order to determine this, the expression of *Sry* must be examined over time. Using this method, some genes have been identified whose disruption alters the expression profile of *Sry*. One of the first of these to be reported was *Map3k4*. This gene, encoding the serine-threonine kinase MAP3K4, was originally found to be required for testis determination following a forward-genetic screen at MRC Harwell. In this study, XY embryos lacking functional MAP3K4 developed ovaries on the C57BL/6J background and showed reduced expression of *Sry* at 11.5 dpc (Bogani et al., 2009). A later study employed comprehensive temporal profiling of *Sry* expression to demonstrate that loss of *Map3k4* was associated with a 14 hour delay in the expression of *Sry* compared with wild-type (WT), resulting in complete XY gonadal sex reversal (Warr et al., 2012).

Another gene, *Gadd45g*, was also reported to be required for appropriate *Sry* timing (Warr et al., 2012). *Gadd45g* encodes growth arrest and DNA damage-inducible protein family member gamma (GADD45γ), a protein known to interact with, and activate, MAP3K4 (Takekawa and Saito, 1998, Lu et al., 2001). Loss of *Gadd45g* was also associated with XY gonadal sex reversal, characterised by similarly delayed *Sry* expression. Further genetic complementation experiments demonstrated that GADD45γ and MAP3K4 are part of a common pathway and form a MAPK signalling module that regulates timing of *Sry* expression (Gierl et al., 2012, Warr et al., 2012).

Importantly, loss of *Map3k4* is associated with a reduction in the level of *Sry* per cell, rather than merely a reduction in the number of *Sry*-expressing cells (Bogani et al., 2009). As a result, GADD45γ-MAP3K4 signalling may be a good candidate for direct

regulation of *Sry* expression in a cell-autonomous manner. This hypothesis is also supported by the spatiotemporal expression profile of *Gadd45g* in gonadal somatic cells, which mirrors, but precedes, that of *Sry*.

1.5.6. Epigenetic regulation of *Sry*

Recent studies are beginning to elucidate some aspects of epigenetic regulation of *Sry*. These have formed two main foci: studies of DNA methylation and studies of chromatin modification.

DNA methylation-mediated control of gene expression is a paradigm of growing complexity and has been shown to be critical for mammalian development (Li and Zhang, 2014, Barreto et al., 2007). DNA methylation typically occurs at cytosines specifically situated immediately 5' to guanine, forming a CpG dinucleotide. In promoters of genes, specific regions of CpG enrichment exist that are differentially methylated in different tissues. These are described as tissue-specific differentially methylated regions (T-DMRs) and have been implicated in regulating cell-type specific gene expression (Ikegami et al., 2009). Typically, the T-DMR of a given gene is hypermethylated in tissues in which the gene is not being actively expressed, and hypomethylated in tissues in which it is being expressed.

Over a decade ago, two putative T-DMRs were identified upstream of *Sry*, which were hypomethylated in gonadal somatic cells in 11.5 dpc gonads and hypermethylated in extra-gonadal tissues (Nishino et al., 2004). The first region (region I), containing 4 CpG dinucleotides, lies approximately 3 kb upstream of the *Sry* open reading frame (ORF) (Fig 1.7) and is thought to reside in the promoter for the circular transcript of *Sry*. This transcript is untranslated but may act as a microRNA (miRNA) sponge (Capel et al., 1993, Hansen et al., 2013). The second T-DMR (region II) is more proximal to the ORF

of the protein-coding transcript and contains 7 CpG dinucleotides (Fig 1.7) (Nishino et al., 2004). Both T-DMRs undergo stage-specific hypomethylation at 11.5 dpc that coincides with the expression of *Sry*. By 15.5 dpc, region II is hypermethylated again whereas region I remains hypomethylated. In a later study, Nishino *et al* also described non-CpG methylation in region II at the internal cytosine of a pentanucleotide sequence 5'-CCTGG-3' (Nishino et al., 2011). The methylation of this non-CpG coincided with the hypomethylation of the CpGs within the region in the 11.5 dpc gonad, however, this did not occur in all the mouse strains examined. Methylation was observed in the ICR-A, BALB-c, DBA2 and C3H strains, but not in B6 or 129. Since non-CpG methylation at this site was also shown to enhance the promoter activity of an *Sry* construct *in vitro*, these strain-specific differences pose interesting questions about the differential regulation of *Sry* between different mouse strains, and therefore about the varying sensitivities of different strains to XY sex reversal. This topic will be revisited in chapters 5 and 6.



Figure 1.7 Tissue specific, differentially methylated regions (T-DMRs) of *Sry*

Two T-DMRs have so far been described upstream of *Sry*. Region I contains 4 CpG dinucleotides (black filled circles) and is situated in the 5' inverted repeat region (orange arrow). Region II contains 7 CpG dinucleotides and is situated close to the *Sry* transcription start site. Numbering of nucleotide positions according to Nishino *et al* (2004).

Whilst hypomethylation of the *Sry* T-DMR is coincident with its expression, so far no regulatory molecules have been identified that directly influence *Sry* methylation during sex determination. One promising candidate for a potential regulator of *Sry* methylation was GADD45 γ , a protein known to be involved in active DNA demethylation mediated by DNA repair (Barreto *et al.*, 2007, Niehrs and Schafer, 2012). Whilst loss of function of GADD45 γ was associated with delayed *Sry* expression, no notable differences in methylation were observed between WT and *Gadd45 γ* or *Map3k4* knockout gonadal somatic cells at 11.5 dpc (Gierl *et al.*, 2012, Warr *et al.*, 2012). This does not exclude the possibility that GADD45 γ may be involved in demethylation elsewhere at the *Sry* locus, or the wider genome, during testis determination, but nonetheless leaves us with a space to fill in the search for regulators of this particular region. Additionally, as many recent reports suggest that MAPKs can also associate with chromatin in a variety of manners, it is plausible that

GADD45 γ and MAP3K4-mediated signalling may regulate *Sry* expression via chromatin modification. Some of these reports are described in section 1.6.

Our knowledge of *Sry* regulation via chromatin modifications is still in its infancy, but significant progress has been made in recent years as techniques for epigenomic analysis of small cell numbers have been developed. The first link between chromatin and *Sry* expression was the report that CBX2, described above, is required for *Sry* expression. However, it is still unclear whether its role is a direct one, or indeed whether CBX2 activates or represses transcription of its targets during sex determination (Katoh-Fukui et al., 2012). More recently, the H3K9 demethylase JMJD1A was found to be required for *Sry* to be expressed sufficiently in the 11.5 dpc gonad (Kuroki et al., 2013). JMJD1A was found to directly regulate the demethylation of H3K9 (H3K9me₂), a histone mark indicative of transcriptional repression (Kooistra and Helin, 2012). Gonadal somatic cells of *Jmjd1a*-deficient 11.5 dpc gonads exhibited significantly higher levels of H3K9me₂ and also showed decreased H3K4me₂, a hallmark of transcriptional activation, compared with WT. The very low levels of *Sry* expression in *Jmjd1a* mutants, as well as the apparent lack of a characteristic wave-like expression pattern, suggested that loss of *Jmjd1a* may be associated with heterochromatinisation or global shutdown of the *Sry* locus. The authors suggested that JMJD1A may mediate deheterochromatinisation of *Sry* and thus facilitate association with a H3K4 methyltransferase and transcription factors (Kuroki et al., 2013).

1.5.7. The search for *Sry* regulatory regions

In addition to the T-DMR, several putative regulatory regions of *Sry* have been reported, however, the role of any specific element in regulating the timely expression

Sry is yet to be clearly demonstrated. Several approaches have been taken to searching for regulatory regions of *Sry*, including:

- Prediction and testing of transcription factor binding sites
- Testing the ability of *Sry* constructs to induce *Sry* expression or XX sex reversal
- Identification of evolutionarily conserved sequences

So far, consensus binding-sites for WT1 and GATA4 have been identified, situated within 500 bp upstream of the transcription start site (TSS) (Miyamoto et al., 2008, Gierl et al., 2012). Importantly, recruitment of GATA4 to chromatin has been demonstrated *in vivo* using a purified population of 11.5 dpc gonadal somatic cells and is reported to require GATA4 to be phosphorylated (Gierl et al., 2012). Binding sites for other regulators of *Sry*, such as SF1, have not yet been identified (Larney et al., 2014).

The original 14.6 kb genomic fragment containing *Sry* that was shown to induce XX sex reversal has been subject to many manipulations in an effort to narrow down the crucial regulatory element. This fragment contained 8 kb of sequence 5' to the *Sry* TSS and 3 kb of 3' sequence. It has since been shown that the 8 kb of 5' sequence is sufficient to drive expression of an enhanced green fluorescent protein (*EGFP*) reporter in a spatiotemporal manner reminiscent of *Sry*'s endogenous expression (Albrecht and Eicher, 2001, Jameson et al., 2012b). Further narrowing down of this region has yielded varying results, and is comprehensively reviewed by Larney *et al* (2014).

Finally, the search for regulatory elements of *Sry* has also focused upon finding regions that are conserved between species. Comprehensive sequence comparison is difficult due to the highly repetitive nature of the sequence surrounding *Sry* (Gubbay et al.,

1992) and no studies have yet performed comparison across multiple species. However, it is clear that, in general, *Sry* and its flanking sequence are poorly conserved, with few regions of conservation between mouse and human. This may suggest that important regulatory elements reside within these discrete conserved regions, however, their functions are yet to be determined. Conversely, a search for conserved regions between the unusual *Sry* of the mouse and other mammalian species may be futile, because murine *Sry* is expressed in such a unique, brief manner. Instead, comparisons between different mouse strains in which *Sry* is differentially expressed may yield more information about important regulatory elements, and this subject will be revisited in detail in Chapter 6.

1.6. MAPKs and the control of gene expression

Since loss of *Map3k4* is associated with delayed *Sry* expression, recent efforts to elucidate the mechanism of MAP3K4 signalling in testis determination now focus on investigating how MAPK signalling may regulate transcription. MAP3K4 is a member of the highly conserved mitogen activated protein kinase family, described below.

1.6.1. MAPK signalling modules

Signalling via MAPKs is a widely employed cellular mechanism for generating specific intracellular responses to extracellular signals. The system principally comprises a highly evolutionarily conserved three-kinase module which transduces signals from the plasma membrane to downstream target molecules by sequential phosphorylation. The first component of the three-kinase module is a MAP kinase kinase kinase (MAP3K or MEKK). These are serine-threonine kinases that phosphorylate specific, conserved serine and threonine residues of their targets, MAP kinase kinases (MAP2Ks or MEKs). In turn MAP2Ks phosphorylate conserved tyrosine and threonine residues of MAPKs. Activated MAPKs can phosphorylate a diverse range of targets that regulate many

cellular processes, including migration, differentiation and proliferation (Craig et al., 2008).

There are four main families of MAPKs in mammals. The extracellular signal-regulated kinase (ERK) group consists of ERK1 and ERK2 (p42/p44 MAPKs) and is the most widely responsive to a large range of stimuli. ERK5 is considered to be a stress-responsive and mitogen-responsive MAPK, and the c-Jun N-terminal kinase family (JNK1-3), as well as the p38 (α , β , γ , δ) family are categorised as stress-responsive (Klein et al., 2013). However, categorisation of MAPK groups has historically been performed using cell lines, and it has become evident that all groups are involved in signal transduction in a vast range of cellular contexts *in vivo* (Aouadi et al., 2006). MAPK pathways are summarised in Figure 1.8. Of particular relevance to this thesis is the module including MAP3K4. MAP3K4 is considered part of the stress-activated MAPK cascade and phosphorylates MAP2K3, MAP2K4 and MAP2K6. MAP2K4 phosphorylates JNK and p38 MAPKs, whereas MAP2K3 and MAP2K6 phosphorylate p38 MAPKs (Craig et al., 2008, McCann-Crosby and Sutton, 2015). Identification of which MAPK acts as the effector MAPK in mouse testis determination began with biochemical experiments that demonstrated a significant reduction in levels of phosphorylated p38 MAPK (p-p38) in 11.5 dpc gonads lacking *Map3k4* (Gierl et al., 2012, Warr et al., 2012). In addition, embryos in which both p38 α and β were conditionally deleted in epiblast-derived tissues exhibited XY gonadal sex reversal, confirming that these p38 MAPKs are required for mouse testis determination (Warr et al., 2012).

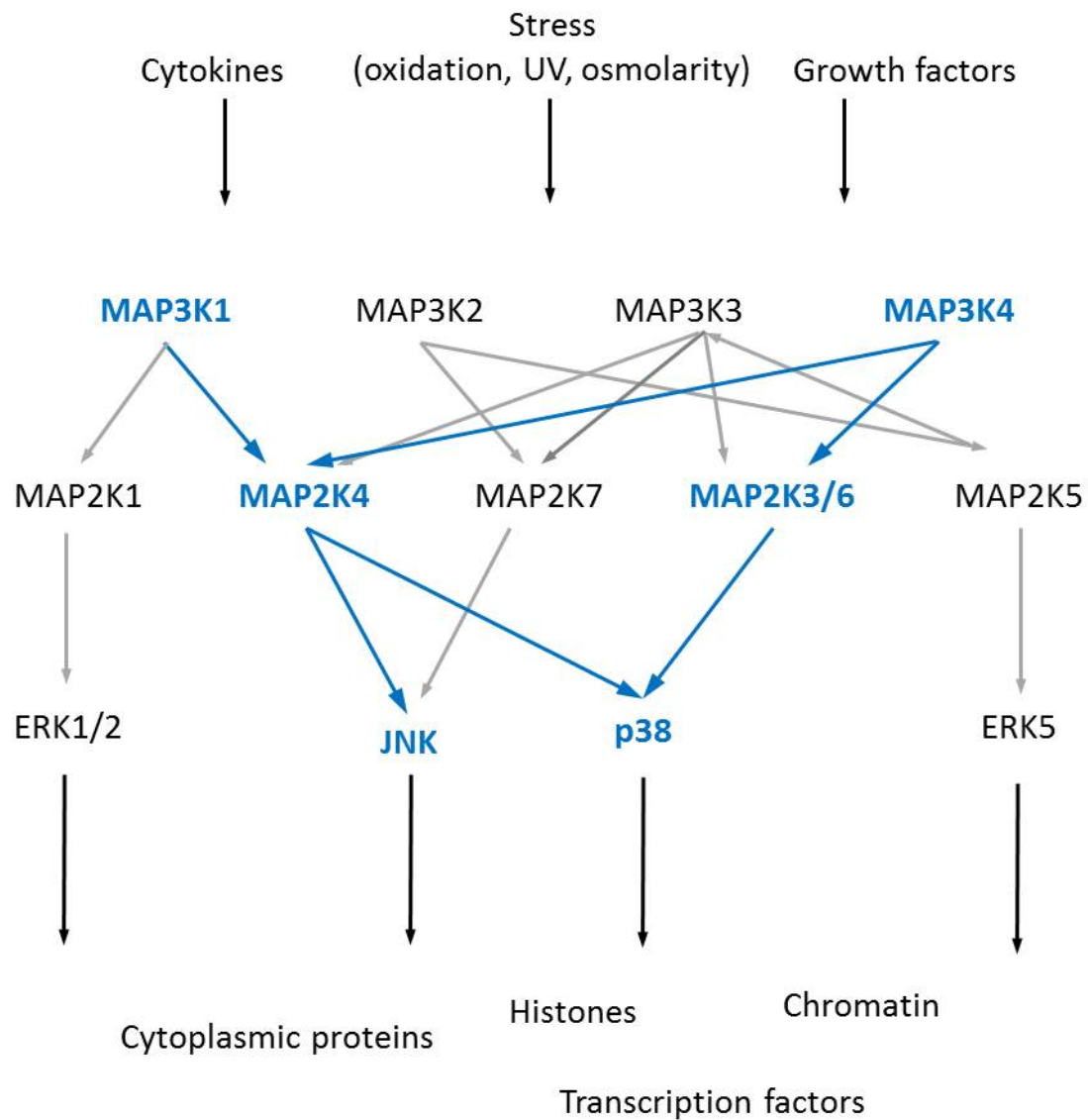


Figure 1.8 MAPK signalling pathways

The MAPK cascade functions to transduce extracellular signals to specific nuclear or cytoplasmic targets. MAP3Ks activate MAP2Ks, which in turn activate MAPKs. The pathways of relevance to this thesis are the MAP3K1-mediated and MAP3K4-mediated signalling pathways, highlighted in blue. Most known MAPK interactions are taken from accumulated *in vitro* data, but some (such as MAP3K4) are supported by mouse knockout studies.

1.6.2. MAPK interactions with chromatin

In the last decade, evidence has mounted that MAPKs are heavily involved in the modification of chromatin. This has added new diversity to the wide range of targets of MAPK signalling, and also provides an intriguing direction of focus for the study of the role of MAP3K4-mediated signalling in *Sry* regulation.

Substrates of p38 and ERK MAPKs, the mitogen-activated and stress-activated protein kinases 1 and 2 (MSK1/2) are able to directly phosphorylate histone 3 (H3) at serine 10 and serine 28, directing a cascade of events that leads to enhanced transcription of the gene *FOSL* (Klein et al., 2013).

Several modes of MAPK-mediated chromatin modification have so far been described. Perhaps the least direct involves MAPK phosphorylation of transcription factors, which then triggers their recruitment of transcriptional machinery to promoters. For example, phosphorylation of the transcription factor MEF2D by p38 MAPK is required for recruitment of a methyltransferase complex to lysine 4 of histone H3 (H3K4) in a mouse myoblast cell line (Rampalli et al., 2007). However, as well as mediating chromatin modification via transcription factors, MAPKs may also directly bind histone modifiers, chromatin remodelling complexes and RNA polymerase II (Pol II). In yeast, the p38 MAPK homolog Hog1 directly recruits a histone deacetylase complex to osmotic stress-inducible promoters, facilitating histone H4 deacetylation and entry of RNA polymerase II (De Nadal et al., 2004). In addition, the histone acetyl transferase p300 is directly phosphorylated by ERK (Chen et al., 2007), and phosphorylation of the SWI-SNF chromatin remodelling complex by p38 is required for its localisation to myogenic promoters (Simone et al., 2004).

A more recent report implicates JNK MAPKs in chromatin modification during neuronal differentiation. Here, JNK was found to be associated with active promoters during neuronal stem cell differentiation, and it was observed that JNK occupancy is associated with increased phosphorylation of histone H3 serine 10 (H3S10) (Tiwari et al., 2012).

Recent profiling of the protein-DNA interactome by Hu et al (2009) demonstrated that ERK2 binds directly to DNA *in vitro* at a defined C/GAAAG/C consensus sequence. In their experiments, Hu *et al* showed that ERK2 represses transcription of interferon- γ by competing for DNA binding with a transcription factor, C/EBP β , in a manner independent of its kinase function (Hu et al., 2009). This finding may transform how we consider the interaction between MAPKs and transcription factors, and prompts further investigation into whether other MAPK family members are able to directly bind DNA to regulate gene expression.

Together, these and many other reports place MAPKs as important regulators of gene expression via chromatin modification and direct DNA interaction. Given the importance of MAPK signalling in the timely expression of *Sry*, this raises intriguing questions about whether MAPK signalling directly regulates *Sry* transcription via one, or more, of these means. Some of these questions shall be addressed within this thesis.

1.7. Disorders of sexual development (DSD)

Discrepancy between an individual's chromosomal and phenotypic sex is described as a disorder of sexual development (DSD) (Hughes et al., 2006). These arise when disruption occurs either in sex determination, described above, or sexual differentiation of the internal and external genitalia, controlled by sex hormones. They

account for approximately 7.5% of all congenital defects in humans and can have significant, lifelong consequences for affected individuals. Not only can DSD affect fertility and the risk of gonadal tumours, it also has a significant impact on psychosocial development, particularly in those individuals with ambiguous phenotypic sex in childhood (McCann-Crosby and Sutton, 2015).

DSDs can arise due to aneuploidy of the sex chromosomes, such as XXY (Klinefelter's syndrome), XYY ("supermales") or XO (Turner Syndrome), or due to disruption of key sex-determining genes. They can also develop after sex determination of the gonads, affecting later sexual differentiation of the internal and external genitalia and development of secondary sexual characteristics. For example, androgen insensitivity syndrome (AIS) occurs as a result of mutations in the androgen receptor and leads to development of female characteristics in XY individuals with testes. The focus of this section, in line with the rest of this thesis, is upon disorders associated with primary sex determination, which fall under the category of 46,XX or 46,XY gonadal dysgenesis (GD).

46,XX GD is mostly caused by translocation of *SRY* onto the X chromosome, resulting in the development of XX males (Abbas et al., 1993). There are also some less common cases of 46,XX GD that have been reported, which are associated with loss-of-function mutations in genes that antagonise the male pathway, including *RSPO1* (Parma et al., 2006) and *WNT4* (Mandel et al., 2008).

46,XY GD describes disorders associated with abnormal testis determination. This can manifest in a range of ways, from incomplete masculinisation or ambiguous genitalia to complete feminisation. At the most severe end of the spectrum, complete GD (CGD) is associated with absence of testes and bilateral streak gonads, with female

internal and external genitalia. Causative mutations are often loss-of-function of critical testis-determining genes, including *SRY* (Cameron and Sinclair, 1997), *SOX9* (Foster, 1996) and *SF1* (Lin and Achermann, 2008), and can also be associated with the gain of function, via duplication, of female-determining genes such as *WNT4* (Jordan et al., 2001).

One intriguing 46,XY GD syndrome is dosage-sensitive sex reversal (DSS), a disorder in which XY individuals carrying duplications of part of the small arm of the X chromosome develop as female. Studies in mice have demonstrated that the causative gene is likely to be the orphan nuclear receptor *Dax1* (also known as *Nr0b1*). Overexpression of *Dax1* in XY mice leads to sex reversal, characterised by reduced expression of *Sox9*, and it was proposed that *Dax1* is an anti-testis gene (Swain et al., 1998, Ludbrook et al., 2012). However, mice lacking *Dax1* develop hypoplastic testes with disorganised architecture, and also express lower levels of *Sox9* (Meeks et al., 2003a, Meeks et al., 2003b). This apparent discordance between phenotypes in loss- and gain-of function studies suggests that there is a critical window of DAX1 activity: too little, or too much DAX1, is detrimental to testis determination (Ludbrook and Harley, 2004). However, it is still unclear what the precise role of *Dax1* is in male or female sex determination.

Analysis of these and other disease-causing variants has facilitated identification of many pathways involved in testis determination, as well as the critically dosage-sensitive nature of sex determination. However the causative mutation is still unknown in around 80% of 46,XY CGD patients (Baxter et al., 2015).

One gene of emerging importance in 46,XY GD is *MAP3K1*. Originally described in two cases of familial 46,XY CGD, this member of the MAPK superfamily is now widely

considered as an important causative gene (Baxter et al., 2015, Loke and Ostrer, 2012, Pearlman et al., 2010). The role of MAP3K1 in testis determination is still somewhat unclear, and will be described in greater detail in Chapter 3. However, that its proper function is required for human testis determination provides an intriguing link between MAP kinases, human and mouse testis determination, which form the main foci of this thesis.

1.8. The mouse as a model of DSD

Sex determination in mammals is highly conserved, with many of the genetic pathways being intact in both humans and mice (Morais da Silva et al., 1996, Swain and Lovell-Badge, 1999). Of particular relevance to this thesis, both human and mouse testis determination require the regulation of *SOX9* by *SRY* (Morais da Silva et al., 1996, Sekido et al., 2004, Sekido and Lovell-Badge, 2008). However, unlike mouse *Sry*, which is expressed transiently, human *SRY* expression continues into adulthood (Clepet et al., 1993). Much of our understanding of the function of key sex-determining genes has been the result of genetic manipulation of the mouse, some of which has been described above. Manipulation of the mouse genome for this purpose has ranged from global knockouts to conditional knockouts (Hammes et al., 2001, Hu et al., 2013, Warr et al., 2012, Gierl et al., 2012, Warr et al., 2011), transgenic overexpression (Koopman et al., 1991, Vidal et al., 2001, Zhao et al., 2015) and also forward genetic screens (Bogani et al., 2009, Siggers et al., 2014).

One particularly powerful tool for modelling DSD in the mouse is our ability to exploit strain-specific differences in susceptibility to sex reversal. Many human DSDs are not only dosage-sensitive but also dependent upon genetic background. For example, fathers who are heterozygous for *SF1* mutations, or hemizygous for *SRY* mutations, can

produce heterozygous or hemizygous XY daughters. Therefore, a given mutation may result in haploinsufficiency for normal testis determination only in certain genetic backgrounds. Likewise, in the mouse, mutations or altered dosage of sex-determining genes can lead to varying severity of sex reversal depending on genetic background (Bouma et al., 2007, Ludbrook et al., 2012). The B6 background is particularly sensitised to sex reversal and as a result is a widely used strain for investigating sex determination (Correa et al., 2012, Munger et al., 2009). In addition, B6 can be further sensitised by the introgression of a *M. domesticus* Y chromosome, as described above. These strains are particularly sensitive to perturbations in testis-determining genes, including *Map3k4*, and have been utilised in many studies to uncover gene networks involved in mammalian testis determination (Bogani et al., 2009, Warr et al., 2012, Warr et al., 2014, Katoh-Fukui et al., 2012, Eicher et al., 1996, Munger et al., 2009, Bouma et al., 2007). With the advent of new technologies, these classical models of sex reversal may gain a new lease of life, as we can begin to learn more about the mechanisms underlying their increased propensity for sex reversal. In particular, investigating the epigenetic basis of these phenomena may facilitate new understanding of the regulation of key testis-determining genes, including *Sry*.

1.9. Aims of this thesis

A key role for MAPK-mediated signalling in testis determination has been identified, both in humans and mice. However, whilst disrupting the function of MAP3K4 in the mouse affects the expression of *Sry*, the mechanism by which MAP3K4-mediated signalling mediates timely *Sry* expression is still unclear. Moreover, the role of MAP3K1 in mouse testis determination, if any, is also unclear. In this thesis, loss- and gain-of-function approaches will be taken to further investigate the role of the MAPK and MAPK-associated pathways in testis determination.

Chapter 2. Materials and Methods

This chapter describes the materials and methods used in this thesis. Where variations to these methods are employed for specific experiments, they are described in the relevant chapter.

2.1. General materials and methods

Standard solutions were prepared according to Sambrook et al (1989) and their composition is given in Table 2.1. Thermocycling reactions were performed using a G-Storm GS4 thermocycler.

Table 2.1 Standard solutions

Reagent	Composition
ES cell lysis buffer	10mM Tris pH7.5, 10mM EDTA, 10mM NaCl, 0.5% sarcosyl, 1mg/ml proteinase K
Laemlli buffer	100 µl 4X LDS (supplier), 2 µl 0.5M EDTA, 10 µl 10X proteinase inhibitor cocktail
Luria Bertani (LB) broth	10 g/l Bactotryptone, 5 g/l Bactoyeast extract, 5 g/l NaCl, pH 7.0 at 25°C
NTMT	(100mM NaCl, 100mM Tris-HCl (pH 9.5) 50mM MgCl ₂ , 0.1% Tween 20 in H ₂ O)
PBT	PBS + 0.1% Tween 20
TBS	50 mM Tris-Cl pH7.5 , 150 mM NaCl
TBST (Western blot)	50 mM Tris-Cl pH8.0, 150 mM NaCl, 0.1 % Tween-20
TBST (WMISH)	50 mM Tris-Cl pH7.5, 150 mM NaCl, 0.1 % Tween-20
TBSTx	TBS + 0.1% Triton X100
TBE (10x)	1 M Tris base, 1 M boric acid, 0.02 M EDTA disodium salt in H ₂ O. diluted to 1x for electrophoresis
SSC (20x)	3M NaCl, 300 mM sodium citrate, in H ₂ O
TE-tween	50 mM Tris-Cl, pH 8.0 / 1 mM EDTA / 0.5% Tween-20
WMISH prehybridisation solution	50% formamide, 5X SSC, 0.1% Tween 20, 50µg/ml Heparin in H ₂ O
WMISH hybridisation solution	Prehybridisation solution + 100 µg/ml yeast RNA, 100 µg/ml salmon sperm DNA
WMISH wash solution (20 ml)	10 ml formamide, 4 ml 20X SSC pH 4.5, 2 ml 10% SDS, 4 ml H ₂ O
Proteinase inhibitor cocktail (10X)	1 tablet PhosStop (Roche), 1 tablet Complete Mini (Roche) in 1 ml H ₂ O

2.2. Mouse Lines and genotyping

Mouse work was conducted under the authority of the Home Office Project License number 30/2877 and all mice were maintained in accordance with the Animals (Scientific Procedures) act 1986 (ASPA) amended 2012. Unless otherwise stated, all lines were housed in a controlled light-dark cycle, with darkness between 7pm and 7am.

All experimental mice were maintained on a C57BL/6J (B6) background, chosen due to its increased sensitivity to disruptions of testis determining genes (Bouma et al., 2007).

2.2.1. Chromosomal sexing

All experimental embryos were genotyped for chromosomal sex using a PCR assay that has been previously described (Warr *et al* 2009). Primers and reaction conditions are given in Table 2.2.

2.2.2. *Gadd45g*

Gadd45g knockout mice were a kind gift from P. Trainor (Stowers Institute for Medical Research, Kansas, USA) and have been previously described (Hoffmeyer et al., 2001). In these mice, most of the coding region of *Gadd45g* is replaced by a neomycin resistance cassette. *Gadd45g* mice were maintained as heterozygotes and genotyped by the MRC Harwell GEMS team (Deen Quailed, Ramakrishna Kurapati). Experimental embryos were genotyped by PCR using a multiplex assay consisting of a forward primer and 2 alternative reverse primers. One anneals to the selection cassette and the other anneals to endogenous sequence, giving bands of different sizes in wild-type and knockout embryos (Table 2.2).

2.2.3. B6.Y^{DOM} Lines

The B6.Y^{DOM} mouse lines B6.Y^{AKR} and B6.Y^{POS} were used in this thesis. These lines are congenic B6, with the exception of the Y chromosome, which is derived from a *Mus domesticus* strain: In both lines, the *Mus domesticus* Y chromosome was maintained on the B6 background by selecting fertile males and crossing with B6 females. Presence of the Y^{DOM} chromosome was confirmed using a PCR assay that differentiates between the *M. domesticus* (Y^{POS}, Y^{AKR}) and *M. musculus* (Y^{B6}) Y chromosomes. In *M. musculus*, *Zfy-2* has an 18 bp deletion so can be detected by PCR using a previously described assay (Prager et al., 1997), which was optimised by Dr Nick Warr (Table 2.2).

2.2.4. B6.Y^{POS}-*Map3k4* Transgenic line

The *Map3k4* BAC transgenic line has been previously described (Warr et al., 2012, Warr et al., 2014). The BAC, a 178.7 kb NOD/MrKTac BAC clone containing the complete *Map3k4* ORF and flanking DNA regions, was originally sourced from the Centre for Applied Genomics (Toronto, Canada) and injected into C57BL/6J 1-cell embryos to produce transgenic founders (Gardiner and Teboul, 2009). The BAC confers a 4-fold increase in expression of *Map3k4* in 11.5 dpc gonads was demonstrated to be functional by its ability to rescue gonadal sex reversal in *Map3k4*^{-/-} embryos (Warr et al., 2012).

The *Map3k4* BAC was maintained on B6.Y^{POS} by generating B6.Y^{POS}-*Map3k4* BAC studs and crossing to B6 females. Genotyping for presence of the BAC was performed by a previously described PCR assay (Warr et al., 2012) using primers that amplify part of the chloramphenicol resistance cassette (Table 2.2).

2.2.5. *Map3k4* knockout

The *Map3k4*^{-/-} line was a kind gift from H. Chi and R. Flavell (Chi et al. 2004). It was generated by the insertion of *LoxP* sites either side of exon 3. Exon 3 was then deleted following homologous recombination driven by a ubiquitously expressed *Cre* transgene, resulting in a frame-shift and premature termination of the coding sequence before the critical kinase domain, creating a null allele. Since *Map3k4*^{-/-} embryos die prenatally, mice were maintained as heterozygotes. Embryos were genotyped using a three-primer PCR reaction: two primers span the location of the 1st *LoxP* site to detect the wild-type allele, while a third anneals downstream of the location of the 2nd *LoxP* site to detect the KO allele.

2.2.6. *Sf1-EGFP*

The *Sf1-EGFP* reporter line was a kind gift from J. Bowles (University of Queensland, AU) (McClelland et al., 2015). Adult *Sf1-EGFP* mice were mated with B6.Y^{POS} to generate embryos for FACS sorting of gonadal somatic cells from B6.Y^{POS} embryos. Genotyping for presence of *EGFP* was performed by PCR using the primers described in Table 2.2.

2.2.7. 110 Region

Presence of the 110 chromosome 11 region (Arboleda et al., 2014) described in Chapter 6 was assayed by PCR followed by Sanger sequencing to detect 3 SNPS located in the region (See Chapter 6). PCR primers sequences were kindly provided by V. Arboleda and the assay optimised at MRC Harwell. Primer sequences and assay conditions are shown Table 2.2.

Table 2.2 Genotyping PCR primers and assay conditions

Assay	forward primer(s)	Reverse primer(s)	PCR conditions	Product sizes	Assay design
XY sexing	TGGATGGTGTGGCCAATG	CACCTGCACGTTGCCCTT	SD* x36, 57°C	X: 253bp Y: 335bp	N. Warr (Warr <i>et al</i> 2009)
<i>Gadd45g</i>	A: CTTGCAAAACCACACTGCTC B: TGGTCTGATCGACTTGGTGA	C: GCCTAGGCCTTCCTGTGCTA	SD* 60°C	WT: 217bp Mut: 120bp	N. Warr (Warr <i>et al</i> 2012)
<i>EGFP</i>	CCTACGGCGTGCCAGTGCTTCAGC	CGGCGAGCTGCACGCGCCGTCCTC	SD* x35, 62°C		G. Carre (personal communication)
<i>Map3k4</i> BAC transgene	GCGTGTTACGGTGAAAACCT	GGGCACCAATAACTGCCTTA	SD* 57°C	314bp	N. Warr (Warr <i>et al</i> 2012)
<i>Map3k4</i> KO	A:TAAGACAGAGGCAGGGATTGGC	B: CCGTTGTTCTCAGAGTTGCTCG C: GAGCAGTATTGTGGATTTCCGAAG	SD* 65°C	WT: 465 bp Mut: 180 bp	Hongbo Chi (optimised by R. Brixey) (Bogani <i>et al</i> 2009)
<i>Zfy2D</i>	CATTAAGACAGAAAAGACCACCG	GTGAGGAAATTTCTTCTGTGG	SD* 57°C	POS/AKR: 202bp B6: 202bp + 184bp	Prager <i>et al</i> (1997)
POSA SNP1 (110 region)	AAAGTGTGCTTCCCAGGAGA	CCTCTCCCTCAACCCCTAAG	SD* 59°C	252bp	V. Arboleda (personal communication)
POSA SNP2 (110 region)	CCACAGCTGGAGGTAGGGTA	CCTAAGATGCCATGGGAAGA	SD* 57°C	176bp	V. Arboleda (personal communication)
POSA SNP3 (110 region)	AGGTGGTGCTGGTTTTGAAG	TTTGACAACCCCTTAGTC	SD* 62°C	139bp	V. Arboleda (personal communication)

*Standard PCR programme “SD”: 95°C for 5 min, 36 cycles of [95°C 30 sec, X°C 30 sec, 72°C 30 sec], then 72°C for 5 min. Variations on annealing temp and number of cycles are stated.

2.3. Tissue collection and processing

2.3.1. Embryo collection

Timed mates were set up for provision of embryos for experiments. Pairs of mice were mated overnight and noon on the day of detection of the vaginal plug was considered 0.5dpc. To obtain embryos at 11.0dpc and 12.0dpc, mice were housed under reversed light-dark conditions, in which darkness was set between 7am and 7pm. In this case, noon on the day of vaginal plug detection was considered to be 1.0dpc. Pregnant females were culled by cervical dislocation (Schedule One) under Home Office regulations. The abdomen was opened and the uterine horns, containing the embryos, were removed. Embryos were removed from the uterus and immediately placed into ice cold PBS (in accordance with Home Office regulations). Embryos 14.5 dpc or older were culled immediately by decapitation. Embryos were dissected in PBS using a Leica M80 dissection microscope.

2.3.2. WMISH

Embryos were harvested between 11.0 dpc and 14.5 dpc. 11.0 dpc to 12.5 dpc embryos were staged by counting the number of tail somites (ts) posterior to the hindlimb. Urogenital ridges (UGR) (11.5-12.5 dpc) or paired gonads and mesonephroi (13.5 dpc onwards) were collected in microcentrifuge tubes and fixed overnight in 4% paraformaldehyde/PBS. Tissue was then dehydrated through methanol (MeOH) serial dilutions by washing for 5 min in each solution (2x PBS, 1x 25% MeOH/PBS, 1x 50%, 1x75%, 2x 100%) and then stored in 100% MeOH at -20°C.

2.3.3. Quantitative real time PCR (qRT-PCR)

Gonadal tissue was collected between 11.5dpc and 14.5dpc. Embryos younger than 13.5dpc were staged by counting tail somites. UGRs were subdissected to separate the

gonad from the mesonephros, and the gonads of each embryo were placed into microcentrifuge tubes in a minimal volume of PBS and immediately frozen on dry ice. Tissue was stored at -70°C until required for RNA extraction.

2.3.4. Western Blots

Gonads of 11.5 dpc to 12.5 dpc embryos were subdissected as described in Section 2.3.3 in PBS containing 1 X Complete Mini protease inhibitor cocktail (Roche) and 1 X Phos-Stop phosphatase inhibitor (Roche), placed into microcentrifuge tubes in a minimal volume of liquid, and snap-frozen on dry ice. Tissue was stored at -70°C until required for protein extraction.

2.3.5. Immunohistochemistry (IHC)

For wax section IHC, 11.5 to 12.5 dpc tissue was collected as per Section 2.3.2 and fixed in 4% PFA/PBS overnight. Tissue was then transferred to PBS and processed into paraffin wax by the MRC Mary Lyon Centre (MLC) Histology team (Adele Austin and Caroline Barker). Tissue was embedded in wax in a transverse orientation and 8 µM sections were cut and placed onto glass slides by the MLC Histology team.

2.3.6. Fluorescence-associated cell sorting (FACS)

FACS requires single-cell suspensions from fresh gonadal tissue, meaning that embryos must be immediately genotyped and organised into the relevant experimental groups before cell dissociation. To achieve this, embryos were removed from uteri and placed into PBS in individual wells of 24 well tissue culture plates on ice. A small piece of tissue was taken for genotyping and immediately processed for genotyping. Genotyping was performed to determine chromosomal sex, presence of the *Map3k4* transgene and also the *Sf1-EGFP* transgene. Meanwhile, UGRs of each embryo were removed by dissection and placed back into individual wells of 24-well tissue culture

plates, and kept on ice in the dark. The presence of EGFP in UGRs was also confirmed by viewing them under a Nikon SMZ1500 microscope with fluorescence lamp.

Once genotyping results were available, tissue was sorted into appropriate experimental groups. The gonads and mesonephroi of 2-3 embryos were pooled together in Eppendorf tubes in a minimal volume of PBS and placed on ice and dissociation was performed immediately.

Cells were dissociated in 250 µl 25% trypsin in phenol red-free DMEM for 5 min at 37°C. The reaction was stopped by adding 25 µl FBS, gently mixed, then 500 µl PBS was added and tissue was centrifuged for 10 min at 1700 rpm, 4°C. Supernatant was removed and replaced with 500 µl fresh PBS. Tissue was then passed several times through pipette tips of decreasing sizes (P1000 and P200) and finally through a 23 G needle attached to a 1 ml syringe. To obtain a uniform single cell suspension, cells were then applied to a Celltrics 30 µm filter (Sysmex) and washed through with 200 µl PBS and collected in round-bottom flow cytometry tubes (Falcon). Tubes were kept on ice and FACS sorted immediately (Section 2.5.1)

2.3.7. Necropsy and histology

Mice were harvested between 8 and 19 weeks and culled by anaesthetic overdose by a member of the MLC Necropsy team (Dave Shipston or Naomi Busk). External genitalia were photographed before removal of the reproductive organs. Complete reproductive tracts were photographed on a black background by Kevin Glover (MRC Harwell Imaging group). Ovaries or testes were then fixed in 10% formalin and processed by the MLC Histology team for paraffin wax sectioning and haematoxylin and eosin (H&E) staining. Slides were photographed using an Olympus BX51 microscope with a digital camera attached.

2.4. Cell biology

2.4.1. Fluorescence-activated cell sorting (FACS)

Dissociated cells were sorted using an S3 cell sorter (Biorad). Cells are passed through a fluid stream and are exposed to excitation by a laser. The light that is transmitted and scattered by the cell is used to determine its size and fluorescence, in order to achieve a population of single, *EGFP*-expressing cells. *EGFP*-positive and *EGFP*-negative cells were collected from each sample by applying gates to sort the cells depending upon their fluorescence (Fig 3.1)

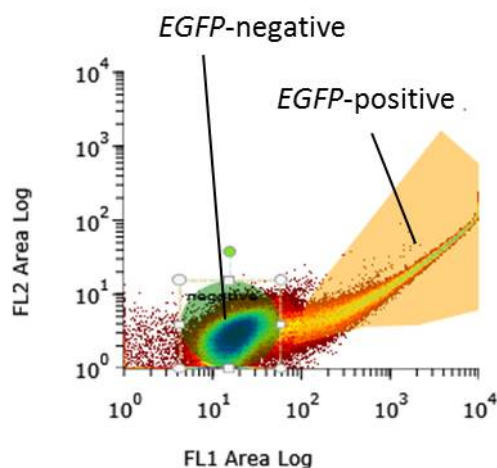


Figure 3.1 Collection of *EGFP*-positive and *EGFP*-negative cells by FACS.

EGFP-positive (orange shaded area) and *EGFP*-negative (green shaded area) were sorted by FACS.

2.5. DNA methods

2.5.1. DNA extraction for NGS

For next-generation sequencing, genomic DNA was extracted from tail tissue using an Illustra Nucleon genomic DNA extraction kit according to the manufacturer's instructions. A 2-4mm piece of tail was cut into small pieces and placed in an Eppendorf tube. To lyse tissue, 350 μ l lysis reagent and 20 μ l 20 mg/ml proteinase K was added to the tissue, mixed and incubated overnight at 55°C. The following day, the tube was vortexed and re-incubated for 30 min at

55°C. 100 µl sodium perchlorate was added to the lysate and mixed by inversion. 600 µl chloroform was then added to the tube and mixed by inversion, before adding 60 µl Nucleon resin, to separate the DNA and protein phases. The tube was centrifuged for 2 min at 2000 rpm and the upper phase carefully removed by pipetting. DNA was then precipitated by adding 2 volumes cold 100% ethanol and centrifuged for 5 min at 13000 rpm. Ethanol was removed and the DNA pellet was washed once with ethanol and centrifuged again. The pellet was then air-dried for 30 min and resuspended in 100 µl Tris-Cl pH 8.5 and stored at 4°C.

2.5.2. DNA extraction for genotyping assays

Genotyping was performed using genomic DNA extracted from ear biopsies of 4 week old animals or embryonic material collected at the time of dissection and stored at -20°C. Tissue was lysed in 25 µl 0.4 mg/ml Proteinase K in TE-Tween for 1-2 hours at 55°C. The enzyme was denatured at 95°C for 5min and the reaction cooled and pelleted by centrifugation at 4000 rpm for 5 min. The supernatant was then diluted 1:10 in ddH₂O to create a working solution for use in genotyping assays.

2.5.3. DNA extraction from ES cells in 96-well plates

ES cells were lysed in 96 well plates by adding 50 µl lysis buffer to each well and incubating overnight at 55°C in a humidified chamber. The next day, 100 µl of an ethanol/NaCl mix (300 µl 5 M NaCl + 20 ml cold absolute ethanol) was added to each well and incubated at RT for 1 hour to precipitate DNA. Plates were then centrifuged for 15 min at 2000 rpm and Ethanol/NaCl was removed by inversion. DNA was then washed 3 times with 70% ethanol (centrifuged for 5 min between each wash, ethanol removed by inversion) and air-dried at RT for 3-4 hours. DNA was then immediately used for restriction digest and Southern blot analysis (Section 2.7).

2.5.4. DNA extraction from FACS-sorted cells

Genomic DNA was extracted from FACS-sorted cells using an Allprep DNA/RNA micro kit (QIAGEN) according to the manufacturer's instructions. This kit enables simultaneous extraction of DNA, RNA and protein from up to 5×10^5 cells by a column-based method. Briefly, cells are lysed by vortexing in lysis buffer containing β -mercaptoethanol and applied to a membrane, which binds genomic DNA whilst allowing RNA and protein to be eluted. DNA is then washed with ethanol to remove salts, and eluted in pre-warmed elution buffer (Buffer EB, QIAGEN). DNA was immediately used for bisulfite conversion (Section 2.6.10).

2.5.5. Polymerase chain reaction

Primers for new assays were designed using the Primer3 online tool (<http://frodo.wi.mit.edu/primer3/>). Primers were synthesised by either Eurofins or Sigma Aldrich. As standard, PCR reactions were performed using the following recipe:

- 1x Maxima Hot Start PCR Mastermix (Thermo-Fisher)
- 0.2 μ M each primer
- 1 μ l working DNA solution (approximately 30 ng)
- H₂O to a total volume 10 μ l.

Maxima Hot Start PCR mastermix consists of Maxima Hot Start Taq polymerase, 2 x Hot Start PCR buffer, 0.4 mM each of dATP, dTTP, dCTP and dGTP and 4 mM Mg²⁺.

2.5.6. Gel electrophoresis

For analysis of DNA, RNA or PCR products, standard gel electrophoresis was used. Nucleic acids were either mixed with loading buffer or (1x TE, pH 8.0 / 10% glycerol / 0.08% Orange G) or, if already in PCR Mastermix containing loading dye, loaded directly onto 0.7-2% agarose (Invitrogen) gels made in TBE containing 1:10,000 SYBR®

Safe gel stain (Invitrogen). Depending upon expected product size, samples were loaded alongside either a 100 bp DNA ladder (NEB) or a 1kb-plus DNA ladder (Invitrogen). Unless otherwise stated, gels were electrophoresed for 30-50 minutes at 90 V and visualised using a UV Gel Doc XR system (BioRad).

2.5.7. Purification of PCR products

If PCR products were required for sequencing or subsequent ligation reactions, they were purified by either of the two following methods, according to the manufacturer's instructions.

QIAquick PCR Purification Kit (QIAGEN): This kit uses a column-based method to remove primers and buffer salts from PCR reactions and can recover products from 100bp up to 10kb. 100 µl PCR reactions were mixed with 5 volumes of dilution buffer, loaded onto columns and centrifuged to bind DNA to the silica membrane. DNA is washed with ethanol to remove salts and then eluted in 30µl elution buffer.

QIAquick Gel Extraction Kit (QIAGEN): this kit was used to isolate specific products of PCR reactions from an agarose gel. The kit follows the same principles as the PCR purification kit above, but the excised gel fragment is first dissolved in a buffer to facilitate binding of DNA to the column membrane.

2.5.8. Cloning

2.5.8.1. TA Cloning

Double-stranded DNA fragments were ligated into pGEM-T vector (Promega) following the manufacturer's instructions. This vector is a linearised vector that has a 3' T overhang on both ends, which prevents recircularisation of the vector and facilitates ligation of a PCR product. DNA amplified by a Taq polymerase has a 3' A overhang and therefore can easily anneal to the complementary T overhang of the vector. Incubation

of PCR product and vector with T4 DNA ligase enables generation of a circular plasmid that can then be transformed into competent cells.

A molar ratio of 3:1 was used to calculate the amount of insert to add to a reaction. The pGEM-T vector is 3 kb in size and 50 ng is required per reaction. The amount of insert was calculated using the following formula:

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

10 µl reactions containing 1X Rapid ligation buffer, 50 ng vector, insert DNA and 3 units T4 DNA ligase were incubated for 1 hour at RT or overnight at 4°C. 2 µl of ligation reaction was used for transformations.

2.5.8.2. Transformations

Plasmid DNA was transformed using subcloning efficiency DH5-α chemically competent cells (Promega) following the manufacturer's instructions. Approximately 10 ng plasmid DNA was added to 50 µl competent cells and incubated for 20 min on ice. Cells were then heat shocked for 20 seconds at 42°C in a water bath, then cooled on ice for 2 min. 950 µl of room temperature LB was added and the samples were incubated for 1 hour at 37°C in a shaking incubator at 150 rpm. Volumes of 50 µl and 100 µl of transformed culture were spread onto duplicate selective LB-agar plates containing 100 µl/ml ampicillin and incubated overnight at 37°C.

2.5.8.3. Purification of plasmid DNA

Colonies for each transformation were selected and used to inoculate 3 ml LB medium containing 100 µl/ml ampicillin. These clonal cultures were incubated at 37°C in a shaking incubator at 225 rpm overnight. Plasmid DNA from these cultures was then either purified using a Miniprep kit (QIAGEN) according to the manufacturer's

instructions, or used to inoculate 30 ml of selective LB at a 1:1000 dilution and incubated for a further night, before being purified using a Midiprep kit (QIAGEN).

The concentration of eluted plasmid DNA was determined using a nanodrop ND-8000 (Thermo Scientific) to measure the absorbance at 260nm.

2.5.9. DNA sequencing

DNA purified from PCR reactions or plasmid DNA was sent to GATC Biotech (Konstanz, Germany) along with 10 μ M specific sequencing primers if required. Sequence data were analysed using DNASTAR Seqman Pro (Lasergene 12, DNASTAR Inc, Wisconsin, USA) or online sequence analysis tools Ensembl BLAST (Ensembl.org) or Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>).

2.5.9.1. *Sry* sequencing

Primers for amplification of *Sry* flanking regions were designed to produce overlapping amplicons (see Fig 6.3). gDNA from B6, B6.Y^{POS} and B6.Y^{AKR} was used to amplify the *Sry* flanking regions of each strain. Details of assays and PCR conditions are given in Table 2.3. PCR conditions are as per SD (Table 2.2) with variation to extension time specified. Sequences of primers used for amplification are given in Table 2.4. PCR products were purified and sequenced as described in sections 2.6.7-2.6.9. Amplicons were sequenced using sequencing primers described in Table 2.5.

