Deciphering the Transcriptional Mechanisms and Function of SOX3 in the Developing Embryonic Mouse Brain and Postnatal Testes

Dale Christopher McAninch

B.Sc. (Molecular Biology), Honours (Biochemistry)

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Department of Molecular and Cellular Biology

School of Biological Sciences

University of Adelaide, Australia

March 2020

TABLE OF CONTENTS

Table of Contents ii
Declaration v
Acknowledgements vi
Abstract vii
Abbreviations viii
Publications and Manuscripts x
Awards and scholarships xii
Presentations xiii
Chapter 1 Introduction 1
1.1.1 Gene transcription and transcription factors 2
1.1.2 Pioneer Factors: transcription factors with an extra twist 5
1.1.3 Characterising transcription factors 8
1.2.1 SOX transcription factor family 11
1.2.2 SOX3 activates and represses transcription 16
1.2.3 Expression of *Sox3* in embryonic and adult neural tissues 18
1.2.4 The function of SOX3 in neurogenesis 21
1.2.5 Binding partners and cofactors of SOX proteins 22
1.2.6 Function of SOX3 in mice 23
1.2.7 Overlapping expression and functional redundancy of SOXB1 proteins 27
1.2.8 *SOX3* and human disorders 31
1.3.1 Spermatogenesis in mice 33
1.3.2 Expression and function of Sox3 in mouse testes 38
1.4 Project rationale and aims 39
Chapter 2 Summary of Results 40
2.1 Characterising the global transcriptional mechanisms of SOX3 in neural progenitor cells 41
2.2 Exploring the molecular mechanisms of SOX3 in mouse testis 42
2.2.1 The role of SOX3 in type A spermatogonia 42
2.2.1 Characterising the global transcriptional mechanisms of SOX3 in mouse testes 43
Chapter 3 Publication I 44
3.0 Statement of Authorship 45
Chapter 4 Manuscript I 56
4.0 Statement of Authorship 57
4.1 Abstract 63
4.2 Introduction 63
4.3 Results 67
4.4 Discussion 86
4.5 Materials and Methods 90

4.6 Acknowledgements ... 92
4.7 Competing Interests Statement ... 93
4.8 Author Contributions ... 93
4.9 Supplementary Figures ... 94
4.10 References ... 97

Chapter 5 Manuscript II ... 102

5.0 Statement of Authorship ... 103
5.1 Abstract ... 105
5.2 Introduction ... 106
5.3 Results ... 108
5.4 Discussion ... 122
5.5 Materials and Methods ... 126
5.6 Funding ... 128
5.7 Competing interests statement ... 128
5.8 Author contributions ... 128
5.9 Supplementary Data ... 129
5.10 References ... 132

Chapter 6 Discussion ... 138

6.1 SOX3's genomic fingerprint in NPCs and SScs ... 139
6.1.1 Evolutionary conservation of SOX3 bound regulatory regions ... 139
6.1.2 Characterising putative SOX3 bound regulatory regions ... 140
6.1.3 Interactions with other proteins ... 141
6.1.4 Loss of SOX3 has a minimal effect on gene expression in two different cell types ... 144
6.1.5 Navigating the epigenetic landscape around SOX3 ... 146
6.2 Exploring SOXB1 functional redundancy ... 149
6.3 Functional analysis of SOX3 in spermatogonia ... 151
6.3.1 Non-SOX TF binding sites within SOX3 ChIP-Seq peaks ... 151
6.3.2 *Ngn3* and the spermatogenic block in Sox3 null mice ... 153
6.3.3 Epitope tagging NGN3 for mechanistic interrogation ... 155
6.3.4 Identifying the pathway leading to increased GFRα1+ cells in SOX3 null mice ... 156
6.3.5 Overcoming the spermatogenic block in *Sox3* null mice ... 157
6.3.6 SOX3 may have a similar role in SSCs and OPCs ... 158
6.3.7 Evidence of human infertility caused by *SOX3* mutations ... 159
6.4 Comparsion of SOX3 in two different tissues ... 161
6.4.1 SOX3 binds *Tex14* in NPCs and postnatal testes ... 161
6.4.2 *Ngn3:* one gene, two distinct SOX3 regulatory regions ... 162
6.4.3 SOX3 may regulate histone expression in NPCs and SPCs ... 163
6.5 Concluding remarks ... 165

References ... 166

This page intentionally left blank.

DECLARATION

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Dale McAninch

March 2020

ACKNOWLEDGEMENTS

It’s done!

First and foremost I must thank Professor Paul Thomas, my supervisor, my mentor, an inspirational leader, for giving me the opportunity to undertake this PhD. Your patience, guidance and unwavering support have been instrumental in getting me through. You can finally cross this one off your list. There aren’t enough words to convey how grateful I am for this opportunity. Thank you.

Thanks to my Co-supervisor Professor Murray Whitelaw, your fundamental passion for science has been inspirational. Going all the way back to undergraduate enrolment day where you told me to cut out all the nonsense from my timetable.

Dr James Hughes, who among many things, taught me that a good work life balance is invaluable. Your constant positive outlook, reliable advice and carefree attitude has helped me grow both personally and as a scientist.

Thank you James Telford, my husband, for putting up with me always having an answer to everything. You may not have known me when I started this journey, but you definitely helped me get across the finish line, and to that I am very thankful. Time for a celebratory glass of wine!

To all the Thomas Lab members, past and present, what a wild ride it has been. Nick and Kristie, thank you for such great friendship over the years, sound advice, and for keeping it real during the good times and the less good. Thanks Ella, Louise, Mel, and Ruby for the support, the wild nights and dinner parties. Thank you for the assistance with the mice Sandie, it wouldn’t have been so straight forward without you. Adi, thanks for always bringing a different perspective. Edi, Daniel, Shaun, Bryan, Sandie, Eva, Pike-See, Chin, and all the placement students in between, thank you, it would not have been the same without you all. To everyone who made up the (former) Department of Biochemistry, you made it an inviting workplace to belong to, nothing else quite compares.

I hope someday to pay it forward, to help others learn, grow, and most of all enjoy, just as everyone here has done for me. It was far from just research, but a journey of self-discovery. I’m far from the same honours student who started his PhD all those years ago.

// ABSTRACT

SOX3 is a transcription factor found within neural progenitor cells (NPC) of the developing and adult vertebrate central nervous system. SOX3 is also found in other tissues, most notably the spermatogonial progenitor/stem cell populations in the testes. Normal brain development in both humans and mice, and sperm production in mice, is reliant on the correct expression and dosage of SOX3. The function of SOX3 has been explored through a number of different cell and mouse model based techniques, however, the mechanisms through which SOX3 acts remain largely unknown.

This thesis explores the genome wide DNA binding profile of SOX3 in both NPCs and postnatal testes, two very different sources of SOX3 expressing cells. We identified 8064 binding sites within NPCs derived from cultured mouse embryonic stem cells, linking SOX3 to a number of different neural development pathways. Additionally, we identified 778 SOX3 binding sites within postnatal day 7 mouse testes, linking SOX3 to the control of histones and histone variants, most of which was also true for NPCs.

We utilised our *Sox3* null mouse model and a number of different marker genes of spermatogenesis to identify that SOX3 is found within the committed progenitor fraction of the undifferentiated spermatogonial pool. We identified that SOX3 is required for the transition from a GFRα1+ state to a NGN3+ committed progenitor state, and in the absence of SOX3 GFRα1+ cells accumulate and spermatogonia fail to differentiate, leading to empty testes with no mature sperm. We provide further evidence that *Ngn3* is a direct target of SOX3 in both NPCs and the testes albeit thought different regulatory regions.

We have generated two invaluable genome wide ChIP-seq datasets that will deepen our understanding of mechanisms by which SOX3 controls context-specific differentiation. Taken together, the data presented in this thesis expand our knowledge of the genomic regions bound by SOX3 and its role in neurogenesis and spermatogenesis.

ABBREVIATIONS

AME	analysis of motif enrichment
ATAC-Seq	assay for transposase-accessible chromatin sequencing
ATP	adenosine triphosphate
bHLH	basic helix-loop-helix
CCA	corpus callosum
CD	campomelic dysplasia
ChIP	chromatin immunoprecipitation
ChIP-Seq	chromatin immunoprecipitation sequencing
CMV	cytomegalovirus
CNS	central nervous system
DAPI	4′,6-diamidino-2-phenylindole
DBD	DNA Binding domain
DHC	dorsal hippocampal commissure
DNA	deoxyribonucleic acid
DPC	days post coitum
EMSA	electrophoretic mobility shift assay
ESC	embryonic stem cell
FACS	fluorescence activated cell sorting
FB	forebrain
FSH	follicle-stimulating hormone
GFP	green fluorescent protein
GH	growth hormone
GO	gene ontology
GREAT	genomic regions enrichment of annotations tool
HB	hindbrain
Hi-C	chromosome conformation capture coupled with high throughput sequencing
HMG	high mobility group box
HOMOCOCO	HOmo sapiens COmprehensive MOdel COllection
HYPO	hypothalamus
ICF	intercerebral fissure
ICM	inner cell mass
ID	intellectual disability

INF	infundibulum
KD	equilibrium dissociation constant
KO	knock out
LH	luteinising hormone
MB	Midbrain
miRNAs	micro RNA
ncRNAs	non-coding RNA
NPC	neural progenitor cell
OPC	oligodendrocyte progenitor cell
ORF	open reading frame
P#	postnatal day #
PCR	polymerase chain reaction
PNS	peripheral nervous system
Pol #	RNA polymerase # (I, II, or III)
qPCR	quantitative polymerase chain reaction
RA	retinoic acid
RNA	ribonucleic acid
RNAi	RNA interference
RP	Rathke's pouch
SC	spinal cord
SOCM	SOX consensus motif
SOX	Sry related HMG box
SSC	spermatogonial stem cell
TESCO	testis-specific enhancer of Sox9 core
TF	transcription factor
TSS	transcription start site
TTS	transcription termination site
TUNEL	terminal deoxynucleotidyl transferase dUTP nick end labelling
UTR	untranslated region
WT	wild type
XH	X-linked hypopituitarism

PUBLICATIONS AND MANUSCRIPTS

This thesis is based on the following:

PUBLICATIONS

I. **<u>Dale McAninch</u>** and Paul Q. Thomas. Identification of Highly Conserved Putative Developmental Enhancers Bound by SOX3 in Neural Progenitors Using ChIP-Seq. Plos One. 2014. doi.org/10.1371/journal.pone.0113361

MANUSCRIPTS

I. **<u>Dale McAninch</u>**, Juho-Antti Mäkelä, Hue M. La, James N. Hughes, Robin Lovell-Badge, Robin M. Hobbs, and Paul Q. Thomas. SOX3 promotes generation of committed spermatogonia in postnatal mouse testes.

Scientific Reports

*Revised manuscript under review: 20/2/2020

II. **<u>Dale McAninch</u>**, Ella P. Thomson and Paul Q. Thomas. Genome wide DNA binding profile of SOX3 in mouse testes.

Reproduction

*Manuscript under review: 13/3/2020

ADDITIONAL PUBLICATIONS FROM WORK PERFORMED DURING THIS PHD

i. Sutton E, Hughes J, White S, Sekido R, Tan J, Arboleda V, Rogers N, Knower K, Rowley L, Eyre H, Rizzoti K, **<u>McAninch D</u>**, Golcalves J, Slee J, Turbitt E, Bruno D, Bengtsson H, Harley V, Vilain E, Sinclair A, Lovell-Badge R, Thomas P. Identification of SOX3 as an XX male sex reversal gene in mice and humans. 2011. Journal of Clinical Investigation. doi.org/10.1172/JCI42580

ii. Hughes J, Piltz S, Rogers N, **McAninch D**, Rowley L, Thomas P. Mechanistic insight into the pathology of polyalanine expansion disorders revealed by a mouse model for x linked hypopituitarism. 2013. PLoS Genetics. doi.org/10.1371/journal.pgen.1003290

iii. Rogers N, **McAninch D**, Thomas P. Dbx1 is a direct target of SOX3 in the spinal cord. 2014. PLoS One. doi.org/10.1371/journal.pone.0095356

iv. Adikusuma F, Pederick D, **McAninch D**, Hughes J, Thomas P. Functional equivalence of the SOX2 and SOX3 transcription factors in the developing mouse brain and testes. 2017. Genetics. doi.org/10.1534/genetics.117.202549

v. Hughes J, Dawson R, Tea M, **McAninch D,** Piltz S, Jackson D, Stewart L, Ricos M, Dibbens L, Harvey N, Thomas P. Knockout of the epilepsy gene Depdc5 in mice causes severe embryonic dysmorphology with hyperactivity of mTORC1 signalling. 2017. Scientific Reports. doi.org/10.1038/s41598-017-12574-2

vi. Pham D, Tan C, Homan C, Kolc K, Corbett M, **McAninch D**, Fox A, Thomas P, Kumar R, Gecz J. Protocadherin 19 (PCDH19) interacts with paraspeckle protein NONO to co-regulate gene expression with estrogen receptor alpha (ERα). 2017. Human Molecular Genetics. doi.org/10.1093/hmg/ddx094

AWARDS AND SCHOLARSHIPS

2009: PhD top-up scholarship – Centre for Stem Cell Research (Robinson Research Institute)

Awarded one of two $18,000 top-up scholarships paid over 3 years.

2012: 1st prize PhD student poster winner. The University of Adelaide, School of Molecular and Biomedical Science, Postgraduate Symposium, Australia. $100

2012: Universities Australia – Discover China Scholarship

Received $2500 in support of a 3 month exchange to Shanghai Jiao Tong University.

PRESENTATIONS

17th International Congress for Developmental Biologists – Cancun 2013

Poster Presentation

McAninch, D., Rogers, N., Thomas, P., Investigating direct targets of mSOX3 in neural progenitor cells

ComBio – Adelaide 2012

Poster Presentation

McAninch, D., Rogers, N., Thomas, P., Investigating direct targets of mSOX3 in neural progenitor cells

CHAPTER 1

INTRODUCTION

Complex organisms develop from single cells though an intricate process of cell division and differentiation requiring highly specific spatial and temporal gene expression. This tightly regulated gene expression is responsible for the development of individual organs, such as the brain, from the same starting genetic material. Gene expression is regulated by a number of different mechanisms such as transcription factors (TF) promoting or repressing expression and chromatin remodelling via epigenetic modifications (e.g. DNA and histone methylation). Additionally there are mechanisms for posttranscriptional regulation such as micro RNAs (miRNA), non-coding RNAs (ncRNAs), and RNA interference (RNAi), which regulate transcripts and translation following transcription. Transcriptional initiation begins at the promoter of genes. The factors required to initiate this process can be found not only at the promoter but also at enhancer elements, which are sometimes great distances away. It is also thought that TFs rarely act alone; rather, they act in tandem with other TFs and can act as adaptors to form protein complexes at specific sites within the genome. This thesis investigates the role of the TF SOX3 in two crucial developmental processes, neurogenesis and spermatogenesis.

1.1.1 GENE TRANSCRIPTION AND TRANSCRIPTION FACTORS

Transcription occurs when RNA polymerase recognises the promoter region of a gene and initiates DNA-dependant RNA synthesis. Three classes of RNA polymerase (Pol) exist [1]; RNA polymerase I (Pol I) - transcribes ribosomal RNAs, RNA polymerase II (Pol II)– transcribes messenger RNAs (mRNA) and ncRNAs, and RNA polymerase III (Pol III)– transcribes transfer RNAs [2]. Tight control of transcription is required to prevent every gene from being expressed all the time. As such, each type of cell will express a unique combination of genes providing cell identity and function. Part of this control is provided by chromatin structure, where closed chromatin prevents Pols from binding DNA and initiating transcription [3]. Compacted “closed” chromatin, consisting of large numbers of nucleosomes (heterochromatin), is inaccessible by Pol. Remodelling the chromatin structure by modifying histones, the protein subunits of nucleosomes, can open up chromatin

making it accessible and allow transcription [4,5]. In addition to chromatin, TFs provide another layer of transcriptional regulation.

Transcription factors are proteins that bind specific sequences of DNA, elicit changes in genomic regulation ultimately regulating gene expression by guiding Pols to promoters [6-9]. They are the second largest group of proteins in humans and mice [10]. In humans there are approximately 1600 known TFs belonging to 100 different TF families [11]. 1198 have an orthologous TF in mouse [11,12]. TFs recruit other proteins through their transactivation domains [13] to regulate promoter accessibility and transcription initiation [14]. While a majority of these TFs bind only to open chromatin regions of DNA, some have the ability to bind closed DNA compacted around nucleosomes, otherwise known as pioneer factors [15], as discussed below.

TFs control gene transcription through binding at proximal promoter regions, or at a distance via enhancers. Enhancers are often found large distances away from their target promoters (1Mbp or more away [16]). However through DNA looping and supercoiling, distant enhancers are brought closer to promoters in order to function. Enhancers are not required to remain in a single orientation, nor do they have to be found on the same chromosome as the promoter that they are affecting [17]. Enhancers often consist of binding sites for a number of different TFs allowing for cooperative gene control, or activation in different tissues with different TFs [18], (Figure 1.1). Combinations of promoters and enhancers, each consisting of different TF binding motifs, provide extra control and fine tuning of gene expression. This delivers cell type or tissue specific control and provides redundancy mechanisms where the loss of a single gene doesn’t always result in the loss of gene expression.

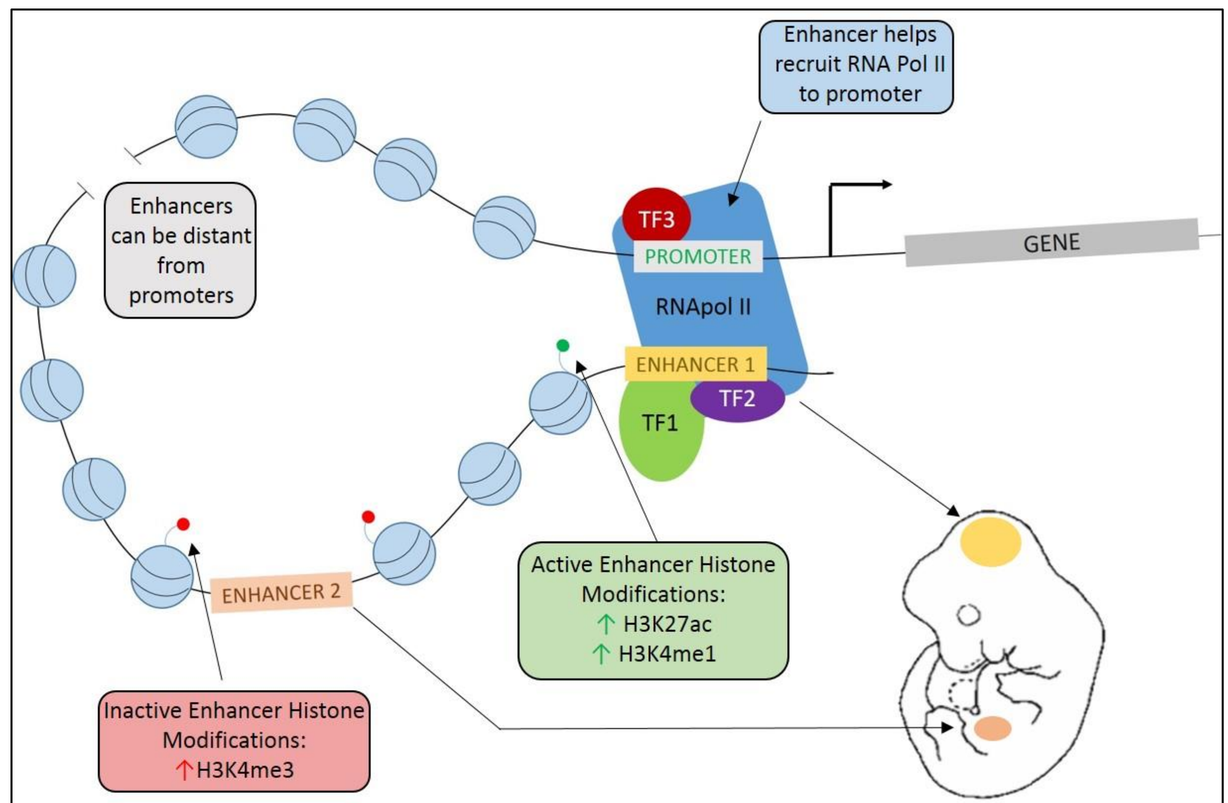


Figure 1.1. Enhancers driving gene expression. Schematic representation of a distant enhancer driving the expression of a gene through DNA looping. Enhancers have been found up to 1Mb away from their target promoters. Enhancers are bound by TFs which aid in the recruitment of Pol II to gene promoters. Histones surrounding active enhancers have increased H3K27ac and H3K4me1 modifications, while those at inactive enhancers tend to have high levels of H3K4me3. Enhancer 1 and enhancer 2 are shown to highlight expression of the same gene in different tissues. Mouse adapted from Richardson *et al.* 2014 [19].

1.1.2 PIONEER FACTORS: TRANSCRIPTION FACTORS WITH AN EXTRA TWIST

Long after the original discovery of TFs and their ability to control gene expression, some TFs were found to have non-canonical functions beyond normal transcriptional regulation. Some TFs can also act as pioneer factors by directly binding compact/condensed chromatin, where most TFs are unable to bind condensed chromatin alone [15]. The chromatin landscape can be difficult for a TF to navigate, as highly compact regions of the genome prevent TFs from binding. This mechanism prevents inappropriate expression of target genes. Classical chromatin remodelling is controlled by histone modifiers (such as histone acetyltransferases, deacetlyases, kinases and methyl transferases) or ATP-dependant remodellers (the SWI/SNF, ISWI, CHD and INO80 families) that physically move or remove nucleosomes from DNA [20,21]. These proteins create long lasting changes to chromatin structure, whereas TFs do not always require permanent access to enhancers. Pioneer factors deliver a transient method for providing access to enhancers [22].

Pioneer factors can bind condensed chromatin, and through mechanisms that are still mostly unknown, remodel the chromatin into a more open state [22]. Pioneer factors such as the FoxA family, do this without requiring ATP or ATP-dependent re-modellers [23]. They have also been linked with altering histones and their post translational modifications [24]. It is currently thought that pioneer factors bind closed chromatin but cannot initiate transcription of their target genes directly while chromatin remains in a closed state. However, once the chromatin has opened up they can recruit other cooperative factors and together they can initiate transcription [24-26] (Figure 1.2). While pioneer factors may be able to bind condensed chromatin to activate silent genes and enhancers, heterochromatin (tightly condensed and transcriptionally inert DNA) generally excludes pioneer factor binding. These regions are often shut down as a mechanism to retain cell fate [27].

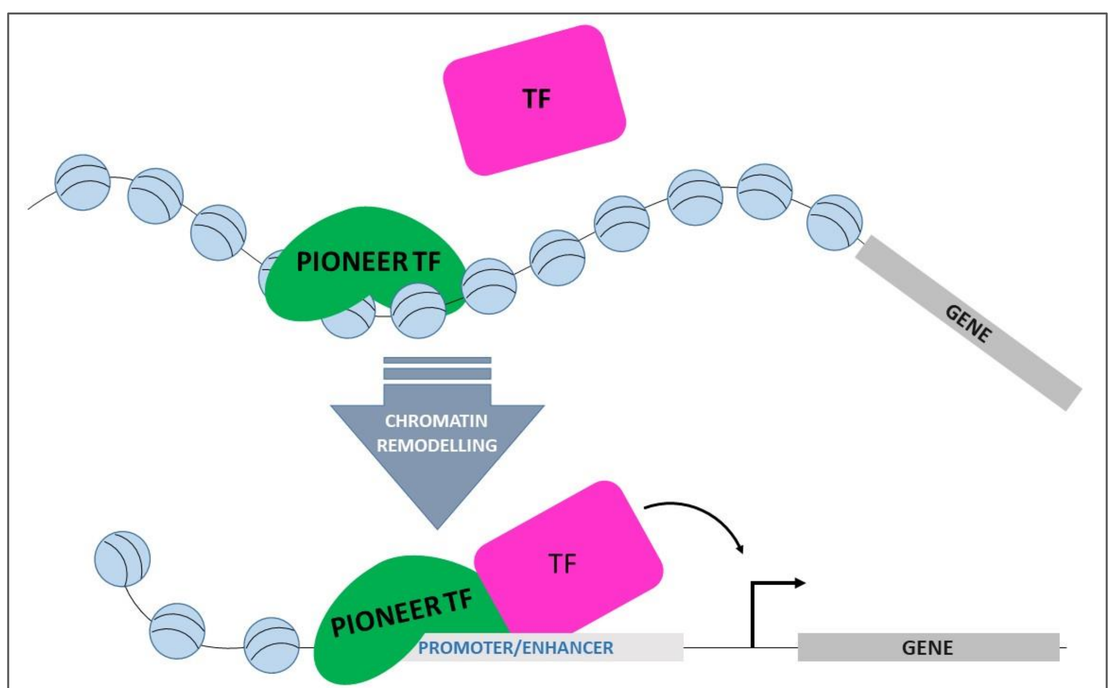


Figure 1.2. Pioneer factors vs transcription factors. Pioneer factors can recognise their DNA binding motifs on the outside of nucleosomes. Partial motif recognition is sufficient for some pioneer factors. Regular TFs cannot bind their DNA binding motifs in closed chromatin. Both TFs and some pioneer factors can bind promoters and enhancers in open chromatin to initiate transcription.

The compact nature of closed chromatin is one of the reasons why most TFs cannot bind. There are two main mechanisms through which pioneer factors can bind closed DNA. Some pioneer factors (such as SOX2, Oct3/4 or KLF4) can target closed chromatin with only part of their DBD on the surface of a nucleosome [28]. Other pioneer factors (such as FOXA) are structurally capable of binding the whole motif when DNA is bound by a nucleosome through a “winged” DBD providing flexibility within the protein structure [29-31]. Once bound to condensed chromatin, these pioneer factors recruit coregulators and/or chromatin remodellers to open up the chromatin making it accessible. FOXA1 can recruit MLL3 to nucleosomes, leading to the addition of mono H3K4 methylation at enhancer elements [32]. Aside from a few specific examples, the molecular mechanisms behind the majority of this remodelling are largely unknown. Most of these pioneer factors can also function as ‘classical’ TFs with a more direct role in influencing gene expression, through binding at open chromatin [33].

The identification of pioneer factors and their activity has added additional layers of complexity to the already intricate mechanisms of transcriptional regulation. It is now known that closed chromatin (once thought to remain silent without permanent chromatin remodelling), can be transiently modified by pioneer factors allowing transcription from within an otherwise silent domain [24]. Additionally, pioneer factors provide cells with the opportunity to prime genes for expression upon differentiation, allowing for rapid changes in gene expression.

1.1.3 CHARACTERISING TRANSCRIPTION FACTORS

DNase footprinting and electrophoretic mobility shift assays (EMSA) [34] are two techniques used to map and characterise DNA binding motifs, as well as quantitate the strength of DNA binding. Additionally, these techniques can be used for the identification of novel TFs when coupled with other methods (N-terminal peptide sequencing or one-hybrid screens). N-terminal peptide sequencing is a chemical based process developed in the 1950s, where a single amino acid is removed from the N-terminus of a peptide and identified by chromatography, before repeating for subsequent amino acids. Proteins that bind DNA as identified by DNase/EMSA can then be identified by sequencing. Yeast one-hybrid screens, employ a library of proteins that are fused to a trans-activation domain, each yeast colony expresses one fusion protein. A reporter construct contains a DNA fragment of interest preceding a reporter gene, the reporter gene is activated only when the fusion protein binds to the DNA fragment. Upon activation of the reporter gene, the plasmid can be isolated and protein identified through sequencing of the plasmid encoding the fusion protein.

These techniques facilitated identification of TFs belonging to the most common TF families, such as homeodomain, basic helix-loop-helix, nuclear hormone receptors and basic leucine zippers, back as long ago as the 1980s [35]. Since then, most TFs have been identified through sequence homology of their DBDs. It should be noted, that although novel TFs can be identified through DBD homology, some TFs identified this way fail to bind DNA, such as the CERS type homeodomain proteins [36].

As mentioned above EMSA and DNase footprinting allow for the characterisation of DNA binding motifs. These systems can be used to introduce point mutations that can inhibit or strengthen DNA binding and be used to identify critical nucleotides. In general DNA binding motifs are only 8-12bp in length [37], as these sequences will often exist in many thousands of locations throughout the genome further characterisation is required to identify where a TF binds throughout the genome. Chromatin immunoprecipitation (ChIP) is a method that has been adapted to do exactly that.

ChIP was originally developed as a technique to identify cooperative interactions between histones and DNA [38,39]. It was initially used to map the locations of post translationally modified histones (including methylated, acetylated and phosphorylated) throughout the genome. A summary of the ChIP process is shown in Figure 1.3. Since its inception, the ChIP assay has been extended to a wide range of protein-DNA interactions [40], including TFs. Initially ChIP was limited by downstream analyses, PCR on a small number of regions, or through targeted microarrays providing probes for basic promoters and known enhancer regions only. In 2007 ChIP was coupled with high throughput next generation sequencing [41-43]. This provided the ability to sequence all DNA fragments that have been precipitated and bound by the protein of interest, providing an invaluable discovery technique for genome wide interrogation.

ChIP-Seq is now a mainstream technique used in genomic and epigenomic profiling and has begun to reshape our understanding of many cellular processes [44] including transcriptional regulation, epigenetic organisation and chromatin modelling [45,46]. One such feature is the genome wide mapping of TF binding sites. Many ChIP-seq datasets exists on a wide range of TFs in a wide range of tissues and cell lines. Some of the initial ChIP-seq experiments demonstrated that the number of unique binding sites an individual TFs have can vary considerably [41-43,47]. Along with identifying where in the genome TFs bind ChIP-seq is also useful in identifying what DNA binding motifs TFs like to bind. Prior to ChIP-seq, the TF Neuron-Restrictive Silencer Factor (NRSF) was known to have a canonical binding motif that existed as two halves, ChIP-Seq helped demonstrate that the spacing between these two halves can be quite variable [48].

ChIP-Seq has been an extremely valuable tool in exploring the binding profiles of TFs. We now know that TFs are significantly more promiscuous than originally thought, bound at hundreds to tens of thousands of unique locations within the genome. However this still leave us questioning what these TFs do at all these sites.

LIBRARY NOTE:
This image has been removed
to comply with copyright.

Figure 1.3. Summary of chromatin immunoprecipitation. DNA and proteins are reversibly cross-linked, DNA is sheared into small fragments usually around 200-1000bp in length, followed by the precipitation of protein-DNA complexes using an antibody specific for the histone/modification of interest. Cross links are reversed using salt and heat, and the DNA is purified. Enriched DNA is then assessed by PCR. Figure adapted from Collas *et al.* 2010 [49].

1.2.1 SOX TRANSCRIPTION FACTOR FAMILY

One of the 100 different TF families is the Sex-determining region-Y (SRY) related high mobility group (HMG) box (SOX) family. The SOX family contains approximately 30 members in vertebrates, 20 of which are orthologous between the mouse and human genomes [50], and is the 5th largest family of transcription factors [11]. Each member exhibits regionally restricted expression during embryogenesis, some of which is maintained beyond embryonic development into the adult. Almost all tissues in the developing embryo express one or more of the SOX TFs [51,52] (Table 1.1).

The SOX family have been grouped by identity within the HMG box. Each SOX member shares a minimum of 50% amino acid identity within this domain with the founding member SRY [51]. The HMG box is a DNA binding domain that binds to specific DNA sequences via the minor groove and, depending on the size and structure of the remainder of the protein, causes DNA to bend by 30-113° [53]. The HMG box of SOX proteins recognise variations of the core consensus sequence AACAAW (W = A or T) [54-57]. Additionally, SOX genes are grouped by amino acid identity outside the HMG domain, resulting in 8 groups, SOXA – SOXH. Each group shares between 70 and 95% amino acid identity outside the HMG box (Figure 1.4) [51,53].

In addition to sharing high amino acid identity, SOX members within groups tend to share similar biological functions [56,58]. The SoxB1 group are all expressed within neural progenitor cells (NPC) and employ similar biological functions to each other [59,60]. The SoxF group are all expressed during vasculogenesis and share the role of controlling vasculature epithelial cell function [61-63]. The SOX family have originated through duplication events that are likely to have included significant regulatory elements controlling cell type specific expression, leading to similar groups being expressed in similar patterns.

Due to the large number of tissues that express one or more of these TFs, it's no surprise that mutations or altered expression of these factors have been implicated in a variety of diseases. SOX10 is expressed within neural crest cells. Heterozygous *SOX10* mutations are associated with Hirschsprung disease, a congenital

aganglionic megacolon disorder [64,65] which is thought to be linked to neural crest cell dysfunction [64].

Both SOX17 and SOX18 are normally expressed during vasculogenesis and mutations in either of these TF genes has been linked to reduced endothelial cell function and regeneration, predisposing individuals to vascular disorders [63]. SOX17 has been linked to intracranial aneurysms [66], while SOX18 has been linked to advanced coronary atherosclerotic lesions [61].

SOX9 is expressed in chondrocytes, where it regulates differentiation. Mutations in and near *SOX9* cause campomelic dysplasia (CD), a genetic disorder resulting in bowing of bones, skeletal and exoskeletal defects [67]. Additionally, 75% of XY males with CD exhibit male-to-female sex reversal [68]. CD is lethal in 95% of neonates, where death occurs as a result of respiratory insufficiency [69]. Due to varying severity of the disease, some individuals (5-10%) with CD survive into adulthood [70].

While there are only a few SOX family TFs, they play important roles in both mouse and human development. Understanding how these TFs work will provide valuable information towards mammalian development.

Table 1.1. SOX family location, function and regions of expression. The table demonstrates the diverse expression of the SOX genes in a range of embryonic and adult tissues. (Collated from Pevny *et al.*, 1997 and Lefebvre *et al.*, 2007.)

Group	Gene	Chromosome	Expression	Function
Group A	*Sry*	Y	Genital ridge, testis, brain [71]	Male sex determination and brain sexual differentiation
Group B1	*Sox1*	8	Embryonic CNS, lens	Forebrain development, chromatin architecture, neuron migration
	Sox2	3	ICM, primitive ectoderm, CNS, PNS, embryonic gut, endoderm	CNS development, neuron fate commitment, embryonic organ development
	Sox3	X	Embryonic CNS, gonads	Specify stem cell identity, gonadogenesis, CNS development
SoxB2	*Sox14*	9	Midbrain, Skeletal muscle	Neurogenesis - counteracts Sox1-3 activity to promote neuron differentiation
	Sox21	14	embryonic CNS	Neurogenesis - counteracts Sox1-3 activity to promote neuron differentiation
Group C	*Sox4*	13	Embryonic heart and spinal cord, adult pre-B and pre-T cells	-
	Sox11	12	Embryonic CNS, post-mitotic neurons	Organ development - lung, stomach, pancreas, spleen, eye and skeleton
	Sox12	2	Fetal Testis	-
Group D	*Sox5*	6	Adult testis	Skeletogenesis, neural crest development, gliogenesis
	Sox6	7	Embryonic CNS, adult testis	Cardiac conduction, skeletogenesis, gliogenesis, erythropoiesis
	Sox13	1	Kidney, Ovary, Pancreas	Lymphopoiesis
Group E	*Sox8*	17	Gliogenesis, Testis development, osteogenesis, neural crest	Cell fate commitment and maturation, CNS development, formation of neural crest and upkeep
	Sox9	11	Chondrocyte, genital ridge and adult testis, CNS, notochord	Male gonad development, cartilage condensation, apoptosis regulation
	Sox10	15	PNS, CNS	Neural crest, inner ear formation
Group F	*Sox7*	14	CNS, heart	Cardio genesis
	Sox17	1	Endoderm, testis, uterus[72]	Endoderm formation, angiogenesis, epithelial-stromal cross talk
	Sox18	2	Heart, lung, spleen, skeletal muscle, liver and brain of adult	Angiogenesis, vasculogenesis
Group G	*Sox15*	11	Pancreas	Skeletal Muscle regeneration
Group H	*Sox30*	11	Heart, brain, lung, testis, mesonephros	-

Abbreviations: CNS - central nervous system, ICM - Inner Cell Mass, PNS - Peripheral Nervous system

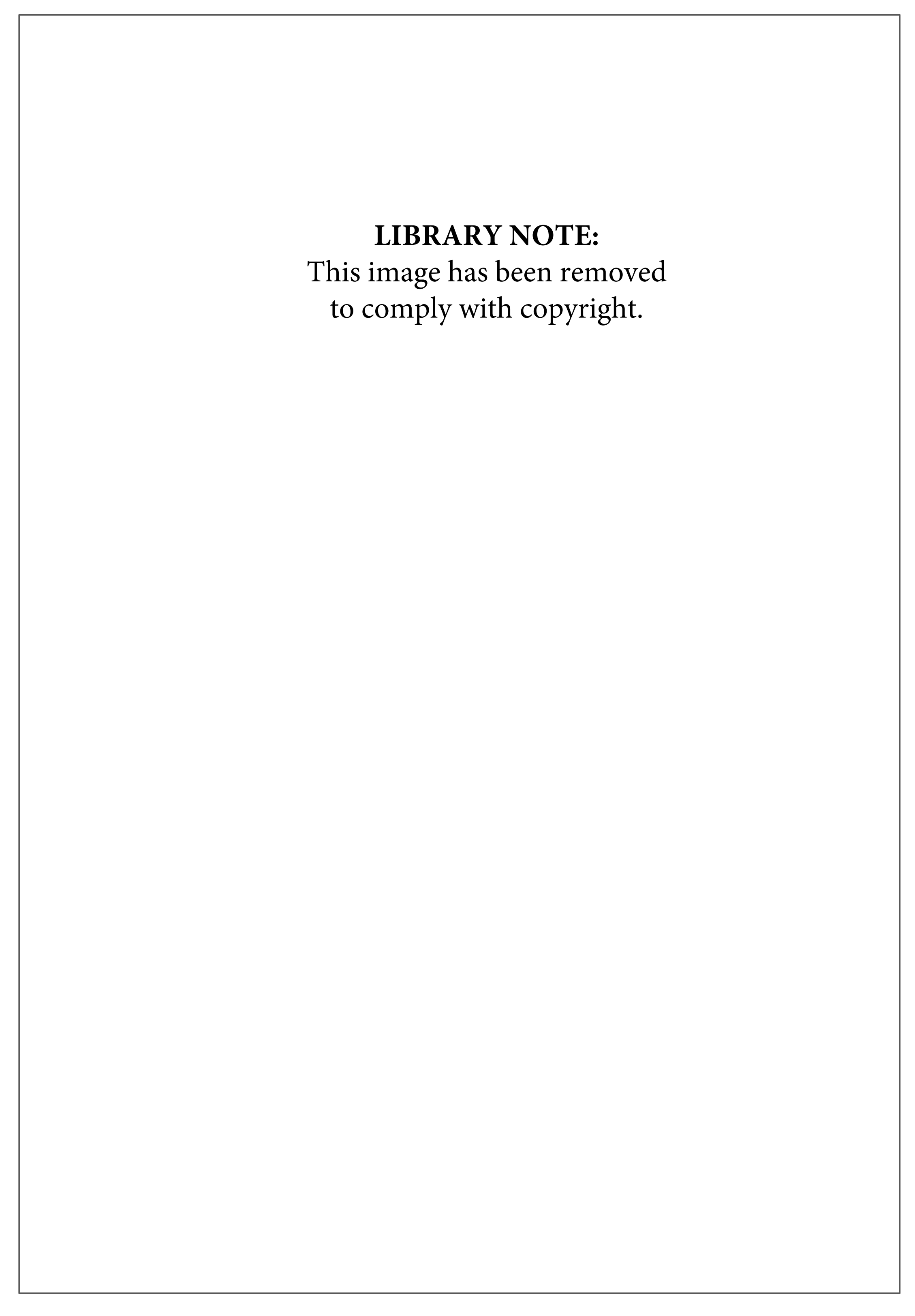


Figure 1.4. The SOX family of transcription factors. Schematic representation of the SOX family and its 8 subgroups. All SOX proteins shown here are based on the mouse variant, except for SoxB2 (Chicken). Group H is not shown. Various structural

features are shown including the HMG box (black box) and known activation and repression domains. Modified from Bowles *et al.* 2000 [56].

1.2.2 SOX3 ACTIVATES AND REPRESSES TRANSCRIPTION

Sox3 is a single exon gene located on the X chromosome and belongs to the SoxB1 subgroup (SOX1/2/3) of the SOX transcription factor family. It is thought that *Sox3* is the gene from which *Sry*, the male-determining gene and first described member of the SOX family, has evolved [73].

SOX3 has been implicated as a transcriptional activator by means of reporter assays, in which increasing the dosage of SOX3 leads to increased expression of a luciferase reporter construct containing either the SOCM or *Hesx1* proximal promoter (known to contain SOX biding regions) [59,74]. Transactivation activity has also been shown *in vitro* using Sox3-Gal4 fusion proteins and a luciferase reporter construct containing Gal4 binding sites [74].

Bylund *et al.* 2003 used chick embryos to investigate the transcriptional activity of SOXB1 proteins. Over expression of WT cSOX1/2/3 results in an inhibition of neural differentiation in the chick neural tube. To identify whether this activity was due to gene activation or repression, HMG domains of cSOX1, cSOX2 or cSOX3 were fused to either the activator domain from the herpes simplex virus VP16 protein, a potent transcriptional activator, or the repressor domain of the *Drosophila* Engrailed represser EnR [59]. Only the VP16 (activator) constructs were able to suppress the generation of differentiated neurons when expressed in the chick neural tube, and therefore replicating WT function [59].

In contrast, Zebrafish SOX3 has also been shown to have transcriptional repressor activities. In Zebrafish the SOX3 HMG:EnR fusion protein mimics the repression of genes such as *boz, sqt* and *gsc* observed with WT Sox3*;* while the SOX3 HMG:VP16 fusion protein does not [75]. The above-mentioned studies suggest that SOX3 has both transactivation and transcriptional repressive abilities, dependant on the cellular context.

The duality of SOX3's transcriptional abilities (acting as both and activator and repressor) is a common feature of TFs, increasing the complexity transcriptional regulation. Transcriptional regulation can be influenced by the affinity of a TF for a binding site [76], the number of binding sites [77,78], and by temporally controlled co-occupancy with partner factors [79-81]. It has been proposed that cooperative

binding of TFs and their binding partners can act as a buffer to variations in protein levels, providing a protection mechanism for important biological functions [82].

In addition to regular TF activities, SOX3 has been shown to have pioneer factor capabilities. SOX3 binds to a large number of genes that are not active in NPCs, but later bound and activated by SOXC TFs [55]. Histone profiling of the genes pre-bound by SOX3 were also consistent with promoter and enhancers in a poised state and ready to be transcribed [55]. While current experimental evidence points to SOX3 having pioneer factor activity, this has not yet been demonstrated directly.