Individual sequence reads from B6.Y^{POS}, B6 and B6.Y^{AKR} were aligned with one another using Clustal Omega (EMBL-EBI, Cambridge, UK) (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) and variations initially identified by eye. Sequence reads for each strain were then assembled into a consensus using SeqMan Pro (DNA star Lasergene) and aligned to one another using Megalign (DNASTAR

Lasergene). Variations from B6 were annotated using the software and validated by re-checking individual sequence reads.

Table 2.3 PCR assays used to amplify *Sry* for sequencing

(For amplicon locations please refer also to Figure 6.3)

Assay name	Forward primer	Reverse primer	Amplicon size (bp)	extension time
A	F5	R2	3019	3 min
B	F4	R2	2020	2 min
C	F3	R2	1498	1 min 30 sec
D	F2	R2	1262	1 min 30 sec
E	F2	R1	2127	2 min
F	F1	R2	889	1 min
G	F0	R1	707	1 min
H	F6	R3	1290	1 min 30 sec
I	F7	R4	1245	1 min 30 sec
J	F8	R5	1345	1 min 30 sec
K	F9	R7	924	1 min
L	F10	R6	620	30 sec
M	F11	R8	1316	1 min 30 sec
CA repeat	CA_F	CA_R	424	30 sec

Table 2.4 Primer sequences for amplification of *Sry* ORF and flanking regions

Primer name	Sequence
R1	CGGGACCACACCATAAATGC
R2	AAAATCATCCAGGTTGAGCATTC
R3	AGGAGAAAAGGCCAGTGTCA
R4	AGGAGAAAAGGCCAGTGTCA
R5	AACTCCAGCTTACTCACCCC
R6	GTCCTTGACCTTTGCCTTGG
F0	CTGCCTGCCGTAGTAGACTA
F1	GAAGTCCAGTAGTGCCAGCT
F4	TAAGCCCCGTCTTCACCATT
F5	TGAAGCTGTCAGGGCCTTAA
F6	TAGTGTTCAAGCCCTACAGCC
F7	GGCATTTTACAGCCTGCAGT
F8	GGCATTTTACAGCCTGCAGT
F9	CGACCAGCAGCTTACCTACT
F10	TGACACTGGCCTTTTCTCCT
CA_F	ATACTTGCCCTCCCCTTCAC
CA_R	AGGGTTTCTCTCTAGCACACA

Table 2.5 Primers used for *Sry* sequencing

Primer name	Sequence
SrySeq_F1	CCCGTCTTCACCATTGTTCT
SrySeq_F10	GTGAAGGGGAGGGCAAGTAT
SrySeq_F11	TAGTTGGGCTGGACTAGGGA
SrySeq_F12	GCACATTTTGGTCAGTGGCT
SrySeq_F13	ACTAGTTGCATTGTTGTGGCA
SrySeq_F14	GGGGTGAGTAAGCTGGAGTT
SrySeq_F15	GGTCCTGCTCCTTCCTTTTG
SrySeq_F2	TGTGCACAAAGATGAAGGAGA
SrySeq_F3	ACATGTAAGTTACTGTTTACA
SrySeq_F4	ATGGCCATTTCATTCATCCA
SrySeq_F5	GCATACTAAAGAGACACAGCCT
SrySeq_F6	CGACCAGCAGCTTACCTACT
SrySeq_F7	GTGTGTAAGGAAGGAGAAAAGCT
SrySeq_F8	GCCACCACCGGAGACATATA
SrySeq_F8	GCCACCACCGGAGACATATA
SrySeq_F9	GGCTGGCTGTCATTTCTTA
SrySeq_R1	TCATCCAGGTTGAGCATTCT
SrySeq_R10	TCCCTAGTCCAGCCCACTA
SrySeq_R11	GGCTAGCACAAAGACAGCAA
SrySeq_R2	AACCCCTGCAACAGTTTCTT
SrySeq_R3	CATTTCAATGCCAACAATCA
SrySeq_R4	CTGCTTAGAGGTGAGTGCTG
SrySeq_R5	TAAGGAAATGACAGCCAGCCC
SrySeq_R6	TGAGGAATTCACAGCTGTCTT
SrySeq_R7	GGGGAGTGTTGGCATAGGTA
SrySeq_R8	AGCTAGCCCTTGAAAACACAGCTT
SrySeq_R9	TCACAGTTGAGGCTCCTTGG
altseq_F1	CACTACAAGACTTTACATACTCA
altseq_F2	TGACCAAGAAAGGAAGCAAAAGTT
altseq_F3	GTTTCTCTTTGCATCACATCTCTG
altseq_F4	AATTTCTCACGTTAGTCCATCCT
altseq_R1	TTGTGGAGGAGAACTAAAATGTCA
altseq_R2	CTTGAGGGGTTTTGCTTGACAGTA
altseq_R3	GACAGAATGATCAAAGAAAAAGCG
altseq_R4	AACTGAAGGTCTGGGTATGTAACA
altseq_R5	ATGGTGATGCTGTAAGGAAATGAC

2.5.10. Next generation sequencing

Next generation sequencing was performed by the High Throughput Genomics facility at Oxford University (Oxford, UK). DNA was prepared from mouse tissue by extraction as described in section 2.5.1 above. DNA concentration and quality was quantified using a nanodrop and diluted to 50 ng/μl. DNA was sent to the facility in a volume of 100 μl Tris-Cl pH 8.5 on dry ice.

Sequence data was aligned to the reference genome (mm10) by the MRC Harwell bioinformatics group (Dr Michelle Simon, Dr Saumya Kumar).

2.5.11. Bisulphite sequencing

Sodium bisulphite converts unmethylated cytosine residues to uracil. Thus, by sequencing the product of bisulphite conversion and comparing to the genomic DNA sequence of a given region, it is possible to determine which cytosine residues are methylated and therefore gauge the level of CpG methylation in a tissue of interest (Callinan and Feinberg, 2006).

100 ng gDNA extracted from 20,000 to 40,000 FACS-sorted gonadal somatic cells was used for bisulphite conversion. DNA was bisulphite converted using an EZ Methylation Direct kit (Zymo Research, California, USA) following the manufacturer's instructions. 20 μl (Approx 200 ng) genomic DNA was added to 130 μl CT conversion reagent. In a 100 μl PCR tube. Tubes were incubated in a thermocycler using the following conditions:

- 98°C for 8 min
- 64°C for 3.5 hours
- 4°C storage until required

DNA was then purified by adding 600 µl M-Binding Buffer (Zymo) to the sample and loading into a spin column. Columns were centrifuged for 30 seconds at 14000 rpm and flow-through was discarded. DNA was then washed by adding 100 µl M-Wash Buffer (Zymo) and centrifuging for another 30 seconds. DNA was desulphonated by adding 200 µl M-Desulphonation Buffer (Zymo) and incubating at RT for 20 minutes, the centrifuged for another 30 seconds. DNA was then washed a further 2 times with 200 µl M-Wash Buffer and centrifuges for 30 seconds between washes. Finally, DNA was eluted by adding 10 µl M-Elution Buffer (Zymo) to the column and centrifuging for 30 seconds at full speed. Converted DNA was then used immediately for PCR amplification steps.

Directly after conversion, 1 µl converted DNA was amplified by 2 rounds of PCR. A first round was performed using primers described in Table ..., under the following conditions:

5 cycles of:

- 30 min at 95°C
- 90 min at °C
- 120 min at 72°C

25 cycles of:

- 30 min at 95°C
- 90 min at ...°C
- 90 min at 72°C

Primers: F1 (AAGGAAATGATAGTTAGTTTTGATT),
R1 (ACCAAAATATACTTATAACAAAATTTTAA)

1 µl product from the first round of PCR was then used for a second round of PCR, under the same conditions as above but using an alternative, semi-nested reverse primer, R2 (TCAAACAAAACCCCTCAAATTTATAAA). The product of this second round of PCR was then purified, ligated into pGEM-T and transformed into DH5-α cells, as described in sections 2.6.7-2.6.9. At least 50 individual colonies per group were picked from selective plates and DNA purified from each colony was sequenced by GATC.

2.5.11.1. Analysis of bisulphite sequencing data

Sequence reads from clones, each containing a ligated fragment of bisulphite-converted DNA, were analysed using the BiQ analyser tool (Max Planck Institute, Germany), which aligns sequence reads and automatically detects CpG dinucleotides (Bock et al., 2005). Lollipop diagrams were generated using the BiQ Analyser's diagram function.

2.6. *Map3k1* gene targeting

2.6.1. L189R construct design

The L189R substitution is caused by a T to G mutation at position 566 of the cDNA sequence of human *MAP3K1* in exon 2, in a highly conserved region (Pearlman et al., 2010). In mouse, the equivalent leucine residue is at position 184 of the MAP3K1 polypeptide which corresponds to nucleotide position 551 of the cDNA. For simplicity, the construct will be referred to as L189R. A PCR-based site-directed mutagenesis strategy for introduction of the single nucleotide variant encoding a Leu-Arg substitution at position 184 was designed by Dr Nick Warr. The mutated nucleotide is

flanked 5' by 4.1 kb homologous *Map3K1* sequence from the targeted strain, C57BL/6N. 3' of the targeted mutation, an unconserved region within intron 2 was chosen for insertion of a *loxP-FRT-neo-FRT-loxP* cassette. This is then flanked on the 3' side by 3.9 kb homologous sequence (Fig 3.3). The L189R targeting construct was ligated into a *Tk-loxP-neo* plasmid vector and cloned into DH5- α cells by a commercial manufacturer (GenScience).

2.6.2. Vector preparation and transfection

The plasmid targeting vector was purified from DH5- α cells as described in section 2.6.8.3. 50 μ g of plasmid DNA was linearised with *ScaI*. Digest completion was checked on a 0.5% agarose gel. Concentration was determined using a Nanodrop ND-8000. 1 X 10^7 JM8 F6 semi-confluent ES cells were electroporated with 10 μ g linearised targeting vector. Transfection of ES cells, antibiotic selection, colony picking and freezing of ES cell clones was performed by the Transgenesis Facility, Mary Lyon Centre, MRC Harwell (Denise Lynch, Dr. Gemma Codner and Dr. Lydia Teboul). Colonies picked for screening were provided frozen in 96-well format.

2.6.3. Strategy for screening ES cells

Two DNA probes complementary to regions just outside the 5' and 3' homology arms of the L189R targeting construct, respectively, were designed by Dr Nick Warr. Identification of homologous recombinants was performed by Southern blotting. In this case, addition of the *Neo* selection cassette introduces additional *SpeI* and *BsaWI* restriction sites (Fig 3.4). Sizes of expected restriction fragments are given in Table 2.6.

Table 2.6: Identification of targeted *Map3k1* recombinants by restriction digest

	Restriction enzyme	Fragment size in WT	Fragment size in L189R
5' probe	<i>SpeI</i>	8.2kb	5.3kb
3' probe	<i>BsaWI</i>	10.1kb	6.6kb

2.6.4. ES Cell screen

Identifying targeted ES clones was performed by screening ES cell colonies grown in media supplemented with G418 (Sigma). G418-resistant colonies were screened by Southern blot analysis using extracted DNA digested with *SpeI* (NEB) and hybridisation of the 5' external probe described above. Both probes were generated by PCR and purified by Dr Nick Warr.

2.6.5. Restriction digests of ES cell gDNA in 96 well plates

Approximately 10µg ES cell genomic DNA was digested in 96 well plates by adding 30µl restriction digest mastermix (1 x NEB Buffer 4, 1mM spermidine, 100 µg/ml BSA, 100 µg/ml RNase, 10 U *SpeI* endonuclease per well) to each well and incubating at 37°C overnight.

2.6.6. Southern blotting

Restriction-digested DNA was electrophoresed on a 0.7% agarose gel in TAE until sufficient resolution was achieved (approx. 6.5 hours, 85 V). The gel was stained with ethidium bromide and photographed against a ladder of known DNA size markers (1 kb ladder, NEB) and a ruler, using a UV transilluminator. The gel was then washed in denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 1 hour at RT on an orbital shaker, then in neutralising solution (0.5 M Tris pH8, 1.5 M NaCl) for 1 hour at RT on an orbital

shaker. The gel was rinsed in 20x SSC for 20 min at RT. DNA was transferred to charged nitrocellulose membranes (Amersham Hibond N+, GE Healthcare, Buckinghamshire, UK) overnight via capillary action using 20x SSC. The next day, positions of the wells were marked on the membrane using a pencil. The membrane was UV cross-linked using a Stratalinker set to the automatic cross link function (Stratagene) and stored at RT.

2.6.7. Southern Hybridisation

Each nitrocellulose membrane was pre-hybridised in 20 ml Quikhyb (Stratagene) solution in a roller tube at 68°C for at least 1 hour. Meanwhile, radiolabelled DNA probe was synthesised using a Megaprime labelling kit (Amersham) according to the manufacturer's instructions. Briefly, 25 ng probe DNA was used as a template for synthesis of α P³²-dCTP-labelled probe appropriate to the assay (Table 2.6). Probe was purified from non-incorporated α P³²-dCTP nucleotides using Illustra Sephadex G-50 DNA Nick columns (GE Healthcare) according to the manufacturer's instructions. The probe was eluted in 400 μ l TE, and 100 μ l sonicated salmon sperm DNA (Stratagene) was added and the probe then denatured at 100°C for 5 min and cooled on ice. Denatured probe was then added to the roller tube containing the membrane and hybridised for 4 hours. The membrane was then washed 2 x 15 min in 2x SSC, 0.1% SDS at RT, followed by 2 x 30 min in 0.1x SSC, 0.1% SDS at 65°C. Membranes were then sealed in polythene wrap, placed in a cassette and exposed to autoradiography film (GE Healthcare) at -80°C at least overnight. Films were developed using a Compact X4 imaging system (Xograph Imaging Systems, Gloucestershire, UK).

2.7. RNA Methods

2.7.1. 3' UTR riboprobe design

New riboprobe (*Pnliprp2*) was designed to hybridise to the 3' untranslated region (UTR) of the gene of interest. This region is unique to each gene, so probes designed using this region are less likely to bind off-target to other genes. A region of 523 bp was amplified from genomic DNA (from ear biopsies) by standard PCR using primers designed in Primer3 (Table 2.7). PCR product was purified as described in Section 2.5.7 and then ligated into the pGEM-T vector and cloned using DH5- α cells (Section 2.5.8). To confirm presence of the cloned region, plasmid DNA was sequenced using M13 forward and reverse sequencing primers, which have a binding sites either side of the cloning region of the vector. sequencing this way also enables determination of the orientation of the insert, so that an appropriate restriction enzyme and RNA polymerase can be used to transcribe the DNA.

Table 2.7. Primers and PCR assay conditions for synthesis of *Pnliprp2* riboprobe

Probe name	Primers	PCR conditions	Product size (bp)
<i>Pnliprp2</i>	F: ATGACTGCTTCCCTTGTCCA	SD*, 57°C	523
	R: GACCAAGTGCATCAGGAACG		

*SD, Standard PCR program (see Table 2.1)

2.7.2. Riboprobe synthesis

Before synthesising riboprobe RNA, the plasmid was first linearised by digestion using a restriction enzyme with a site located at the 5' end of the sense strand (Table 2.8). 15 µl plasmid (at approximately 1 µg / µl) was added to 5 µl 10X buffer, 2 µl restriction enzyme and 0.5 µl BSA (if required) (all reagents from NEB) in a total volume of 50 µl (made up with H₂O). The reaction was incubated for 2 hours at 37°C. Restriction digest completion was checked by electrophoresis in a 1% agarose/TBE gel alongside uncut plasmid DNA. Linearised DNA was precipitated by adding 10% volume 3 M sodium acetate (pH 4.5) and 2.5 volumes 100% ethanol. The solution was incubated for 30 min on dry ice and then centrifuged at 14,000rpm, 4°C. Supernatant was removed and replaced with 200 µl 70% ethanol, then centrifuged again for 10 min at 4°C. Ethanol was removed and the DNA pellet briefly air-dried, then resuspended in 20 µl H₂O.

Table 2.8. Conditions for RNA probe synthesis

Probe name	Restriction enzyme	RNA polymerase	Reference
<i>Pnliprp2</i>	<i>NotI</i>	T7	-
<i>Sox9</i>	<i>HindIII</i>	T7	Wright et al 1995
<i>Sry</i>	<i>SacI</i>	T3	Bogani et al 2009
<i>Stra8</i>	<i>HindIII</i>	T7	Bogani et al 2009

3 µl linearised DNA was then added to a 20 µl reaction containing 2 µl 10x transcription buffer, 2 µl 10 X DIG RNA labelling mix, 1 µl 40 U/ µl Rnasin, 2 µl 20 U/ µl RNA polymerase (SP6, T7 or T3) (all reagents from Roche) and H₂O. The reaction was incubated for 2 hours at 37°C. 2 µl 20 U/ µl DNaseI was then added and the reaction

incubated at 37°C for a further 20 min to remove plasmid DNA. The reaction was stopped and RNA precipitated by addition of 2 µl 200mM EDTA, 2.5 µl 4M lithium chloride and 75 µl 100% ethanol. RNA was precipitated overnight at -20°C or for 30 min on dry ice. The reaction was centrifuged at 14,000rpm, 4°C for 20 min and supernatant discarded. The pellet was washed with 200 µl ice cold 70% ethanol and centrifuged for a further 2 min. The ethanol was removed and the pellet resuspended in 100 µl H₂O. Probe quality was checked by running 5 µl on a 1% agarose gel. All riboprobes were stored at 20°C until required.

2.7.3. Wholemout *in situ* Hybridisation (WMISH)

WMISH was performed using 12-well culture dishes with Netwell (Fisher) inserts. All steps were performed at RT unless otherwise stated.

Tissue collected as described in Section 2.3.2 was first rehydrated through a methanol series for 5 min per wash, into PBST (1x75% methanol, 1x 50%, 1x 25%, 2x PBST). Tissue was bleached with 6% hydrogen peroxide in PBST for 40 min then washed 3 x 5 min in PBST. Tissue was digested with 10µg/ml proteinase K (Roche) for 20min and stopped with 2mg/ml glycine for 5 min, followed by 2 x 5 min PBST washes then post fixation in 0.2% gluteraldehyde in 4% Paraformaldehyde (PFA) in PBST for 20 min. After 2 X 5 min washes in PBST, tissue was incubated in pre-hybridisation solution at 70°C for at least 1 hour. RNA probe was denatured at 70°C for 10 min, cooled before adding 5µl per ml hybridisation solution. Tissue was transferred into hybridisation solution containing probe and incubated overnight at 70°C.

The next day, tissue was washed in WMISH wash solution (Table 2.1) for 2 X 30 min at 70°C then 2 X 30 min at 65°C. tissue was then washed 3 x in TBST at RT, before being transferred into 10% heat inactivated sheep serum (HISS) in TBST to pre-block for a

minimum of 2.5 hours. Antibody solution was prepared by first heat inactivating embryo powder for 30 min at 70°C in 1ml 1 x TBST, the pulsed down to a pellet and TBST removed. Embryo powder was then resuspended in 1% HISS in TBST and 1µl anti-DIG-alkaline phosphatase conjugated Fab fragments (Roche) per ml of reaction was added to the embryo powder and mixed on a rotator for at least 1 hour at 4°C. After a 5 min spin, supernatant containing the antibody was removed and transferred to wells containing 1% HISS in TBST, and samples incubated in the antibody-containing solution overnight at 4°C.

The following day, tissue was washed in TBST for 3 x 5 min at RT, then 5 x 1 hour at RT, before being left overnight in TBST at 4°C. the following day, colour was developed by first washing tissue in NTMT for 3 x 10min and transferring to glass dishes and incubated in alkaline phosphatase stain (20µl NBT/BCIP [Roche] in NTMT) in the dark until colour had sufficiently developed. The reaction was the stopped with PBST and samples washed in PBST for up to three days, before being post-fixed in 4% PFA for 1 hour. Samples were then stored in 0.4% PFA in PBST.

WMISH samples were imaged on 1% agarose in dishes flooded with PBS using a Leica MZ16 dissection microscope with a digital camera attached.

2.7.4. RNA extractions

Total RNA was extracted from frozen embryonic gonads using an RNeasy Plus™ Micro Kit (QIAGEN) according to the manufacturer's instructions. Briefly, lysis buffer containing 1% β-mercaptoethanol is added to samples and tissue us then lysed by vortexing. For RNA-seq, tissue from 15-20 embryos was first pooled in ice cold PBS, briefly centrifuged and supernatant removed before adding lysis buffer. Lysate is then loaded onto "gDNA-eliminator" spin columns, which bind DNA but allow RNA to be

passed through. Lysate containing RNA is eluted by centrifugation and then precipitated with 70% ethanol. This lysate is then applied to a second spin column, which binds RNA but allows protein to be eluted. Bound RNA is then washed several times to remove salts, before being eluted in 16µl RNase-free water (QIAGEN). To improve yield for RNA-seq, 2 elutions were performed by applying the same 16µl RNase-free water twice to the RNeasy column. RNA concentration and quality was determined by measuring absorption at 260 nm using a Nanodrop ND-8000.

2.7.5. Reverse transcription

cDNA was synthesised from 200-220 ng RNA using a High Capacity Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's instructions. Synthesis was performed in a thermocycler under the following conditions: 25°C for 10 min, 37°C for 120 min, 85°C for 5 min. Following synthesis, cDNA was diluted 1:6 with H₂O.

2.7.6. qRT-PCR

qRT-PCR was performed using the Fast SYBR® Green assay system, which enables quantitation of cDNA based on incorporation of the SYBR® Green dye into double stranded DNA during Taq polymerase-mediated amplification using specific primers. As the SYBR green dye is incorporated into each newly formed double-stranded DNA molecule, it emits excitation energy which is detected as fluorescence at 521 nm by the thermocycler. The level of fluorescence is recorded at the end of each amplification cycle and directly correlates to the amount of cDNA present in the reaction. qRT PCR measures the number of cycles it takes to reach a specific threshold of fluorescence and this is called the cycle threshold (Ct). Samples in which the mRNA

(and therefore cDNA) of interest is more abundant will reach this threshold more quickly, and have a smaller Ct.

Primer sequences are given in Table 2.9. Primers for new qRT-PCR assays were designed using Primer3 and synthesised by Eurofins.

Table 2.9 Sequences of qRT-PCR primers

Gene name	Forward primer	Reverse primer
<i>Hprt1</i>	CATTATGCCGAGGATTTGGAA	CACACAGAGGGCCACAATGT
<i>Sox9</i>	AAGAAAGACCACCCGATTACA	CAGCGCCTTGAAGATAGCATT
<i>Sry</i>	TTATGGTGTGGTCCCGTGGT	GGCCTTTTTTCGGCTTCTGT
<i>Gadd45g</i>	GGAGACCTGCATTGCATCCTCATTT	ACTCGGGAAGGGTGATGCTGG
<i>Map3k4</i>	GAGAGTGGACGAGGAGCCAAC	CTTCAGTGCTTTGTCCACGA
<i>Sf1</i>	TTCGTCTGTCTCAAGTTCCTCATC	CCTTTACGAGGCTGTGGTTGTT
<i>Wt1</i>	AGGTTTTCTCGCTCAGACCA	CCTGGTGTGGGTCTTCAGAT
<i>Tgm3</i>	GTTGCAAGCGGATGATGTCT	GCCAATTCGATTGTGCTGC
<i>Pnliprp2</i>	CCGCGGAAGTTCGGTTGGAC	ATCCAGATGGCCGACCTTTTG
<i>Col4a1</i>	CCTGGCACAAAAGGGACGA	ACGTGGCCGAGAATTTACC
<i>Col4a2</i>	GGACCGAGTGCGTTCAAA	ACCGCAGGGCACATCCAAC
<i>Cldn6</i>	TGTCACCCTCCTCATTGTCC	CACCAGACGAGACTTGGAGT
<i>Wnt9b</i>	GCACCAAGTTCCTCAGCAAC	CTCTTCACAGCCTTGATGCC
<i>Pou3f3</i>	CCATGCCTAAAGATTCCGCC	GCAGAAGTCAAGGGAAGTGC
<i>Tfcp2l1</i>	CAAGGTGTTCAAGCCCAAGG	CACTCGGTAAGGATGGTGGT
<i>Gja5</i>	GTGCGATACCATTGAGCCTG	GTGTCCACGCCTTCTCTAGT

2.7.6.1. Assay efficiency

For new qRT-PCR assays that had not been previously validated in the lab, assay efficiency was calculated. Using a range of concentrations of cDNA from 5 pooled WT gonads, standard curves were generated. Seven relative dilution points were used: 1:1, 1:3, 1:9, 1:27, 1:81, 1:243, 1:729. The standard curve was generated by plotting Ct over log of the cDNA dilution and the efficiency of the assay was calculated using the gradient of the linear regression (equation 1).

Equation 1: Assay efficiency (%)

$$\left(10^{\left(\frac{-1}{\text{slope}}\right)} - 1\right) \times 100$$

A slope of -3.32 would confer efficiency of 100%. Assays with efficiency of 90-110% are deemed acceptable. Efficiencies calculated for assays designed for this thesis are given in Table 2.10.

Table 2.10 qRT-PCR assay efficiency

Assay name	Slope	Efficiency
<i>Cldn6</i>	-3.27	98%
<i>Pnliprp2</i>	-3.59	97%
<i>Pou3f3</i>	-3.07	97%
<i>Gja5</i>	-3.37	96%
<i>Tfcp2l1</i>	-3.10	99%
<i>Col4a1</i>	-3.88	98%
<i>Col4a2</i>	-3.50	97%
<i>Wnt9b</i>	-3.87	96%
<i>Tgm3</i>	-3.05	98%

2.7.6.2. Plate setup

Plates were prepared by first making a master mix for all reactions. 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 0.2 µM each specific primer (Table 2.1). 15 µl master mix was then added to each well of a Microamp fast optical reaction plate (Applied Biosystems) and 5 µl (approximately 9 ng) cDNA added to each well. All samples were assayed in duplicate. Real-time thermocycling and fluorescence detection was performed using the 7500-Fast real time PCR system (Applied Biosystems).

2.7.6.3. Analysis

Following completion of the reaction, data was analysed by first calculating the cycle threshold (Ct) for each sample in the reaction plate, using 7500 Fast system software. All following data analysis was performed using Excel 2010 (Microsoft) or Graphpad Prism (Graphpad Software Inc, California, USA). Relative expression levels were calculated using the comparative Ct method, which gives the ratio (R) of the Ct value of the target gene versus the endogenous gene *Hprt1* (equation 2).

Equation 2: Relative expression (R)

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

The mean R of all the samples in a particular age group (n = 2 to 6 pairs of gonads) was calculated to produce plots of relative expression. The standard deviation (Stdev) of each age group was calculated and used to estimate error by calculating the Standard error of the mean (SEM) (equation 3).

Equation 3: Standard error of the mean

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

2.7.7. RNA Sequencing

RNA was extracted from gonads and mesonephroi of 11.5 dpc embryos according to section 2.7.4. 1.3 – 3 µg RNA was sent to Oxford Gene Technology (OGT) for cDNA library synthesis and sequencing. Data analysis was performed by the Bioinformatics group at MRC Harwell (Dr Michelle Simon and Siddarth Sethi). Briefly, reads were aligned to the reference genome

(mm10) using the alignment tool TopHat (<https://ccb.jhu.edu/software/tophat/index.shtml>). Differential expression was then tested using three different bioinformatics tools: Cufflinks (<http://cole-trapnell-lab.github.io/cufflinks/>), DEseq (<http://www-huber.embl.de/users/anders/DESeq/>), EdgeR (Robinson et al., 2010) and NOISeq (<https://www.bioconductor.org/packages/release/bioc/html/NOISeq.html>). A report was produced by Siddharth Sethi and the top 44 differentially expressed genes are presented in this thesis. Differential expression is presented in reads per kilobase per million (RPKM), which is the number of sequence reads per kilobase of transcript, per million mapped reads.

2.8. Protein methods

2.8.1. Western blots

2.8.1.1. Protein extraction

Protein was extracted from 11.5 dpc gonads collected for Western blotting (Section 2.3.4). For each experimental sample 3 pairs of gonads were pooled together in 15 μ l PBS containing 1X Complete Mini protease inhibitor cocktail (Roche) and 1X Phos-Stop phosphatase inhibitor (Roche). Protein was extracted by lysing tissue in 15 μ l Laemmli Buffer, plus 2 μ l β -Mercaptoethanol and 10 μ l H₂O. Lysate was sonicated in a Bioruptor Plus (Diagenode) for 20 cycles of 15s on/off at 4°C to break up genomic DNA. Samples were then denatured at 95°C for 3 min and cooled on ice. Samples were then immediately prepared for immunoblotting or stored at -20°C.

2.8.1.2. Protein electrophoresis and blotting

5 μ l protein lysate was mixed with 4 μ l LDS buffer and 1 μ l reducing agent (Invitrogen). Protein was size-separated using NuPage 4-12% linear gradient bis-tris ready-poured polyacrylamide gels (Invitrogen) with 1 X MES electrophoresis running buffer (Invitrogen). Samples were electrophoresed alongside a SeeBlue Plus protein standard protein ladder (188 kDa to 3 kDa, Invitrogen) to allow estimation of molecular weight.

Protein was then transferred to a nitrocellulose membrane using the iBlot dry electrotransfer system by passing 20 V across the iBlot stack for 7 min.

2.8.1.3. Immunoblotting

All steps were performed at RT unless otherwise stated. Membranes were rinsed in 1X TBST and blocked in 5% BSA for 1 hour. Membranes were then incubated with primary antibody at the appropriate concentration in 5 ml 5% BSA (Table 2.11) overnight at 4°C on a rolling mixer.

The following day, membranes were washed 3 x 10 min in 1X TBST with shaking, then incubated with the appropriate horseradish peroxidase (HRP) conjugated secondary antibody (Table 2.11) in 5% BSA for 1 hour. Membranes were given 3 further washes for 10 min in 1X TBST. Membranes were incubated with Pierce ECL 2 reagent (Life Technologies) for 5 min, according to the manufacturers' instructions. Excess fluid was drained and the membranes placed into autoradiography cassettes and covered with acetate. Membranes were then exposed to Hyperfilm ECL autoradiography film (GE Healthcare) and exposed films were developed using a Compact X4 imaging system (Xograph).

For re-probing with a loading control antibody, membranes were washed 3 times in 1X TBST and then re-blocked with 5% BSA for 1 hour. Membranes were then incubated with the appropriate loading control antibody overnight at 4°C, and the wash, secondary incubation and chemiluminescence steps above repeated.

Table 2.11 Antibodies used for Immunoblotting

		Source	Dilution
Primary antibodies	Rabbit anti- β -Actin	Gift from H. Tateossian (MRC Harwell)	1:3000
	Rabbit anti-GATA4	Abcam Ab32298	1:2000
	Rabbit anti- phospho-GATA4	Abcam AB63398	
	Rabbit anti- p38 MAPK	Cell Signalling Technologies #9212	1:1000
	Rabbit anti- phospho-p38 MAPK	Cell Signalling Technologies #9215	1:1000
	Mouse anti α -tubulin	MRC Harwell protein core	1:3000
Secondary antibodies	Goat anti-rabbit HRP conjugate	Abcam AB6721	1:100,000
	Rabbit anti-mouse HRP conjugate	MRC Harwell protein core	1:10,000

2.8.1.4. Quantification and normalisation of Western Blots

To enable the normalisation of protein samples in each experiment, 5 μ l lysate was first loaded as a test run. Films from these blots, performed using anti-Tubulin antibody as described above, were imaged with a Geldoc XR (Biorad) and quantification was performed by manually drawing around each band and calculating band density using Imagemag (Biorad) software. Each sample in the experiment was then normalised against the sample with the lowest relative protein amount by diluting in LDS buffer accordingly before electrophoresis.

Experimental blots were quantified as described above. For each blot performed, three biological replicates (each containing 3 pairs of subdissected gonads) were used. The

mean density of bands for each sample was calculated relative to loading control (β -actin or α -tubulin) using Microsoft Excel.

2.8.2. Immunohistochemistry

The antibodies used for analysis of protein expression on wax sections of embryonic gonads are detailed in table 2.12.

Table 2.12 Antibodies used for immunohistochemistry

Antibody	Supplier	Dilution	Species	Secondary antibody
anti-SRY	Gift from D. Wilhelm	1:50	Guinea pig	anti-guinea pig Alexa fluor 488 (Fisher) 1:200
anti-SOX9	Abcam AB5535	1:50	Rabbit	anti-rabbit Alexa fluor 594 (Fisher) 1:200

Slides were dewaxed in 1X Declere solution (Sigma) in a steamer for 45 min. slides were then transferred to pre-warmed fresh Declere and left to cool for 30 min. slides were rinsed in TBSTx and bleached with 0.3% hydrogen peroxide for 10 min and washed twice with TBSTx for 5 min. All further steps take place at RT unless otherwise stated.

A wax pen was used to draw around sections and 500 μ l block (2% skimmed milk powder [Sigma], 0.1% Triton X100 in TBS) was placed onto each slide and incubated for 1 hour. Block was poured off and replaced with 200 μ l block containing the two primary antibodies at their respective dilutions and incubated at 4°C overnight.

The next day, slides were washed three times in TBST on a shaker. Slides were then incubated with 500 μ l block containing the secondary antibodies at their appropriate dilutions, for 1 hour. Slides were washed for a further 3 x 5 min in TBST and then mounted with cover slips, using Vectashield (Vectalabs, California, USA) mounting

medium with DAPI. Slides were then imaged using a Zeiss LSM710 upright multiphoton microscope equipped with ZEN software.

Chapter 3. Generation of mice bearing a human sex-reversing mutation in *Map3k1*

3.1. Introduction

The kinase MAP3K4 has been shown to be essential for mouse testis determination and the regulation of *Sry* expression (Bogani et al., 2009, Gierl et al., 2012, Warr et al., 2012). However, so far, no *MAP3K4* mutations have been associated with 46,XY DSD in humans. This may be because loss of function of MAP3K4 is not compatible with life, since mice lacking *Map3k4* die perinatally and exhibit *spina bifida* and/or exencephaly (Bogani et al., 2009). However, it may also be due to discrepancies in signalling pathway utilisation between mice and humans. Mutations in another MAP3K superfamily member, *MAP3K1*, have been described in several cases of human 46,XY DSD, including complete gonadal dysgenesis (CGD), supporting the notion that MAPK signalling is an important regulator of sex determination in humans as well as mice. However it is unclear whether MAP3K1 adopts a similar role in humans to that of MAP3K4 in mice, or whether the two are members of separate MAPK signalling pathways that together contribute to testis determination in both species.

MAP3K1, encoded by *MAP3K1* and also known as MEKK1, is developmentally important for the migration of epithelial cells and crucial for coordinating signals for eyelid closure in the mouse embryo (Geh et al., 2011). It has been strongly linked with the cytoskeleton and has been shown to promote growth factor-induced cell migration (Xia et al., 2000) and the formation of actin stress fibres (Zhang et al., 2005). MAP3K1 binds and activates MAP2K1 and MAP2K4 by phosphorylation, which in turn activate JNK and ERK kinases as well as p38 α MAPK (Craig et al., 2008). MAP3K1 is also known to physically associate with the GTPase RHOA, through a conserved binding motif (Gallagher et al., 2004). RHOA has been

implicated in repression of *Sox9* expression and also the transcriptional activity of its protein product in chondrocytes (Woods et al., 2005, Kumar and Lassar, 2009).

3.1.1. *MAP3K1* mutations in human DSD

In 2003, a family was reported with multiple cases of 46, XY DSD across several generations. This included seven females with 46, XY CGD and four males with genital abnormalities including hypospadias and cryptorchidism (Jawaheer et al., 2003). Linkage analysis identified a critical region on chromosome 5, and later the causative gene was identified as *MAP3K1* (Pearlman et al 2010). Sequencing *MAP3K1* in this family identified a splice-site mutation that led to aberrant splicing of the RNA and insertion of six nucleotides, which resulted in an in-frame insertion of two amino acids (isoleucine and glycine) into the protein (Pearlman *et al* 2010). In this study, *MAP3K1* mutations were also identified in a second family and in two sporadic cases of 46, XY CGD. These were missense mutations leading to amino acid substitutions: Gly616Arg, Leu189Pro and Leu189Arg, respectively (Figure 3.1).

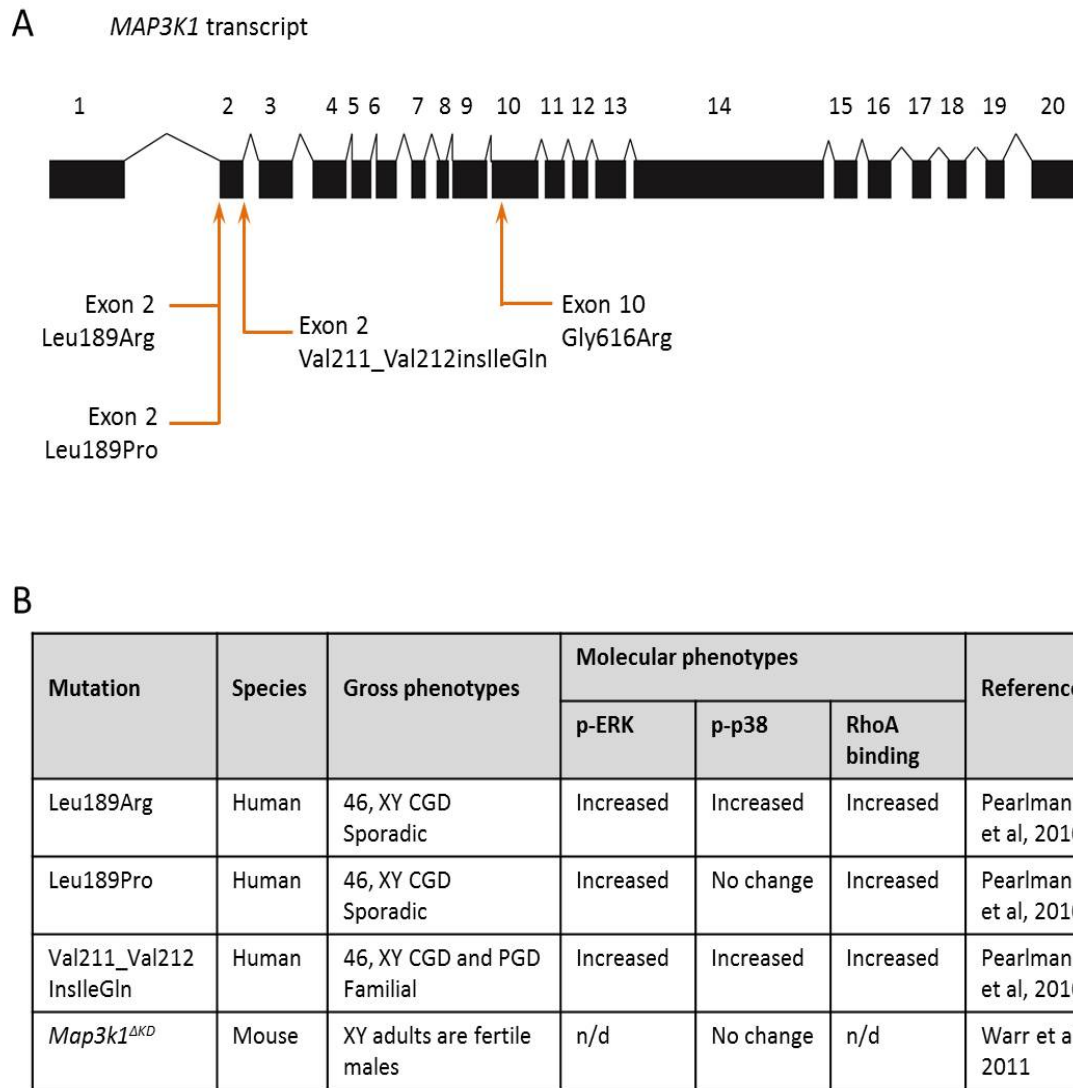


Figure 3.1 Human sex-reversing *MAP3K1* mutations

(A) Locations of reported mutations on *MAP3K1* transcript. Numbered boxes represent exons. Mutations that have been identified in exon 2 include Leu189Arg, Leu189Pro and Val211_Val212inslleGln. A further missense mutation, Gly616Arg, has been reported in exon 10. (B) Summary of gonadal phenotypes associated with *MAP3K1* mutations in humans and mice (n/d, not determined). Data obtained from Pearlman et al (2010) and Warr et al (2011).

Pearlman *et al* went on to investigate the molecular mechanisms underlying these *MAP3K1* mutations, using patient-derived lymphoblastoid cell lines carrying the mutations. For the Leu189Arg (L189R) mutation, in particular, they found that levels of phosphorylated ERK (p-ERK) and phosphorylated p38 (p-p38) were elevated in mutant cells, indicating that disruption to MAPK signalling could underlie disrupted sex determination seen in patients carrying this mutation. Moreover, the mutations arose in the N-terminal region of MAP3K1, which is highly conserved and has been shown to be the site of binding of RHOA (Gallagher et al., 2004). In immunoprecipitation assays using lysate containing L189R mutant protein, RHOA showed increased binding compared to WT lysate.

Whilst these cell-based studies provided clues as to the molecular effects of *MAP3K1* mutations, without a tractable model system it is unclear what the significance of this altered signalling is in the context of gonad development. To address this, Warr *et al* (2011) performed analysis of testis determination in embryos and adult mice lacking functional MAP3K1. On the B6 background, XY *Map3k1*^{Δkd/Δkd} embryos developed testes and expressed appropriate markers of testis determination at 13.5dpc, including *Sox9*, although mutant XY gonads were slightly longer than WT. Moreover, no difference was observed in levels of p-p38 between control and homozygous mutant gonads at 12.5dpc. The *Map3k1*^{ΔKD} mutation is mostly embryonic lethal in homozygotes, but some homozygous males survived to adulthood and were fertile (Warr et al., 2011).

Taken together with the data reported by Pearlman *et al* (2010), these results indicate that the sex-reversing human *MAP3K1* mutations may not be loss-of-function alleles but instead may confer gain-of-function. Therefore it is necessary

to find a relevant model system in which to further investigate this hypothesis and confirm whether the observations made in cell lines are recapitulated *in vivo*. Since testis determination is such a crucially time-sensitive event, it would be highly informative to examine the molecular effect of these mutations in gonadal tissue at the time of testis determination. Introduction of a human sex-reversing mutation into the mouse would enable analysis of the impact of a specific *Map3k1* disruption on testis determination, and would also facilitate further biochemical studies of the underlying mechanism in a relevant tissue at a developmentally appropriate stage. To achieve this, a targeted mutagenesis strategy was devised by Dr Nick Warr (MRC Harwell), with the aim of introducing the L189R amino acid substitution into mouse MAP3K1.

3.1.2. Targeted mutagenesis in the mouse

At the time of beginning this study, the most efficient and cost-effective method for introducing a targeted mutation into the mouse genome was to use site-directed mutagenesis in embryonic stem (ES) cells. This method, pioneered in the 1980s, has since been used to generate thousands of knockout mouse lines, as well as lines carrying point mutations (Capecchi, 1989, Rubinstein et al., 1993, Ayadi et al., 2012).

Targeted mutagenesis first requires introduction of the desired mutation into ES cells. This is performed by transfection of a plasmid targeting vector. The vector contains the mutated sequence along with an antibiotic selection cassette, and is flanked by arms of sequence homologous to the gene of interest to maximise the likelihood of integration into the correct genomic location. The targeting vector

inserts randomly into the genome, but in a few cells will pair with the complementary chromosomal DNA sequence of the homology arms and integrates into the required site by homologous recombination. Cells are grown in medium containing the appropriate selective antibiotic to enrich for homologous recombinants. A positive selection marker is usually the gene for neomycin phosphotransferase (*neo*), which confers resistance to the drug G418. Therefore culturing cells in G418 selects for cells that have integrated the construct.

Accurate screening for homologous recombinants also relies upon the ability to differentiate between cells that have integrated the selection cassette randomly and cells that have integrated it in the correct position, since enrichment using antibiotic selection does not guarantee all surviving colonies have undergone correct integration events. Therefore, a simple assay to identify the presence of *neo* will not suffice. Instead, the targeting vector is designed such that its integration into the genome will introduce novel restriction sites to the targeted region. By digesting ES cell DNA with appropriate restriction enzymes and resolving on an agarose gel, it is possible to identify clones that have integrated the vector in the correct place by the size of restriction-digested fragments in randomly versus correctly targeted cells. To visualise the electrophoresed fragments, Southern blotting is used (Southern, 1975). DNA within the agarose is first transferred to a charged nitrocellulose membrane and can then be hybridised with DNA probes complementary to approximately 500 bp of sequence just outside the homology arms of the targeting vector. The probes will bind different sized fragments depending upon the presence or absence of the additional restriction sites within the vector, so it is possible to determine whether it has been integrated in the

correct position by visualising the hybridised probes. This is commonly achieved by radiolabelling them with $\alpha^{32}\text{P}$, and then exposing the hybridised membrane to X-ray film.

Following identification of homologous recombinants, the targeted ES cell clones are then microinjected into pre-implantation blastocysts, which are transferred into pseudopregnant mice (Fig 3.2). Pups that develop from injected blastocysts to which ES cells have successfully contributed are chimaeras, containing a mixture of cells derived from the targeted ES cells and from the WT cells of the host blastocyst. Chimeric mice are then bred to wild-type mice. Chimeras that have integrated the targeted allele into the germ line will produce progeny that are heterozygous for the allele in all cells.

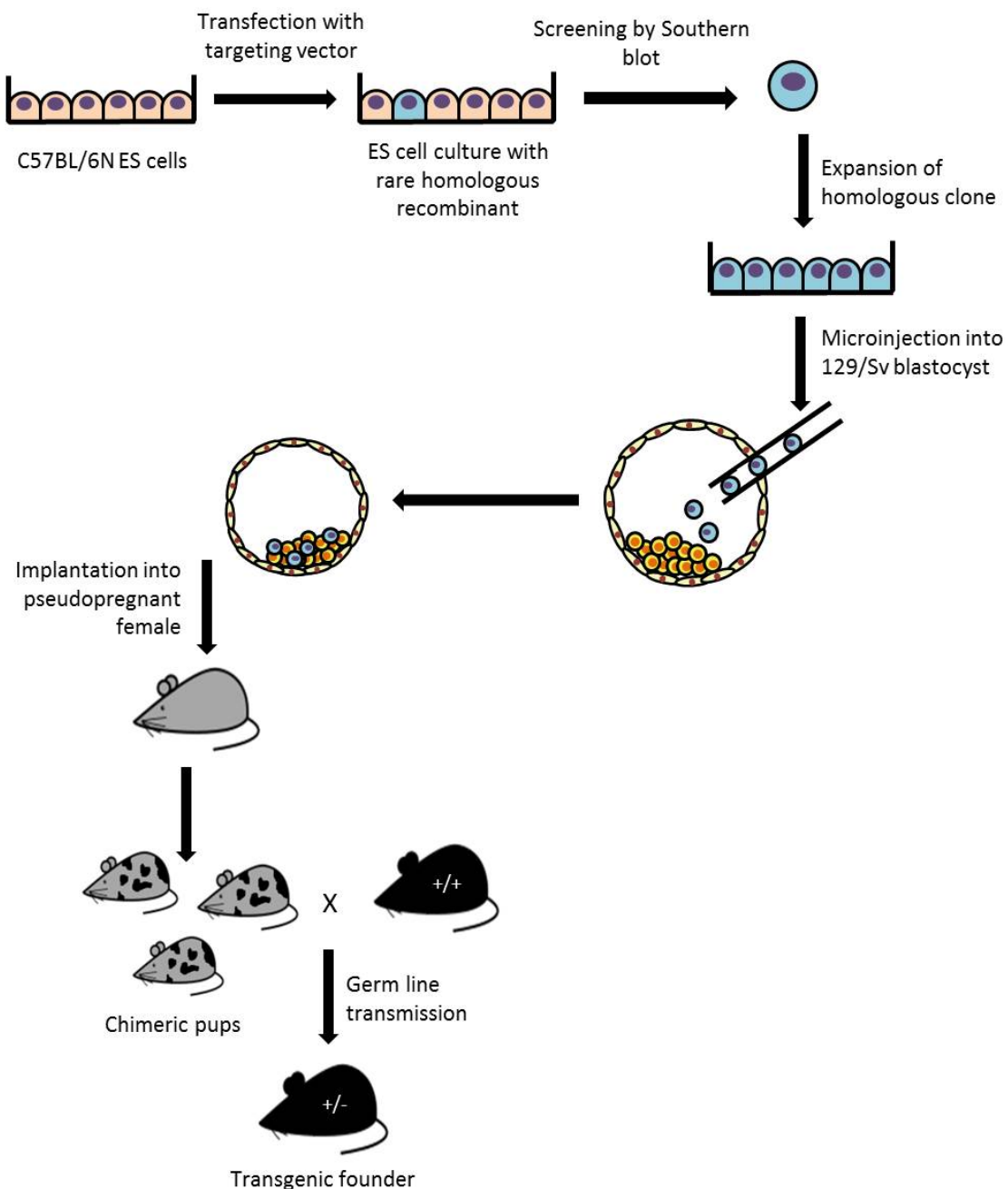


Figure 3.2 Generation of mice carrying a targeted mutation derived in ES cells

Transfection of the targeting vector into mouse ES cells results in many random integrations, and rare homologous recombinants (blue cells). These are screened for by Southern blot analysis and then expanded to produce a pure population of the selected clone. Targeted cells are microinjected into blastocysts, which are implanted into a foster mother. The resulting chimeric pups are mated with WT mice, and chimeras that have integrated the mutation into the germ line will produce progeny heterozygous for the mutated allele in all cells.

3.1.3. Aims

The aim of this chapter is to determine whether the human sex-reversing MAP3K1 amino acid substitution, L189R, causes sex-reversal in mice by introducing the causative mutation into the mouse *Map3k1* gene. This will involve generation of a targeted knock-in allele by homologous recombination in ES cells, followed by injection of targeted cells into mouse blastocysts. Following germ line transmission of the targeted allele, heterozygous and homozygous embryos bearing L189R will be bred and harvested at 14.5 dpc to analyse gonadal development.

3.1.4. The L189R construct

The L189R substitution is caused by a T to G mutation at position 566 of the cDNA sequence of human *MAP3K1* in exon 2, in a highly conserved region (Pearlman et al., 2010). In mouse, the equivalent leucine residue is at position 184 of the MAP3K1 polypeptide which corresponds to nucleotide position 551 of the cDNA. For simplicity, the construct will be referred to as L189R throughout this chapter. A PCR-based site-directed mutagenesis strategy for introduction of the single nucleotide variant encoding a Leu-Arg substitution at position 184 was designed by Dr Nick Warr. The mutated nucleotide is flanked 5' by 4.1 kb homologous *Map3K1* sequence from the targeted strain, C57BL/6N. 3' of the targeted mutation, an unconserved region within intron 2 was chosen for insertion of a *loxP-FRT-neo-FRT-loxP* cassette. This is then flanked on the 3' side by 3.9 kb homologous sequence (Fig 3.3). The L189R targeting construct was ligated into a *Tk-loxP-neo* plasmid vector and cloned into DH5- α cells by using a commercial supplier (GenScript).

Following electroporation of ES cells with the targeting construct (Section 2.6), identification of homologous recombinants was performed by Southern blotting. In this case, addition of the *Neo* selection cassette introduces additional *SpeI* and *BsaWI* restriction sites, meaning that electrophoresis of ES cell DNA would yield bands of differing size in WT versus targeted colonies (Fig 3.4).

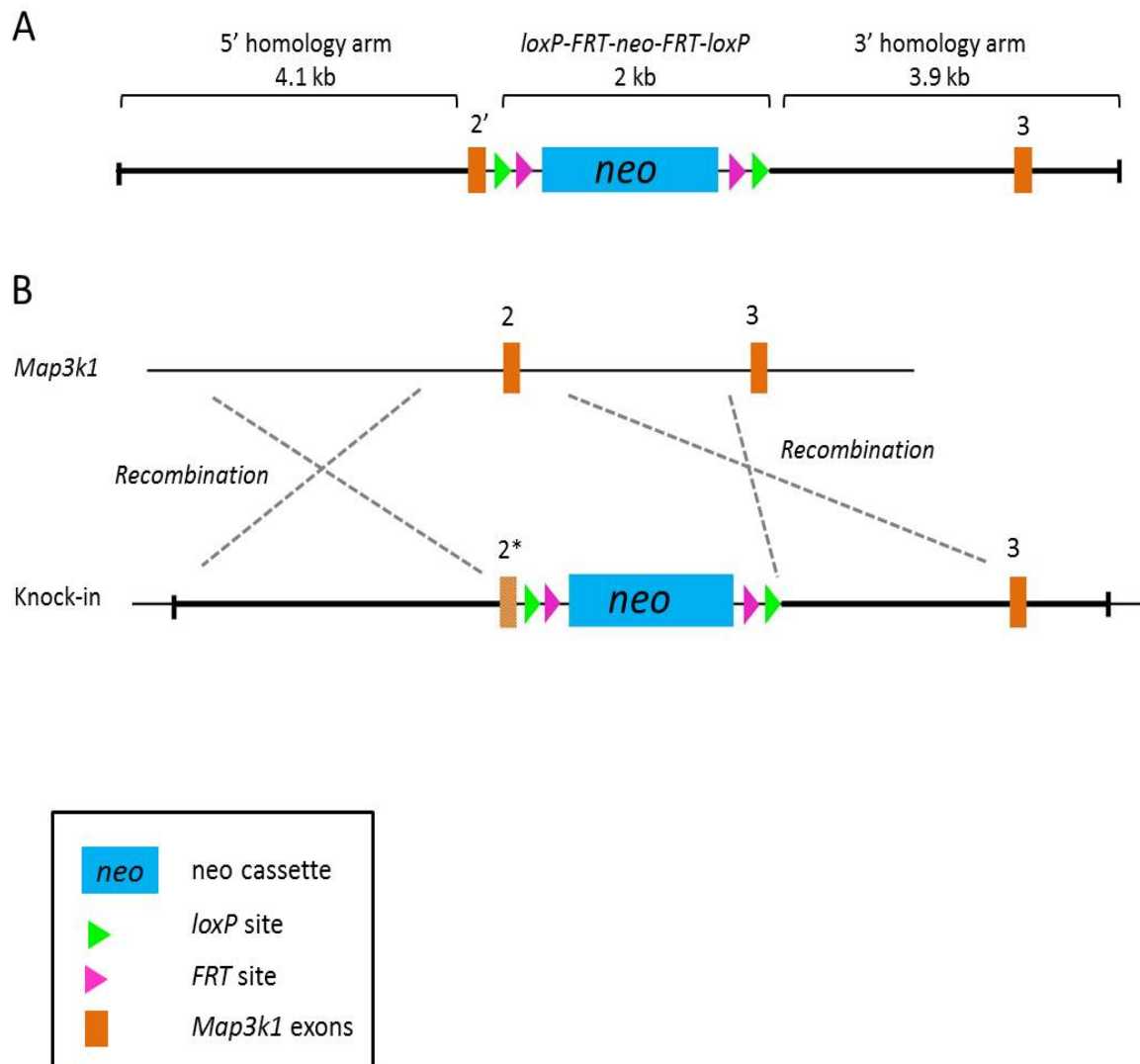


Figure 3.3 Design of the L189R targeting construct

(A) The L189R targeting construct consists of a 4.1 kb 5' arm of homology to part of intron 1 of *Map3k1*, followed by the engineered Exon 2 containing the T-G base change encoding a Leu-Arg substitution (2'). A *FRT-loxP-neo-FRT-loxP* cassette is inserted into intron 2 in a region of little conservation. The 3' arm of homology is 3.9 kb in length and contains exon 3 as well as some intronic sequence. (B) In ES cells, homologous recombination with the targeting vector results in replacement of WT sequence with the construct. Construct design was performed by Dr Nick Warr.

3.2. Results

3.2.1. Identification of targeted ES cell clones

Following visualisation of hybridised DNA fragments on X-ray film, homologous recombinants were identified by the presence of both the WT and recombinant bands. The L189R targeting construct generated 5 correctly targeted ES cell clones out of 144 that were screened (Fig 3.5a). Recombinants were confirmed by measuring the migrated distance of the WT and putative mutant band from the wells of the gel, which had been marked on the membrane after blotting. For each gel, the migrated distance of each band of the 1 kb ladder was plotted against the Log of its size (in kb) and a line of best fit calculated using GraphPad Prism. The linear equation of each graph was then used to estimate the size of the labelled mutant and WT bands (Fig 3.5b).

The stoichiometry of the fragments in the Southern blots (Fig 3.5 a) suggests that all the colonies apart from 5C were mixed clones. If the samples were clonal, the band intensities of the WT and targeted fragment should be highly similar due to each cell carrying 1 of each allele. However the upper, WT band is much more intense in all the samples except for 5C. This indicates an over-representation of WT alleles and therefore potentially mixed clones. Nonetheless, it was considered worthwhile to expand all the targeted clones, because mixed clones could later be re-plated to single colonies in an attempt to purify correctly targeted colonies.

The 5 correctly targeted clones were expanded; however, only clone 5C grew successfully. Restriction-digested DNA from this clone was screened using both the 3' and 5' probes (see Fig 3.3) to confirm correct targeting. Unfortunately, following this second screen, clone 5C was found to be WT for the targeted locus (Fig 3.5c).

A second transfection was performed, but attempts to extract gDNA from picked colonies were unsuccessful and, due to time constraints and considerations of priority, the project was terminated at this stage.

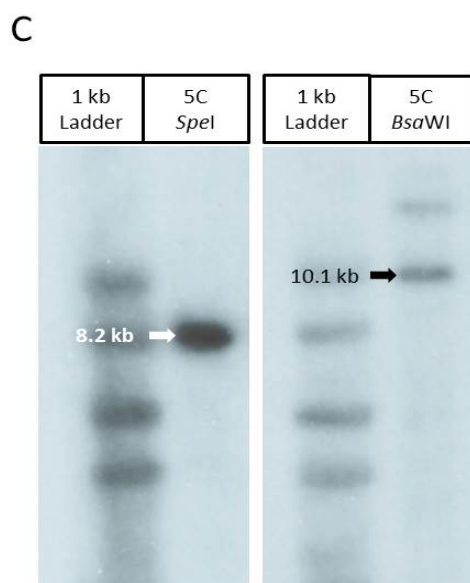
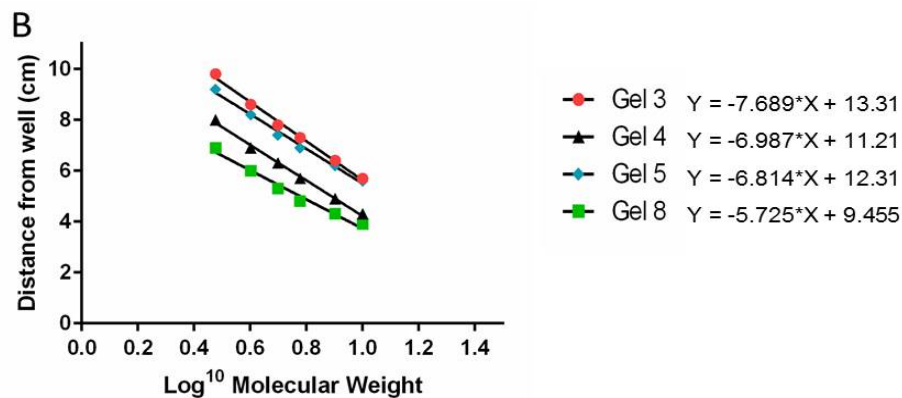
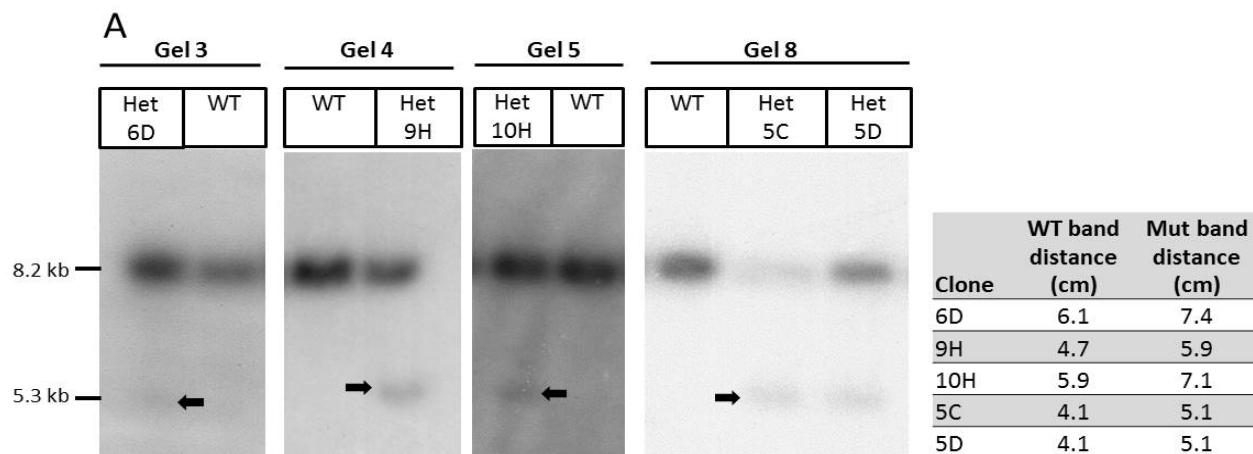


Figure 3.5 Screening ES cells for L189R homologous recombinants

For the initial 144-colony screen (A), DNA was extracted from ES cells seeded on 96-well plates and digested with *SpeI*. Fragments separated by gel electrophoresis were transferred to nitrocellulose by Southern blotting and hybridised with radiolabelled 5' probe. The screen yielded 5 putative homologous recombinants: 6D, 9H, 10H, 5C and 5D. For each gel, a fragment size reference was created by plotting migrated distance of the 1 kb ladder against Log^{10} size (kb) of each ladder band. The distance of the WT and mutant (arrows) bands from the wells was then measured and the linear equation of each gel's standard plot ($y = mx + c$) used to calculate band size. Clone 5C was expanded and subjected to secondary screening with both 5' and 3' probes (C). The clone was found to be WT.

3.3. Discussion

3.3.1. Summary

A screen of ES cells transfected with the L189R targeting vector yielded five correctly targeted clones. However, of these five, only one colony was successfully expanded and this clone was found to be WT on secondary screening. This may have been for several reasons. If a mixed colony was picked following transfection, cells that had incorporated the construct at other positions in the genome may have outcompeted the targeted cells and resulted in a loss of signal from them. Alternatively, whilst every care was taken to accurately label each lane of the 144 lanes loaded onto agarose gels for the screen, it cannot be ruled out that an incorrect identity was assigned to clone 5C, and therefore it was in fact a WT neighbour that was expanded for re-screening. Due to the amount of time taken to expand this one clone to sufficient density for the second screen, it was not feasible to repeat the process for neighbouring clones to check this.

3.3.2. *MAP3K1* mutations are a common cause of human DSD

Since beginning this study, further *MAP3K1* variants have been described in patients with 46,XY DSD, and *MAP3K1* mutations are now accepted as a common cause of human DSD (Loke et al., 2014, Baxter et al., 2015). In particular, exome sequencing of 40 46,XY DSD patients identified four cases with *MAP3K1* mutations, two of which had a previously reported variant, Gly616Arg (Baxter et al., 2015). A further two *de novo* variants were identified in this study, Arg339Gln and Pro257Leu. Interestingly, these patients did not all have CGD; some had DSD with complex ambiguous genitalia and there was no clear correlation between

phenotypic severity and the position of the mutation in the *MAP3K1* gene. This suggests that the role of MAP3K1 is a complex one or that it may be modified by genetic background. Nevertheless, its emergence as an important regulator of human testis determination is undeniable.

3.3.3. Future prospects

Given the large body of evidence supporting a role for *MAP3K1* in human sex determination, it is arguable that modelling disease-causing mutations in the mouse is still worthwhile. Not only might this facilitate further molecular understanding of the basis of human DSD, it would also contribute significantly to elucidating the molecular basis of sex determination in the mouse, and in mammals more generally.

3.3.3.1. CRISPR-Cas: a revolutionary tool for mouse genetics

The advent of the clustered regulatory interspaced short palindromic repeats (CRISPR)-Cas9 genome editing system makes the prospect of producing the L189R mouse more feasible. This highly efficient technique has revolutionised genetics since its use was first reported by Jinek and colleagues in 2012 (Jinek et al., 2012) and has since been successfully employed to introduce targeted mutations into the mouse genome in a single step, abrogating the need for ES cell targeting and complicated targeting construct design (Menke, 2013). CRISPR-Cas systems are based upon the nuclease, Cas9, from *Streptococcus pyogenes*. This nuclease relies upon RNA to direct it to the desired DNA sequence and therefore can be targeted to specific sequences by following RNA-DNA base-pairing rules. Cas9 introduces a double strand break (DSB) at the target sequence, which stimulates DNA repair.

Repair mediated via non-homologous end joining (NHEJ) can result in insertions and deletions, interrupting the gene of interest. For the purpose of introducing specific mutations, the DSB should be repaired via homology-directed repair (HDR), which uses the homologous DNA strand as a template. By saturating the system with template DNA that contains a mutation of interest, HDR can be employed to repair the DSB and simultaneously introduce a targeted mutation in the gene of interest.

Using this novel, efficient method of genome editing, it will be possible to generate mice bearing disease-causing *Map3k1* mutations more quickly than by conventional targeting methods, and could facilitate relatively fast analysis of gonadal phenotypes in L189R animals. As such, it may be possible to also analyse gonadal phenotypes of mice bearing other *Map3k1* mutations and enable tandem analysis and comparison of different *Map3k1* variants *in vivo*.

3.3.4. Future experiments

If the L189R mouse line were to be generated, two possible scenarios may ensue. Firstly, the animals may have a gonadal phenotype. One may predict that, if the L189R amino acid substitution is causative of 46,XY CGD in humans, we may observe testicular defects in mice. This may range from complete XY sex reversal to the presence of some ovarian tissue in XY gonads at 14.5 dpc, identified by WMISH staining for a marker of an ovarian cell-type, such as the meiotic germ cell marker *Stra8*. If a testicular defect is observed, genetic complementation experiments could be performed to assess whether MAP3K1 acts in a pathway with other known testis-determining kinases such as MAP3K4, or whether it is involved in an alternative pathway involving other testis-determining genes. Secondly, the L189R

mouse line may have no gonadal phenotype, or have a phenotype reminiscent of *Map3k1* KO mice (Warr et al., 2011). This would indicate that the human and mouse testis determining pathways have diverged, that MAP kinases have differing functions in the two species or simply that the amino-acid substitution is tolerated in the context of the mouse polypeptide.

If the L189R mouse did exhibit a gonadal phenotype, biochemical assays could help to ascertain the molecular events underlying any observed defect. In particular, analysis of the effect of L189R upon phosphorylation of downstream kinases such as JNK, ERK and p38 MAPK, may indicate whether previously reported observations in lymphoblastoid cells (Pearlman et al., 2010) are recapitulated in embryonic mouse gonads. If the presence of L189R does, as suggested, lead to elevated p-p38 and p-ERK, the amino acid substitution may indeed confer gain of MAP3K1 function. This would suggest that an excess of MAP3K1 may inhibit testis determination. Whether this is due to elevated p38 MAPK signalling is unclear, yet intriguing, since it may indicate another example of dose-dependent effects on sex determination. It would also be important to determine whether there is some level of cross-talk between MAP3K1- and MAP3K4-mediated signalling during sex determination underlying any phenotyping.

Furthermore, ascertaining whether MAP3K1 is involved in testis determination upstream or downstream of *Sry* expression would be necessary if a phenotype were observed. Cell-based reports of mutant MAP3K1 interacting with RHOA, which may inhibit SOX9, are of interest (Loke et al., 2014). However, without a relevant, tractable model in which to study any potential role of MAP3K1 in gonad development, it is unhelpful to speculate.

3.3.5. Final remarks

In summary, an attempt was made to generate mice bearing a human sex-reversing amino acid substitution in mouse MAP3K1, by targeted mutagenesis in ES cells. This was unsuccessful and therefore the project was abandoned in favour of more promising studies. However, given the importance of MAP3K1 in human CGD, it may be beneficial to our understanding of these disorders to generate mice bearing these mutations in the future.

Chapter 4. Identifying transcriptional targets of GADD45 γ -mediated signalling in the developing mouse gonad by RNA-seq

4.1. Introduction

4.1.1. *Gadd45g* in the mouse gonad

Gadd45g encodes the small nuclear protein GADD45 γ , a member of the Growth Arrest and DNA Damage-inducible protein family, which also includes GADD45 α and GADD45 β . These proteins were first identified in yeast two-hybrid screens for proteins that interact with MAP3K4 (Takekawa and Saito, 1998) and have since been shown to activate MAP3K4 via disruption of its autoinhibitory domain, thereby facilitating MAP3K4 autophosphorylation (Miyake et al., 2007). Specifically, GADD45 γ has also been shown to activate the p38 MAPK and JNK pathways in T cells (Lu et al., 2001).

In the mouse embryo, *Gadd45g* is expressed in a variety of tissues including the brain, heart and gonad at 10.5 dpc (Kaufmann et al., 2011). It was recently demonstrated that *Gadd45g* is expressed in the somatic cells of the gonad with a distinct spatiotemporal profile reminiscent of *Sry*, but preceding *Sry* onset by approximately 6 hours (Warr et al., 2012). Loss- and gain-of function experiments provided evidence that *Gadd45g* is crucial for mouse testis determination on the sensitised C57BL/6J background. XY mice lacking *Gadd45g* undergo complete gonadal sex reversal that is characterised by severely delayed expression of *Sry* (Gierl et al., 2012, Warr et al., 2012). *Gadd45g*^{+/-}/*Map3k4*^{+/-} double heterozygotes exhibit more severe disruption to testis determination at 14.5 dpc than do single heterozygotes of either gene. This demonstrates that *Gadd45g* and *Map3k4* interact genetically (Warr et al., 2012). In the somatic cells of the gonad, loss of GADD45 γ or MAP3K4 is associated with reduced phosphorylation of p38 MAPK and GATA4, and it has been proposed that a GADD45 γ -MAP3K4-p38 MAPK signalling

module facilitates timely expression of *Sry*, potentially via association of phosphorylated GATA4 with the *Sry* promoter (Gierl et al., 2012, Warr et al., 2012) (Fig 4.1). However, it is still unclear whether the output of this pathway is a direct, cell-autonomous effect at the *Sry* promoter, or whether GADD45 γ has additional non-MAP3K4-dependent functions in testis determination. One possibility is the up-regulation, at other loci, of transcripts that act as intermediaries or co-activators of *Sry* expression. We also cannot exclude the possibility that GADD45 γ -mediated signalling affects transcription of other genes in gonadal somatic cells that may play other roles in testis development, in a non-*Sry* dependent manner. Identification of such transcripts in mouse testis determination would add depth to our knowledge of the control of *Sry* expression and provide further insight into role of GADD45 γ in development of the mouse gonad.

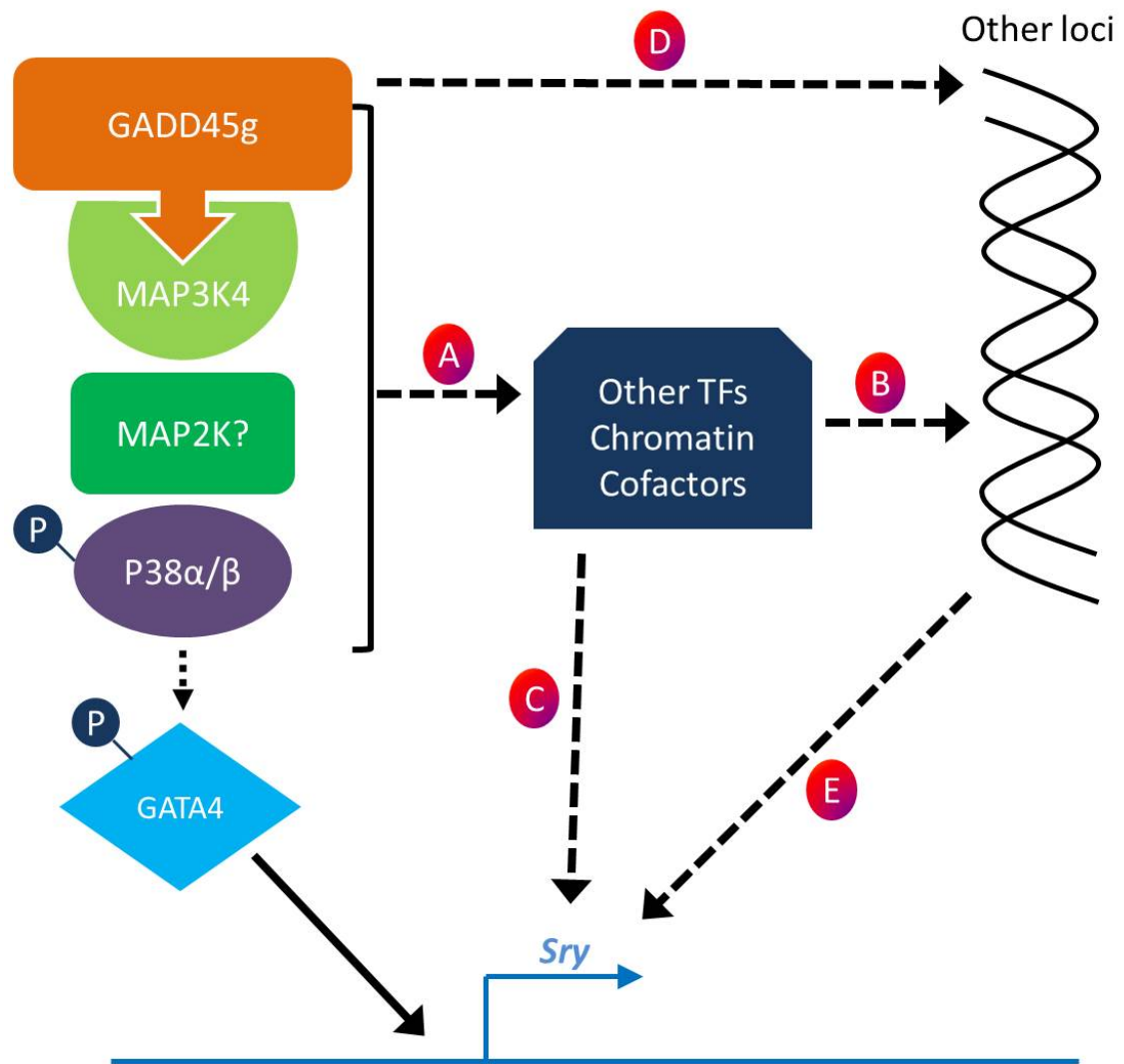


Figure 4.1 Potential roles for GADD45 γ in testis determination

In the somatic cells of the gonad, GADD45 γ activates MAP3K4, which activates p38 α/β via a MAP2K(s) that is yet to be identified. This results, directly or indirectly, in phosphorylation of GATA4 (dotted arrow), and in turn leads to early expression of *Sry*, potentially via association of p-GATA4 with the *Sry* promoter (Gierl *et al* 2012). The GADD45 γ -MAP3K4-p38 module may also activate other transcription factors, cofactors, or facilitate modification of chromatin (A). This may, in turn, activate transcription at other as yet uncharacterised loci (B), or at *Sry* itself (C). In addition, GADD45 γ may also act to promote transcription at other loci via non-MAP3K4-dependent mechanisms (D). Furthermore, products of transcription at other uncharacterised loci may regulate expression of *Sry* in *trans* (E). This figure is adapted from Warr *et al* (2012).

The aim of this chapter is to identify transcripts whose expression is disrupted by the loss of *Gadd45g* and, thereby, potentially novel genes involved in regulating testis or ovary determination. Comparing the transcriptomes of WT gonads with those lacking a key determinant of testis determination, such as *Gadd45g*, may enable identification of such transcripts. Moreover, since GADD45 γ is implicated in the regulation of *Sry* timing, identification of its transcriptional targets in the gonad may also shed light upon how GADD45 γ facilitates timely *Sry* expression.

4.1.2. RNA-Sequencing

The development of Next-Generation sequencing (NGS) technology has revolutionised how we think about gene expression. It has paved the way for the many “Seq” applications that now exist, including ChIP-Seq, DNase-Seq, methyl-Seq and RNA-Seq (Park, 2009, Song and Crawford, 2010, Li and Tollefsbol, 2011, Wang et al., 2009). Several large-scale studies have utilised these technologies to reveal the dynamic and complex nature of gene regulation, including the encyclopaedia of mouse DNA elements (Mouse ENCODE) project (Mouse et al., 2012) (<http://www.mouseencode.org/>). This comprehensive, worldwide collaboration has generated a wealth of data about the regulation of mouse genome and epigenome.

Of the NGS technology applications, RNA-Seq is perhaps the most complex. It enables visualisation of differential transcript expression, differential splicing, allele-specific expression and the detection of novel transcripts - information that it was not previously possible to obtain using hybridisation-based approaches such as microarrays (Ozsolak and Milos, 2011). As such, it has the potential to provide an

abundance of information about the transcriptional landscape of the developing mouse gonad.

RNA-seq comprises several steps, beginning with the isolation of total RNA from the tissue of interest. Since total RNA consists of up to 95% ribosomal RNA (Lindberg and Lundeberg, 2010), this must first be depleted to enrich for Poly-A and non-Poly-A mRNA. Following rRNA depletion, a double-stranded cDNA library is prepared from fragmented RNA. In order to preserve the directionality of the subsequent transcript information, cDNA library synthesis can be performed using strand-specific protocols such as the dUTP protocol described in Fig 4.2a. This method utilises uracil incorporation during reverse transcription, followed by ligation of end-specific adaptors to enable differentiation between RNAs originating from the forward or reverse strand. Following library preparation, next-generation sequencing is performed by loading the library onto a flow cell that immobilises each fragment by hybridisation (Fig 4.2b). Fluorescently-labelled nucleotides are then incorporated as each fragment is amplified, and the emission wavelength and intensity of each incorporated base enables it to be recorded (Fig 4.2b). The cycle is repeated to produce sequence reads, which are then aligned to a reference genome (Illumina, 2015; www.illumina.com/technology/next-generation-sequencing.html).

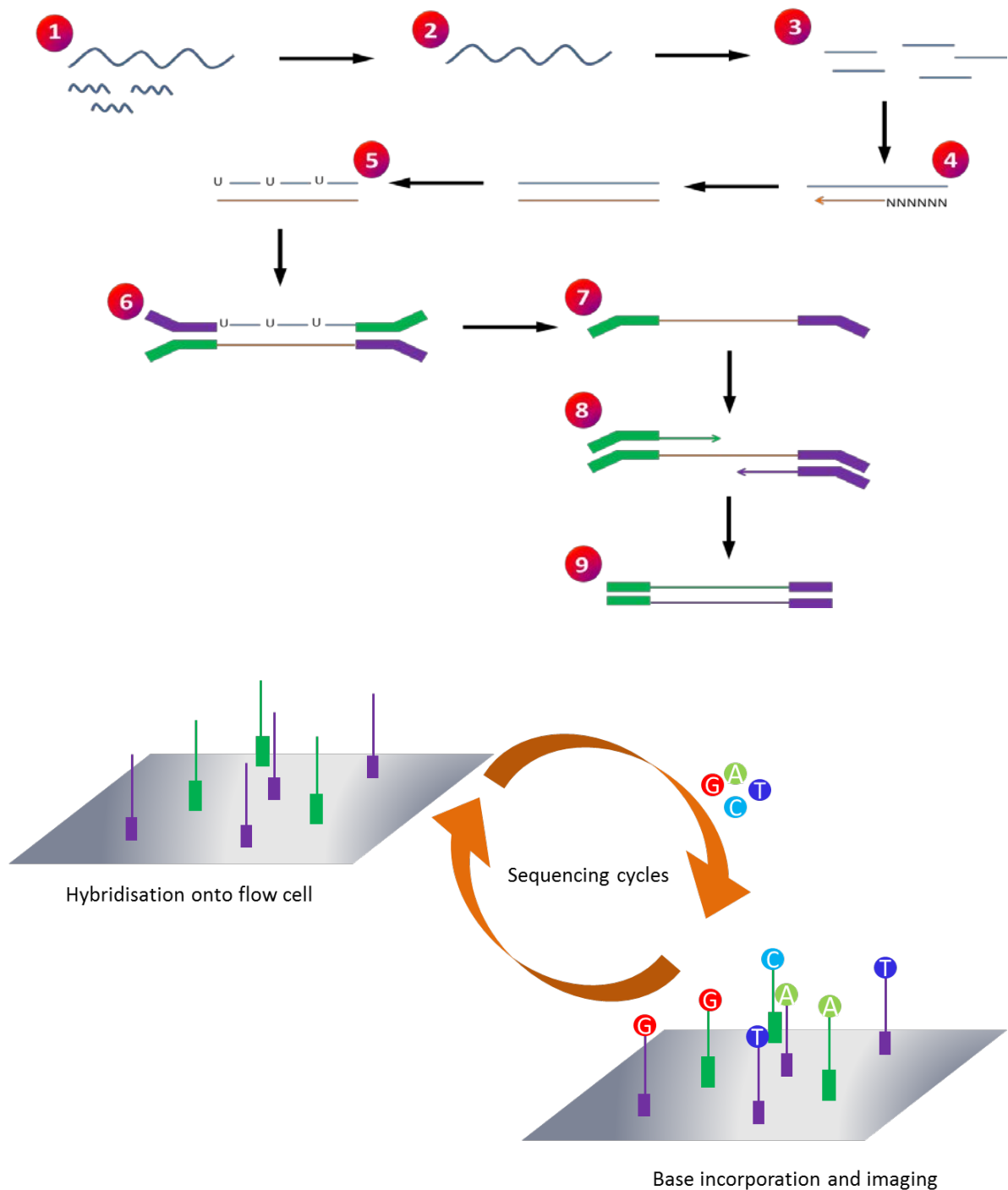


Figure 4.2 cDNA library preparation and sequencing for RNA-seq

(A) Total RNA is isolated from cells or tissues (1). PolyA+ RNA is enriched by depletion of rRNA (2) then fragmented (3). First-strand reverse transcription is then performed using random primers (4), followed by second-strand synthesis using dUTP instead of dTTP (5). The cDNA fragments are then ligated to y-shaped adaptors (6) and the dUTP strand is then degraded (7). cDNA is then PCR-amplified (8). Finally, a dsDNA library is created and can be used for next-generation sequencing (9). (B) The cDNA library is loaded onto flow cells that contain a lawn of oligonucleotides complementary to the adaptors. The cDNA fragments hybridise and are immobilised on the flow cell surface. Each fragment is amplified into a clonal cluster (not depicted) and sequenced by incorporation of fluorescently-labelled dNTPs.

Previous RNA-seq experiments in the context of mouse sexual development have focused on postnatal ovaries and testes, or on germ cells (Gong et al., 2013, Margolin et al., 2014). Although numerous microarray-based studies have already made significant contributions to our understanding of the transcriptional networks underlying sex determination (Klattig et al., 2007, Coveney et al., 2008, Munger et al., 2009, Jameson et al., 2012b, Munger et al., 2013), RNA-seq-based transcriptomes of gonads at the crucial time of sex determination are not yet available. By exploiting the *Gadd45g* knockout line, we can gain a large amount of information about the transcriptional events taking place around the crucial time of testis determination. It is, therefore, crucially important to select an appropriate age range at which to perform RNA-seq on embryonic gonads. Tissue that is expressing *Gadd45g* but has not yet reached the peak of *Sry* expression would be an ideal choice, because this would facilitate identification of transcripts that may be involved in regulating *Sry* itself. It would also limit the detection of sexually dimorphic genes that are to be expected as the gonad begins to differentiate into testis or ovary, thereby reducing confounding differential expression that arises as a consequence of *Sry* expression. *Gadd45g* is detectable from 9ts onwards (Warr et al., 2012), and *Sry* peaks around 17ts in B6 (Bullejos and Koopman, 2001, Bullejos and Koopman, 2005), therefore tissue should be selected that is somewhere between these two stages. To enable the maximum number of *Gadd45g*-expressing cells to be obtained, and to allow relatively straightforward dissection, a window of 15-16ts (approximately 11.25 dpc) was chosen for tissue collection for this experiment. Based on previous comparisons of XX and XY transcriptomes at these early stages (Jameson et al., 2012b, Munger et al., 2013), it is predicted that there will not be a large number of differentially expressed genes.

4.2. Results

At the time of the experiment, RNA requirements for sequencing were a minimum of 1µg total RNA, with 260/280nm ratio greater than 1.8 and 260/230nm ratio greater than 1.5. In order to obtain the required quantity of RNA it was necessary to pool gonads and mesonephroi from 15-20 stage-matched embryos. Since *Gadd45g* is only expressed in the gonad, it was not expected that including mesonephric tissue would affect the detection of differentially expressed transcripts, although some dilution of signal might occur.

2-5µg RNA was purified for each of four experimental groups: XX *Gadd45g*^{+/+}, XY *Gadd45g*^{+/+}, XX *Gadd45g*^{-/-}, XY *Gadd45g*^{-/-}. Table 4.1 shows the quantity and quality of each sample.

Table 4.1 RNA-seq sample quality and quantity

Sample name	Quantity (µg)	260/280nm ratio	260/230nm ratio
XX ^{+/+}	2.2	2.08	0.91
XY ^{+/+}	5.7	2.09	0.96
XX ^{-/-}	2.4	2.12	0.60
XY ^{-/-}	1.4	2.05	0.34

Samples were sent to Oxford Gene Technology, where additional quality control was performed. cDNA library preparation, sequencing and data analysis were performed by OGT and raw data were also passed on to the MRC Harwell Bioinformatics core for analysis. In both cases, sequence reads were aligned to the reference genome using the read alignment tool TopHat (Section 2.8.7). A report of differential expression was provided for validation in the lab.

4.2.1. Proof-of-principle

To first confirm that analysis of the RNA-seq data could identify expected changes in expression, sequence read alignments for *Gadd45g* were analysed (Fig 4.3). No reads aligned with the coding region of *Gadd45g* in both XX *Gadd45g*^{-/-} and XY *Gadd45g*^{-/-} samples, indicating that the full-length transcript is not present in mutant gonads and confirming that RNA-seq could successfully identify differential expression between the knockout and control samples. Large numbers of reads did, however, align to the untranslated region of *Gadd45g*. This has been observed in several knockout lines, but has not yet been explained (M. Simon, personal communication).

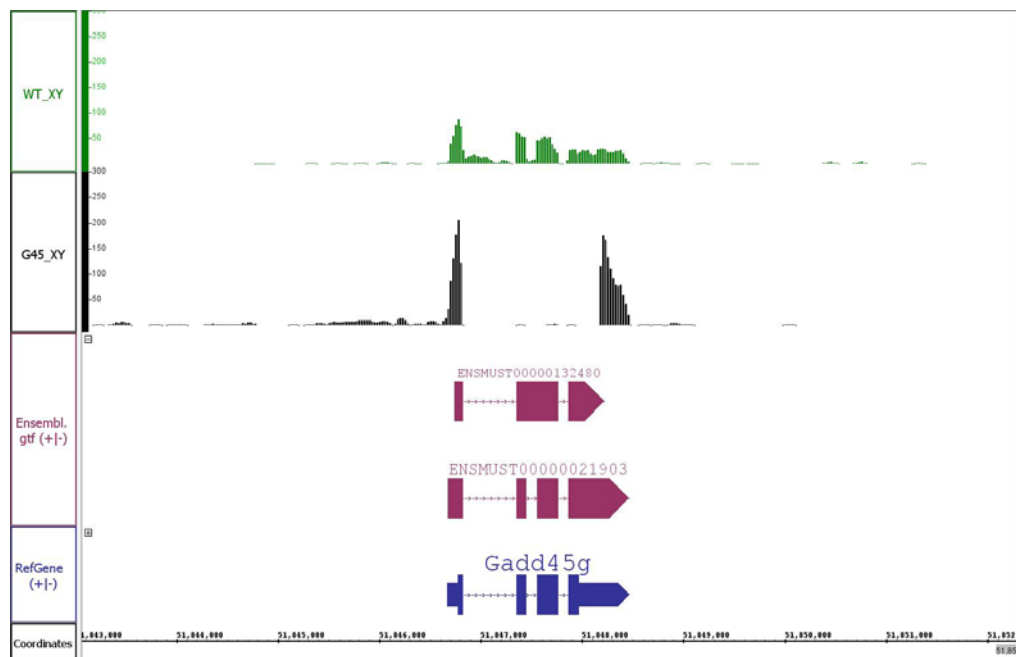


Fig 4.3 RNA-seq proof-of-principle

Snapshot from the TopHat read alignment showing the distribution of sequence reads across the *Gadd45g* locus. Alignment of sequence reads to the reference genome in XY WT (green, top) and XY *Gadd45g*^{-/-} (black, bottom) samples is shown. The protein coding region is annotated in purple (Ensembl) and blue (Refgene). Density of sequence reads mapped to the region is represented by the green or black peaks. Reads can be seen aligned to the *Gadd45g* coding region in WT, but not in *Gadd45g*^{-/-}.

To further confirm that expected changes in gene expression could be detected with the analysis used, expression in WT XX and XY samples was compared (Table 4.2). This revealed significantly differential expression of several sex-linked genes, including *Xist*, the major effector of the X-inactivation process (Lyon, 1996).

Table 4.2 Differentially expressed genes in XX vs XY WT samples

Gene name	Log2 fold change (XX ^{+/+} /XY ^{+/+})	<i>p</i> value	<i>q</i> value
<i>Hist1h2br</i>	-7.30	0.00005	0.038
<i>Alox5ap</i>	-3.20	0.00005	0.038
<i>Xist</i>	-4.70	0.00005	0.038
<i>Ube1y1</i>	infinity	0.00005	0.038
<i>Eif2s3y</i>	7.22	0.00005	0.038
<i>Uty</i>	3.40	0.00005	0.038

4.2.2. Differential expression: XY *Gadd45g*^{+/+} vs XY *Gadd45g*^{-/-}

Transcript abundance, calculated using TopHat, was used to test for differential expression using four separate tools: DESeq, Cufflinks, EgeR and NOISeq (Section 2.8.7). A list of the top 44 differentially expressed transcripts was produced by the Harwell Bioinformatics Core team (Siddharth Sethi, and Dr Michelle Simon), based on a log2 fold-change of greater than 1.2 and *p* value less than 0.005. These are shown in Table 4.3. For simplicity, WT and *Gadd45g*^{-/-} samples will be referred to as XX/XY^{+/+} and XX/XY^{-/-} respectively, for the remainder of this chapter.

Table 4.3 Top 44 differentially expressed genes ranked by p-value

Gene name	Up-regulated or down-regulated	Log2 fold change (XY ^{+/+} /XY ^{-/-})	p value
<i>Snora74a</i>	up	2.54	6.5 x 10 ⁻⁸
<i>Gm26316</i>	up	2.50	2.2 x 10 ⁻⁷
<i>Gm25395</i>	up	2.37	4.3 x 10 ⁻⁷
<i>Gm24407</i>	up	2.29	4.4 x 10 ⁻⁷
<i>Gm22973</i>	up	2.27	4.9 x 10 ⁻⁷
<i>Gm25360</i>	up	2.18	9.7 x 10 ⁻⁷
<i>Tgm3</i>	down	-2.30	7.0 x 10 ⁻⁵
<i>Tfcp2l1</i>	down	-1.80	9.3 x 10 ⁻⁵
<i>Pou3f3</i>	down	-2.10	0.0001
<i>Gja5</i>	down	-1.94	0.0001
<i>Ttn</i>	up	1.90	0.0001
<i>Pnliprp2</i>	down	-2.13	0.0001
<i>Wnt9b</i>	down	-1.81	0.0001
<i>Cpa1</i>	down	-2.02	0.0002
<i>Cldn6</i>	down	-1.71	0.0002
<i>Prl7a1</i>	up	5.63	0.0002
<i>Scarna6</i>	up	1.70	0.0003
<i>Pkhd1l1</i>	up	1.62	0.0004
<i>Col4a2</i>	down	-1.49	0.0006
<i>Kel</i>	up	1.64	0.0007
<i>Col4a1</i>	down	-1.44	0.0009
<i>Itgb6</i>	down	-2.28	0.0012
<i>Ank1</i>	up	1.50	0.0013
<i>Chrm4</i>	down	-1.58	0.0013
<i>Cdh16</i>	down	-1.45	0.0015
<i>Slc4a1</i>	up	1.39	0.0015
<i>Sptbn2</i>	down	-1.47	0.0019
<i>Reln</i>	up	1.41	0.0022
<i>Lipg</i>	down	-1.49	0.0023
<i>Ripk4</i>	down	-1.91	0.0024
<i>Lama1</i>	down	-1.33	0.0024
<i>Hmgcs2</i>	down	-1.76	0.0027
<i>Lamc2</i>	down	-1.40	0.0029
<i>Fam129a</i>	down	-1.51	0.0030
<i>Rhag</i>	up	1.45	0.0034
<i>Sel1l3</i>	down	-1.37	0.0039
<i>Rhd</i>	up	1.33	0.0039
<i>Actc1</i>	down	-1.57	0.0042
<i>Vwf</i>	down	-1.29	0.0044
<i>Cd93</i>	down	-1.25	0.0045
<i>St14</i>	down	-1.50	0.0046
<i>Sgpp2</i>	down	-1.51	0.0046
<i>Frem2</i>	down	-1.25	0.0047
<i>Lrp2</i>	down	-1.26	0.0049

28 transcripts were down-regulated whereas 16 were up-regulated. Interestingly, the most significantly differentially expressed transcripts were up-regulated in XY^{-/-} tissue, indicating that they may be pro-ovarian genes that are normally repressed in XY gonads at this stage. In the search for potential pro-testis genes, the down-regulated transcripts were initially taken forward for further analysis. The top 10 most significantly down-regulated expressed transcripts were taken forward for further analysis (Table 4.4).

Table 4.4 Top 10 downregulated transcripts in XY^{-/-} gonads

Gene	Protein name	Location
<i>Tgm3</i>	transglutaminase 3, E polypeptide	Chr2:130012349-130050399
<i>Tfcp2l1</i>	transcription factor CP2-like 1	Chr1:118627945-118685168
<i>Pou3f3</i>	POU domain, class 3, transcription factor 3	Chr1:42695768-42703176
<i>Gja5</i>	gap junction protein, alpha 5	Chr3:97032416-97053634
<i>Pnliprp2</i>	pancreatic lipase-related protein 2	Chr19:58759723-58777533
<i>Wnt9b</i>	wingless-type MMTV integration site family, member 9B	Chr11:103727364-103749821
<i>Cpa1</i>	carboxypeptidase A1, pancreatic	Chr6:30639218-30645363
<i>Cldn6</i>	claudin 6	Chr17:23679365-23682446
<i>Col4a2</i>	collagen, type IV, alpha 2	Chr8:11312805-11449287
<i>Col4a1</i>	collagen, type IV, alpha 1	Chr8:11198423-11312826

4.2.3. Further analysis of downregulated transcripts

Expression of the downregulated transcripts from Table 4.4 as assayed by qRT-PCR using subdissected gonads at 14-16ts. One gene from the top 10, *Cpa1* was unfortunately not included in this analysis, due to time constraints in designing an appropriate qRT-PCR assay.

Primers for qRT-PCR SYBR® Green assays were designed using Primer3 (Section 2.8.6) and the efficiency of novel assays was tested on WT cDNA before progressing to the experiment (Table 2.10). qRT-PCR was performed on cDNA from 1-5 pairs of subdissected gonads of each genotype.

Five transcripts were significantly differentially expressed in subdissected gonads: *Tgm3*, *Pnliprp2*, *Gja5*, *Cldn6* and *Col4a1* (Fig 4.4a - g). *Col4a2*, *Pou3f3*, *Tfcp2l1* and *Wnt9b* were not significantly differentially expressed (Fig 4.4h - k).

Tgm3 and *Pnliprp2* were significantly downregulated in *XY^{-/-}* gonads (Fig 4.4a-c). Both these genes encode enzymes: *Tgm3* encodes transglutaminase 3 (also known as epidermal transglutaminase or Tgase) and *Pnliprp2* encodes PLRP2 (pancreatic lipase-related protein-2). Neither of these genes have previously been described in connection with mouse gonad development, but transglutaminase has been detected in homogenate from adult rat Leydig cells (Dias, 1985) and has been implicated in hormone processing by Sertoli cells (Dias, 1986).

Gja5 was also significantly downregulated in *XY^{-/-}* gonads (Fig 4.4d). *Gja5* encodes connexin 40, a gap junction protein that is important for development of conductivity of the heart (Delmar and Makita, 2012) and has not previously been described in connection with gonadal development. *Gja5* was not sexually

dimorphically expressed in gonads at the stage examined, and was not differentially expressed in $XX^{-/-}$ gonads compared with $XX^{+/+}$, suggesting potentially XY-specific regulation of this gene at this stage.

Cldn6 showed a conflicting, yet significant trend in differential expression and was significantly upregulated in $XY^{-/-}$ gonads (Fig 4.4e), opposite to the trend indicated by RNA-seq analysis (table 4.4). *Cldn6* encodes the transmembrane protein Claudin 6, which is found in tight junctions (Furuse and Moriwaki, 2009) and has recently been identified as a marker of pluripotency in mouse cells (Wang et al., 2012). Intriguingly, expression of *Cldn6* in $XY^{-/-}$ gonads was highly reminiscent of expression in WT XX gonads, and expression in WT XY gonads very similar to $XX^{-/-}$. This suggests that it may act in a pathway that both promotes testis determination and also inhibits ovarian differentiation.

Col4a1 was also significantly upregulated in $XY^{-/-}$ gonads (Fig 4.4f, g). This is the opposite trend to that identified by analysis of RNA-seq data. Both *Col4a1* and *Col4a2* have been implicated in the maintenance of testis cord integrity at 14.5dpc and are downregulated in embryonic testes lacking *Wt1* (Chen et al., 2013a). Their importance in earlier gonadogenesis is so far unknown. However, *Col4a2* was not significantly differentially expressed in subdissected gonads (Fig 4.4h). There was a trend towards upregulation of *Col4a1* in $XX^{-/-}$ gonads, however this was not significant.

In addition, there was also a trend towards upregulation of *Wnt9b* in $XY^{-/-}$ gonads (Fig 4.4k). Although this is the opposite to the trend suggested by analysis of RNA-seq data, it suggests that *Wnt9b* may be a pro-ovarian gene. *Wnt4* is a well-known

ovary-determining gene and the involvement of other WNT molecules cannot be discounted.

As a control, expression of *Gadd45g* was also analysed in the same samples used for expression analysis of the putative DEGS (Fig 4.4I). *Gadd45g* was significantly downregulated in $XY^{-/-}$ samples, confirming that this analysis could recapitulate expected changes in gene expression.