1.2.3 EXPRESSION OF *SOX3* IN EMBRYONIC AND ADULT NEURAL TISSUES

Successful development of mouse brain is a highly complex process requiring precise expression of thousands of genes. Transcription factors are just one of numerous control mechanisms that define correct cell fate, function and spatial migration. Neural cell identity emerges from the ectoderm, and brain development continues throughout embryogenesis through to early postnatal life [83]. During embryogenesis there are large numbers of NPCs, undifferentiated cells that give rise to all radial glial and neuronal cells. Some NPCs can also be found in the adult brain, particularly in the sub-ventricular and –granular zones, providing a source of cells for neurogenesis in the adult brain [84-88]. In addition to these NPCs, the mouse brain comprises many different terminally differentiated cell types, including astrocytes, oligodendrocytes and neurons, that all develop at different times throughout embryogenesis and early postnatal life.

In mice, *Sox3* expression has been detected from embryonic day (E) 7.5 in the neural plate [74,89,90] from which the entire central nervous system (CNS) is derived (Figure 1.5A). At E10.5 *Sox3* is highly expressed within the developing brain, specifically within the telencephalon, that forms the cerebral cortex, within the diencephalon that forms the presumptive hypothalamus, thalamus and optic nerve (Figure 1.5B) [91]. At E11.5 *Sox3* is expressed within the infundibular recess, which together with Rathke's pouch, develops into the pituitary [92] (Figure 1.5C). By E13.5 *Sox3* expression is restricted to the sub-ventricular cell layer and is generally down regulated in differentiated neurons [93]. At E14.5, *Sox3* is still expressed within the sub-ventricular progenitor cell layer lining the lateral ventricles (Figure 1.5D). Together these data demonstrate *Sox3* expression is restricted to NPCs.

LIBRARY NOTE:
This image has been removed
to comply with copyright.

Figure 1.5. Expression of *Sox3* throughout embryogenesis. *In situ* hybridisation of Sox3 in the developing embryo from 7.5dpc to 14.5dpc. (A) Expression at 7.5dpc indicating *Sox3* is found within the neural plate. (B) Whole mount *in situ* shows *Sox3* is expressed right throughout the developing CNS, including the forebrain (FB), midbrain (MB), hindbrain (HB), and spinal cord (SC). (C) Sagittal section of an 11.5 mouse embryo showing *Sox3* expression in the infundibulum (INF), and the presumptive hypothalamus (HYPO), expression is not detected in Rathke's pouch (RP). (D) Transverse section of a 14.5dpc mouse embryo, *Sox3* is restricted to the sub-ventricular layer of the lateral ventricles (Thomas, P. Unpublished). Figure adapted from Solomon *et al.* 2007 and Brunelli *et al.* 2003 [90,91].

In addition to the well-characterised embryonic expression of *Sox3,* two recent publications have characterised the expression of SOX3 in the adult mouse brain. SOX3 expression can be detected in the neurogenic zones of the hippocampus and the sub-ventricular zone of the lateral ventricles within the adult brain [93]. Furthermore, SOX3 protein is detected in the adult cerebellum (as well as the developing embryonic cerebellum), and is restricted within the cerebellar glial cell system [94].

These data provide a comprehensive expression profile, at both the RNA and protein level, pertaining to the timing and location of SOX3's expression. *Sox3* is found in numerous regions of the developing brain, indicating that is likely to play an important role in the successful formation of the brain. The function of SOX3 has been assessed through a number of *in vitr*o and *in vivo* model systems, discussed below.

1.2.4 THE FUNCTION OF SOX3 IN NEUROGENESIS

A number of functional studies have been performed in different tissues and model organisms (including mice, chicken, and xenopus) in order to identify the function of SOX3. cSOX3, along with the other SOXB1 genes cSOX1 and cSOX2, was shown to regulate neurogenesis in the chick neural tube by maintaining NPCs in an undifferentiated state [59]. Overexpression of any of the cSoxB1 genes in the neural tube inhibited neurogenesis, as seen by the absence of neuron markers, p27, NeuN and Tuj1. Conversely, the repression of SoxB1 genes prompted premature neuronal cells differentiation [59]. Whilst it has been demonstrated that SOX3 has both activating and repressive capabilities (discussed above), only the activating function appear to be required in the chicken neural tube.

Another study demonstrated a functional relationship between Notch and SoxB1 genes. Neuronal differentiation was blocked through the overexpression of either Notch or SOX3 within the chick spinal cord, however Notch's ability to inhibit differentiation is dependent on the simultaneous expression of functional SOX3 [95]. These data help solidify cSOX3's role in the maintenance of neural progenitors and suggest that cSOX3/B1 can act alone, or in conjunction with Notch.

In mice, *Sox2* and *Sox3* (but not *Sox1*) are expressed in oligodendroglial cells, where they act as differentiation factors [96]. In mouse embryos lacking either *Sox2* or *Sox3* individually or the combined loss of both *Sox2* and *Sox3*, there is a reduction in the number of cells expressing myelin genes indicating a lack of terminal differentiation in oligodendrocytes [96]. It is also proposed that SOX2/3 act as pioneer factors in oligodenroglial progenitor cells (OPCs), where they pre-bind genes in progenitor cells that are activated later by other SOX families.

These studies have highlighted two different functions of SOX3, one maintaining cells in a progenitor state, and the second acting as a pro-differentiation factor. There are a few potential mechanisms that might explain how one TF can have multiple functions. SOX3 may be acting as a pioneer factor in one cell type and not the other, altering the cellular outcome. Alternatively, the simplest explanation is that different cell types will express different cofactors and it is these different combination of cofactors which helps determine which role SOX3 plays.

1.2.5 BINDING PARTNERS AND COFACTORS OF SOX PROTEINS

As mentioned above, many TFs do not act alone, and often require specific partner factors to perform their desired function. TF (and many proteins in general) can form dimers, either homo- or hetero-typic. Some require dimerisation to occur before they can bind DNA, while others require the presence of DNA and their DNA binding motifs in order to dimerise [97]. Partner proteins can strengthen DNA binding specificity by increasing the number of nucleotides recognised by the TF complex. Changes in partner protein concentrations can influence TF activity independent of the primary TF abundance [98]. They also act to reduce off target binding and transcriptional noise, resulting in more resourceful transcription [99,100].

The SOX TFs all recognise and bind similar DNA motifs. Given this similarity in binding motif preferences, it has been proposed that SOX genes acquire target specificity through cofactors [53,101]. Studies interrogating the binding partners of SOX2, SOX10 and SOX18 have shown that the HMG domain of SOX proteins can function as both a DNA binding domain and a protein binding interface [62,102-104]. SOX2 and PAX6 have been shown to regulate the DC5 enhancer element of the δ-crystallin gene cooperatively [105]. ChIP analyses have demonstrated that these TFs can only bind the regulatory element when both factors are present and thus cofactors are required for their function [105].

In vitro reporter assays have been used to demonstrate SOX2 binds cooperatively with POU family TFs to activate the Nes30 enhancer, an enhancer found in intron 2 of *Nestin [106]*. TFs shown to interact cooperatively with SOX2 include POU class II, III and IV members including BRN1, BRN2, BRN3, OCT2, OCT3 and OCT6. Additionally, SOX1 and SOX3 were also shown to cooperatively interact with BRN2 drive expression from the Nes30 enhancer [106].

This information provides insight into the mechanism by which SOX proteins obtain target specificity. Currently the only information regarding SOX3's cofactors exits from *in vitro* assays, there is no information identifying or confirming *in vivo* cofactors of SOX3.

1.2.6 FUNCTION OF SOX3 IN MICE

As discussed above, we know *Sox3* is found within the NPCs of the developing mouse brain and two adult neurogenic zones in the adult mouse brain. A robust method for the identification of cell function is through deleting the gene in mice. To date, the function of SOX3 has been investigated through two independent *Sox3* null (*or* Knock Out) mouse models.

One KO mouse line was generated using homologous recombination, where the *Sox3* open reading frame (ORF) was replaced with the green fluorescent protein (GFP) ORF [92]. These *Sox3* null mice exhibit hypothalamic hypoplasia as well as variable dwarfism [92]. *Sox3* null mouse embryos (from E11.5 through E14.5) show a bifurcation and expansion of the dorsal side of Rathke's pouch (Figure 1.6A & 1.6B) [92], while 3 week old mice have some severe CNS malformations including dysgenesis of the corpus callosum and failure of the dorsal hippocampus to cross the midline (Figure 1.6C & 1.6D). Taken together this indicates the importance of *Sox3* expression in the mouse CNS for both embryonic and postnatal brain development. Although *Sox3* expression is detected in the embryonic and adult cerebellum, *Sox3* null mice do no exhibit any cerebellar phenotypes [94]. In addition to these neural defects, *Sox3* null mice also exhibit a block in spermatogenesis (discussed in further detail below).

LIBRARY NOTE:
This image has been removed
to comply with copyright.

Figure 1.6. CNS defects in *Sox3* null mice. (A, B) Sagittal sections 11.5dpc brains from WT (A) and *Sox3*-null embryos (B). Rathke's pouch is dorsally expanded and bifurcated and the evagination of the infundibulum was less pronounced (arrow) in the *Sox3* null embryos. Transverse section 3-week-old brains from WT (A) and *Sox3*-null mice (B). Dysgenesis of the corpus callosum is seen in the *Sox3* null brains (arrow). Hyp, presumptive hypothalamus; RP, Rathke's pouch; Inf, infundibulum; ICF, intercerebral fissure; Cca, corpus callosum; Dhc dorsal hippocampal commissure. Adapted from Rizzoti *et al.* 2004 [92].

A second study has also generated an independent *Sox3* null mouse line to study the effect of the loss of *Sox3*. LoxP sites were introduced to flank *Sox3* by homologous recombination, followed by the excision of *Sox3* by breeding the chimeras onto a Cre recombinase expressing line; no reporter gene was introduced [107]. These mice exhibited overgrown and misaligned teeth, variable dwarfism, and infertility [107]. No change in growth hormone levels was observed in these *Sox3* null mice. Additionally it was reported that brain development was grossly normal, however this data is not presented [107,108]. Male *Sox3* null mice showed postnatal hypogonadism and were identified as being sterile [108]. Further analysis into the effect of the loss of *Sox3* on gonad function identified a role of *Sox3* in the progression of spermatogenesis [108] (the role of *Sox3* in spermatogenesis will be expanded on below). A germ cell specific *Sox3* null mouse line confirmed the reduction in spermatogenesis was caused directly by the loss of *Sox3* in the testis and was not secondary to a pituitary hormone deficiency [109]. Unlike the above-mentioned study, no pituitary defects were identified in this line of *Sox3* null mice [108].

The difference in phenotype observed by the two different groups is most likely caused by the differences in genetic background [110]. It has been proposed that a mixed genetic background provides the quickest way to identify most strain dependent phenotypes [110]. This phenomenon is quite evident when comparing the two different *Sox3* null mouse models. The *Sox3* null model generated by Rizzoti *et al.* was produced and maintained on a mixed genetic background; 129/SvEv, C57BL/6J, and MF1 mouse strains. The *Sox3* null line produced by Weiss *et al.* was produced on a 129/SvJ (and an unreported background from the transgenic CMV-Cre mouse) background followed by backcrossing the null allele onto the C57BL/6J strain [107]. Neural phenotypes observed within the mixed background *Sox3* null mice, were not present on the 129/SvJ or C57BL/6J backgrounds. Although different phenotypes are observed between the two *Sox3* KO lines, the observed phenotypes are always within *Sox3* expressing areas.

Ultimately the CNS defects observed in *Sox3* null mice appear to be quite minor, given its extensive expression in the developing CNS. This suggests that functional

redundancy, a functional backup mechanism through which the loss of a single gene is mitigated by another closely related gene, is occurring in *Sox3* KO mice.

1.2.7 OVERLAPPING EXPRESSION AND FUNCTIONAL REDUNDANCY OF SOXB1 PROTEINS

For functional redundancy to occur the expression of the redundant factors needs to either overlap, or be upregulated in response to the loss of the gene of interest. In the case of the SoxB1 subgroup, they are expressed in largely overlapping regions during early development. *Sox2* is the first member of the SoxB1 family to be expressed during development and is detectable in the blastocyst inner cell mass and primitive embryonic endoderm, the pluripotent cell population within the embryo [111,112]. Like *Sox3*, by E7.5 *Sox2* is expressed within the anterior neuroectoderm, but not within the posterior ectoderm [112]. By E9.5 *Sox2* is expressed in the brain and neural tube as well as in some regions of the branchial arches and gut endoderm [113]. As cells begin to differentiate *Sox2* is down regulated and becomes restricted to the ventricular zone [112].

Sox1 is initially expressed at E8 within the neural plate [60]. By E9.5 *Sox1* is detected throughout the entire length of neural tube [114]. At E12.5 *Sox1* is expressed in the ventricular and sub ventricular zones as well as the lens [114].

As outlined above, the expression pattern of *Sox1* and *Sox2* show extensive overlap during early development, particularly between E7.5-E10.5. This is also the case for the expression of *Sox3*. Although we see very similar patterns of expression between the three SoxB1 group members, this overlap is not complete as there are distinct regions of expression of only 1 or 2 of the genes (Figure 1.7). For example, *Sox1* is expressed uniquely in the lens from E12.5 [115].

The SoxB1 proteins have been shown to share functional similarities including the observation that over expression of cSOX1, cSOX2 or cSOX3 in chick neural tube leads to the same effect of maintaining cells in a progenitor state (described above) [59]. The SoxB1 proteins have also been shown to bind the same DNA sequence, the SOX consensus motif (SOCM), through gel shift assays [89]. SOX1 and SOX3 where shown to bind the SOCM containing probe with the same affinity while SOX2 was shown to bind with a greater affinity [89].

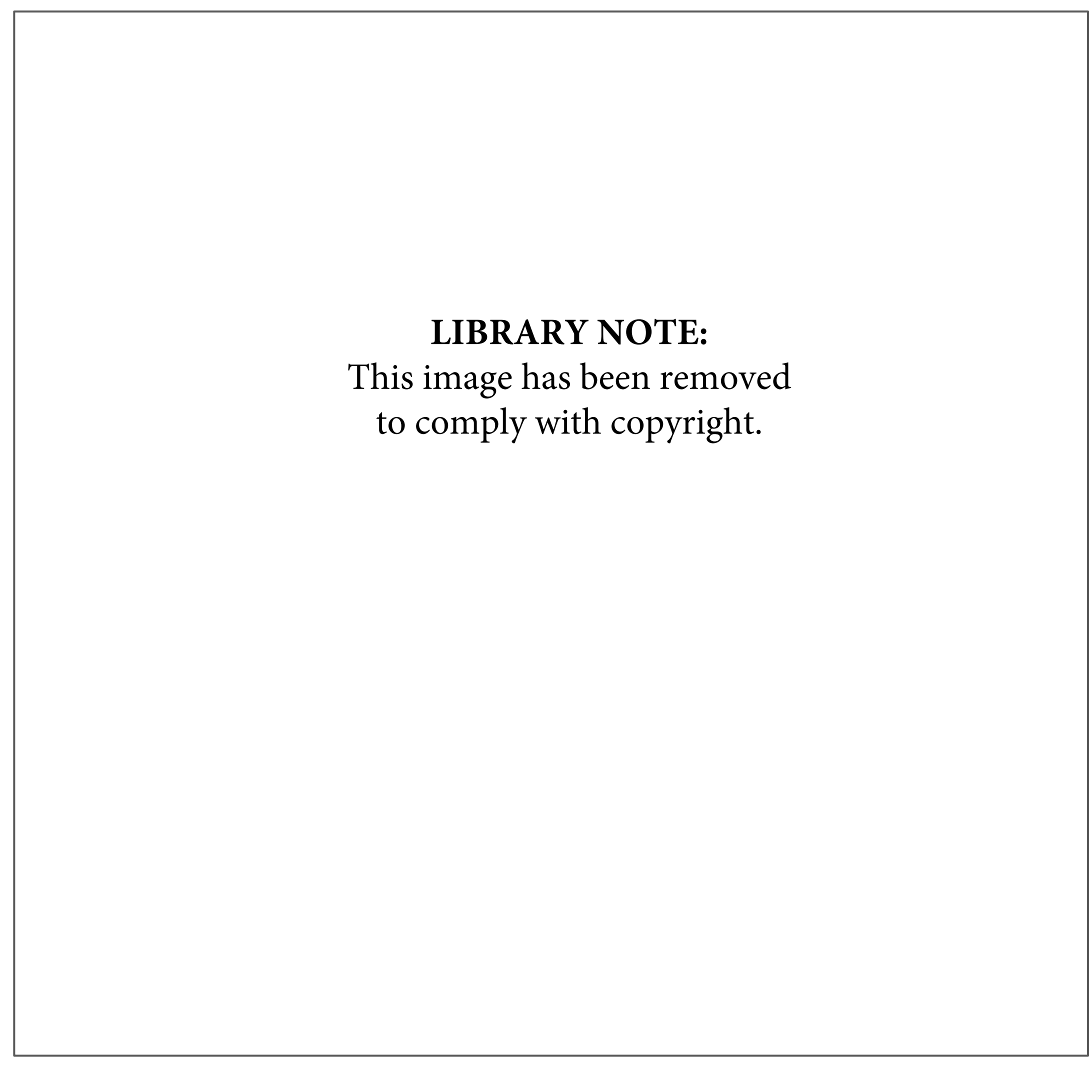


Figure 1.7. Overlapping SoxB1 expression during embryogenesis. Whole mount *in situ* hybridisation of mouse embryos ranging from 8-9.5dpc (1-14 somites). Embryos from 1-6 somites have been flattened. Extensive overlap of SoxB1 expression can be observed at all stages, with some small unique regions of expression for each TF. Key: Arrow heads, primitive streak; small arrows, ectoderm; large arrow, foregut, r, rhombomere; d, dorsal; v, ventral; h, heart; fg, foregut; se, surface ectoderm; ov, optic vesicle; np, nasal placode; op, otic placode. Figure adapted from Wood and Episkopou, 1999 [60].

Due to the extensive overlap in expression and functional similarities between the SoxB1 proteins it is believed that functional redundancy is likely to occur between these proteins [89,92,115,116]. *Sox3* is expressed widely throughout the developing CNS, however defects caused by the loss of SOX3 (as described above) might be ameliorated, due to the potential compensatory roles of SOX1 and SOX2. Similarly, CNS defects caused in relation to the loss of either SOX1 or the conditional loss of SOX2 in the CNS is also relatively mild, supporting this model [92,117]. The ablation of *Sox2* in mice leads to early embryonic lethality [113], however the reduction of *Sox2* expression through the deletion of a neuronal enhancer found within the *Sox2* locus leads to grossly normal development [117]. The embryonic lethality caused by the loss of *Sox2* occurs during the stages of development where no other SoxB1 members are expressed, which further supports the model of functional redundancy.

Embryonic stem cells (ESC) are by definition pluripotent, a unique feature that allows them to become any cell in the organism. *Sox2* is one of the genes required to maintain this pluripotency, as loss of *Sox2* expression in ESCs results in the loss of pluripotency [118]. Expressing either *Sox1* or *Sox3* in ESCs that lack *Sox2* results in the rescue of ESC pluripotency [119]. Interestingly, this functional redundancy was also observed with *Drosophila* SoxN, both in ESCs and in chimeric mice [119]. These experiments help strengthen the argument for SoxB1 functional redundancy, providing more *in vitro* and *in vivo* evidence. However cells alone and chimeric mice still leave many questions unanswered, without experimental evidence it is uncertain whether functional redundancy occurs in all tissues with overlapping SoxB1 expression. The chimeric mice generated by Niwa *et al.* 2016 do not completely answer the question of functional redundancy, there are WT cells present in these chimeras that may mask the true extent of SoxN's ability to replace SOX2. Breeding these chimeras to create a SoxN knock-in mouse model would help strengthen this argument.

Recently an ORF swap experiment was performed to answer these questions. *Sox3* ORF was replaced with *Sox2* ORF in mice, utilizing CRISPR/Cas9 [120]. These mice expressed SOX2 in place of SOX3 and concluded that the expression of SOX2 in place of SOX3 can almost completely rescue the *Sox3* null phenotype observed in

both the pituitary and testis [120]. Expression of genes in the spermatogenesis pathway were restored enough to rescue the phenotype, however not all genes returned to WT levels (eg. *Ngn3* expression was still lower than WT). One possible explanation is that SOX2 may not have been present at endogenous SOX3 levels in WT mice. This decreased protein amount may explain the reduction in gene expression rescue, and as described earlier mice can be sensitive to incorrect SoxB1 dosage. Alternatively, SOX2 may not be able to regulate SOX3 specific targets as well as it can regulate SOX2/SOXB1 targets. As regions of *Sox3* specific expression have evolved, it is possible that SOX3 has developed unique ways to regulate *Sox3* specific target gene that SOX2 cannot match.

These gene swap experiments help provide further *in vitro* and *in vivo* evidence towards the redundant roles of the SoxB1 family and highlight the importance of maintaining tight control over expression of the SoxB1 genes.

1.2.8 *SOX3* AND HUMAN DISORDERS

X-linked hypopituitarism (XH) is a male specific disorder characterised by growth hormone (GH) deficiency as well as variable deficiencies in other pituitary hormones [121]. Individuals with this disorder are short in stature due to GH deficiency (dwarfism results if not treated with GH replacement therapy), and have variable degrees of intellectual disability (ID) [90]. XH patients also have structural abnormalities of the hypothalamic region as well as an ectopic posterior pituitary [74,122]. To date there have been a number of individual families with unique duplication events of varying size that include the *SOX3* locus [122]. The most recent investigation into XH has identified the smallest duplication event, 323.8kb in length. Critically, the only brain gene included in these duplication events is SOX3 [74,121,122].

Duplication events including the *SOX3* locus have also been linked to XX female-to-male sex reversal. There are currently 5 reported individuals exhibiting sex reversal and varying degrees of pituitary hormone deficiencies [123-127]. Female-to-male sex reversal was also observed in transgenic mice with extra copies of *Sox3* leading to ectopic expression in the developing bi-potential gonads [125]. SOX3 was proposed to act through mechanisms similar to that of SRY, resulting in the initiation of the male gonad pathway in the absence of a Y chromosome [125].

As well as duplication events, mutations that cause the expansion of SOX3 polyalanine tracts have also been shown to cause XH [74,128,129]. Polyalanine tracts have been associated with diseases in numerous different proteins [130]. *In vitro* experiments have shown the expansions cause the formation of cytoplasmic protein aggregates, as well as reducing transactivation capabilities [74,131]. However the expression of SOX3 polyalanine expansion mutants *in vivo* point towards it acting as a partial loss of function allele due to misfolding and clearance by the proteasome [130].

In contrast to the genomic duplications and polyalanine expansions of *SOX3* that lead to XH, there have been several independent deletion events that lead to varying levels of ID. Three independent patients with deletions encompassing *SOX3* have presented with varied levels of ID and haemophilia B, one with a 6Mb deletion

[132,133], a second patient with a 2Mb deletion [134] and a third with a 2.3Mb deletion [135]. In addition there was a patient reported with a 2.1Mb deletion encompassing *SOX3* and presenting with ID and mild facial abnormities [136].

Together these data highlight the importance for tightly controlled expression of *SOX3*, where both over and under dosage of *SOX3* can lead to CNS defects.

1.3.1 SPERMATOGENESIS IN MICE

Spermatogenesis is the process of sperm formation in the male testis, and is a complex process requiring precise timing and control of mitosis, meiosis and cell differentiation. It is through this complex process that SSCs, which comprise only 0.03% of total germ cells in mice [137], amplify, differentiate and reduce to a haploid genome, resulting in millions of mature sperm each day.

During embryogenesis, primordial germ cells migrate to the genital ridge, followed by differentiation into gonocytes [138]. In mice, this gonocyte population differentiates into spermatogonial stem cells by postnatal day 6 [138]. This process is rapid in mice when compared to humans due to the difference in pre-pubertal length, an average of 42 days in mice and 11.5 years in humans [139].

During the first 6 postnatal days, gonocytes differentiate into a heterogeneous population of undifferentiated spermatogonia. This pool of undifferentiated spermatogonia consists of type A-single (A_s) spermatogonia or cysts of 2 or more spermatogonia termed A-paired (A_{pr}) or A-aligned (A_{al}) respectively, and are located on the basal membrane of the seminiferous tubule [140-143]. A_s spermatogonia undergo either symmetrical cell division to form two identical A_s cells, maintaining the spermatogonia pool, or undergo incomplete cytokinesis during cell division to form A_{pr} spermatogonia [142,144]. A_{pr} continue this incomplete cell division process forming chains of A_{al} spermatogonia which can be 4, 8, 16 or sometimes 32 cells in length. Longer chains of A_{al} spermatogonia are believed to be committed to differentiation into A1 spermatogonia. Type A1 spermatogonia differentiate through a series of different differentiating spermatogonia including A2, A3, A4, intermediate (I), and Type B [143]. These cells differentiate in a synchronised manner and are located within the seminiferous tubules. Type B spermatogonia undergo meiosis to form secondary spermatocytes, followed by a second round of meiosis giving rise to haploid spermatids [143]. These spermatids finally undergo spermiogenesis, which involves a number of significant morphological changes to become spermatozoa, before they are ready to be released as mature sperm [145] (Figure 1.8).

LIBRARY NOTE:
This image has been removed
to comply with copyright.

Figure 1.8. Overview of mouse spermatogenesis [146].

An overview of spermatogenesis in mice, from undifferentiated stem spermatogonia through to mature sperm, including genes that have been shown to mark the different cell stages. Modified from Fayomi *et al.,* 2018 [146].

Type A and B spermatogonia were originally identified and distinguished by the amount of heterochromatin present in the nuclei of each type, type A have little to no heterochromatin, while type B nuclei consist of large quantities of heterochromatin. Heterochromatin is heavily condensed chromatin that is found in silenced genomic regions, nuclei with large amounts of heterochromatin are usually more differentiated, while those without are usually more undifferentiated. The entire spermatogenesis pathway revolves around intricate and rapid changes to chromatin state [147]. These changes in chromatin state are required to maintain chromosomal integrity throughout the large number of mitotic, meiotic and differentiation steps that occur. Errors during this process can lead to apoptosis, cell cycle arrest, and ultimately infertility [148].

Currently there are two different models regarding amplification of the spermatogonial stem cell population, the "A_s model" and the "fragmentation model". The A_s model relies upon A_s spermatogonia undergoing either symmetrical cell division generating two identical A_s daughter cells, or undergoing incomplete cell division generating A_{pr} spermatogonia. The outcome of cell division is dependent on the location and migration of the dividing A_s cell within the seminiferous tubule and their mitotic activity. Overall A_s cells have the same chance of self-renewal or differentiation [143]. This model implies that A_s cells are a homogenous population which contradicts studies that indicate A_s cells are a heterogeneous population [149-151].

The "fragmentation model" has been proposed following long term lineage tracing of GFRα1-GFP and NGN3-GFP expressing spermatogonia and accounts for a heterogeneous A_s population [152]. Live cell imaging identified that a majority of A_s cell divisions were incomplete giving rise to A_{pr} cells, and that A_{al-4} cells fragment to repopulate the A_s population [153]. This model proposes that all undifferentiated spermatogonia, not just A_s cells, have the potential to belong to the stem cell pool [152].

Undifferentiated spermatogonia are characterised by the expression of PLZF and E-CAD and do not express c-KIT while, conversely, differentiating spermatogonia are characterised by the expression of c-KIT and not PLZF and E-CAD [154-157].

GFRα1 and NGN3 are not only found heterogeneously within the undifferentiated spermatogonia pool, but they are also reciprocally expressed. GFRα1+ cells are usually A_s and A_{pr} (with a small number of A_{al} being GFRα+), while NGN3+ cells are usually A_{al} spermatogonia (with a small number of A_s and A_{pr} being NGN3+) [152,158-160]. Most NGN3+ spermatogonia will differentiate into c-Kit+ spermatogonia and subsequently continue through the differentiation pathway to become spermatozoa [159-161]. These two sub-populations of undifferentiated spermatogonia, either GFRα1+ or NGN3+, also exhibit different differentiation capabilities in response to retinoic acid stimulation due to the expression of RARγ in NGN3+ cells [162]. The exposure of NGN3+/RARγ+ cells to retinoic acid promotes differentiation. Seminiferous tubules have an abundant amount of retinoic acid present at stage VIII [163] promoting the differentiation of NGN3+ cells but not GFRα+ cells.

1.3.2 EXPRESSION AND FUNCTION OF SOX3 IN MOUSE TESTES

As mentioned earlier, *Sox3* is expressed within the adult testis in mice, specifically within undifferentiated spermatogonia [108]. *Sox3* deletion in mice leads to a block in spermatogenesis, however the mechanism behind this block remains unclear [92,107,108]. The severity of the spermatogenic block is dependent on the genetic background of the mouse strain, whereby C57Bl/6J mice exhibit the strongest phenotype. Mice from the *Sox3/Sox2* ORF swap model by Adikusuma *et al.* 2017, demonstrate that the expression of *Sox2* in place of *Sox3* is sufficient to overcome the morphological and phenotypic alteration in the *Sox3* null model. However, it should be noted that expression levels of key players in the spermatogenesis pathway, in particular *Ngn3*, did not return to normal wild type levels. This blockage highlights the importance of *Sox3* expression during spermatogenesis, gaining a clearer understanding of the function of *Sox3* in these cells is a critical step towards deciphering the delicate gene expression code that controls mouse spermatogenesis.

1.4 PROJECT RATIONALE AND AIMS

As described above, *Sox3* is a dosage sensitive regulator of CNS development, where altered expression of *Sox3* leads to CNS defects observed in both humans and mice. Additionally, SOX3 plays a key role in spermatogenesis, as demonstrated by the spermatogenic block observed in *Sox3* null mice. To date there have been a number of studies looking at how SOX3 functions, both endogenously and ectopically within cells and cell lines, as well as within two independently generated *Sox3* null mouse models. While these functions are of key interest, the mechanisms surrounding SOX3's transcriptional activity still remain unclear, in particular the transcriptional targets of SOX3. We set out to develop two highly comparable SOX3 ChIP-seq datasets from two different SOX3 expressing cell types using the same antibody. Surprisingly, the mechanism/s through which SOX3 controls the progression of spermatogenesis remains largely unknown. These aspects were therefore the focus of the research in this thesis, with the specific aims of;

1. Characterising the global binding profile SOX3 and its impact on transcription in mouse neural progenitor cells
2. Interrogate the role of SOX3 in spermatogenesis by:
    a) Using the *Sox3* null mouse model to identify the mechanisms through which SOX3 controls spermatogenesis and clarify which cell type/s express SOX3
    b) Characterising the global binding profile of SOX3 and its impact on transcription in mouse testes

CHAPTER 2

SUMMARY OF RESULTS

2.1 CHARACTERISING THE GLOBAL TRANSCRIPTIONAL MECHANISMS OF SOX3 IN NEURAL PROGENITOR CELLS

SOX3 is a transcription factor expressed within most NPCs of the developing CNS, and is required for normal brain development in both mice and humans. The molecular mechanisms underlining the function of SOX3 in NPCs is addressed in Publication I, entitled “Identification of highly conserved putative developmental enhancers bound by SOX3 in neural progenitors using ChIP-Seq”.

For the study in Publication I, we utilized the highly specific SOX3 antibody, validated in our lab [93], to generate a ChIP-Seq data set from ESC derived NPCs. Publication I identified 8064 binding sites within the mouse genome, a majority of which are located at either intronic or intergenic locations. We demonstrated that SOX3 binds extensively to evolutionarily conserved sequences located in or near neurodevelopmental genes. We identified numerous potential large-scale interactions where SOX3 links the promoters and enhancers of distant genes on a genome wide scale. Additionally, we cross examined our dataset with a SOX2 and an additional SOX3 ChIP-seq dataset from NPCs to explore potential mechanisms behind SoxB1 functional redundancy.

2.2 EXPLORING THE MOLECULAR MECHANISMS OF SOX3 IN MOUSE TESTIS

The main research focus surrounding the function of SOX3 relates to its expression within NPC populations, both *in vivo* and *in vitro. Sox3* expression can also be found within other tissues such as the heart or testis. Previous publications have shown that *Sox3* is expressed within the spermatogonial progenitor cells located within the testis. Loss of *Sox3* from spermatogonial progenitor cells leads to a variable spermatogenic block dependent of the genetic background of the mice, leading to sub fertile or infertile mice. Currently the SOX3's role in spermatogonia and the mechanisms behind this spermatogenic block are poorly understood. We set out to explore this phenotype with two different approaches. The first being a closer look at SOX3 in the mouse testes, investigating which sub-types of Type A spermatogonia express SOX3, and examining the SOX3 *null* mice model to identify when and why the block occurs. Our second approach was using ChIP-seq to map where SOX3 is found in the genome of mouse testes.

2.2.1 THE ROLE OF SOX3 IN TYPE A SPERMATOGONIA

In Manuscript I, "SOX3 promotes generation of committed spermatogonia in postnatal mouse testes", we investigate the molecular mechanisms surrounding the spermatogenic block as caused by a loss of *Sox3* expression in 6th generation backcrossed 129Sv mice. It was found that SOX3 expression was restricted to the late committed subset of type A spermatogonia. SOX3 was found to be highly co-expressed with PLZF, yet very little overlap is observed with GFRα1+ spermatogonia. This places *Sox3* expression within a very specific subset of Type A spermatogonia. Furthermore, we noted an increase in GFRα1+ spermatogonia in *Sox3* null mouse testis at P7, indicating that without SOX3, Type A spermatogonia fail to differentiate and remain in an uncommitted stem cell state.

2.2.1 CHARACTERISING THE GLOBAL TRANSCRIPTIONAL MECHANISMS OF SOX3 IN MOUSE TESTES

In Manuscript II, "DNA binding profile of SOX3 in mouse testes", we sought to examine the transcriptional role of SOX3 in a previously unexplored tissue. Whilst spermatogonial progenitor cells are unique in that they express only *Sox3* and not *Sox1* or *Sox2*, it is unknown whether SOX3 performs a different function in SPCs than it does in NPCs (Manuscript I). On one hand, the lack of functional redundancy should allow for easier identification of the role of SOX3 in SPCs, however SPCs may not provide significant insight into understanding SOX3s function in neural cell types. In this manuscript we assess SOX3's genome wide binding sites with ChIP-seq, using our highly specific SOX3 antibody in testes from P7 wild type mice. Manuscript II identifies 778 SOX3 bound regions, which through gene ontology analyses appear to implicate SOX3 in histone regulation during spermatogenesis. We also highlight that this may be SOX3's core role, as many of these histone related binding sites are also bound by SOX3 in NPCs.

CHAPTER 3

PUBLICATION I

3.0 STATEMENT OF AUTHORSHIP

Statement of Authorship

Title of Paper	Identification of Highly Conserved Putative Developmental Enhancers Bound by SOX3 in Neural Progenitors Using ChIP-Seq
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	McAninch D, Thomas P (2014) Identification of Highly Conserved Putative Developmental Enhancers Bound by SOX3 in Neural Progenitors Using ChIP-Seq. PLOS ONE 9(11): e113361. https://doi.org/10.1371/journal.pone.0113361

Principal Author

Name of Principal Author (Candidate)	Dale Christopher McAninch		
Contribution to the Paper	Conceived and designed the study, performed the experiments, analysed the data, wrote and reviewd the paper		
Overall percentage (%)	80		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	25/02/2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Paul Thomas		
Contribution to the Paper	Conceived, designed and evaluated the study, revised and edited the manuscript		
Signature		Date	18.3.2020

Name of Co-Author			
Contribution to the Paper			
Signature		Date	

Please cut and paste additional co-author panels here as required.

OPEN ACCESS Freely available online

Identification of Highly Conserved Putative Developmental Enhancers Bound by SOX3 in Neural Progenitors Using ChIP-Seq

Dale McAninch, Paul Thomas*

Department of Biochemistry, School of Molecular & Biomedical Science and Robinson Research Institute, The University of Adelaide, Adelaide, Australia

Abstract

The transcription factor SOX3 is expressed within most neural progenitor (NP) cells of the vertebrate central nervous system (CNS) and is essential for normal brain development in mice and humans. However, despite the widespread expression of *Sox3*, CNS defects in null mice are relatively mild due to functional redundancy with the other SOXB1 sub-group members Sox1 and Sox2. To further understand the molecular function of SOX3, we investigated the genome-wide binding profile of endogenous SOX3 in NP cells using ChIP-seq. SOX3 binding was identified at over 8,000 sites, most of which were intronic or intergeneic and were significantly associated with neurodevelopmental genes. The majority of binding sites were moderately or highly conserved (phastCons scores >0.1 and 0.5, respectively) and included the previously characterised, SOXB1-binding *Nestin* NP cell enhancer. Comparison of SOX3 and published ChIP-Seq data for the co-activator P300 in embryonic brain identified hundreds of highly conserved putative enhancer elements. In addition, we identified a subset of highly conserved putative enhancers for CNS development genes common to SOXB1 members in NP cells, all of which contained the SOX consensus motif (ACAAWR). Together these data implicate SOX3 in the direct regulation of hundreds of NP genes and provide molecular insight into the overlapping roles of SOXB1 proteins in CNS development.

Citation: McAninch D, Thomas P (2014) Identification of Highly Conserved Putative Developmental Enhancers Bound by SOX3 in Neural Progenitors Using ChIP-Seq. PLoS ONE 9(11): e113361. doi:10.1371/journal.pone.0113361

Editor: Deyou Zheng, Albert Einstein College of Medicine, United States of America

Received August 6, 2014; **Accepted** October 22, 2014; **Published** November 19, 2014

Copyright: © 2014 McAninch, Thomas. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All ChIP-Seq files are available from the NCBI GEO database (accession number GSE57186).

Funding: This work was fully funded by the Australian Research Council. Funding received by PT Grant #: DP140103965, URL: http://www.arc.gov.au/. No other sources of funding were used for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: paul.thomas@adelaide.edu.au

Introduction

The SOX (Sry-related HMG box) family of transcription factors (TFs) are expressed in most if not all developing tissues and have critical roles in stem/progenitor cell induction, maintenance and differentiation [1,2] SOX proteins bind to the minor groove of DNA via an HMG box that has at least 50% identity to the founding member SRY and recognise variations of the core consensus sequence AACAAW (W = A or T) [2–4]. In vivo, SOX factor binding typically occurs in association with partner proteins, many of which belong to other major TF families including POU-Oct and zinc finger proteins [5].

Twenty SOX genes have been identified in mammals, which have been divided into groups based on their overall sequence homology. *Sox3*, *Sox2* and *Sox1* belong to the SOXB1 subgroup. These genes are expressed in neural progenitor (NP) cells throughout the vertebrate neuroaxis and are generally downregulated during NP differentiation [6]. In vitro and in vivo data indicate that SOX3 acts predominantly as a transcriptional activator, although there is also evidence supporting repressive activity [1,7,8]. Enforced expression of SOX3 in neural progenitors (NP) actively represses their differentiation functioning at least in part to repress Notch signalling [9]. Recent data also suggests that SOX3 may function as a pioneer factor through binding to neuronal-specific genes, priming them for subsequent activation by SOX11 [1,2]. Despite the widespread expression of *Sox3* in the developing CNS, *Sox3* null mice exhibit relatively mild neurodevelopmental defects, which are restricted to the hypothalamic-pituitary axis, the corpus callosum and the hippocampus [10,11]. CNS deletion of the other SoxB1 genes is also relatively mild [12,13]. Together, these data, coupled with overexpression analysis, indicate that SOXB1 proteins functionally interchangeable. This is supported by the recent observation that SOX3 binds to 96% of the known SOX2 binding sites within NP cells [2].

The development of ChIP-seq technology in recent years has provided invaluable insight into TF biology [14–16]. These data have highlighted the complexity of transcription factor activity by demonstrating TFs can have tens of thousands of binding sites within a single cell population. While it has been known for many years that TFs can act over long distances, a recent RNAPII ChIP-PET study has added to this complexity by providing further evidence for transcription factor mediated interchromsomal interactions [17]. Many TF binding sites are found at enhancers, promoting gene expression through the recruitment of TFs, cofactors (such as CBP/P300) and RNA Polymerase II (RNAPII) while looping DNA to the target promoter [18]. The ENCODE project has identified ~400,000 putative enhancer regions in human cell lines based on genomic traits including chromatin

methylation and acetylation status, evolutionary conservation and TF binding motifs [19]. Given the human and mouse genomes are in the same order of magnitude, it seems likely that there are a similar number of enhancers. By combining existing data for enhancer regions with TF binding site locations identified using ChIP-seq, we can identify putative enhancers for transcription factors such as SOX3, and begin to understand the functional significance of the vast expanses of non-coding genomic regions.

Identifying SOX3 binding sites and enhancers is crucial for complete understanding of the role of SOXB1 proteins in neural development. Here we present a genome-wide analysis of SOX3 binding in NP cells using ChIP-Seq. Through integration of this data with additional existing datasets we provide evidence that SOX3 and its SOXB1 partners activate hundreds of neurodevelopmental genes through binding to evolutionarily conserved sequences located principally within intergeneic regions. We also identify a putative multi-gene transcriptional hub, implicating SOX3 in interchromosomal transcriptional regulation.

Results

Identification of SOX3 binding sites in Neural Progenitor cells

To identify genomic binding sites of endogenous SOX3 protein, we performed ChIP-Seq analysis of NP cells generated from embryonic stem cells by N2B27 neuroinduction [20]. We have shown previously that these NP cells exhibit robust SOX3 expression [21] and that the SOX3 antibody used for ChIP has specific activity in immunohistochemistry [6] and Western blot analyses [22]. A total of 8067 common binding sites were identified across three independent samples (Figure 1A; Table S1). ChIP-Seq data was validated using ChIP-qPCR on independently generated samples, with all but one of the SOX3 binding sites (SBS) tested showing enrichment (Figure 1B). A *de novo* analysis of the full set of ChIP peaks was performed to identify enriched DNA motifs. Comparison with the JASPAR database [23] confirmed that the most common motif was a SOX binding motif (Figure 1C) (with at least one occurrence within >70% of peaks, p-value less than 10^{-6}, and an expected background occurrence of 39%), which was similar to the motif identified in a recently published SOX3 ChIP dataset [2]. The second most common motif features paired SOX/POU binding sites separated by a single nucleotide (Figure 1C) (with at least one occurrence within >40% of peaks). Motifs for other neural TF classes, such as the *Zic, Klf* and *Engrailed* families, were also enriched within 1215, 979 and 647 peaks respectively (all with p-values less than 0.0001). Together, these data indicate successful immunoprecipitation of SOX3-associated chromatin.

Alignment of the peak sequences to the genome revealed that they were associated with 4636 unique nearest neighbouring genes (approximately 20% of genes from the mm9 genome [24]). The majority (60.5%) were located in intergenic regions, 28.5% were located within introns and the remaining 11% were located near the proximal promoter (Figure 1A). GO term analysis revealed that peaks located in intergenic regions were most significantly associated with genes involved in forebrain neuron development (including *Gli3*, *Pax6* and *Wnt8b*) and hindbrain development (including *Hoxa1, Smo* and *Wnt8a*) (Figure 1D; Table S2). Peaks located within intronic regions were highly associated with neural tube development and formation (such as *Shroom3*, *Wnt3* and *Zeb2*). In contrast, peaks located at promoters were significantly associated with genes involved in chromatin modification (including a range of Histone 1, 2 and 4 variants and *Rbbp4*) (Figure 1D; Table S2).

Conservation of SOX3 binding sites

To assess the evolutionary conservation of each peak, we calculated the average phastCons score from data generated from 30 placental mammals provided from the UCSC database. Scores range from 0 to 1 where a value greater than 0.1 shows some conservation between placental mammals, while a value greater than 0.5 is considered highly conserved [25]. The majority (more than 56%) of peaks showed had an average score above 0.1, while more than 20% of peaks exhibited a high level of conservation with a value greater than 0.5 (Figure 2A). When peaks were sorted by genomic location those located in intronic or intergenic regions showed similar levels of conservation (more than 50% above 0.1), while more than 70% of peak located at promoters scored above 0.1. All three locations had a similar percentage of highly conserved peaks scoring more than 0.5 (approximately 20% each). For example, the SOX3 peak within intron 2 of *Dbx1*, a known target of SOX3 [21] has a conservation score of 0.97 and is located within a region of high conservation (Figure 2B). The SOX3 peak within intron 2 of *Nestin*, a well-characterised enhancer shown to bind SOXB1 proteins [26], has an average phastCons score of 0.51 where half the peak is highly conserved while the remaining half is not (Figure 2C).

Identification of putative neural enhancers that bind SOX3

To identify possible enhancers bound by SOX3 in NP cells, we overlayed our ChIP-Seq dataset with ChIP-Seq data for the coactivator protein P300 generated from 11.5 dpc embryonic mouse forebrain and midbrain [27]. Peaks were classified as overlapping if the midpoints of each peak were within 300 bp. This comparison revealed that SOX3 bound to approximately 20% and 29% of P300 enhancer regions in the forebrain and midbrain, respectively (Figure 3A, B). Although relatively small, the overlap between these datasets is highly significant, with p-values less than 10^{-5}. In contrast, comparison of SOX3 ChIP-seq data with P300 sites from the 11.5 dpc mouse limb bud revealed less that 5% overlap (a non-significant overlap, $P<0.6$) (Figure 3C). Interestingly, greater than 85% of the common peaks in fore- and midbrain samples had conservation scores above 0.1 while 32% and 41% of fore- and midbrain samples, respectively, were highly conserved (>0.50; Figure 3D, E).

Identification of conserved SOXB1 binding sites

Given the overlapping expression and functional redundancy of *Sox1*, *Sox2* and *Sox3*, we next attempted to identify common binding sites for SOXB1 proteins in the developing CNS. Comparison of our SOX3 ChIP-Seq dataset to existing ChIP-Seq data for SOX2 and SOX3 generated from a similar NP cell type [2] revealed 648 binding sites that were common to all three datasets (Figure 4A). Strikingly, MEME-ChIP analysis identified two variants of the SOX consensus motif, one of which is common to all 648 peaks (Figure 4B i), and the second present within 285 peaks (Figure 4B ii). 50% of these peaks showed high conservation, with phastCons scores greater than 0.5, while more than 80% have a score of more than 0.1 (Figure 4C).

Only 2.5% of the SOXB1 binding sites were located in promoter regions, whereas 34% were located within introns and the remaining 63.5% were located within intergenic regions (Figure 4D). GO term enrichment for the 648 SOXB1 peaks indicated that transcription factors were the most common genes regulated by these binding sites (Figure 4E). Intronic sites tend to regulate genes involved in neurogenesis (such as *Fezf2, Robo1*, and *Slit1*) while intergenic sites bind near transcription factors (such as

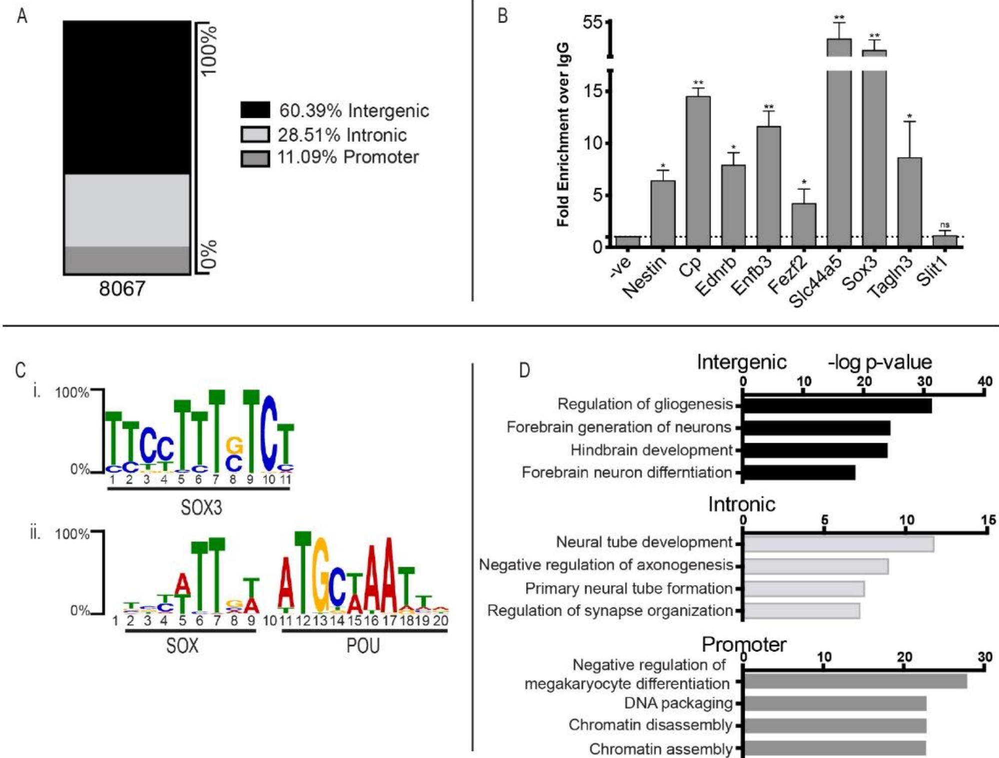


Figure 1. Overview of SOX3 ChIP-Seq data from mouse neural progenitor cells. (A) Genomic classification of SOX3 binding sites relative to nearest transcriptional start sites. (B) Validation of SOX3 ChIP by qPCR. Fold change is relative to both input DNA and IgG control values for the same genomic location. Error bars correspond to standard deviation of three independent sample replicates, P-values indicated as <0.05 (ns), >0.05 (*) and >0.001 (**). (C) Highest enriched DNA motifs identified by MEME-ChIP as i. a SOX motif and ii. a SOX-POU motif. (D) Enriched Gene Ontology terms associated with subsets of SOX3 ChIP peaks.
doi:10.1371/journal.pone.0113361.g001

Irx, *Nkx*, and *SOX* family members; Figure 4F). Together, these data define a core set of SOXB1 target sites that appear to have evolutionally conserved roles in NP cells.

Identification of SOX3 binding sites in SOX3 target genes

Although SOXB1 proteins have highly overlapping functions, comparison of genome-wide expression profiles of WT and *Sox3* null NP cells has identified a set of 19 genes with significantly different expression levels, suggesting that a small subset of SOXB1 targets are particularly sensitive to the loss of SOX3 [21]. To investigate whether these genes are direct SOX3 targets, we examined our ChIP-Seq data for binding sites with their intronic and flanking sequences. Thirteen of the 19 differentially expressed genes (68%, with an expected random frequency of 24%, and a p value of less than 0.0001) featured at least one ChIP peak (Table 1), two of which were located at promoters, nine within introns and the remaining seven within intergenic regions. Together, these data suggest that small subset of SOX3 direct target genes require SOX3 (and not other SOXB1 members) for normal expression.

SOX3 interaction with an interchromosomal transcriptional network

A recent study has published a dataset for chromatin interaction analysis with paired end tagging (ChIA-Pet) of RNA polymerase II in neural stem cells [17]. They identified more than 5,000 putative enhancers linked to the promoter of genes on different chromosomes (interchromosomal interactions), as well as more than 10,000 enhancers linked to distant genes on the same chromosome. We sought to identify whether SOX3 could be linked to any of these putative inter- or intra- chromosomal enhancers identified from a neural stem cell population. From the 8067 SOX3 peaks identified, 97 overlapped with potential long-range enhancers (a significant overlap, P<0.001, with an expected overlap of 34 by chance) that can be linked to 304 and 246 inter- and intrachromosomal promoters, respectively. For example, SOX3 binds an intronic enhancer within *Tex14* (Figure 5) that can be linked to 263 different promoters and enhancers. This putative enhancer has a phastCons score of 0.12, moderate evolutionary conservation, and features a single SOX binding site. These data suggest that SOX3 may be involved in complex, long-range gene regulation.

Discussion

This study has identified 8067 regions within the genome of murine NP cells that are associated with SOX3. The majority of the SOX3 binding sites are not located at the proximal promoter of genes but rather are within intronic or intergenic regions, suggesting an extensive regulatory (enhancer-binding) role for SOX3 in NP cells. Interestingly, the most common motif within

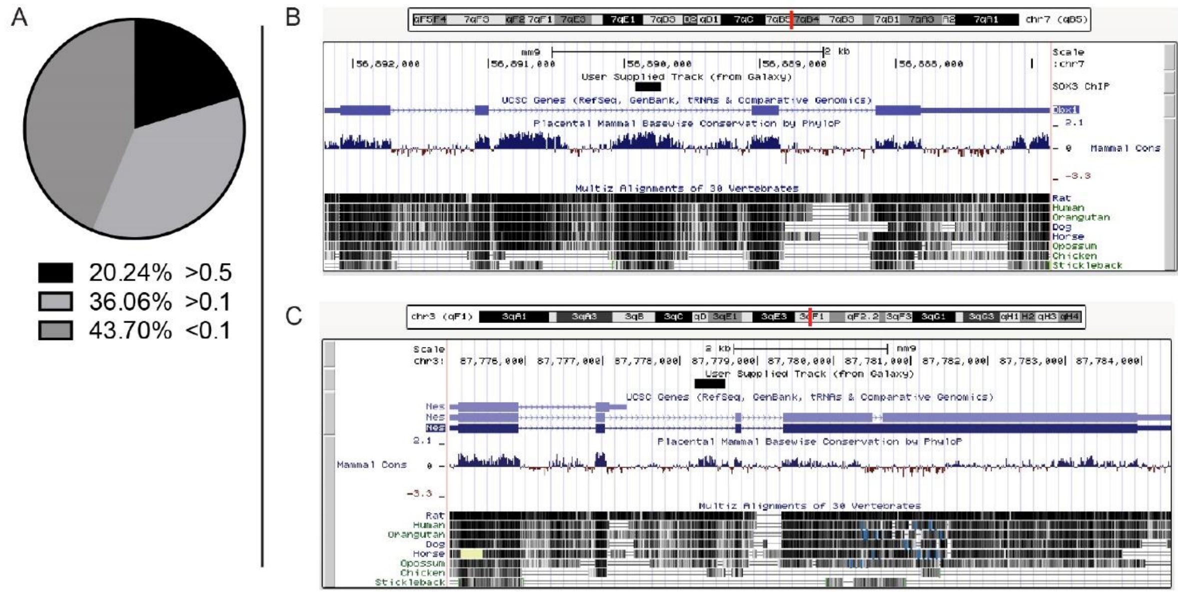


Figure 2. Evolutionary conservation of SOX3 bound regions. (A) The average phastCons score of each SOX3 bound peak showing 20% of peaks are highly conserved across 30 placental mammals. (B) A highly conserved peak within the second intron of *Dbx1* giving the highest conservation score of 0.97 (compared to all peaks). (C) A peak within the second intron of the neural gene *Nestin*, with an average phastCons score of 0.51.

doi:10.1371/journal.pone.0113361.g002

these peaks was a variation of the standard SOX DNA binding motif that contained a strong preference for a C residue at position 10 (C_{10}) instead of a T (T_{10}). Significant enrichment for C_{10} was also identified by Bergsland et al. 2011 using an independent SOX3 antibody, although not to the same degree as in our study. Previous structural studies of SOX2 binding to a T_{10}-containing binding site indicate that T_{10} makes contact with arginine 5 of the HMG domain as well as the c-terminal tail. However, as residues within these regions are completely conserved throughout SOXB1 members, it remains unclear why SOX3 prefers to bind the C_{10} sequence *in vivo*. Given that partner factors can influence SOXB1 binding site preference [28], it is possible that the strong enrichment for the C_{10} site may reflect SOX3 partner protein usage in NPCs.

SOX3-POU cooperation

SOX transcription factors cooperate extensively with specific partner factors, including OCT, BRN, or other SOX family

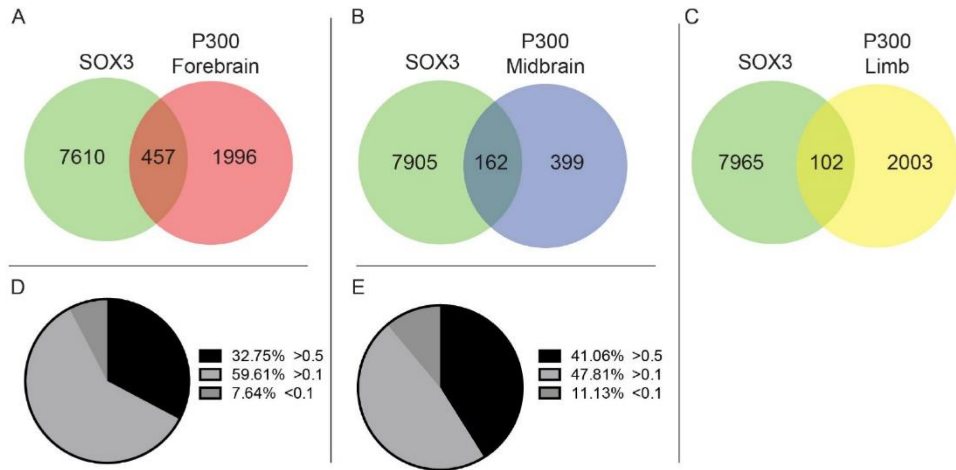


Figure 3. SOX3 binding sites at enhancer regions. The overlap of SOX3 peaks with P300 binding sites identified from 11.5 dpc mouse; forebrain (A), midbrain (B), and limb (C), showing a high degree of overlap in the developing brain and not within the limb. Average phastCons score of the common peaks between SOX3 and P300 forebrain (D) and midbrain (E) binding sites.

doi:10.1371/journal.pone.0113361.g003

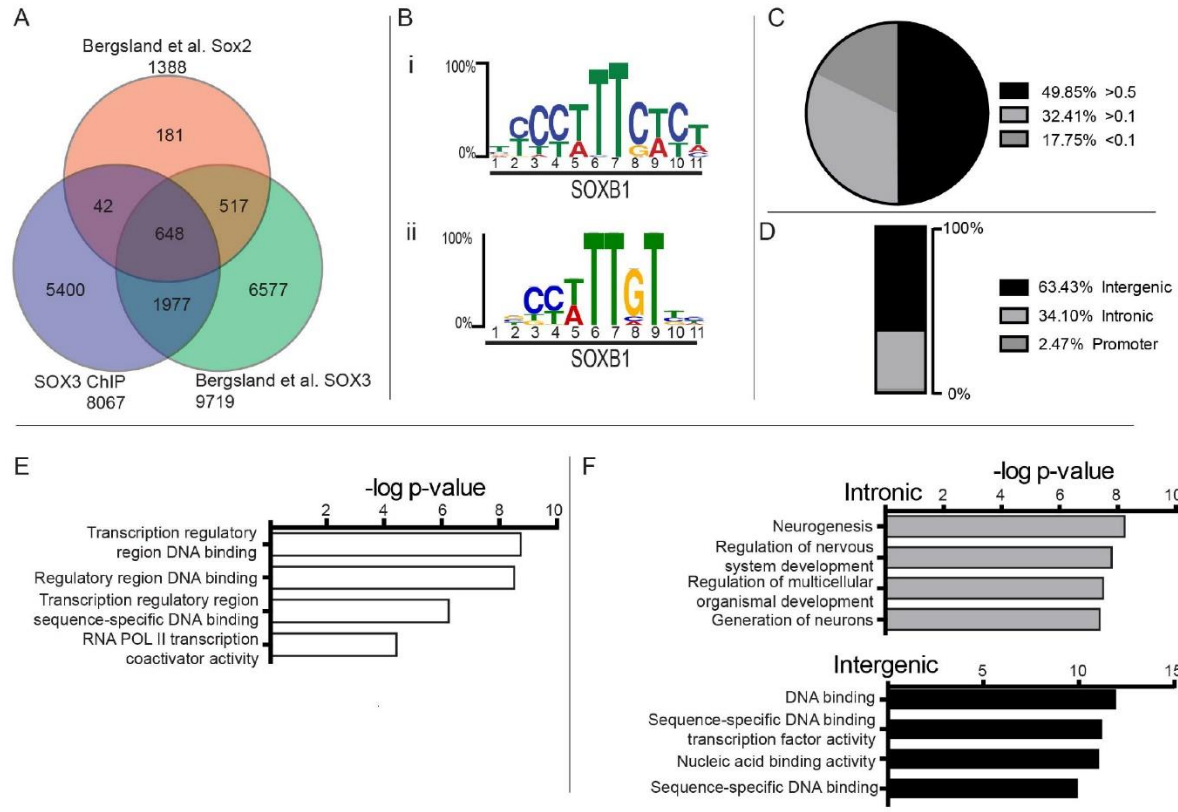


Figure 4. Identification of common SOXB1 regulatory regions. (A) Overlap of this SOX3 ChIP-seq dataset with previously published SOX2 and SOX3 datasets from similar NPCs (2). (B) SOX motif identified with MEME-ChIP present in (i) all 648 SOXB1 common peaks and (ii) 295 peaks. (C) Average phastCons scores for the 648 SOXB1 peaks, showing more than 80% of peaks are either moderately (32%) or highly (42%) conserved. (D) Genomic localisation of the common SOXB1 peaks. Enriched Gene Ontology terms for (E) all 648 SOXB1 peaks and (F) both intronic and intergenic peaks.

doi:10.1371/journal.pone.0113361.g004

members [29]. Co-factor DNA binding is a common developmental mechanism that provides exquisite temporal and spatial expression of target genes [30]. The second most common motif identified from the complete SOX3 ChIP-Seq dataset (within 1998/8067 peaks) was a combined SOX/POU DNA binding motif (Figure 1). Similar enrichment of this motif was also identified in the SOXB1 subset (126/648 peaks; p value < 0.0001) [31]. The nucleotides comprising the POU component of the SOX/POU motif show minimal variation, while the nucleotides comprising the SOX motif show greater variability, suggesting the POU-DNA interaction is more sensitive to sequence composition. To assign functional significance to these binding sites it would be useful to delete these binding sites either completely or each of the SOX and POU binding sites separately [32]. This could provide information as to the importance of these linked binding sites indicating whether both factors required or if one is sufficient.

Many SOX3 binding sites exist within biologically important enhancers

Although the preferential binding of SOX3 in or near many known neurodevelopmental genes is suggestive of a wide-ranging regulatory role in NP cells, it is difficult to identify the functional significance of these sites from binding data alone. To address this, we assessed the evolutionary conservation of each ChIP peak, with the rationale that a highly conserved peak has the potential to be more biologically relevant if selection pressure has maintained sequence conservation throughout evolution. The proposed link between conservation and function is supported by previous studies that have demonstrated that developmental enhancers can be reliably identified solely on sequence conservation [33,34]. Our data show over 50% of SOX3 peaks are either moderately (36%) or highly (20%) conserved, having PhastCons scores above 0.1 or 0.5, respectively. Amongst the highly conserved peaks we identified the well-characterised intronic *Nestin* enhancer that has been shown previously to bind SOXB1 proteins *in vitro* and drives NP cell expression of reporter genes in the developing CNS [26]. We also identified a SOX3 peak within the second intron of *Dbx1* that had the highest overall conservation score and has been linked to the expression of *Dbx1* in cultured NPCs and the spinal cord of 9.5 dpc *Sox3* null mice [21]. In addition, SOX3 peaks were identified at independently identified *cis* regulatory motifs (CRM) shown to respond to SOXB1 transcription factors including *Olig2, Dbx2* and *Nkx2.2*, although not at the remaining CRMs identified in this study [35]. Further comparison of our data to ChIP-Seq data for the co-activator P300 from mouse 11.5 dpc forebrain and midbrain, identified 457 (19.85%) and 162 (28.88%) common peaks, respectively. Many of these P300

Table 1. Differentially expressed genes from *Sox3* null NPCs with nearby SOX3 ChIP binding sites.

Gene	Fold Change	RefSeq ID	Peak Coordinates	Location
Tmem163	−1.45	NM_028135	chr1:129472320-129472556	Intron
Slc44a5	1.44	NM_001081263	chr3:153836591-153836875	Intron
Fgfr3	1.65	NM_008010	chr5:34047895-34048514	Intergenic
Cpv1	1.41	NM_025817	chr6:53860009-53860356	Intron
			chr6:53984529-53984835	Intergenic
Dbx1	−2.35	NM_001005232	chr7:56889725-56889913	Intron
Gpr56	1.56	NM_018882	chr8:97524576-97524923	Intron
Cspg5	1.49	NM_013884	chr9:110154883-110154975	Intron
			chr9:110183673-110183904	Intergenic
Ctgf	1.45	NM_010217	chr10:24310458-24310740	Promoter
Flrt2	1.46	NM_201518	chr12:95662522-95662885	Intergenic
Ednrb	1.53	NM_007904	chr14:104243418-104243673	Promoter
			chr14:104298351-104298531	Intergenic
			chr14:104298536-104298789	Intergenic
Tagln3	1.54	NM_019754	chr16:45724870-45725203	Promoter
Slit1	1.87	NM_015748	chr19:41745438-41745636	Intron
			chr19:41773461-41773782	Intron
			chr19:41790683-41791049	Intron
Sox3	−4.10	NM_009237	chrX:57972960-57973254	Intergenic

doi:10.1371/journal.pone.0113361.t001

associated regions have been shown to function as CNS enhancers *in vivo* [27]. Interestingly, overlapping SOX3/P300 peaks showed an increased enrichment for high conservation with >32% (forebrain) and >41% (midbrain) of peaks being highly conserved. Overlapping P300 peaks from the limb also showed high conservation, approximately 42%, which is to be expected given P300's high level of association with enhancers. Although only a small percentage of the total number of SOX3 ChIP peaks, this is only data pertaining to one co-activator (P300), it is likely that other peaks feature alternate co-activators or other proteins required for SOX3 to act as a pioneering factor [2]. Taken together, these data suggest that many SOX3 peaks are likely to correspond to important NP enhancers.

Previous ChIP-Seq analysis of SOX3 in a similar population of NP cells also identified extensive binding across the genome (9719 SOX3 binding sites). Comparison to our dataset identifies more than 15,000 unique SOX3 binding sites, 2625 of which are common to both studies. Although both analyses produced a similar number of binding sites individually, the overlap in binding sites between datasets was lower than anticipated. This variability is likely to be caused by use of different NPCs. Although both studies used NPCs that were generated from ES cells by N2B27 induction [20], the Bergsland et al. culture conditions also included bFGF, SHH, and retinoic acid [2]. These additional factors have major roles in CNS patterning and are therefore likely to influence the identity of the overall NP cell population. It has been observed previously the degree of overlap of binding sites between different cell lines can vary, from less than 50% (SRF peaks from 3 different human cell lines [36]) to more than 80% overlap (E2F4 peaks from multiple primary mouse tissues and cell lines [37]). It is also important to note that different SOX3 antibodies were used in the ChIP-seq studies. Although both have been shown to be specific they likely recognise different SOX3 epitopes, potentially introducing variation in the SOX3 binding sites identified.

Given the functional redundancy of SOXB1 proteins, we sought to identify genomic regions bound by SOX3 and SOX2 in NP cells. The 648 common SOXB1 binding sites showed greater enrichment for conserved peaks than any of the datasets alone with approximately 50% having high conservation. Remarkably, two variants of the SOX motif were identified through *de novo* screening of the 648 SOXB1 peaks. The most common motif was found in all 648 peaks with a highly significant E value (4.7e-371). This motif is very similar to the one identified from the complete SOX3 dataset (although there was no nucleotide preference observed at position 10), and essentially identical across positions 6, 7 and 9 which are known to make contact with the HMG domain [4]. The second SOX motif identified was present in 285 peaks (with an E value of 1.5e-23) and was more similar to the SOX3 motif, with an increased nucleotide preference at position 10. GO term analysis indicated that SOXB1 common peaks bind nearest to transcription factors, in particular members of the SOX family (including *Sox1, Sox2, Sox4, Sox5, Sox6, Sox9, Sox11*), *Nkx2.1 and Nkx2.2* all with known roles in neural development. The intronic peaks are enriched for genes involved in neural development, including *Fzd3* and *Fzd6* both of which are required for correct ventricle formation in the developing midbrain [38]. Together these data give evidence that SOXB1 binding sites have been highly conserved throughout evolution and potentially regulate genes important for neural development, and also suggests extensive cross regulation of SOX factors consistent with analyses of individual factors [39,40].

The expression level of most NP genes is not affected by the loss of *Sox3* [21], presumably due to functional redundancy with other SOXB1 members. However, we have previously identified nineteen genes that have altered expression in *Sox3* null NP cells, suggesting that some genes are particularly sensitive to SOX3 loss

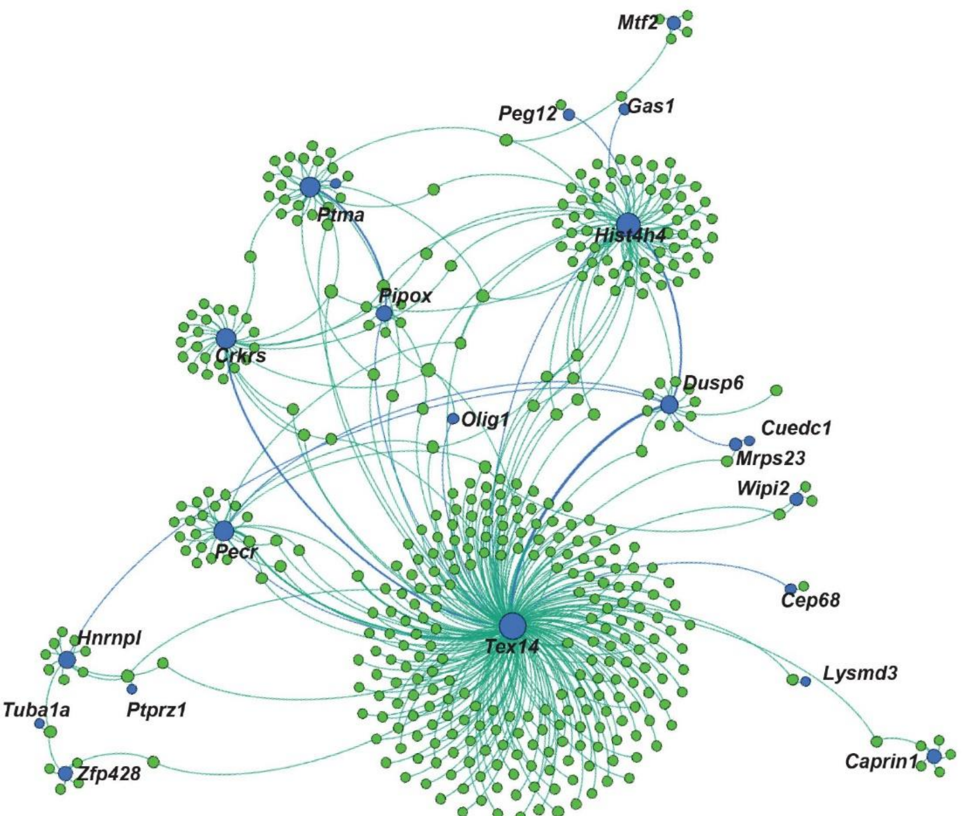


Figure 5. SOX3 is implicated in an interchromosomal regulatory network. Graphical representation of potential long range intra- and interchromosomal interactions as identified by overlap of SOX3 peaks with RNAPII ChIA-PET peaks. Blue circles are sites present in both SOX3 and RNAPII datasets, while green circles are the linked genomic regions present only in the RNAPII dataset. The nearest TSS is labelled for the common binding sites.
doi:10.1371/journal.pone.0113361.g005

of function. Here we show that the majority of these genes (68%) are directly bound by SOX3. The high frequency of binding sites in differentially expressed genes appears significant given that SOX3 binding sites on average are found in only 20% of genes genome wide. The mechanism that underpins the sensitivity of these direct targets to the loss of SOX3 alone is currently unclear. Given that the HMG box sequences of the SOXB1 proteins are not identical, one possibility is that the binding sites in these differentially expressed genes have a higher affinity for SOX3 than other SOXB1 proteins. Alternatively, these genes may be sensitive to the overall dosage of SOXB1 protein rather than SOXB1 itself. Additional studies such as SoxB1 gene swap experiments are required to further investigate this issue.

Finally we have identified a number of potential genome wide interactions, linking enhancers and promoters of genes regulated by SOX3. It was observed that an intronic enhancer within *Tex14* could be linked to many other enhancers and promoters, forming a transcriptional hub. These networks may aid in ensuring transcription of specific genes occurs at similar time, potentially allowing for quicker responses. As noted by Zhang et al. 2013 the RNAPII data identifies regions of pre-initiation events; as such not all genes will be actively transcribed. Further to this, it is likely that there are other long-range interactions that do not involve RNAPII. It could be informative to perform a ChIA-PET against SOX3 in NPCs to see what long-range interactions occur independent of RNAPII. It would also be interesting to see whether deleting one of the binding sites that form part of a transcriptional hub, such as *Tex14*, has an effect on the transcription at connected genomic locations.

Conclusions

In conclusion, we have shown that SOX3 binds extensively to evolutionarily conserved sequence in or near known neurodevelopmental genes that are likely to function as enhancers. Further functional validation within an *in vivo* system is required to assess the functional importance of these data. With the recent emergence of genome editing tools, CRISPR and TALENs, the generation of genetically modified mice has become significantly more streamlined and efficient. This technology will allow for the simple deletion or modification of binding sites in mice to highlight the importance of each binding site providing biological significance.

Materials and Methods

NPC generation

Mouse R1 ES cells, as described previously [10,22], were passaged without feeders in DMEM (Gibco) in the presence of LIF

and foetal calf serum. ES cell monolayers were cultured in N2B27 media for 4 days to produce NPCs as described previously [20].

SOX3 ChIP-seq

NPCs were fixed in 1% formaldehyde for 8 minutes at room temperature, lysed and sonicated (Bioruptor, Diagenode) for 15 minutes in 1-minute pulses on ice. SOX3 bound chromatin was immunoprecipitated by a goat polyclonal antibody raised against human SOX3 (R&D systems, AF2569). DNA was recovered by reversing crosslinks, and purified by PCR clean-up kit (QIAGEN). Three independent DNA libraries were produced with the Illumina TrueSeq library kit as per manufacturer's instructions, and libraries were sequenced on the Illumina HiSeq producing 50 bp single end reads. A control sample (without SOX3 antibody) was run as input for background control. ChIP samples were validated by qPCR (StepOne Plus, Applied Biosystems) using Fast SYBR (Life Technologies). Signals were considered positive when Ct_{sample} (non-enriched region) - Ct_{sample} (peak region) was >2 following normalisation to Ct_{IgG}.

Peak Calling

Bowtie [41] was used to align reads to the mouse genome (mm9). Peaks were called for each biological replicate using MACS [42], with bandwidth of 300, a model fold of 10–30, using input sample as background control, and a p-value threshold of 1e-5. Only peaks present in all 3 biological replicates were retained.

Gene ontology

Gene ontology was performed using GREAT (http://great.stanford.edu/) for regions associated with SOX3 bound enhancers [43]. Default 'basal plus' parameters were used to define gene locus, and the whole genome was used for background regions.

De Novo motif enrichment

MEME (http://meme.sdsc.edu/) was used to identify DNA enriched motifs from SOX3 bound peaks [44]. WebLogo [45] was used to visualise *de novo* motifs generated by MEME-ChIP.

Published ChIP-Seq data

ChIP-Seq data on SOX2 and SOX3 in NPCs were obtained from GSE33024, P300 data from GSE10516, ChIA-PET data from GSE44067.

Primers for qPCR

Primers for validating ChIP samples by qPCR are listed in a 5′ to 3′ direction as follows: *Nes* F- GCCCCAGTCAGTCTTCT-GAG, R- GCTGGTGACAGACAAAAGCA,

Cp F- CTCACACTGTGCTGGGCTAA, R- AGGAAGT-fTGTGCAACTCTGGA,

Ednrd F- CTAAACAGGCCTCTCGCAAC, R- TTGTC-TGGGGACAGCAAAG,

Enfb3 F- ATGCTCAGCACCTCATTGG, R- GGCACGT-GACTGGTGGTAG,

Fezf2 F- GGTCGTCTTTTCTTCCTGTCC, R- ATCCA-CAGAACCAGCATCACT,

Slc44a5 F- CTGCCTGGATGTCAGGATTT, R- CCCA-CAGTGTTTGTAGGAACG,

Sox3 F- ATGAGTTTCCGGAATGTTGC, R- CTTCTCA-CTTCCTGCCCTTG,

Tagln3 F- CCTCTCCTAGACAGGCCAGA, R- GTGG-GGCCTCAGATACAATG,

Slit1 F- AGACGGACCTGGGAAATTCT, R- CCAGAAAG-CAGGATTTGCAT

Supporting Information

Table S1 A list of all SOX3 peaks identified by ChIP-seq. Featuring mm9 coordinates, nearest gene, RefSeq gene ID and distance from centre of peak to nearest transcriptional start site.
(XLSX)

Table S2 A list of the top 20 GO terms associated with the intergenic, intronic and promoter SOX3 ChIP-seq peaks.
(XLS)

Acknowledgments

Thanks to Jonas Muhr for providing valuable information regarding ChIP positive controls and to the Adelaide Cancer Research Foundation for performing the ChIP library preparations and next generation sequencing reactions.

Author Contributions

Conceived and designed the experiments: DM PT. Performed the experiments: DM. Analyzed the data: DM PT. Contributed reagents/materials/analysis tools: DM PT. Wrote the paper: DM PT.

References

1. Bylund M, Andersson E, Novitch BG, Muhr J (2003) Vertebrate neurogenesis is counteracted by Sox1-3 activity. Nat Neurosci 6: 1162–1168.
2. Bergsland M, Ramskold D, Zaouter C, Klum S, Sandberg R, et al. (2011) Sequentially acting Sox transcription factors in neural lineage development. Genes Dev 25: 2453–2464.
3. Bowles J, Schepers G, Koopman P (2000) Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. Dev Biol 227: 239–255.
4. Remenyi A, Lins K, Nissen LJ, Reinbold R, Scholer HR, et al. (2003) Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. Genes Dev 17: 2048–2059.
5. Yoshida M, Uchikawa M, Rizzoti K, Lovell-Badge R, Takemoto T, et al. (2014) Regulation of mesodermal precursor production by low-level expression of B1 Sox genes in the caudal lateral epiblast. Mech Dev 132: 59–68.
6. Rogers N, Cheah PS, Szarek E, Banerjee K, Schwartz J, et al. (2013) Expression of the murine transcription factor SOX3 during embryonic and adult neurogenesis. Gene Expr Patterns 13: 240–248.
7. Graham V, Khudyakov J, Ellis P, Pevny L (2003) SOX2 functions to maintain neural progenitor identity. Neuron 39: 749–765.
8. Shih YH, Kuo CL, Hirst CS, Dee CT, Liu YR, et al. (2010) SoxB1 transcription factors restrict organizer gene expression by repressing multiple events downstream of Wnt signalling. Development 137: 2671–2681.
9. Holmberg J, Hansson E, Malewicz M, Sandberg M, Perlmann T, et al. (2008) SoxB1 transcription factors and Notch signaling use distinct mechanisms to regulate proneural gene function and neural progenitor differentiation. Development 135: 1843–1851.
10. Rizzoti K, Brunelli S, Carmignac D, Thomas PQ, Robinson IC, et al. (2004) SOX3 is required during the formation of the hypothalamo-pituitary axis. Nat Genet 36: 247–255.
11. Rizzoti K, Lovell-Badge R (2005) Early development of the pituitary gland: induction and shaping of Rathke's pouch. Rev Endocr Metab Disord 6: 161–172.
12. Favaro R, Valotta M, Ferri AL, Latorre E, Mariani J, et al. (2009) Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. Nat Neurosci 12: 1248–1256.
13. Miyagi S, Masui S, Niwa H, Saito T, Shimazaki T, et al. (2008) Consequence of the loss of Sox2 in the developing brain of the mouse. FEBS Lett 582: 2811–2815.
14. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, et al. (2007) High-resolution profiling of histone methylations in the human genome. Cell 129: 823–837.
15. Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science 316: 1497–1502.

16. Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, et al. (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods 4: 651–657.
17. Zhang Y, Wong CH, Birnbaum RY, Li G, Favaro R, et al. (2013) Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature 504: 306–310.
18. Krivega I, Dean A (2012) Enhancer and promoter interactions-long distance calls. Curr Opin Genet Dev 22: 79–85.
19. Consortium EP, Bernstein BE, Birney E, Dunham I, Green ED, et al. (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57–74.
20. Ying QL, Smith AG (2003) Defined conditions for neural commitment and differentiation. Methods Enzymol 365: 327–341.
21. Rogers N, McAninch D, Thomas P (2014) Dbx1 Is a Direct Target of SOX3 in the Spinal Cord. PLoS One 9: e95356.
22. Hughes J, Piltz S, Rogers N, McAninch D, Rowley L, et al. (2013) Mechanistic insight into the pathology of polyalanine expansion disorders revealed by a mouse model for X linked hypopituitarism. PLoS Genet 9: e1003290.
23. Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, et al. (2014) JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res 42: D142–147.
24. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, et al. (2002) The human genome browser at UCSC. Genome Res 12: 996–1006.
25. Fan R, Bonde S, Gao P, Sotomayor B, Chen C, et al. (2012) Dynamic HoxB4-regulatory network during embryonic stem cell differentiation to hematopoietic cells. Blood 119: e139–147.
26. Kawaguchi A, Miyata T, Sawamoto K, Takashita N, Murayama A, et al. (2001) Nestin-EGFP transgenic mice: visualization of the self-renewal and multipotency of CNS stem cells. Mol Cell Neurosci 17: 259–273.
27. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, et al. (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457: 854–858.
28. Kondoh H, Kamachi Y (2010) SOX-partner code for cell specification: Regulatory target selection and underlying molecular mechanisms. Int J Biochem Cell Biol 42: 391–399.
29. Kamachi Y, Uchikawa M, Kondoh H (2000) Pairing SOX off: with partners in the regulation of embryonic development. Trends Genet 16: 182–187.
30. Spitz F, Furlong EE (2012) Transcription factors: from enhancer binding to developmental control. Nat Rev Genet 13: 613–626.
31. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, et al. (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37: W202–208.
32. Bery A, Martynoga B, Guillemot F, Joly JS, Retaux S (2013) Characterization of Enhancers Active in the Mouse Embryonic Cerebral Cortex Suggests Sox/Pou cis-Regulatory Logics and Heterogeneity of Cortical Progenitors. Cereb Cortex.
33. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, et al. (2006) In vivo enhancer analysis of human conserved non-coding sequences. Nature 444: 499-502.
34. Visel A, Prabhakar S, Akiyama JA, Shoukry M, Lewis KD, et al. (2008) Ultraconservation identifies a small subset of extremely constrained developmental enhancers. Nat Genet 40: 158–160.
35. Oosterveen T, Kurdija S, Enstero M, Uhde CW, Bergsland M, et al. (2013) SoxB1-driven transcriptional network underlies neural-specific interpretation of morphogen signals. Proc Natl Acad Sci U S A 110: 7330–7335.
36. Cooper SJ, Trinklein ND, Nguyen L, Myers RM (2007) Serum response factor binding sites differ in three human cell types. Genome Res 17: 136–144.
37. Conboy CM, Spyrou C, Thorne NP, Wade EJ, Barbosa-Morais NL, et al. (2007) Cell cycle genes are the evolutionarily conserved targets of the E2F4 transcription factor. PLoS One 2: e1061.
38. Stuebner S, Faus-Kessler T, Fischer T, Wurst W, Prakash N (2010) Fzd3 and Fzd6 deficiency results in a severe midbrain morphogenesis defect. Dev Dyn 239: 246–260.
39. Kamachi Y, Kondoh H (2013) Sox proteins: regulators of cell fate specification and differentiation. Development 140: 4129–4144.
40. Lee KE, Seo J, Shin J, Ji EH, Roh J, et al. (2014) Positive feedback loop between Sox2 and Sox6 inhibits neuronal differentiation in the developing central nervous system. Proc Natl Acad Sci U S A 111: 2794–2799.
41. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.
42. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, et al. (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9: R137.
43. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, et al. (2010) GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 28: 495–501.
44. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2: 28–36.
45. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14: 1188–1190.

This page intentionally left blank.

CHAPTER 4

MANUSCRIPT I

4.0 STATEMENT OF AUTHORSHIP

Statement of Authorship

Title of Paper	SOX3 promotes generation of committed spermatogonia in postnatal mouse testes.
Publication Status	☐ Published ☐ Accepted for Publication ☒ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Manuscript reviewed by *Scientific Reports*, revisions submitted 20/02/2020

Principal Author

Name of Principal Author (Candidate)	Dale Christopher McAninch		
Contribution to the Paper	Conceived and designed study, performed experiments, analysed and interpreted data, wrote and revised manuscript		
Overall percentage (%)	50		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	03/03/2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Juho-Antti Mäkelä		
Contribution to the Paper	Performed experiments, helped analyse and interpret data, reviewed manuscript.		
Signature		Date	3/3/2020

Name of Co-Author	Hue M. La		
Contribution to the Paper	Performed experiments, helped analyse and interpret data, reviewed manuscript.		
Signature		Date	

Statement of Authorship

Title of Paper	SOX3 promotes generation of committed spermatogonia in postnatal mouse testes.		
Publication Status	☐ Published ☒ Submitted for Publication	☐ Accepted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style	
Publication Details	Manuscript reviewed by *Scientific Reports*, revisions submitted 20/02/2020		

Principal Author

Name of Principal Author (Candidate)	Dale Christopher McAninch		
Contribution to the Paper	Conceived and designed study, performed experiments, analysed and interpreted data, wrote and revised manuscript		
Overall percentage (%)	50		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	03/03/2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Juho-Antti Mäkelä		
Contribution to the Paper	Performed experiments, helped analyse and interpret data, reviewed manuscript.		
Signature		Date	

Name of Co-Author	Hue M. La		
Contribution to the Paper	Performed experiments, helped analyse and interpret data, reviewed manuscript.		
Signature		Date	03/03/2020

Name of Co-Author	**James N. Hughes**		
Contribution to the Paper	Helped with data analysis and interpretation, reviewed manuscript		
Signature		Date	4/3/2020

Name of Co-Author	**Robin Lovell-Badge**		
Contribution to the Paper	Helped with data analysis and interpretation, reviewed manuscript		
Signature		Date	

Name of Co-Author	**Robin M. Hobbs**		
Contribution to the Paper	Conceived and designed the study, helped with data analysis and interpretation, wrote and edited manuscript		
Signature		Date	

Name of Co-Author	**Paul Q. Thomas**		
Contribution to the Paper	Conceived and designed the study, helped with data analysis and interpretation, wrote and edited manuscript		
Signature		Date	

Name of Co-Author	**James N. Hughes**		
Contribution to the Paper	Helped with data analysis and interpretation, reviewed manuscript		
Signature		Date	

Name of Co-Author	**Robin Lovell-Badge**		
Contribution to the Paper	Helped with data analysis and interpretation, reviewed manuscript		
Signature		Date	11/03/2020

Name of Co-Author	**Robin M. Hobbs**		
Contribution to the Paper	Conceived and designed the study, helped with data analysis and interpretation, wrote and edited manuscript		
Signature		Date	

Name of Co-Author	**Paul Q. Thomas**		
Contribution to the Paper	Conceived and designed the study, helped with data analysis and interpretation, wrote and edited manuscript		
Signature		Date	

Name of Co-Author	**James N. Hughes**		
Contribution to the Paper	Helped with data analysis and interpretation, reviewed manuscript		
Signature		Date	

Name of Co-Author	**Robin Lovell-Badge**		
Contribution to the Paper	Helped with data analysis and interpretation, reviewed manuscript		
Signature		Date	

Name of Co-Author	**Robin M. Hobbs**		
Contribution to the Paper	Conceived and designed the study, helped with data analysis and interpretation, wrote and edited manuscript		
Signature		Date	3/3/2020

Name of Co-Author	**Paul Q. Thomas**		
Contribution to the Paper	Conceived and designed the study, helped with data analysis and interpretation, wrote and edited manuscript		
Signature		Date	18.3.2020

SOX3 promotes generation of committed spermatogonia in postnatal mouse testes.