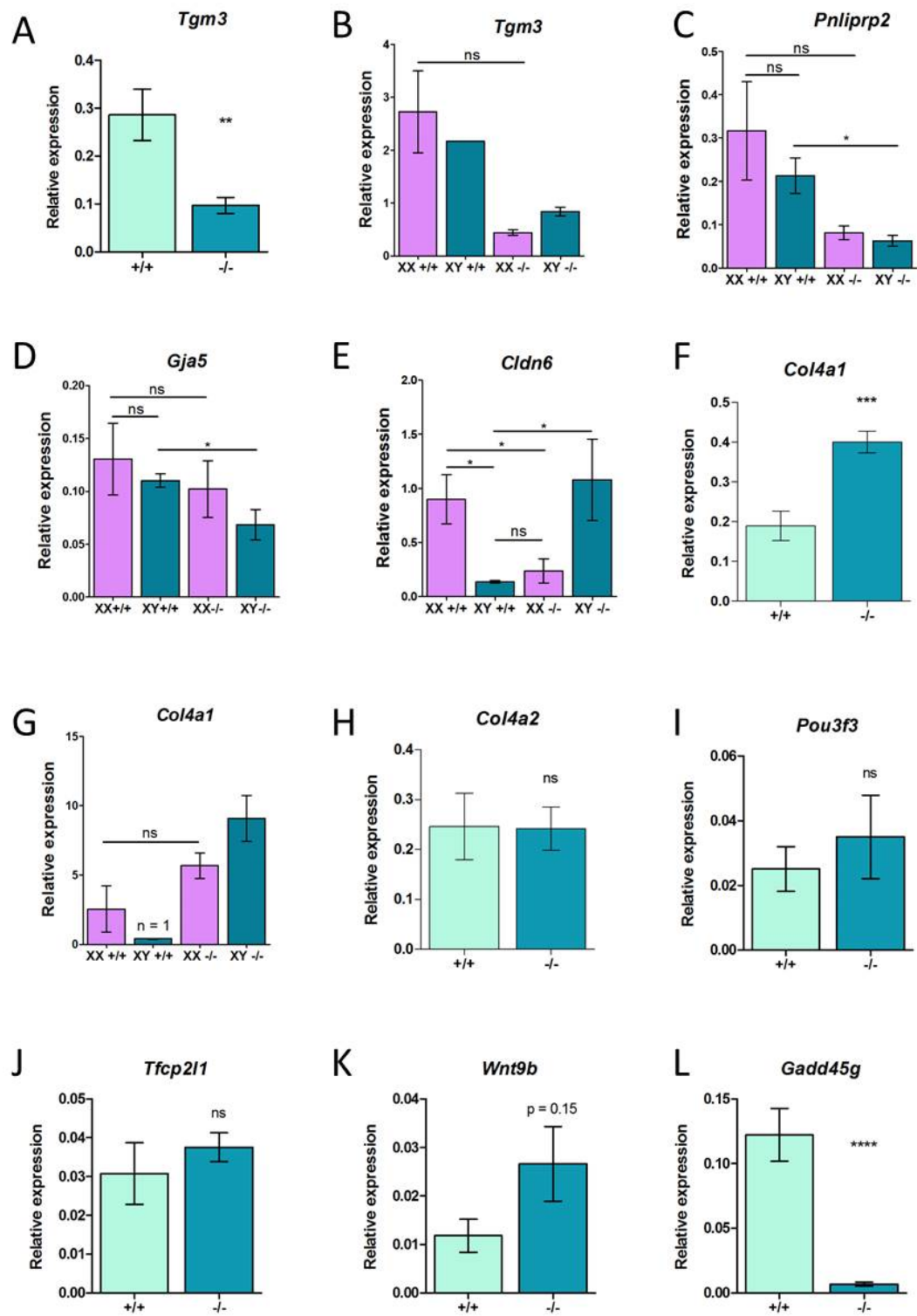


Figure 4.4 qRT-PCR analysis of top 9 downregulated genes from RNA-seq analysis
For legend please see page 115

Figure 4.4 qRT-PCR analysis of top 9 downregulated genes from RNA-seq analysis

Tgm3 (A, B) and *Pnliprp2* (C) were significantly downregulated in XY^{-/-} gonads compared with XY^{+/+} ($p = 0.006$ and 0.016 respectively). Both *Tgm3* and *Pnliprp2* also showed a trend towards downregulation in XX^{-/-} gonads compared with XX^{+/+}, although this was not significant ($p = 0.206$ and 0.107 respectively). *Gja5* (D) was significantly downregulated in XY^{-/-} gonads ($p = 0.024$) but not XX^{-/-} ($p = 0.529$). *Cldn6* (E) and *Col4a1* (F, G) were significantly upregulated in XY^{-/-} gonads ($p = 0.046$ and 0.0004 respectively). *Cldn6* expression was significantly decreased in XX^{-/-} gonads compared with XX^{+/+} ($p = 0.047$) and in addition, expression was significantly lower in XY^{+/+} than XX^{+/+} ($p = 0.028$). There was a trend towards upregulation of *Col4a1* in XX^{-/-} gonads compared with XX^{+/+}, but this was not significant ($p = 0.224$). *Col4a2* (H), *Pou3f3* (I) and *Tfcp2l1* (J) were not significantly differentially expressed in XY^{-/-} gonads ($p = 0.959$, 0.495 and 0.437 respectively). *Wnt9b* (K) showed a trend towards upregulation in XY^{-/-} gonads but this was not significant ($p = 0.157$). Control qRT-PCR of *Gadd45g* (L) expression confirmed downregulation in XY^{-/-} samples ($p = 0.0001$). Bars, mean RQ relative to *Hprt1* endogenous control. Error bars, SEM. Statistical analysis, unpaired t test.

4.2.4. Further analysis of *Pnliprp2*

Having identified five significantly differentially expressed genes in XY^{-/-} gonads by qRT-PCR, it was also important to analyse their expression by WMISH. This would provide information about the spatial expression of each gene and, potentially, cell-type specificity. Unfortunately, as time became short, it was only possible to design and synthesise a WMISH probe for *Pnliprp2*.

Pnliprp2 and the protein it encodes, PLRP2, are commonly associated with lipid digestion. Mice lacking PLRP2 are deficient in dietary fat absorption as neonates and gain weight more slowly than WT littermates (Lowe et al 1998). No gonadal phenotype has so far been described. However, that it was one of the most significantly downregulated genes in XY^{-/-} gonads was intriguing and warranted further investigation.

In order to determine where *Pnliprp2* is expressed, WMISH was performed on 15ts UGRs from XY^{+/+} and XY^{-/-} embryos (Fig 4.5). In XY^{+/+} gonads, *Pnliprp2* expression had a highly punctate pattern and appeared to be located specifically in the gonad. The location of expression was strikingly reminiscent of that of *Sry* but was very weak at this stage. In XY^{-/-} gonads, expression was much weaker.

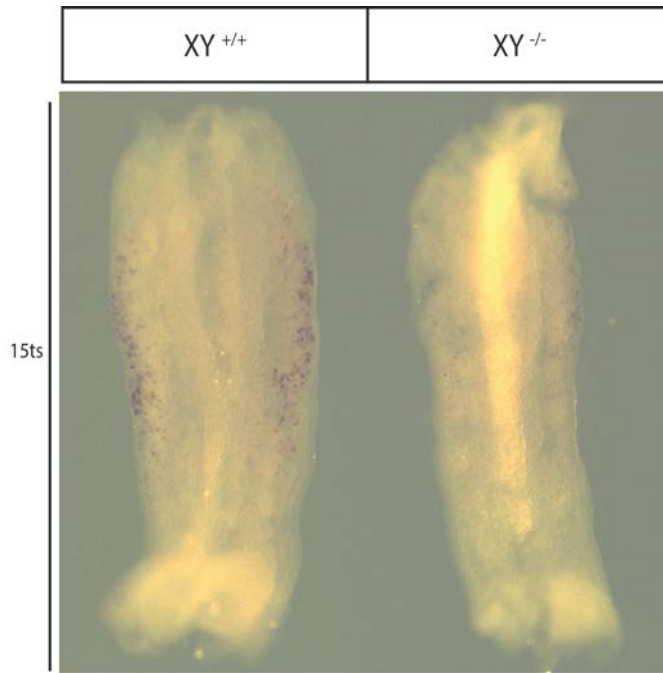


Figure 4.5 Analysis of *Pnliprp2* expression by WMISH

WMISH was performed on 5 XY WT and 5 XY *Gadd45g*^{-/-} UGRs, stage-matched at 15ts. *Pnlip2p2* is specifically expressed in the gonad and is greatly reduced in XY^{-/-}.

As other projects took priority and time became short, it was not possible to investigate the expression profile or function of *Pnliprp2* any further. However, proposals for future experiments are detailed in the Discussion.

4.3. Discussion

Comparison of the transcriptomes of 11.5dpc XY^{+/+} and XY^{-/-} gonads by RNA-seq has identified a minimal number of differentially expressed transcripts. Of these, three are significantly downregulated and two significantly upregulated following further characterisation by qRT-PCR. A further six upregulated transcripts identified by RNA-seq remain to be characterised (Table 4.4).

The most significantly down-regulated gene, *Tgm3*, is expressed in seminiferous tubules at 12.5dpc (Zhang et al 2005) but so far has not been described in earlier stages of gonad development. The gene product, transglutaminase 3, is correlated with differentiation of epidermal tissues (Zhang et al 2005) and therefore may play

a role in directing the epithelial fate of the supporting cells of the testis. Its downregulation in $XY^{-/-}$ gonads supports this hypothesis. However, it is also downregulated in $XX^{-/-}$ gonads. This suggests that it may not function in a male-specific manner, but in differentiation more generally, but given that gonad development is not affected in $XX^{-/-}$ mice, this seems unlikely. It would be interesting to use WMISH to examine the location of *Tgm3* expression in 11.5dpc WT gonads of both sexes, and also to analyse expression of transglutaminase protein by immunofluorescence to observe its cellular localisation.

Also down-regulated in $XY^{-/-}$ gonads was *Pnliprp2*. This gene has no previously described role in differentiation or cell fate decisions, and is more commonly linked with neonatal lipid digestion (Lowe et al., 1998, D'Agostino and Lowe, 2004). However, its strikingly genotype-specific expression pattern at 15ts warrants further investigation. *Pnliprp2* was also downregulated in $XX^{-/-}$ gonads, suggesting that it, like *Tgm3*, may have a role that is not XY-specific. *Gadd45g* itself is not expressed in a sexually dimorphic manner, however $XX^{-/-}$ mice are normal, fertile females, suggesting that *Gadd45g* is dispensable for female gonad development (Gierl et al., 2012, Warr et al., 2012). Of course, if the only target of GADD45g-mediated signalling is *Sry* expression, then *Pnliprp2*, like *Gadd45g*, might have a male-specific role despite its expression in XX gonads.

It would be very interesting to further characterise the expression of both *Pnliprp2* and *Tgm3*. In particular, it would be interesting to identify the gonadal cell population in which they are expressed. WMISH has provided clues that *Pnliprp2* may be specifically expressed in the gonad (Fig 4.5), but it is not clear which cell

population it is expressed in. FACs sorting of gonads expressing an *Sf1-EGFP* reporter (McClelland et al., 2015) followed by qRT-PCR could help identify whether *Pnliprp2* and *Tgm3* are enriched in somatic or germ cells. Additionally, analysis of gonad sections would help identify whether they are expressed within the gonad proper, the coelomic epithelium or in mesenchyme. Dual-labelling with antibodies marking gonadal somatic cells, such as GATA4, and germ cells, such as OCT4 (Warr et al., 2012) would help determine the cell type within which the protein products PLRP2 and transglutaminase 3 are present, as well as their cellular localisation in the context of the gonad. If these genes are expressed at later stages, once testis cords have formed, identification of cell-types exhibiting expression becomes simpler. Ultimately, loss-of-function experiments that ablate these genes in the developing gonad, perhaps conditionally in the appropriate cell-type, would reveal any function in sex determination. *Tgm3* knockout mice have recently been described and exhibit impaired hair development and disrupted epidermal barrier integrity (John et al., 2012, Bognar et al., 2014), but so far no abnormalities in sexual development have been described.

Cldn6 was significantly upregulated in XY^{-/-} gonads with a fold change of 2.99 (Fig 4.4). *Cldn6* is a member of the claudin family, which comprises some 24 proteins in mice (Lal-Nag and Morin, 2009). The claudins are a group of small, transmembrane proteins that are important components of tight junctions and have wide-ranging roles in development. Claudin 6 has been implicated in very early morphogenesis and blastocyst formation, and has been described as a novel marker of mouse pluripotent stem cells (Wang et al., 2012). This is intriguing because it suggests that

its upregulation as a result of loss of *Gadd45g* could be associated with a tendency towards pluripotency and a delay in differentiation. This supports the idea that *Gadd45g* may be important in driving differentiation. A recent publication describes the induction of differentiation by GADD45 γ , via MAP3K4, in haematopoietic stem cells (Thalheimer et al., 2014), so it is plausible that it could have a similar role in the gonad. Claudin 11 has been implicated in maintenance of the epithelial fate of mouse Sertoli cells via the maintenance of cell-cell adhesions (Mazaud-Guittot et al., 2010), but it is not yet known whether this or other claudins are important earlier in testis development.

Whilst *Tgm3* and *Pnliprp2* were downregulated in both XY^{-/-} and XX^{-/-} gonads, *Cldn6* did not show any change in expression in XX^{-/-} gonads compared to WT, suggesting that it is regulated in an XY-specific manner. Again, it would be interesting to analyse expression of *Cldn6* by WMISH and immunostaining to ascertain the cell types in which it is present.

Col4a1 was also significantly upregulated in XY^{-/-} gonads. Along with *Col4a2*, *Col4a1* has been implicated in WT1-mediated maintenance of testis cord integrity in 15.5 dpc mouse embryos (Chen et al., 2013a). Its role in earlier gonad development has not yet been described.

What was particularly striking when analysing the putative DEGS arising from RNA-seq data, was the low number that were significantly differentially expressed when assayed by qRT-PCR in subdissected gonads. There could be several reasons for this. Firstly, the parameters set for selection of the top 44 genes were not very stringent. This is because, in an initial analysis provided by OGT there were no

significantly differentially expressed genes. In this analysis, a Q value threshold of 0.05 was used to identify differentially expressed genes (J. Cooper, OGT, personal communication). A q value is a p value that is adjusted to take into account false detection rates of an experiment, making it a more stringent statistical indicator to use for identifying true-positive results (Arava et al., 2003). Therefore, in order to identify any differentially expressed transcripts at all, it was necessary for the Harwell Bioinformatics group to adjust their parameters to only take p value into account. As a result of this, there is a greater chance of detecting false-positives.

Secondly, due to the mixed population of cells used in this analysis, it is likely that some dilution of signal may have occurred, making differences in expression more difficult to detect. At the time of the experiment, RNA-seq required relatively large amounts of RNA and due to the size of the 11.5 dpc gonad, it was necessary to include mesonephros in the samples. Since then, advances have been made enabling RNA-seq to be performed on minute amounts of RNA and even single cells. In addition, the acquisition of a FACS cell sorter at Harwell has introduced the possibility of generating RNA from samples of purified gonadal somatic cells, using fluorescent reporter lines. This is likely to enable more powerful analysis of the transcriptomes of cells in the developing gonad and potentially identify novel targets of GADD45 γ .

Another intriguing observation was that some genes were differentially expressed in both RNA-seq and qRT-PCR analyses, but with opposite trends. For example, *Cldn6* and *Col4a1* were both significantly downregulated in XY *Gadd45g*^{-/-} samples compared to WT, but were significantly up-regulated in XY *Gadd45*^{-/-} subdissected

gonads analysed by qRT-PCR. This has been previously reported in other studies analysing concordance between RNA-seq and microarray data and has been shown to be related to low transcript abundance (Guo et al., 2013, Wang et al., 2014) but may also be related to high homology between closely-related transcripts, for example *Col4a1* and *Col4a2* (M. Simon, personal communication). Such discrepancies highlight the importance of independent verification of potentially differentially expressed genes.

Interestingly, the genes that were differentially expressed with the highest fold change were all up-regulated in the XY *Gadd45g*^{-/-} sample. One gene in particular, *Prl7a1*, had a strikingly large fold change in expression. *Prl7a1* encodes a member of the prolactin family of protein hormones, also known as prolactin-related protein E (PRLPE). In the mouse, the prolactin family consists of over 20 members, all clustered within a 1 Mb region of chromosome 13 (Muller et al., 1998). *Prl7a1* is expressed in placenta (Lin et al., 1997), but information on its function or expression in other tissues is lacking. That there is such a large fold change in expression indicates that *Prl7a1* is highly expressed in the 11.5 dpc gonad or mesonephros. However, further examination of *Prl7a1* expression is required to ascertain whether it is significantly up-regulated in XY *Gadd45g*^{-/-} subdissected gonads.

The large number of up-regulated genes in XY *Gadd45g*^{-/-} gonads (Table 4.4) suggests that GADD45γ-mediated signalling may play a role in inhibiting the ovarian pathway even at this early stage. It has been suggested that, as early as 10.5 dpc, gonadal somatic cells are “primed” for an ovarian fate, meaning that a robust

program of repression is required for commitment to the male pathway (Jameson et al., 2012b). Since *Gadd45g* expression precedes *Sry* expression by some six hours (Warr et al., 2012), it is plausible that it could play a role in inhibiting an early program of pro-ovarian gene expression. However, since the stage examined in this study was one during which *Sry* is already being expressed, it is impossible to determine whether these putative pro-ovarian genes are being repressed in an *Sry*-dependent or *Sry*-independent manner.

GADD45 family proteins have been implicated in the regulation of gene expression by active DNA demethylation at specific sites, and have already been shown to be important for regulating gene expression during development (Ma et al., 2009, Sen et al., 2010, Zhang et al., 2011). Whilst a link between GADD45 γ , DNA demethylation and *Sry* transcription has not yet been made, it is possible that GADD45 may regulate transcription at other loci by facilitating demethylation at target sites. It is currently unclear how GADD45 γ 's proposed function in DNA methylation is related to MAP3K4 activation. The two may be distinct processes that independently lead to transcription of target genes, or they may indeed be linked. Transcriptome wide analysis of *Map3k4*^{-/-} gonads may facilitate identification of potential targets of a common pathway, by comparing transcriptomes to identify genes that are differentially expressed in both *Gadd45g*^{-/-} and *Map3k4*^{-/-} gonads.

Chapter 5. The role of *Map3k4* in B6.Y^{POS} sex reversal

5.1. Introduction

5.1.1. Genetic background affects sensitivity to XY sex reversal

The interplay between the dominant male determinant, *Sry*, and its genomic context, plays a key role in driving the bias of gene expression in the gonad towards the male fate. It has been well established that differences exist in strain-specific susceptibility to perturbations in testis-determining factors, and that these can be exploited to enable us to understand the complex dichotomy between the testis- and ovary-determining pathways (Eicher and Washburn, 2001, Munger et al., 2009, Wilhelm et al., 2009, Warr et al., 2014, Nikolova et al., 2008). The B6 background is particularly sensitive to perturbation of many autosomal testis-determining genes and as such has become the strain of choice for identification of important testis-determining loci. Whilst disruption of some testis determining genes, including *Cbx2* and *Map3k4*, have minimal effect on testis determination on a mixed background, testis determination is increasingly affected with successive backcrosses to B6 (Bogani et al., 2009, Katoh-Fukui et al., 2012). Other inbred strains such as DBA and 129/SvImJ are resistant to sex reversal induced by disruption of such genes (Eicher and Washburn, 2001, Munger et al., 2009, Nikolova et al., 2008).

A microarray comparison of gene expression, between B6 and 129/SvImJ gonads at 11.5 dpc, has shown that many ovary-determining genes, including *Fst* and *Wnt4*, are more highly expressed in 11.5 dpc B6 gonads (Munger et al., 2009). More recent temporal transcriptional profiling of gonadal somatic cells has revealed that they are primed with a slight pro-ovarian bias from an early stage (Jameson et al., 2012b), and that activation of pro-testis pathways in B6 cells is delayed relative to 129 (Munger et

al., 2013). Together, these reports suggest that the effects of disruption to testis-determining genes may have a more drastic effect on B6 than other strains.

It is increasingly evident that *Sry*^{B6} is expressed within a particularly short, early window (Hiramatsu et al., 2009). This early onset of expression may have evolved as a strategy to circumvent precocious B6 ovarian genes. As such, it follows that even subtle perturbations in the expression of *Sry* in the B6 background may be sufficient to preclude upregulation of *Sox9* and therefore enable the ovarian pathway to proceed in XY gonads. This may be achieved via 2 means - either by disrupting genes required for *Sry* expression, or by introducing an alternative *Sry* allele that is not expressed in the characteristic B6 manner.

5.1.2. B6.Y^{DOM}: a tool for studying *Sry* regulation

A well-studied phenomenon in mouse genetics is that of B6.Y^{DOM} sex reversal. Originally described in the 1980s, B6.Y^{DOM} sex reversal refers broadly to disrupted testis development in B6 mice which possess the Y chromosome from *Mus musculus domesticus* strains (Y^{DOM}) in place of the indigenous B6 Y chromosome (Y^{B6}). Depending upon the strain of origin of the Y chromosome, XY^{DOM} mice exhibit varying degrees of XY gonadal sex reversal. *M. domesticus* Y chromosomes are placed into three categories depending upon the severity of sex reversal that they confer (Biddle and Nishioka, 1988, Biddle et al., 1991, Nagamine et al., 1999). The first class, which includes the Y chromosome from the FVB/N strain (Y^{FVB}), do not cause sex reversal. A second class, including Y chromosomes from the AKR/J strain (Y^{AKR}) and RFJ strain (Y^{RFJ}), cause transient sex reversal, characterised by the development of ovarian tissue in embryos but the presence of normal testes at birth (Nagamine et al., 1987c,

Nagamine et al., 1987a). The third, most severe group, is associated with sex reversal, including the development of completely sex-reversed XY females. This group consists of the Y chromosomes from wild-derived strains captured in Val Poschiavo, Switzerland (Y^{POS}), and Tirano, Italy (Y^{TIR}). Of these, Y^{POS} is perhaps the best studied. Around 75% of B6. Y^{POS} mice develop as females, with the remaining 25% as hermaphrodites (Eicher 1982).

Other strains that have been subjected to introgression of Y^{POS} , such as DBA/2J and 129/svImJ (129), are resistant to sex reversal, further highlighting the sensitivity of the B6 strain to disruption of the testis-determining pathway (Eicher et al., 1996, Eicher and Washburn, 2001). Most attempts to explain B6. Y^{POS} sex reversal have focused upon *Sry*, since it resides upon the Y chromosome, which is the only genetic difference between B6. Y^{POS} and B6. Y^{B6} . Intriguingly, it has been reported that the expression of *Sry* in B6. Y^{POS} and B6. Y^{TIR} embryos is delayed relative to B6. Y^{B6} (Albrecht et al., 2003, Bullejos and Koopman, 2005, Nagamine et al., 1999). This suggests that altered *Sry* expression may be a key cause of sex reversal in these animals, and supports the hypothesis that the window of opportunity for *Sry* to initiate testis development is particularly short in B6 compared with other strains.

It has also been suggested that SRY protein isoform differences play a role in B6. Y^{DOM} sex reversal. Whilst the HMG domains are highly conserved between strains, part of the C-terminal of SRY ^{DOM} is truncated relative to the *M. musculus* isoform. A substantial part of this truncation lies within the region known as the polyQ domain, which contains multiple glutamine residues. In SRY ^{DOM} this consists of a substantially shorter region of repeats, and the number of repeats varies somewhat between

different *M. domesticus*-derived strains (Coward et al., 1994). It has been suggested that this may contribute to B6.Y^{DOM} sex reversal, however, it cannot completely explain it, since no correlation has so far been found between the number of glutamine repeats and the severity of sex reversal (Carlisle et al., 1996, Albrecht and Eicher, 1997). Furthermore, a transgene containing coding sequence from *M. poschiavinus* *Sry*, situated within flanking sequence from the 129 strain, is able to cause male development of female embryos, suggesting that there is no inherent defect in the protein (Albrecht et al., 2003). It is therefore likely that B6.Y^{DOM} sex reversal is the result of a combination of altered *Sry* expression and differing properties of the *M. domesticus* and *M. musculus* protein isoforms.

Delayed *Sry* expression in B6.Y^{POS} is suggestive of misregulation, perhaps caused by incompatibility between the *Sry*^{POS} allele and autosomal or X-linked modifiers of *Sry*. Since many genes that regulate *Sry* expression have now been identified, it is plausible that these may have the potential to modify B6.Y^{POS} sex reversal. Employing a loss- and gain-of-function approach may facilitate further understanding of the importance of B6.Y^{POS} sex reversal, as well as improving our understanding of how *Sry* is regulated.

5.1.3. *Map3k4* and *T*-associated sex reversal (*Tas*)

Another example of B6.Y^{DOM} sex reversal is *T*-associated sex reversal (*Tas*). This was identified when mice carrying the hairpin-tail (*T^{hp}*) mutation on proximal chromosome 17 were crossed onto the B6 background and XY embryos carrying *T^{hp}* were found to possess ovarian tissue (Washburn and Eicher, 1989). This dominant mutation is caused by a deletion of approximately 6 Mb and heterozygotes exhibit a characteristic hairpin

tail phenotype. The T^{hp} mutation can only be transmitted via males, because the deletion also contains the imprinted, dominant lethal *Tme* (*T*, maternal effect) locus (Cattanach, 1986, Johnson, 1974). Because of this, and because the mutation arose on the inbred AKR/J strain, all $T^{hp/+}$ males also carry Y^{AKR} (Johnson, 1974). This Y chromosome causes transient ovotestis development in B6 embryos (Nagamine et al., 1987a) (Table 5.1). However, when the T^{hp} deletion is also present, all XY embryos develop ovaries or ovotestes (Washburn and Eicher, 1983). Thus, the presence of the T^{hp} deletion on the B6. Y^{AKR} background exacerbates a mild form of B6. Y^{DOM} sex reversal. Importantly, expression of *Sry* is delayed in $T^{hp/+}$ B6. Y^{AKR} embryos (Warr et al., 2014).

Several explanations for *Tas* have been proposed. It has been postulated that the disruption to testis determination may be due to generally delayed development, because $T^{hp/+}$ embryos are delayed compared to WT littermates (Cattanach 1987). However, it was also suggested that the underlying cause may be haploinsufficiency for a gene, situated within the deleted region, that is required for primary sex determination. One candidate for this was *Map3k4*. *Map3k4* resides on chromosome 17, within the region that is deleted in T^{hp} , and is required for timely *Sry* expression (Warr et al., 2012). The Greenfield lab recently reported that a functional *Map3k4* BAC transgene is able to rescue *Tas*, providing strong evidence that it is indeed haploinsufficiency for *Map3k4* that underlies the phenomenon (Warr et al., 2014). The BAC conferred an approximate 4-fold increase in expression of *Map3k4* in 11.5 dpc gonads (Warr et al., 2014) and in XY $T^{hp/+}$ transgenic embryos, overexpression of *Map3k4* was associated with enhanced expression of *Sry*.

Given that overexpression of *Map3k4* rescued *Tas*, the question arose as to whether similar overexpression could rescue other cases of B6.Y^{DOM} sex reversal. Being one of the most severe cases, and also reported to be associated with delayed *Sry*, B6.Y^{POS} was chosen for another *Map3k4* gain-of-function study. The *Map3k4* BAC was therefore crossed onto the B6.Y^{POS} background and embryos were harvested at 14.5 dpc by Dr Nick Warr and analysed by WMISH (Fig 5.1).

All B6.Y^{POS} *Map3k4* BAC embryos had bilateral testes that expressed *Sox9* (fig 5.1e), indicating that overexpression of *Map3k4* was able to rescue B6.Y^{POS} sex reversal. Interestingly, however, some B6.Y^{POS} *Map3k4* BAC gonads expressed small amounts of the ovary-specific germ cell marker *Stra8* at the poles (Fig 5.1i), indicating that the rescue is not complete and that testis determination may be somewhat delayed despite *Map3k4* overexpression. When B6.Y^{POS} *Map3k4* BAC mice were aged to sexual maturity, they were found to be fertile males, suggesting that the development of ovarian tissue was transient (N. Warr, unpublished data).

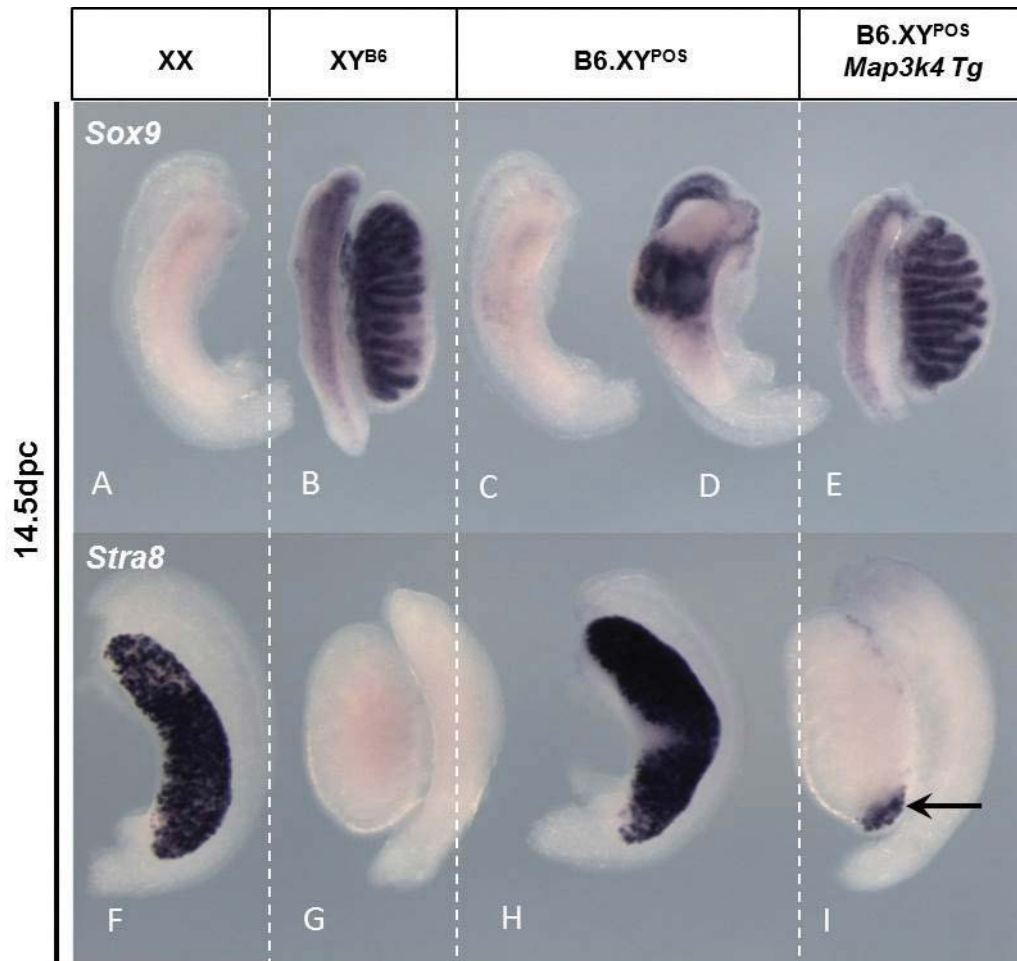


Figure 5.1 Overexpression of *Map3k4* rescues B6.Y^{POS} sex reversal

WMISH was performed on 14.5 dpc gonads from the strains indicated above the panels. The top panel (A-E) shows expression of the Sertoli cell marker *Sox9* and the bottom panel (F-I) shows expression of the marker of ovarian meiotic germ cells, *Stra8*. In B6, XY gonads (B, G) express *Sox9* in the testis cords and do not express *Stra8*, whereas XX gonads show the opposite pattern (A, F). B6.Y^{POS} embryos develop either ovaries (C) or ovotestes (D, H). Note restriction of *Sox9* expression to the central region of the gonad (row of asterisks in D) and abundant *Stra8* signal in H. B6.Y^{POS} embryos expressing the *Map3k4* transgene develop testes of near-normal appearance, showing appropriate expression of *Sox9*. In some gonads a region of *Stra8*-expressing cells was observed at the caudal pole (arrow in I). Data courtesy of Dr Nick Warr.

5.1.4. Mouse strain nomenclature

To aid clarity for the reader, the following terms will be used when referring to the mouse strains described in the remainder this thesis:

- B6.Y^{B6} shall be referred to as B6
- B6.Y^{POS} mice transgenically expressing the *Map3k4* BAC shall be denoted B6.Y^{POS} Tg
- 129/SvImJ shall be referred to as 129.

5.1.5. Aims

The aims of this chapter are to further investigate the role of *Map3k4* in B6.Y^{POS} sex reversal. This will be achieved by the following approaches:

- Analysis of adult B6.Y^{POS} Tg gonads
- Analysis of expression of *Sry* and *Sox9* in B6.Y^{POS} Tg gonads
- Further modification of *Map3k4* dosage in B6.Y^{POS} embryos

5.2. Results

5.2.1. Analysis of adult B6.Y^{POS} Tg mice

Following the observation that overexpression of *Map3k4* is able to restore testis determination in B6.Y^{POS} embryos at 14.5 dpc and leads to the development of fertile adult males, it was necessary to further examine the mice to confirm that they developed into adults with normal morphology.

Mice were sexed at weaning by analysis of the anogenital distance by Lee Kent or Jackie Harrison (Mary Lyon Centre, MRC Harwell). Ear tissue from each animal was then genotyped by PCR to determine chromosomal sex. All B6 animals were phenotypic males, whereas 35% B6.Y^{POS} animals were scored as male and 65% as female. In contrast, all B6.Y^{POS} *Map3k4* Tg animals were scored as male (Fig 5.2s).

Analysis of the gross anatomy and histological morphology of B6.Y^{POS} *Map3k4* Tg reproductive organs was then performed (Fig 5.2). Adult animals were sacrificed between 16 and 20 weeks of age. Reproductive organs were removed by dissection and photographed, then gonads were fixed in 10% formalin and processed for sectioning by the MRC Harwell Histology team (Adele Austin, Jenny Corrigan, Caroline Barker).

As has previously been described, B6.Y^{POS} animals exhibited a range of gonadal phenotypes. Some possessed testes that were small, yet normal in appearance (Fig 5.2e). Analysis of H&E sections of these testes revealed normal testicular architecture (Fig 5.2l). However, others ranged from having unilateral testicular agenesis (Fig 5.2f) to bilateral ovaries (Fig 5.2.i, j, p, q). Ovaries of B6.Y^{POS} sex-reversed females frequently contained cysts (Fig 5.2j, q).

In contrast, all B6.Y^{POS}-*Map3k4* Tg mice examined (n = 3) exhibited normal gross anatomy of the reproductive tract (Fig 5.2c, g, n, r). No abnormalities were observed in testis sections.

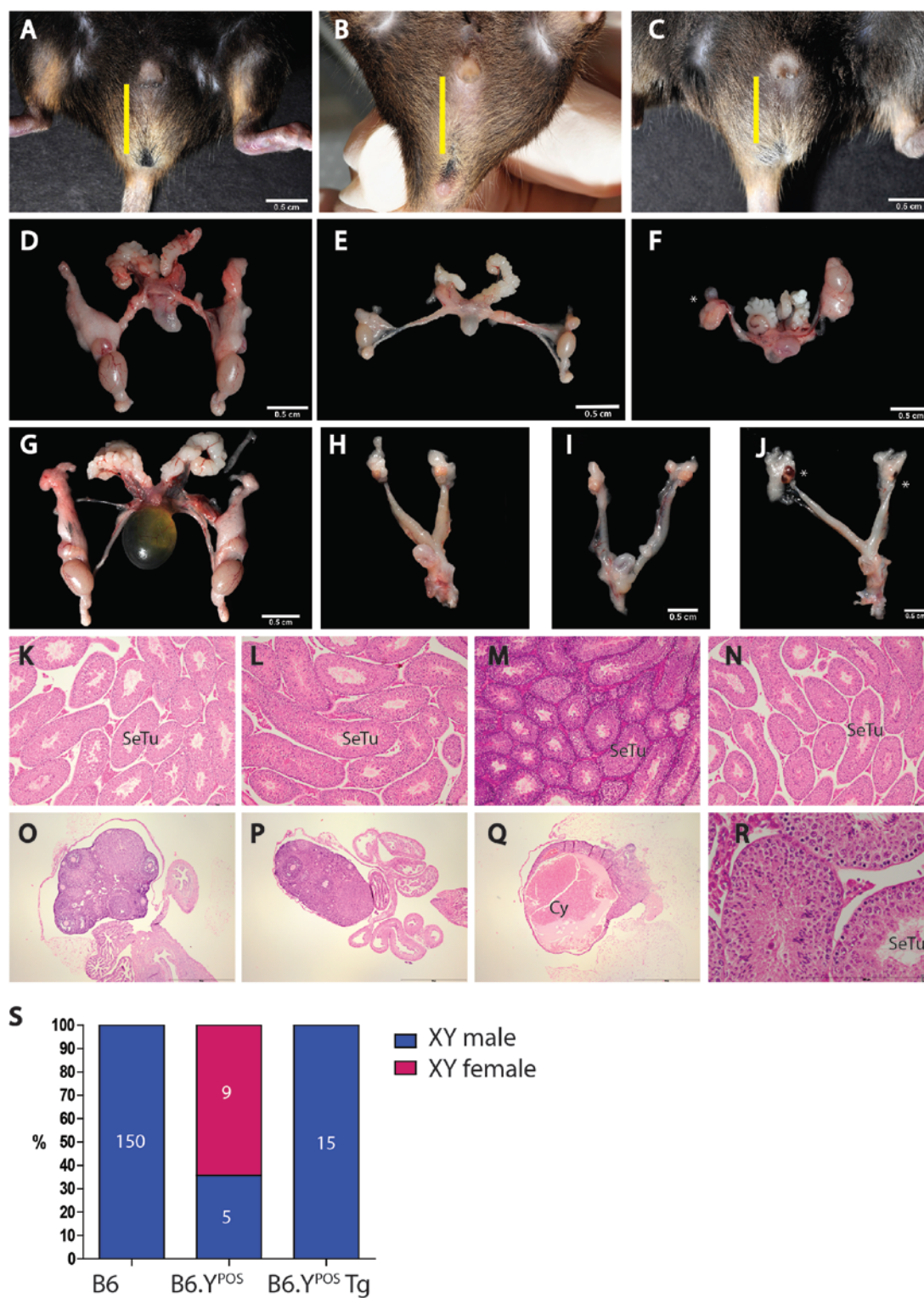


Figure 5.2 Analysis of adult B6.Y^{POS} Tg animals
For legend please see Page 136

Figure 5.2 Analysis of adult B6.Y^{POS} Tg animals

Adult animals were sacrificed for necropsy and histology at 16-20 weeks, with the exception of the animal in B, which was photographed alive at 24 weeks. Anogenital distances of B6 (A) and B6.Y^{POS} Tg (C) animals were measured and no difference was observed. Examination of reproductive organs (D-J) revealed a range of phenotypes in B6.Y^{POS} males (n = 8) – 6 mice possessed bilateral testes (E) that appeared smaller than B6 (D). 2 mice had one testis with a contralateral testis that was either very small or completely absent (asterisk in F). B6.Y^{POS} XY females (n = 2) possessed ovaries that were normal in appearance (I, H is XX B6 for comparison) or that contained cysts (asterisks in J). B6.Y^{POS} Tg animals (n = 3) all possessed testes of normal appearance (G) Examination of H&E-stained sections of gonads indicated normal appearance of some B6.Y^{POS} testes (L, K is B6 for comparison) but one animal showed abnormal staining of the interstitium between seminiferous tubules (M). Ovaries of B6.Y^{POS} XY females appeared normal (P, O is XX B6 for comparison) or contained cysts (Q). B6.Y^{POS} Tg testes appeared histologically normal (N, R). SeTu, seminiferous tubule; Cy, cyst. (S) Proportions of XY males and females within the B6, B6.Y^{POS} and B6.Y^{POS} Tg colonies, based on anogenital examination at weaning. The B6.Y^{POS} colony consists of 35% XY males and 65% XY females, whereas all B6.Y^{POS} Tg animals are male. n numbers are indicated within the chart.

5.2.2. *Sry* expression

It has previously been reported that *Sry* is delayed in B6.Y^{POS} gonads compared with B6.Y^{B6} (Bullejos and Koopman, 2005). Given that *Map3k4* has been implicated in the regulation of timely *Sry* expression in the mouse (Warr et al., 2012, Gierl et al., 2012, Warr et al., 2014) we proposed that some degree of restoration of *Sry* expression may underlie the rescued testis determination observed in B6.Y^{POS} Tg embryos. In order to assess this, stage-matched gonads from B6.Y^{B6}, B6.Y^{POS} and B6.Y^{POS} Tg embryos were collected and expression profiling of *Sry* was performed, by both WMISH and qRT-PCR.

For WMISH, 4-5 UGRs were used per stage and a representative gonad was selected for photography. In B6, *Sry* transcript was detectable from 14ts, peaking at 18ts and declining to low levels by 26ts (Fig 5.3 A-G). As has previously been reported, *Sry* transcript appeared delayed in B6.Y^{POS} gonads, peaking instead around 22ts (Fig 5.2 H-N). *Sry*^{POS} expression followed the characteristic spatial profile of *Sry*^{B6}, initiating in the centre of the gonad, before spreading to the poles. B6.Y^{POS} Tg embryos expressed *Sry* more strongly than both B6.Y^{POS} and B6 at earlier stages, with peak expression being around 20-22ts (Fig 5.3 O-U).

qRT-PCR was performed on subdissected gonads between 15ts and 28ts. 4 to 6 pairs of gonads were used per tail somite stage and the mean relative expression, calculated according to section 2.8.6.3, of all samples for each stage was plotted (Fig 5.3 V). In B6, *Sry* transcript rapidly rises to a sharp peak at 17ts, then declines to low levels by 28ts. In B6.Y^{POS}, *Sry* expression is at an equivalent level to B6 at 16ts, but appears not to begin rising until after 17ts. It then rises between 18-20ts, plateauing around 19-20ts

with no definitive peak, after which it declines. *Sry* expression in B6.Y^{POS} is significantly lower than B6 at 17ts, but significantly higher at 19ts, 22ts, 24ts, and 28ts.

In B6.Y^{POS} Tg gonads, *Sry* expression is consistently higher than B6.Y^{POS} and B6 from 15-21ts. Whilst expression is significantly higher than in B6 at 16ts, 19ts, 20ts, 21ts, 22ts and 26ts, it is only significantly higher than B6.Y^{POS} at 16ts and 21ts. Significance of comparative expression was determined using a Student's t-test. A summary of *p* values for these expression data is given in Table 5.1.

Table 5.1 Summary of *p* values from qRT-PCR profiling of *Sry* expression

Stage (tail somites)	<i>p</i> value B6 vs Y ^{POS}		<i>p</i> value B6 vs Y ^{POS} Tg		<i>p</i> value Y ^{POS} vs Y ^{POS} Tg	
16	0.398	ns	0.026	*	0.046	ns
17	0.007	**	-	-	-	-
18	0.137	ns	0.194	ns	0.581	ns
19	0.045	*	0.060	*	0.455	ns
20	0.130	ns	0.010	**	0.326	ns
21	0.348	ns	0.024	*	0.081	ns
22	0.041	*	0.001	***	0.533	ns
24	0.019	*	0.077	ns	0.670	ns
26	0.239	ns	0.045	*	0.331	ns
28	0.025	*	0.090	ns	0.123	Ns

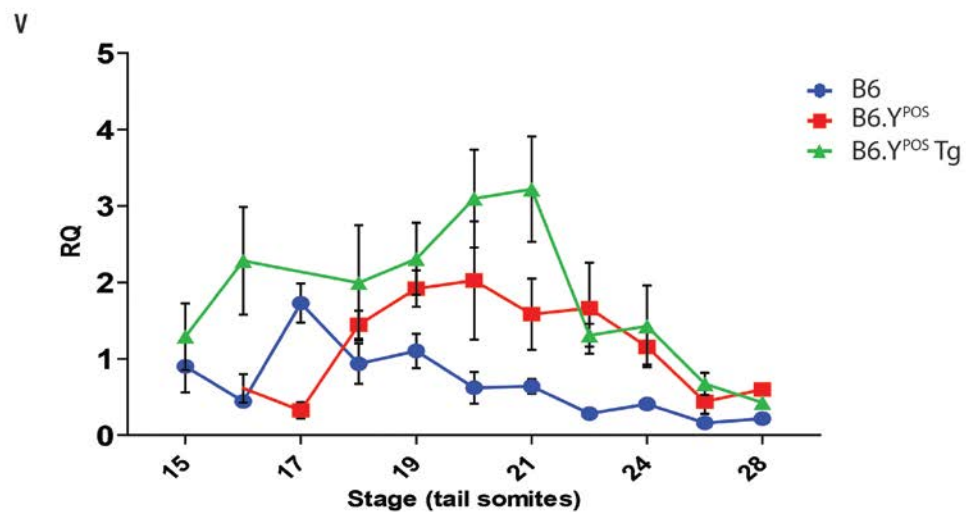
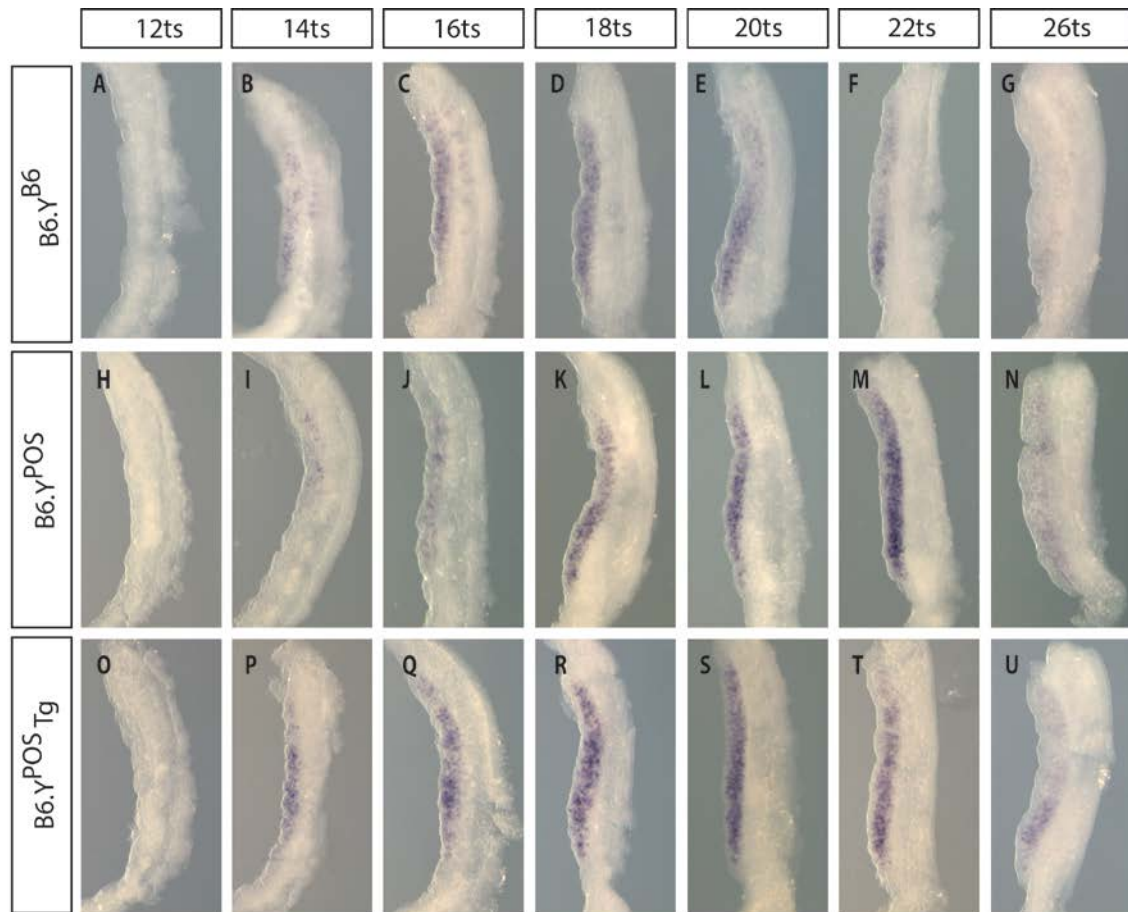


Figure 5.3 Delayed *Sry* expression in B6-Y^{POS} embryos is ameliorated by overexpressing *Map3k4*

Sry expression was examined using WMISH (A-U) and qRT-PCR (V). In B6.Y^{POS} gonads, *Sry* transcript is detectable from 14ts, reaches a peak at 18ts and then declines, becoming restricted to the caudal pole by 26ts (A-G). In B6.Y^{POS}, *Sry* is more weakly expressed at 14-16ts (I-J), then rises to peak intensity at 22ts (K-M). At 26ts, *Sry* is still present throughout the length of the gonad but has declined to low levels (N). Overexpression of *Map3k4* is associated with stronger *Sry* expression (O-U) and peak expression being shifted to 20ts (S). Images are representative of 4-5 UGRs. Examination of relative transcript levels by qRT-PCR on subdissected gonads (V) reflects the pattern observed using WMISH, with peak expression in B6.Y^{POS} appearing shifted relative to that of B6. In B6.Y^{POS} Tg gonads, relative *Sry* expression is higher throughout the critical period of testis determination. Each data point represents the mean RQ of 4-6 pairs of subdissected gonads. Error bars, SEM.

5.2.3. *Sox9* expression

Sox9, expressed in pre-Sertoli and Sertoli cells, is directly upregulated by SRY and is an excellent readout of SRY activity and the degree of “maleness” of a developing gonad. To assess whether the observed changes in *Sry* expression in B6.Y^{POS} Tg gonads are fed through to effects on downstream pro-testis genes, *Sox9* expression was profiled in B6.Y^{B6}, B6.Y^{POS} and B6.Y^{POS} Tg gonads, by WMISH and qRT-PCR.

For WMISH, 4-5 UGRs were used per stage and a representative gonad was selected for photography. *Sox9* is strongly expressed in B6 gonads from 18ts onwards and testis cords can clearly be seen by 28ts (Fig 5.4a-e). B6.Y^{POS} gonads exhibit marked reduction in *Sox9* expression, which becomes restricted to the central region of the gonad at later stages (Fig 5.4i, j). Testis cords cannot be clearly seen.

Overexpression of *Map3k4* was associated with enhanced *Sox9* expression relative to B6.Y^{POS}, although still appeared somewhat delayed (Fig 5.4k, l). At 31ts, cords can clearly be seen but *Sox9* is not detectable in the caudal pole of the gonad (asterisk in Fig 5.4o), indicative of delayed testis determination.

qRT-PCR was performed on the same subdissected gonads that were used in Fig 5.3v above (Fig 5.4p). *Sox9* expression was significantly lower in B6.Y^{POS} gonads compared to B6 at 17ts, 19ts, 21ts, 24ts and 26ts, indicating that expression failed to be upregulated. B6.Y^{POS} Tg gonads expressed higher levels of *Sox9* than B6.Y^{POS}, and this was significant at 16ts, 18ts, 20ts, 26ts and 28ts. Intriguingly, although *Sry* expression was generally higher in B6.Y^{POS} Tg gonads than in B6 gonads (Fig 5.3), *Sox9* expression did not reflect the same pattern of elevated expression relative to B6, and instead was expressed at slightly lower levels than B6 at all stages examined.