Dale McAninch[1], Juho-Antti Mäkelä[2,3], Hue M. La[2,3], James N. Hughes[1], Robin Lovell-Badge[4], Robin M. Hobbs[2,3,*] and Paul Q. Thomas[1,*]

[1]School of Biological Sciences and Robinson Research Institute, University of Adelaide, Adelaide, SA 5005, Australia

[2]Australian Regenerative Medicine Institute, Monash University, Melbourne, VIC 3800, Australia

[3]Development and Stem Cells Program, Monash Biomedicine Discovery Institute and Department of Anatomy and Developmental Biology, Monash University, Melbourne, VIC 3800, Australia

[4]Laboratory of Stem Cell Biology and Developmental Genetics, The Francis Crick Institute, London NW1 1AT

[*]Authors for correspondence (paul.thomas@adelaide.edu.au; robin.hobbs@monash.edu)

Short Title

Role of SOX3 in spermatogonia

Key words

SOX3 spermatogonia, testes, mouse, *Ngn3*, GFRα1

4.1 ABSTRACT

SOX3 is a transcription factor expressed within the developing and adult nervous system where it mostly functions to help maintain neural precursors. *Sox3* is also expressed in other locations, notably within the spermatogonial stem/progenitor cell population in postnatal testis. Independent studies have shown that *Sox3* null mice exhibit a spermatogenic block as young adults, the mechanism of which remains poorly understood. Using a panel of spermatogonial cell marker genes, we demonstrate that *Sox3* is expressed within the committed progenitor fraction of the undifferentiated spermatogonial pool. Additionally, we use a *Sox3* null mouse model to define a potential role for this factor in progenitor cell function. We demonstrate that *Sox3* expression is required for transition of undifferentiated cells from a GFRα1+ self-renewing state to the NGN3+ transit-amplifying compartment. Critically, using chromatin immunoprecipitation, we demonstrate that SOX3 binds to a highly conserved region in the *Ngn3* promoter region *in vivo*, indicating that *Ngn3* is a direct target of SOX3. Together these studies indicate that SOX3 functions as a pro-commitment factor in spermatogonial stem/progenitor cells.

4.2 INTRODUCTION

SOX3 is a member of the SOX (Sry-related HMG box) family of transcription factors (TFs), of which there are 20 members in mammals. SOX TFs are expressed within many tissues of the mouse embryo and regulate a range of important cellular activities including self-renewal, specification and differentiation [164,165]. SOX TFs bind to variants of the SOX consensus motif (A/TA/TCAAA/TG) *via* a highly conserved HMG domain that shares at least 50% sequence identity with the founding member SRY. *Sox1, Sox2* and *Sox3* have high similarity across their entire open reading frame and together comprise the *SoxB1* subgroup. *SoxB1* genes are

expressed in NPCs throughout the entire vertebrate neuroaxis and are generally down regulated during differentiation [59,93]. Loss-of-function and overexpression experiments in a range of vertebrate systems indicate important and overlapping roles for SOXB1 factors in the generation and maintenance of neural stem/progenitor cells [92,115,117,166]. SOX3 is also expressed in progenitor cells outside of the nervous system, including the postnatal testis. However, the role of SOX3 in stem/progenitor cell maintenance in these tissues is less well understood.

Spermatogenesis is the fundamental biological process required for the generation of sperm from progenitor cells via mitosis, meiosis, and a complex program of cellular differentiation. Importantly, in mammals, as in many other animals, sustained spermatogenesis in the adult is dependent on a resident population of germline cells with self-renewal potential. In the mouse testis, this stem cell activity is contained within a heterogeneous population of germ cells known as undifferentiated spermatogonia that develop from gonocytes (foetal germ cells) during the first week of postnatal development. The undifferentiated pool is located in the basal layer of the seminiferous tubules, and is composed of cells of distinct topologies; isolated type A-single spermatogonia (A_s) and interconnected chains of 2 or more cells formed from incomplete cytokinesis during cell division referred to as A-paired (A_{pr}) and A-aligned (A_{al}) spermatogonia, respectively[142]. Upon commitment to differentiate, cells convert to type A1 spermatogonia, which then undergo a series of rapid mitotic divisions prior to meiosis and sperm formation. Besides having distinct cell division kinetics, differentiating spermatogonia can be distinguished from undifferentiated cells by expression of the receptor tyrosine kinase c-KIT plus DNA methyltransferases 3A and 3B (DNMT3A/DNMT3B) [156,167].

All cells within the undifferentiated pool may possess self-renewal potential [168]. However, only a small subset of this population act as stem cells in the steady state tissue, with a majority of undifferentiated cells being primed to differentiate and therefore acting as committed progenitor/transit-amplifying cells [169]. The fate tendencies of undifferentiated cells correlate with gene expression patterns and

chain length. Specifically, steady-state stem cells express *Gfra1*, encoding a co-receptor for the key niche-derived growth factor glial cell line derived neurotrophic factor (GDNF) and exist primarily as A_s, A_{pr} and some short-chained A_{al} cells [153]. A primitive subset of GFRα1+ A_s and A_{pr} spermatogonia with potent stem cell activity and marked by transcription factors EOMES and PDX1 has also recently been described [170,171]. In contrast, the majority of A_{al} spermatogonia express *Ngn3* and *Rarg* and are usually differentiation-committed [161,162,172,173]. Interestingly, lineage-tracing studies have demonstrated that a small fraction of the NGN3+ population is still capable of forming stable long-lived clones within the testis [161]. Moreover, NGN3+ A_{al} cells occasionally fragment to shorter chains plus A_s cells and may revert gene expression patterns to a GFRα1+ state, demonstrating the dynamic nature of the stem cell pool [159,171]. This limited contribution of NGN3+ cells to the steady-state self-renewing pool is also enhanced under conditions of tissue regeneration [161]. However, in contrast to GFRα1+ spermatogonia, NGN3/RARγ+ undifferentiated cells are sensitive to retinoic acid, a key endogenous differentiation stimulus, which promotes a differentiation-committed fate [162].

As transition from the GFRα1+ to NGN3+ state switches the predominant fate of undifferentiated cells from self-renewal to differentiation, it must be tightly regulated to ensure tissue homeostasis. A limited number of factors have been directly implicated in regulation of this transition. For instance, the SOHLH1/2 transcription factors and mTORC1-signalling pathway promote exit from a GFRα1+ state while the NANOS2 RNA binding protein prevents the GFRα1+ to NGN3+ transition *via* direct inhibition of both *Sohlh2* mRNA translation and mTORC1 activation [173-177]. Despite the importance of such factors and pathways in fate transitions within the undifferentiated pool, the relevant downstream effectors remain poorly characterised.

Sox3 is one of a number of identified target genes of SOHLH1/2 within the testis and is reported to play a role in spermatogenesis, whereby *Sox3* deletion causes a block in spermatogenesis that is most severe in mice bred on the C57Bl/6 genetic

background [108,109,175]. However, the exact nature of this spermatogenic block and the underlying molecular mechanisms are not fully understood. Through use of a *Sox3*-GFP knock-in mouse model we now confirm that *Sox3* is specifically expressed within the committed progenitor fraction of the undifferentiated pool and we define its critical role in the GFRα1+ to NGN3+ spermatogonial transition.

4.3 RESULTS

4.3.1 SOX3 EXPRESSION IS RESTRICTED TO COMMITTED SPERMATOGONIAL PROGENITOR CELLS

Previous studies have shown that *Sox3* expression in the testis is restricted to spermatogonial populations within the basal layer of the seminiferous tubules [92,108,109]. Through use of a SOX3-specific antibody [93], we confirmed by immunofluorescence (IF) analysis that SOX3 protein is restricted to spermatogonia of wild type (WT) adult testis (Fig. 1A). Spermatogenesis occurs in a coordinated, cyclic process that can be divided into 12 stages in the mouse. Sections of tubules at a given stage contain defined populations of spermatogonia, spermatocytes and developing spermatids at distinct differentiation and maturation steps [142,178]. Importantly, SOX3-positive spermatogonia were present at all stages of the seminiferous epithelium cycle, indicating that *Sox3* is expressed in the undifferentiated population (Fig. 1B). Moreover, spermatogonia expressing *Sox3* displayed low levels of Cyclin D1 (*Ccnd1*), a marker predominantly expressed by differentiating spermatogonia (Fig. 1A) [155,179]. In contrast to the stage-independent presence of SOX3-positive cells, spermatogonia expressing CCND1 were predominantly found in stage I-VI tubules in which populations of differentiating spermatogonia are the most abundant (Fig. 1B).

In order to better define the identity of the SOX3-positive population we next performed IF analysis of prepubertal, WT postnatal day 7 (P7) testis, which is relatively enriched in undifferentiated spermatogonia. Comparison of *Sox3* expression with that of *promyelocytic leukaemia zinc finger* (*Plzf*), which is expressed throughout the undifferentiated pool [154,180], revealed extensive overlap (Fig. 1C,D). Analysis of adult and juvenile testis therefore suggested that SOX3 is broadly expressed in the undifferentiated population. To investigate this further, we compared SOX3 expression with that of GFRα1, a marker for the self-renewing fraction of undifferentiated spermatogonia [153]. In comparison to the pronounced co-expression of SOX3 with PLZF, we found a limited overlap in expression of SOX3 and GFRα1 (Fig. 1E), suggesting that within the undifferentiated

population, SOX3 is preferentially expressed in the committed progenitor fraction [175].

To further characterise the expression of SOX3 in spermatogonia, we also performed IF analysis on cultures of undifferentiated spermatogonia derived from juvenile WT mice [180]. SOX3 expression was compared to that of PLZF and the transcription factor SALL4, which is expressed in undifferentiated and differentiating spermatogonia [181]. Consistent with previous studies, PLZF and SALL4 were expressed by essentially all cells within the colonies of cultured cells (Fig. 1F)[181]. In contrast, SOX3 expression was clearly heterogeneous. Given that these cultures are considered to contain a mix of both stem and committed progenitor cells[182], our data support the conclusion that SOX3 expression is restricted to a subset of undifferentiated cells, likely representing the committed progenitor fraction, consistent with previous reports [92,108,109].

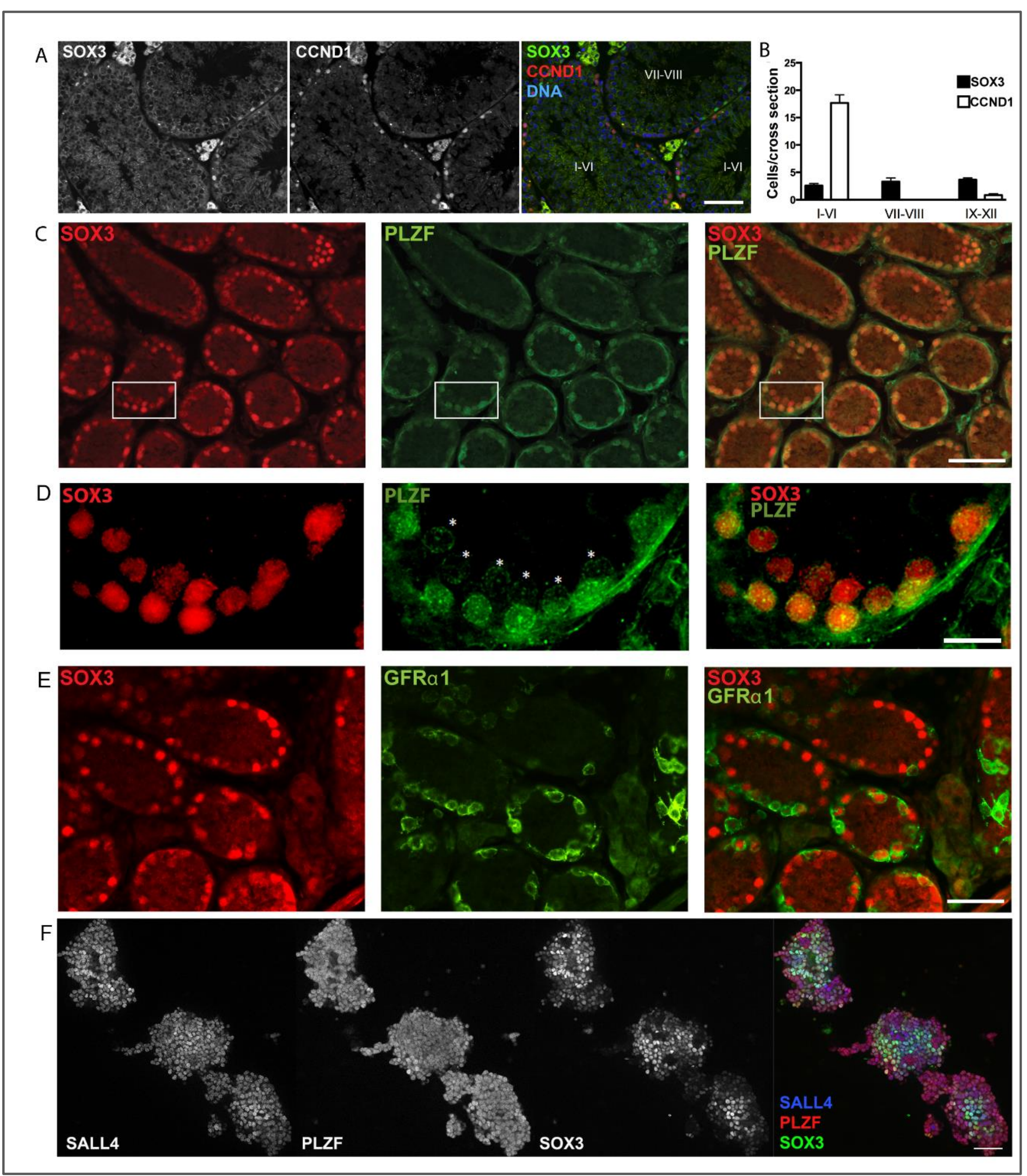


Figure 1. SOX3 expression is found within a subset of undifferentiated spermatogonia. (A) IF of adult testis cross sections showing expression of SOX3 is restricted to spermatogonia. Note the staining in Leydig cells is non-specific. (B) The number of SOX3+ spermatogonia remains constant, independent of the stage of spermatogenesis, whereas CCDN1+ spermatogonia are more prevalent in stage I-VI tubules. (C) IF of SOX3 and PLZF in P7 testis cross sections demonstrates an extensive overlap, with only a few SOX3-/PLZF+ spermatogonia present. (D) 60X magnification of white box in (C) highlighting variations in PLZF expression from

low (*) to high. (E) IF of SOX3 and GFRα1 in P7 testis cross-sections showing limited overlap of expression. (F) IF of SOX3, PLZF and SALL4 in cultures of undifferentiated spermatogonia derived from juvenile mice, PLZF and SALL4 are detected within essentially all cells, while SOX3 is only present within a subset of cells. Scale bars are 50μm. (A,C,E&F), 5um (D)

Notably, from IF analysis of P7 testis sections, we confirmed that essentially all SOX3+ cells were positive for spermatogonial marker SALL4 but that *Sox3* expression was restricted to a subset of the SALL4+ population (32.7 ± 8.0% of SALL4+ cells were SOX3+, mean ± s.e.m, *n* = 3 mice, >70 tubule cross-sections scored per sample) (Figure 2A). In contrast, no overlap in expression was found between SOX3 and EOMES, a marker of the most primitive cells within the GFRα1+ self-renewing fraction (Figure 2B) [170,171]. Combined, our results indicate that *Sox3* expression delineates a subset of differentiation committed spermatogonia.

Finally, to confirm the expression pattern of SOX3 within mature stem and progenitor spermatogonial populations, we compared its expression with that of a number of distinct spermatogonial markers by wholemount analysis of adult seminiferous tubules (Fig. 3). Importantly, IF for GFRα1 and SOX3 in adult tubules demonstrated that expression of these genes largely marked distinct spermatogonial populations, in agreement with our previous results (Fig. 1D). GFRα1 primarily labelled A_s and A_{pr} cells while SOX3 was present in GFRα1-negative SALL4-positive A_{al} (Fig. 3A). Antibodies to GFRα1 and SOX3 used in wholemount analysis are both raised in goat and are visualized in the same fluorescence channel but can be distinguished by cell membrane vs. nuclear staining pattern (Fig. 3A and S1). Rather, we found that SOX3 was co-expressed in chains of undifferentiated spermatogonia plus cells at early differentiation stages with genes that mark committed progenitor fractions, including RARγ and LIN28A (Fig. 3B,C and data not shown) [162,183]. In contrast, chains of c-KIT+ spermatogonia at mid-to-late stages of differentiation (A_3 to B) expressed low or undetectable levels of SOX3 (Fig. 3D) [155,156,167,171].

Taken together, our data demonstrate that SOX3 expression is selectively upregulated upon transition from the self-renewing to the differentiation-primed progenitor state within the undifferentiated population of both juvenile and adult testes. It is subsequently downregulated as they differentiate further.

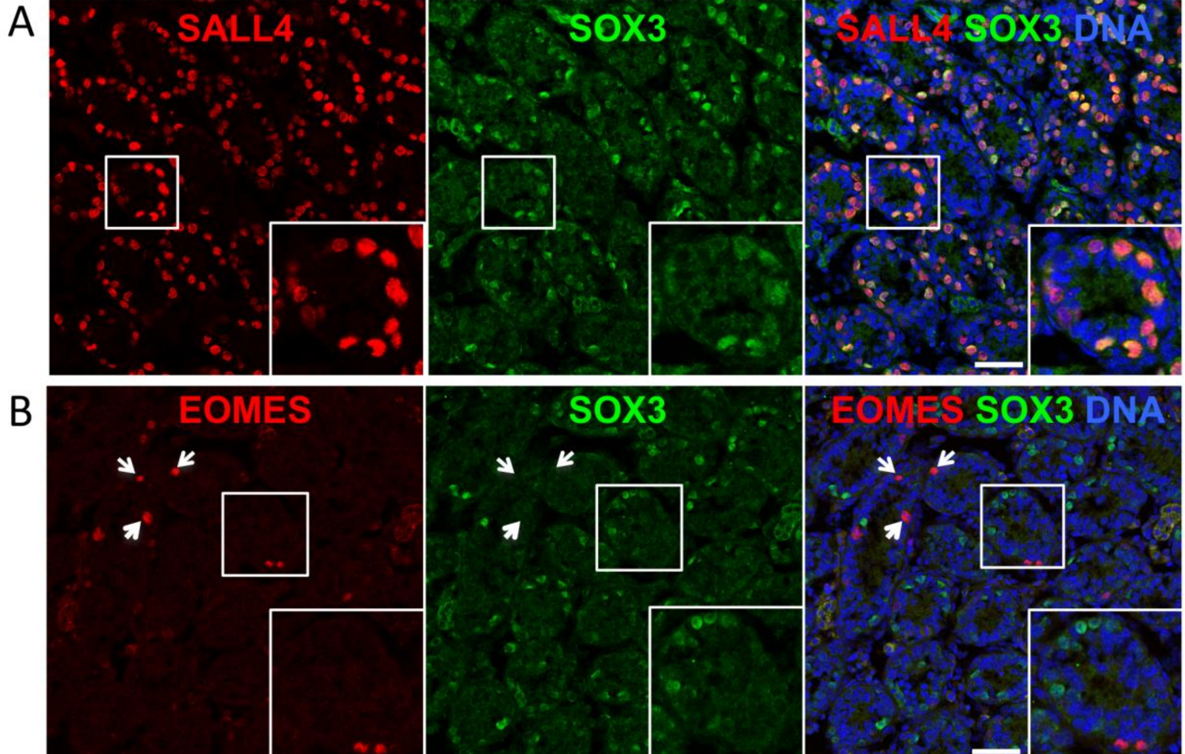


Figure 2. Characterising the expression of SOX3 in juvenile testis. (A,B) Representative IF analysis of SOX3 and markers of spermatogonia (SALL4) and primitive self-renewing cells (EOMES) in testis sections from P7 mice (n = 3 mice). Insets show higher magnification details of indicated regions. Arrowheads indicate EOMES+ SOX3- spermatogonia. Scale bars are 50µm.

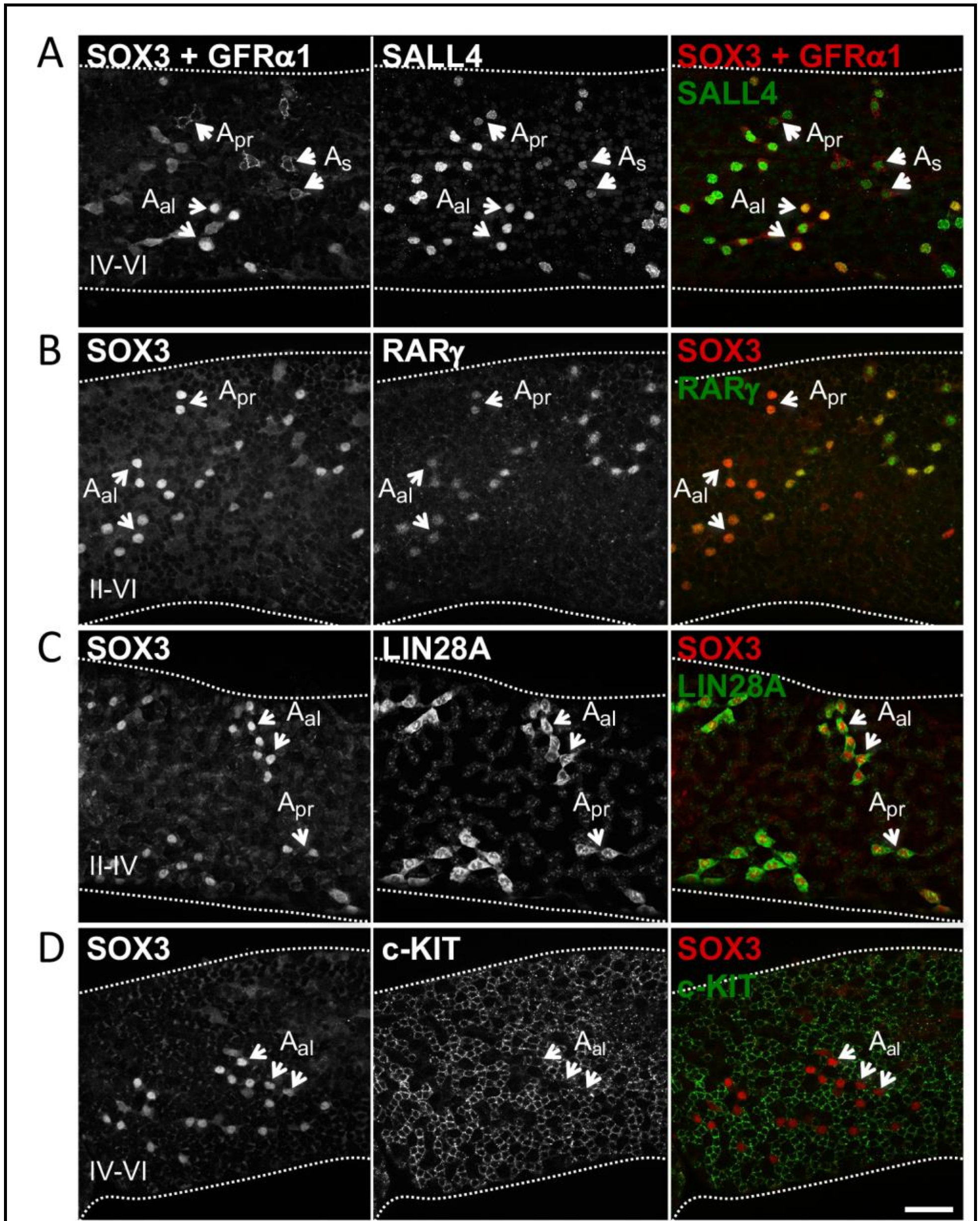


Figure 3. Characterising the expression of SOX3 in adult testis. Assessing co-expression of SOX3 with a panel of undifferentiated and differentiated spermatogonia cell markers in adult seminiferous tubules by wholemount IF. SOX3 shows limited overlap in expression with stem cell-associated marker GFRα1 (A) but is co-expressed with progenitor markers RARγ (B) and LIN28A (C) within the undifferentiated pool. SOX3 is expressed at low levels or is undetectable in mid-late stage differentiating spermatogonia marked with c-KIT (D). Note that antibodies to SOX3 and GFRα1 are both raised in goat and detected in the same fluorescence channel. Cell staining patterns are used to distinguish GFRα1 (cell membrane) and SOX3 (nucleus). Selected undifferentiated cells and stages of tubules are indicated. Representative images are shown (n = 4 mice). Scale bars are 50μm.

4.3.2 EARLY POSTNATAL DEFECT IN 129/SVJ *SOX3* NULL TESTES

Spermatogenesis phenotypes have been reported in *Sox3* null mice on mixed and C57BL/6 inbred backgrounds [92,107,109]. To investigate whether spermatogenesis is also affected on a 129/Svj inbred background, *Sox3* null mice were backcrossed and examined for testis abnormalities. No significant difference in testis weight or morphology (Fig. S2A,B) was detected in *Sox3* null mice at postnatal day 7 (P7; Fig. S2 A,B; WT = 5.3 ± 0.2mg; KO = 5.1 ± 0.4mg). Histological analysis also failed to reveal any gross defects at this stage (Fig. S3). In contrast, *Sox3* null testes were significantly smaller than WT testis at P14, P21 (Fig. S2C-F) and at 6 months of age. Histological analysis revealed an abundance of empty tubules in *Sox3* null testes consistent with previous reports [92,107,109] (Fig. S3E,F). To address whether elevated programmed cell death might account for testis hypoplasia in *Sox3* null mice, we performed TUNEL staining on testis sections. No difference in the number of apoptotic cells was observed between WT or *Sox3* null testis at P7 (Fig. S3C,D).

To further investigate the cellular mechanism that underpins the spermatogenic block in *Sox3* null testes, we compared expression of spermatogonial cell markers in WT and *Sox3* null testis at P7. We reasoned that although *Sox3* null testes were not hypoplastic at this stage, the primary molecular/cellular defect(s) underpinning the defect would likely be present, given that *Sox3* is expressed at this time point (Fig. 1). Expression of *Oct4*/ *Pou5f1*, *Plzf*, *Id4* and *Ecad* were not significantly different, suggesting that the number of Type A spermatogonia is not grossly affected by the absence of *Sox3* [157,180]. Further, expression of these markers was not significantly altered in KO testis at P14, P21 and P28. In contrast, *Ngn3*, a marker of committed undifferentiated spermatogonia, was substantially reduced at P7 (Fig. 4A) and remained significantly lower throughout postnatal testis development (P14, P21 and P28) (Fig. 4B, C and D), as observed previously in B6 KO mice [108]. Conversely, expression of *Gfra1*, a marker for the self-renewing fraction of undifferentiated spermatogonia, was significantly higher in *Sox3* null testis compared to WT (Fig. 4). *Gfra1* expression remained significantly elevated at P14 and P21 although it had apparently normalised by postnatal week 4 (Fig. 4B, C and D). Interestingly, expression of *Id4*, a marker associated with transplantable stem cell activity that is detected in a subset of the GFRα1-positive population in adults,

but is more broadly expressed by spermatogonia during postnatal development, was not significantly altered by loss of *Sox3* (Fig. 4A-D) [184]. While the significance of unaltered *Id4* expression is unclear, the increase in *Gfra1* and reduction in *Ngn3* expression in response to *Sox3* deletion suggests a relative increase in self-renewing spermatogonial subsets and a corresponding decrease in differentiation-primed progenitors.

To confirm these observations in the mature spermatogonial pool, we isolated undifferentiated cells (EpCAM+ α6-integrin+ c-KIT–) from wildtype control and *Sox3* KO adults and analysed gene expression by RT-qPCR (Fig. 5) [185]. Within the EpCAM+ germ cell fraction of *Sox3* KO testis a significantly greater proportion of cells were in the α6-integrin+ c-KIT– undifferentiated cell gate compared to controls (Fig. 5B,C), suggesting that spermatogonial differentiation was disrupted in the *Sox3* KO and undifferentiated cells accumulated with age. In agreement with analysis of total testis extracts during postnatal development, *Sox3* KO adult undifferentiated spermatogonia exhibited significantly lower expression of *Ngn3* while other stem and progenitor-associated markers were mostly comparable to controls. Expression of *Gfra1* was increased in *Sox3* KO undifferentiated cells, although not significantly so ($P = 0.0569$), consistent with previous results (Fig. 4 and 5D). Combined, our results support the involvement of SOX3 in stem-to-progenitor transition in the male germline and specifically link SOX3 with expression of *Ngn3*, a marker of differentiation-primed undifferentiated spermatogonia [159].

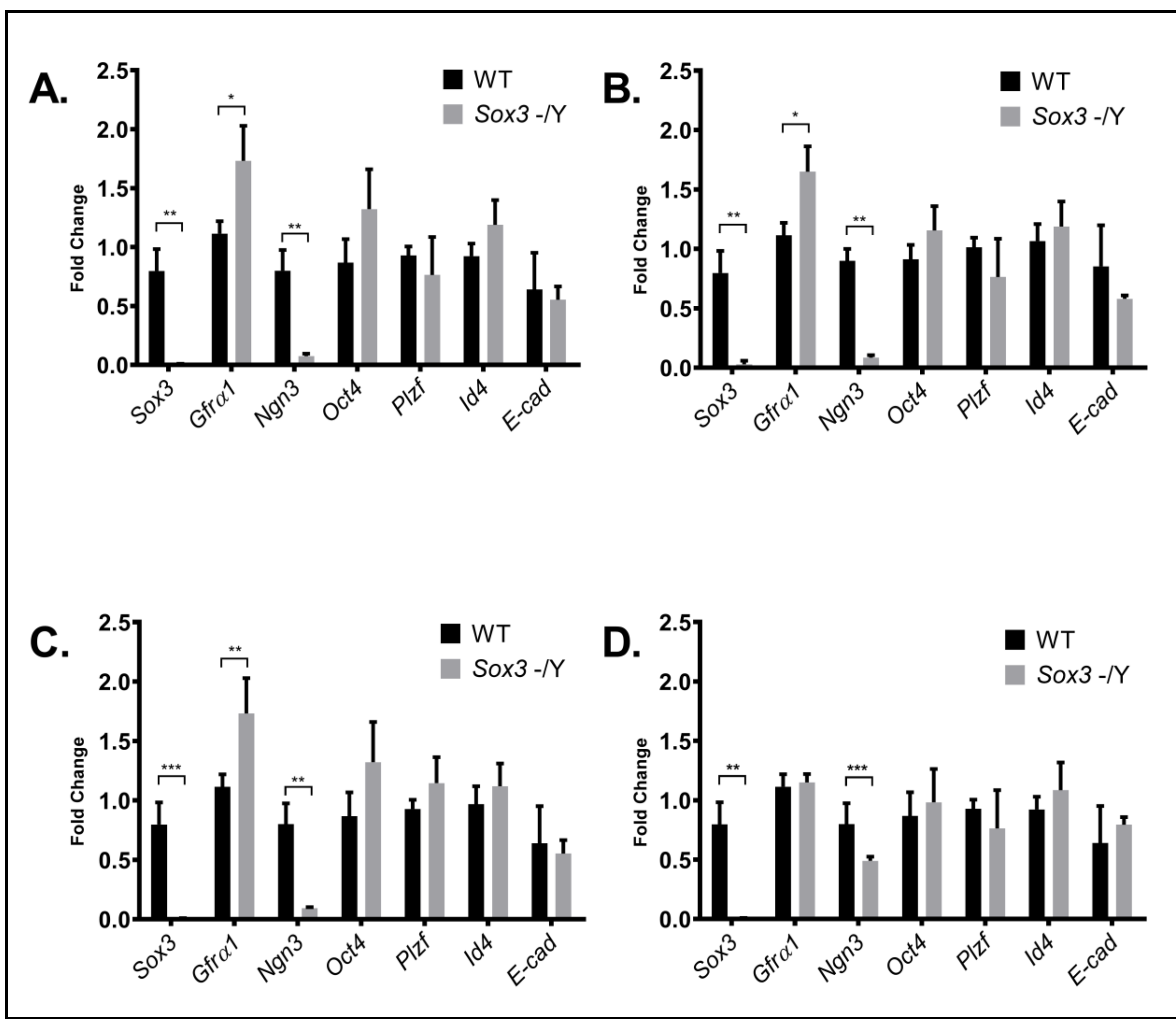


Figure 4. Comparative analysis of *Sox3*$^{-/Y}$ and WT testis gene expression. Expression analysis of spermatogonia cell markers *Sox3*, *Gfra1*, *Ngn3*, *Oct4*, *Plzf*, *Id4* and *E-Cad* in P7 (A), P14 (B), P21(C) and P28 (D) testes. n = 4 for each genotype, $P<0.05$ (*), $P<0.01$ (**), $P<0.001$ (***) as determined by t-test. Data represented as mean±s.d.

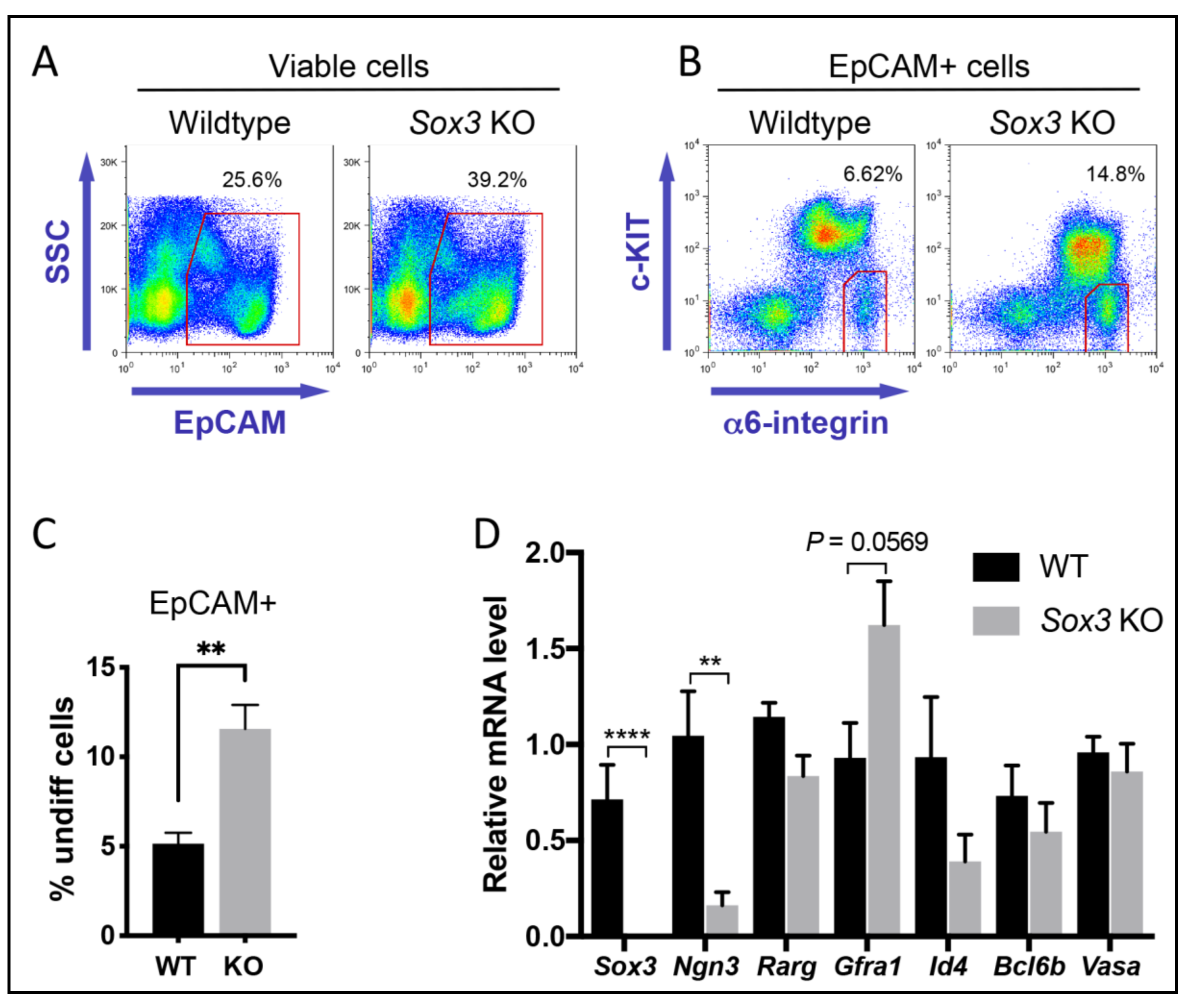


Figure 5. Isolation and analysis of *Sox3* null adult undifferentiated spermatogonia. (A,B) Flow-sorting strategy for isolation of undifferentiated spermatogonia (EpCAM+ c-KIT− α6-integrin+) for gene expression analysis. Percentages of cells within gates are indicated. (C) Graph shows percentage of EpCAM+ germ cells within c-KIT− α6-integrin+ undifferentiated cell gate from flow cytometry analysis of wildtype and *Sox3* null adult testis of A and B. Mean values ± s.e.m. are shown (n = 5 wildtype and n = 6 *Sox3* null mice). Significance was determined by two-tailed unpaired *t*-test, $P < 0.01$ (**). (D) RT-qPCR analysis of sorted undifferentiated spermatogonia (EpCAM+ c-KIT− α6-integrin+) from wildtype and *Sox3* null adult testis. Expression levels are corrected to β-actin and normalized to a control sample. Mean values ± s.e.m. shown (n = 5 wildtype and n = 6 *Sox3* null mice). Significance was determined by two-tailed unpaired *t*-test and selected P values are indicated, $P < 0.01$ (**), $P < 0.0001$ (****).

4.3.3 SOX3 promotes exit from the GFRα1-positive spermatogonial state

T To further investigate the cellular mechanism that underpins elevated expression of *Gfra1* in *Sox3* null testis, we first performed IF analysis of P7 testis sections. Importantly, we observed a significant increase in the total number of GFRα1 positive cells per tubule cross-section in *Sox3* null testis compared to WT testis (mean of 12.3 ± 0.28 versus 6.2 ± 0.19 cells/tubule, $P < 0.0001$; Fig. 6A). This relative expansion of the GFRα1-positive population may suggest that *Sox3* deficiency resulted in a differentiation block at the stem-to-progenitor cell transition or increased self-renewing divisions of *Gfra1*-expressing cells. Alternatively, given the dynamics of cell transitions within the undifferentiated compartment [159], it might represent an inability to maintain a stable progenitor population and increased reversion of committed cells to a self-renewing state.

To investigate these distinct possibilities, we took advantage of a unique feature of our *Sox3*-deficient model. Namely, that *Sox3* deletion is accompanied by insertion of a GFP reporter gene into the *Sox3* locus [92]. Analysis of GFP expression in *Sox3* null testis therefore allows identification of committed progenitor populations that are usually marked by *Sox3* expression in the WT setting. From WT and *Sox3* null testis we therefore scored the number of uncommitted stem cells (GFRα1+/SOX3- and GFRα1+/GFP- respectively), transitional/early committed progenitor cells (GFRα1+/SOX3+ and GFRα1+/GFP+) and committed progenitor cells (GFRα1-/SOX3+ and GFRα1-/GFP+) (Fig. 6B). Importantly, no significant difference was detected in the total number of cells expressing these distinct marker combinations. However, significant differences were detected in the relative abundance of two subpopulations. Firstly, the number of cells that exhibited a committed progenitor phenotype (GFRα1-/GFP+ *vs*. GFRα1-/SOX3+) was significantly reduced in *Sox3* null testes as compared to WT controls. Secondly, there was a significant increase in the number of uncommitted (GFRα1+/GFP- *vs*. GFRα1+/SOX3-) stem cells in the *Sox3* null testes. While populations of transitional cells (GFRα1+/GFP+ *vs*. GFRα1+/SOX3+) were not significantly altered upon loss of *Sox3* at this early postnatal age, transitional stem-progenitor cells were found to accumulate in *Sox3* null adults (see below). Our results indicate that while the total population of stem and progenitor spermatogonia is relatively unaffected in the juvenile *Sox3* null

testis, there is a shift in balance from stem to committed progenitor cell phenotype. This suggests that *Sox3* promotes the stable transition from stem to progenitor cell states and is consistent with changes in *Gfra1* and *Ngn3* expression in *Sox3* null testis (Figs. 4 and 5).

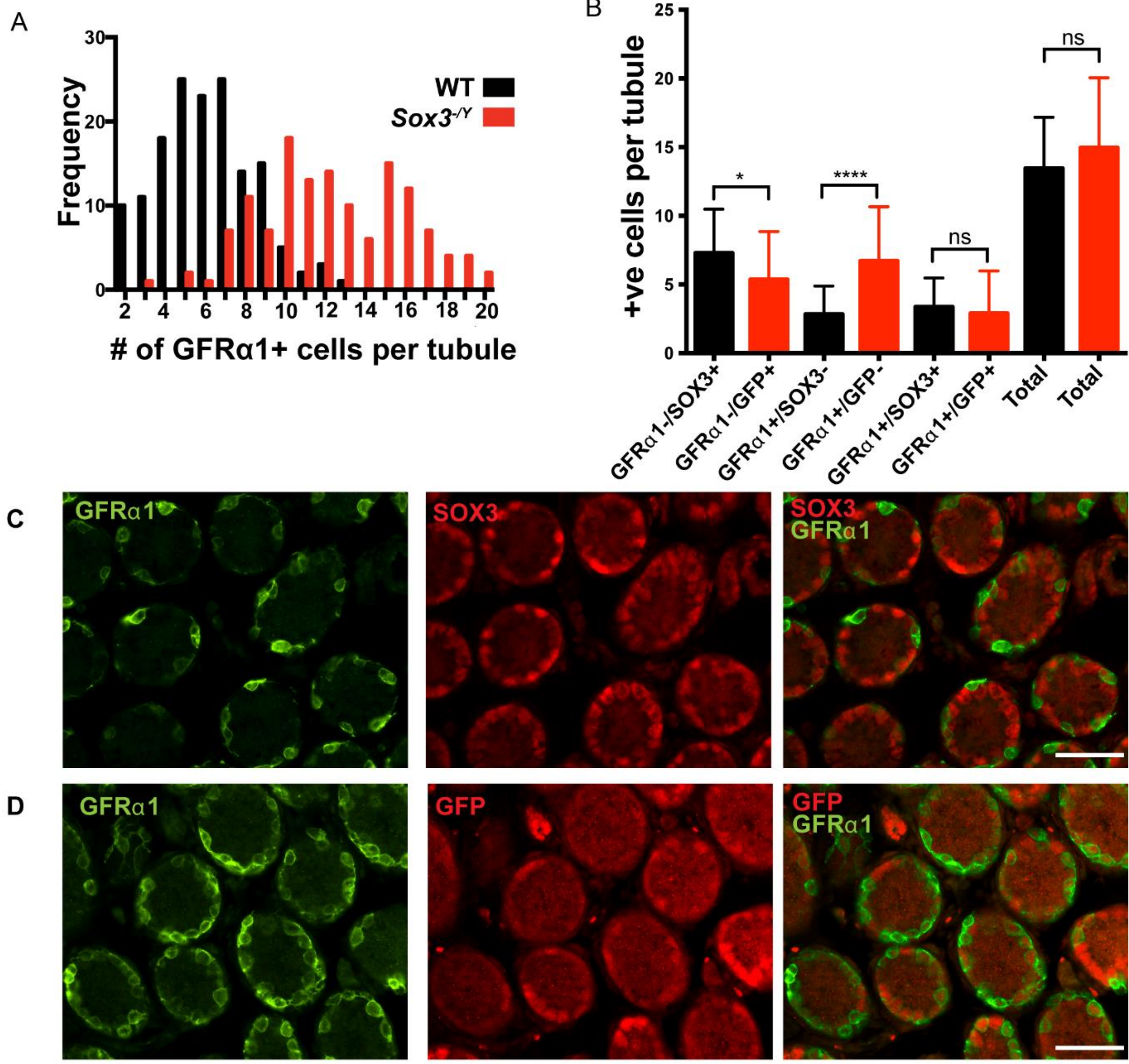


Figure 6. Increased number of GFRα1+ spermatogonia in *Sox3*$^{-/Y}$ testis. (A) Frequency plot showing more GFRα1+ cells per tubule in *Sox3* null testis cross-sections (12.3+/- 0.28) compared to WT (6.2+/-0.19), P<0.0001. (B) Cell counts of the different spermatogonia cell identities as marked by SOX3/GFP and or GFRα1 staining. IF of WT testis cross sections at P7 (C), and *Sox3* null testis sections (D), staining for SOX3 (red) and GFRα1 (green). Minimum 10 fields of view per counted section, 3 sections per testis, n=3 for each genotype, not significant (ns), P<0.05 (*), P<0.0001(****) as determined by Mann-Whitney test. Data represented as mean±s.d., scale bar is 50μm (C,D).

To further investigate this phenotype and the consequences of inefficient stem to progenitor cell conversion on the spermatogenic pathway, we analysed *Sox3* deficient adult testis by whole mount IF of seminiferous tubules (Fig. 7A). In this analysis, SALL4 was used as a marker of both undifferentiated and differentiating spermatogonia, while expression of DNMT3B identified differentiating spermatogonia [167,181]. The distinct stages of the seminiferous tubule areas were then defined according to the differentiation stage of resident spermatogonial populations as described [178,186]. Consistent with analysis of juvenile testis sections, the basal layer of adult *Sox3* null tubules demonstrated pronounced increases in the density of GFRα1-positive cells throughout the seminiferous epithelium cycle (Fig. 7A). Notably, while in WT tubules GFRα1-positive cells were most frequently present as A_s and A_{pr}, in *Sox3* null testis they were often found as A_{al} chains of 4 or more cells. The increased GFRα1-positive chain length in *Sox3* null testis was accompanied by evidence of disruption to the spermatogonial differentiation pathway. Specifically, at stages II-V, when WT tubules contained abundant populations of GFRα1-/SALL4+/DNMT3B- A_{al} cells; *Sox3* deficient A_{al} often abnormally retained *Gfra1* expression (Fig. 7A). Moreover, between stages V-XI, during which essentially all WT A_{al} cells initiate differentiation and generate GFRα1-/SALL4+/DNMT3B+ differentiating A-type spermatogonia, a substantial fraction of A_{al} cells in the *Sox3* deficient testis persisted in a GFRα1+/SALL4+/DNMT3B- undifferentiated state (Fig. 7A). Consequently, the density of differentiating spermatogonia at times appeared lower in *Sox3* null tubules as compared to WTs.

This data indicates that *Sox3*-deficient A_{al} cells, while morphologically resembling differentiation-primed undifferentiated spermatogonia do not appropriately down-regulate GFRα1, a key functional marker of the stem cell state [187]. Moreover, these GFRα1+ A_{al} cells subsequently enter the differentiation pathway less efficiently than their WT GFRα1– A_{al} counterparts. Consistent with our observations, the majority of GFRα1+ A_{al} that accumulated in *Sox3* null adults did not express RARγ (Fig. 7B), a marker of committed progenitors required for differentiation commitment (Gely-Pernot *et al.*, 2012; Ikami *et al.*, 2015). Interestingly, however, a minor subset of GFRα1+ A_{al} cells in the knockout did express RARγ, indicating an accumulation of transitional undifferentiated cells exhibiting both stem and progenitor

characteristics. Such a stem-progenitor transitional state is rarely detected in WT adult testis, as is presumably short-lived but persists in the absence of *Sox3,* as undifferentiated cells appear unable to adopt a stable committed progenitor state (Fig. 7B). Note that RARγ+ GFRα1– progenitors were still found in *Sox3* null testis indicating that stem to progenitor conversion was not completely abrogated, in agreement with our analysis of juvenile testis (Figs. 6B and 7B).

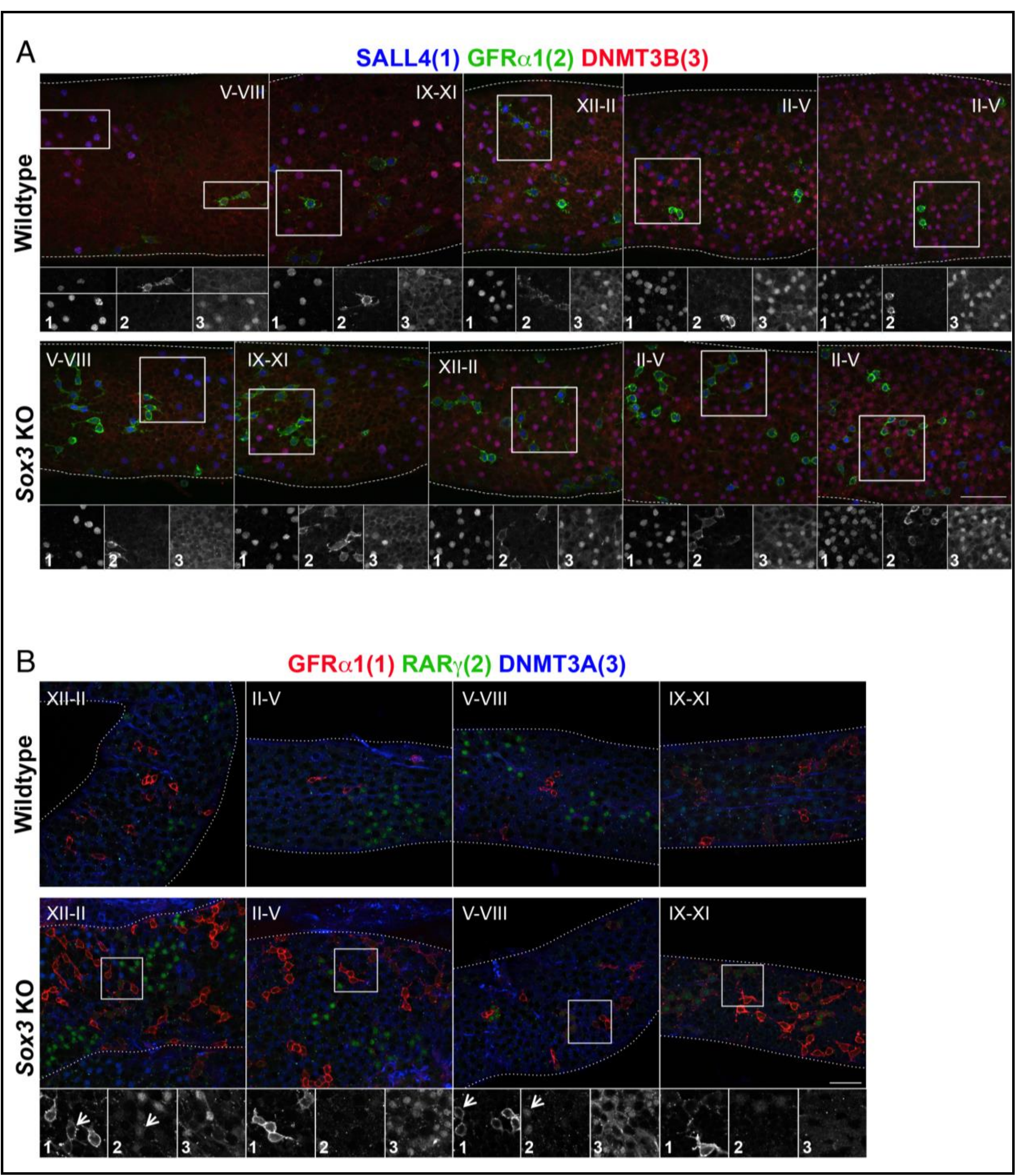
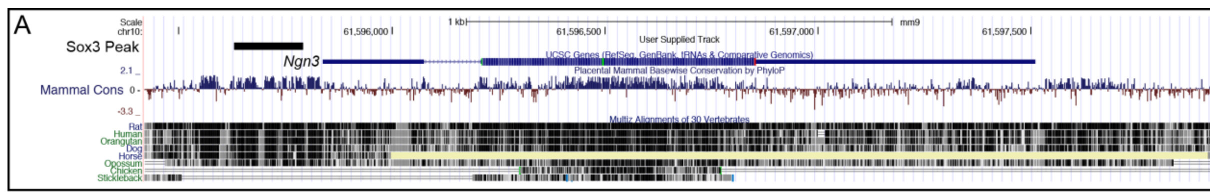


Figure 7. Analysis of *Sox3* null adult seminiferous tubules by whole mount IF. Wholemount IF analysis of WT and *Sox3* knockout adult seminiferous tubules for SALL4, DNMT3B, GFRα1 (A) and GFRα1, RARγ, DNMT3A (B) at the indicated spermatogenic stages. Grayscale panels show individual immunostaining within the indicated area. Arrowheads in (B) indicate GFRα1+/RARγ+ transitional cells found relatively frequently in *Sox3* knockout tubules but rarely in WT samples. Scale bars are 50μm.

4.3.4 NGN3 IS A DIRECT TARGET OF SOX3

Given the near complete failure to activate *Ngn3* expression in *Sox3* null postnatal testes, we investigated the possibility that *Ngn3* can be a relevant target of SOX3 in germline cells. Previous ChIP-Seq studies of neural progenitors[55,188] identified a SOX3 binding peak and single SOX consensus motif (SOCM) approximately 500bp upstream of the transcriptional start site of *Ngn3* (Fig. 8A). This genomic region is highly conserved in mammals suggesting it may function as an enhancer. To investigate SOX3 binding to this region, we performed SOX3 ChIP-PCR on P7 mouse testis. Significant enrichment of SOX3 binding to this region was detected, suggesting that *Ngn3* is directly activated by SOX3 *in vivo* (Fig. 8B). As a positive control, we confirmed that SOX3 in testis cells also binds the SOX3 target gene *Dbx1* previously identified in a neural system[189].



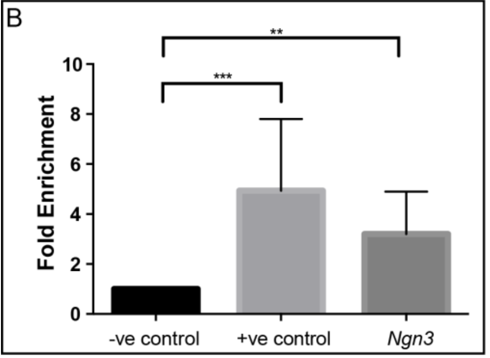


Figure 8. *Ngn3* is a direct SOX3 target gene in spermatogonia. UCSC Genome Browser view of published SOX3 ChIP-seq data[188] from NPCs showing a SOX3 binding site within the upstream promoter region of *Ngn3* (A). ChIP-PCR for SOX3 in P7 testis (B), demonstrating SOX3 binding upstream of the *Ngn3* promoter (*Ngn3*), $P<0.05$ (*), $P<0.01$ (**) as determined by t-test, data represented as mean±s.d. –ve control represents a site within *Notch1* known not to bind SOX3, +ve control represents a site within intron 2 of *Dbx1* previously validated as a binding site of SOX3 [55].

4.4 DISCUSSION

Previous studies have shown that SOX3 is expressed in Type A spermatogonia [92,108,109]; however, these range from stem cells through to committed progenitors and overtly differentiating cells and it was unclear how the expression and function of SOX3 related to this progression. Here we show extensive overlap between SOX3 and multiple markers of undifferentiated spermatogonia (PLZF, LIN28A and SALL4) indicating that *Sox3* is expressed within this compartment. Comparison of SOX3 with spermatogonial stem cell markers GFRα1 and EOMES revealed limited or no overlap respectively, indicating that the SOX3 negative spermatogonial population are *bona fide* stem cells. Taken together, we conclude that *Sox3* is not expressed by the steady-state stem cell population (GFRα1-positive) but is rapidly upregulated during initial differentiation commitment of these cells. The lack of *Sox3* expression in DNMT3A/B, c-KIT and CCND1 positive A_3 to B spermatogonia indicates that *Sox3* is subsequently downregulated at later stages of differentiation. Thus, *Sox3* expression closely mirrors that of *Ngn3*, a key marker of committed progenitor spermatogonia [190]. Definitive identification of *Ngn3* expressing cells *in vivo* is hampered by the lack of a specific antibody and therefore requires knock-in GFP reporter mice. Therefore, given the availability of commercial specific antibodies, SOX3 provides a useful alternative marker for committed type A spermatogonia.

Within the undifferentiated spermatogonial pool, GFRα1+ cells have a high self-renewal potential while NGN3+ cells are primed to differentiate [153,159]. The distinct fates of these undifferentiated cell subsets are defined by relative sensitivity to retinoic acid (RA), a key endogenous regulator of spermatogonial differentiation. Specifically, NGN3+ cells are responsive to RA due to expression of the RA receptor RARγ, while GFRα1+ cells lack RARγ expression and are consequently unresponsive to RA [162,173]. Combined, our data demonstrate that SOX3 is critically required for formation of GFRα1-/NGN3+ differentiation-primed progenitors from the GFRα1+/NGN3- stem cell pool. Moreover, we find that *Sox3* is specifically

upregulated during this stem-to-progenitor transition and SOX3 directly targets *Ngn3* (see below). This suggests that SOX3 plays a central and cell-autonomous role in promoting exit from a GFRα1+ stem cell state. *Sox3* deletion therefore results in an accumulation of GFRα1+ cells and a substantially reduced capacity to generate RA-sensitive, differentiation-primed GFRα1- progenitors within the undifferentiated pool. Consequently, this leads to a reduction in production of differentiating spermatogonia and ineffective or blocked spermatogenesis (Fig. 9). In adults, loss of *Sox3* was also accompanied by an accumulation of transitional undifferentiated cells expressing both stem and progenitor markers, reflecting inefficient stem to progenitor conversion.

It is also possible that in the absence of SOX3, differentiation-primed GFRα1-negative progenitors are still successfully generated within the undifferentiated pool but are not stably maintained and rapidly revert to a GFRα1+ state, mimicking the cellular dynamics observed during testis regeneration [159]. This possibility could be tested by fate-mapping progenitor populations upon *Sox3* deletion, through use of *Ngn3-CreER* plus reporter lines or by developing an equivalent *Sox3-CreER* transgenic model [161]. Based on current models [159,169], progenitor-to-stem cell reversion upon *Sox3* deletion would be predicted to be accompanied by fragmentation of longer undifferentiated progenitor chains back to A_s and A_{pr} GFRα1+ cells. However, in the *Sox3* null adult, we find GFRα1+ cell chain lengths to be increased compared to WT controls, arguing against this reversion model. Rather, our results suggest that the physiological process of differentiation priming that normally occurs as chain length increases within the undifferentiated pool does not occur in the *Sox3* null testis (Fig. 8). Consequently, a large proportion of the *Sox3*-deficient A_{al} population abnormally retains gene expression patterns and properties of the stem cell state, including being refractory to differentiation stimuli such as RA. This GFRα1+ A_{al} population generally lacks expression of RARγ, a marker of progenitor cells required for RA-dependent spermatogonial differentiation [162,172].

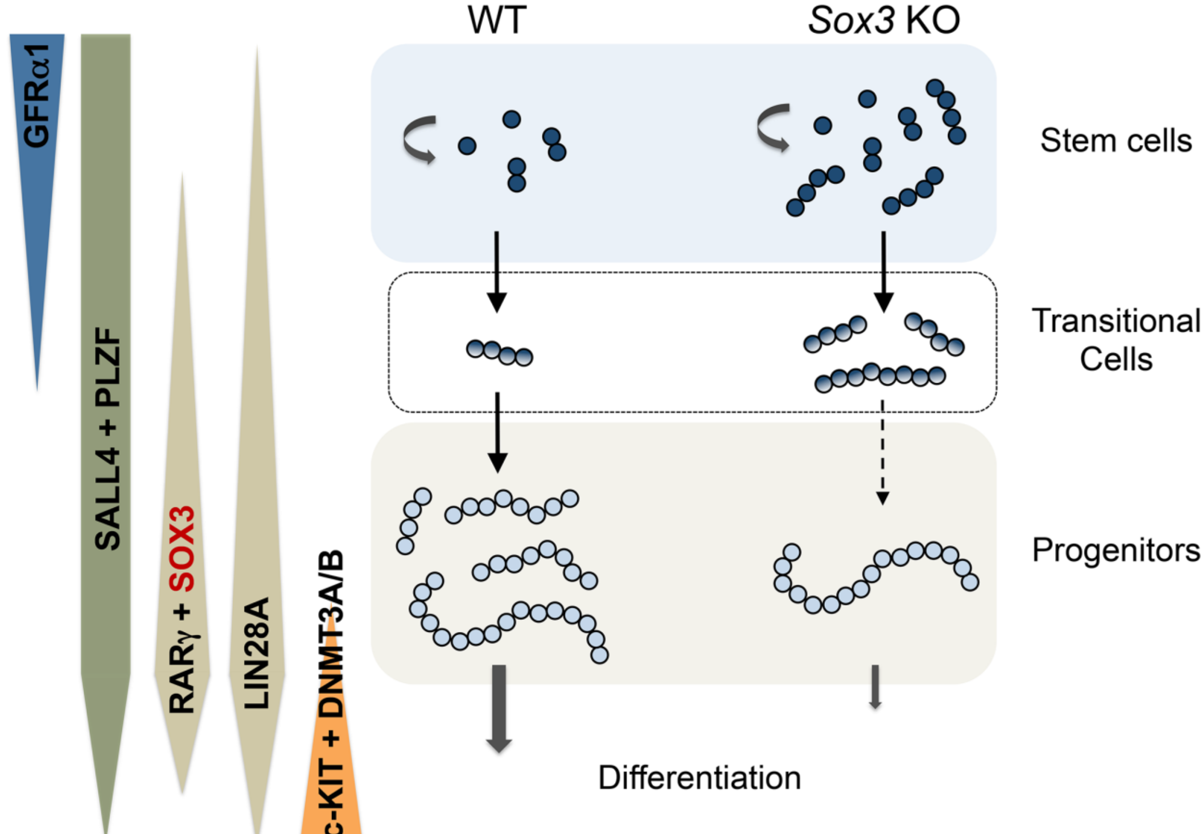


Figure 9. Model of spermatogenesis in WT and *Sox3* KO adult testis. GFRα1 is found within the spermatogonial stem cell pool of WT testis, before being switched off as the cells progress into a progenitor cell state. Conversely, SOX3 is not found within the stem cell population and expression is switched on upon conversion into progenitor cells. SOX3+ progenitor cells differentiate and continue through the spermatogenesis pathway. Additional markers delineating the distinct differentiation stages of spermatogonia are indicated. In *Sox3* knockout (KO) testis, the loss of SOX3 leads to an accumulation of GFRα1+ cells, and a reduction in the capacity to form differentiation-primed GFRα1- progenitor cells. This ultimately leads to a reduction in the production of differentiating spermatogonia. In adult WT testis, GFRα1+ stem cells are mostly present as A_s and A_{pr} while GFRα1+ A_{al} cells are frequently found in *Sox3* null testis. Stem-progenitor transitional A_{al} cells expressing both GFRα1 and RARγ also accumulate in *Sox3* knockouts but are rarely found in WT tubules. We propose that SOX3 plays an essential role in efficient stem to progenitor conversion within the undifferentiated spermatogonial population.

We have provided evidence that *Ngn3* is a direct target gene of SOX3 in spermatogonia. NGN3 is a prototypical marker of the committed progenitor state and knockdown of *Ngn3* expression in cultured undifferentiated spermatogonia disrupts differentiation upon transplantation *in vivo* [159,191]. Therefore, the loss of *Ngn3* expression upon *Sox3* deletion may be responsible for reduced efficiency of the stem cell population to progress to a committed progenitor state. Collectively our data define the window in which SOX3 acts during spermatogenesis, as well as one critical direct target gene, *Ngn3*. Further studies of both the regulation and the role of SOX3 will help reveal how type A spermatogonia progress from stem cells to committed precursors and the degree to which these processes are shared between spermatogonial and neuronal progenitors.

4.5 MATERIALS AND METHODS

4.5.1 ANIMALS

Generation of the *Sox3* null mice on a mixed genetic background (129 sv/eV, MF1, CBA and C57Bl/6) has been described previously [92]. Mixed background *Sox3* null mice were crossed onto the 129/SvJ genetic background and mice from the F6 generation or from a 129/SvJ, CBA mixed background were used in these studies.

4.5.2 TESTIS HISTOLOGY AND IMMUNOFLUORESCENCE (IF)

Testes were isolated and fixed in 4% formaldehyde overnight, followed by equilibration of tissues in 30% sucrose. Testis were set in OCT compound (Tissue-Tek) and cryosectioned on a Leica CM1900 at 16μm thickness. For IF, sections were permeablised with 1% Triton in PBS for 10 minutes, blocked with 10% horse serum for 1 hour, followed by overnight incubation with primary antibodies overnight at 4°C. Sections were washed with PBS and incubated in secondary antibody solution for 1-2 hours, washed in PBS, mounted in ProLong Gold Anitfade Mountant with DAPI (Life Technologies, P-36931) before being cover slipped. Images were captured on a Nikon Ti-E inverted microscope with NIS software (NIKON) using a 20x objective lens. For histology, sections were dehydrated in an increasing methanol series from 10% to 100%. Sections were stained with haematoxylin and eosin. Whole mount IF of seminiferous tubules was performed as described [171,173].

4.5.3 ANTIBODIES

Antibodies are detailed previously [155,171,173,180,181], including α SOX3 (R&D Systems AF2569), α DNMT3A (Novus Biologicals 64B1446), α DNMT3B (Novus Biologicals 52A1018), α c-KIT (Cell Signaling Technologies #3074), α LIN28A (Cell Signaling Technologies #8641) and α Cyclin D1 (Novus Biologicals SP4).

4.5.4 TUNEL ASSAY

TUNEL assays were performed as per manufacturer's instructions (Roche, 11684795910). Briefly, tissue was permeabilised in 0.1% Triton X-100 for 2 minutes on ice, washed twice with PBS, then incubated with TUNEL reaction mixture for 60 mins

at 37C in a humidified chamber. Tissue was washed 3 times with PBS followed by mounting with ProLong Gold Anitfade Mountant with DAPI.

4.5.5 QPCR

RNA from testis of all mouse ages was extracted by mechanically dissociating tissue in Trizol (Life Technologies) for 10 minutes as per manufacturer's instructions. RNA was run on a 1.5% RNase free agarose gel to assess integrity. cDNA was generated using the High Capacity RNA to cDNA kit (Life Technologies). Expression profiling was performed on four WT and four *Sox3* null testes at each age, using *B-Actin* and *Eif4a* as a reference genes. qRT-PCR was performed using Fast Sybr Green Master Mix (Life Technologies), and run on the ABI 7500 StepOnePlus system, all reactions were completed in triplicate. Isolation and gene expression analysis of adult undifferentiated spermatogonia was performed as described [185].

4.5.6 PRIMER SEQUENCES

qPCR primers used for gene expression analyses were as follows:

Sox3: F 5'-GAACGCATCAGGTGAGAGAAG-3', R 5'-GTCGGAGTGGTGCTCAGG-3'

Gfra1: F 5'-ATCGGGCAGTACACATCTCTG-3', R 5'-TGTGGTTATGTGGCTGGAGG-3'

Plzf: F 5'-CCTGGACAGTTTGCGACTGA-3', R 5'-GCCATGTCCGTGCCAGTAT-3'

Ngn3: F 5'-CCCCAGAGACACAACAACCT-3', R 5'-AGTCACCCACTTCTGCTTCG-3'

Oct4: F 5'-CCCAGGCCGACGTGG-3', R 5'-GATGGTGGTCTGGCTGAACAG-3'

Id4: F 5'-CTCACCCTGCTTTGCTGAGA-3', R 5'-TCACCCTGCTTGTTCACGG-3'

E-Cad: F 5'-TTGCAAGTTCCTGCCATCCT-3', R 5'-CATCATCTGGTGGCAGCAG-3'

β-Actin: F 5'-CTGCCTGACGGCCAGG-3', R 5'-GATTCCATACCCAAGAAGGAAGG-3'

Eif2: F 5'-TGATGGCACTGGCCCCAACAT-3', R 5'-GCGCCCTCCTTAGTAGCCCAC-3'

qPCR primers used for ChIP-PCR analyses were as follows:

Ngn3: F 5'-GAGAGTTGCTGGGACTGAGC-3', R 5'-AGCTGGATTCCGGACAAAG-3'

Dbx1: (+ve control) F 5'-CTTTGGTCTCCACAAGCTTTCT-3',
R 5'-GAATGTGGCCTTTAACAACTCAC-3'

Notch1: (-ve control) F 5'-TGTTGTGCTCCTGAAGAACG-3',
R 5'-GCAACACTTTGGCAGTCTCA-3'

4.5.7 CHIP-PCR

P7 mouse testis were extracted and tissue was mechanically disassociated in 1% formaldehyde for 8 minutes at room temperature, followed by neutralisation with 125mM glycine for 5 minutes at 4°C. Cells were lysed followed by sonication (Bioruptor, Diagenode) for 15 minutes with 30s cycling pulses on ice. SOX3 bound chromatin was immunoprecipitated by a goat polyclonal antibody raised against human SOX3 (R&D systems, AF2569). DNA was recovered after reversing crosslinks and purified by PCR clean-up kit (QIAGEN). Three independent ChIP samples were generated each from the material of 4 testis. Each sample was analysed by qRT-PCR as described above, using 1% input as a reference and ChIP samples from *Sox3* null testis as a negative control.

4.5.8 ETHICS STATEMENT

Animal experiments were subject to approval by the Animal Ethics Committees of the University of Adelaide and Monash University. All studies were conducted within the principles of animal replacement and reduction and experimental refinement. Animals were monitored daily for evidence of illness and, if distressed, were culled immediately by cervical dislocation by an experienced investigator/animal technician. All the experiments were performed in accordance with the approved guidelines and regulations.

4.6 ACKNOWLEDGEMENTS

We thank Monash Micro Imaging Facility for technical assistance. We are especially grateful for the experimental support provided by Eva Chan and Mia De Seram. An Australian Research Council grant awarded to PT and RH supported this research

(#DP140103965) HML is supported by an Australian Government Research Training Program (RTP) Scholarship and JM was supported by a fellowship from the Sigrid Jusélius Foundation. RLB acknowledges support from the Francis Crick Institute. The Australian Regenerative Medicine Institute is supported by grants from the State Government of Victoria and Australian Government.

4.7 COMPETING INTERESTS STATEMENT

The authors have no conflicts of interest to declare.

4.8 AUTHOR CONTRIBUTIONS

D.M. R.H. and P.T. conceived the study. Experiments were performed by D.M., J.M. and H.L. and analysed by D.M., J.M., H.L., J.H., R.L.B, R.H. and P.T.; D.M., R.H. and P.T. wrote the manuscript; and all authors revised the manuscript.

4.9 SUPPLEMENTARY FIGURES

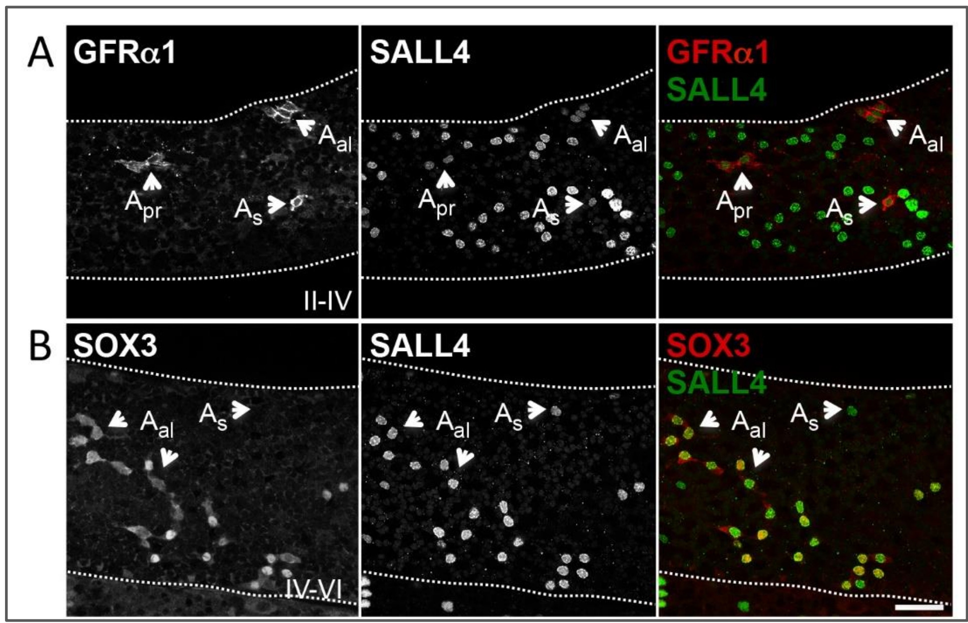


Figure S1. Characterising expression pattern of SOX3 in adult testis. Wholemount IF analysis of adult seminiferous tubules demonstrating distinct subcellular localisation patterns of GFRα1 (A) (cell membrane) and SOX3 (B) (predominantly nuclear) within undifferentiated spermatogonia. Demonstrates single stain antibody staining controls for GFRα1 + SOX3 co-stain within same fluorescence channel (antibodies are both raised in goat). Spermatogonia are counterstained with SALL4. Selected undifferentiated cells and stages of tubules are indicated. Representative images are shown (n = 4 mice). Scale bar, 50μm.

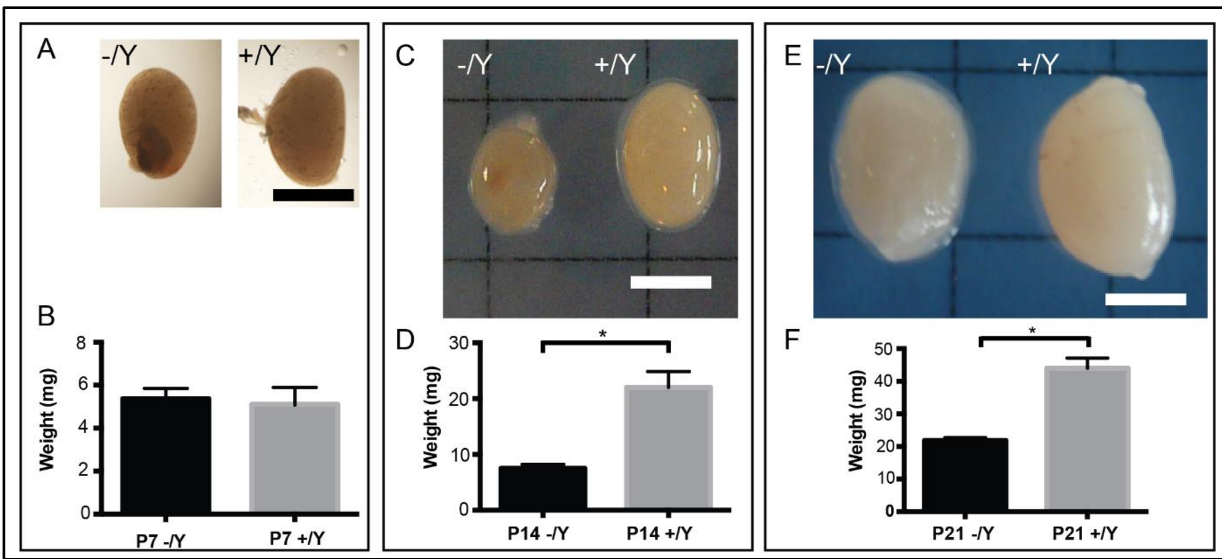


Figure S2. Comparison of whole *Sox3* -/Y testes with WT mouse testes. No difference was observed in either size (A) or weight (B) of *Sox3*-/Y testes at P7. *Sox3*-/Y testes appear smaller at both P14 (C) and P21 (E), and weigh significantly less than WT testes (D & F). Scale bars are 2.5mm.

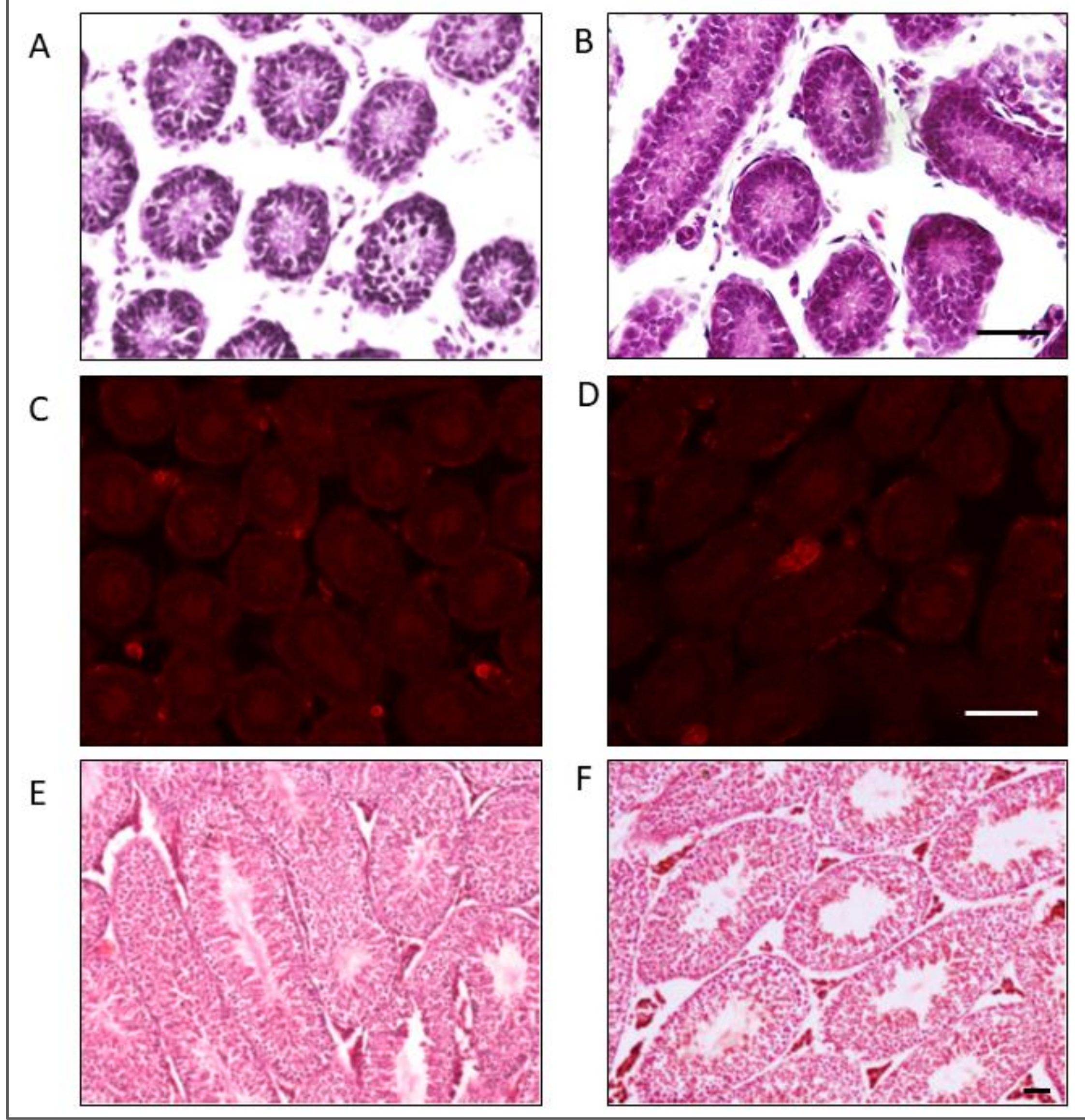


Figure S3. Testes morphology. Representative H&E images of P7 WT (A) and *Sox3* null (B) testes Adult testes, indicating a grossly normal phenotype. TUNEL assay on P7 WT (C) and *Sox3* null (D) testis, showing no difference in amount of cell death. Representative H&E images of adult WT (E) and *Sox3* null (F) testes sections showing *Sox3* null tissue is largely void of mature sperm. Scale bars 50um.

4.10 REFERENCES

55. Bergsland, M.; Ramskold, D.; Zaouter, C.; Klum, S.; Sandberg, R.; Muhr, J. Sequentially acting sox transcription factors in neural lineage development. *Genes & development* **2011**, *25*, 2453-2464.

59. Bylund, M.; Andersson, E.; Novitch, B.G.; Muhr, J. Vertebrate neurogenesis is counteracted by sox1-3 activity. *Nature neuroscience* **2003**, *6*, 1162-1168.

92. Rizzoti, K.; Brunelli, S.; Carmignac, D.; Thomas, P.Q.; Robinson, I.C.; Lovell-Badge, R. Sox3 is required during the formation of the hypothalamo-pituitary axis. *Nature genetics* **2004**, *36*, 247-255.

93. Rogers, N.; Cheah, P.S.; Szarek, E.; Banerjee, K.; Schwartz, J.; Thomas, P. Expression of the murine transcription factor sox3 during embryonic and adult neurogenesis. *Gene expression patterns : GEP* **2013**, *13*, 240-248.

107. Weiss, J.; Meeks, J.J.; Hurley, L.; Raverot, G.; Frassetto, A.; Jameson, J.L. Sox3 is required for gonadal function, but not sex determination, in males and females. *Mol Cell Biol* **2003**, *23*, 8084-8091.

108. Raverot, G.; Weiss, J.; Park, S.Y.; Hurley, L.; Jameson, J.L. Sox3 expression in undifferentiated spermatogonia is required for the progression of spermatogenesis. *Developmental biology* **2005**, *283*, 215-225.

109. Laronda, M.M.; Jameson, J.L. Sox3 functions in a cell-autonomous manner to regulate spermatogonial differentiation in mice. *Endocrinology* **2011**, *152*, 1606-1615.

115. Nishiguchi, S.; Wood, H.; Kondoh, H.; Lovell-Badge, R.; Episkopou, V. Sox1 directly regulates the gamma-crystallin genes and is essential for lens development in mice. *Genes & development* **1998**, *12*, 776-781.

117. Ferri, A.L.; Cavallaro, M.; Braida, D.; Di Cristofano, A.; Canta, A.; Vezzani, A.; Ottolenghi, S.; Pandolfi, P.P.; Sala, M.; DeBiasi, S.*, et al.* Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development* **2004**, *131*, 3805-3819.

142. de Rooij, D.G.; Grootegoed, J.A. Spermatogonial stem cells. *Current opinion in cell biology* **1998**, *10*, 694-701.

153. Hara, K.; Nakagawa, T.; Enomoto, H.; Suzuki, M.; Yamamoto, M.; Simons, B.D.; Yoshida, S. Mouse spermatogenic stem cells continually interconvert between equipotent singly isolated and syncytial states. *Cell stem cell* **2014**, *14*, 658-672.

154. Buaas, F.W.; Kirsh, A.L.; Sharma, M.; McLean, D.J.; Morris, J.L.; Griswold, M.D.; de Rooij, D.G.; Braun, R.E. Plzf is required in adult male germ cells for stem cell self-renewal. *Nature genetics* **2004**, *36*, 647-652.

155. Costoya, J.A.; Hobbs, R.M.; Barna, M.; Cattoretti, G.; Manova, K.; Sukhwani, M.; Orwig, K.E.; Wolgemuth, D.J.; Pandolfi, P.P. Essential role of plzf in maintenance of spermatogonial stem cells. *Nature genetics* **2004**, *36*, 653-659.

156. Schrans-Stassen, B.H.; van de Kant, H.J.; de Rooij, D.G.; van Pelt, A.M. Differential expression of c-kit in mouse undifferentiated and differentiating type a spermatogonia. *Endocrinology* **1999**, *140*, 5894-5900.

157. Tokuda, M.; Kadokawa, Y.; Kurahashi, H.; Marunouchi, T. Cdh1 is a specific marker for undifferentiated spermatogonia in mouse testes. *Biol Reprod* **2007**, *76*, 130-141.

159. Nakagawa, T.; Sharma, M.; Nabeshima, Y.; Braun, R.E.; Yoshida, S. Functional hierarchy and reversibility within the murine spermatogenic stem cell compartment. *Science* **2010**, *328*, 62-67.

161. Nakagawa, T.; Nabeshima, Y.; Yoshida, S. Functional identification of the actual and potential stem cell compartments in mouse spermatogenesis. *Dev Cell* **2007**, *12*, 195-206.

162. Ikami, K.; Tokue, M.; Sugimoto, R.; Noda, C.; Kobayashi, S.; Hara, K.; Yoshida, S. Hierarchical differentiation competence in response to retinoic acid ensures stem cell maintenance during mouse spermatogenesis. *Development* **2015**, *142*, 1582-1592.

164. Kamachi, Y.; Kondoh, H. Sox proteins: Regulators of cell fate specification and differentiation. *Development* **2013**, *140*, 4129-4144.

165. Reiprich, S.; Wegner, M. From cns stem cells to neurons and glia: Sox for everyone. *Cell Tissue Res* **2015**, *359*, 111-124.