These data indicate that whilst overexpression of *Map3k4* is sufficient to significantly elevate *Sry* expression to supra-B6 levels, this effect is not completely recapitulated downstream and *Sox9* is unable to reach B6-like expression levels. It is, however, significantly elevated compared with B6.Y^{POS}. This may account for the incomplete rescue of testis determination observed at 14.5dpc (Fig 5.1). *p* values from statistical analysis of these data are summarised in Table 5.2.

Table 5.2 Summary of *p* values from qRT-PCR profiling of *Sox9* expression

Stage (tail somites)	<i>p</i> value B6 vs Y ^{POS}		<i>p</i> value B6 vs Y ^{POS} Tg		<i>p</i> value Y ^{POS} vs Y ^{POS} Tg	
16	0.119	ns	0.240	ns	0.0002	***
17	0.010	**	-	-	-	-
18	0.058	ns	0.807	ns	0.041	*
19	0.001	***	0.065	ns	0.169	ns
20	0.078	ns	0.360	ns	0.011	*
21	0.008	**	0.111	ns	0.111	ns
22	0.027	*	0.270	ns	0.241	ns
24	0.016	*	0.264	ns	0.038	*
26	0.054	ns	0.986	ns	0.026	*

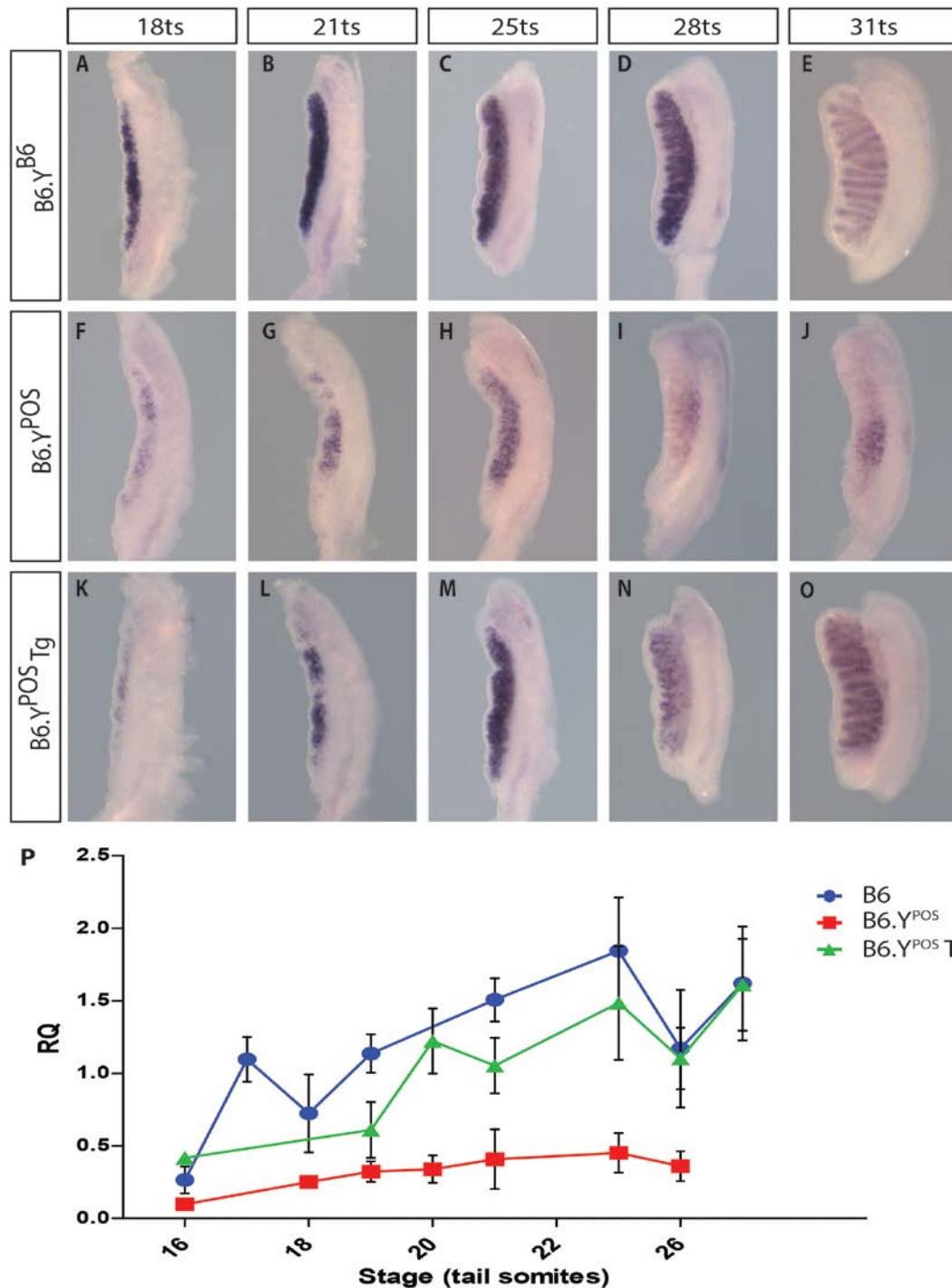


Fig 5.4 Delayed *Sox9* expression in B6.Y^{POS} embryos is ameliorated by overexpressing *Map3k4*

Sox9 is strongly expressed in B6 gonads from 18ts onwards (A-E). B6.Y^{POS} gonads exhibit marked reduction in *Sox9* expression, which becomes restricted to the central region of the gonad at later stages (I, J). Overexpression of *Map3k4* is associated with enhanced *Sox9* expression, although it is still somewhat delayed (K, L). At 31ts, *Sox9* is undetectable in the caudal pole of the gonad, indicative of delayed testis determination. Images are representative of 4-5 UGRs. qRT-PCR analysis of subdissected gonads (P) indicates the same trend, with significantly lower relative expression of *Sox9* at 18ts, compared with B6, which is enhanced in embryos expressing the *Map3k4* BAC but does not reach B6-like levels. Each data point represents the mean RQ of 4-6 pairs of subdissected gonads. Error bars, SEM.

5.2.4. Immunohistochemical analysis of SRY and SOX9 expression

Having established that overexpression of *Map3k4* is able to modulate *Sry* and *Sox9* expression in B6.Y^{POS} gonads, it was necessary to examine whether expression of SRY and SOX9 protein are also affected. A previous report suggests that SRY expression is similar in B6.Y^{POS} and B6 at 11.5 dpc but that SRY is downregulated later in B6.Y^{POS} (Wilhelm et al., 2009). To assess whether this is the case in the B6.Y^{POS} colony at Harwell, and whether overexpression of *Map3k4* also affects abundance of SRY, fluorescence immunohistochemistry was performed. Longitudinal sections from stage-matched B6.Y^{B6}, B6.Y^{POS} and B6.Y^{POS} Tg gonads were probed with anti-SRY and anti-SOX9 antibodies, followed by hybridisation with Alexa Fluor secondary antibodies (See Section 2.9.2). Representative sections were photographed.

Contrary to the report by Wilhelm *et al*, SRY was reduced in B6.Y^{POS} gonads at 18ts and only detectable in the centre of the gonad (Fig 5.5). Expression was then equivalent to B6 at 20ts, being detectable along the length of the gonad. By 28ts, expression had diminished in B6 gonads to undetectable levels, whereas it was still expressed throughout the gonad in B6.Y^{POS}. B6.Y^{POS} Tg gonads expressed SRY at similar levels to B6 at 18ts and 20ts, but, similarly to B6.Y^{POS} gonads, SRY was still present at 28ts. This is in agreement with the broad expression profile observed by profiling expression of *Sry*^{POS} mRNA in Figure 5.4 and previous reports that SRY^{POS} is downregulated later than SRY^{B6} (Wilhelm et al., 2009).

SOX9 expression also appeared delayed in B6.Y^{POS} gonads (Fig 5.6). At 18ts, SOX9 was barely detectable in B6.Y^{POS} gonads whereas it was strongly expressed in B6. At 20ts,

B6.Y^{POS} gonads also showed few SOX9-expressing cells that were restricted to the central region of the gonad. At 28ts, SOX9 expression in B6 was strong and testis cords were clearly evident, whereas in B6.Y^{POS} gonads the beginnings of ovotestis development were observed as SOX9 signal was restricted to the central region of the gonad. SOX9 was primarily expressed in the centre of the gonad and no cords were observed. SOX9 was somewhat less abundant in B6.Y^{POS} Tg gonads at 18ts but reached equivalent levels by 20ts. Interestingly, at 28ts, expression of SOX9 was strong but testis cords were less evident, although the SOX9-expressing cells did appear to be beginning to arrange into cord-like structures. This is indicative of delayed testis determination and supports the finding that ovarian tissue can develop at the poles of some B6.Y^{POS} Tg gonads.

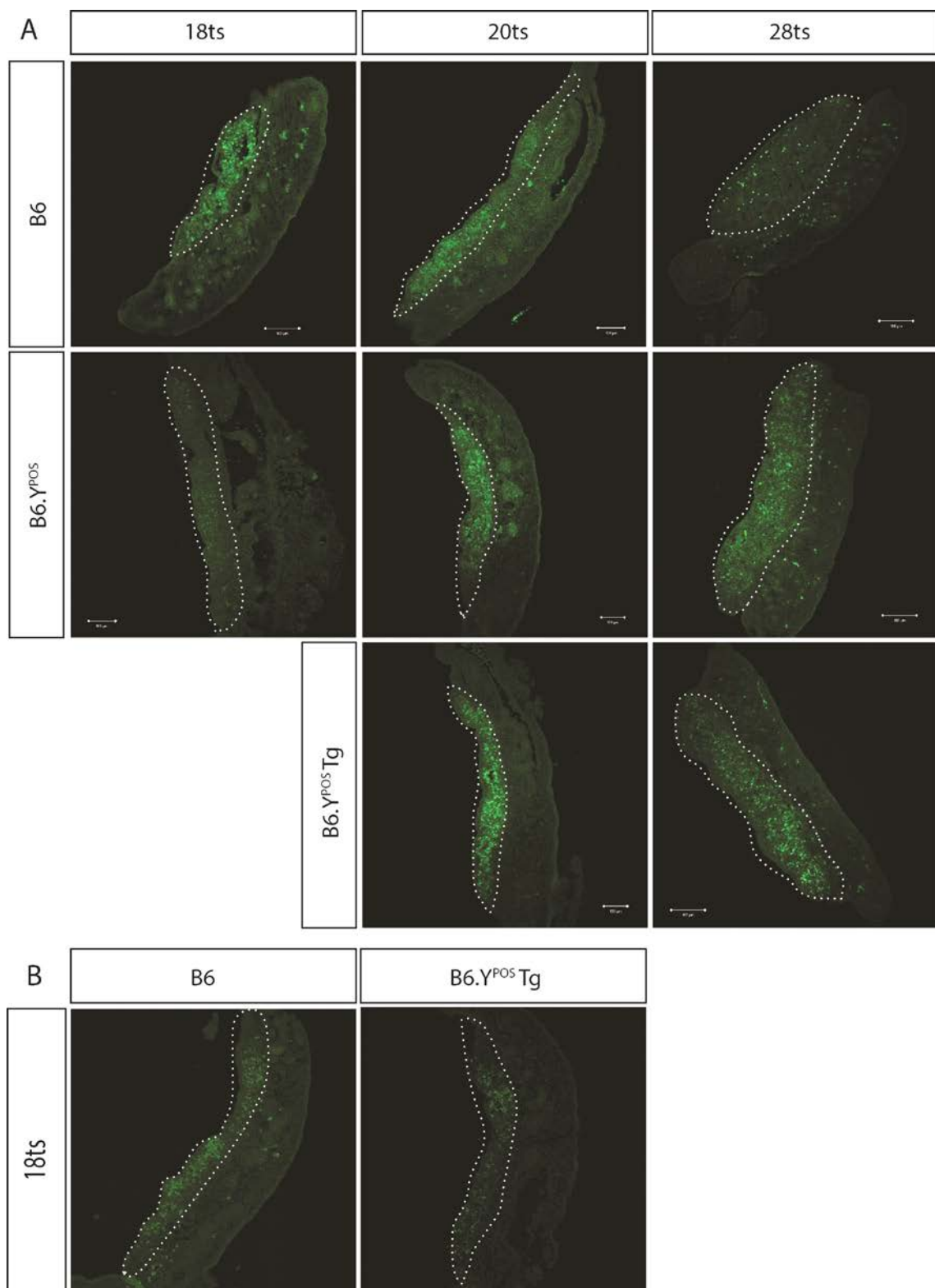


Figure 5.4 SRY expression in B6.Y^{POS}Tg gonads

For legend see page 146

Figure 5.4 SRY expression in B6.Y^{POS} Tg gonads

Immunofluorescence staining using anti-SRY antibody on wax sections of 11.5 dpc – 12.5 dpc gonads. Data from two separate experiments (A and B) is shown. Location of the gonad is indicated by the dotted line. In B6 (top panel in A), SRY expression is strong at 18ts and 20ts, then by 28ts has declined to undetectable levels. Some autofluorescence can be observed in blood cells in the 28ts gonad. In B6.Y^{POS} (middle panel in A), SRY is weakly expressed at 18ts but has increased by 20ts and remains expressed at 28ts, indicative of a delayed expression profile. B6.Y^{POS} Tg gonads (bottom panel in A) also express SRY strongly at 20ts and 28ts. Comparison of SRY between B6 and B6.Y^{POS} Tg expression at 18ts (B) indicates that SRY is expressed at equivalent levels. Scale bar, 100µm.

Figure 5.5 SOX9 expression in B6.Y^{POS} Tg gonads

Immunofluorescence staining using anti-SOX9 antibody on wax sections of 11.5 dpc – 12.5 dpc gonads. Data from two separate experiments (A and B) is shown. Location of the gonad is indicated by the dotted line. In B6 (top panel in A), SOX9 expression is observed at 18ts and 20ts and is maintained at 28ts. Formation of early testis cords can be seen (white arrow). In B6.Y^{POS} (middle panel in A), SOX9 is barely detectable at 18ts and very weak at 20ts. Expression has increased by 28ts but is low to absent at the poles of the gonad (yellow arrows). B6.Y^{POS} Tg gonads (bottom panel in A) also express SOX9 strongly at 20ts and 28ts. Some arrangement of cells into cord-like structures can be seen (white arrow) but the gonad has a narrower shape than B6, indicative of delayed testis determination. Comparison of SOX9 between B6 and B6.Y^{POS} Tg expression at 18ts (B) indicates that SOX9 is barely detectable at 18ts. Scale bar, 100µm.

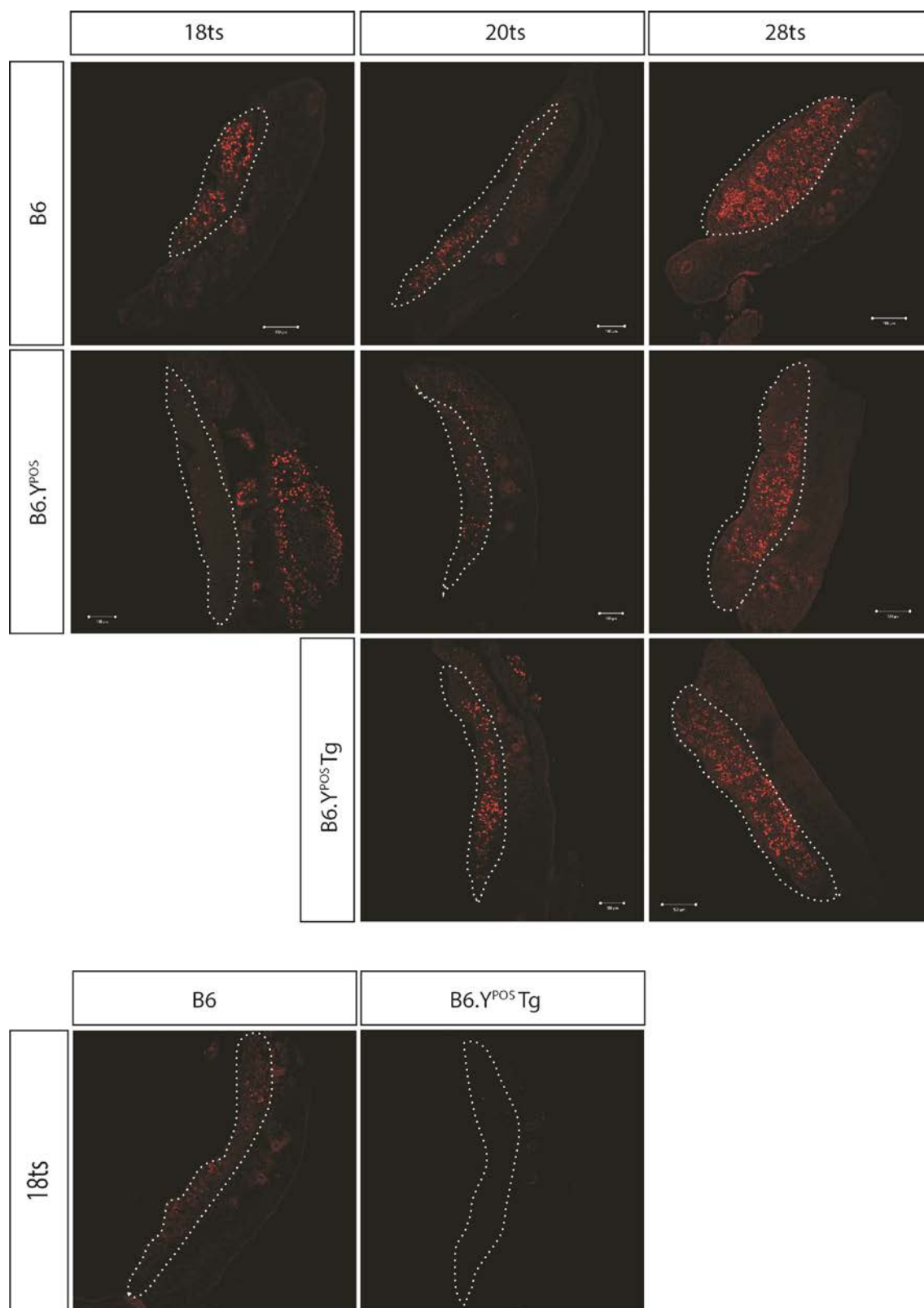


Figure 5.4 SOX9 expression in B6.Y^{POS}Tg gonads
 For legend see page 146

5.2.5. Effect of loss of one copy of *Map3k4* on the gonadal phenotype of B6.Y^{POS} embryos

Together, the data presented in sections 5.2 to 5.5 support a model in which *Map3k4* overexpression is capable of restoring testis determination in B6.Y^{POS} embryos via increased expression of *Sry*/SRY and *Sox9*/SOX9. This, along with previously published reports that loss of *Map3k4* is associated with delayed testis determination and reduced *Sry* expression, supports the hypothesis that *Map3k4* is a highly dosage-sensitive gene in the context of murine testis determination. Thus, it follows that it may be possible to exacerbate B6.Y^{POS} sex reversal by reducing dosage of *Map3k4*. To investigate this, B6.Y^{POS}-*Map3k4*^{+/-} embryos were generated by crossing B6.Y^{POS} males with *Map3k4*^{+/-} females. Embryos were harvested at 14.5dpc and gonads were scored as testes, ovaries or ovotestes (Table 5.3). WMISH was then performed on 10 fixed gonads per genotype to assess expression of *Sox9* and *Stra8* (Fig 5.6). As previously reported, B6.Y^{POS} gonads either resembled ovotestes, which expressed *Sox9* only in the centre of the gonad and *Stra8* at the poles (Fig 5.6c, h), or ovaries, which expressed *Stra8* but exhibited no detectable *Sox9* expression (Fig 5.6d, i). Conversely, all B6.Y^{POS}-*Map3k4*^{+/-} gonads (n = 5 embryos) resembled ovaries, in which *Sox9* was undetectable yet *Stra8* was expressed strongly (Fig 5.6 e, j). Therefore, loss of a copy of *Map3k4* exacerbates B6.Y^{POS} sex reversal.

Table 5.3 Phenotypes of 14.5 dpc gonads from B6.Y^{POS} *Map3k4*^{+/+} and B6.Y^{POS} *Map3k4*^{+/-} embryos

	B6 XY	B6 XX	B6.Y ^{POS}	B6.Y ^{POS}
	+/+	+/+	+/+	+/-
Bilateral testes	10	0	0	0
Bilateral ovotestes	0	0	3	0
Ovary and ovotestis	0	0	13	0
Bilateral ovaries	0	10	12	10
Total	10	10	28	10

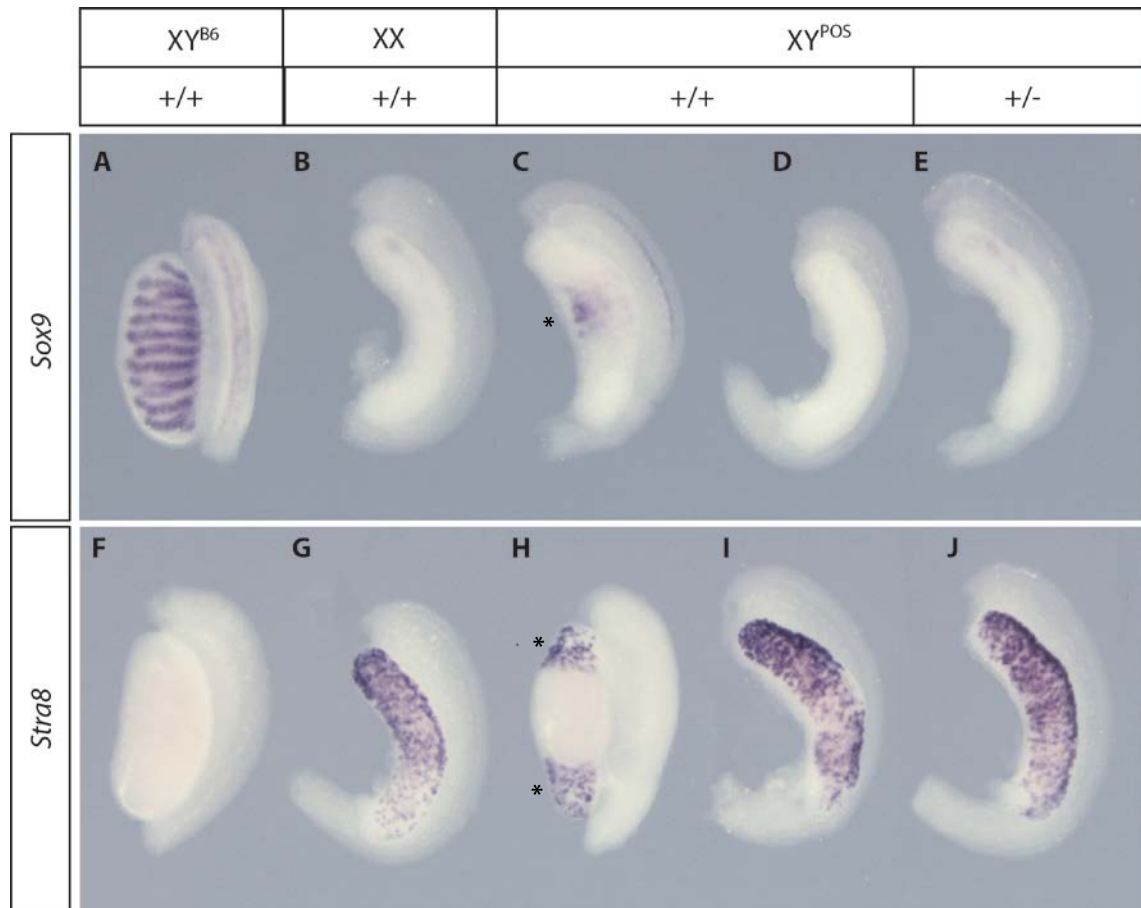


Figure 5.6 Loss of a copy of *Map3k4* exacerbates B6.Y^{POS} sex reversal

WMISH was performed on 14.5 dpc gonads from the strains indicated above the panels, where *Map3k4* zygosity is also indicated. The top panel (A-E) shows expression of the Sertoli cell marker *Sox9* and the bottom panel (F-J) shows expression of the marker of ovarian meiotic germ cells, *Strat8*. In B6, XY gonads (A, F) express *Sox9* in the testis cords and do not express *Strat8*, whereas XX gonads show the opposite pattern (B, G). B6.Y^{POS} embryos (n = 5) develop either ovaries (C) or ovotestes (D, H). Note restriction of *Sox9* expression to the central region of the gonad (asterisk in D) and *Strat8* signal at the poles (asterisks in H). All *Map3k4*^{+/-} B6.Y^{POS} embryos (n=5) developed bilateral ovaries.

5.3. Discussion

5.3.1. B6.Y^{POS} Tg mice develop as normal males

To further support findings that all B6.Y^{POS} Tg mice are born as phenotypic males and are able to breed successfully (N. Warr, unpublished data), a more detailed analysis of adult animals was performed. Testes appeared morphologically normal and histological examination did not identify ovarian tissue. Thus, it can be deduced that the development of a small amount of ovarian tissue in embryos (Fig 5.1i) is transient. This is reminiscent of testis development in B6 embryos carrying Y^{AKR} (Nagamine et al., 1987c), which also develop transient ovotestes but have testes at birth.

5.3.2. Overexpression of *Map3k4* leads to enhanced *Sry* expression in B6.Y^{POS} embryos

Sry expression was substantially elevated in B6.Y^{POS} Tg embryos. Importantly, it is significantly elevated over both B6 and B6.Y^{POS} at 17ts, but not at later stages. This supports the idea that the crucial period for *Sry* expression to reach a functional threshold in B6 is early, and that higher expression at later stages is not able to compensate for insufficient expression during this critical window. These data also suggest that delayed *Sry* expression is the principal cause of B6.Y^{POS} sex reversal. This agrees with previous reports that expression of *Sry*^{B6} follows a particularly narrow expression profile, with an early onset and peak compared with other strains. This may have evolved to overcome the precocious expression of pro-ovary genes in B6 and, therefore, introduction of a later-expressing *Sry* allele disrupts *Sry* function sufficiently during the critical window that its capacity to circumvent the ovarian pathway is compromised. Also in support of this model are experiments performed by Eicher *et al*

(1995), who successfully corrected B6.Y^{POS} sex reversal by introduction of an *Sry* transgene from the 129 strain (Eicher et al., 1995).

Additionally, overexpression of *Map3k4* is associated with elevated SRY expression in B6.Y^{POS} embryos, indicating that changes to expression of *Sry* mRNA are fed-through to changes in abundance of the protein. Interestingly, though, SRY^{POS} expression was more abundant than SRY^{B6} at 28ts, even in the presence of the *Map3k4* BAC. This was also observed at the mRNA level, and suggests that distinct processes govern the onset and downregulation of *Sry* expression. Clearly, *Map3k4* plays a key role in augmenting *Sry*'s timely expression, however, the question of how it is downregulated remains to be answered.

5.3.3. *Sox9* is modestly elevated in B6.Y^{POS} Tg embryos

Whilst *Sry* expression was elevated to supra-B6 levels at most of the stages examined, *Sox9* expression in B6.Y^{POS} Tg embryos was only modestly upregulated over that of B6.Y^{POS}. There are several possible explanations for this. Firstly, it may be that although the expression of *Sry*^{POS} is elevated in the presence of the *Map3k4* BAC, it is still a *domesticus*-derived *Sry* allele. It was recently demonstrated in HEK293 cells that a construct containing the closely related *Sry*^{TIR} had reduced ability to transactivate a TESCO luciferase reporter (Zhao et al., 2014). This was shown to be related to the truncated polyQ repeat domain in the *domesticus* protein isoform. Therefore, it is plausible that SRY^{POS} may possess reduced transcriptional-activator capability and therefore, even increased expression is not able to fully compensate for sub-optimal activation of *Sox9* expression by the SRY^{POS} protein.

5.3.4. Losing a copy of *Map3k4* exacerbates B6.Y^{POS} sex reversal

That removal of one copy of *Map3k4* is sufficient to completely feminise all B6.Y^{POS} embryos supports the hypothesis that *Map3k4* is an important, dose-dependent modifier of sensitivity to B6.Y^{POS} sex reversal. It also highlights the sensitivity of the B6 background to perturbations in dosage of key testis-determining genes, emphasising the complex interactions between the mutually antagonistic male and female pathways.

5.3.5. Variation and the analysis of B6.Y^{POS} gene expression

One feature of B6.Y^{POS} gonads, demonstrated in this chapter and that has also been well documented previously, is the variation in the severity of sex reversal observed between B6.Y^{POS} embryos. 14.5 dpc B6.Y^{POS} gonads exhibit varying degrees of sex reversal including bilateral ovaries, an ovary and contralateral ovotestis, or bilateral ovotestes (Eicher et al., 1982, Nagamine et al., 1999). When sampling tissue at 11.5 dpc it is not possible to ascertain the phenotypic destiny of a selected gonad and, therefore, it is possible that samples may, by chance, contain gonads that are over representative of a particular fate. Therefore, since sample sizes are relatively small due to the availability of embryonic gonads at specific tail somite stages, there is the chance that data points may become biased. Attempts were made to reduce bias by utilising as many samples as feasibly possible for each data point, but the possibility still cannot be excluded and so must be taken into account when considering gene expression data from B6.Y^{POS} gonads.

There is evidence to support the hypothesis that variation in gonadal phenotypes of B6.Y^{POS} animals, and other B6.Y^{DOM} strains, may be the result of the inherently

stochastic nature in the testis- and ovary-determining pathways. Principally, this evidence stems from the observation that B6.Y^{POS} individuals can have different gonadal phenotypes in each gonad. Since phenotypic variation occurs between two isogenic gonads, there is a strong suggestion that a level of stochasticity is responsible. However, genetic factors contributing to the likelihood of developing testes or ovaries in B6.Y^{POS} mice have also been reported, and will be discussed in the next chapter.

5.3.6. Summary and future directions

Taken together, the data presented above support a key role for *Map3k4* in the regulation of *Sry* expression and, therefore, the susceptibility of the B6 strain to sex reversal in the presence of the Y^{POS} chromosome. Having this unique tool, with which to modify expression of a *domesticus* *Sry* allele, will facilitate further investigation of the control of *Sry* expression and is discussed in the next chapter. This loss-and gain-of-function approach could also provide clues as to the loci of importance in mediating strain-specific susceptibility to sex reversal if applied to other genes of interest.

Genetic complementation experiments could help to support the idea that the ovary-determining pathway is particularly precocious in the B6 strain. If, indeed, a major contributing factor to the propensity of the B6 background to sex reversal is hypermorphic alleles of pro-ovarian genes, could we modify B6.Y^{POS} sex reversal by altering dosage of these? Microarray comparison between B6 and 129 XY 11.5 dpc gonads found that *Wnt4* was upregulated in B6 by some 1.53-fold (Munger et al., 2009), so it could be a promising candidate for dosage-modification.

The remainder of this thesis will focus on investigating why *Sry*^{POS} is delayed compared with *Sry*^{B6}, in the context of the B6 genome, and how overexpressing *Map3k4* is

sufficient to overcome this. Answering both questions is important and may contribute significantly to our understanding of *Sry* regulation. By doing so it may be possible shed further light upon the intriguing B6.Y^{POS} sex reversal phenomenon as well as the control of *Sry* expression.

Chapter 6. Molecular studies of B6.Y^{POS} sex reversal

6.1. Introduction

6.1.1. Questions arising from Chapter 5

In the previous chapter, *Map3k4*-mediated rescue of B6.Y^{POS} sex reversal, via effects on *Sry*, is described. This highlights the importance of delayed *Sry* expression as a key defect underlying sex reversal in B6.Y^{POS} gonads. Two key questions have arisen from these studies:

- Why is *Sry* expression delayed in B6.Y^{POS} embryos?
- How does overexpressing *Map3k4* overcome this?

It is important to note that these questions may have the same answer, but that they also may not. For example, a deficit of factor X in B6.Y^{POS} may not be directly rectified by *Map3k4* overexpression. Instead, high levels of MAP3K4 may circumvent the lack of factor X via a separate factor, Y, to achieve timely *Sry* expression. Therefore, in designing experiments it is always important to consider this. However, by utilising the *Map3k4* transgene as a tool for modifying B6.Y^{POS} sex reversal, it may be possible to answer both questions simultaneously through careful analysis of experimental data.

A third question, that has arisen as a result of recent publications of other autosomal modifiers of B6.Y^{POS} sex reversal, is the question of variable penetrance of B6.Y^{POS} sex reversal in an apparently consomic strain. This question will also be addressed in this chapter.

6.1.2. Approaches to understanding *Sry* regulation

By analysing changes to key molecular events in B6.Y^{POS} embryos with and without the *Map3k4* transgene, we may be able to further elucidate the mechanism of MAP3K4-

mediated regulation of *Sry* expression. Furthermore, by comparing such markers between B6 and B6.Y^{POS} gonads, we may also be able to identify features that are responsible for delayed *Sry* expression in B6.Y^{POS} embryos. There are a number of approaches that could be taken to investigate the mechanism of delayed *Sry* expression in B6.Y^{POS} and its rescue by overexpression of *Map3k4*.

6.1.2.1. Biochemical approach

Since MAP3K4 is a kinase, it would be prudent to analyse levels of phosphorylation of targets of the MAP3K4 pathway, particularly those that have previously been implicated in *Sry* regulation. There is evidence that the GADD45-MAP3K4 pathway facilitates timely *Sry* expression via the phosphorylation of p38 MAPK (Warr et al., 2012, Gierl et al., 2012). Therefore, analysis of levels of p-p38 in B6.Y^{POS}-Tg gonads at the critical testis-determining stage may enable us to determine whether elevated MAP3K4 is also associated with elevated p-p38 at this time, and therefore whether this is likely to be an underlying cause of enhanced *Sry* expression. Likewise, it may also indicate whether there is a difference in p-p38 expression in B6.Y^{POS} gonads compared with B6. If the latter were to be observed, it may indicate that some other Y-encoded factor may be acting to modulate proteins that regulate *Sry*, or that p38 is also involved in testis determination post-*Sry* expression.

Another key molecule whose phosphorylation may be mediated via the MAP3K4 pathway is GATA4. Loss of MAP3K4 function is associated with reduced levels of p-GATA4 (Warr et al., 2012, Gierl et al., 2012) and GATA4 has been shown to directly associate with the *Sry* promoter in a MAPK-dependent manner (Gierl et al., 2012). Given this evidence, it is plausible that enhanced MAP3K4-mediated signalling in the

gonad may lead to increased phosphorylation of GATA4, and that this may, in turn, facilitate enhanced expression of *Sry* in B6.Y^{POS}-Tg gonads.

6.1.2.2. Genomic approach

It has previously been shown that *M. musculus*- and *M. domesticus*-derived alleles encode different SRY protein isoforms (Coward et al., 1994, Miller et al., 1995). However, this alone cannot account for the differing degrees of sex reversal observed in different B6.Y^{DOM} strains (Carlisle et al., 1996, Albrecht and Eicher, 1997, Albrecht et al., 2003). Therefore, strain-specific differences in sequences flanking *Sry* may in some way be involved in the misregulation of its expression. These could be polymorphisms in regions previously implicated in *Sry* expression, such as CpG dinucleotides or binding motifs for factors known to bind the *Sry* promoter, such as GATA4 (Gierl et al., 2012). They may also affect the recruitment of chromatin modifying proteins, such as the histone demethylase JMJD1A (Kuroki et al., 2013). Additionally, polymorphisms could be identified that are associated with as yet unidentified regulators of *Sry* expression and may therefore facilitate identification of these.

At the time of writing, much of the flanking sequence of *Sry*^{POS} remains unpublished. In 1996, Coward *et al* sequenced 8 kb of sequence surrounding the closely related *Sry*^{TIR} ORF. In this study, 2.7 kb of this sequence was compared with sequence from the 129 inbred strain, which possesses a *molossinus*-derived Y chromosome. This study identified only 3 polymorphisms in the 5' UTR of *Sry*^{TIR}. Intriguingly, all of these reside within the region that we now refer to as the *Sry* T-DMR and two of them result in ablation of CpG dinucleotides. Whilst the Tirano and *poschiavinus* strains are closely related, it cannot be assumed that the sequence of this region is identical in the two.

Moreover, sequence comparisons more distal to the *Sry* ORF may enable identification of other potentially important differences between *domesticus* and *molossinus*-derived Y chromosomes. Finally, genetic drift can occur within isolated populations of mice over time. Therefore, a strategy for sequencing *Sry* and the flanking regions was implemented for the Harwell B6.Y^{POS} colony.

A second genetic approach to understanding the mechanism of delayed *Sry* expression in B6.Y^{POS} embryos is the analysis of expression of genes that are implicated in the regulation of *Sry*. This may help to elucidate two things. Firstly, if there is no difference in the expression of such genes, it suggests that delayed *Sry* expression in B6.Y^{POS} embryos is a direct result of a *cis* effect at the *Sry* locus itself, i.e. an altered *Sry* regulatory element. However, if, for example, there was lower expression of *Gadd45g* in B6.Y^{POS} embryos, it suggests that the delay may instead due to the effect of a strain-specific Y-encoded factor that is able to alter expression of *Sry*-regulating autosomal genes, thereby affecting *Sry* in *trans*. Several genes have been implicated in the regulation of *Sry*, and when insufficiently expressed, some are associated with reduced or delayed expression. Most notably, loss of either *Gadd45g* or *Map3k4* leads to severely delayed expression of *Sry* (Warr et al., 2012, Gierl et al., 2012). It is plausible that the rescue of sex reversal in B6.Y^{POS} embryos by the *Map3k4* transgene is simply a case of overcoming an insufficiency in the MAP3K4 pathway in B6.Y^{POS} embryos. Additionally, it would also be necessary to assess levels of expression of other genes required for normal expression of *Sry*, such as *Sf1*, *Wt1*, *Six1/4* or *Cbx2*.

Secondly, assessing expression of these genes may also enable us to suggest a mechanism by which overexpression of *Map3k4* is able to enhance *Sry*^{POS} expression. If

expression of any of the above genes is elevated in gonadal cells from B6.Y^{POS} Tg embryos, it suggests that the MAP3K4 pathway regulates *Sry* expression in an indirect manner, perhaps by increasing expression of other pro-*Sry* genes.

6.1.2.3. Epigenetic approach

Our understanding of the epigenetic control of *Sry* has made significant progress in the last decade. Since the first report of the *Sry* T-DMR in by Nishino *et al* (2004), and more recent chromatin immunoprecipitation (ChIP) experiments identifying factors that bind the *Sry* promoter, including GATA4 (Gierl *et al.*, 2012), the focus of studying *Sry* has shifted from the genetic to the epigenetic. Gathering data about the epigenetic landscape of *Sry*^{POS} during testis determination may facilitate elucidation of the mechanism of its delayed expression on the B6 background. Furthermore, analysis of the epigenetic events underlying expression of two differentially expressed *Sry* alleles in the same genomic context (B6) is an invaluable tool for enabling better understanding of the regulation of *Sry*, of gene expression on the Y chromosome and how sequence context affects gene regulation more generally.

One epigenetic event that may underlie the delayed *Sry* expression observed in B6.Y^{POS} gonads is differences in methylation of the *Sry* T-DMR. This region, lying closely upstream of the *Sry* ORF (Fig 1.7) undergoes a dynamic process of demethylation that is coincident with *Sry* expression (Nishino *et al.*, 2004, Gierl *et al.*, 2012). However, no sex reversal study has yet reported altered *Sry* methylation being associated with a change in expression of *Sry*. The delayed expression of *Sry*^{POS} on the B6 background is a good candidate for investigating whether CpG methylation is altered in a model of *Sry* misregulation. Moreover, altered CpG methylation at the *Sry* T-DMR may also underlie

rescued *Sry* expression in B6.Y^{POS} Tg embryos, although a link between MAP3K4 and DNA methylation is yet to be established.

CpG methylation can be analysed using bisulphite sequencing. This technique pairs sodium bisulphite conversion of DNA with targeted sequencing. Sodium bisulphite converts unmethylated cytosines to uracil, meaning that they can be distinguished from methylated cytosines by sequencing bisulphite-treated DNA and comparing the sequence with untreated DNA. Following bisulphite treatment, a region of interest is amplified by PCR, cloned and then sequenced. This technique has previously been employed to analyse *Sry* methylation in *Gadd45g* and *Map3k4* knockout gonads (Warr et al., 2012).

One caveat of studying CpG methylation in any tissue containing several different cell populations is the presence of cells that do not express the gene of interest. In the case of the gonad, this group consists mostly of germ cells, and, to a lesser extent, the somatic progenitors that will develop into Leydig cells and other interstitial cell types. In order to reduce noise from cells not expressing *Sry*, it is necessary to enrich for the cell population of interest, the somatic progenitors. This can be achieved by sorting gonadal cells that express a fluorescent lineage marker by fluorescence-assisted cell sorting (FACS). This technology detects cells expressing the fluorescent marker and can sort them, enabling collection of a sample enriched with the cell type of interest. One previously-employed marker of the somatic lineage is *Sf1-EGFP*. By generating mouse lines expressing *Sf1-EGFP*, it may be possible to compare *Sry* methylation between B6 and B6.Y^{POS} cells at crucial time points in testis determination.

6.1.3. Varying severity of B6.Y^{POS} sex reversal

The question of variation in the severity of B6.Y^{POS} sex reversal is an intriguing one and warrants further investigation. It has previously been suggested that phenotypic variation in B6.Y^{POS} gonads may be the result of stochasticity of the testis-determining pathway. Because of this, small differences in the number of *Sry*-expressing cells at the critical time of testis determination may have great impact upon downstream events, and therefore have a large phenotypic readout. However, it has also been suggested that the presence of genetic modifiers may also be an underlying cause of this variation (Eicher and Washburn, 2001). A reported region identified on chromosome 11 provides protection against B6.Y^{POS} sex reversal (Nikolova et al., 2008). This region was later narrowed down to a 1.62 Mb segment and is believed to contain a strong *Sox9* enhancer, as presence of the region confers enhanced *Sox9* expression in B6.Y^{POS} gonads and results in almost all XY individuals developing testes (Arboleda et al., 2014, Nikolova et al., 2008). The region is thought to be derived from the POSA strain from which the B6.Y^{POS} line was derived, and has been maintained through selection of fertile males over at least 20 generations (Arboleda et al., 2014). This protective region, denoted the “110 region”, offers an alternative explanation for variation in the severity of sex reversal in B6.Y^{POS} embryos, since only 50% of the progeny of an animal heterozygous for the region will inherit it. It has been postulated that, if this region were eliminated from a B6.Y^{POS} colony, all XY animals would undergo complete sex reversal and the Y^{POS} chromosome would eventually be lost. In support of this, many labs maintain their B6.Y^{POS} colonies using an *Sry* transgene, as sufficient numbers of fertile males cannot be obtained via conventional breeding (Arboleda et al., 2014). The MRC Harwell B6.Y^{POS} colony is maintained through conventional breeding, with approximately 35% of XY^{POS} individuals developing as males (Chapter 5, Figure 5.2).

This suggests that the 110 region may be present in the colony and is conferring some protection against B6.Y^{POS} sex reversal.

6.1.4. Aims

The aims of this chapter are to investigate the mechanism underlying delayed *Sry* expression in B6.Y^{POS} embryos and its amelioration by overexpressing *Map3k4*. This will be achieved using the following approaches:

- Analysis of abundance of phosphoproteins known to be associated with MAP3K4 activation and *Sry* expression
- Identification of sequence polymorphisms between the flanking regions *Sry*^{POS} and *Sry*^{B6}
- Analysis of CpG methylation in B6.Y^{POS} gonadal somatic cells at 11.5 dpc.

A further aim of this chapter is to determine whether the 110 region has been maintained in the Harwell B6.Y^{POS} colony and, therefore, whether its presence or absence may partially explain the variation in observed severity of B6.Y^{POS} sex reversal. This will be achieved by sequencing DNA from B6.Y^{POS} animals to identify POSA-specific SNPs within the putative 110 region.

6.2. Results

6.2.1. Analysis of p38 and GATA4 phosphorylation

To investigate whether p38 and GATA4 phosphorylation are affected by overexpression of *Map3k4*, Western blots were performed on lysate from subdissected gonads from B6, B6.Y^{POS} and B6.Y^{POS} Tg embryos (Fig 6.1). Two stage-ranges were selected for examination. An early stage, 15-16ts, was chosen to facilitate analysis of phosphorylation during the earlier stages of *Sry* expression. A later stage, 17-18ts, which represents the peak of *Sry* expression in B6, was also selected.

At 17-18ts, there was no significant difference in the expression of p38 in B6.Y^{POS} gonads compared to B6.Y^{B6} (Fig 6.1 a, b). Likewise, B6.Y^{POS} Tg gonads expressed similar levels of p38 (Fig 6.1 a b). Expression of p-p38 also did not differ significantly between B6.Y^{B6}, B6.Y^{POS} and B6.Y^{POS} Tg gonads at this stage (Fig 6.1 a,c). Due to lack of lysate, it was not possible to obtain data for p38 expression in 15-16ts gonads.

There was no significant difference in the expression of GATA4 or p-GATA4 in B6.Y^{POS} gonads compared to B6.Y^{B6} at 15-16ts (Fig 6.1 d-f). Likewise, B6.Y^{POS} Tg gonads also expressed similar levels of GATA4 and p-GATA4. At 17-18ts, there was no significant difference in expression of 55kD GATA4 between the three strains (Fig 6.1 g, h). However, the 46kD GATA4 isoform was significantly more highly expressed in B6.Y^{POS} compared to B6.Y^{B6} ($p = 0.03$). There was a trend towards further increased expression of 46kD GATA4 in B6.Y^{POS} Tg gonads, but this was not significant ($p = 0.10$). There was no significant difference in levels of p-GATA4 at 17-18ts between any of the three strains (Fig 6.1 g-i).

That there were no detectable differences in p38 or GATA4 phosphorylation at the stages examined, despite increased GATA4 expression at 15-16ts, indicates that the *Map3k4* transgene may not be acting via a phospho-p38- or phospho-GATA4-dependent pathway. Perhaps the MAPK signalling pathway is already functioning at maximal capacity in gonads at this stage, meaning that overexpressing *Map3k4* is inconsequential for phosphorylation of downstream components. It also cannot be excluded that MAP3K4 is acting via a different effector MAPK, such as JNK or ERK, in this context.

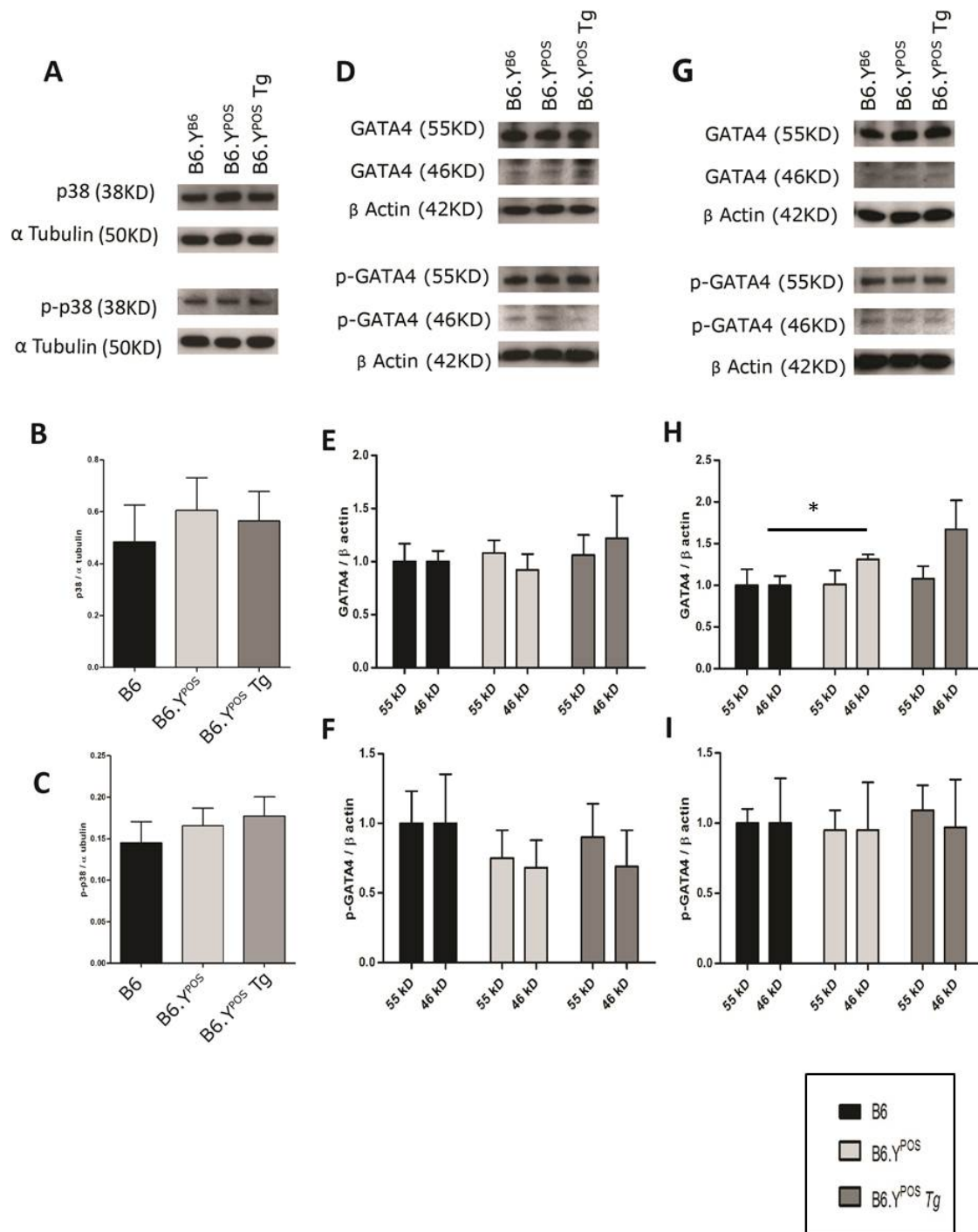


Figure 6.1 Western blot analysis of p38 and GATA4
Please see page 168 for legend.

Figure 6.1 Western blot analysis of p38 and GATA4

Western blots were performed on lysate from subdissected gonads. p38 expression was analysed in 17-18ts gonads (A-C). There was no significant difference in expression of p38 between B6, B6.Y^{POS} or B6.Y^{POS} Tg gonads. Likewise, there was no difference in levels of phospho-p38 (p-p38) between genotypes.

GATA4 expression was analysed at 15-16ts (D-F) and 17-18ts (G-I). At 15-16ts, there was no significant difference in expression of GATA4 between B6, B6.Y^{POS} or B6.YPOS Tg gonads. There was no difference in levels of phospho-GATA4 (p-GATA4) between any of the genotypes (F). At 17-18ts (G-I), the 46 KD isoform of GATA4 was significantly elevated in B6.Y^{POS} gonads compared to B6 (H, $p=0.03$). There was a trend towards increased expression of 46 kD GATA4 in B6.Y^{POS} Tg gonads compared to B6, but this was not significant ($p=0.10$). There was no difference in expression of p-GATA4 between any of the genotypes at 17-18ts (I).

Graphs show the mean of 3-5 separate experiments, each performed in triplicate. Band intensity is expressed relative to tubulin (p38) or actin (GATA4) loading control. Error bars, SEM. Statistical analysis, unpaired t test.

6.2.2. Expression of known regulators of *Sry* in B6.Y^{POS} gonads

To determine whether expression of regulators of *Sry* is altered in B6.Y^{POS} embryos, qRT-PCR was performed on subdissected gonads between 15ts and 18ts. Due to limited tissue availability, only four genes were assayed: *Wt1*, *Sf1*, *Gadd45g* and *Map3k4* (fig 6.2). *Wt1* was not significantly differentially expressed in B6.Y^{POS} or B6.Y^{POS} Tg gonads compared to B6 (fig 6.2a), although there was a slight trend towards increased expression in B6.Y^{POS} or B6.Y^{POS} Tg. *Sf1* was not significantly differentially expressed in B6.Y^{POS} gonads compared to B6 (Fig 6.2b). There was a trend towards decreased expression in B6.Y^{POS} Tg gonads, however, this was not significant. *Gadd45g* was not differentially expressed in any of the genotypes (Fig 6.2c). Likewise, *Map3k4* was not significantly differentially expressed in B6.Y^{POS} gonads (Fig 6.2b). To confirm previous reports (Warr et al., 2012, Warr et al., 2014), *Map3k4* was elevated by approximately 4-fold in B6.Y^{POS} Tg gonads.

This preliminary analysis of expression indicates that expression of components of the MAPK pathway are not dramatically altered in B6.Y^{POS} embryos. Additionally, *Wt1* and *Sf1* expression are unaltered in B6.Y^{POS} embryos. Given time to gather more tissue, it would be interesting to analyse expression of several more of the key genes implicated in *Sry* regulation, to further rule out a *trans*-effect upon *Sry* expression in B6.Y^{POS} embryos. This would be particularly informative if performed using sorted gonadal somatic cells.

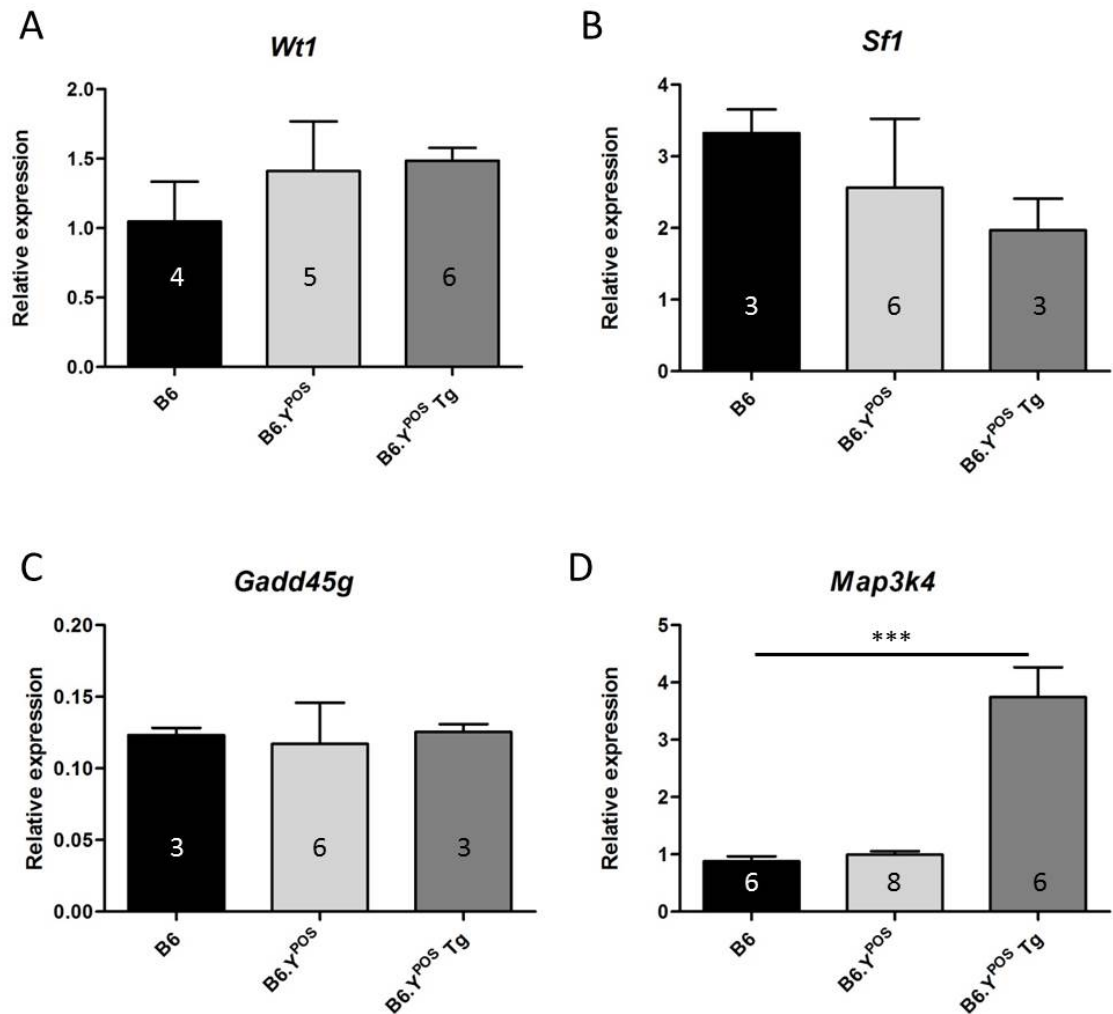


Figure 6.2 Expression of known regulators of *Sry* in B6, B6.Y^{POS} and B6.Y^{POS} Tg gonads

Expression of *Wt1*, *Sf1*, *Gadd45g* and *Map3k4* was examined by qRT-PCR in subdissected 15-18ts gonads. *Wt1* was not significantly differentially expressed in B6.Y^{POS} or B6.Y^{POS} Tg gonads compared to B6 ($p = 0.47, 0.12$ respectively), although there was a slight trend towards increased expression in B6.Y^{POS} or B6.Y^{POS} Tg. *Sf1* was not significantly differentially expressed in B6.Y^{POS} gonads compared to B6 ($p = 0.61$). There was a trend towards decreased expression in B6.Y^{POS} Tg gonads, however, this was not significant ($p = 0.09$). *Gadd45g* was not differentially expressed in any of the genotypes compared with B6 ($p = 0.84, 0.79$ respectively). *Map3k4* was not significantly differentially expressed in B6.Y^{POS} gonads ($p = 0.28$) but was significantly elevated in B6.Y^{POS} Tg gonads ($p = 0.0003$). Error bars, SEM. Statistical analysis, unpaired t test.

6.2.3. Sequencing *Sry*^{POS}

Recent reports of a non-B6 region on chromosome 11 conferring protection against sex reversal via enhanced expression of *Sox9* suggest B6.Y^{POS} animals may not be completely consomic. However, the presence of a strong *Sox9* enhancer merely circumvents delayed *Sry* expression, it does not explain it. If the strains are genetically identical other than the 110 region, it follows that any strain-specific polymorphisms resulting in delayed *Sry* expression must reside upon the Y chromosome. One key way in which *Sry*^{POS} may differ from B6 is the sequence of the putative regulatory regions. It follows that any sequence variants in these, or other as yet unidentified regions, may account for the misregulation of *Sry*^{POS} in the context of the B6 genome.

6.2.3.1. Sequencing strategy

Approximately 7 kb of the Y^{POS} chromosome, encompassing the *Sry* ORF and surrounding sequence (Fig 6.3), was sequenced by Sanger sequencing. The *Sry* ORF is located within approximately 800 bp of unique sequence, outside which is a large inverted repeat region (Gubbay et al., 1992, Dubin et al., 1995). Due to this, any sequence amplified from within the inverted repeat region will also be amplified from the reverse strand on the opposite side of the gene. It is therefore impossible to know which side sequence has been amplified from, making assembly of sequence reads challenging. To tackle this, a PCR amplification strategy was implemented using sense primers within the repeat region paired with antisense primers nested within the last few nucleotides of the unique sequence either side of *Sry*. PCR amplicons were sequenced with specific sequencing primers to achieve a read depth of at least 3. Reads were assembled in groups and the consensus sequences for each group were

then assembled together to create the final *Sry*^{POS} sequence assembly. All *Sry*^{POS} sequence reads are provided in Appendix 1.

Sequencing was performed using purified PCR products from the pooled genomic DNA of at least 5 animals. Sanger sequencing was performed by GATC as described in Section 2.6.9. Reads were assembled using SeqMan Pro (Lasergene DNASTar) and consensus sequences aligned using ClustalO version 1.2.1 (<http://www.ebi.ac.uk/Tools/msa/clustalo/>).

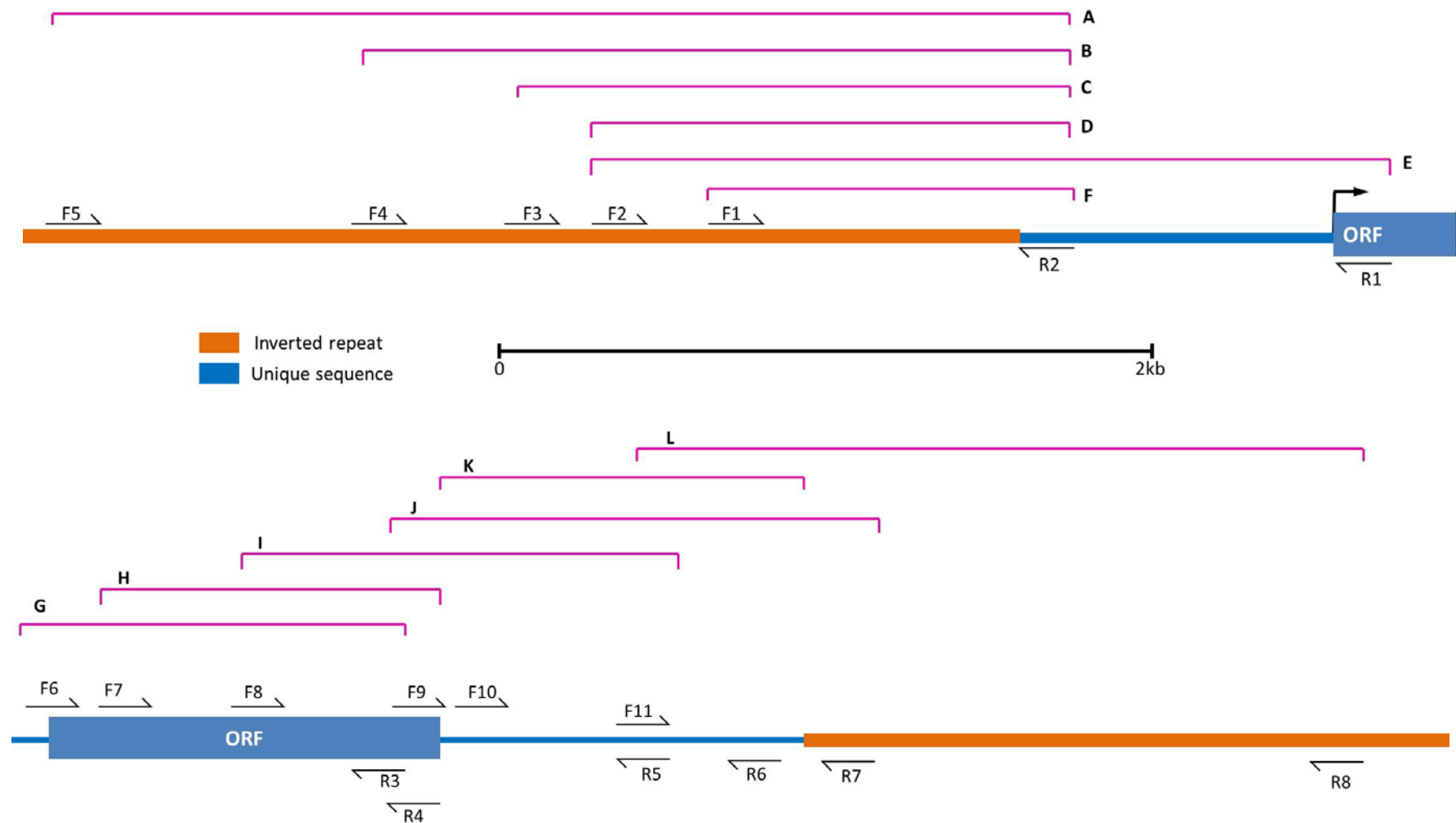


Figure 6.3 Strategy for sequencing the flanking regions of *Sry*

Locations of primers and amplified fragments of *Sry* 5' flanking region (A), ORF (B) and 3' flanking region (B). Primers are indicated by half arrows and amplicons are represented by pink brackets.

6.2.3.2. *Sry*^{POS} open reading frame

To determine similarity of *Sry*^{POS} to previously reported sequence of *Sry*^{TIR} (Carlisle et al., 1996), the consensus sequence of the *Sry*^{POS} ORF of was compared with the publicly available sequence for the Tirano strain (www.ncbi.nlm.nih.gov). Alignment of the sequences confirmed that the open reading frames are identical in these two strains (Appendix 2). Since the ORF of *Mus domesticus Sry* has already been described in detail (Coward et al., 1994, Miller et al., 1995, Carlisle et al., 1996, Albrecht et al., 2003), it will not be discussed further in this chapter.

6.2.3.3. Y^{POS} sequence assembly

A full assembly of a 7 kb fragment of Y^{POS} was produced from Sanger sequencing data.

To identify sequence polymorphisms, the assembled sequence was aligned to the corresponding B6 sequence, downloaded from the publicly available reference sequence for the C57BL/6J strain (www.ensembl.org). Polymorphisms in the 5' and 3' flanking regions are shown in figure 6.4 and summarised in Table 6.1. The full pairwise sequence alignment is provided in Appendix 3.

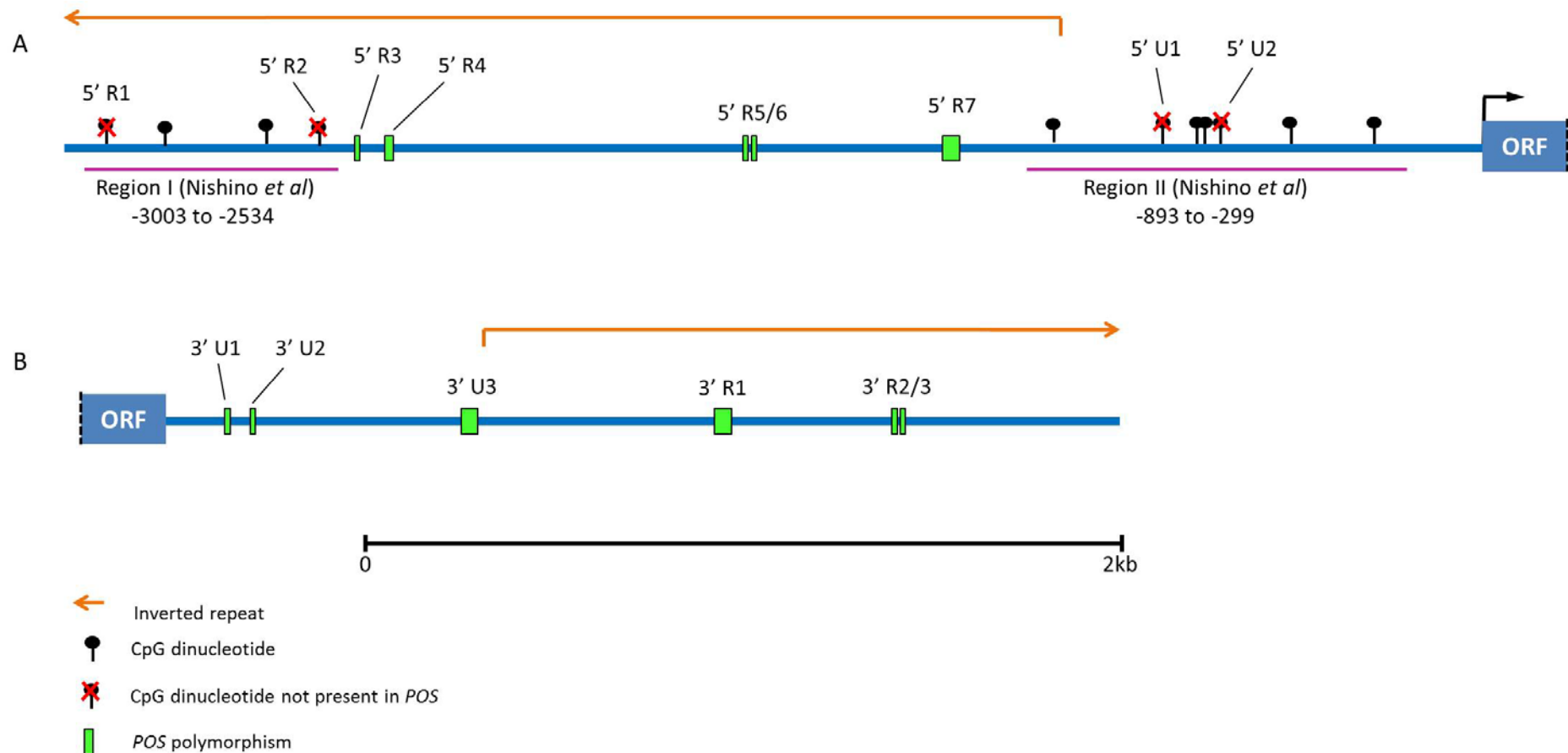


Figure 6.4 Summary of sequence polymorphisms in *Sry*^{POS} flanking regions

Schematic representation of *Sry*, showing the locations of polymorphisms identified between *Sry*^{POS} and *Sry*^{B6} in the 5' (A) and 3' (B) flanking regions. Polymorphisms are denoted 5' or 3' depending upon position relative to *Sry* ORF. The letters R and U denote polymorphisms located in the repeat region or unique region, respectively. Polymorphisms are also described in Table 6.1.

Table 6.1 *Sry*^{B6}-*Sry*^{POS} polymorphisms

Name	Position*	B6-POS	Coverage	Comments
5' R1	-2972	C-T	4x	CpG lost
5' R2	-2510	C-T	4x	CpG lost
5' R3	-2435	G-A	9x	
5' R4	-2386	GGAGGA deletion	4x	
5' R5	-1528	C-T	13x	Predicted GATA4 site lost
5' R6	-1523	C-T	13x	Predicted GATA4 site lost
5' R7	-1136	CA(9) deletion	19x	
5' U1	-648	G-T	11x	CpG lost. Also reported in Carlisle <i>et al</i> (1996)
5' U2	-536	G-A	9x	CpG lost. Also reported in Carlisle <i>et al</i> (1996)
3' U1	1331	T-C	7x	Also reported in Carlisle <i>et al</i> (1996)
3' U2	1360	A-C	7x	Also reported in Carlisle <i>et al</i> (1996)
3' U3	1886	ATATATGTGTAT deletion	3x	
3' R1	2221	GT(10) insertion	10x	
3' R2	2606	G-A	6x	Predicted GATA4 site lost (antisense strand)
3' R3	2611	G-A	6x	Predicted GATA4 site lost (antisense strand)

*Nucleotide position relative to ATG, denoting start of *Sry* open reading frame (Nishino et al., 2004)

There is minimal sequence variation between *Sry*^{B6} and *Sry*^{POS} in the region sequenced in this study. In total, within the 5.5 kb sequence flanking the ORF, 15 polymorphisms were identified. Of these, 4 are deletions and the remaining 11 are single nucleotide polymorphisms (SNPs). A total of 7 polymorphisms were identified in the 5' repeat region (5'R) and 2 in the 5' unique region (5'U). In the 3' unique region (3'U), 3 polymorphisms were identified, and a further 3 identified in the 3' repeat region (3'R). Each polymorphism identified is in a region of at least 3x sequence coverage, containing both sense and antisense reads.

6.2.3.4. Deletions in *Sry*^{POS}

In close proximity to the distal limit of the *Sry*^{B6} 5' unique region, is a group of 30 CA tandem repeats. In *Sry*^{POS}, there are 9 fewer CA repeats. The repeats are also found in complement (GT) within the 3' inverted repeat region, however, at this position there are 21 repeats in *Sry*^{B6}, and 31 in *Sry*^{POS}. To investigate similarity to other *domesticus* strains, the region containing the CA repeats was also sequenced in B6.Y^{AKR} animals using the same primers, and aligned against the *Sry*^{B6} and *Sry*^{POS} sequences (Fig 6.5). In *Sry*^{AKR}, 15 CA repeats are deleted relative to *Sry*^{B6}. This is intriguing, and suggestive of a high mutation rate in this repetitive region. These polymorphisms may either be inconsequential, or they could be linked to the differential *Sry* expression observed in the three strains. Sequence reads for *Sry*^{AKR} are provided in appendix 4.

```

AKR      TTCATAGAACAACTGGGCTTTGCACATTGTGGAGGAGAACTAAAATGTCATACACACACA
B6       TtCatAGAACAACTGGGCTTTGCACATTGTGGAGGAGAACTAAAATGTCATACACACACA
POS      TtCATAGAACAACTGGGCTTTGCACATTGTGGAGGAGAActAAAATGTcaTACACACACA
          *****

AKR      CACACACA-----CACACACACACACTCACAAG
B6       CACACACACACACACACACACACACACACACACACACACACACTCACAAG
POS      CACACACACACACAC-----ACACACAACACACACACACACTCACAAG
          *****

AKR      ACTTTACATACTCATGATGTAGTTTTGTGTGCTAGAGAGAAACCTGATAAACAGAAAGA
B6       ACTTTACATACTCATGATGTAGTTTTGTGTGCTAGAGAGAAACCTGATAAACAGAAAGA
POS      ACTTTACATACTCATGATGTAGTTTTGTGTGCTAGAGAGAAACCTGATAAACAGAAAGA
          *****

AKR      TGCATTTGATCAACAAAGGAGTGTTTTGCATACTAAAGAGACACAGCCTGTGTTTTCACA
B6       TGCATTTGATCAACAAAGGAGTGTTTTGCATACTAAAGAGACACAGCCTGTGTTTTCACA
POS      TGCATTTGATCAACAAAGGAGTGTTTTGCATACTAAAGAGACACAGCCTGTGTTTTCACA
          *****

AKR      GTATTAAGTTTGGTATGTTTTTCAAAGATGAGTATTGAAGACTTACAGATGGTTTGCATG
B6       GTATTAAGTTTGGTATGTTTTTCAAAGATGAGTATTGAAGACTTACAGATGGTTTGCATG
POS      GTATTAAGTTTGGTATGTTTTTCAAAGATGAGTATTGAAGACTTACAGATGGTTTGCATG
          *****

```

Figure 6.5 Alignment of Sry^{B6} and Sry^{POS} and Sry^{AKR} CA repeat regions

Sequence reads were assembled using SeqMan Pro. The consensus sequences were aligned using the ClustalO online tool. The region contains 30 CA repeats (highlighted in blue) in Sry^{B6} , 21 in Sry^{POS} and 15 in Sry^{AKR} .

6.2.3.5. Single nucleotide polymorphisms in Sry^{POS}

The most striking differences between the Sry^{POS} and Sry^{B6} 5' flanking regions are four polymorphisms that result in the loss of CpG dinucleotides in Y^{POS} . The 2 most proximal, 5' U1 and 5' U2, were previously reported by Coward *et al* (1994), but their functional relevance has not since been investigated. Both these CpG dinucleotides are members of the 7 CpGs comprising the *Sry* T-DMR proximal to the linear promoter of *Sry* (Nishino et al., 2004). Therefore, their absence in Sry^{POS} could be of functional significance.

The 2 more distal CpG-ablating polymorphisms are housed within the 5' inverted repeat (5' R1 and 5' R2). Both these CpGs are part of the T-DMR that is proximal to the circular promoter of *Sry* (Nishino et al., 2004), meaning that their absence in Sry^{POS}

may be important for the regulation of *Sry*. Loss of CpG dinucleotides is likely to impact levels of promoter methylation, which may then be associated with delayed initiation of expression. These polymorphisms may therefore underlie delayed *Sry* expression in B6.Y^{POS} embryos.

To investigate whether these ablated CpGs could be related to delayed *Sry* expression, the same region of *Sry*^{AKR} was sequenced (Fig 6.6). B6.Y^{AKR} embryos exhibit transient sex reversal associated with only modestly delayed *Sry* expression. All four of the CpGs found to be absent in Y^{POS} were also absent in Y^{AKR}, suggesting that changes to these CpGs may not be sufficient to explain the differential expression profiles of *Sry* in B6.Y^{AKR} and B6.Y^{POS} embryos.

```

AKR      GTGATGCTGTAAGGAAATGACAGCCAGCCCTGACCCGATGGCACTGTGTTTTGCCTGCT
B6       GTGATGCTGTAAGGAAATGACAGCCAGCCCTGACCCGATGGCACTGTGTTTTGCCTGCT
POS      GTGATGCTGTAAGGAAATGACAGCCAGCCCTGACCCGATGGCACTGTGTTTTGCCTGCT
          *****

AKR      CAGAAATGAACACTACTGCATCCCAGTCATTAGAATGCTCAACCTGGATGATTTTTACCATG
B6       CAGAAATGAACACTACTGCATCCCAGTCATTAGAATGCTCAACCTGGATGATTTTTACCATG
POS      CAGAAATGAACACTACTGCATCCCAGTCATTAGAATGCTCAACCTGGATGATTTTTACCATG
          *****

AKR      AAAGAATTTTGGCTATTATGTAAGAAATTTGTATACATTACTTTTATGATGCATTGATAT
B6       AAAGAATTTTGGCTATTATGTAAGAAATTTGTATACATTACTTTTATGATGCATTGATAT
POS      AAAGAATTTTGGCTATTATGTAAGAAATTTGTATACATTACTTTTATGATGCATTGATAT
          *****

AKR      GACCAAGAAAGGAAGCAAAAGTTCTTTTATGGTGTAGATATTTATACTGGGCTATAGAGC
B6       GACCAAGAAAGGAAGCAAAAGTTCTTTTATGGTGTAGATATTTATACTGGGCTATAGAGC
POS      GACCAAGAAAGGAAGCAAAAGTTCTTTTATGGTGTAGATATTTATACTGGGCTATAGAGC
          *****

AKR      CAGAACAGACTATTTCTTAGTTTTTCCATGCTGCCTGCCTTAGTAGACTATGATACTAA
B6       CAGAACAGACTATTTCTTAGTTTTTCCATGCTGCCTGCCCGTAGTAGACTATGATACTAA
POS      CAGAACAGACTATTTCTTAGTTTTTCCATGCTGCCTGCCTTAGTAGACTATGATACTAA
          *****

AKR      TGATAGGTTCCATGCTTATCATTTTTAACCCTTTCCACTACTTTTGCAAGCTCGTTAATT
B6       TGATAGGTTCCATGCTTATCATTTTTAACCCTTTCCACTACTTTTGCAAGCTCGTTAATT
POS      TGATAGGTTCCATGCTTATCATTTTTAACCCTTTCCACTACTTTTGCAAGCTCGTTAATT
          *****

AKR      TCTCACGTAGTCCATCCTAAACATTAGGTACTTTACAGGGCAAAGAACTCATCAGTAT
B6       TCTCACGTAGTCCATCCTAAACATTAGGTACCGTTACAGGGCAAAGAACTCATCAGTAT
POS      TCTCACGTAGTCCATCCTAAACATTAGGTACTTTACAGGGCAAAGAACTCATCAGTAT
          *****

AKR      GTTCTTTGTATCTTAATACTTGTGTGACAGAATGATCAAAGAAAAAGCGATATATTTTT
B6       GTTCTTTGTATCTTAATACTTGTGTGACAGAATGATCAAAGAAAAAGCGATATATTTTT
POS      GTTCTTTGTATCTTAATACTTGTGTGACAGAATGATCAAAGAAAAAGCGATATATTTTT
          *****

AKR      AAGGAGTGTGAAAGTGCAGTCTCTAACATTACAAACTTGAGGGGTTTTGCTTGACAGTA
B6       AAGGAGTGTGAAAGTGCAGTCTCTAACATTACAAACTTGAGGGGTTTTGCTTGACAGTA
POS      AAGGAGTGTGAAAGTGCAGTCTCTAACATTACAAACTTGAGGGGTTTTGCTTGACAGTA
          *****

AKR      TCTAGGTTTCATGGGATGGCTTCAAAGTAGATTAGTTGGGCTGGACTAGGGAGGTCCTGAA
B6       TCTAGGTTTCATGAGATGGCTTCAAAGTAGATTAGTTGGGCTGGACTAGGGAGGTCCTGAA
POS      TCTAGGTTTCATGGGATGGCTTCAAAGTAGATTAGTTGGGCTGGACTAGGGAGGTCCTGAA
          *****

AKR      GGAGGAGGGATAAATATTTTCTTACACACGTAAATATTAAAACTTTTGTACAAGCACA
B6       GGAGGAGGGATAAATATTTTCTTACACACGTAAATATTAAAACTTTTGTACAAGCACA
POS      GGAGGAGGGATAAATATTTTCTTACACACGTAAATATTAAAACTTTTGTACAAGCACA
          *****

```

Figure 6.6 Alignment of sequences from *Sry*^{B6}, *Sry*^{AKR} and *Sry*^{POS} containing CpG dinucleotides of the *Sry* T-DMR

Sry^{AKR} and *Sry*^{POS} sequence reads were assembled using SeqMan Pro and consensus sequences for each strain were aligned using the ClustalO online tool. CpG dinucleotides are underlined and coloured red. Polymorphisms are highlighted in blue. At the location of the 2nd and 5th CpG in *Sry*^{B6} there are not CpG dinucleotides in either *Sry*^{POS} or *Sry*^{AKR}.

A further two polymorphisms of interest are 5' R5 and 5' R6, which are only 5 bp apart. That these polymorphisms exist in such close proximity was highly intriguing, so a 47 bp sequence containing this mutated region was analysed using the JASPAR (<http://jaspar.binf.ku.dk/>) database, which scans for transcription factor binding motifs. Analysis of the B6 variant predicted binding sites for GATA4, SOX9 and SRY (Table 6.2). When the Y^{POS} variant of the sequence was scanned, the GATA4 site was no longer predicted and an SRY binding site was altered, although still predicted (iv, v). The mutated sequence also led to prediction of additional SOX9 and SRY binding sites on the reverse strand (Table 6.2 iii,vi).

Table 6.2 Predicted transcription factor binding

Number	Model ID	Name	JASPAR score	Strand	Predicted sequence	B6	POS
i	MA0482.1	GATA4	4.729	1	CCTTAGCCTTAA	Y	-
ii	MA0077.1	SOX9	6.196	-1	AGATTGTTA	Y	Y
iii	MA0077.1	SOX9	8.502	1	CCATTCTTT	-	Y
iv	MA0084.1	SRY	10.294	1	CTTAACAAT	Y	-
v	MA0084.1	SRY	10.854	1	TTTAACAAT	-	Y
vi	MA0084.1	SRY	6.453	-1	CTAAAGAAT	-	Y
Sequence alignment		B6 POS	GTGGGTAAC TTTGAACCATTCCTTAGCCTTAACAATCTGGCAGTTGAG GTGGGTAAC TTTGAACCATTCCTTAGCTTTAACAATCTGGCAGTTGAG *****				

Polymorphisms are highlighted in blue. An alignment of the region scanned using JASPAR is shown below the table.

To assess whether this may be of functional importance, the same region was also sequenced in Y^{AKR} (Fig 6.7). Interestingly, Y^{AKR} possessed the B6 variant at each of the two SNP sites, suggesting that this polymorphism may be unique to Y^{POS}.

AKR	AACTTGTTTGTAAACAAGAGTGGGTAACTTTGAACCATTCCTTAGCTTAACAATCTGG
B6	AACTTGTTTGTAAACAAGAGTGGGTAACTTTGAACCATTCCTTAGCTTAACAATCTGG
POS	AACTTGTTTGTAAACAAGAGTGGGTAACTTTGAACCATTCCTTAGCTTAACAATCTGG

Figure 6.7 Alignment of Y^{AKR}, Y^{B6} and Y^{POS} fragments containing polymorphisms 5' R5 and 5' R6

Sequence reads were assembled using Seqman Pro (Lasergene DNA Star) and aligned using Clustal Omega. Whilst nucleotides at positions 5' R5 and 5' R6 (highlighted) are cytosine in Y^{AKR} and Y^{B6}, both are thymine in Y^{POS}.

Polymorphisms 5'R5 and 5'R6 reside in the 5' repeat region, and therefore are recapitulated in complement in the 3' repeat region. Polymorphisms 3'R6 and 3'R5 correspond to the complement of 5'R5 and 5'R6, respectively. This means that the predicted GATA4 consensus site also resides upon the antisense strand at this position in the 3' repeat, and therefore that the site is also abolished in B6.Y^{POS}. The potential functional importance of this is highly intriguing and warrants further investigation.

That there are minimal sequence polymorphisms in the flanking sequence of *Sry* could mean two things. Firstly, it could mean that those differences that are present are important for the regulation of *Sry* and its differential expression in B6.Y^{POS} vs B6. Alternatively, it could mean that the most important sequence lies elsewhere, perhaps further upstream in an as-yet unidentified enhancer. However, the observation that, of the few polymorphisms that are present, so many alter CpG dinucleotides, is

suggestive of some functional importance. It is necessary to further examine the changes identified in this study to determine whether they could be significant, or whether sequence analyses should search more distally.

6.2.4. Analysis of CpG methylation in B6.Y^{POS} gonadal somatic cells

Having identified several intriguing CpG polymorphisms in *Sry*^{POS}, it was important to assess whether CpG methylation in the *Sry* T-DMR differed between B6 and B6.Y^{POS} at the crucial time of *Sry* expression in the embryonic gonad. This would be the first information gathered about the epigenomic landscape of *Sry*^{POS} at the crucial time of *Sry* expression. Additionally, analysis of CpG methylation of this region, in a strain that is missing some of the CpGs, may enable us to estimate the relative importance of individual CpG dinucleotides. In particular, it will help to estimate whether the CpGs that are absent in Y^{POS} and Y^{AKR} affect methylation of those remaining.

6.2.4.1. Collection of gonadal somatic cells by FACS

In order to achieve a sample of purified gonadal somatic cells, an *Sf1-EGFP* reporter was used to sort somatic cells from gonads. To generate the line, B6.Y^{POS} Tg males were crossed with *Sf1-EGFP* females to generate B6.Y^{POS}-*Map3k4* Tg-*Sf1-EGFP* Tg studs. These were then crossed with B6 females to generate experimental embryos. Gonads from B6, B6.Y^{POS} and B6.Y^{POS} Tg embryos were collected at 17-18ts. Cells from dissociated gonads were sorted by FACS, as described in Section 2.5.

6.2.4.2. Gene expression profiles of sorted cells

To confirm that FACS-sorted samples were enriched for somatic cells, expression of genes associated with somatic cells was assayed by qRT-PCR. If gonadal somatic cells

were purified, it would be expected that markers of gonadal somatic cells such as *Sox9* and *Sry* would be enriched in GFP-positive samples compared with GFP-negative. Conversely, germ cell markers such as *Oct4* are predicted to be enriched in GFP-negative samples. At the time of extracting gDNA from the sorted cells used for methylation analysis, RNA was also purified and used to synthesise cDNA. Additionally, expression was also analysed in samples of subdissected gonad and mesonephros, collected as per Section 2.3.3.

To first confirm enrichment of *Sf1-EGFP*-expressing cells, expression of *Sf1* was analysed (Fig 6.8). As predicted, *Sf1* was enriched in *EGFP*-positive samples. *Sf1* was also markedly enriched in *EGFP*-positive cells compared with subdissected gonads, which is expected because the subdissected gonad contains a mixed population of somatic cells, germ cells and epithelial cells. Expression of *Sry* was also enriched in all *EGFP*-positive samples compared with the corresponding *EGFP*-negative cells. *Sox9* was enriched in B6 and B6.Y^{POS}-Tg *EGFP*-positive cells but not in B6.Y^{POS} cells. This is reflective of the significantly reduced *Sox9* expression observed in B6.Y^{POS} gonads at this stage (Section 5, Fig 5.4). Unfortunately, due to technical issues with the assay, it was not possible to obtain expression data for *Oct4*, a primordial germ cell marker.

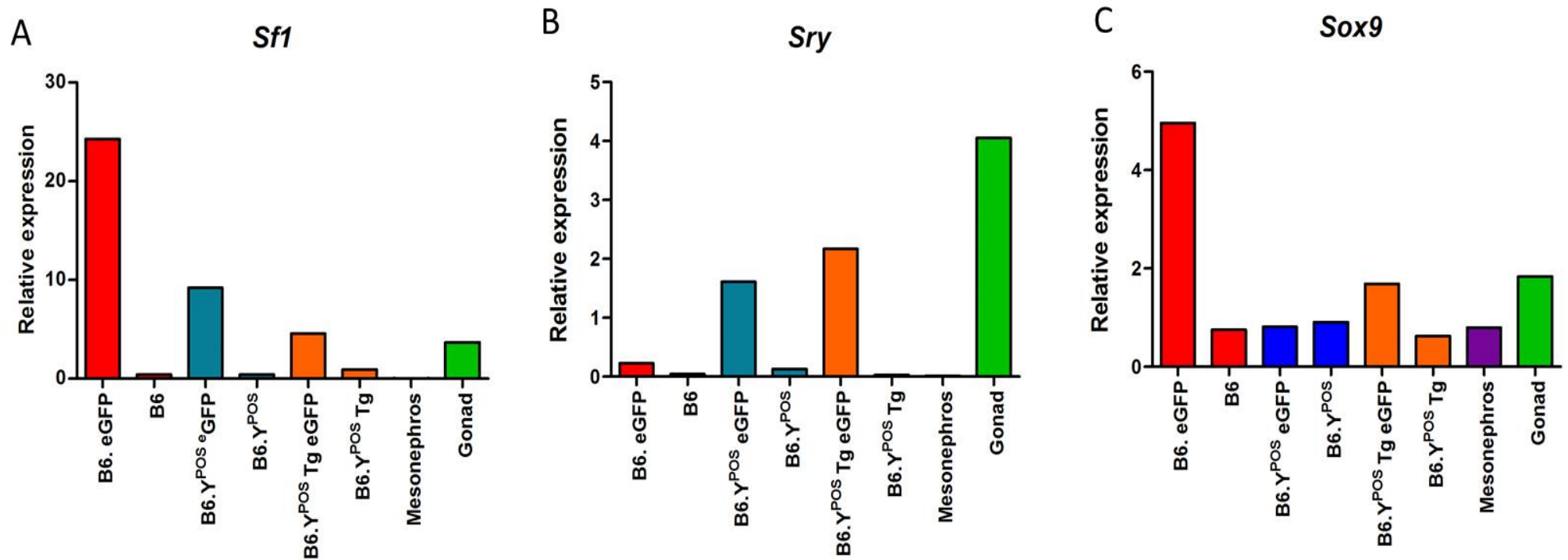


Figure 6.8 qRT-PCR analysis of sorted cells

qPCR was performed on cDNA synthesised from RNA extracted from FACS-sorted cells or from subdissected gonads and mesonephroi. *Sf1* (A) was enriched in the *EGFP*-positive populations of cells from B6, B6.Y^{POS} and B6.Y^{POS} Tg gonads, as well as in subdissected gonads, and was not detected in mesonephros. *Sry* (B) was also enriched in *EGFP*-positive cells from B6, B6.Y^{POS} and B6.Y^{POS} Tg gonads and subdissected gonads, and was not detected in mesonephros. *Sox9* (C) was enriched in B6 *EGFP*-positive cells and somewhat enriched in B6.Y^{POS} Tg *EGFP*-positive cells but not in B6.Y^{POS} *EGFP*-positive cells, and was detected in both gonad and mesonephros.

6.2.4.3. Bisulphite sequencing

Extracted DNA from sorted cells was then subjected to sodium bisulphite conversion and PCR amplification of the 537 bp region containing 6 CpG dinucleotides, as described by Warr *et al* (2012) (Fig 6.9a). PCR, subsequent ligation, cloning and DNA sequencing were performed as described in Section 2.6. Although up to 90 colonies were picked following ligation and cloning of bisulphite-treated DNA from each genotype, very few clones were successfully sequenced. Additionally, due to unsuccessful bisulphite conversion of DNA from B6.Y^{POS}-Tg cells, it was not possible to proceed to ligation and cloning for this genotype, so the data below refer only to B6 and B6.Y^{POS}.

Fig 6.9 shows methylation analysis for the reads that were produced in the time available. The most telling piece of information from this experiment is an indication of the relative importance of each CpG dinucleotide in this region. In particular, CpGs 2 and 5, which are not CpGs in *Sry*^{POS}. These two CpGs are largely unmethylated in *Sry*^{B6} at 17-18ts. At CpGs 2 and 5, it is the guanine component of the CpG that is substituted in *Sry*^{POS}, leaving behind a non-CpG cytosine. Analysis of the converted sequence reveals that these cytosines are unmethylated in *Sry*^{POS} (Fig 6.9b). Although it is unclear how an unmethylated CpG differs from a non-CpG cytosine in the context of gene regulation, these data suggest that CpGs 2 and 5 may be dispensable for timely expression of *Sry*. The finding that these two CpGs are also absent in *Sry*^{AKR}, which is less delayed than *Sry*^{POS}, supports this.

When B6 and B6.Y^{POS} were compared, there were no striking differences in methylation of any of the CpG dinucleotides analysed in this initial experiment (Fig

6.9c). There was a trend towards hypermethylation of the 1st and 6th CpGs in B6.Y^{POS}, but with so few reads it is not possible to determine whether this is significant.

It was not possible to generate further bisulphite sequencing data due to severe time constraints.

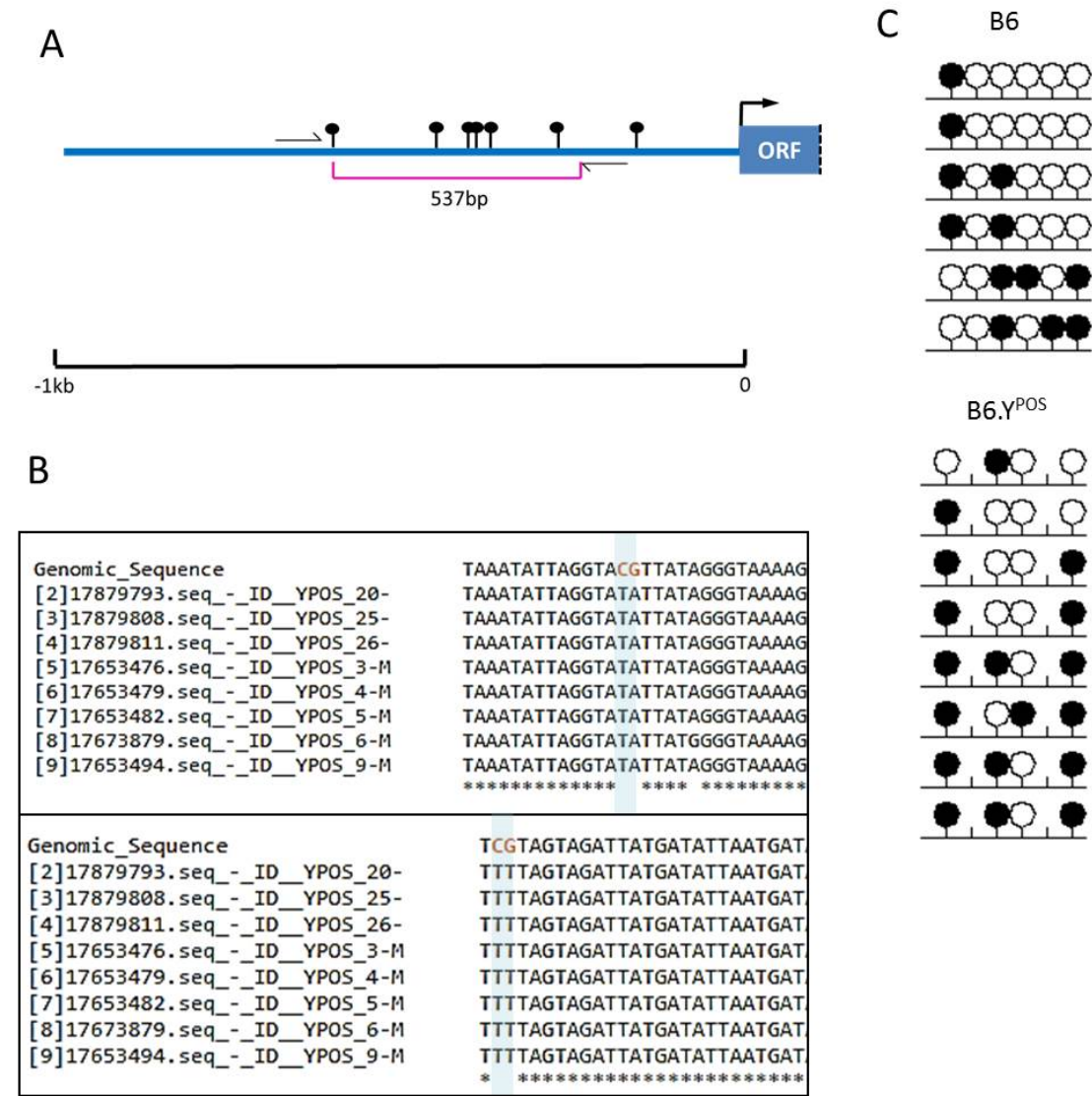


Figure 6.9 Analysis of methylation of the Sry T-DMR in B6.Y^{POS} gonadal somatic cells
 Bisulphite sequencing was performed on DNA from 17-18ts FACS-sorted gonadal somatic cells. (A) Location of the sequenced region, which includes 6 CpG dinucleotides in B6. (B) Sequence analysis of bisulphite-treated B6.Y^{POS} DNA confirms that 2 of the CpGs in B6 are not CpGs in Sry^{POS}. The cytosines of these missing CpGs are unmethylated as they have been converted to uracil during bisulphite treatment. (C) Preliminary comparison of the methylation of this region between B6 and B6.Y^{POS} does not reveal any striking differences.

6.2.5. Assaying for presence of the 110 region in Harwell

B6.Y^{POS} mice

To determine whether the 110 region identified by Arboleda *et al* (2014) is present in the Harwell B6.Y^{POS} colony, genomic DNA was extracted from ear biopsies of all available XY animals. DNA was sequenced using primers that were kindly provided by V. Arboleda (David Geffen School of Medicine, UCLA, California), which amplify fragments containing POSA-specific SNPs located within the 110 region

Table 6.3 Assays to determine presence of the 110 region

dbSNP ID	Assay name	B6	POSA	Search string
rs27019103	POSA_SNP1	C	T	TGGAGG
rs28240850	POSA_SNP2	G	A	AGGGAAG
rs27029785	POSA_SNP3	C	A	AAGGCT

Purified PCR product was sent to GATC for Sanger sequencing. In total, 35 Harwell B6.Y^{POS} animals were tested. As a negative control, a B6 male animal was also sequenced for all three SNPs. Control DNA from two UCLA B6.Y^{POS} mice, one homozygous (110 hom) and one heterozygous (110 het) for the 110 region, was kindly provided by V. Arboleda.

All Harwell B6.Y^{POS} animals sequenced were found to be homozygous B6 for all 3 SNPs (Fig 6.10) Sequencing of DNA from the previously validated heterozygous and homozygous 110 animals confirmed that the assay used was able to identify both. Therefore, it can be concluded that the Harwell B6.Y^{POS} colony does not harbour the 110 region.