166. Favaro, R.; Valotta, M.; Ferri, A.L.; Latorre, E.; Mariani, J.; Giachino, C.; Lancini, C.; Tosetti, V.; Ottolenghi, S.; Taylor, V.*, et al.* Hippocampal development and neural stem cell maintenance require sox2-dependent regulation of shh. *Nature neuroscience* **2009**, *12*, 1248-1256.

167. Shirakawa, T.; Yaman-Deveci, R.; Tomizawa, S.; Kamizato, Y.; Nakajima, K.; Sone, H.; Sato, Y.; Sharif, J.; Yamashita, A.; Takada-Horisawa, Y.*, et al.* An epigenetic switch is crucial for spermatogonia to exit the undifferentiated state toward a kit-positive identity. *Development* **2013**, *140*, 3565-3576.

168. Simons, B.D.; Clevers, H. Strategies for homeostatic stem cell self-renewal in adult tissues. *Cell* **2011**, *145*, 851-862.

169. La, H.M.; Hobbs, R.M. Mechanisms regulating mammalian spermatogenesis and fertility recovery following germ cell depletion. *Cell Mol Life Sci* **2019**, *76*, 4071-4102.

170. Sharma, M.; Srivastava, A.; Fairfield, H.E.; Bergstrom, D.; Flynn, W.F.; Braun, R.E. Identification of eomes-expressing spermatogonial stem cells and their regulation by plzf. *Elife* **2019**, *8*.

171. La, H.M.; Makela, J.A.; Chan, A.L.; Rossello, F.J.; Nefzger, C.M.; Legrand, J.M.D.; De Seram, M.; Polo, J.M.; Hobbs, R.M. Identification of dynamic undifferentiated cell states within the male germline. *Nat Commun* **2018**, *9*, 2819.

172. Gely-Pernot, A.; Raverdeau, M.; Celebi, C.; Dennefeld, C.; Feret, B.; Klopfenstein, M.; Yoshida, S.; Ghyselinck, N.B.; Mark, M. Spermatogonia differentiation requires retinoic acid receptor gamma. *Endocrinology* **2012**, *153*, 438-449.

173. Hobbs, R.M.; La, H.M.; Mäkelä, J.A.; Kobayashi, T.; Noda, T.; Pandolfi, P.P. Distinct germline progenitor subsets defined through tsc2-mtorc1 signaling. *EMBO Rep* **2015**, *In press*.

174. Sada, A.; Suzuki, A.; Suzuki, H.; Saga, Y. The rna-binding protein nanos2 is required to maintain murine spermatogonial stem cells. *Science* **2009**, *325*, 1394-1398.

175. Suzuki, H.; Ahn, H.W.; Chu, T.; Bowden, W.; Gassei, K.; Orwig, K.; Rajkovic, A. Sohlh1 and sohlh2 coordinate spermatogonial differentiation. *Developmental biology* **2012**, *361*, 301-312.

176. Zhou, Z.; Shirakawa, T.; Ohbo, K.; Sada, A.; Wu, Q.; Hasegawa, K.; Saba, R.; Saga, Y. Rna binding protein nanos2 organizes post-transcriptional buffering system to retain primitive state of mouse spermatogonial stem cells. *Dev Cell* **2015**, *34*, 96-107.

177. Busada, J.T.; Chappell, V.A.; Niedenberger, B.A.; Kaye, E.P.; Keiper, B.D.; Hogarth, C.A.; Geyer, C.B. Retinoic acid regulates kit translation during spermatogonial differentiation in the mouse. *Developmental biology* **2015**, *397*, 140-149.

178. Phillips, B.T.; Gassei, K.; Orwig, K.E. Spermatogonial stem cell regulation and spermatogenesis. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **2010**, *365*, 1663-1678.

179. Beumer, T.L.; Roepers-Gajadien, H.L.; Gademan, I.S.; Kal, H.B.; de Rooij, D.G. Involvement of the d-type cyclins in germ cell proliferation and differentiation in the mouse. *Biol Reprod* **2000**, *63*, 1893-1898.

180. Hobbs, R.M.; Seandel, M.; Falciatori, I.; Rafii, S.; Pandolfi, P.P. Plzf regulates germline progenitor self-renewal by opposing mtorc1. *Cell* **2010**, *142*, 468-479.

181. Hobbs, R.M.; Fagoonee, S.; Papa, A.; Webster, K.; Altruda, F.; Nishinakamura, R.; Chai, L.; Pandolfi, P.P. Functional antagonism between sall4 and plzf defines germline progenitors. *Cell stem cell* **2012**, *10*, 284-298.

182. Nagano, M.C.; Yeh, J.R. The identity and fate decision control of spermatogonial stem cells: Where is the point of no return? *Current topics in developmental biology* **2013**, *102*, 61-95.

183. Chakraborty, P.; Buaas, F.W.; Sharma, M.; Snyder, E.; de Rooij, D.G.; Braun, R.E. Lin28a marks the spermatogonial progenitor population and regulates its cyclic expansion. *Stem Cells* **2014**, *32*, 860-873.

184. Chan, F.; Oatley, M.J.; Kaucher, A.V.; Yang, Q.E.; Bieberich, C.J.; Shashikant, C.S.; Oatley, J.M. Functional and molecular features of the id4+ germline stem cell population in mouse testes. *Genes & development* **2014**, *28*, 1351-1362.

185. Chan, A.L.; La, H.M.; Legrand, J.M.D.; Makela, J.A.; Eichenlaub, M.; De Seram, M.; Ramialison, M.; Hobbs, R.M. Germline stem cell activity is sustained by sall4-dependent silencing of distinct tumor suppressor genes. *Stem Cell Reports* **2017**, *9*, 956-971.

186. de Rooij, D.G. Stem cells in the testis. *International journal of experimental pathology* **1998**, *79*, 67-80.

187. Sada, A.; Hasegawa, K.; Pin, P.H.; Saga, Y. Nanos2 acts downstream of glial cell line-derived neurotrophic factor signaling to suppress differentiation of spermatogonial stem cells. *Stem Cells* **2012**, *30*, 280-291.

188. McAninch, D.; Thomas, P. Identification of highly conserved putative developmental enhancers bound by sox3 in neural progenitors using chip-seq. *PLoS One* **2014**, *9*, e113361.

189. Rogers, N.; McAninch, D.; Thomas, P. Dbx1 is a direct target of sox3 in the spinal cord. *PLoS One* **2014**, *9*, e95356.

190. Yoshida, S.; Sukeno, M.; Nakagawa, T.; Ohbo, K.; Nagamatsu, G.; Suda, T.; Nabeshima, Y. The first round of mouse spermatogenesis is a distinctive program that lacks the self-renewing spermatogonia stage. *Development* **2006**, *133*, 1495-1505.

191. Kaucher, A.V.; Oatley, M.J.; Oatley, J.M. Neurog3 is a critical downstream effector for stat3-regulated differentiation of mammalian stem and progenitor spermatogonia. *Biol Reprod* **2012**, *86*, 164, 161-111.

This page intentionally left blank.

CHAPTER 5

MANUSCRIPT II

5.0 STATEMENT OF AUTHORSHIP

Statement of Authorship

Title of Paper	Genome wide DNA binding profile of SOX3 in mouse testes
Publication Status	☐ Published ☐ Accepted for Publication ☒ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Submitted to Reproduction 13/03/2020

Principal Author

Name of Principal Author (Candidate)	Dale Christopher McAninch		
Contribution to the Paper	Concieved the study, performed the experiments, analysed the data, wrote and edited the manuscript		
Overall percentage (%)	80		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	03/03/2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Ella Thomson		
Contribution to the Paper	Analysed the data, editied the manuscript		
Signature		Date	3/3/2020

Name of Co-Author	Paul Thomas		
Contribution to the Paper	Conceived the study, guided analysis of data, evaluated and edited manuscript		
Signature		Date	18.3.2020

Genome wide DNA binding profile of SOX3 in mouse testes

Dale McAninch[1], Ella P Thomson[1] and Paul Q Thomas[1,2,3 *]

[1]School of Biological Sciences and Robinson Research Institute, University of Adelaide, Adelaide, SA 5005, Australia

[2]Adelaide Medical School, University of Adelaide, Adelaide, SA 5005, Australia

[3]Precision Medicine Theme, South Australia Health and Medical Research Institute, Adelaide, SA 5000, Australia

*Corresponding author. South Australia Health and Medical Research Institute, North Terrace, Australia, SA 5000.

Email: Paul.thomas@adelaide.edu.au

Short title: SOX3's DNA binding profile in mouse testes

Keywords: Transcription factor, ChIP, testes, histones

5.1 ABSTRACT

Spermatogenesis is the male version of gametogenesis, where germ cells are transformed into haploid spermatozoa through a tightly controlled series of mitosis, meiosis and differentiation. This process is reliant on precisely timed changes in gene expression controlled by a number of different hormonal and transcriptional mechanisms. One important transcription factor is SOX3 which is transiently expressed within the uncommitted spermatogonial stem cell population in the testes. Loss of *Sox3* in mouse testes results in a block in spermatogenesis, leading to infertility or subfertility. However, molecular role of SOX3 during spermatogonial differentiation remains poorly understood because the genomic regions targeted by this transcription factor have not been identified. To address this issue, we used ChIP-seq to identify and characterise the endogenous genome wide binding profile of SOX3 in mouse testes at post-natal day 7. As we have shown previously in neural progenitor cells, *Ngn3* is directly targeted by SOX3 in spermatogonial stem cells, albeit via a different binding site in the *Ngn3* promoter. We also implicate SOX3 for the first time in direct regulation of histone gene expression and demonstrate that this function is shared by both neural progenitors and testes, and with another important transcription factor required for spermatogenesis, PLZF. Altogether, these data provide new insight into the function of SOX3 in different stem cell contexts.

5.2 INTRODUCTION

The SOX family of transcription factors (TF) are widely expressed during embryogenesis and play a role the development of most, if not all organs [164,165]. The SOXB1 family comprises three members - *Sox3,* along with the closely-related genes *Sox1* and *Sox2*. SOXB1 TFs have been shown to act redundantly in central nervous system (CNS) development, where they act to inhibit differentiation of neural progenitor cells (NPC) [59]. Expression of the SOXB1 TFs overlap extensively, particularly within NPCs of the developing mouse CNS [60]. Despite this overlap, *Sox3* exhibits a small number of unique zones of expression. Interestingly, development of these unique zones is compromised in *Sox3* null mice, resulting in aberrant pituitary formation, corpus callosum defects and impaired spermatogenesis [92,107-109,120].

Within the testes, *Sox3* is the only SOXB1 gene expressed within the undifferentiated pool of spermatogonial stem cells (SSCs) (D McAninch, JA Mäkelä, HM La, JN Hughes, R Lovell-Badge, RM Hobbs, & PQ Thomas, manuscript under review) [92,108,109]. *Sox3* null mice exhibit a transient block in spermatogenesis resulting in a lack of mature sperm and reduced fertility [108]. While this phenotype has been reported in two independently generated *Sox3* null mouse models, the function of SOX3 in spermatogonial stem cells has not yet been fully elucidated [107]**.**

Spermatogenesis is a complex biological process that drives the differentiation of SSCs into mature sperm, via a tightly controlled combination of mitosis and meiosis. Sustained spermatogenesis occurs throughout the adult life of mammals. Many genes, including *Sox3* (as detailed above), *Ngn3* and *Plzf* are required for successful completion of spermatogenesis [155,161]. *Plzf* is a zinc finger TF with transcriptional repressor and activator activity depending on the cellular context [192]. *Plzf* is expressed within A_s, A_{pr} and A_{al} spermatogonia [154,155] and is proposed to maintain SSCs in an undifferentiated progenitor like state [193]. *Ngn3* is basic helix-loop-helix TF expressed within late committed SSCs and is required to promote their differentiation into Type 1 spermatogonia [194].

A significant amount of chromatin remodeling occurs throughout spermatogenesis, whereby standard histones are removed and replaced by testes-specific histone

variants. The testes-specific histones are subsequently replaced by protamines, an alternative form of chromatin condensing protein [195], allowing for tighter compaction of the nucleus**.** Various knockout studies performed in mice have shown that altered regulation of histone and protamine genes causes defects in spermatogenesis leading to infertility [196] indicating that chromatin repackaging is an essential process that must be undertaken for viable sperm production [197]**.**

Here, we set out interrogate the role of SOX3 in spermatogenesis by profiling SOX3 binding sites in mouse postnatal testes at genome-wide scale using ChIP-Seq. We identified 778 putative SOXB1 binding regions, including sites close to *Ngn3, H3f4*, *Hist1h2aa and Hist1h2ab*, genes known to play important roles regulating sperm development. Further, we show SOX3 and PLZF binding near a large number of histone genes, suggesting they share a common regulatory role in histone gene regulation in testes.

5.3 RESULTS

5.3.1 CHARACTERISTICS OF SOX3 BOUND REGIONS IN P7 TESTES

To identify physiologically-relevant SOX3 target sites in the postnatal mouse testes, three replicate ChIP experiments were performed on independent P7 testes using a previously validated SOX3 specific antibody [188]. *Sox3* is expressed in spermatogonial progenitor cells at P7, although no testis phenotype is evident in null mice at this stage [108]. Independent replicate datasets were overlain and overlapping regions identified from the peak sequences. Across the three experiments, 2912, 2598 and 2734 peaks were identified. Only 241 regions were present in all samples, whereas 778 regions were present in at least 2 of 3 samples. This group of 778 was used for further analysis.

To determine the genomic context of SOX3-bound regions, HOMER ChIP-Seq analysis was performed (Figure 1A). This revealed that 50% of the sites were intergenic (388/778), 31% are intronic (243/778), 11% are in the promoter region (86/778), 4% are within exons (29/778), 2% are at transcriptional start and stop sites (18/778) and 2% are within 3' or 5' UTRs (14/778).

ChIP peaks often represent regulatory regions that act upon their nearest neighboring gene, either proximal or distal to each peak location. GREAT (Genomic Regions Enrichment of Annotations Tool) uses ChIP-Seq-identified peaks and genomic annotations to predict statistically enriched cis-regulatory regions and gene ontology (GO) functions [198]. The top five terms associated with SOX3 ChIP peaks were "negative regulation of megakaryocyte differentiation", "nucleosome assembly", "nucleosome organization", "chromatin assembly" and "chromatin assembly or disassembly" (Figure 1B). There is a highly significant enrichment for functional association with nucleosomes and chromatin environment (Supplementary Table 1). Interestingly, the majority of the peaks are located close to various histone, and histone modifying genes such as *Brd2*, *Setd2* and *H2afx*.

We calculated the average phastCons score of all SOX3 bound regions, based on the conservation between 30 placental mammals (< 0.1 low conservation, 0.1-0.5 moderate conservation, and >0.5 high conservation) (Figure 1C). The majority of the sequences were poorly conserved (65%), with 23% moderately conserved and

12% highly conserved. Assessing the genomic location of the 92 highly conserved SOX3 peaks indicated that the largest proportion of peaks were either within a promoter region or overlapping an exon.

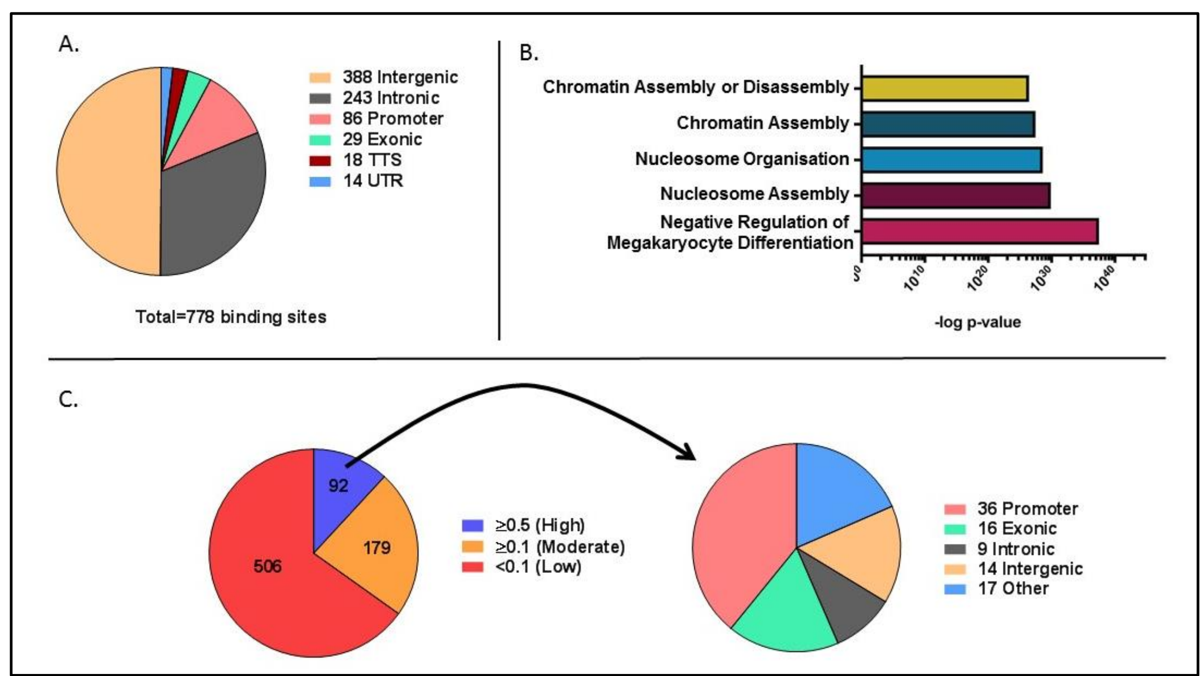


Figure 1. Overview of SOX3 ChIP-Seq biding sites in mouse P7 Testis. (A) The genomic context of all SOX3 peaks in mouse P7 testis. Transcriptional start and stop sites (TTS) and both 5' and 3' untranslated regions (UTR). (B) Enriched Gene Ontology terms identified by GREAT based on the nearest neighbouring gene to each SOX3 peak. (C) PhastCons score distribution of all 778 testes peaks. Most peaks are located in regions of low evolutionary conservation. The 92 highly conserved peaks were assessed for their genomic location, with the largest proportion located near promoters.

5.3.2 MOTIF DISCOVERY IN CHIP-SEQ PEAKS

Previously, we identified SOX3 binding motifs in NPCs, as well as a core SOXB1 motif, which highlighted the central ACAA nucleotides as critical for SOXB1 DNA binding [188]. To compare SOX3 binding in NPCs and spermatogonial stem/progenitors, we performed *de novo* motif analysis on the SOX3 testes ChIP peak sequences. We identified a SOX3 motif that is shorter than the one identified in NPCs (Figure 2A) in 421/778 (53%) of all peaks (E-value of 8.6e-457 (Figure 2B). The *de novo* testes motif is located close to the center of most ChIP-seq peaks (Figure 2C).

Comparison of the *de novo* SOX3 motif with the SOX3 motif identified in NPCs reveals a number of differences: Position 2 favours an A in testes and a G in NPCs, position 4 favours a C in testes and either a G or C in NPCs, position 7 accepts an A in both cell types as well as a T in the testes. Positions 9, 10 and 11 are not present in the testes motif. The *de novo* SOX3 motif identified in testes is very similar to the SOXB1 motif identified in McAninch and Thomas 2014, with the addition of 2 A nucleotides at the start of the recognition motif.

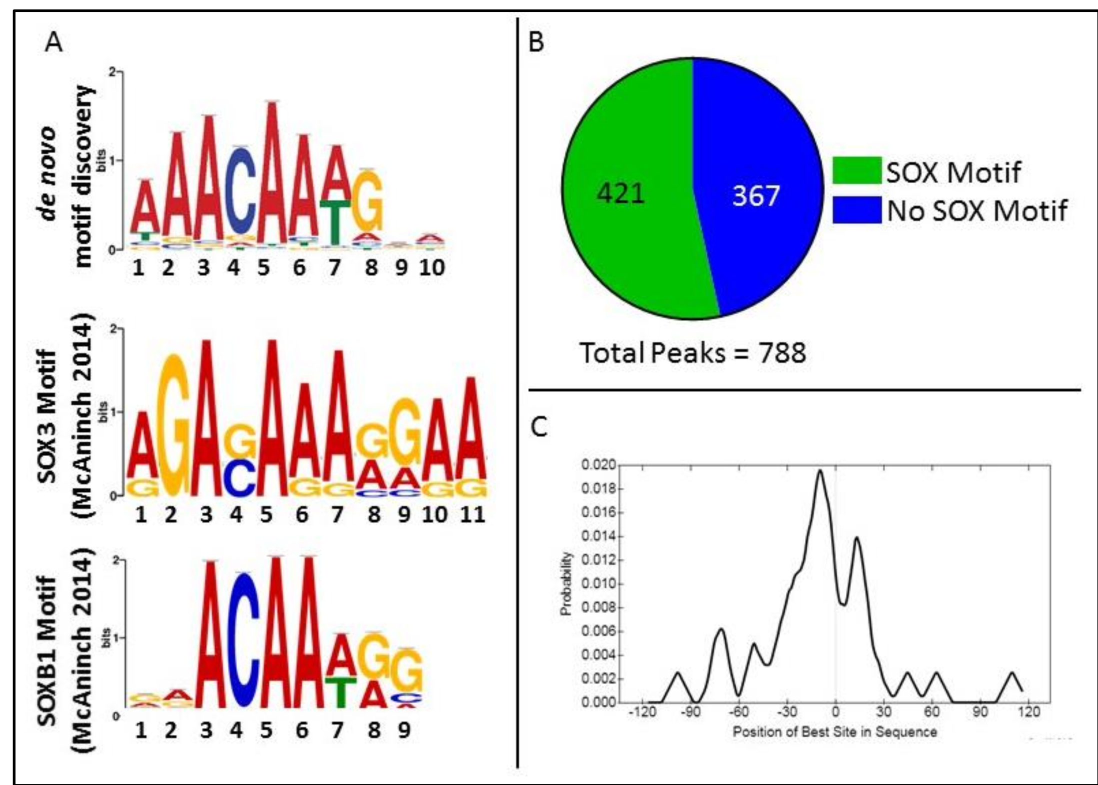


Figure 2. Motif identification in SOX3 bound regions. (A) *De novo* identification of SOX3 binding motifs within the peak sequences using MEME (E-value 5.6e-11). Comparison of the *de novo* SOX3 motif discovered in testes ChIP-Seq peaks with the SOX3 and SOXB1 motifs discovered in NPC (McAninch 2014). (B) 421 of the 778 peaks contain the de novo SOX3 motif identified in (A). (C) Positional analysis of the *de novo* SOX3 motif in all 778 peaks showing that the motif is highly central.

5.3.3 IDENTIFICATION OF PUTATIVE SOX3 PARTNER PROTEINS

SOX proteins have been shown to co-regulate expression of target genes with varying partner factors, dependent upon the tissue context [53,102,104,105,164]. We used motif enrichment analyses to identify the most significantly enriched motifs present within the 778 SOX3 peak sequences. Using AME (Analysis of motif enrichment) [199] and the HOCOMOCOv11 Mouse motif database [200], we identified more than 20 significantly enriched motifs, aside from SOX motifs. The top 6 most significantly enriched binding sites included motifs for SOX-POU, DMRT1, TEAD2, STAT1, GFI1B and NRF5A2, with P values less than 2e-10 (Table 1). Aside from the SOX-POU motif, none of these proteins have been shown to be SOX-protein partners. Consistent with this, none of the motifs, excluding SOX-POU, were in close proximity to the SOX3 DNA binding motif, suggesting coregulation but not direct interaction.

5.3.4 IDENTIFYING DIRECT SOX3 TARGET GENES THROUGH CHIP-SEQ AND MICROARRAY ANALYSES

A microarray study comparing postnatal mouse testes of wild type and *Sox3* null mice was performed previously to identify differences in gene expression caused by the loss of *Sox3* [120]. From this data, 18 genes were found to be significantly dysregulated in response to the lack of SOX3. To identify putative SOX3 direct target genes, we compared these genes to the ChIP-Seq data. Only 1 peak sequence had a nearest neighbour that belonged to the 18 significantly altered genes - neurogenin3 (*Neurog3,* also known as *Ngn3*). The microarray showed *Ngn3* was downregulated 2.9 fold, compared to wild type. The peak sequence for *Ngn3* is approximately 3kb upstream of the TSS, and is within a larger region of conservation, potentially a large regulatory region (Figure 3). We have previously identified a SOX3 binding peak directly upstream of the *Ngn3* TSS in NPCs [188]. However, the peak identified in the testes is further upstream indicating SOX3 may regulate *Ngn3* expression through different regulatory regions depending on cell type (Figure 3).

Table 1. Summary of the top 6 most significant motifs found within SOX3 bound ChIP-seq peaks.

Gene	Motif	p-Value	Positive Peaks	Total Peaks
SOX-POU	POU SOX	5.90E-32	65	778
DMRT1		5.13E-30	84	778
TEAD2		7.07E-14	30	778
STAT1		1.25E-11	31	778
GFI1B		2.22E-10	39	778
NR5A2		8.09E-09	50	778

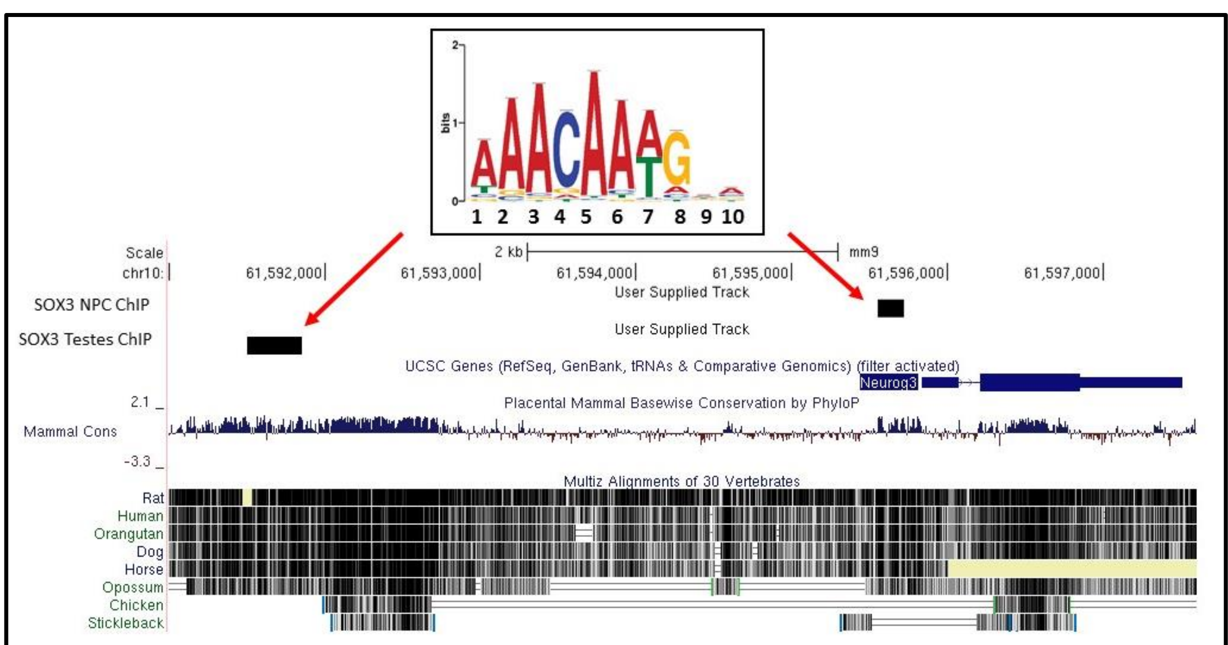


Figure 3. *Ngn3* is a direct target of SOX3 in testes. Comparison of the ChIP and microarray datasets identified *Ngn3* as a direct target of SOX3 in P7 mouse testes. UCSC browser view of *Ngn3* and the promoter region, overlayed with the testes and NPC ChIP peaks. Both peaks contained the same SOX3 binding motif, and both exist within highly conserved regions of the genome.

5.3.5 SOX3 BOUND REGIONS PRESENT IN TESTES AND NPCS

Our previous ChIP-Seq analysis of mouse NPCs (using the same SOX3 antibody) identified 8064 regions bound by SOX3 [188]. To further understand the context-specific role of SOX3 in stem cells, we compared the ChIP targets from testes and NPCs. A total of 218 regions were bound by SOX3 in both tissues (Figure 4), suggesting some overlap of SOX3 function exists between these two distinct types of stem/progenitor cells. We used GREAT to identify enriched GO terms for three separate groups; peaks common to both data sets, NPCs only, P7 testes only.

Significant enrichment for terms involving “chromatin modification”, “nucleosome and chromatin assembly”, as well as the “negative regulation of megakaryocyte differentiation” were identified for the 218 common peaks (Figure 4). Peaks found only in NPCs were associated with “maintenance of cell numbers”, “epigenetic gene regulation”, “retinoic acid signaling”, and “gliogenesis”. The 570 peaks unique to P7 testes showed no significant enrichment for any GO terms (Figure 4). These data suggest SOX3 may have a greater number of unique regulatory roles in NPCs, compared to the testes.

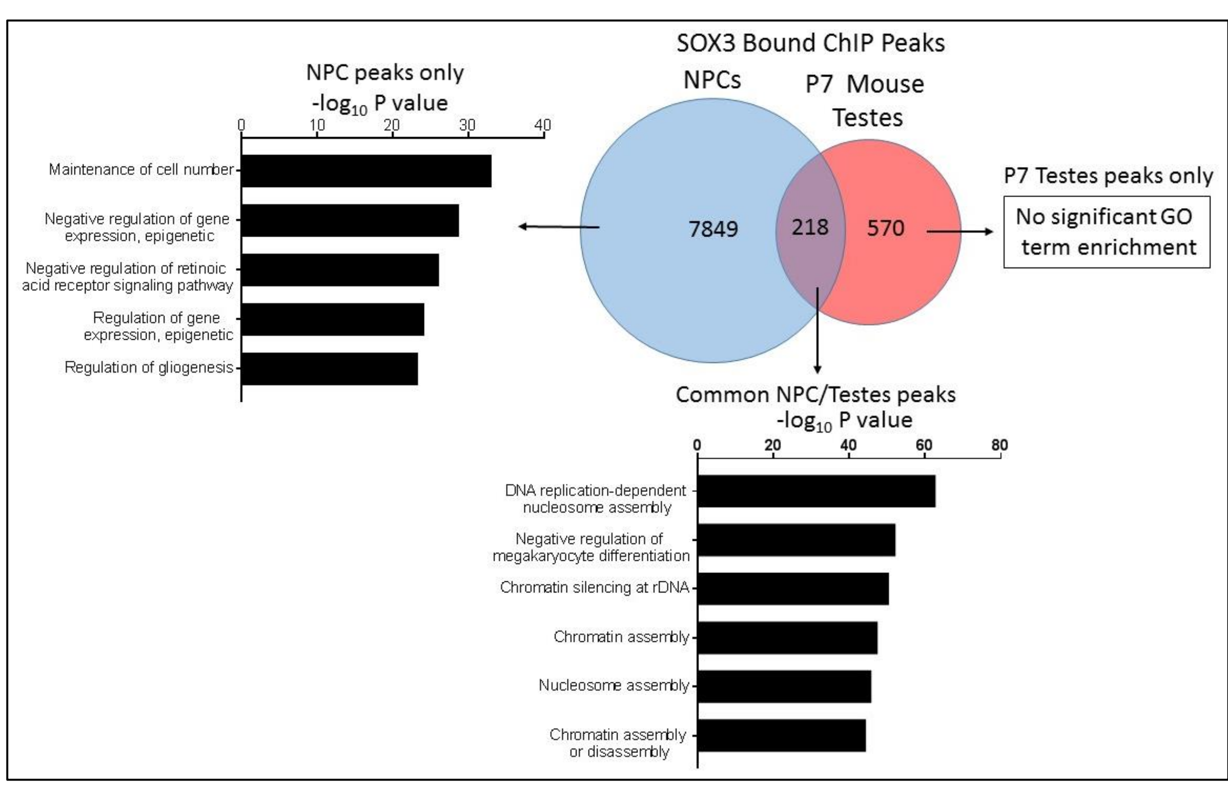


Figure 4. Comparison of NPC and P7 testes SOX3 ChIP-seq peaks. (A) Comparison of the NPC and testes datasets show that 218 peaks are common to both tissues. GO enrichment analysis of the different groups of peaks indicates a core function of SOX3 exists common to both tissue types.

5.3.6 SOX3, PLZF AND SALL4 TRANSCRIPTIONAL MECHANISMS IN SSCS

During spermatogenesis, SOX3 expression overlaps extensively with PLZF and SALL4, two zinc finger TFs expressed within a subset of undifferentiated SSCs [201] (D McAninch, JA Mäkelä, HM La, JN Hughes, R Lovell-Badge, RM Hobbs, & PQ Thomas, manuscript under review). Loss of *Plzf* results in the depletion of SSCs and spermatogenic arrest [154,155] while loss of *Sall4* results in a failure of SSC maintenance and differentiation [202]. Given the loss of any of these three TFs results in a spermatogenic defect, we wanted to identify if they share common regulatory targets.

Comparison of three ChIP-seq datasets, SOX3 from testes, and PLZF and SALL4 from isolated THY1+ SSCs [201], identified 52 common peaks between SOX3 and PLZF and 51 between SOX3 and SALL4 (Supplementary Figure 1). All three factors are located at 20 common genomic regions. GREAT analysis identified one enriched GO term for the nearest genes associated with the 20 common peaks, "Chromatin organisation" (P = 2.39e-7). Interestingly, 80% of the binding sites associated with the GO term we found within 5kb of a promoter (Supplementary Figure 1).

5.3.7 COMMON TARGETS OF SOX3 AND PLZF IN TESTES AND SOX3 IN NPCS

Due to similar GO terms being identified in both SOX3 testes/NPC and SOX3/PLZF comparisons, we further compared SOX3 in testes, SOX3 in NPCs and PLZF from SSCs. While the majority of SOX3 and PLZF bound regions are unique to each TF, there is a core group of 34 genomic regions present in all three ChIP-seq datasets (Figure 5A). GREAT analysis identified a number of significantly enriched terms, of which "chromatin assembly" was the most significantly enriched (Figure 5 B). Peaks associated with the GO term "chromatin assembly" were almost exclusively located within 5kb of the transcription start site (Figure 5 C), most were within 500bp.

Due to the large number of histone genes with nearby SOX/PLZF binding regions we examined the 3 mouse histone clusters, Hist1, Hist2 and Hist3 for ChIP peaks from all three datasets (Figure 5 D). Histone genes in Hist1 are almost all bound by both SOX3 (testes and NPCs) and PLZF, with a few exceptions, e.g. *Hist1h4b* is only bound by SOX3 and not PLZF. The Hist2 cluster is exclusively bound by SOX3 in both testes and NPCs, while the SOX3 binding site near *Hist2h4* is only found in

NPCs. Hist3 is almost exclusively bound by PLZF, with the exception of *H3f4* that is also bound by SOX3 in the testes but not NPCs. In addition to the histone clusters, there are ChIP peaks located at two additional testes specific histone genes, *Hist1h2aa* and *Hist1h2ab* (Figure 5 E).

These data suggest common regulatory mechanisms exist between SOX3 and PLZF with a focus on chromatin regulation. This common mode of action by SOX3 also exists within two different cell types, NPCs and testes.

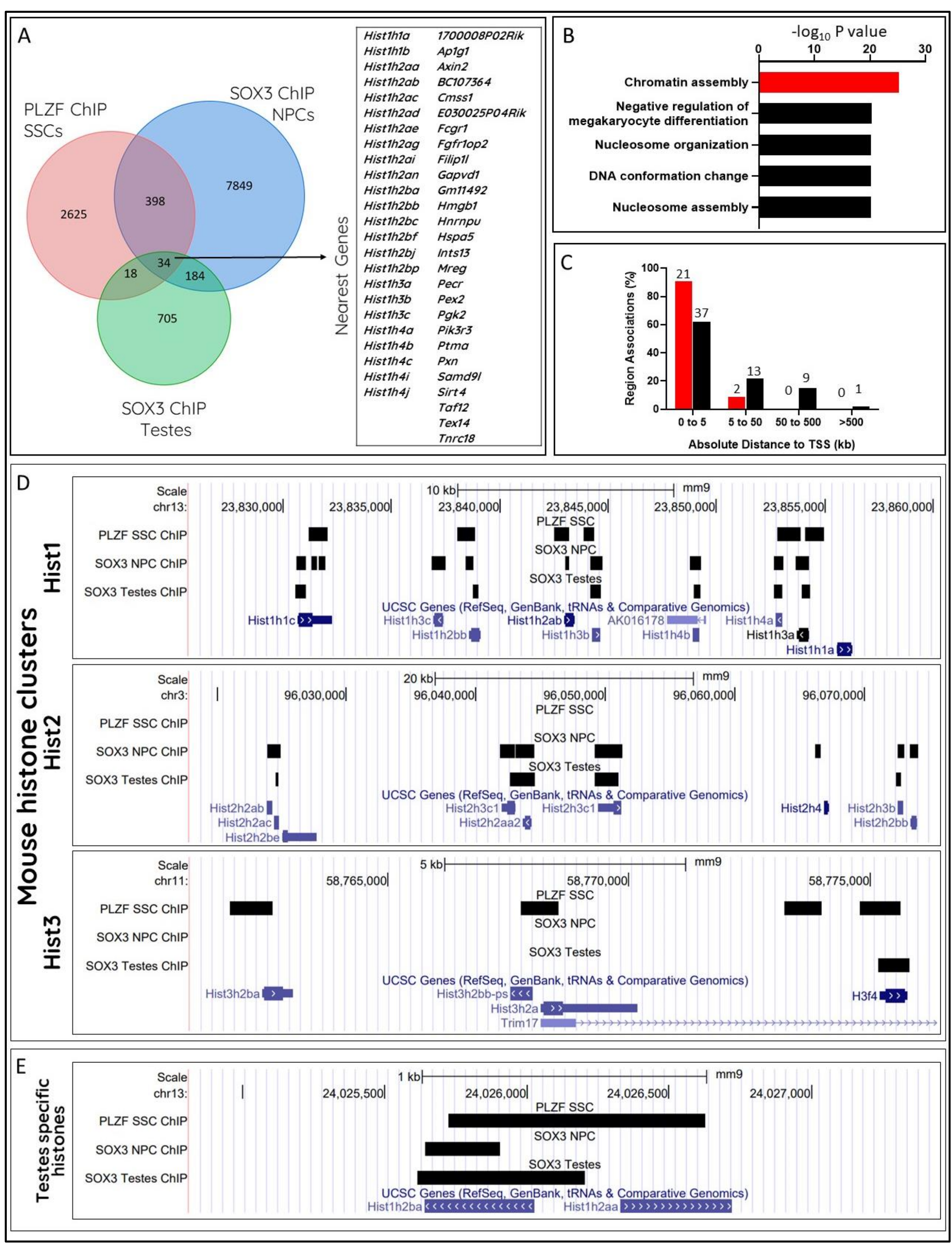


Figure 5. Comparison of NPC and P7 testes SOX3 ChIP-seq peaks with SSC PLZF ChIP-seq peaks. (A) Comparison of two SOX3 and one PLZF ChIP-seq datasets show 34 common peaks between all 3 datasets. SOX3 shares 218 peaks between two different tissue types. (B) GO enrichment analysis of the nearest genes to the

34 common peaks identified in (A). (C) Proximity of the peaks to the TSS of the nearest gene enriched in the “chromatin assembly” GO term. (D) UCSC genome browser view of the three ChIP-seq datasets looking at the Hist1, Hist2 and Hist3 clusters. (E) Genome browser view of the ChIP peaks near testes specific histone variants.

5.4 DISCUSSION

SOX3 is expressed within the undifferentiated pool of spermatogonial stem/progenitor cells in postnatal testes, and whilst it is known that loss of this TF leads to spermatogenesis defects and sterility, its transcriptional targets in this context were previously unknown [107,108]. Using ChIP-Seq analysis we have shown that SOX3 binds to 778 genomic regions within postnatal testes, and through integration with other data have identified putative SOX3 bound regulatory regions that may be important for the initiation and progression of spermatogenesis, as well as the regulation of histone gene expression. Peaks identified from testes SOX3 ChIP-seq were mostly found within intergenic and intronic regions, consistent with other SOX3 and SOXB1 TF ChIP-seq analyses [188,203,204], suggesting a regulatory role via enhancer binding. Conservation analysis via PhastCons scoring showed that 12% of bound regions had high conservation, indicative of their importance within the mammalian genome.

While the majority of the peaks we identified contained a SOX motif, 367 out of the 778 peaks did not. There is often flexibility in TF binding motifs, which can be influenced by partner factor binding, and may reduce the ability to confidently identify shorter motifs [53,205]. Alternatively, these may be representative of DNA looping events and incorporating partner protein binding sites. Chromosome conformation capture experiments would be required to tease out these fragment identities and confirm that they were physically connected at the time of fixation and not just background DNA pulldown [206].

De novo motif discovery identified a similar SOX3 binding motif to those we previously identified in NPCs. This motif was featured within a majority of the 778 peaks, and was significantly associated with the center of the ChIP peaks supporting *bona fide* SOX3:DNA interaction [207]. The motif we identified in the testis is shorter and has a number of differences when compared to the motifs identified in NPCs. This may be due in part to SOX3 binding to DNA with different partner factors, a common mechanism of SOX TFs [53]. Our comparison here shows that SOX3 binds consistently to a core motif where the central ACAA nucleotides appear to be the most critical.

Dozens of DNA binding motifs were enriched with the peaks with high significance, the top 6 presented in Table 1. Most motifs were too distantly spaced from the SOX3 motif for direct protein:protein interaction except for the SOX-POU motif, which was the most significantly enriched. SOX3 has been shown to bind co-operatively with OCT4/POU proteins in the nervous system [101,106,208], and this was further supported by a SOXB1 ChIP-Seq experiment in mouse NPCs [188]**.** OCT4 is known to be expressed within the gonocytes and spermatogonial stem cells along with SOX3, and is required for the self-renewal of the SSC population [209,210]. Thus, it seems likely that OCT4 is a partner protein for SOX3 in the mouse testes.

While almost all cells package DNA using histones, mature sperm package DNA using protamines. During the progression of spermatogonia through spermatogenesis there is complete reorganization of the chromatin structure within the nucleus [211]. Initially, standard histones are replaced with testes-specific histone variants. As spermatogonia undergo spermiogenesis these variants are replaced with transition proteins before finally being replaced by protamines as the sperm mature. Here we found SOX3 and PLZF bind to three different testes-specific histone variants *H3f4*, *Hist1h2aa* and *Hist1h2ba.* H3T (encoded by *H3f4*) has been shown to be essential for entry into spermatogenesis [197]. The H3T histone variant is incorporated into the genome prior to spermatogenesis, displacing the canonical H3 variant. Similar to *Sox3* knockout mouse models, loss of *H3f4* leads to reduced testes weight and a loss of germ cells [197]. Further analysis of SOX3 binding at this region is warranted, and may reveal a link between the loss of SOX3 leading to infertility through reduced or lack of H3T replacement of canonical H3 within the spermatogonial stem cells.