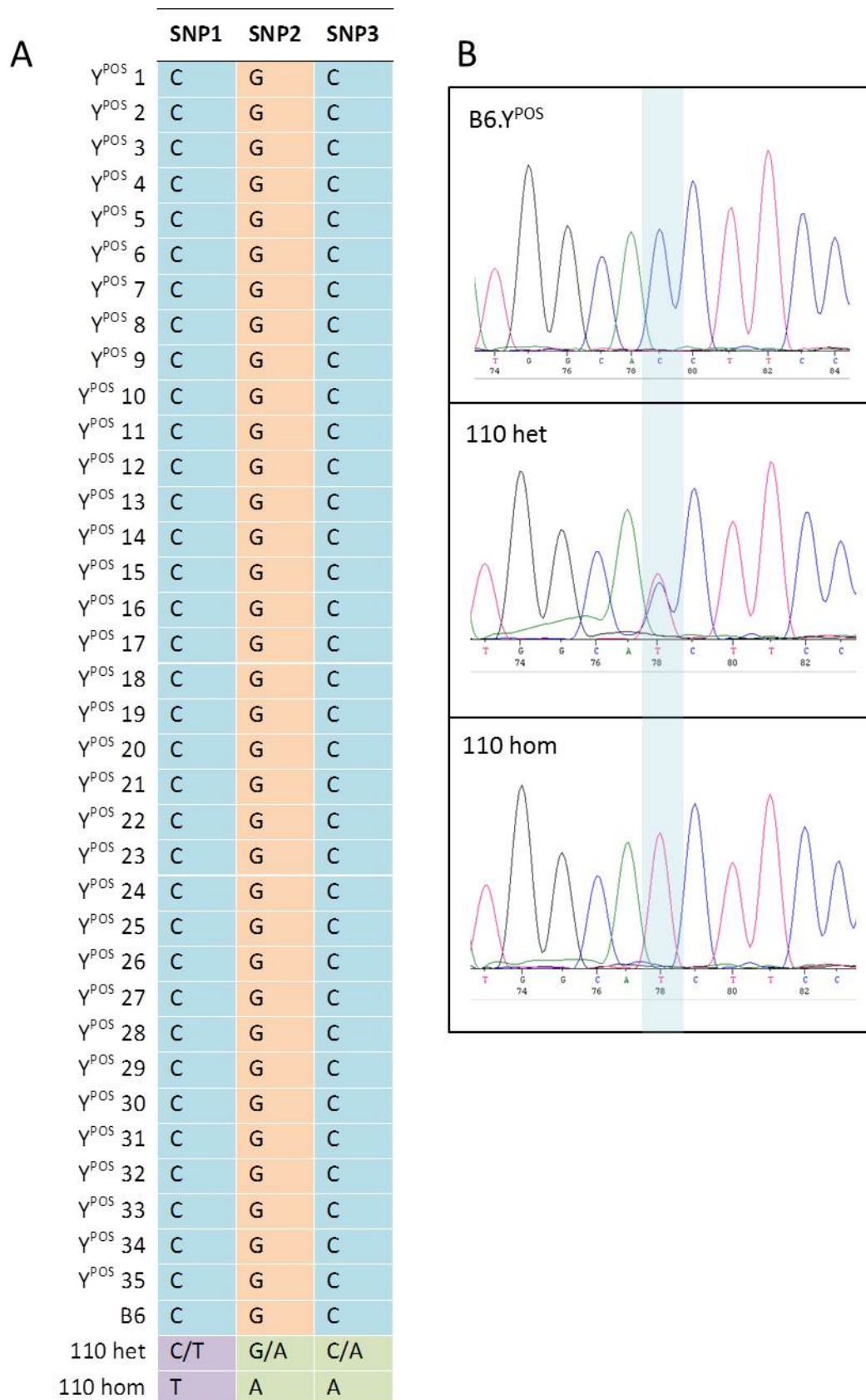


Figure 6.10 Genotyping for the 110 region in Harwell B6.Y^{POS} mice

Please see page192 for legend

Figure 6.10 Genotyping for the 110 region in Harwell B6.Y^{POS} mice

(A) Animals were genotyped by PCR amplification and sequencing of 3 regions, each containing a SNP associated with the 110 region (Arboleda et al 2014). All 35 B6.Y^{POS} animals tested possessed the B6 variant at all 3 positions. A B6 animal was also genotyped to confirm this. As a positive control for the sequencing assay, DNA from one heterozygous (110 het) and one homozygous (110 hom) animal from the colony at UCLA was also sequenced. (B) Representative chromatograms for an example SNP (SNP2). Analysis of chromatograms confirms it is possible to differentiate a heterozygote from a homozygote at this position. Note: These chromatograms are in reverse complement as the region was sequenced using an antisense primer.

6.2.5.1. Next-generation sequencing

Since the Harwell B6.Y^{POS} colony does produce fertile male animals, it was postulated that they may possess an alternative region of non-B6 origin that protects against sex reversal and promotes fertility. This may have been introduced historically, during outcrosses with other strains and maintained by selection of fertile males, or alternatively could also be derived from POSA. To identify any such region, genomic DNA from a proven fertile male was sent to the High Throughput Genomics facility (University of Oxford) for next-generation sequencing.

Analysis of sequence reads from NGS was performed by the Bioinformatics team at MRC Harwell (Dr Michelle Simon and Dr Saumya Kumar). Sequence data was aligned to the B6 reference genome and compared against panels of known strain-specific SNPs. At the time of writing, a cluster of non-B6 SNPs has been identified on chromosome 13. In the interest of time, further investigation of this region was passed on to Dr Nick Warr and validation is in progress.

6.3. Discussion

In this chapter, the mechanism of restored *Sry* expression in B6.Y^{POS} embryos was investigated. Additionally, evidence of a minimal level of variation between the flanking regions of *Sry*^{B6} and *Sry*^{POS} is presented.

6.3.1. p38 and GATA4 phosphorylation are not affected by overexpressing *Map3k4*

The lack of any observable change in levels of either phospho-p38 or phospho-GATA4 in B6.Y^{POS}Tg gonads (Fig 6.1) suggests that MAP3K4 signalling via these two molecules is not affected by overexpressing *Map3k4*. This could be because, at this stage, all available MAP2K, p38 MAPK or GATA4 is already phosphorylated, creating a bottleneck and meaning that increased amounts of MAP3K4 in the cell are inconsequential for this

pathway. Alternatively, MAP3K4 signalling via another MAPK may be increased, such as ERK or JNK MAPKs. In order to investigate this it would be necessary to perform further Western blots to quantify phosphorylation of these molecules. In addition, immunofluorescence on sections of gonads at the two stages examined (16-16ts and 17-18ts) may indicate whether other changes in these molecules are associated with overexpression of *Map3k4*, such as altered cellular localisation.

6.3.2. *Wt1* and *Sf1* are not differentially expressed in B6.Y^{POS} gonads, but GATA4 is elevated

Analysis of expression of *Wt1* and *Sf1* suggests a subtle trend towards altered expression in B6.Y^{POS} gonads (Fig 6.2). It is necessary to repeat this experiment to confirm these findings, but is suggestive of subtle differences in the expression of genes upstream of *Sry*. Likewise, GATA4 was significantly elevated in B6.Y^{POS} gonads at 17-18ts, and was also elevated, although not significantly, in B6.Y^{POS} Tg gonads (Fig 6.1). This suggests that there may be some other Y-linked factor, differing between Y^{POS} and Y^{B6}, that drives expression of GATA4. Alternatively, it may be the result of a feedback mechanism, whereby delayed *Sry* expression somehow results in cells trying to compensate by upregulating GATA4. However, before investigating a possible mechanism, it would first be necessary confirm whether these observations are seen at the transcript level, by analysing expression of *Gata4* by qRT-PCR.

Given the discordance between these variations in expression, it is difficult to suggest an explanation. However, if they are confirmed in repeated experiments, it suggests that there is another Y-encoded factor influencing transcription of genes upstream of *Sry*. However, in a recent comprehensive analysis of transcription during early stages of sex determination, no gene on the Y chromosome has been reported to be

expressed (Munger et al., 2013). Nevertheless, these are intriguing findings that warrant further investigation.

6.3.3. There are modest sequence differences between the *Sry*^{B6} and *Sry*^{POS} 5' flanking regions

Sequencing the flanking regions of *Sry*^{POS} has revealed a small number of polymorphisms. Polymorphisms that affect CpG dinucleotides have previously been described in *Y*^{POS} and in the outbred CD1 strain (Carlisle et al., 1996, Nishino et al., 2004), and in this study have been shown to be shared with *Y*^{AKR}. Since CD1 mice are derived from a large breeding stock, making it difficult to determine a single origin of their Y chromosome, the absence of these CpG dinucleotides in this strain is not very informative. That they are also missing in another *domesticus* strain indicates that this characteristic may be shared among *domesticus* Y chromosomes. A logical next step would be to sequence the region containing these CpG dinucleotides in a variety of other inbred mouse strains and *domesticus*-derived strains. The Mary Lyon Centre maintains colonies of many inbred strains, making DNA readily available, whereas obtaining DNA from *domesticus*-derived animals would require obtaining tissue from groups that currently study them.

The second potentially significant finding is the close proximity of the 5'U 1 and 5'U 2 polymorphisms (Table 6.2). These two SNPs are present in *Y*^{POS} but not *Y*^{AKR}, suggesting that they may be able to account for the difference in *Sry* expression in the two strains. However, a great deal of work is required to determine whether these polymorphisms are of any significance. This would require, firstly, ascertaining whether the predicted transcription factor binding sites actually bind GATA4, SRY or SOX9 *in vitro* as well as *in vivo*, a challenging undertaking that would require chromatin immunoprecipitation

from minute amounts of gonadal tissue. However, initial studies to ascertain whether this polymorphism occurs in any other mouse strains may help to direct interest, by determining whether its presence or absence correlates with *Sry* expression. It could be argued that, in order to do this, Y chromosomes from these other inbred strains would first need to be congenic on the B6 background, to give a comparable picture of *Sry* expression in the same context. However, it has been suggested that *Sry*^{POS} also shows a delayed expression profile (relative to B6) on a mixed background, suggesting that there is an innate feature of the Y^{POS} chromosome that leads to a later onset, longer duration expression pattern (Palmer and Burgoyne, 1991).

Finally, the polymorphism in the region of CA repeats is also of some interest. This region was found to vary in all three strains examined, however there does not appear to be any correlation with severity of sex reversal. There are fewer CA repeats in Y^{AKR} than there are in Y^{POS}, but Y^{B6} has more than both of them. This region has been suggested as a putative site of SRY binding to its own promoter *in vitro* (Vriz et al., 1995) and therefore may be important for SRY autoregulation.

Together, the parts flanking regions of *Sry*^{B6} and *Sry*^{POS} sequenced in this study indicate little variation that could obviously account for the severely delayed expression reported in B6.Y^{POS} embryos. More work is required to determine whether any of the small number of polymorphisms identified are linked with delayed *Sry* expression. In addition, these findings suggest that there may be an as-yet unidentified enhancer of *Sry*, located at a more distal location.

6.3.4. A preliminary study of *Sry*^{POS} methylation

This study has provided a very preliminary look at the methylation of *Sry* in 11.5 dpc B6.Y^{POS} gonadal somatic cells. Whilst currently there are no clear differences between

methylation of this region in B6.Y^{POS} compared with B6, only one experiment was performed and a small number of sequence reads were obtained. The experiment must be repeated to give a clearer picture of *Sry* methylation at this stage. In addition, since *Sry* expression is dynamic, it might be expected that differences in dynamics of its methylation are observed over time. Therefore, to build a complete picture of *Sry* methylation in B6.Y^{POS}, it is important to examine at least two stages of development. By doing so, it would indicate whether any observed differences are resolved over time, as *Sry* reaches its later peak of expression at around 22ts. In addition to this, examining *Sry* methylation in B6.Y^{POS} Tg gonads may also indicate whether any observed differences are overcome by overexpressing *Map3k4*. Not only would this help to explain delayed *Sry* expression in B6.Y^{POS}, it may also suggest a mechanism for its rescue by *Map3k4* overexpression.

6.3.5. The Harwell B6.Y^{POS} colony does not possess the protective 110 region

Genotyping for the 110 region of chromosome 11 has indicated that the Harwell B6.Y^{POS} colony does not possess this region. However, NGS indicates that it may possess a different non-B6 region on chromosome 13. Validation of this region is in progress but its presence suggests that the Harwell B6.Y^{POS} colony and others such as the UCLA B6.Y^{POS} colony have diverged through generations of selective breeding. The two regions may have originated from the same POSA strain, but perhaps have been segregated at some point during selection for fertile males. Alternatively Harwell B6.Y^{POS} mice may have been outcrossed to maintain sex ratios at some point during their long history, which spans over 40 years (Cattanach et al., 1972). Whilst no record of this has yet been found (N. Warr, personal communication), the possibility cannot be excluded.

Following validation of the putative protective region, analysis of whether its presence affects the likelihood of B6.Y^{POS} sex reversal can ensue. In order to do this the colony must be segregated and the region eliminated from one group. Very preliminary investigations have identified some B6.Y^{POS} males that do not possess the region, suggesting that in some cases they are able to develop testes. However, these mice have not sired pups so may be infertile, although testing of this has not yet been performed (N. Warr, personal communication). Given the possibility that males lacking the region may not be fertile, generation of a segregated colony lacking it may require maintenance via the sex-reversal-rescuing *Map3k4* transgene. This would enable generation of studs lacking the region, which therefore would not pass it on to their offspring. These offspring could be analysed at 14.5 dpc for disruptions to testis determination and molecular studies performed at earlier stages.

Furthermore, following validation of the region it is necessary to identify candidate genes that could account for protection against B6.Y^{POS} sex reversal. These may be pro-testis genes with enhanced expression or pro-ovarian genes with reduced expression. Narrowing down the region by generating lines with varying-sized fragments of the non-B6 region may help to elucidate a critical region and enable identification of a minimal number of candidate genes for further investigation.

Chapter 7. General Discussion

7.1. Overview

This thesis has examined the role of MAPK signalling in testis determination, in an endeavour to further elucidate how MAPK-related pathways are involved in the regulation of testis determination and, more specifically, *Sry* expression. This is a system of ever-growing complexity. Since *SRY* was first discovered in 1990, much focus has been placed on how *SRY* coordinates expression of its own targets, as well as how *Sry* itself is regulated. A growing number of genes have been implicated in the control of its expression, but we are yet to identify a direct link between a regulatory molecule and the transcription of *Sry*. It is likely that multiple processes act in concert to ensure the appropriate expression of *Sry* in the gonad, and within this thesis, one particular process has been investigated.

The main aims of this thesis were to further elucidate the role of MAPK-related signalling in testis determination, with two distinct directions of focus forming the body of results. The first, transcriptomic analysis of potential targets of *GADD45γ*, has identified novel genes that are dependent on *Gadd45γ* for gonadal expression. The second has delved into the complex phenomenon of B6.Y^{DOM} sex reversal and confirmed *Map3k4* as a dose-dependent modifier of sensitivity to sex reversal in this context.

7.2. Key findings

7.2.1. *Pnliprp2* is expressed in the 11.5 dpc gonad and is downregulated in gonads lacking *Gadd45g*

The undertaking of a transcriptome-wide analysis of the developing mouse gonad was not without significant challenges and few differentially expressed were identified, potentially due to the early stage of gonadogenesis at which comparisons were made. However, two differentially expressed genes were validated by qRT-PCR: *Tgm3* and

Pnliprp2. Characterisation of the expression of *Tgm3* by WMISH remains to be performed, but *Pnliprp2* was found to be expressed in a punctate pattern reminiscent of a marker of gonadal somatic cells at this stage, and is an interesting candidate for further characterisation.

The data presented in Chapter 4 highlights the importance of sample composition when performing whole-transcriptome analysis of gonads during testis-determination. In this experiment, a sample containing gonad and mesonephros was used, and therefore contained a broad mix of cell types including germ cells, epithelial cells, mesenchyme, connective tissue and rare somatic supporting cell and steroidogenic cell progenitors. Since the cells expressing *Gadd45g* are within a relatively rare population, transcript abundance of any GADD45γ targets is likely to be low. This enhances the possibility of detection of false positives and discordance of trends in expression when validating genes by qRT-PCR.

Furthermore, due to constraints on the cost of RNA-seq, only one sample per genotype was sequenced. Although this contained pooled RNA of up to 20 pairs of gonads, the absence of a replicate made the use of bioinformatic tools to detect differential expression difficult (M. Simon, personal communication). Therefore, in future, samples should be sequenced at least in duplicate, to better facilitate reliable detection of differential expression.

7.2.2. *Map3k4* modifies sensitivity to B6.Y^{POS} sex reversal via an effect on *Sry* expression

In chapter 5, rescue of B6.Y^{POS} sex reversal by a *Map3k4* transgene is described. Comprehensive transcriptional profiling showed that overexpression of *Map3k4* is associated with enhanced *Sry* expression in B6.Y^{POS} gonads. Removal of a copy of

Map3k4 exacerbated B6.Y^{POS} sex reversal, demonstrating that *Map3k4* modifies sensitivity to B6.Y^{POS} sex reversal in a dose-dependent manner. Furthermore, the findings described in Chapter 5 support a key role for MAP3K4 in the regulation of *Sry* expression, in particular its timing, in B6.Y^{POS} sex reversal.

7.2.3. Delayed testis determination in B6.Y^{POS} Tg embryos

Despite enhanced, early expression of *Sry*, B6.Y^{POS} Tg embryos still show signs of subtle delays to testis determination. This is most clearly evident at 14.5 dpc with the presence of some *Stra8*-positive ovarian germ cells at the caudal poles of some gonads. Due to the wave-like spatiotemporal pattern of *Sry* expression, the cells at the caudal pole see SRY last, and are therefore the last cells to up-regulate *Sox9*. Development of ovarian tissue in this region of the gonad indicates that there has been sufficient delay in signal for the ovary-determining pathway to dominate. Analysis of *Sox9* expression in B6.Y^{POS} gonads indicates that *Sox9* is only modestly elevated, is still somewhat delayed and fails to spread to the caudal pole of the gonad by 31ts. This suggests that, even in the face of enhanced SRY, *Sox9* is still not sufficiently upregulated in time to fully restore testis determination in B6.Y^{POS} Tg embryos. A contribution to the observed delay is also potentially made by the reduced transactivation activity of the *M. poschiavinus* isoform of SRY relative to its *M. musculus* counterpart (Zhao et al., 2014) and presents an intriguing phenomenon of uncoupling of *Sry* and *Sox9* expression. In other words, increasing *Sry* does not necessarily lead to a proportional increase in *Sox9*. It would be interesting to determine whether further enhanced expression of *Sry* is able to fully restore testis determination in B6.Y^{POS} embryos. A report by Albrecht et al (2003) suggests that this may not be possible. In their study, the authors generated a construct containing the coding sequence of

Sry^{POS} within 129-derived regulatory regions, and transgenically expressed it in B6 XX embryos. The transgene conferred *Sry* expression levels 5 times higher than XY^{B6} and was able to cause male development of XX embryos. However, when examined at 13.5 dpc these embryos exhibited delayed testis determination, morphologically resembling transiently sex-reversed B6.Y^{AKR} testes that develop ovarian tissue at the gonadal poles (Albrecht et al., 2003). This suggests that even highly elevated *Sry* expression may not be sufficient to completely overcome reduced activity of the SRY^{POS} protein, consistent with data described here, and indicates that B6.Y^{POS} sex reversal cannot fully be explained by delayed timing of *Sry*.

7.2.4. No clear mechanism for delayed *Sry* expression in B6.Y^{POS} embryos

The question still remained, however, why *Sry* is delayed in B6.Y^{POS} embryos. This was the main focus of chapter 6. In summary, these experiments failed to identify a clear mechanism for the delayed expression observed or its rescue by the *Map3k4* transgene. However they do confirm that overexpression of *Map3k4* is not associated with overtly enhanced p38 MAPK or GATA4 phosphorylation. Attempts to characterise methylation of the *Sry* T-DMR in B6.Y^{POS} have generated only a small amount of preliminary data, and so these efforts must be continued in order to elucidate any differences in methylation dynamics in B6.Y^{POS} gonads.

7.2.5. POS-specific *Sry* polymorphisms

Sequence analysis of *Sry*^{POS} revealed modest differences between it and *Sry*^{B6}. Of these, most were also shared with another *M. domesticus*-derived *Sry* allele, *Sry*^{AKR}. B6.Y^{POS} embryos exhibit more severe sex reversal than B6.Y^{AKR} embryos, and furthermore *Sry* expression is more delayed in B6.Y^{POS} than in B6.Y^{AKR}. Therefore,

polymorphisms shared between these two strains, such as the missing CpGs, cannot alone explain the severely delayed *Sry* expression observed in B6.Y^{POS}. In this study, two polymorphisms that were unique to *Sry*^{POS} were identified. That they are situated so close together, and also that they lead to loss of a predicted binding motif for GATA4, is highly interesting. Whilst these data are merely indicative at this stage, I predict that these polymorphisms may be of some importance.

7.3. Future work

7.3.1. Further RNA-seq experiments

The future of the use of RNA-seq for the study of testis determination has been made brighter by the development of technologies that enable the technique to be performed using tiny amounts of tissue and even individual cells, opening up the possibility of more sophisticated analyses (Saliba et al., 2014, Miyamoto et al., 2015, Blakeley et al., 2015). Therefore, for future experiments it may not be necessary to pool tissue or include mesonephros in samples. In addition to comparisons between genotypes, sexes and stages it will be possible look at variation between individual cells in distinct regions of the gonad. A centre-to-pole dynamic profile of *Gadd45g* has been reported (Warr et al., 2012) and it would be possible to examine the consequences of this profile for expression in these regions. Of course, NGS is still very costly, so prioritisation would be important. Furthermore, by exploiting fluorescent reporters such as *Sf1-EGFP* (McClelland et al., 2015), *Sox9-EGFP* (Nel-Themaat et al., 2009) or *Sry-EGFP* (Albrecht and Eicher, 2001), gonadal somatic cell populations can be purified and their transcriptomes analysed with minimal noise from unwanted cell types. This would reduce false-detection rates and improve the specificity of prediction.

Microarrays have been extensively used for characterisation of gene expression in the developing testis, and it has been shown that the two techniques are highly concordant in their findings (Wang et al., 2014). However, RNA-seq may be more sensitive for the detection of rare transcripts as well as novel splice variants and alternative transcription start sites (Wang et al., 2009). Therefore, it is worth pursuing it as a potential means of uncovering novel members of the testis-determining pathway.

One further question remaining from this RNA-seq data is whether the putative targets of GADD45 γ are related to, or independent of, the interaction of GADD45 γ with MAP3K4. To assess this, comparison of transcriptome of gonads lacking *Map3k4* with that of gonads lacking *Gadd45g*, would enable identification of differentially expressed transcripts shared between the two. This may help to define targets of GADD45 γ -MAP3K4 signalling and also suggest novel targets of GADD45 γ acting via other means, perhaps via its role in DNA demethylation.

7.3.2. The role of MAPKs in Y^{POS} and beyond

Given more time, there are several key experiments that may help to shed light upon the phenomenon of delayed *Sry* expression in B6.Y^{POS} embryos and its rescue by *Map3k4* overexpression. Firstly, continuing with analysis of methylation will build a picture of the events occurring at the *Sry* T-DMR during the crucial period of testis determination. Not only would this require generating more bisulphite sequencing data for the stage already examined, it would also be necessary to analyse a range of stages to assess whether there are differences in the progressive demethylation of the T-DMR that coincide with delayed *Sry* expression. This is a technically challenging

technique due to the requirement for purified gonadal somatic cells, as well as time-consuming cloning steps, but nonetheless may prove highly informative.

Given the reported importance of histone modifications in *Sry* regulation (Katoh-Fukui et al., 2012, Kuroki et al., 2013), it would also be of great interest to examine these in B6.Y^{POS} embryos. Analysis of the deposition and loss of positive and negative histone marks such as H3K4me2 and H3K9me2, binding of key transcription factors such as GATA4, and of other chromatin-associated proteins and chromatin remodelling enzymes, should be performed by chromatin immunoprecipitation (ChIP). Indeed, it will be important to determine whether MAP3K4, GADD45 γ and p38 MAPK are recruited to *Sry* chromatin. Again, this is highly technically challenging due to the limiting number of somatic cells in the 11.5 dpc gonad and the lack of ChIP-grade antibodies against some key proteins, such as GATA4, MAP3K4 and GADD45 γ . Moreover, it is still unclear how the timing of chromatin modification relates to the onset of transcription. Presumably it precedes it by some hours, however, we do not yet know precisely when. This adds further complication to the technical difficulty of the technique, as the need to look at earlier gonadal stages means that cell numbers are even lower. However, the potential for identifying a direct link between the chromatin dynamics of *Sry* in two contexts in which its expression pattern differs, is tantalising. Ultimately, what is required is an account of the delay to *Sry* expression, either in *Gadd45g*-deficient or B6.Y^{POS} gonads, in terms of a delay to changes at *Sry*^{POS} chromatin. These chromatin changes themselves must presumably be explained by sequence variants at *Sry*^{POS}.

7.3.3. Closing remarks

In summary, this thesis has examined potential roles of MAPK-related pathways in testis determination. The data described herein indicate that studies of MAPK-related processes in mice offer an important model for understanding a significant cause of DSD in humans, namely, *MAP3K1* mutations. In addition to their translational potential, the genetic models described here form a firm foundation for ongoing studies of the epigenetic basis of *Sry* expression and, more broadly, mammalian sex determination.

References

- ABBAS, N., MCELREAVEY, K., LECONIAT, M., VILAIN, E., JAUBERT, F., BERGER, R., NIHOUL-FEKETE, C., RAPPAPORT, R. & FELLOUS, M. 1993. Familial case of 46,XX male and 46,XX true hermaphrodite associated with a paternal-derived SRY-bearing X chromosome. *C R Acad Sci III*, 316, 375-83.
- AKIYAMA, H. 2008. Control of chondrogenesis by the transcription factor Sox9. *Mod Rheumatol*, 18, 213-9.
- ALBRECHT, K. H. & EICHER, E. M. 1997. DNA sequence analysis of Sry alleles (subgenus Mus) implicates misregulation as the cause of C57BL/6J-Y(POS) sex reversal and defines the SRY functional unit. *Genetics*, 147, 1267-77.
- ALBRECHT, K. H. & EICHER, E. M. 2001. Evidence that Sry is expressed in pre-Sertoli cells and Sertoli and granulosa cells have a common precursor. *Dev Biol*, 240, 92-107.
- ALBRECHT, K. H., YOUNG, M., WASHBURN, L. L. & EICHER, E. M. 2003. Sry expression level and protein isoform differences play a role in abnormal testis development in C57BL/6J mice carrying certain Sry alleles. *Genetics*, 164, 277-88.
- AOUADI, M., BINETRUY, B., CARON, L., LE MARCHAND-BRUSTEL, Y. & BOST, F. 2006. Role of MAPKs in development and differentiation: lessons from knockout mice. *Biochimie*, 88, 1091-8.
- ARAVA, Y., WANG, Y., STOREY, J. D., LIU, C. L., BROWN, P. O. & HERSCHLAG, D. 2003. Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*, 100, 3889-94.
- ARBOLEDA, V. A., FLEMING, A., BARSEGHYAN, H., DELOT, E., SINSHEIMER, J. S. & VILAIN, E. 2014. Regulation of sex determination in mice by a non-coding genomic region. *Genetics*, 197, 885-97.
- ARCECI, R. J., KING, A. A., SIMON, M. C., ORKIN, S. H. & WILSON, D. B. 1993. Mouse GATA-4: a retinoic acid-inducible GATA-binding transcription factor expressed in endodermally derived tissues and heart. *Mol Cell Biol*, 13, 2235-46.
- ARMSTRONG, J. F., PRITCHARD-JONES, K., BICKMORE, W. A., HASTIE, N. D. & BARD, J. B. 1993. The expression of the Wilms' tumour gene, WT1, in the developing mammalian embryo. *Mech Dev*, 40, 85-97.
- AYADI, A., BIRLING, M. C., BOTTOMLEY, J., BUSSELL, J., FUCHS, H., FRAY, M., GAILUS-DURNER, V., GREENAWAY, S., HOUGHTON, R., KARP, N., LEBLANC, S., LENGGER, C., MAIER, H., MALLON, A. M., MARSCHALL, S., MELVIN, D., MORGAN, H., PAVLOVIC, G., RYDER, E., SKARNES, W. C., SELLOUM, M., RAMIREZ-SOLIS, R., SORG, T., TEBOUL, L., VASSEUR, L., WALLING, A., WEAVER, T., WELLS, S., WHITE, J. K., BRADLEY, A., ADAMS, D. J., STEEL, K. P., HRABE DE ANGELIS, M., BROWN, S. D. & HERAULT, Y. 2012. Mouse large-scale phenotyping initiatives: overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project. *Mamm Genome*, 23, 600-10.
- BAGHERI-FAM, S., SIM, H., BERNARD, P., JAYAKODY, I., TAKETO, M. M., SCHERER, G. & HARLEY, V. R. 2008. Loss of Fgfr2 leads to partial XY sex reversal. *Dev Biol*, 314, 71-83.
- BARRETO, G., SCHAFER, A., MARHOLD, J., STACH, D., SWAMINATHAN, S. K., HANDA, V., DODERLEIN, G., MALTRY, N., WU, W., LYKO, F. & NIEHRS, C. 2007. Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature*, 445, 671-5.
- BARRIONUEVO, F., BAGHERI-FAM, S., KLATTIG, J., KIST, R., TAKETO, M. M., ENGLERT, C. & SCHERER, G. 2006. Homozygous inactivation of Sox9 causes complete XY sex reversal in mice. *Biol Reprod*, 74, 195-201.
- BAXTER, R. M., ARBOLEDA, V. A., LEE, H., BARSEGHYAN, H., ADAM, M. P., FECHNER, P. Y., BARGMAN, R., KEEGAN, C., TRAVERS, S., SCHELLEY, S., HUDGINS, L., MATHEW, R. P., STALKER, H. J., ZORI, R., GORDON, O. K., RAMOS-PLATT, L., PAWLIKOWSKA-HADDAL, A., ESKIN, A., NELSON, S. F., DELOT, E. & VILAIN, E. 2015. Exome sequencing for the

- diagnosis of 46,XY disorders of sex development. *J Clin Endocrinol Metab*, 100, E333-44.
- BEUCHLE, D., STRUHL, G. & MULLER, J. 2001. Polycomb group proteins and heritable silencing of *Drosophila* Hox genes. *Development*, 128, 993-1004.
- BIDDLE, F. G., EALES, B. A. & NISHIOKA, Y. 1991. A DNA polymorphism from five inbred strains of the mouse identifies a functional class of domesticus-type Y chromosome that produces the same phenotypic distribution of gonadal hermaphrodites. *Genome*, 34, 96-104.
- BIDDLE, F. G. & NISHIOKA, Y. 1988. Assays of testis development in the mouse distinguish three classes of domesticus-type Y chromosome. *Genome*, 30, 870-8.
- 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., PORTER, F. D. & WESTPHAL, H. 2000. The LIM homeobox gene *Lhx9* is essential for mouse gonad formation. *Nature*, 403, 909-13.
- BLAKELEY, P., FOGARTY, N. M., DEL VALLE, I., WAMAITHA, S. E., HU, T. X., ELDER, K., SNELL, P., CHRISTIE, L., ROBSON, P. & NIAKAN, K. K. 2015. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development*, 142, 3151-65.
- BOCK, C., REITHER, S., MIKESKA, T., PAULSEN, M., WALTER, J. & LENGAUER, T. 2005. BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics*, 21, 4067-8.
- BOGANI, D., SIGGERS, P., BRIXEY, R., WARR, N., BEDDOW, S., EDWARDS, J., WILLIAMS, D., WILHELM, D., KOOPMAN, P., FLAVELL, R. A., CHI, H., OSTRER, H., WELLS, S., CHEESEMAN, M. & GREENFIELD, A. 2009. Loss of mitogen-activated protein kinase kinase 4 (MAP3K4) reveals a requirement for MAPK signalling in mouse sex determination. *PLoS Biol*, 7, e1000196.
- BOGNAR, P., NEMETH, I., MAYER, B., HALUSZKA, D., WIKONKAL, N., OSTORHAZI, E., JOHN, S., PAULSSON, M., SMYTH, N., PASZTOI, M., BUZAS, E. I., SZIPOCS, R., KOLONICS, A., TEMESVARI, E. & KARPATI, S. 2014. Reduced inflammatory threshold indicates skin barrier defect in transglutaminase 3 knockout mice. *J Invest Dermatol*, 134, 105-11.
- BOUMA, G. J., WASHBURN, L. L., ALBRECHT, K. H. & EICHER, E. M. 2007. Correct dosage of *Fog2* and *Gata4* transcription factors is critical for fetal testis development in mice. *Proc Natl Acad Sci U S A*, 104, 14994-9.
- BOWLES, J., COOPER, L., BERKMAN, J. & KOOPMAN, P. 1999. *Sry* requires a CAG repeat domain for male sex determination in *Mus musculus*. *Nat Genet*, 22, 405-8.
- BOWLES, J., KNIGHT, D., SMITH, C., WILHELM, D., RICHMAN, J., MAMIYA, S., YASHIRO, K., CHAWENGSAKSOPHAK, K., WILSON, M. J., ROSSANT, J., HAMADA, H. & KOOPMAN, P. 2006. Retinoid signaling determines germ cell fate in mice. *Science*, 312, 596-600.
- BRADFORD, S. T., WILHELM, D., BANDIERA, R., VIDAL, V., SCHEDL, A. & KOOPMAN, P. 2009. A cell-autonomous role for *WT1* in regulating *Sry* in vivo. *Hum Mol Genet*, 18, 3429-38.
- BRENNAN, J. & CAPEL, B. 2004. One tissue, two fates: molecular genetic events that underlie testis versus ovary development. *Nat Rev Genet*, 5, 509-21.
- BRENNAN, J., TILMANN, C. & CAPEL, B. 2003. *Pdgr*-alpha mediates testis cord organization and fetal Leydig cell development in the XY gonad. *Genes Dev*, 17, 800-10.
- BULLEJOS, M. & KOOPMAN, P. 2001. Spatially dynamic expression of *Sry* in mouse genital ridges. *Dev Dyn*, 221, 201-5.
- BULLEJOS, M. & KOOPMAN, P. 2005. Delayed *Sry* and *Sox9* expression in developing mouse gonads underlies B6-Y(DOM) sex reversal. *Dev Biol*, 278, 473-81.
- CALLINAN, P. A. & FEINBERG, A. P. 2006. The emerging science of epigenomics. *Hum Mol Genet*, 15 Spec No 1, R95-101.
- CAMERON, F. J. & SINCLAIR, A. H. 1997. Mutations in *SRY* and *SOX9*: testis-determining genes. *Hum Mutat*, 9, 388-95.
- CAPECCHI, M. R. 1989. Altering the genome by homologous recombination. *Science*, 244, 1288-92.

- CAPEL, B., SWAIN, A., NICOLIS, S., HACKER, A., WALTER, M., KOOPMAN, P., GOODFELLOW, P. & LOVELL-BADGE, R. 1993. Circular transcripts of the testis-determining gene *Sry* in adult mouse testis. *Cell*, 73, 1019-30.
- CARLISLE, C., WINKING, H., WEICHENHAN, D. & NAGAMINE, C. M. 1996. Absence of correlation between *Sry* polymorphisms and XY sex reversal caused by the M. m. domesticus Y chromosome. *Genomics*, 33, 32-45.
- CARRE, G. A. & GREENFIELD, A. 2014. Characterising novel pathways in testis determination using mouse genetics. *Sex Dev*, 8, 199-207.
- CATTANACH, B. M. 1986. Parental origin effects in mice. *J Embryol Exp Morphol*, 97 Suppl, 137-50.
- CATTANACH, B. M., WILLIAMS, C. E. & BAILEY, H. 1972. Identification of the linkage groups carried by the metacentric chromosomes of the tobacco mouse (*Mus poschiavinus*). *Cytogenetics*, 11, 412-23.
- CHABOISSIER, M. C., KOBAYASHI, A., VIDAL, V. I., LUTZKENDORF, S., VAN DE KANT, H. J., WEGNER, M., DE ROOIJ, D. G., BEHRINGER, R. R. & SCHEDL, A. 2004. Functional analysis of *Sox8* and *Sox9* during sex determination in the mouse. *Development*, 131, 1891-901.
- CHEN, S. R., CHEN, M., WANG, X. N., ZHANG, J., WEN, Q., JI, S. Y., ZHENG, Q. S., GAO, F. & LIU, Y. X. 2013a. The Wilms tumor gene, *Wt1*, maintains testicular cord integrity by regulating the expression of *Col4a1* and *Col4a2*. *Biol Reprod*, 88, 56.
- CHEN, Y. J., WANG, Y. N. & CHANG, W. C. 2007. ERK2-mediated C-terminal serine phosphorylation of p300 is vital to the regulation of epidermal growth factor-induced keratin 16 gene expression. *J Biol Chem*, 282, 27215-28.
- CHEN, Y. S., RACCA, J. D., SEQUEIRA, P. W., PHILLIPS, N. B. & WEISS, M. A. 2013c. Microsatellite-encoded domain in rodent *Sry* functions as a genetic capacitor to enable the rapid evolution of biological novelty. *Proc Natl Acad Sci U S A*, 110, E3061-70.
- CLEPET, C., SCHAFER, A. J., SINCLAIR, A. H., PALMER, M. S., LOVELL-BADGE, R. & GOODFELLOW, P. N. 1993. The human *SRY* transcript. *Hum Mol Genet*, 2, 2007-12.
- COLVIN, J. S., GREEN, R. P., SCHMAHL, J., CAPEL, B. & ORNITZ, D. M. 2001. Male-to-female sex reversal in mice lacking fibroblast growth factor 9. *Cell*, 104, 875-89.
- COMBES, A. N., WILHELM, D., DAVIDSON, T., DEJANA, E., HARLEY, V., SINCLAIR, A. & KOOPMAN, P. 2009. Endothelial cell migration directs testis cord formation. *Dev Biol*, 326, 112-20.
- COOL, J., DEFALCO, T. J. & CAPEL, B. 2011. Vascular-mesenchymal cross-talk through *Vegf* and *Pdgfr* drives organ patterning. *Proc Natl Acad Sci U S A*, 108, 167-72.
- CORREA, S. M., WASHBURN, L. L., KAHN, R. S., MUSSON, M. C., BOUMA, G. J., EICHER, E. M. & ALBRECHT, K. H. 2012. Sex reversal in C57BL/6J XY mice caused by increased expression of ovarian genes and insufficient activation of the testis determining pathway. *PLoS Genet*, 8, e1002569.
- COVENEY, D., ROSS, A. J., SLONE, J. D. & CAPEL, B. 2008. A microarray analysis of the XX *Wnt4* mutant gonad targeted at the identification of genes involved in testis vascular differentiation. *Gene Expr Patterns*, 8, 529-37.
- COWARD, P., NAGAI, K., CHEN, D., THOMAS, H. D., NAGAMINE, C. M. & LAU, Y. F. 1994. Polymorphism of a CAG trinucleotide repeat within *Sry* correlates with B6.YDom sex reversal. *Nat Genet*, 6, 245-50.
- CRAIG, E. A., STEVENS, M. V., VAILLANCOURT, R. R. & CAMENISCH, T. D. 2008. MAP3Ks as central regulators of cell fate during development. *Dev Dyn*, 237, 3102-14.
- D'AGOSTINO, D. & LOWE, M. E. 2004. Pancreatic lipase-related protein 2 is the major colipase-dependent pancreatic lipase in suckling mice. *J Nutr*, 134, 132-4.
- DE NADAL, E., ZAPATER, M., ALEPUZ, P. M., SUMOY, L., MAS, G. & POSAS, F. 2004. The MAPK Hog1 recruits Rpd3 histone deacetylase to activate osmoresponsive genes. *Nature*, 427, 370-4.

- DELMAR, M. & MAKITA, N. 2012. Cardiac connexins, mutations and arrhythmias. *Curr Opin Cardiol*, 27, 236-41.
- DIAS, J. A. 1985. Transglutaminase activity in testicular homogenates and serum-free Sertoli cell cultures. *Biol Reprod*, 33, 835-43.
- DIAS, J. A. 1986. Effect of transglutaminase substrates and polyamines on the cellular sequestration and processing of follicle-stimulating hormone by rat Sertoli cells. *Biol Reprod*, 35, 49-58.
- DUBIN, R. A., KAZMI, M. A. & OSTRER, H. 1995. Inverted repeats are necessary for circularization of the mouse testis Sry transcript. *Gene*, 167, 245-8.
- DUBIN, R. A. & OSTRER, H. 1994. Sry is a transcriptional activator. *Mol Endocrinol*, 8, 1182-92.
- EICHER, E. M., SHOWN, E. P. & WASHBURN, L. L. 1995. Sex reversal in C57BL/6J-YPOS mice corrected by a Sry transgene. *Philos Trans R Soc Lond B Biol Sci*, 350, 263-8; discussion 268-9.
- EICHER, E. M. & WASHBURN, L. L. 2001. Does one gene determine whether a C57BL/6J-Y(POS) mouse will develop as a female or as an hermaphrodite? *J Exp Zool*, 290, 322-6.
- EICHER, E. M., WASHBURN, L. L., SCHORK, N. J., LEE, B. K., SHOWN, E. P., XU, X., DREDGE, R. D., PRINGLE, M. J. & PAGE, D. C. 1996. Sex-determining genes on mouse autosomes identified by linkage analysis of C57BL/6J-YPOS sex reversal. *Nat Genet*, 14, 206-9.
- EICHER, E. M., WASHBURN, L. L., WHITNEY, J. B., 3RD & MORROW, K. E. 1982. Mus poschiavinus Y chromosome in the C57BL/6J murine genome causes sex reversal. *Science*, 217, 535-7.
- FORD, C. E., JONES, K. W., POLANI, P. E., DE ALMEIDA, J. C. & BRIGGS, J. H. 1959. A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). *Lancet*, 1, 711-3.
- FOSTER, J. W. 1996. Mutations in SOX9 cause both autosomal sex reversal and campomelic dysplasia. *Acta Paediatr Jpn*, 38, 405-11.
- FUJIMOTO, Y., TANAKA, S. S., YAMAGUCHI, Y. L., KOBAYASHI, H., KUROKI, S., TACHIBANA, M., SHINOMURA, M., KANAI, Y., MOROHASHI, K., KAWAKAMI, K. & NISHINAKAMURA, R. 2013. Homeoproteins Six1 and Six4 regulate male sex determination and mouse gonadal development. *Dev Cell*, 26, 416-30.
- FURUSE, M. & MORIWAKI, K. 2009. The role of claudin-based tight junctions in morphogenesis. *Ann N Y Acad Sci*, 1165, 58-61.
- GALLAGHER, E. D., GUTOWSKI, S., STERNWEIS, P. C. & COBB, M. H. 2004. RhoA binds to the amino terminus of MEKK1 and regulates its kinase activity. *J Biol Chem*, 279, 1872-7.
- GARDINER, W. J. & TBOUL, L. 2009. Overexpression transgenesis in mouse: pronuclear injection. *Methods Mol Biol*, 561, 111-26.
- GEH, E., MENG, Q., MONGAN, M., WANG, J., TAKATORI, A., ZHENG, Y., PUGA, A., LANG, R. A. & XIA, Y. 2011. Mitogen-activated protein kinase kinase kinase 1 (MAP3K1) integrates developmental signals for eyelid closure. *Proc Natl Acad Sci U S A*, 108, 17349-54.
- GIERL, M. S., GRUHN, W. H., VON SEGGERN, A., MALTRY, N. & NIEHRS, C. 2012. GADD45G functions in male sex determination by promoting p38 signaling and Sry expression. *Dev Cell*, 23, 1032-42.
- GONG, W., PAN, L., LIN, Q., ZHOU, Y., XIN, C., YU, X., CUI, P., HU, S. & YU, J. 2013. Transcriptome profiling of the developing postnatal mouse testis using next-generation sequencing. *Sci China Life Sci*, 56, 1-12.
- GUBBAY, J., COLLIGNON, J., KOOPMAN, P., CAPEL, B., ECONOMOU, A., MUNSTERBERG, A., VIVIAN, N., GOODFELLOW, P. & LOVELL-BADGE, R. 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature*, 346, 245-50.
- GUBBAY, J., VIVIAN, N., ECONOMOU, A., JACKSON, D., GOODFELLOW, P. & LOVELL-BADGE, R. 1992. Inverted repeat structure of the Sry locus in mice. *Proc Natl Acad Sci U S A*, 89, 7953-7.

- GUO, Y., SHENG, Q., LI, J., YE, F., SAMUELS, D. C. & SHYR, Y. 2013. Large scale comparison of gene expression levels by microarrays and RNAseq using TCGA data. *PLoS One*, 8, e71462.
- HAINES, B., HUGHES, J., CORBETT, M., SHAW, M., INNES, J., PATEL, L., GECZ, J., CLAYTON-SMITH, J. & THOMAS, P. 2015. Interchromosomal insertional translocation at Xq26.3 alters SOX3 expression in an individual with XX male sex reversal. *J Clin Endocrinol Metab*, 100, E815-20.
- HAMMES, A., GUO, J. K., LUTSCH, G., LEHESTE, J. R., LANDROCK, D., ZIEGLER, U., GUBLER, M. C. & SCHEDL, A. 2001. Two splice variants of the Wilms' tumor 1 gene have distinct functions during sex determination and nephron formation. *Cell*, 106, 319-29.
- HANSEN, T. B., JENSEN, T. I., CLAUSEN, B. H., BRAMSEN, J. B., FINSEN, B., DAMGAARD, C. K. & KJEMS, J. 2013. Natural RNA circles function as efficient microRNA sponges. *Nature*, 495, 384-8.
- HARLEY, V. R., LOVELL-BADGE, R. & GOODFELLOW, P. N. 1994. Definition of a consensus DNA binding site for SRY. *Nucleic Acids Res*, 22, 1500-1.
- HIRAMATSU, R., MATOBA, S., KANAI-AZUMA, M., TSUNEKAWA, N., KATOH-FUKUI, Y., KUROHMARU, M., MOROHASHI, K., WILHELM, D., KOOPMAN, P. & KANAI, Y. 2009. A critical time window of Sry action in gonadal sex determination in mice. *Development*, 136, 129-38.
- HOFFMEYER, A., PIEKORZ, R., MORIGGL, R. & IHLE, J. N. 2001. Gadd45gamma is dispensable for normal mouse development and T-cell proliferation. *Mol Cell Biol*, 21, 3137-43.
- HOSSAIN, A. & SAUNDERS, G. F. 2001. The human sex-determining gene SRY is a direct target of WT1. *J Biol Chem*, 276, 16817-23.
- HU, S., XIE, Z., ONISHI, A., YU, X., JIANG, L., LIN, J., RHO, H. S., WOODARD, C., WANG, H., JEONG, J. S., LONG, S., HE, X., WADE, H., BLACKSHAW, S., QIAN, J. & ZHU, H. 2009. Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell*, 139, 610-22.
- HU, Y. C., OKUMURA, L. M. & PAGE, D. C. 2013. Gata4 is required for formation of the genital ridge in mice. *PLoS Genet*, 9, e1003629.
- HUGHES, I. A., HOUK, C., AHMED, S. F., LEE, P. A. & LAWSON WILKINS PEDIATRIC ENDOCRINE SOCIETY/EUROPEAN SOCIETY FOR PAEDIATRIC ENDOCRINOLOGY CONSENSUS, G. 2006. Consensus statement on management of intersex disorders. *J Pediatr Urol*, 2, 148-62.
- IKEGAMI, K., OHGANE, J., TANAKA, S., YAGI, S. & SHIOTA, K. 2009. Interplay between DNA methylation, histone modification and chromatin remodeling in stem cells and during development. *Int J Dev Biol*, 53, 203-14.
- JACOBS, P. A. & STRONG, J. A. 1959. A case of human intersexuality having a possible XXY sex-determining mechanism. *Nature*, 183, 302-3.
- JAMESON, S. A., LIN, Y. T. & CAPEL, B. 2012a. Testis development requires the repression of Wnt4 by Fgf signaling. *Dev Biol*, 370, 24-32.
- JAMESON, S. A., NATARAJAN, A., COOL, J., DEFALCO, T., MAATOUK, D. M., MORK, L., MUNGER, S. C. & CAPEL, B. 2012b. Temporal transcriptional profiling of somatic and germ cells reveals biased lineage priming of sexual fate in the fetal mouse gonad. *PLoS Genet*, 8, e1002575.
- JAWAHEER, D., JUO, S. H., LE CAIGNEC, C., DAVID, A., PETIT, C., GREGERSEN, P., DOWBAK, S., DAMLE, A., MCELREAVEY, K. & OSTRER, H. 2003. Mapping a gene for 46,XY gonadal dysgenesis by linkage analysis. *Clin Genet*, 63, 530-5.
- JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337, 816-21.
- JOHN, S., THIEBACH, L., FRIE, C., MOKKAPATI, S., BECHTEL, M., NISCHT, R., ROSSER-DAVIES, S., PAULSSON, M. & SMYTH, N. 2012. Epidermal transglutaminase (TGase 3) is required

- for proper hair development, but not the formation of the epidermal barrier. *PLoS One*, 7, e34252.
- JOHNSON, D. R. 1974. Hairpin-tail: a case of post-reductional gene action in the mouse egg. *Genetics*, 76, 795-805.
- JORDAN, B. K., MOHAMMED, M., CHING, S. T., DELOT, E., CHEN, X. N., DEWING, P., SWAIN, A., RAO, P. N., ELEJALDE, B. R. & VILAIN, E. 2001. Up-regulation of WNT-4 signaling and dosage-sensitive sex reversal in humans. *Am J Hum Genet*, 68, 1102-9.
- KAMACHI, Y. & KONDOH, H. 2013. Sox proteins: regulators of cell fate specification and differentiation. *Development*, 140, 4129-44.
- KASHIMADA, K. & KOOPMAN, P. 2010. Sry: the master switch in mammalian sex determination. *Development*, 137, 3921-30.
- KATOH-FUKUI, Y., MIYABAYASHI, K., KOMATSU, T., OWAKI, A., BABA, T., SHIMA, Y., KIDOKORO, T., KANAI, Y., SCHEDL, A., WILHELM, D., KOOPMAN, P., OKUNO, Y. & MOROHASHI, K. 2012. Cbx2, a polycomb group gene, is required for Sry gene expression in mice. *Endocrinology*, 153, 913-24.
- KATOH-FUKUI, Y., TSUCHIYA, R., SHIROISHI, T., NAKAHARA, Y., HASHIMOTO, N., NOGUCHI, K. & HIGASHINAKAGAWA, T. 1998. Male-to-female sex reversal in M33 mutant mice. *Nature*, 393, 688-92.
- KAUFMANN, L. T., GIERL, M. S. & NIEHRS, C. 2011. Gadd45a, Gadd45b and Gadd45g expression during mouse embryonic development. *Gene Expr Patterns*, 11, 465-70.
- KIM, Y., BINGHAM, N., SEKIDO, R., PARKER, K. L., LOVELL-BADGE, R. & CAPEL, B. 2007. Fibroblast growth factor receptor 2 regulates proliferation and Sertoli differentiation during male sex determination. *Proc Natl Acad Sci U S A*, 104, 16558-63.
- KIM, Y., KOBAYASHI, A., SEKIDO, R., DINAPOLI, L., BRENNAN, J., CHABOISSIER, M. C., POULAT, F., BEHRINGER, R. R., LOVELL-BADGE, R. & CAPEL, B. 2006. Fgf9 and Wnt4 act as antagonistic signals to regulate mammalian sex determination. *PLoS Biol*, 4, e187.
- KLATTIG, J., SIERIG, R., KRUSPE, D., MAKKI, M. S. & ENGLERT, C. 2007. WT1-mediated gene regulation in early urogenital ridge development. *Sex Dev*, 1, 238-54.
- KLEIN, A. M., ZAGANJOR, E. & COBB, M. H. 2013. Chromatin-tethered MAPKs. *Curr Opin Cell Biol*, 25, 272-7.
- KOCER, A., REICHMANN, J., BEST, D. & ADAMS, I. R. 2009. Germ cell sex determination in mammals. *Mol Hum Reprod*, 15, 205-13.
- KOOISTRA, S. M. & HELIN, K. 2012. Molecular mechanisms and potential functions of histone demethylases. *Nat Rev Mol Cell Biol*, 13, 297-311.
- KOOPMAN, P., GUBBAY, J., VIVIAN, N., GOODFELLOW, P. & LOVELL-BADGE, R. 1991. Male development of chromosomally female mice transgenic for Sry. *Nature*, 351, 117-21.
- KOUBOVA, J., HU, Y. C., BHATTACHARYYA, T., SOH, Y. Q., GILL, M. E., GOODHEART, M. L., HOGARTH, C. A., GRISWOLD, M. D. & PAGE, D. C. 2014. Retinoic acid activates two pathways required for meiosis in mice. *PLoS Genet*, 10, e1004541.
- KOUBOVA, J., MENKE, D. B., ZHOU, Q., CAPEL, B., GRISWOLD, M. D. & PAGE, D. C. 2006. Retinoic acid regulates sex-specific timing of meiotic initiation in mice. *Proc Natl Acad Sci U S A*, 103, 2474-9.
- KREIDBERG, J. A., SARIOLA, H., LORING, J. M., MAEDA, M., PELLETIER, J., HOUSMAN, D. & JAENISCH, R. 1993. WT-1 is required for early kidney development. *Cell*, 74, 679-91.
- KUMAR, D. & LASSAR, A. B. 2009. The transcriptional activity of Sox9 in chondrocytes is regulated by RhoA signaling and actin polymerization. *Mol Cell Biol*, 29, 4262-73.
- KUO, C. T., MORRISEY, E. E., ANANDAPPA, R., SIGRIST, K., LU, M. M., PARMACEK, M. S., SOUDAIS, C. & LEIDEN, J. M. 1997. GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev*, 11, 1048-60.
- KUROKI, S., MATOBA, S., AKIYOSHI, M., MATSUMURA, Y., MIYACHI, H., MISE, N., ABE, K., OGURA, A., WILHELM, D., KOOPMAN, P., NOZAKI, M., KANAI, Y., SHINKAI, Y. & TACHIBANA, M. 2013. Epigenetic regulation of mouse sex determination by the histone demethylase Jmjd1a. *Science*, 341, 1106-9.

- KUSAKA, M., KATOH-FUKUI, Y., OGAWA, H., MIYABAYASHI, K., BABA, T., SHIMA, Y., SUGIYAMA, N., SUGIMOTO, Y., OKUNO, Y., KODAMA, R., IIZUKA-KOGO, A., SENDA, T., SASAOKA, T., KITAMURA, K., AIZAWA, S. & MOROHASHI, K. 2010. Abnormal epithelial cell polarity and ectopic epidermal growth factor receptor (EGFR) expression induced in Emx2 KO embryonic gonads. *Endocrinology*, 151, 5893-904.
- LAL-NAG, M. & MORIN, P. J. 2009. The claudins. *Genome Biol*, 10, 235.
- LARNEY, C., BAILEY, T. L. & KOOPMAN, P. 2014. Switching on sex: transcriptional regulation of the testis-determining gene Sry. *Development*, 141, 2195-205.
- LI, E. & ZHANG, Y. 2014. DNA methylation in mammals. *Cold Spring Harb Perspect Biol*, 6, a019133.
- LI, Y. & TOLLEFSBOL, T. O. 2011. DNA methylation detection: bisulfite genomic sequencing analysis. *Methods Mol Biol*, 791, 11-21.
- LI, Y., ZHENG, M. & LAU, Y. F. 2014. The sex-determining factors SRY and SOX9 regulate similar target genes and promote testis cord formation during testicular differentiation. *Cell Rep*, 8, 723-33.
- LIM, H. N., BERKOVITZ, G. D., HUGHES, I. A. & HAWKINS, J. R. 2000. Mutation analysis of subjects with 46, XX sex reversal and 46, XY gonadal dysgenesis does not support the involvement of SOX3 in testis determination. *Hum Genet*, 107, 650-2.
- LIN, J., POOLE, J. & LINZER, D. I. 1997. Two novel members of the prolactin/growth hormone family are expressed in the mouse placenta. *Endocrinology*, 138, 5535-40.
- LIN, L. & ACHERMANN, J. C. 2008. Steroidogenic factor-1 (SF-1, Ad4BP, NR5A1) and disorders of testis development. *Sex Dev*, 2, 200-9.
- LINDBERG, J. & LUNDEBERG, J. 2010. The plasticity of the mammalian transcriptome. *Genomics*, 95, 1-6.
- LOKE, J. & OSTRER, H. 2012. Rapidly screening variants of uncertain significance in the MAP3K1 gene for phenotypic effects. *Clin Genet*, 81, 272-7.
- LOKE, J., PEARLMAN, A., RADI, O., ZUFFARDI, O., GIUSSANI, U., PALLOTTA, R., CAMERINO, G. & OSTRER, H. 2014. Mutations in MAP3K1 tilt the balance from SOX9/FGF9 to WNT/beta-catenin signaling. *Hum Mol Genet*, 23, 1073-83.
- LOVELL-BADGE, R., CANNING, C. & SEKIDO, R. 2002. Sex-determining genes in mice: building pathways. *Novartis Found Symp*, 244, 4-18; discussion 18-22, 35-42, 253-7.
- LOWE, M. E., KAPLAN, M. H., JACKSON-GRUSBY, L., D'AGOSTINO, D. & GRUSBY, M. J. 1998. Decreased neonatal dietary fat absorption and T cell cytotoxicity in pancreatic lipase-related protein 2-deficient mice. *J Biol Chem*, 273, 31215-21.
- LU, B., YU, H., CHOW, C., LI, B., ZHENG, W., DAVIS, R. J. & FLAVELL, R. A. 2001. GADD45gamma mediates the activation of the p38 and JNK MAP kinase pathways and cytokine production in effector TH1 cells. *Immunity*, 14, 583-90.
- LUDBROOK, L. M., BERNARD, P., BAGHERI-FAM, S., RYAN, J., SEKIDO, R., WILHELM, D., LOVELL-BADGE, R. & HARLEY, V. R. 2012. Excess DAX1 leads to XY ovotesticular disorder of sex development (DSD) in mice by inhibiting steroidogenic factor-1 (SF1) activation of the testis enhancer of SRY-box-9 (Sox9). *Endocrinology*, 153, 1948-58.
- LUDBROOK, L. M. & HARLEY, V. R. 2004. Sex determination: a 'window' of DAX1 activity. *Trends Endocrinol Metab*, 15, 116-21.
- LUO, X., IKEDA, Y. & PARKER, K. L. 1994. A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. *Cell*, 77, 481-90.
- LYON, M. F. 1996. X-chromosome inactivation. Pinpointing the centre. *Nature*, 379, 116-7.
- MA, D. K., GUO, J. U., MING, G. L. & SONG, H. 2009. DNA excision repair proteins and Gadd45 as molecular players for active DNA demethylation. *Cell Cycle*, 8, 1526-31.
- MANDEL, H., SHEMER, R., BOROCHOWITZ, Z. U., OKOPNIK, M., KNOFF, C., INDELMAN, M., DRUGAN, A., TIOSANO, D., GERSHONI-BARUCH, R., CHODER, M. & SPRECHER, E. 2008. SERKAL syndrome: an autosomal-recessive disorder caused by a loss-of-function mutation in WNT4. *Am J Hum Genet*, 82, 39-47.

- MARGOLIN, G., KHIL, P. P., KIM, J., BELLANI, M. A. & CAMERINI-OTERO, R. D. 2014. Integrated transcriptome analysis of mouse spermatogenesis. *BMC Genomics*, 15, 39.
- MATSON, C. K., MURPHY, M. W., SARVER, A. L., GRISWOLD, M. D., BARDWELL, V. J. & ZARKOWER, D. 2011. DMRT1 prevents female reprogramming in the postnatal mammalian testis. *Nature*, 476, 101-4.
- MAZAUD-GUITTOT, S., MEUGNIER, E., PESENTI, S., WU, X., VIDAL, H., GOW, A. & LE MAGUERESSE-BATTISTONI, B. 2010. Claudin 11 deficiency in mice results in loss of the Sertoli cell epithelial phenotype in the testis. *Biol Reprod*, 82, 202-13.
- MCCANN-CROSBY, B. & SUTTON, V. R. 2015. Disorders of sexual development. *Clin Perinatol*, 42, 395-412, ix-x.
- MCCLELLAND, K. S., BELL, K., LARNEY, C., HARLEY, V. R., SINCLAIR, A. H., OSHLACK, A., KOOPMAN, P. & BOWLES, J. 2015. Purification and Transcriptomic Analysis of Mouse Fetal Leydig Cells Reveals Candidate Genes for Specification of Gonadal Steroidogenic Cells. *Biol Reprod*.
- MEEKS, J. J., CRAWFORD, S. E., RUSSELL, T. A., MOROHASHI, K., WEISS, J. & JAMESON, J. L. 2003a. Dax1 regulates testis cord organization during gonadal differentiation. *Development*, 130, 1029-36.
- MEEKS, J. J., WEISS, J. & JAMESON, J. L. 2003b. Dax1 is required for testis determination. *Nat Genet*, 34, 32-3.
- MENKE, D. B. 2013. Engineering subtle targeted mutations into the mouse genome. *Genesis*, 51, 605-18.
- MILLER, K. E., LUNDRIGAN, B. L. & TUCKER, P. K. 1995. Length variation of CAG repeats in Sry across populations of *Mus domesticus*. *Mamm Genome*, 6, 206-8.
- MIYAKE, Z., TAKEKAWA, M., GE, Q. & 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, 2765-76.
- MIYAMOTO, D. T., ZHENG, Y., WITTNER, B. S., LEE, R. J., ZHU, H., BRODERICK, K. T., DESAI, R., FOX, D. B., BRANNIGAN, B. W., TRAUTWEIN, J., ARORA, K. S., DESAI, N., DAHL, D. M., SEQUIST, L. V., SMITH, M. R., KAPUR, R., WU, C. L., SHIODA, T., RAMASWAMY, S., TING, D. T., TONER, M., MAHESWARAN, S. & HABER, D. A. 2015. RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science*, 349, 1351-6.
- MIYAMOTO, Y., TANIGUCHI, H., HAMEL, F., SILVERSIDES, D. W. & VIGER, R. S. 2008. A GATA4/WT1 cooperation regulates transcription of genes required for mammalian sex determination and differentiation. *BMC Mol Biol*, 9, 44.
- MOALEM, S., BABUL-HIRJI, R., STAVROPOLOUS, D. J., WHERRETT, D., BAGLI, D. J., THOMAS, P. & CHITAYAT, D. 2012. XX male sex reversal with genital abnormalities associated with a de novo SOX3 gene duplication. *Am J Med Genet A*, 158A, 1759-64.
- MOLKENTIN, J. D., LIN, Q., DUNCAN, S. A. & OLSON, E. N. 1997. Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev*, 11, 1061-72.
- MONIOT, B., DECLOSMENIL, F., BARRIONUEVO, F., SCHERER, G., ARITAKE, K., MALKI, S., MARZI, L., COHEN-SOLAL, A., GEORG, I., KLATTIG, J., ENGLERT, C., KIM, Y., CAPEL, B., EGUCHI, N., URADE, Y., BOIZET-BONHOURE, B. & POULAT, F. 2009. The PGD2 pathway, independently of FGF9, amplifies SOX9 activity in Sertoli cells during male sexual differentiation. *Development*, 136, 1813-21.
- MORAIS DA SILVA, S., HACKER, A., HARLEY, V., GOODFELLOW, P., SWAIN, A. & LOVELL-BADGE, R. 1996. Sox9 expression during gonadal development implies a conserved role for the gene in testis differentiation in mammals and birds. *Nat Genet*, 14, 62-8.
- MOUSE, E. C., STAMATOYANNOPOULOS, J. A., SNYDER, M., HARDISON, R., REN, B., GINGERAS, T., GILBERT, D. M., GROUDINE, M., BENDER, M., KAUL, R., CANFIELD, T., GISTE, E., JOHNSON, A., ZHANG, M., BALASUNDARAM, G., BYRON, R., ROACH, V., SABO, P. J., SANDSTROM, R., STEHLING, A. S., THURMAN, R. E., WEISSMAN, S. M., CAYTING, P.,

- HARIHARAN, M., LIAN, J., CHENG, Y., LANDT, S. G., MA, Z., WOLD, B. J., DEKKER, J., CRAWFORD, G. E., KELLER, C. A., WU, W., MORRISSEY, C., KUMAR, S. A., MISHRA, T., JAIN, D., BYRSKA-BISHOP, M., BLANKENBERG, D., LAJOIE, B. R., JAIN, G., SANYAL, A., CHEN, K. B., DENAS, O., TAYLOR, J., BLOBEL, G. A., WEISS, M. J., PIMKIN, M., DENG, W., MARINOV, G. K., WILLIAMS, B. A., FISHER-AYLOR, K. I., DESALVO, G., KIRALUSHA, A., TROUT, D., AMRHEIN, H., MORTAZAVI, A., EDSALL, L., MCCLEARY, D., KUAN, S., SHEN, Y., YUE, F., YE, Z., DAVIS, C. A., ZALESKI, C., JHA, S., XUE, C., DOBIN, A., LIN, W., FASTUCA, M., WANG, H., GUIGO, R., DJEBALI, S., LAGARDE, J., RYBA, T., SASAKI, T., MALLADI, V. S., CLINE, M. S., KIRKUP, V. M., LEARNED, K., ROSENBLOOM, K. R., KENT, W. J., FEINGOLD, E. A., GOOD, P. J., PAZIN, M., LOWDON, R. F. & ADAMS, L. B. 2012. An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol*, 13, 418.
- MULLER, H., ORWIG, K. E. & SOARES, M. J. 1998. Identification of two new members of the mouse prolactin gene family. *Biochim Biophys Acta*, 1396, 251-8.
- MUNGER, S. C., AYLOR, D. L., SYED, H. A., MAGWENE, P. M., THREADGILL, D. W. & CAPEL, B. 2009. Elucidation of the transcription network governing mammalian sex determination by exploiting strain-specific susceptibility to sex reversal. *Genes Dev*, 23, 2521-36.
- MUNGER, S. C., NATARAJAN, A., LOOGER, L. L., OHLER, U. & CAPEL, B. 2013. Fine time course expression analysis identifies cascades of activation and repression and maps a putative regulator of mammalian sex determination. *PLoS Genet*, 9, e1003630.
- NAGAMINE, C. M., MOROHASHI, K., CARLISLE, C. & CHANG, D. K. 1999. Sex reversal caused by *Mus musculus domesticus* Y chromosomes linked to variant expression of the testis-determining gene *Sry*. *Dev Biol*, 216, 182-94.
- NAGAMINE, C. M., TAKETO, T. & KOO, G. C. 1987a. Morphological development of the mouse gonad in *tda-1* XY sex reversal. *Differentiation*, 33, 214-22.
- NAGAMINE, C. M., TAKETO, T. & KOO, G. C. 1987c. Studies on the genetics of *tda-1* XY sex reversal in the mouse. *Differentiation*, 33, 223-31.
- 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., PARKER, K. L. & VASSALLI, J. D. 2005. Gene expression during sex determination reveals a robust female genetic program at the onset of ovarian development. *Dev Biol*, 287, 361-77.
- NEL-THEMAAT, L., VADAKKAN, T. J., WANG, Y., DICKINSON, M. E., AKIYAMA, H. & BEHRINGER, R. R. 2009. Morphometric analysis of testis cord formation in *Sox9*-EGFP mice. *Dev Dyn*, 238, 1100-10.
- NIEHRS, C. & SCHAFFER, A. 2012. Active DNA demethylation by *Gadd45* and DNA repair. *Trends Cell Biol*, 22, 220-7.
- NIKOLOVA, G., SINSHEIMER, J. S., EICHER, E. M. & VILAIN, E. 2008. The chromosome 11 region from strain 129 provides protection from sex reversal in XYPOS mice. *Genetics*, 179, 419-27.
- NISHINO, K., HATTORI, N., SATO, S., ARAI, Y., TANAKA, S., NAGY, A. & SHIOTA, K. 2011. Non-CpG methylation occurs in the regulatory region of the *Sry* gene. *J Reprod Dev*, 57, 586-93.
- NISHINO, K., HATTORI, N., TANAKA, S. & SHIOTA, K. 2004. DNA methylation-mediated control of *Sry* gene expression in mouse gonadal development. *J Biol Chem*, 279, 22306-13.
- OZSOLAK, F. & MILOS, P. M. 2011. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet*, 12, 87-98.
- PALMER, S. J. & BURGOYNE, P. S. 1991. The *Mus musculus domesticus* *Tdy* allele acts later than the *Mus musculus musculus* *Tdy* allele: a basis for XY sex-reversal in C57BL/6-YPOS mice. *Development*, 113, 709-14.
- PANNETIER, M., TILLY, G., KOCER, A., HUDRISIER, M., RENAULT, L., CHESNAIS, N., COSTA, J., LE PROVOST, F., VAIMAN, D., VILOTTE, J. L. & PAILHOX, E. 2006. Goat *SRY* induces testis development in XX transgenic mice. *FEBS Lett*, 580, 3715-20.

- PARK, P. J. 2009. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*, 10, 669-80.
- PARMA, P., RADI, O., VIDAL, V., CHABOISSIER, M. C., DELLAMBRA, E., VALENTINI, S., GUERRA, L., SCHEDL, A. & CAMERINO, G. 2006. R-spondin1 is essential in sex determination, skin differentiation and malignancy. *Nat Genet*, 38, 1304-9.
- PEARLMAN, A., LOKE, J., LE CAIGNEC, C., WHITE, S., CHIN, L., FRIEDMAN, A., WARR, N., WILLAN, J., BRAUER, D., FARMER, C., BROOKS, E., ODDOUX, C., RILEY, B., SHAJAHAN, S., CAMERINO, G., HOMFRAY, T., CROSBY, A. H., COUPER, J., DAVID, A., GREENFIELD, A., SINCLAIR, A. & OSTRER, H. 2010. Mutations in MAP3K1 cause 46,XY disorders of sex development and implicate a common signal transduction pathway in human testis determination. *Am J Hum Genet*, 87, 898-904.
- PRAGER, E. M., BOURSOT, P. & SAGE, R. D. 1997. New assays for Y chromosome and p53 pseudogene clines among East Holstein house mice. *Mammalian Genome*, 8, 279-281.
- RAMPALLI, S., LI, L., MAK, E., GE, K., BRAND, M., TAPSCOTT, S. J. & DILWORTH, F. J. 2007. p38 MAPK signaling regulates recruitment of Ash2L-containing methyltransferase complexes to specific genes during differentiation. *Nat Struct Mol Biol*, 14, 1150-6.
- RAYMOND, C. S., MURPHY, M. W., O'SULLIVAN, M. G., BARDWELL, V. J. & ZARKOWER, D. 2000. Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. *Genes Dev*, 14, 2587-95.
- ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139-40.
- RUBINSTEIN, M., JAPON, M. A. & LOW, M. J. 1993. Introduction of a point mutation into the mouse genome by homologous recombination in embryonic stem cells using a replacement type vector with a selectable marker. *Nucleic Acids Res*, 21, 2613-7.
- SALIBA, A. E., WESTERMANN, A. J., GORSKI, S. A. & VOGEL, J. 2014. Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res*, 42, 8845-60.
- SEKIDO, R., BAR, I., NARVAEZ, V., PENNY, G. & LOVELL-BADGE, R. 2004. SOX9 is up-regulated by the transient expression of SRY specifically in Sertoli cell precursors. *Dev Biol*, 274, 271-9.
- SEKIDO, R. & LOVELL-BADGE, R. 2008. Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature*, 453, 930-4.
- SEKIDO, R. & LOVELL-BADGE, R. 2009. Sex determination and SRY: down to a wink and a nudge? *Trends Genet*, 25, 19-29.
- SEN, G. L., REUTER, J. A., WEBSTER, D. E., ZHU, L. & KHAVARI, P. A. 2010. DNMT1 maintains progenitor function in self-renewing somatic tissue. *Nature*, 463, 563-7.
- SIGGERS, P., CARRE, G. A., BOGANI, D., WARR, N., WELLS, S., HILTON, H., ESAPA, C., HAJIHOSEINI, M. K. & GREENFIELD, A. 2014. A novel mouse Fgfr2 mutant, hobbyhorse (hob), exhibits complete XY gonadal sex reversal. *PLoS One*, 9, e100447.
- SIMONE, C., FORCALES, S. V., HILL, D. A., IMBALZANO, A. N., LATELLA, L. & PURI, P. L. 2004. p38 pathway targets SWI-SNF chromatin-remodeling complex to muscle-specific loci. *Nat Genet*, 36, 738-43.
- SINCLAIR, A. H., BERTA, P., PALMER, M. S., HAWKINS, J. R., GRIFFITHS, B. L., SMITH, M. J., FOSTER, J. W., FRISCHAUF, A. M., LOVELL-BADGE, R. & 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, 240-4.
- SONG, L. & CRAWFORD, G. E. 2010. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc*, 2010, pdb prot5384.
- SOUTHERN, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol*, 98, 503-17.
- SUTTON, E., HUGHES, J., WHITE, S., SEKIDO, R., TAN, J., ARBOLEDA, V., ROGERS, N., KNOWER, K., ROWLEY, L., EYRE, H., RIZZOTI, K., MCANINCH, D., GONCALVES, J., SLEE, J., TURBITT,

- E., BRUNO, D., BENGTSSON, H., HARLEY, V., VILAIN, E., SINCLAIR, A., LOVELL-BADGE, R. & THOMAS, P. 2011. Identification of SOX3 as an XX male sex reversal gene in mice and humans. *J Clin Invest*, 121, 328-41.
- SWAIN, A. & LOVELL-BADGE, R. 1999. Mammalian sex determination: a molecular drama. *Genes Dev*, 13, 755-67.
- SWAIN, A., NARVAEZ, V., BURGOYNE, P., CAMERINO, G. & LOVELL-BADGE, R. 1998. Dax1 antagonizes Sry action in mammalian sex determination. *Nature*, 391, 761-7.
- TAKEKAWA, M. & SAITO, H. 1998. A family of stress-inducible GADD45-like proteins mediate activation of the stress-responsive MTK1/MEKK4 MAPKKK. *Cell*, 95, 521-30.
- TAM, P. P. 1981. The control of somitogenesis in mouse embryos. *J Embryol Exp Morphol*, 65 Suppl, 103-28.
- TEVOSIAN, S. G., ALBRECHT, K. H., CRISPINO, J. D., FUJIWARA, Y., EICHER, E. M. & 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, 4627-34.
- THALHEIMER, F. B., WINGERT, S., DE GIACOMO, P., HAETSCHER, N., REHAGE, M., BRILL, B., THEIS, F. J., HENNIGHAUSEN, L., SCHROEDER, T. & RIEGER, M. A. 2014. Cytokine-regulated GADD45G induces differentiation and lineage selection in hematopoietic stem cells. *Stem Cell Reports*, 3, 34-43.
- TIWARI, V. K., STADLER, M. B., WIRBELAUER, C., PARO, R., SCHUBELER, D. & BEISEL, C. 2012. A chromatin-modifying function of JNK during stem cell differentiation. *Nat Genet*, 44, 94-100.
- UHLENHAUT, N. H., JAKOB, S., ANLAG, K., EISENBERGER, T., SEKIDO, R., KRESS, J., TREIER, A. C., KLUGMANN, C., KLASSEN, C., HOLTER, N. I., RIETHMACHER, D., SCHUTZ, G., COONEY, A. J., LOVELL-BADGE, R. & TREIER, M. 2009. Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. *Cell*, 139, 1130-42.
- VIDAL, V. P., CHABOISSIER, M. C., DE ROOIJ, D. G. & SCHEDL, A. 2001. Sox9 induces testis development in XX transgenic mice. *Nat Genet*, 28, 216-7.
- VRIZ, S., GRIFFITHS, B. L., HARLEY, V., GOODFELLOW, P. & LOVELL-BADGE, R. 1995. The SRY protein, like HMG 1, recognizes (CA)_n sequences, an abundant repeat sequence in vertebrates. *Biochem Mol Biol Int*, 37, 1137-46.
- WANG, C., GONG, B., BUSHEL, P. R., THIERRY-MIEG, J., THIERRY-MIEG, D., XU, J., FANG, H., HONG, H., SHEN, J., SU, Z., MEEHAN, J., LI, X., YANG, L., LI, H., LABAJ, P. P., KREIL, D. P., MEGHERBI, D., GAJ, S., CAIMENT, F., VAN DELFT, J., KLEINJANS, J., SCHERER, A., DEVANARAYAN, V., WANG, J., YANG, Y., QIAN, H. R., LANCASHIRE, L. J., BESSARABOVA, M., NIKOLSKY, Y., FURLANELLO, C., CHERICI, M., ALBANESE, D., JURMAN, G., RICCADONNA, S., FILOSI, M., VISINTAINER, R., ZHANG, K. K., LI, J., HSIEH, J. H., SVOBODA, D. L., FUSCOE, J. C., DENG, Y., SHI, L., PAULES, R. S., AUERBACH, S. S. & TONG, W. 2014. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat Biotechnol*, 32, 926-32.
- WANG, L., XUE, Y., SHEN, Y., LI, W., CHENG, Y., YAN, X., SHI, W., WANG, J., GONG, Z., YANG, G., GUO, C., ZHOU, Y., WANG, X., ZHOU, Q. & ZENG, F. 2012. Claudin 6: a novel surface marker for characterizing mouse pluripotent stem cells. *Cell Res*, 22, 1082-5.
- WANG, Z., GERSTEIN, M. & SNYDER, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 10, 57-63.
- WARR, N., BOGANI, D., SIGGERS, P., BRIXEY, R., TATEOSSIAN, H., DOPPLAPUDI, A., WELLS, S., CHEESEMAN, M., XIA, Y., OSTRER, H. & GREENFIELD, A. 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., TEDESCO, M., DEL BARCO BARRANTES, I., NEBREDA, A. R., TRAINOR, P. A. & GREENFIELD, A. 2012. Gadd45gamma and Map3k4

- interactions regulate mouse testis determination via p38 MAPK-mediated control of Sry expression. *Dev Cell*, 23, 1020-31.
- WARR, N., SIGGERS, P., CARRE, G. A., BOGANI, D., BRIXEY, R., AKIYOSHI, M., TACHIBANA, M., TEBOUL, L., WELLS, S., SANDERSON, J. & GREENFIELD, A. 2014. Transgenic expression of Map3k4 rescues T-associated sex reversal (Tas) in mice. *Hum Mol Genet*, 23, 3035-44.
- WASHBURN, L. L. & EICHER, E. M. 1983. Sex reversal in XY mice caused by dominant mutation on chromosome 17. *Nature*, 303, 338-40.
- WASHBURN, L. L. & EICHER, E. M. 1989. Normal testis determination in the mouse depends on genetic interaction of a locus on chromosome 17 and the Y chromosome. *Genetics*, 123, 173-9.
- WILHELM, D. & ENGLERT, C. 2002. The Wilms tumor suppressor WT1 regulates early gonad development by activation of Sf1. *Genes Dev*, 16, 1839-51.
- WILHELM, D. & KOOPMAN, P. 2006. The makings of maleness: towards an integrated view of male sexual development. *Nat Rev Genet*, 7, 620-31.
- WILHELM, D., MARTINSON, F., BRADFORD, S., WILSON, M. J., COMBES, A. N., BEVERDAM, A., BOWLES, J., MIZUSAKI, H. & KOOPMAN, P. 2005. Sertoli cell differentiation is induced both cell-autonomously and through prostaglandin signaling during mammalian sex determination. *Dev Biol*, 287, 111-24.
- WILHELM, D., PALMER, S. & KOOPMAN, P. 2007. Sex determination and gonadal development in mammals. *Physiol Rev*, 87, 1-28.
- WILHELM, D., WASHBURN, L. L., TRUONG, V., FELLOUS, M., EICHER, E. M. & KOOPMAN, P. 2009. Antagonism of the testis- and ovary-determining pathways during ovotestis development in mice. *Mech Dev*, 126, 324-36.
- WOODS, A., WANG, G. & BEIER, F. 2005. RhoA/ROCK signaling regulates Sox9 expression and actin organization during chondrogenesis. *J Biol Chem*, 280, 11626-34.
- XIA, Y., MAKRI, C., SU, B., LI, E., YANG, J., NEMEROW, G. R. & KARIN, M. 2000. MEK kinase 1 is critically required for c-Jun N-terminal kinase activation by proinflammatory stimuli and growth factor-induced cell migration. *Proc Natl Acad Sci U S A*, 97, 5243-8.
- ZHANG, L., DENG, M., PARTHASARATHY, R., WANG, L., MONGAN, M., MOKKENTIN, J. D., ZHENG, Y. & XIA, Y. 2005. MEKK1 transduces activin signals in keratinocytes to induce actin stress fiber formation and migration. *Mol Cell Biol*, 25, 60-5.
- ZHANG, R. P., SHAO, J. Z. & XIANG, L. X. 2011. GADD45A protein plays an essential role in active DNA demethylation during terminal osteogenic differentiation of adipose-derived mesenchymal stem cells. *J Biol Chem*, 286, 41083-94.
- ZHAO, L., NG, E. T., DAVIDSON, T. L., LONGMUSS, E., URSCHITZ, J., ELSTON, M., MOISYADI, S., BOWLES, J. & KOOPMAN, P. 2014. Structure-function analysis of mouse Sry reveals dual essential roles of the C-terminal polyglutamine tract in sex determination. *Proc Natl Acad Sci U S A*, 111, 11768-73.
- ZHAO, L., SVINGEN, T., NG, E. T. & KOOPMAN, P. 2015. Female-to-male sex reversal in mice caused by transgenic overexpression of Dmrt1. *Development*, 142, 1083-8.

Appendix 1: Sry^{POS} sequence reads

[illegible]

	ATAGGAGGGAAAAAAATCACAAATGACATGGAGAAACAATAGCAGAAACTCTCCAGAGAAAGAGACATAAGCACCCATTGGGATGGATATATTTGTATTCATTTGTGTTAGAAAATCTATCCAAAGGCTTA TAAAGTCTTATTGATTGTTGGCATTGAAATGGTGCTGAAACATACCCCTTAAAAGAAAACTAACCACCTAGAAAGTCCAGTAGTGCCAGCTAACTGTTCCATTGAATTCAGAGAACAGTCTTTAACTGATTT ACAGAAAAACAATACATGTAAGTTACTGTTACAATACATAATTTGAAAACTCACTAAAACCTCAAAGGTCTTAAGACCATGTAGAACAGTAGTTAATTGTTTCCAAGAGTTCAATACATCAGATAATACACC ATGTTGTTGTATAGTATTTAGTGATAAGATGTCACAGTTGAGGCTCCTTGGCTGCTATTTCTTCCAACCTATGTTTATAGAGTTGACAGGGCTACATAATCAATCAAAGGTCATGGGTAACAGAGTTATTA TGAGGGAAGCATCGAGGGAGATTAATGTGATGAAATTGGAACAGGCTAACTGAAGGTCTGGGTATGTAACATGGTGTTAACAAGCCTATGTGTAGTTCCTTGGTCTTTCTTAAAGAACTGTTGCAGGGGT TGCTGTGTACACTCTTTATGTCTCTTTAACTTGTGTTGTAAGAACAAGAGTGGGTAACCTTGAACCACTCTTTAGTTTAACTCTGGCAGTTGAGTTAATGTGCAGATGGCCATTCAATCATCCACATATACT TGCCCTCCCTTACCTTTTGTATTCCAAATCATTTTGTATATGTGGTGTCTTAAGTGTTGGTACATCCACTTAGCTTACTATTTGTCCATTTTGTCTGATTCTTAGCTAGCCCTTGAAAAACAGCTTTGC TGTATGTCAATAAAACAGTCTATTCTTTAGAGATAGATAATTA
	GTCGTACCGTGGTGTTAATGAATTTGATGGTCTGCCCTTGGTTTGGATTTCTTGGTGTGGCCCATGAAAAGTGCATAGCAGGAAGTATCAGCAGAGGTATGTGTGTAGGGGGATGGCATCTAACAAAAGA ACAGCATCTCATGAGCTCTCCAAGAATAGCCACCAATCCCAGCACTACCTCTAAGCAGATAGTCGAGTGAACATTCTGGAGAATGCTAAAGGTATATAAATCTTAGGCAGAACACTGCTTGCAAAATTCAA GCCAGAATTAACCTGAAAGCTCTCTTGACACAAGGATACTCTGTGCACAAAGATGAAGGAGAGAGAGGAGGAGGAAGAAGAAAAGAGGAGAAGGAGGAGGAGGAAGAAGAGGAGAAAGAAAGAGG AGGAAGAAAAGAGGAGGAGGAGGAGAGAGGAGGAAGATATGTAGCAAGTAGGTACAGGTGAAAAAGAGGAAGCCTGTGGAATGCCAAAGGATCAATGAGGTCCTTGTCTGTATCTATGGCTCTCTGCAT ACAATTCACACAGAGGAGAATAAAATCCACAAATGACATGGAGAAACAATAGCAGAACTCTCCAGAGAAAGAGACATAAGCACCCATTGGGATGGATATATTTGTATTCATTTGTGTTAGAAAATCTATCC AAAGGCTTATAAAGTCTTATTGATTGTTGGCATTGAAATGGTGCTGAAACATACCCCTTAAAAGAAAACTAACCACCTAGAAAGTCCAGTAGTGCCAGCTAACTGTTCCATTGAATTCAGAGAACAGTCTTTT AACTGATTTACAGAAAACAATACATGTAAGTTACTGTTACAATACATAATTTTAAAAATCACTAAAACCTCAAAGGTCTTAAGACCATGTAGAACAGTAGTTAATTGTTTCCAAG
	GTGTAGTTCCTTGGTCTTTCTTAAAGAACTGTTGCAGGGTGTGCTGTGTACACTCTTTATGTCTCTTTAACTTGTTTGAAGAACAAGAGTGGGTAACCTTTGAACCATTCTTTAGTTTAACTCTGGCAGT TGAGTTAATGTGCAGATGGCCATTCAATCATCCACATATACTTGCCCTCCCTTACCTTTTGTATTATCCAAATCATTTTGTATATGTGGTGTCTTAAGTGTTGGTACATCCACTTAGCTTACTATTTGTCCA TTTTGTTTCTGATTCTTAGCTAGCCCTTGAAAAACAGCTTTGCTGTATGTCAATAAAACAGTCTATTCTTTAGAGATAGATAATAAAGAGAGAAGCAAAATGTGACTCCTTATGCATAAAGGCTATTTTACCA AAAGATATAGAGAGCTCATAGCATAGTTTATAGAACAACCTGGGCTTTGCACATTGTGGAGGAGAACTAAAATGTCATACACACACACACACACACACACACACACACACACACACACACACACAC
	GGGAAGAAGAGAGGAGAGGAGGAGGAGGAGGAAGAAGAGGAGAAAGAAAGAGGAGGAAGAAGAGGAGGAGGAGGAGGAGACAGAGGAAGATATGTAGCAAGTAGGTACAGGTGAAAAAGAGGAAG CCTGTGGAATGCCAAAGGATCAATGAGGTCCTTGTCTGTATCTATGGCTCTCTGCATACGATTACACAGAGGAGAATAAAATCCACAAATGACATGGAGAAACAATAGCAGAACTCTCCAGAGAAAGA GACATAAGCACCCATTGGGATGGATATATTTGTATTCATTTGTGTTAGAAAATCTATCCAAAGGCTTATAAAGTCTTATTGATTGTTGGCATTGAAATGGTGCTGAAACATACCCCTTAAAAGAAAACTAACCA CTTAGAAGTCCAGTAGTGCCAGCTAACTGTTCCATTGAATTCAGAGAACAGTCTTTAACTGATTTACAGAAAACAATACATGTAAGTTACTGTTACAGATACATAATTTTAAAAATCACTAAAACCTCAAAG GTCTTAAGACCATGTAGAACAGTAGTTTAATTGTTTCCAAGAGTTCAATACATCAGATAATACACCATGTTGTTGTATAGTATTTAGTGATAAGATGTACAGTTGAGGCTCCTTGGCTGCTATTTTCTTCCA ACCTATGTTTTAGAGTTGACAGGGCTACATAATCAATTCAAAGTTTATGGGTAACAGAGTTATTATGAGGGAAGCATCGAGGGAGATTAATGTGATGAAATTGGAACAGGCTAACTGAAGGTCTGGGTAT GTAACATGGTGTTAACGAGCCTATGTGTAGTTCCTTGGTCTTTCTTAAAGAACTGTTGCAGGGGTTGCTGTGTACACTCTTTATGTCTCTTTAACTTGTTTGAAGAAC
	ATTTGATACTAACTCAAAGGCTCTTAAGACCATGTAGAACAGTAGTTAATTGTTTCCAAGAGTTCAATACATCAGATAATACACCATGTTGTTGTATAGTATTTAGTGATAAGATGTCACAGTTGAGGCT CCTTGGCTGCTATTTTCTTCCAACCTATGTTTATAGAGTTGACAGGGCTACATAATCAATCAAAGGTCATGGGTAACAGAGTTATTATGAGGGAAGCATCGAGGGAGATTAATGTGATGAAATTGGAACAG GCTAACTGAAGGTCTGGGTATGTAACATGGTGTTAACAAGCCTATGTGTAGTTCCTTGGTCTTTCTTAAAGAACTGTTGCAGGGGTTGCTGTGTACACTCTTTATGTCTCTTTAACTTGTTTGAAGAACA GAGTGGGTAACCTTTGAACCACTCTTATGTTTAACTCTGGCAGTTGAGTTAATGTGCAGATGGCCATTCAATCATCCACATATACTTGCCCTCCCTTACCTTTTGTATTCCAAATCATTTTGTATATGTG GTGTCTTAAGTGTTGGTACATCCACTTAGCTTACTATTTGTCCATTTTGTCTGATTCTTAGCTAGCCCTTGAAAAACAGCTTTGCTGTATGTCAATAAAACAGTCTATTCTTTAGAGATAGATAATAAAG AGAAGCAAAATGTGACTCCTTATGCATAAAGGCTATTTTACCAAAAGATATAGAGAGCTCATAGCATAGTTTATAGAACAACCTGGGCTTTGCACATTGTGGAGGAGAACTAAAATGTCATACACACACACA CACACACACACACACACACA

[illegible]

	GTGGCATTTAATCCACAGAAGAAACCAGAGATTTACACTCAGATCCAATTTAGATGAAATGCATTTATTCTTAGGCTAGCACAAAGACAGCAAACAGCTTTGTGGTCAATACAGCATGCCATGCTAAAGTGG GGGCTAGGGGTGAGTAAGCTGGANT
--	---

Appendix 2: Alignment of Sry^{POS} and Sry^{TIR} open reading frames

TIR	GACAATTGTCTAGAGGCATGGAGGGCCATGTCAAGCGCCCCATGAATGCATTTATGGTG
POS	-----ATGGAGGGCCATGTCAAGCGCCCCATGAATGCATTTATGGTG

TIR	TGGTCCCGTGGTGAGAGGCACAAGTTGGCCCAGCAGAATCCAGCATGCAAAATACAGAG
POS	TGGTCCCGTGGTGAGAGGCACAAGTTGGCCCAGCAGAATCCAGCATGCAAAATACAGAG

TIR	ATCAGCAAGCAGCTGGGATGCAGGTGAAAAGCCTTACAGAAGCCGAAAAAGGCCCTTT
POS	ATCAGCAAGCAGCTGGGATGCAGGTGAAAAGCCTTACAGAAGCCGAAAAAGGCCCTTT

TIR	TTCCAGGAGGCACAGAGATTGAAGACCCCTACACAGAGAGAAATACCCAAACTATAAATAT
POS	TTCCAGGAGGCACAGAGATTGAAGACCCCTACACAGAGAGAAATACCCAAACTATAAATAT

TIR	CAGCCTCATCGGAGGGCTAAAGTGTACAGAGGAGTGGCATTTCACAGCCTGCAGTTGCC
POS	CAGCCTCATCGGAGGGCTAAAGTGTACAGAGGAGTGGCATTTCACAGCCTGCAGTTGCC

TIR	TCAACAAAACGTACAACCTTCTGCAGTGGGACAGGAACCCACATGCCATCACATACAGG
POS	TCAACAAAACGTACAACCTTCTGCAGTGGGACAGGAACCCACATGCCATCACATACAGG

TIR	CAAGACTGGAGTAGAGCTGCACACCTGTACTCCAAAACCAGCAAAGCTTTTATTTGCAG
POS	CAAGACTGGAGTAGAGCTGCACACCTGTACTCCAAAACCAGCAAAGCTTTTATTTGCAG

TIR	CCTGTCGATATCCCCACTGGGCACCCGACGAGCAGCAGCAGCAGCAGCAGCAGTTCAT
POS	CCTGTCGATATCCCCACTGGGCACCCGACGAGCAGCAGCAGCAGCAGCAGCAGTTCAT

TIR	AACCACCACCAGCAGCAACAGCAGTTCTATGACCACCAGCAGCAGCAGCAGCAGCAGCAG
POS	AACCACCACCAGCAGCAACAGCAGTTCTATGACCACCAGCAGCAGCAGCAGCAGCAGCAG

TIR	CAGCAGCAGTTCCATGACCACCACCAGCAGAAGCAGCAGTTTCATGACCACCACCAGCAG
POS	CAGCAGCAGTTCCATGACCACCACCAGCAGAAGCAGCAGTTTCATGACCACCACCAGCAG

TIR	CAACAGCAGTTCCATGACCACCACCACCACCAGCAGCAGCAGCAGTTCCATGACCACCAC
POS	CAACAGCAGTTCCATGACCACCACCACCACCAGCAGCAGCAGCAGTTCCATGACCACCAC

TIR	CAGCAGCAACAGCAGTTCCATGACCACCAGCAGCAGCAGCAGCAGCAGTAGCAGCAGCAG
POS	CAGCAGCAACAGCAGTTCCATGACCACCAGCAGCAGCAGCAGCAGCAGTAGCAGCAGCAG

TIR	TTCCATGACCACCACCAGCAGAAGCAGCAGTTCCATGACCACCACCACCACCAACAGCAG
POS	TTCCATGACCACCACCAGCAGAAGCAGCAGTTCCATGACCACCACCACCACCAACAGCAG

TIR	CAGCAGTTC-CATGACCACCACCAGCAGCAGCAGCAGCAGTTCATGACCACCACCAGCA
POS	CAGCAGTTCATGACCACCACCAGCAGCAGCAGCAGCAGTTCATGACCACCACCAGCA

TIR	GCAGCAGCAGCAGTTCATGACCCTCCAGCAGAAGCAGCAGTTCATGACCACCCCA
POS	GCAGCAGCAGCAGTTCATGACCCTCCAGCAGAAGCAGCAGTTCATGACCACCCCA

TIR	GCAGCAACAGCAGTTCATGACCACCACCACCAGCAGCAACAGCAGTTCATGACCACCA
POS	GCAGCAACAGCAGTTCATGACCACCACCACCAGCAGCAACAGCAGTTCATGACCACCA

TIR	CCACCAGCAGCAGCAGAAGCAGCAGTTCATGACCACCACCAGCAGAAGCAGCAGTTC
POS	CCACCAGCAGCAGCAGAAGCAGCAGTTCATGACCACCACCAGCAGAAGCAGCAGTTC

[illegible]

[illegible]

B6 ATGACTGGGATGCAGTAGTTTCATTTCTGAGCAGGCCAAAAACACAGTGCCATCGGGTCAGG
POS ATGACTGGGATGCAGTAGTTTCATTTCTGAGCAGGCCAAAAACACAGTGCCATCGGGTCAGG

B6 GCTGGCTGTCATTTCTTACAGCATCACCATGCAAACCATCTGTAAGTCTTCAATACTCA
POS GCTGGCTGTCATTTCTTACAGCATCACCATGCAAACCATCTGTAAGTCTTCaATACTCA

B6 TCTTTGAAAAACATACCAAACCTTAATACTGTGAAAAACACAGGCTGTGTCTCTTTAGTATG
POS TCCTTTGAAAAACATACCAAACCTTAATACTGTGAAAAACACAGGCTGTGTCTCTTTAGTATG

B6 CAAACACTCCTTTGTTGATCAAATGCATCTTTCTGTTTATCAGGGTTTCTCTAGCAC
POS CAAACACTCCTTTGTTGATCAAATGCATCTTTCTGTTTATCAGGGTTTCTCTAGCAC

[illegible]

B6
POS

B6 AATGTGCAAGCCCAGTTGTTCTATGAACATGCTATGAGCTCTCTATATCTTTTGGTAA
POS AATGTGCAAGCCCAGTTGTTCTATGAACATGCTATGAGCTCTCTATATCTTTTGGTAA

B6 AATAAGCCTTTATGCATAAGGAGTCACATTTTGCTTCTCTTTATTATCTATCTCTAAAG
POS AATAAGCCTTTATGCATAAGGAGTCACATTTTGCTTCTCTTTATTATCTATCTCTAAAG

B6 AATAGACTGTTTTATTGACATACAGCAAAGCTGTGTTTTCAAGGGCTAGCTAAGAATCAG
POS AATAGACTGTTTTATTGACATACAGCAAAGCTGTGTTTTCAAGGGCTAGCTAAGAATCAG

B6 AAACAAATGGACAAATAGTCAAGCTAAGTGGATGTACCAACACTTAAGACACCACATAT
POS AAACAAATGGACAAATAGTCAAGCTAAGTGGATGTACCAACACTTAAGACACCACATAT

B6 ACAAATGATTGGAATAAACAAAAGGTGAAGGGGAGGGCAAGTATATGTGGGATGAATG
POS ACAAATGATTGGAATAAACAAAAGGTGAAGGGGAGGGCAAGTATATGTGGGATGAATG

B6 AATGCCATCTGCACATTAACtCAACTGCCAGATTGTTAAgCTAAgGAATGGTTCAAAGT
POS AATGCCATCTGCACATTAACtCaACTGCCAGATTGTTAACTAAAGAATGGTTCAAAGT

B6 TACCCACTCTTGTTCTTACAAACAAGTTTAAAGAGACATAAAGAGTGTACACAGCAACCC
POS TACCCACTCTTGTTCTTACAAACAAGTTTAAAGAGACATAAAGAGTGTACACAGCAACCC

B6 CTGCAACAGTTTCTTTAAGAAAGACCAAGGAACACACATAGGCTTGTTAACACCATGTT
POS CTGCAACAGTTTCTTTAAGAAAGACCAAGGAACACACATAGGCTTGTTAACACCATGTT

B6 ACATACCCAGACCTTCAGTTAGCCTGTTCCAATTTTCATCACATTAATCTCCCTCGATGCT
POS ACATACCCAGACCTTCAGTTAGCCTGTTCCAATTTTCATCACATTAATCTCCCTCGATGCT

B6 TCCCTCATAATAACTCTGTTACCCATGAACTTTTGAATTGATTATGTAGCCCTGTCAACT
POS TCCCTCATAATAACTCTGTTACCCATGAACTTTTGAATTGATTATGTAGCCCTGTCAACT

B6 CTA AACATAGGTTGGAAGAAAATAGCAGCCAAGGAGCCTCAACTGTGACATCTTATACA
POS CTA AACATAGGTTGGAAGAAAATAGCAGCCAAGGAGCCTCAACTGTGACATCTTATACA

B6 CTAATACTATACAACAACATGGTGTATTATCTGATGTATTGAACTCTTGGAACAATTA
POS CTAATACTATACAACAACATGGTGTATTATCTGATGTATTGAACTCTTGGAACAATTA

B6 AACTACTGTTCTACATGGTCTTAAGACCTTTGAGTTTTAGTTGATTTTCAAATTATGTA
POS AACTACTGTTCTACATGGTCTTAAGACCTTTGAGTTTTAGTTGATTTTCAAATTATGTA

B6 TTGTGAACAGTAAC TTACATGTATTGTTTCTGTAAATCAGTTAAAAGACTGTTCTCTAG
POS TTGTGAACAGTAAC TTACATGTATTGTTTCTGTAAATCAGTTAAAAGACTGTTCTCTAG

B6 AATTCAATGGAACAGTTAGCTGGCACTACTGGACTTCTAAGTGTTAGTTTTCTTTTAA
POS AATTCAATGGAACAGTTAGCTGGCACTACTGGACT-----

Appendix 4: SRY^{AKR} sequence reads

[illegible]

	GTGAAACACAGGCTGTGTCTCTTTAGTATGCAAAACACTCCTTTGTTGATCAAATGCATCTTTCTGTTTATCAGGGTTTCTCTCTAGCACACAAAACACTACATCATGAGTATGTAAAGTCTT GTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGtnngacattttannTctcctCcacnnTgtGCA
	TtnncTCCCTCGATGCTTCCCTCaTAATAACTCTGTTACCCATGAACTTTTGAATTGATtAtGTAGCCCTGTCAACTCTAAAACATAGGTTGGAAGAAAATAGCAGCCAAGGAGCCTCAACT GTGACATCTTATACACTAAATACTATAACAACATGGTGTATTATCTGATGTATTGAACTCTTGGAAACAATTAACTACTGTTCTACATGGTCTTAAGACCTTTGAGTTTTAGTTGATTT TCAAAATTATGTATTGTGAACAGTAACTTACATGTATTGTTTTCTGTAAATCAGTAAAAGACTGTTCTCTAGAATTCAATGGAACAGTTAGCTGGCac caTACCaAACTTAATACTGTGAAACACAGGCTGTGTCTCTtAgTAtGCAAAACACTCCTTTGTTGATCAAATGCATCTTTCTGTTTATCAGGGTTTCTCTCTAGCACACAAAACACTACATC ATGAGTATGTAAAGTCTTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGtatgAcaTTTtatctCcCctccncngtGtGcaaagC