We identified a single ChIP-Seq peak, common to three datasets (PLZF in SSCs, and SOX3 in NPCS and testes) at the *Hist1h2aa* and *Hist1h2ba* loci, encoding TH2A and TH2B, respectively. These testis-specific histone variants replace the canonical core histone proteins H2A and H2B within the early spermatocytes and are encoded by genes that have a common promoter [212,213]. The replacement of H2B with TH2B is not essential for spermatogenesis [213]. *Hist1h2ba* knockout mice show no obvious spermatogenic defects, H2B expression is maintained until the end of spermatogenesis, rather than shutting down [213]. However, loss of both *Hist1h2ba*

and *Hist1h2aa* leads to defects in spermatogenesis and infertile males [214]. Neither *Hist1h2aa* nor *Hist1h2ba* have been identified as being differentially expressed either *Sox3* null mouse testes or *Plzf* null spermatogonia when compared to their respective controls [120,155]. The lack of any significant differences in expression of either of these genes in both knockout models suggest that SOX3 and PLZF (or other TFs) may be able to compensate for the loss of one another. qPCR analysis of type A spermatogonia from Sox3 or Plzf null testes could help determine if there are subtle differences in gene expression, not detected by microarray. Additionally, SSC specific double knockout mouse model of *Sox3* and *Plzf* would help confirm not only if they regulate *Hist1h2aa* and *Hist1h2ba,* but also demonstrate the existence of compensatory mechanisms.

A number of genomic regulatory regions and promoters were observed to be bound by multiple TFs, including SOX3, PLZF, and or SALL4. Multiple TF binding sites within a single element provide opportunities not just for synergistic amplification of transcription, or increased control, but also provide an additional layer of redundancy [97,99,100]. Genes like *His1h2aa* and *Hist1h2ba* feature SOX3 and PLZF binding sites at their shared promoter, however no change in gene expression is observed in single knockout mouse models of either *Sox3* or *Plzf* [120,154,155,201]. Conversely, genes such as *Ngn3,* bound uniquely by SOX3, or *Fos,* bound uniquely by PLZF, do result in reduced gene expression in their respective knock out mouse models [120,154,155,201]. These direct, uniquely bound targets appear likely to play a role in generating the phenotypes observed in the knockout mouse models [92,107,109,155,201,202].

Overlapping ChIP data for PLZF in SSCs and SOX3 in testes and NPCs indicate they share 34 binding sites. These sites include the promoters of 24 histone genes across three different sample types. The GO term "Negative regulation of megakaryocytes" (a term consisting mostly of histone genes) is identified as significant in each individual dataset, and also enriched for the peaks common to all three. Binding of all three TFs is observed within the Hist1 cluster, whereas the Hist2 and Hist3 clusters appear to be bound exclusively by either SOX3 or PLZF respectively. Most histones within the 3 clusters are replication dependent histones [215], implicating both SOX3 and PLZF in cell replication mechanisms, however the

reason only SOX3 binds Hist2 and only PLZF is found at Hist3 remains unclear. ChIP-Seq datasets are unlikely to represent a complete list of TF binding sites, due to both experimental and analytical challenges. We included peaks found in a minimum of 2 out of 3 datasets in order to capture as many regions as possible. Alternatively, ChIP-seq performed on SOX3 and PLZF from the same source material, or a SOX3/PLZF re-ChIP experiment would help strengthen this comparison. Previously, SOXB1 TFs have been linked with the posttranslational modification of histones their pioneer factor activity [55]. However, no studies to date have implicated any SOXB1 TF in the regulation of histone gene expression. These data suggest that one of the core functions of SOX3 (and PLZF) is to regulate histone gene expression.

In summary, we used ChIP-Seq to identify 778 SOX3 bound regions of the genome in the postnatal mouse testes, including a number of putative enhancers implicated in spermatogenesis. Additionally, we have described a number of common genes and promoters linking SOX3 and PLZF to histone gene regulation in postnatal mouse testes/SSCs. These data provide a platform for further functional analysis to investigate the function of SOX3 within the testes. This work has provided the first evidence of a core, tissue independent role for SOX3 in regulation of histone genes.

5.5 MATERIALS AND METHODS

5.5.1 ETHICS STATEMENT

Animal experiments were subject to approval by the Animal Ethics Committees of the University of Adelaide. All studies were conducted within the principles of animal replacement and reduction and experimental refinement. Animals were monitored daily for evidence of illness and, if distressed, were culled immediately by cervical dislocation by an experienced investigator/animal technician. All the experiments were performed in accordance with the approved guidelines and regulations.

5.5.2 CHROMATIN IMMUNOPRECIPITATION

Testes from postnatal day 7 wild type mice (129/SvJ) were collected transferred to ice cold 1xPBS containing protease inhibitors. Testes were mechanically dissociated before crosslinking in 1% freshly prepared formaldehyde for 8 minutes at room temperature. Fixed testes tissue was then lysed and sonicated (Bioruptor, Diagenode) for 15 minutes in 30 second pulses on ice. SOX3 bound chromatin was immunoprecipitated by a goat polyclonal antibody raised against human SOX3 (R&D systems, AF2569), previously used for ChIP-Seq [188] and shown to be SOX3 specific [93]. DNA was recovered by reversing crosslinks, and purified by PCR product purification kit (QIAGEN). Three independent DNA libraries were produced from testis from three independent mice with the Illumina TrueSeq library kit as per manufacturer's instructions, and libraries were sequenced using the Illumina HiSeq, producing 50 bp single end reads. A control sample (without SOX3 antibody) was run as input for background control.

5.5.3 PEAK CALLING

As published in McAninch & Thomas 2014, Bowtie [216] was used to align reads to the mouse genome (mm9). Peaks were called for each biological replicate using MACS [217], with bandwidth of 300, a model fold of 10–30, using input sample as background control, and a p-value threshold of 1e-5. Only peaks present in at least 2 of 3 biological replicates were retained.

5.5.4 GENE ONTOLOGY

Gene ontology was performed using the GREAT webtool [198] on the mouse mm9 genome. The default settings of 'basal plus' were used to define the genomic regions and the whole genome was used for background regions.

5.5.5 DE NOVO MOTIF ANALYSIS

MEME-ChIP was used to identify DNA enriched motifs from SOX3 bound peaks, searching for any number of repetitions of non-overlapping motifs. Default parameters were used and the HOCOMOCO v11 database of known mouse and human motifs [200]. WebLogo was used to visualise de novo motifs generated by MEME-ChIP [218]. FIMO (Find Individual Motif Occurrences), part of MEME-Suite, was used to confirm peak regions containing the motifs identified previously [219,220].

5.5.6 CONSERVATION ANALYSIS

PhastCons scores were generated for each peak sequence [221]. These were then ranked, and sorted into high, moderate and low conservation based on their mean score [221].

5.5.7 PUBLISHED CHIP-SEQ AND MICROARRAY DATA

Raw and processed data were submitted to NIH Gene Expression Omnibus (GEO) and Sequence Read Archive (SRA) databases under accession number GSE146706. SOX3 ChIP-Seq data from NPCs was obtained from NCBI GEO database GSE57186 [188]. Microarray data of *Sox3* null testes obtained from NCBI GSE96805 [120]. PLZF and SALL4 ChIP-seq was obtained from GSE73390 [201] and PLZF microarray expression data was obtained from GSE14222 [222].

5.6 FUNDING

This work was supported by the Australian Research Council (grant numbers DP140103965)

5.7 COMPETING INTERESTS STATEMENT

The authors have no conflicts of interest to declare.

5.8 AUTHOR CONTRIBUTIONS

D.M. and P.T. conceived the study. Experiments were performed by D.M., and analysed by D.M., E.T., and P.T. Manuscript was written by D.M., E.T. and P.T.; all authors revised the manuscript.

5.9 SUPPLEMENTARY DATA

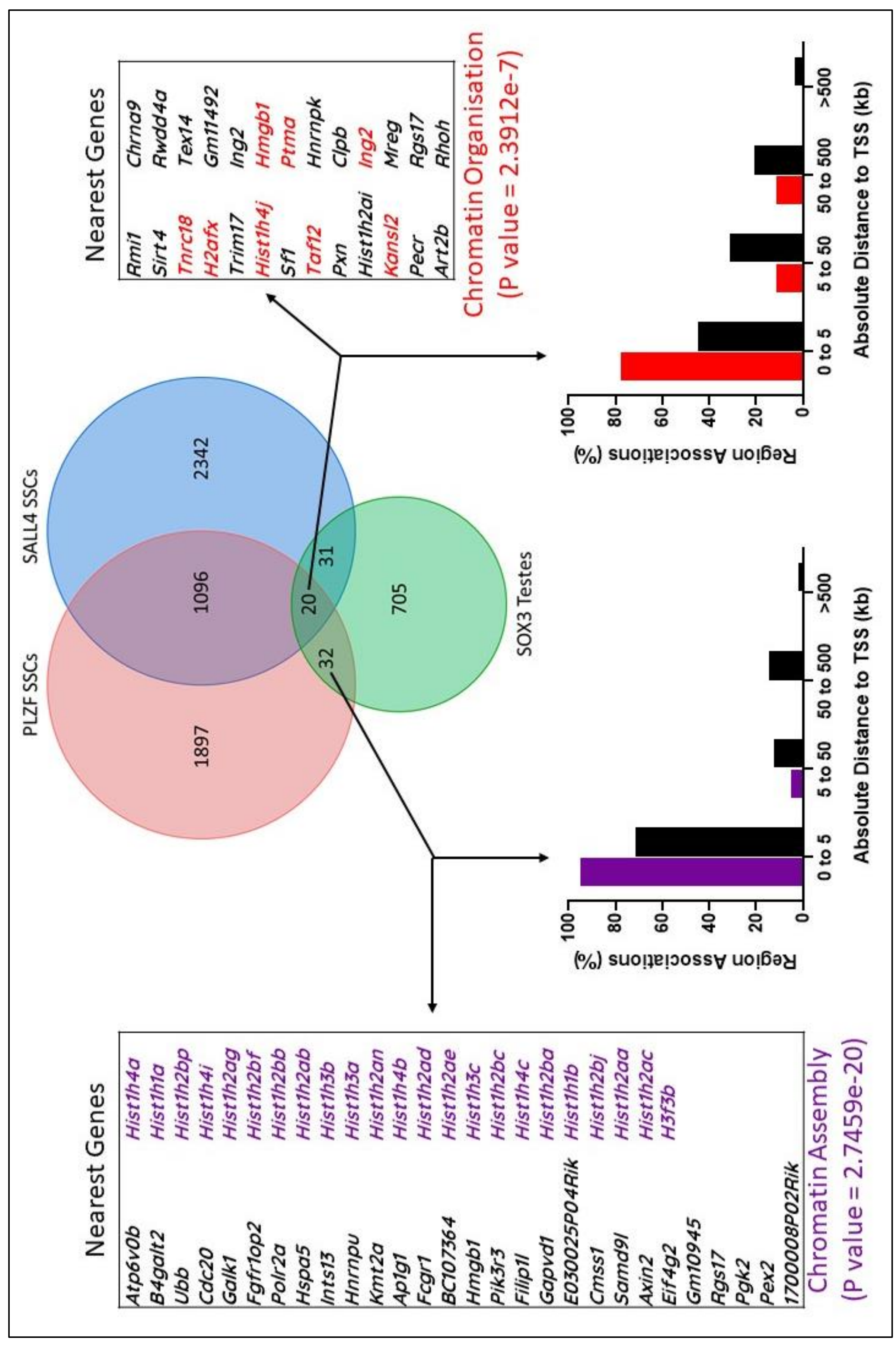


Figure S1. Comparison of P7 testes SOX3 ChIP-seq with PLZF and SALL4 SSC ChIP-seq peaks. Comparison of testes and SSC ChIP-seq datasets for SOX3, PLZF and SALL4 show 20 peaks common to all three datasets. GO analysis of the genes neighbouring the 20 common peaks identified a single enriched GO term, "chromatin organisation". Almost all histone neighbouring peaks do not overlap with SALL4. GO enrichment identified "chromatin assembly" as the only enriched term for the 32 peaks common to SOX3 and PLZF datasets. The peaks associated with the two GO terms identified were almost exclusively found within 5kb of a transcriptional start site.

Supplementary Table 1. GO term summary and associated genes

Rank	Desc	BinomP	Binom BonfP	Binom FdrQ	Exp Regions	Obs Regions	Obs Genes	Total Genes	Genes
1	**GO:0006335:** DNA replication-dependent nucleosome assembly	6.16E-47	7.91E-43	7.91E-43	3.06E-01	29	16	27	*Hist1h3a,Hist1h3b,Hist1h3c,Hist1h4a,Hist1h4b,Hist1h4c,Hist1h4f,Hist1h4h,Hist1h4i,Hist1h4j,Hist1h4k,Hist1h4m,Hist1h4n,Hist2h3b,Hist2h3c1,Hist2h3c2*
2	**GO:0045653:** negative regulation of megakaryocyte differentiation	5.10E-41	6.55E-37	3.28E-37	1.07E-01	21	10	15	*Hist1h4a,Hist1h4b,Hist1h4c,Hist1h4f,Hist1h4h,Hist1h4i,Hist1h4j,Hist1h4k,Hist1h4m,Hist1h4n*
3	**GO:0031497:** chromatin assembly	2.12E-31	2.72E-27	6.80E-28	5.41E+00	50	44	119	*Bahcc1,Brd2,H1fx,H2afy,H3f3b,Hist1h1a,Hist1h1b,Hist1h1c,Hist1h1d,Hist1h1e,Hist1h2ba,Hist1h2bb,Hist1h2bc,Hist1h2be,Hist1h2bf,Hist1h2bj,Hist1h2bk,Hist1h2bp,Hist1h2bq,Hist1h2br,Hist1h3a,Hist1h3b,Hist1h3c,Hist1h4a,Hist1h4b,Hist1h4c,Hist1h4f,Hist1h4h,Hist1h4i,Hist1h4j,Hist1h4k,Hist1h4m,Hist1h4n,Hist2h2aa1,Hist2h2aa2,Hist2h2be,Hist2h3b,Hist2h3c1,Hist2h3c2,Hmgb1,Nap1l3,Smarca5,Tnrc18,Zfp469*
4	**GO:0006333:** chromatin assembly or disassembly	5.85E-30	7.51E-26	1.50E-26	6.42E+00	52	46	137	*Bahcc1,Brd2,Chd1,H1fx,H2afy,H3f3b,Hist1h1a,Hist1h1b,Hist1h1c,Hist1h1d,Hist1h1e,Hist1h2ba,Hist1h2bb,Hist1h2bc,Hist1h2be,Hist1h2bf,Hist1h2bj,Hist1h2bk,Hist1h2bp,Hist1h2bq,Hist1h2br,Hist1h3a,Hist1h3b,Hist1h3c,Hist1h4a,Hist1h4b,Hist1h4c,Hist1h4f,Hist1h4h,Hist1h4i,Hist1h4j,Hist1h4k,Hist1h4m,Hist1h4n,Hist2h2aa1,Hist2h2aa2,Hist2h2be,Hist2h3b,Hist2h3c1,Hist2h3c2,Hmgb1,Nap1l3,Smarca5,Suv39h2,Tnrc18,Zfp469*
5	**GO:0006334:** nucleosome assembly	1.56E-28	2.00E-24	3.34E-25	4.56E+00	44	40	101	*Brd2,H1fx,H2afy,H3f3b,Hist1h1a,Hist1h1b,Hist1h1c,Hist1h1d,Hist1h1e,Hist1h2ba,Hist1h2bb,Hist1h2bc,Hist1h2be,Hist1h2bf,Hist1h2bj,Hist1h2bk,Hist1h2bp,Hist1h2bq,Hist1h2br,Hist1h3a,Hist1h3b,Hist1h3c,Hist1h4a,Hist1h4b,Hist1h4c,Hist1h4f,Hist1h4h,Hist1h4i,Hist1h4j,Hist1h4k,Hist1h4m,Hist1h4n,Hist2h2aa1,Hist2h2aa2,Hist2h2be,Hist2h3b,Hist2h3c1,Hist2h3c2,Nap1l3,Smarca5*

5.10 REFERENCES

53. Kamachi, Y.; Uchikawa, M.; Kondoh, H. Pairing sox off: With partners in the regulation of embryonic development. *Trends in genetics : TIG* **2000**, *16*, 182-187.

55. Bergsland, M.; Ramskold, D.; Zaouter, C.; Klum, S.; Sandberg, R.; Muhr, J. Sequentially acting sox transcription factors in neural lineage development. *Genes & development* **2011**, *25*, 2453-2464.

59. Bylund, M.; Andersson, E.; Novitch, B.G.; Muhr, J. Vertebrate neurogenesis is counteracted by sox1-3 activity. *Nature neuroscience* **2003**, *6*, 1162-1168.

60. Wood, H.B.; Episkopou, V. Comparative expression of the mouse sox1, sox2 and sox3 genes from pre-gastrulation to early somite stages. *Mech Dev* **1999**, *86*, 197-201.

92. Rizzoti, K.; Brunelli, S.; Carmignac, D.; Thomas, P.Q.; Robinson, I.C.; Lovell-Badge, R. Sox3 is required during the formation of the hypothalamo-pituitary axis. *Nature genetics* **2004**, *36*, 247-255.

93. Rogers, N.; Cheah, P.S.; Szarek, E.; Banerjee, K.; Schwartz, J.; Thomas, P. Expression of the murine transcription factor sox3 during embryonic and adult neurogenesis. *Gene expression patterns : GEP* **2013**, *13*, 240-248.

97. Amoutzias, G.D.; Robertson, D.L.; Van de Peer, Y.; Oliver, S.G. Choose your partners: Dimerization in eukaryotic transcription factors. *Trends Biochem Sci* **2008**, *33*, 220-229.

99. Bundschuh, R.; Hayot, F.; Jayaprakash, C. The role of dimerization in noise reduction of simple genetic networks. *J Theor Biol* **2003**, *220*, 261-269.

100. Palena, C.M.; Gonzalez, D.H.; Chan, R.L. A monomer-dimer equilibrium modulates the interaction of the sunflower homeodomain leucine-zipper protein hahb-4 with DNA. *Biochem J* **1999**, *341 (Pt 1)*, 81-87.

101. Wilson, M.; Koopman, P. Matching sox: Partner proteins and co-factors of the sox family of transcriptional regulators. *Curr Opin Genet Dev* **2002**, *12*, 441-446.

102. Ambrosetti, D.C.; Basilico, C.; Dailey, L. Synergistic activation of the fibroblast growth factor 4 enhancer by sox2 and oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. *Mol Cell Biol* **1997**, *17*, 6321-6329.

104. Yuan, H.; Corbi, N.; Basilico, C.; Dailey, L. Developmental-specific activity of the fgf-4 enhancer requires the synergistic action of sox2 and oct-3. *Genes & development* **1995**, *9*, 2635-2645.

105. Kamachi, Y.; Uchikawa, M.; Tanouchi, A.; Sekido, R.; Kondoh, H. Pax6 and sox2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes & development* **2001**, *15*, 1272-1286.

106. Tanaka, S.; Kamachi, Y.; Tanouchi, A.; Hamada, H.; Jing, N.; Kondoh, H. Interplay of sox and pou factors in regulation of the nestin gene in neural primordial cells. *Mol Cell Biol* **2004**, *24*, 8834-8846.

107. Weiss, J.; Meeks, J.J.; Hurley, L.; Raverot, G.; Frassetto, A.; Jameson, J.L. Sox3 is required for gonadal function, but not sex determination, in males and females. *Mol Cell Biol* **2003**, *23*, 8084-8091.

108. Raverot, G.; Weiss, J.; Park, S.Y.; Hurley, L.; Jameson, J.L. Sox3 expression in undifferentiated spermatogonia is required for the progression of spermatogenesis. *Developmental biology* **2005**, *283*, 215-225.

109. Laronda, M.M.; Jameson, J.L. Sox3 functions in a cell-autonomous manner to regulate spermatogonial differentiation in mice. *Endocrinology* **2011**, *152*, 1606-1615.

120. Adikusuma, F.; Pederick, D.; McAninch, D.; Hughes, J.; Thomas, P. Functional equivalence of the sox2 and sox3 transcription factors in the developing mouse brain and testes. *Genetics* **2017**, *206*, 1495-1503.

154. Buaas, F.W.; Kirsh, A.L.; Sharma, M.; McLean, D.J.; Morris, J.L.; Griswold, M.D.; de Rooij, D.G.; Braun, R.E. Plzf is required in adult male germ cells for stem cell self-renewal. *Nature genetics* **2004**, *36*, 647-652.

155. Costoya, J.A.; Hobbs, R.M.; Barna, M.; Cattoretti, G.; Manova, K.; Sukhwani, M.; Orwig, K.E.; Wolgemuth, D.J.; Pandolfi, P.P. Essential role of plzf in maintenance of spermatogonial stem cells. *Nature genetics* **2004**, *36*, 653-659.

161. Nakagawa, T.; Nabeshima, Y.; Yoshida, S. Functional identification of the actual and potential stem cell compartments in mouse spermatogenesis. *Dev Cell* **2007**, *12*, 195-206.

164. Kamachi, Y.; Kondoh, H. Sox proteins: Regulators of cell fate specification and differentiation. *Development* **2013**, *140*, 4129-4144.

165. Reiprich, S.; Wegner, M. From cns stem cells to neurons and glia: Sox for everyone. *Cell Tissue Res* **2015**, *359*, 111-124.

188. McAninch, D.; Thomas, P. Identification of highly conserved putative developmental enhancers bound by sox3 in neural progenitors using chip-seq. *PLoS One* **2014**, *9*, e113361.

192. Koubi, M.; Poplineau, M.; Vernerey, J.; N'Guyen, L.; Tiberi, G.; Garciaz, S.; El-Kaoutari, A.; Maqbool, M.A.; Andrau, J.C.; Guillouf, C., *et al.* Regulation of the positive transcriptional effect of plzf through a non-canonical ezh2 activity. *Nucleic Acids Res* **2018**, *46*, 3339-3350.

193. Filipponi, D.; Hobbs, R.M.; Ottolenghi, S.; Rossi, P.; Jannini, E.A.; Pandolfi, P.P.; Dolci, S. Repression of kit expression by plzf in germ cells. *Mol Cell Biol* **2007**, *27*, 6770-6781.

194. Song, H.W.; Wilkinson, M.F. Transcriptional control of spermatogonial maintenance and differentiation. *Semin Cell Dev Biol* **2014**, *30*, 14-26.

195. Rathke, C.; Baarends, W.M.; Awe, S.; Renkawitz-Pohl, R. Chromatin dynamics during spermiogenesis. *Biochim Biophys Acta* **2014**, *1839*, 155-168.

196. Hoghoughi, N.; Barral, S.; Vargas, A.; Rousseaux, S.; Khochbin, S. Histone variants: Essential actors in male genome programming. *J Biochem* **2018**, *163*, 97-103.

197. Ueda, J.; Harada, A.; Urahama, T.; Machida, S.; Maehara, K.; Hada, M.; Makino, Y.; Nogami, J.; Horikoshi, N.; Osakabe, A.*, et al.* Testis-specific histone variant h3t gene is essential for entry into spermatogenesis. *Cell Rep* **2017**, *18*, 593-600.

198. McLean, C.Y.; Bristor, D.; Hiller, M.; Clarke, S.L.; Schaar, B.T.; Lowe, C.B.; Wenger, A.M.; Bejerano, G. Great improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **2010**, *28*, 495-501.

199. Buske, F.A.; Boden, M.; Bauer, D.C.; Bailey, T.L. Assigning roles to DNA regulatory motifs using comparative genomics. *Bioinformatics* **2010**, *26*, 860-866.

200. Kulakovskiy, I.V.; Vorontsov, I.E.; Yevshin, I.S.; Sharipov, R.N.; Fedorova, A.D.; Rumynskiy, E.I.; Medvedeva, Y.A.; Magana-Mora, A.; Bajic, V.B.; Papatsenko, D.A.*, et al.* Hocomoco: Towards a complete collection of transcription factor binding models for human and mouse via large-scale chip-seq analysis. *Nucleic Acids Res* **2018**, *46*, D252-D259.

201. Lovelace, D.L.; Gao, Z.; Mutoji, K.; Song, Y.C.; Ruan, J.; Hermann, B.P. The regulatory repertoire of plzf and sall4 in undifferentiated spermatogonia. *Development* **2016**, *143*, 1893-1906.

202. Sakaki-Yumoto, M.; Kobayashi, C.; Sato, A.; Fujimura, S.; Matsumoto, Y.; Takasato, M.; Kodama, T.; Aburatani, H.; Asashima, M.; Yoshida, N.*, et al.* The murine homolog of sall4, a causative gene in okihiro syndrome, is essential for embryonic stem cell proliferation, and cooperates with sall1 in anorectal, heart, brain and kidney development. *Development* **2006**, *133*, 3005-3013.

203. Liu, C.F.; Lefebvre, V. The transcription factors sox9 and sox5/sox6 cooperate genome-wide through super-enhancers to drive chondrogenesis. *Nucleic Acids Res* **2015**, *43*, 8183-8203.

204. Ohba, S.; He, X.; Hojo, H.; McMahon, A.P. Distinct transcriptional programs underlie sox9 regulation of the mammalian chondrocyte. *Cell Rep* **2015**, *12*, 229-243.

205. Siggers, T.; Gordan, R. Protein-DNA binding: Complexities and multi-protein codes. *Nucleic Acids Res* **2014**, *42*, 2099-2111.

206. Han, J.; Zhang, Z.; Wang, K. 3c and 3c-based techniques: The powerful tools for spatial genome organization deciphering. *Mol Cytogenet* **2018**, *11*, 21.

207. Ma, W.; Noble, W.S.; Bailey, T.L. Motif-based analysis of large nucleotide data sets using meme-chip. *Nat Protoc* **2014**, *9*, 1428-1450.

208. Ng, C.K.; Li, N.X.; Chee, S.; Prabhakar, S.; Kolatkar, P.R.; Jauch, R. Deciphering the sox-oct partner code by quantitative cooperativity measurements. *Nucleic Acids Res* **2012**, *40*, 4933-4941.

209. Li, R.; Vannitamby, A.; Zhang, J.G.; Fehmel, E.L.; Southwell, B.R.; Hutson, J.M. Oct4-gfp expression during transformation of gonocytes into spermatogonial stem cells in the perinatal mouse testis. *J Pediatr Surg* **2015**, *50*, 2084-2089.

210. Dann, C.T.; Alvarado, A.L.; Molyneux, L.A.; Denard, B.S.; Garbers, D.L.; Porteus, M.H. Spermatogonial stem cell self-renewal requires oct4, a factor downregulated during retinoic acid-induced differentiation. *Stem Cells* **2008**, *26*, 2928-2937.

211. Boskovic, A.; Torres-Padilla, M.E. How mammals pack their sperm: A variant matter. *Genes & development* **2013**, *27*, 1635-1639.

212. Huh, N.E.; Hwang, I.W.; Lim, K.; You, K.H.; Chae, C.B. Presence of a bi-directional s phase-specific transcription regulatory element in the promoter shared by testis-specific th2a and th2b histone genes. *Nucleic Acids Res* **1991**, *19*, 93-98.

213. Montellier, E.; Boussouar, F.; Rousseaux, S.; Zhang, K.; Buchou, T.; Fenaille, F.; Shiota, H.; Debernardi, A.; Hery, P.; Curtet, S.*, et al.* Chromatin-to-nucleoprotamine transition is controlled by the histone h2b variant th2b. *Genes & development* **2013**, *27*, 1680-1692.

214. Shinagawa, T.; Huynh, L.M.; Takagi, T.; Tsukamoto, D.; Tomaru, C.; Kwak, H.G.; Dohmae, N.; Noguchi, J.; Ishii, S. Disruption of th2a and th2b genes causes defects in spermatogenesis. *Development* **2015**, *142*, 1287-1292.

215. Singh, R.; Bassett, E.; Chakravarti, A.; Parthun, M.R. Replication-dependent histone isoforms: A new source of complexity in chromatin structure and function. *Nucleic Acids Res* **2018**, *46*, 8665-8678.

216. Langmead, B. Aligning short sequencing reads with bowtie. *Curr Protoc Bioinformatics* **2010**, *Chapter 11*, Unit 11 17.

217. Feng, J.; Liu, T.; Zhang, Y. Using macs to identify peaks from chip-seq data. *Curr Protoc Bioinformatics* **2011**, *Chapter 2*, Unit 2 14.

218. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. Weblogo: A sequence logo generator. *Genome Res* **2004**, *14*, 1188-1190.

219. Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The meme suite. *Nucleic Acids Res* **2015**, *43*, W39-49.

220. Grant, C.E.; Bailey, T.L.; Noble, W.S. Fimo: Scanning for occurrences of a given motif. *Bioinformatics* **2011**, *27*, 1017-1018.

221. Siepel, A.; Bejerano, G.; Pedersen, J.S.; Hinrichs, A.S.; Hou, M.; Rosenbloom, K.; Clawson, H.; Spieth, J.; Hillier, L.W.; Richards, S.*, et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **2005**, *15*, 1034-1050.

222. Oatley, J.M.; Oatley, M.J.; Avarbock, M.R.; Tobias, J.W.; Brinster, R.L. Colony stimulating factor 1 is an extrinsic stimulator of mouse spermatogonial stem cell self-renewal. *Development* **2009**, *136*, 1191-1199.

This page intentionally left blank.

CHAPTER 6

DISCUSSION

6.1 SOX3'S GENOMIC FINGERPRINT IN NPCS AND SSCS

6.1.1 EVOLUTIONARY CONSERVATION OF SOX3 BOUND REGULATORY REGIONS

SOX3's genome wide binding sites have now been mapped in two tissues (NPCs and testes) in three independent ChIP-seq experiments, Bergsland *et al.* 2011, McAninch and Thomas 2014 (Publication I), and Manuscript II. SOX3 shows an overwhelming propensity for binding intergenic and intronic regulatory regions in preference over binding directly at proximal promoters. Publication I, highlights that more than 50% of these non-promoter binding sites are located in regions of high evolutionary conservation in NPC, while Manuscript II shows only 12% are highly conserved in testes. We highlighted the probability that many of these sites are highly conserved neural enhancers in NPCs. This high level of conservation indicates these regions are under significant selective pressure to remain unchanged [223]. Conversely, the low degree of conservation seen in the testes peaks suggests that these regions are subject to different selective pressures.

SSCs undergoing spermatogenesis undergo a large number of cell divisions before they reach mature sperm – and each division provides the opportunity to introduce alterations to the genome. These changes can increase or decrease the genetic fitness of resultant sperm, with the ultimate goal being egg fertilization. As a consequence, mutations that increase sperm fitness are more likely to be passed to the next generation. As sexual reproduction selects for the "fittest" sperm, mutations that lead to increased sperm numbers, changes in morphology and organisation within the testes are often favoured. These mutations can be present within protein coding regions to change gene function, or within regulatory regions (promoters and enhancers) to modify gene expression levels or domains of expression. In general, reproductive genes, such as those required for spermatogenesis, evolve at a much faster rate than non-reproductive genes [224-226]. It is possible that the same is true for their regulatory regions, which may explain the low degree of conservation of SOX3 peaks in the testes.

6.1.2 CHARACTERISING PUTATIVE SOX3 BOUND REGULATORY REGIONS

Between the three SOX3 ChIP-seq datasets, there are more than 15,000 SOX3 binding regions genome wide. ChIP-seq data only provides a snapshot of information for where TFs were located at the time of fixation. No further information is provided pertaining to activity at each location, ie. activating or repressing gene expression, regulating epigenetic marks, or scanning DNA for an appropriate binding site. Traditional methods for identifying areas of expression controlled by enhancers have involved making transgenic mice, with a LacZ reporter gene under the control of a genomic region of interest. Alternatively, enhancer trap reporter lines have been used as a non-targeted approach to enhancer discovery [227]. While these approaches provide some information on the control of gene control, moving enhancers from their normal genomic locus removes any context-dependant regulation.

Given the recent emergence of CRISPR/Cas9 genome editing technologies, it is now possible to take the identification and classification of TF binding sites further than before. Simple genomic modifications, such as TF binding site mutations, can be created to assess the role of virtually any TF binding site on endogenous gene expression. Binding site mutations within cell lines and/or mice, can be created and examined for effects on gene expression, TF binding via ChIP-PCR, and phenotype characterisation, providing functional information for TF binding sites.

Investigating the function of the SOXB1 bound neural enhancers, identified in Publication I, would be of significant interest. Binding site and enhancer deletion mutants could provide precise information as to the importance of individual binding sites and enhancers for gene expression. Analysis of embryos, or adult mice for gene expression by real time PCR, whole mount *in situ* hybridisation, and immunohistochemistry could give spatial information about regions of expression controlled by a single TF binding site or entire enhancer. With the rise of whole genome sequencing, disease-causing mutations within enhancers are an emerging area [228-230]. Analysing SOXB1 bound enhancers may help to tease apart the precise transcriptional mechanisms by which SOX3 acts, ultimately linking enhancers to regions of gene expression. Indeed, research performed by others in the Thomas lab has shown that deletion of the SOX-bound Nes enhancer results in

a significant reduction in Nestin expression in the developing CNS (Thomson, E.P., PhD Thesis, 2019).

6.1.3 INTERACTIONS WITH OTHER PROTEINS

The SOX/POU interaction has been the focus of many previous studies, including x-ray crystallography to identify how these proteins physically interact [57]. It has been demonstrated *in vitro* that co-operative interactions between SOX and POU proteins lead to greater responses than either TF alone [53,105,106,231]. The binding strength of SOX proteins is quite low, with a KD between 10^{-8}-10^{-9}M, while strong DNA binding domains are around 10-^{11}M [53]. With such low binding affinities it is unlikely that SOX TFs are able to activate enhancers alone. The SOX/POU binding motif was identified in both NPCs and testes ChIP-seq peaks. Interestingly, in both cell types' nucleotides closest to the centre of the SOX/POU motif were not as conserved as those on the outside of the motif, when compared to the SOX3 only DNA binding motif. This is consistent with the crystal structure of the SOX/POU DNA interaction that indicates the outer nucleotides of the SOX motif are the only nucleotides that contact the HMG domain [57]. It is likely that cooperative binding between SOX3 and POU TFs alters how SOX3 binds DNA, likely through changing the physical interaction between the HMG domain and DNA. To date, there is no crystal structure of the SOX HMG domain without a binding partner, likely due to the weak DNA binding properties of the HMG domain. Developing a crystal structure of the SOX HMG domain with a different binding partner might help identify whether SOXB1 proteins have altered DNA binding properties depending on their binding partner.

Aside from the SOX/POU motif, both of our ChIP-seq datasets (Publication I, Manuscript II) failed to identify any significant enrichment for nearby partner factor binding sites. This may suggest that SOX3 doesn't have much/any preference for cofactors, and will co-operatively influence gene expression with any available partner proteins. During this PhD a yeast 2 hybrid screen was performed in an attempt to identify potential co-factors for SOX3. The N-terminal region of SOX3 and its HMG domain was used as the bait, (as the activator domain interferes with the assay), and the library was generated from 11 day mouse embryos (Clontech).

Unfortunately, none of the potential interacting partners identified by this assay were able to be validated. We also performed a co-immunoprecipitation assay, where SOX3 was pulled down out of cell lysates using an antibody for SOX3. The resulting proteins were separated by polyacrylamide gel electrophoresis, and potential interacting partners were identified by mass spectrometry. Histones were the only proteins identified, and were likely identified as an experimental artefact as SOX3 is bound to DNA a majority of the time. These data suggest that SOX3 may be a weak protein interactor, as it is with DNA.

SOX3 partner factors remain elusive, and pulldown/non-endogenous interactions may be too harsh to produce useful data. Proximity based labelling may be a useful technique to identify these interactions, as it doesn't require the interactions to remain in place during the harsher conditions of pull down experiments. BioID is a proximity based labelling technique, successfully performed with TFs including SOX2 [232], that uses a modified Biotin Ligase attached to a protein of interest. Proteins in close proximity to the bait are then modified with biotin on exposed lysine residues. Biotinylated proteins can then be captured using a streptavidin matrix and identified by liquid chromatography-tandem mass spectrometry [233]. This method would allow for the identification of neighbouring proteins, even if there is no strong binding affinity between them.

Due to the extensive number of ChIP-seq datasets in existence, there are numerous databases of known DNA binding motifs for a number of different TFs and DNA binding proteins. We scanned our ChIP-seq peaks for enriched binding sites that were in close proximity to the SOX3 motif, with only the SOX-POU motif identified conclusively. These methods look for potential binding partners in the same location, however we know that DNA looping can play a large role in enhancer based gene expression control [234,235]. Thus, binding partners and cofactors may be located at a distant site rather than in close proximity to SOX3.

Both of our ChIP-seq data sets contained peaks that featured no identifiable SOX3 or SOX motif (Publication I, Manuscript II). This is common for ChIP-seq datasets and generally arise from either; DNA linked by protein:protein interactions, and therefore indirectly bound by the target TF or alternatively, from non-specific

antibody pulldown of either non-specific proteins or fragments of DNA [236,237]. Recently, it has been proposed that some biological artefacts may arise due to protein complexes, such as Cohesin and Polycomb, linking otherwise unrelated regions of genomic DNA and showing up as enriched peaks in ChIP-seq datasets [238]. Ultimately DNA pulled down due to protein:protein interactions (such as SOX3 and a binding partner) would be of most interest. Unfortunately, due to the complexity of these interactions and the simplicity of the ChIP-seq datasets, these interactions cannot be identified without further experimentation, such as chromatin conformation capture assays, as discussed below.

6.1.4 Loss of SOX3 has a minimal effect on gene expression in two different cell types

Knowing the genome wide binding profile of a TF is critical, yet insufficient, when it comes to understanding the function of TFs. Understanding what happens after a TF binds to DNA is just as important. We performed genome wide transcriptional profiling of *Sox3* null NPCs [189] and *Sox3* null testes by microarray [120]. We set out to characterise how the transcriptional profile of NPCs and testes change in the absence of SOX3, as discussed in Rogers *et al.* 2013 and Adikusuma *et al.* 2017 [120,189]. 19 genes were identified as being significantly differentially expressed in *Sox3* null NPCs, and 17 in *Sox3* null testes when compared to their WT controls. It was surprising to see so few changes in gene expression, given the large number of SOX3 binding sites identified by ChIP-seq. Originally it was proposed that the small number of changes to gene expression in NPCs might be due to functional redundancy between the highly similar SOXB1 proteins. However a similar number of differentially expressed genes were identified in the testes, where SOX1/2 are not present and therefore cannot act in a functionally redundant manner. Perhaps functional redundancy extends beyond SoxB1 proteins and highly important gene expression is protected by more than one TF/family. Although it is important to note that despite the few gene expression changes in the testes, there are observable phenotypes in *Sox3* null mice.

Although there were fewer differentially expressed genes than expected, one of particular interest in NPC was *Dbx1*, which was found to be decreased in *Sox3* null NPCs by microarray analysis [189]. This reduction in *Dbx1* was similarly reflected at the protein level, with fewer DBX1+ cells present in *Sox3* null NPC populations. This reduction in both RNA and protein was also observed *in vivo* in 9.5 dpc mouse embryos. We identified numerous SOX3 binding sites located near *Dbx1*, two within the second intron and three further upstream, indicating it is likely *Dbx1* is a direct target of SOX3. We know that SOX3 binds to these elements *in vivo*, and we know *Dbx1* expression is reduced in the absence of SOX3, however more information is required to confirm direct regulation. As discussed above, CRISPR/Cas9 genome editing tools can be used to create SOX3 binding site mutations in NPCs to confirm this direct regulation. Either the entire binding peak can be deleted, or the SOX

binding site mutated in ESCs, and NPCs generated from these cells assessed for *Dbx1* expression by qPCR. Further, the CRISPR/Cas9 system could be used in mice to generate the same mutations and confirm their importance *in vivo* in 9.5 dpc mouse embryos. Additionally, ChIP-PCR performed on these same embryos could confirm a loss of SOX3 binding at these sites.

6.1.5 NAVIGATING THE EPIGENETIC LANDSCAPE AROUND SOX3

A major step towards understanding SOX3's function would be to survey the chromatin landscape in NPCs or SSCs with and without the presence of SOX3. These methods will help distinguish which binding sites are important now and which sites will be important as the cells differentiate. The techniques below are best performed on populations of single cell types to help reduce noise and classify the landscape specifically in SOX3 expressing cells. Cultured NPCs should be suitable, however isolation of SSCs would aid in simplifying the datasets. Protocols exist to isolate SSCs via magnetic cell sorting [239].

Histones make up the protein core of nucleosomes; 4 different histones in duplicate make up the octamer core, which is wrapped in DNA to form chromatin. These nucleosomes help compact DNA, allowing it to fit within the nucleus of a cell. Posttranslational modifications to the tails of these histones, a short amino acid chain that extends out from the nucleosome, are associated with different states of chromatin condensation, accessibility and function. Some of the better characterised histone marks include; H3K9me3, trimethylation at lysine 9 of histone 3, found at condensed chromatin [240,241]; H3K4me3, trimethylation at lysine 4 of histone 3, found at active gene promoters [242,243].

The histone code at enhancers can distinguish between 3 different states; active, primed and poised [244-247]. An active enhancer is one that is found in open chromatin, able to bind TFs and aid in the recruitment of Pol initiating strong gene transcription, these enhancers are enriched for H3K27ac and H3K4me1. Primed enhancers are enriched for H3K4me1 and not H3K27ac, and will drive basal expression of their target genes. Poised enhancers are enriched for H3K4me1 (activator modification) and H3K27me3 (repressive modification) that hold the enhancer in an inactive state. Upon differentiation, the trimethyl group is removed from H3K27 and is acetylated to become H3K27ac, inducing strong expression of its target gene. Poised enhancers, which are a relatively new concept, are only required after the cell differentiates, implying that the use of a poised enhancer has been predetermined. Poised enhancers allow for the rapid expression of genes upon differentiation.

ATAC-Seq (Assay for Transposase-Accessible Chromatin) is a relatively new method developed to generate a global landscape of chromatin compaction [248]. It is a sensitive alternative to DNase- and MNase-Seq, providing both chromatin accessibility and nucleosome position information and can be performed with limited starting material. The method uses a hyperactive mutant of the transposase Tn5, which cuts DNA and ligates in adapters in regions of accessible chromatin. High numbers of sequencing reads in a single area indicate the chromatin is more accessible to transposase and therefore in a more open chromatin state. Reads of less that 150bp indicate that region is protected by nucleosomes, providing information on the nucleosomal landscape. Given SOX3's potential pioneer activity combining ChIP-seq and ATAC-seq from the same cell types will help to clarify which binding sites are normal TF sites and those that are pioneer sites (if any). Overlapping regions of closed chromatin, i.e. minimal to no sequencing reads and ChIP-seq peaks would suggest pioneer activity.

Alternatively, chromosome conformation capture coupled with high throughput sequencing (Hi-C) would provide experimental evidence towards which regions are directly bound by SOX3 and those connected thought cooperative interactions [249,250]. Hi-C works by crosslinking protein and DNA, as you would in ChIP, followed by restriction digest of the DNA providing compatible DNA ends. The DNA is then ligated while still attached to the TF, this is performed at low DNA concentrations to minimise ligations between fragments bound to different TF molecules. Crosslinks are reversed, and DNA is sequenced. Mapping of DNA sequences provides information regarding which regions of the genome are linked. Together these data would provide more conclusive evidence towards which enhancers are bound by SOX3 and which genes these enhancers are targeting.

Combining ChIP-Seq, ATAC-Seq, Hi-C and histone ChIP-seq data sets for the same tissue/cell types would help provide a clearer picture of the epigenetic landscape and facilitate identification of SOX3's function at each binding region. Comparison with *Sox3* null tissue will also help identify modifications that are SOX3 dependent, providing more information towards genomic regions where SOX3 is acting as a pioneer factor. These datasets would also provide more detailed information as to the state of each enhancer, if they are active or poised. It would also be interesting

to separate SSCs into GFRα1+ and GFRα1- populations (by FACS [251]) and comparing these profiles to identify what differences define the two populations. Changes in chromatin state and histone marks between the two similar cell populations will help identify differential use of enhancers and promoters, potentially providing the tools to pinpoint subtle changes required for the progression of spermatogenesis.

6.2 EXPLORING SOXB1 FUNCTIONAL REDUNDANCY

SoxB1 functional redundancy has been discussed widely and is believed to provide a mechanism to ensure successful development of the CNS [74,89,93,94,107,119,189,252,253]. While functional redundancy may be important for minimising the damage caused by the loss or mutation of a family member, it creates a difficult scenario when it comes to identifying the specific role of an individual TF. Our lab has investigated this functional redundancy in mice, through the replacement of the *Sox3* ORF with the ORF of *Sox2* [120]. The 5' and 3' UTRs were left untouched to ensure minimal disruption of any potential regulatory sequences. These *Sox2* knock-in mice displayed an almost complete rescue of the *Sox3* null phenotype, although some mild phenotypes remained. These include minor clefting of the adult pituitary, and small differences in gene expression in the testis [120]. Despite the altered gene expression, fertility was restored in these mice [120]. These residual phenotypes were minor and possibly due to slight reduction in the amount of SOX2 expressed from the *Sox3* locus compared to WT SOX3 levels. Overall, these mice demonstrate that SOX2 can biochemically and functionally replace SOX3 in both the developing embryonic brain and the postnatal testis.

We now know that SOX2 can replace SOX3 in terms of functional equivalence, which does not necessarily indicate that SOX3 (or another SOX TF) can replace SOX2. Niwa *et al.* 2016 demonstrated that SOX1 and SOX3 can restore pluripotency in ESCs that is lost when *Sox2* is removed. They also established that replacing *Sox2* with *Drosophila* SoxN allows for the generation of chimeric mice, indicating some evolutionary conservation of SoxB1 and SoxN [119]. All cells in chimeric mice do not arise from the introduced genetically modified ESCs, and therefore does not completely address whether SOX2 can be completely replaced. As *Sox3* is the founding member of the SoxB1 family, *Sox1* and *Sox2* arose from duplication events providing them with the ability to compensate for SOX3. Since their generation however, they may have gained additional functionality for which SOX3 may not be able to compensate.

Replacing the ORF of *Sox2* with *Sox3 (Sox2^{Sox3KI})*, using CRISPR/Cas9 and attempting to establish a colony of gene swapped mice would help address this question. As *Sox2* null mice exhibit peri-implantation lethality, if SOX3 cannot compensate for SOX2 before implantation, no *Sox2^{Sox3KI}* mice could be created. To verify and validate this finding, the one-step two-cell embryo microinjection method could be employed [254], injecting one cell of a two cell embryo, allowing the generation of chimeric mice. If these chimeric mice cannot be bred to create homozygous *Sox2^{Sox3KI}* mice then this would indicate that SOX3 is not completely functionally redundant.

It is interesting to note that whilst a majority of the SOX2 NPC peaks [55] are also bound by SOX3 [55,188], there's a substantial number (13,954 peaks) of SOX3 only regions (181 SOX2+/SOX3- and 1,207 SOX2+/SOX3+ peaks [55]). Given the extent of the functional redundancy observed between SOX2 and SOX3, it could be expected that a majority of the peaks between the two sets would be shared. As SOX2 can bind both active and poised enhancers in addition to acting as a pioneer factor, just like SOX3, the number of identified binding sites appears low. The peaks identified by SOX2 ChIP-seq did not have central enrichment of a SOX motif. Together this suggests that the conditions, or the antibody used were suboptimal or unable to precipitate SOX2 in all of its functional states. Alternately, these are the real binding sites of SOX2 in NPCs (and ESCs) and SOX2 had developed a more refined mechanism of targeting DNA. Without further validation with an independent antibody it is impossible to draw any significant conclusions.

While we know that functionally SOX2 can replace SOX3 in mice, it would be of interest to assess SOX2's DNA binding profile in mouse testes. This comparison would be best performed using knock-in epitope tagged SOXB1 proteins, thus removing variability in ChIP experiments introduced through the use of two unique antibodies. ChIP-seq comparing tagged endogenous SOX3 with tagged *Sox2^{Sox3KI}* mice in testes would help decode the mechanisms of their functional equivalence.

6.3 FUNCTIONAL ANALYSIS OF SOX3 IN SPERMATOGONIA

6.3.1 NON-SOX TF BINDING SITES WITHIN SOX3 CHIP-SEQ PEAKS

Many of the SOX3 peaks featured binding sites of other TFs with known roles in spermatogenesis or mouse fertility. Co-localisation of SOX3 motifs with other TF motifs was rarely spaced close enough to imply direct binding partner interactions, however their presence help strengthen the argument that these regions are important for mouse fertility.

The DMRT1 motif was found within 84 of the 778 peaks. *Dmrt1* is expressed in the primordial gonads of embryonic mice at 10.5dpc. *Dmrt1* expression increases in the testes as development progresses, and is maintained in Sertoli cells and spermatogonial stem cells [255-257]. DMRT1 has been shown to be required for maintaining and replenishing spermatogonia as stem cells, where the loss of *Dmrt1* leads to a decrease in PLZF and ultimately the differentiation of SSCs without replenishing the pool [258]. Overlapping phenotypes of *Sox3* and *Dmrt1* knockout mice supports the two genes sharing some targets. While there may be some overlap in terms of function, they cannot completely compensate for one another, as both *Sox3* null and *Dmrt1* null male mice are infertile [259].

Tead2 encodes a TF that plays a key role in the Hippo signaling pathway, regulating cell proliferation, migration as well as epithelial to mesenchymal transition [260]. *Tead2* expression is restricted to Type A spermatogonia in the developing mouse testes, both early undifferentiated spermatogonia along with *Sox3* as well as late stage undifferentiated spermatogonia [259]. *Tead2* null mice do not show any obvious phenotype and have normal fertility [261]. It is thought that Tead1/Tead2 are functionally redundant as double knockouts are embryonic lethal (do not survive past 9.5dpc). As a consequence, the fertility of Tead1/Tead2 null mice has not been assessed and would require conditional deletion in the gonads [261].

Nr5a2 is expressed within spermatogonia. *Nr5a2* null mice are embryonic lethal, and do not survive beyond 6.5dpc [262]. To date three different conditional knockout mice have been generated to assess the function of *Nr5a2* in neural

development, T-cell function and ovarian development [263-265]. NR5A2 promotes neurogenesis through directly activating *Prox1*, leading to an inhibition of Notch signaling [263]. T-Cell specific deletion of *Nr5a2* leads to a reduction in the number of mature T-Cells [264]. NR5A2 appears to have a broad role in many tissues, based on extensive cell-specific knockout data it would be likely that NR5A2 also plays an important role in spermatogenesis. Testes specific knockouts of *Nr5a2* (and *Tead1/*2), would help provide further evidence that these genes all belong to part of the same regulatory network or spermatogenesis related genes in the mouse testes.

The enrichment of these DNA binding motifs of TFs expressed in the testes, within the SOX3 peaks identified by ChIP-seq, provides further evidence that these genomic regions are of importance during spermatogenesis. While it is unlikely that any of these TFs are direct binding partners in this context, they may still be connected through larger regulatory complexes.

6.3.2 *NGN3* AND THE SPERMATOGENIC BLOCK IN *SOX3* NULL MICE

In Manuscript I and Manuscript II, we demonstrate that SOX3 binds upstream of *Ngn3.* Additionally we demonstrated that *Ngn3* expression is decreased in *Sox3* null testes, together demonstrating that *Ngn3* is a direct target of SOX3. We see that there appears to be two SOX3 regulatory regions in the promoter of *Ngn3,* one bound by SOX3 in NPCs and the other bound by SOX3 in testes. It would be interesting to assess the chromatin landscape of these regions in the two different cell populations as SOX3 binding at these regions may be controlled by chromatin modifications, or alternatively other proteins may be bound preventing SOX3s access depending on the cell type.

Sox3 null mice exhibit a reduction in *Ngn3* expression, starting off with a complete loss at P7 and returning to 70% WT levels by P28. *Ngn3* expression has been shown to be controlled by other TFs expressed in spermatogonia, such as SOHLH1 and SOHLH2 which belong to the bHLH family of TFs. Similar to *Sox3* null mice, a loss of *Sohlh1* leads to a block in spermatogenesis and a reduction in *Ngn3* expression at P7 [266]. It is possible that both SOX3 and SOHLH1 act synergistically to initiate *Ngn3* transcription, where the loss of just one of these TFs is sufficient to inhibit SSC differentiation. Individually, SOX3 or SOHLH1 may induce basal levels of *Ngn3* expression, which over time allows for partial recovery of spermatogenesis (Figure 6.1).

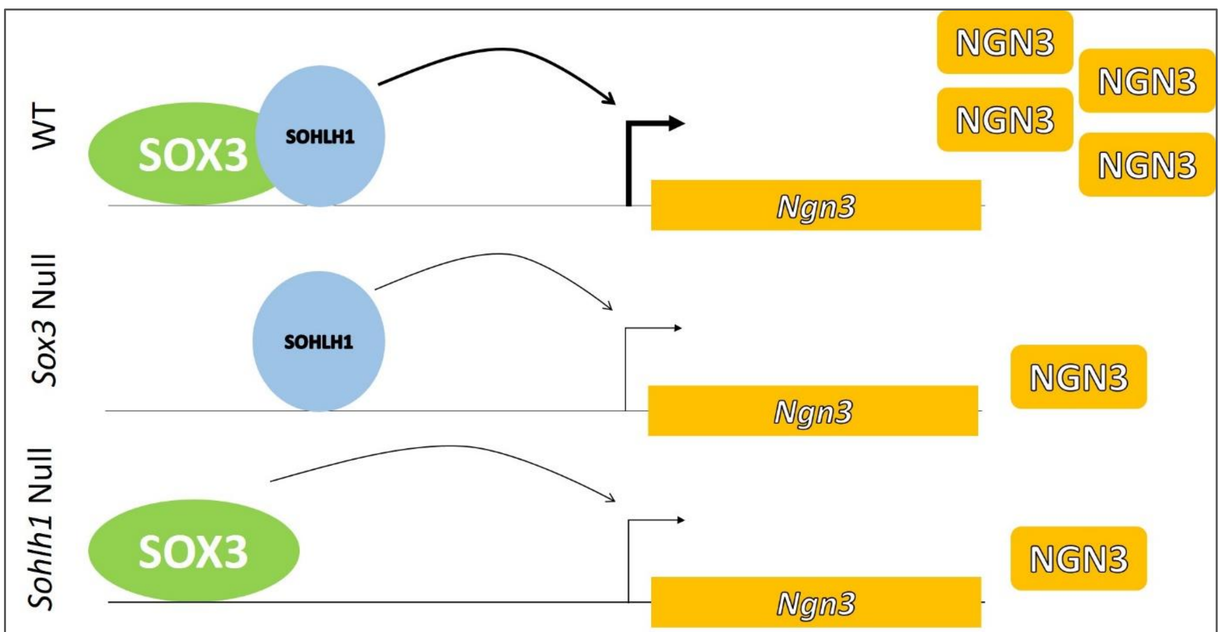


Figure 6.1. Model of *Ngn3* expression in WT, *Sox3* null and *Sohlh1* null testes. Under normal conditions SOX3 and SOHLH1 may synergistically drive *Ngn3* expression in the testes, ensuring SSCs commit to differentiation and allow spermatogenesis to progress. In either *Sox3* or *Sohlh1* null mice *Ngn3* expression is reduced dramatically, as the testes mature some SSCs may accumulate sufficient NGN3 and overcome this block, restoring some fertility.

6.3.3 Epitope tagging NGN3 for mechanistic interrogation

A significant hurdle preventing a complete understanding of the spermatogenesis pathway is the lack of a functional NGN3 antibody. To date there are no antibodies that reliably identify NGN3+ cells via immunohistochemistry. Although there are published data suggesting an NGN3 antibody is functional [108], these results have yet to be repeated successfully. While there is an NGN3/GFP reporter strain, this does not provide direct information regarding endogenous NGN3 expression, protein half-life or function. To overcome this, it is now significantly easier to tag endogenous genes [267], such as NGN3, utilizing the CRISPR/Cas9 genome editing system. Once tagged, identifying NGN3+ cells will no longer be an issue. Tagged NGN3 would allow for the identification of which cells express NGN3 and at which stage in spermatogenesis, when NGN3 expression is turned on/off, what other markers NGN3+ cells also express. Protein and transcript data for NGN3 can be combined and used to tease apart the recovery of spermatogenesis that *Sox3* null mice exhibit.

Aside from allowing us to be certain of the cells expressing NGN3, epitope tagging the endogenous NGN3 will also provide the ability to perform ChIP-Seq on NGN3. NGN3 is a basic helix-loop-helix (bHLH) TF, as identified by sequence homology within the bHLH domain [268-270]. However, with no useful antibodies developed against NGN3 understanding its mode of action is limited to knockout studies. ChIP-Seq against NGN3 would help identify which genes and transcriptional pathways drive mouse spermatogenesis.

6.3.4 Identifying the pathway leading to increased GFRα1+ cells in SOX3 null mice

In Manuscript I, we identified an increase in the number of GFRα1+ spermatogonia in P7 *Sox3* null mouse testis. SOX3 is required for the successful transition of early (GFRα1+/NGN3-) to late (GFRα1-/NGN3+) committed spermatogonia. We proposed that in *Sox3* null mouse testis, type A spermatogonia revert back from a late committed state (GFRα1-) to an early committed state (GFRα1+) after failing to express SOX3. This hypothesis could be confirmed through a fate mapping experiment.

Our model suggests that two distinct populations of GFRα1+ spermatogonia exist, those that have *not* attempted to switch on *Sox3* expression, and those that *have* attempted (but failed) to switch on *Sox3* expression. A Rosa26 Flox-STOP-Flox-LacZ reporter mouse strain can be crossed with *Sox3* null mice expressing *Cre* recombinase under the control of the *Sox3* promoter. Cells that attempt to express *Sox3* will instead express CRE, which will remove the stop codon in front of *LacZ*, in turn allowing the expression of *LacZ*. Immunohistochemical analysis of the testes (looking for GFRα1, LACZ and GFP) would provide evidence towards the mechanism behind the spermatogenic block. The presence of GFRα1+/LacZ+/GFP- spermatogonia would indicate the cells have reverted back to an early committed stage.

6.3.5 Overcoming the spermatogenic block in *Sox3* null mice

We know that the spermatogenic block in *Sox3* null mice is more severe on a pure inbred strain, such as C57BL/6J or 129/SvJ, compared to a mixed background. The ability to overcome this block may be caused by subtle differences in mouse strains, such as modifier genes [271,272], epigenetic modifications [272] and/or hormonal expression and responses [273,274]. Modifier genes are genetic variants that lead to a change phenotype at an independent locus, and can have subtle or profound effects [271]. It would be interesting to investigate what is causing this phenotypic rescue in mixed background *Sox3* null mice.

Pituitary gonadotropins follicle-stimulating hormone (FSH) and luteinising hormone (LH) are both required to initiate spermatogenesis. In their absence, spermatogenesis cannot initiate and spermatogonia remain in their undifferentiated state [275]. Although FSH and LH levels are normal in *Sox3* null mice [108], spermatogonia may still respond to increasing hormone levels within the testes and initiate differentiation. Gonadotropin or testosterone could be injected into juvenile mouse testes at regular intervals for a number of months before assessing sperm production. If sperm production is restored in *Sox3* null mice on a C57BL/6J mice then this would suggest that hormone response may be compromised, yet able to be overcome.

6.3.6 SOX3 MAY HAVE A SIMILAR ROLE IN SSCS AND OPCS

SOX3 functions in NPCs to maintain progenitor like state, while in SPCs SOX3 is required to complete the transition from early to late committed progenitor cells. While the function of SOX3 appears to be different between NPCs and SSCs, SOX3's role in SSCs and oligodendroglial progenitor cells (OPCs) may be similar. As mentioned earlier, SOX3 acts as a differentiation factor in OPCs, and its absence leads to fewer terminally differentiated oligodendrocytes [96]. Given the similar function for SOX3 between these two cell types it would be interesting to compare ChIP-seq of SOX3 and SOX2 in OPCs and SSCs. There are current culture methods allowing for the culture of OPCs from ESCs by culturing ESC-derived embryoid bodies in defined media containing retinoic acid and purmorphamine, resulting in an 80% pure OPC population [276]. The comparison of SOX3 ChIP-Seq between the two cell types would highlight any common genes and regulatory regions that SOX3 uses to promote progenitor cell differentiation. Adding in SOX2 ChIP-seq would expand on the functional redundancy mechanisms but in a different cell type, while performing different function.

6.3.7 EVIDENCE OF HUMAN INFERTILITY CAUSED BY *SOX3* MUTATIONS

As described above, *Sox3* null mice exhibit both brain and fertility defects. Predictably, these phenotypes exist within *Sox3* expressing regions within the mouse. While *Sox3* null mice exhibit fertility defects, this phenotype has not yet been linked with mutations in human *SOX3*. Stevanovic *et al.* 1993 identified a male patient with a deletion encompassing *SOX3* (and two other males from the same pedigree) exhibiting partial primary testicular failure [133]. However, it was concluded that SOX3 was not required for human testis formation as the testes still partially developed. However, in mice we still see the formation of testes, albeit empty. These *SOX3* deletion patients showed a reduction in endocrine function, however a sperm count was not reported [133]. It is still possible that these men had reduced sperm production caused by the loss of *SOX3*. Given current knowledge, assessing *SOX3* deletion patients for mature sperm production would help link *SOX3* to spermatogenesis in humans. While it is common for congenital pituitary hormone deficiencies to result in infertility [277], there have been no reported cases of impaired fertility in patients XH caused by mutations in *SOX3* [122].

Raverot *et al.* 2004 screened 56 human males with idiopathic oligoazoospermic infertility for mutations in *SOX3*. Only 3 were found to have a single polymorphic variants within the coding region, which lead to synonymous amino acid changes [278]. It is possible that screening larger cohorts of infertile, or even sub-fertile, men may lead to the identification of additional causative *SOX3* mutations. *Sox3* null mice have been shown to be able to overcome the spermatogenic block, depending on the genetic background of the mouse, it is possible that humans can overcome this block as well (if a block even occurs). Mice on a mixed background are less affected by the block in spermatogenesis, it is possible that the increased genetic diversity these mice feature, compared to an inbred strain, provides the ability to minimise genetic defects. Humans exhibit significant amounts of genetic diversity which define and modify their response to a range of challenges including chemical exposures and genetic abnormalities [279,280]. This genetic variation observed in humans may be sufficient to mitigate some genetic phenotypes, such as the spermatogenic block observed in *Sox3* null mice [281]. Alternatively, given the rapid evolution of reproductive genes [224-226], SOX3 may not play the same

role in human spermatogonia that it does in mice and therefore mutations in *SOX3* may have no effect on spermatogenesis in humans.

6.4 COMPARSION OF SOX3 IN TWO DIFFERENT TISSUES

6.4.1 SOX3 BINDS *TEX14* IN NPCS AND POSTNATAL TESTES

The SOX/POU DNA binding motif was observed in 65/218 peaks common to both NPC and testes datasets. Two of these peaks were identified within the cluster of peaks surrounding the first exon of the predominant *Tex14* transcript. Previously, we proposed that *Tex14* may belong to a large transcriptional hub in NPCs [188]. *Tex14* is essential for spermatogenesis and *Tex14* null mice are infertile due to the inability of germ cells to form intercellular bridges [282]. Similar to *Sox3* null mice, the testis of *Tex14* null mice are smaller, and lack mature sperm [283]. However, unlike *Sox3* null testis, *Tex14* male mice do not overcome this block in spermatogenesis [283]. *Tex14* was not observed to be downregulated in *Sox3* null testes [120]. This does not rule SOX3 out as being a direct regulator of *Tex14*, as there may be redundant mechanisms in place to maintain *Tex14* expression in the absence of SOX3. The microarray analysis was performed at P14, 7 days later than this ChIP-seq experiment. It would be interesting to profile the expression of *Tex14* in WT and *Sox3* null testes across a range of time points to see if there is any observable change in gene expression. The similarity in phenotypes observed, and the lack of recovery of spermatogenesis, indicates that SOX3 likely acts upstream of *Tex14.* The loss SOX3 can be compensated for, allowing for the partial rescue of spermatogenesis.

We have identified 6 SOX3 binding sites around the promoter of *Tex14* in both NPCs and testes. To date, *Tex14* expression has not been shown in the embryonic or adult mouse brain, and no neural phenotype has been described in the *Tex14* null mouse model. Our NPC microarray data shows *Tex14* is expressed in NPCs [189], however gene expression does not change following the loss of *Sox3,* much like in the testes. Considering multiple isoforms of *Tex14* exist, it is possible that subtle changes in the expression of individual isoforms have been missed by microarray. To overcome this, isoform specific primers could be developed to asses *Tex14* expression by qPCR. Thus, there is mounting evidence that SOX3 regulates *Tex14* in two different cell types, although additional functional experiments are required to clarify this connection.

6.4.2 *NGN3:* ONE GENE, TWO DISTINCT SOX3 REGULATORY REGIONS

Another gene that may be regulated by SOX3, is *Ngn3,* which encodes a basic Helix Loop Helix TF, and known mostly for its role in the differentiation of endocrine cells and NPCs [284,285]. *Ngn3* mirrors the expression pattern of SOX3 during early spermatogenesis, being found in A_s, A_{pr} and A_{al} spermatogonia [160,190], and its expression is abolished in *Sox3* null mice [108]. This, coupled with the fact that other members of the SOXB1 family are known to regulate the expression of other Neurogenin members during neuronal differentiation [286], led Raverot and colleagues (2005) to propose a functional link between SOX3 and NGN3 in the differentiation of spermatogonia. In the microarray data of *Sox3* null testes [120], its expression was reduced to 50% of normal levels. This reduced expression indicates a strong reliance on SOX3, either through direct regulation or through an indirect signaling pathway involving SOX3, or both. We have identified two different SOX3 bound regions within the promoter of *Ngn3*, one found only in testes, and one found only in NPCs. These two different regions may indicate SOX3 uses two unique binding sites to regulate the expression of *Ngn3* depending on cellular context. Control of these two binding sites may be provided by different combinations of binding partners, or different chromatin states, however further experimentation would be required. *In silico* analyses comparing the two binding regions could look for evolutionarily conserved partner binding sites to identify potential control binding sites.

Examination of the region upstream of the *Ngn3* promoter indicates two distinct highly conserved regions, each containing one of the two SOX3 bound sites. It would be interesting to assess whether either of these regions can act as enhancers through use of a reporter assay. Each highly conserved region could be used to drive the expression of a LacZ reporter system. Different cell types could be used to provide further information as to the cellular context in which these putative enhancers may be functional (eg NPCs and telomerase-immortalized mouse type A spermatogonial cells). Comprehensive ChIP profiling of the *Ngn3* promoter region for TFs including OCT4 can help tease apart what other TFs/cofactors are required for *Ngn3* expression.

6.4.3 SOX3 MAY REGULATE HISTONE EXPRESSION IN NPCS AND SPCS

During the process of cell replication, genomic DNA is duplicated requiring large numbers of *de novo* histones (H1, H2A & H2B, H3 and H4) which are integrated into and help compact the newly synthesized DNA [287]. Histone expression can control the rate of replication, if there are insufficient histones within the cell the replication rate is decreased and can ultimately cause premature exit from the cell cycle [287]. Regulation of core histones is not reliant on single transcriptional mechanisms, rather multiple redundant mechanisms exist to ensure the continuous and appropriate supply of new histones when required [288,289].

In Paper I and Manuscript II we found that SOX3 at the promoters of a large number of histone gene variants, and chromatin modifier genes in both NPCs and postnatal testes. SOX3 has not previously been reported to regulate histone gene expression. However there is mounting evidence that SOX3 may be involved in cell cycle regulation. SOXB1 TFs have been shown to help maintain neural cells in a progenitor like state, and functionally act to repress PROX1-mediated cell cycle exit and neurogenesis [59,290]. SOX3 binding at histone promoters may be indicative of a novel cell cycle control mechanism, in addition to the PROX1 pathway.

Differential expression has not been reported in histone genes when comparing WT and *Sox3* null tissues. Ultimately it is not histone RNA levels that control cell cycle exit, rather, histone protein levels. Quantitative histone analyses could be used to investigate whether changes in histone protein levels exist between WT and *Sox3* null SSCs or NPCs. Methods such as HiHiMap could be employed, where cells are incubated with DAPI, anti-CyclinA, anti-Histone of interest antibodies, the fluorescent intensities are used to quantitate histone levels [291]. Differences in total histone protein proportions within SSCs would imply SOX3 plays a critical role in histone protein regulation.

This proposed core role of SOX3 is based on preliminary DNA binding data, further investigation is required to confirm or disprove this hypothesis. Upon confirmation of SOX3 binding at the promoters of these histones in a greater number of samples, functional assays would be required to identify SOX3's function at these promoters. To date SOX3's ability to influence the cell cycle has not been addressed in mouse cells/tissues. The ability of WT and *Sox3* null NPCs/SSCs to navigate and progress

through the cell cycle could be compared using a variety of cell cycle markers, looking for premature exit of the cell cycle in *Sox3* null cells.

Throughout spermatogenesis, SSCs undergo a large number of changes, including cell morphology, nuclear compaction, and ploidy level, as they differentiate and form mature sperm. Mature sperm do not contain regular histones - they are swapped out for protamines, allowing for increased levels of nuclear compaction. This process of swapping histones for protamines proceeds through transitional phases that utilize testes-specific histone variants. It is possible that in SSCs, SOX3 is required to help with these transitions through the control of histone gene expression.

6.5 CONCLUDING REMARKS

The experiments and data produced during the course of this PhD have led to the generation of two unique SOX3 ChIP-Seq data sets in two different SOX3+ cell populations/tissues. We have identified a number of evolutionary conserved enhancers in NPCs, as well as refining a list of potential base transcriptional targets that may be common to SOX3+ cells regardless of cell type. Aside from the greater understanding of how SOX3 functions, we have developed two large ChIP-seq data sets which will be useful for us and other researchers in years to come.

We have narrowed down the specific sub population of spermatogonial cells that express SOX3, helping us understand the function of SOX3 in spermatogonia. We have highlighted potential pathways that likely include *Ngn3* by which SOX3 maintains spermatogonia in an uncommitted progenitor state. Further studies are required to reveal the complete mechanisms that drive stem cells into committed precursors. CRISPR/Cas9 genome editing tools will likely play a key role in teasing apart this mechanism.

Additionally, we have identified a number of targets of SOX3 common to both NPCs and SSCs. The majority of these sites are at the promoters of histones or chromatin related genes. This may indicate the existence of a novel core function for SOX3 relating to chromatin regulation, regardless of cell type.

Together these data help shape our knowledge of SOX3 transcriptional targets and facilitate a more complete understanding of the mechanisms that underpin SOX3's activity.

REFERENCES

1. Roeder, R.G.; Rutter, W.J. Multiple forms of DNA-dependent rna polymerase in eukaryotic organisms. *Nature* **1969**, *224*, 234-237.

2. Sentenac, A. Eukaryotic rna polymerases. *CRC Crit Rev Biochem* **1985**, *18*, 31-90.

3. Fuda, N.J.; Ardehali, M.B.; Lis, J.T. Defining mechanisms that regulate rna polymerase ii transcription in vivo. *Nature* **2009**, *461*, 186-192.

4. Lorch, Y.; Kornberg, R.D. Chromatin-remodeling for transcription. *Q Rev Biophys* **2017**, *50*, e5.

5. Lorch, Y.; LaPointe, J.W.; Kornberg, R.D. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell* **1987**, *49*, 203-210.

6. Mulvihill, E.R.; LePennec, J.P.; Chambon, P. Chicken oviduct progesterone receptor: Location of specific regions of high-affinity binding in cloned DNA fragments of hormone-responsive genes. *Cell* **1982**, *28*, 621-632.

7. Payvar, F.; Wrange, O.; Carlstedt-Duke, J.; Okret, S.; Gustafsson, J.A.; Yamamoto, K.R. Purified glucocorticoid receptors bind selectively in vitro to a cloned DNA fragment whose transcription is regulated by glucocorticoids in vivo. *Proc Natl Acad Sci U S A* **1981**, *78*, 6628-6632.

8. Engelke, D.R.; Ng, S.Y.; Shastry, B.S.; Roeder, R.G. Specific interaction of a purified transcription factor with an internal control region of 5s rna genes. *Cell* **1980**, *19*, 717-728.

9. Dynan, W.S.; Tjian, R. The promoter-specific transcription factor sp1 binds to upstream sequences in the sv40 early promoter. *Cell* **1983**, *35*, 79-87.

10. Jiang, D.; Jia, Y.; Jarrett, H.W. Transcription factor proteomics: Identification by a novel gel mobility shift-three-dimensional electrophoresis method coupled with southwestern blot and high-performance liquid chromatography-electrospray-mass spectrometry analysis. *J Chromatogr A* **2011**, *1218*, 7003-7015.

11. Lambert, S.A.; Jolma, A.; Campitelli, L.F.; Das, P.K.; Yin, Y.; Albu, M.; Chen, X.; Taipale, J.; Hughes, T.R.; Weirauch, M.T. The human transcription factors. *Cell* **2018**, *172*, 650-665.

12. Blake, J.A.; Eppig, J.T.; Kadin, J.A.; Richardson, J.E.; Smith, C.L.; Bult, C.J.; the Mouse Genome Database, G. Mouse genome database (mgd)-2017: Community knowledge resource for the laboratory mouse. *Nucleic Acids Res* **2017**, *45*, D723-D729.

13. Ptashne, M.; Gann, A. Transcriptional activation by recruitment. *Nature* **1997**, *386*, 569-577.

14. Fong, Y.W.; Cattoglio, C.; Yamaguchi, T.; Tjian, R. Transcriptional regulation by coactivators in embryonic stem cells. *Trends Cell Biol* **2012**, *22*, 292-298.

15. Zhu, F.; Farnung, L.; Kaasinen, E.; Sahu, B.; Yin, Y.; Wei, B.; Dodonova, S.O.; Nitta, K.R.; Morgunova, E.; Taipale, M.*, et al.* The interaction landscape between transcription factors and the nucleosome. *Nature* **2018**, *562*, 76-81.

16. Furlong, E.E.M.; Levine, M. Developmental enhancers and chromosome topology. *Science* **2018**, *361*, 1341-1345.

17. Robson, M.I.; Ringel, A.R.; Mundlos, S. Regulatory landscaping: How enhancer-promoter communication is sculpted in 3d. *Mol Cell* **2019**, *74*, 1110-1122.

18. Reiter, F.; Wienerroither, S.; Stark, A. Combinatorial function of transcription factors and cofactors. *Curr Opin Genet Dev* **2017**, *43*, 73-81.

19. Richardson, L.; Venkataraman, S.; Stevenson, P.; Yang, Y.; Moss, J.; Graham, L.; Burton, N.; Hill, B.; Rao, J.; Baldock, R.A.*, et al.* Emage mouse embryo spatial gene expression database: 2014 update. *Nucleic Acids Res* **2014**, *42*, D835-844.

20. Teif, V.B.; Rippe, K. Predicting nucleosome positions on the DNA: Combining intrinsic sequence preferences and remodeler activities. *Nucleic Acids Res* **2009**, *37*, 5641-5655.

21. Tyagi, M.; Imam, N.; Verma, K.; Patel, A.K. Chromatin remodelers: We are the drivers!! *Nucleus* **2016**, *7*, 388-404.

22. Zaret, K.S.; Carroll, J.S. Pioneer transcription factors: Establishing competence for gene expression. *Genes & development* **2011**, *25*, 2227-2241.

23. Cirillo, L.A.; Lin, F.R.; Cuesta, I.; Friedman, D.; Jarnik, M.; Zaret, K.S. Opening of compacted chromatin by early developmental transcription factors hnf3 (foxa) and gata-4. *Mol Cell* **2002**, *9*, 279-289.

24. Iwafuchi-Doi, M.; Zaret, K.S. Cell fate control by pioneer transcription factors. *Development* **2016**, *143*, 1833-1837.

25. Iwafuchi-Doi, M.; Donahue, G.; Kakumanu, A.; Watts, J.A.; Mahony, S.; Pugh, B.F.; Lee, D.; Kaestner, K.H.; Zaret, K.S. The pioneer transcription factor foxa maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol Cell* **2016**, *62*, 79-91.

26. Carroll, J.S.; Liu, X.S.; Brodsky, A.S.; Li, W.; Meyer, C.A.; Szary, A.J.; Eeckhoute, J.; Shao, W.; Hestermann, E.V.; Geistlinger, T.R.*, et al.* Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein foxa1. *Cell* **2005**, *122*, 33-43.

27. Becker, J.S.; Nicetto, D.; Zaret, K.S. H3k9me3-dependent heterochromatin: Barrier to cell fate changes. *Trends in genetics : TIG* **2016**, *32*, 29-41.

28. Soufi, A.; Garcia, M.F.; Jaroszewicz, A.; Osman, N.; Pellegrini, M.; Zaret, K.S. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **2015**, *161*, 555-568.

29. Cirillo, L.A.; McPherson, C.E.; Bossard, P.; Stevens, K.; Cherian, S.; Shim, E.Y.; Clark, K.L.; Burley, S.K.; Zaret, K.S. Binding of the winged-helix transcription factor hnf3 to a linker histone site on the nucleosome. *EMBO J* **1998**, *17*, 244-254.

30. Ramakrishnan, V.; Finch, J.T.; Graziano, V.; Lee, P.L.; Sweet, R.M. Crystal structure of globular domain of histone h5 and its implications for nucleosome binding. *Nature* **1993**, *362*, 219-223.

31. Clark, K.L.; Halay, E.D.; Lai, E.; Burley, S.K. Co-crystal structure of the hnf-3/fork head DNA-recognition motif resembles histone h5. *Nature* **1993**, *364*, 412-420.

32. Jozwik, K.M.; Chernukhin, I.; Serandour, A.A.; Nagarajan, S.; Carroll, J.S. Foxa1 directs h3k4 monomethylation at enhancers via recruitment of the methyltransferase mll3. *Cell Rep* **2016**, *17*, 2715-2723.

33. Mayran, A.; Drouin, J. Pioneer transcription factors shape the epigenetic landscape. *J Biol Chem* **2018**, *293*, 13795-13804.

34. Garner, M.M.; Revzin, A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: Application to components of the escherichia coli lactose operon regulatory system. *Nucleic Acids Res* **1981**, *9*, 3047-3060.

35. Johnson, P.F.; McKnight, S.L. Eukaryotic transcriptional regulatory proteins. *Annu Rev Biochem* **1989**, *58*, 799-839.

36. Mesika, A.; Ben-Dor, S.; Laviad, E.L.; Futerman, A.H. A new functional motif in hox domain-containing ceramide synthases: Identification of a novel region flanking the hox and tlc domains essential for activity. *J Biol Chem* **2007**, *282*, 27366-27373.

37. Larsson, E.; Lindahl, P.; Mostad, P. Helicis: A DNA motif discovery tool for colocalized motif pairs with periodic spacing. *BMC Bioinformatics* **2007**, *8*, 418.

38. O'Neill, L.P.; Turner, B.M. Histone h4 acetylation distinguishes coding regions of the human genome from heterochromatin in a differentiation-dependent but transcription-independent manner. *EMBO J* **1995**, *14*, 3946-3957.

39. O'Neill, L.P.; Turner, B.M. Immunoprecipitation of chromatin. *Methods Enzymol* **1996**, *274*, 189-197.

40. Kuo, M.H.; Allis, C.D. In vivo cross-linking and immunoprecipitation for studying dynamic protein:DNA associations in a chromatin environment. *Methods* **1999**, *19*, 425-433.

41. Robertson, G.; Hirst, M.; Bainbridge, M.; Bilenky, M.; Zhao, Y.; Zeng, T.; Euskirchen, G.; Bernier, B.; Varhol, R.; Delaney, A.*, et al.* Genome-wide profiles of stat1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* **2007**, *4*, 651-657.

42. Barski, A.; Cuddapah, S.; Cui, K.; Roh, T.Y.; Schones, D.E.; Wang, Z.; Wei, G.; Chepelev, I.; Zhao, K. High-resolution profiling of histone methylations in the human genome. *Cell* **2007**, *129*, 823-837.

43. Mikkelsen, T.S.; Ku, M.; Jaffe, D.B.; Issac, B.; Lieberman, E.; Giannoukos, G.; Alvarez, P.; Brockman, W.; Kim, T.K.; Koche, R.P.*, et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **2007**, *448*, 553-560.

44. Nakato, R.; Shirahige, K. Recent advances in chip-seq analysis: From quality management to whole-genome annotation. *Brief Bioinform* **2017**, *18*, 279-290.

45. Hansen, P.; Hecht, J.; Ibrahim, D.M.; Krannich, A.; Truss, M.; Robinson, P.N. Saturation analysis of chip-seq data for reproducible identification of binding peaks. *Genome Res* **2015**, *25*, 1391-1400.

46. Jeppsson, K.; Carlborg, K.K.; Nakato, R.; Berta, D.G.; Lilienthal, I.; Kanno, T.; Lindqvist, A.; Brink, M.C.; Dantuma, N.P.; Katou, Y.*, et al.* The chromosomal association of the smc5/6 complex depends on cohesion and predicts the level of sister chromatid entanglement. *PLoS Genet* **2014**, *10*, e1004680.

47. Chen, X.; Xu, H.; Yuan, P.; Fang, F.; Huss, M.; Vega, V.B.; Wong, E.; Orlov, Y.L.; Zhang, W.; Jiang, J.*, et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **2008**, *133*, 1106-1117.

48. Johnson, D.S.; Mortazavi, A.; Myers, R.M.; Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **2007**, *316*, 1497-1502.

49. Collas, P. The current state of chromatin immunoprecipitation. *Mol Biotechnol* **2010**, *45*, 87-100.

50. Schepers, G.E.; Teasdale, R.D.; Koopman, P. Twenty pairs of sox: Extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. *Dev Cell* **2002**, *3*, 167-170.

51. Lefebvre, V.; Dumitriu, B.; Penzo-Mendez, A.; Han, Y.; Pallavi, B. Control of cell fate and differentiation by sry-related high-mobility-group box (sox) transcription factors. *Int J Biochem Cell Biol* **2007**, *39*, 2195-2214.

52. Pevny, L.H.; Lovell-Badge, R. Sox genes find their feet. *Curr Opin Genet Dev* **1997**, *7*, 338-344.

53. Kamachi, Y.; Uchikawa, M.; Kondoh, H. Pairing sox off: With partners in the regulation of embryonic development. *Trends in genetics : TIG* **2000**, *16*, 182-187.

54. Harley, V.R.; Lovell-Badge, R.; Goodfellow, P.N. Definition of a consensus DNA binding site for sry. *Nucleic Acids Res* **1994**, *22*, 1500-1501.

55. Bergsland, M.; Ramskold, D.; Zaouter, C.; Klum, S.; Sandberg, R.; Muhr, J. Sequentially acting sox transcription factors in neural lineage development. *Genes & development* **2011**, *25*, 2453-2464.

56. Bowles, J.; Schepers, G.; Koopman, P. Phylogeny of the sox family of developmental transcription factors based on sequence and structural indicators. *Developmental biology* **2000**, *227*, 239-255.

57. Remenyi, A.; Lins, K.; Nissen, L.J.; Reinbold, R.; Scholer, H.R.; Wilmanns, M. Crystal structure of a pou/hmg/DNA ternary complex suggests differential assembly of oct4 and sox2 on two enhancers. *Genes & development* **2003**, *17*, 2048-2059.

58. Wegner, M. All purpose sox: The many roles of sox proteins in gene expression. *Int J Biochem Cell Biol* **2010**, *42*, 381-390.

59. Bylund, M.; Andersson, E.; Novitch, B.G.; Muhr, J. Vertebrate neurogenesis is counteracted by sox1-3 activity. *Nature neuroscience* **2003**, *6*, 1162-1168.

60. Wood, H.B.; Episkopou, V. Comparative expression of the mouse sox1, sox2 and sox3 genes from pre-gastrulation to early somite stages. *Mech Dev* **1999**, *86*, 197-201.

61. Garcia-Ramirez, M.; Martinez-Gonzalez, J.; Juan-Babot, J.O.; Rodriguez, C.; Badimon, L. Transcription factor sox18 is expressed in human coronary atherosclerotic lesions and regulates DNA synthesis and vascular cell growth. *Arterioscler Thromb Vasc Biol* **2005**, *25*, 2398-2403.

62. Hosking, B.M.; Wang, S.C.; Chen, S.L.; Penning, S.; Koopman, P.; Muscat, G.E. Sox18 directly interacts with mef2c in endothelial cells. *Biochem Biophys Res Commun* **2001**, *287*, 493-500.

63. Lee, S.; Kim, I.K.; Ahn, J.S.; Woo, D.C.; Kim, S.T.; Song, S.; Koh, G.Y.; Kim, H.S.; Jeon, B.H.; Kim, I. Deficiency of endothelium-specific transcription factor sox17 induces intracranial aneurysm. *Circulation* **2015**, *131*, 995-1005.

64. Pingault, V.; Bondurand, N.; Kuhlbrodt, K.; Goerich, D.E.; Prehu, M.O.; Puliti, A.; Herbarth, B.; Hermans-Borgmeyer, I.; Legius, E.; Matthijs, G.*, et al.* Sox10 mutations in patients with waardenburg-hirschsprung disease. *Nature genetics* **1998**, *18*, 171-173.

65. Kuhlbrodt, K.; Schmidt, C.; Sock, E.; Pingault, V.; Bondurand, N.; Goossens, M.; Wegner, M. Functional analysis of sox10 mutations found in human waardenburg-hirschsprung patients. *J Biol Chem* **1998**, *273*, 23033-23038.

66. Yasuno, K.; Bilguvar, K.; Bijlenga, P.; Low, S.K.; Krischek, B.; Auburger, G.; Simon, M.; Krex, D.; Arlier, Z.; Nayak, N., *et al.* Genome-wide association study of intracranial aneurysm identifies three new risk loci. *Nature genetics* **2010**, *42*, 420-425.

67. Jo, A.; Denduluri, S.; Zhang, B.; Wang, Z.; Yin, L.; Yan, Z.; Kang, R.; Shi, L.L.; Mok, J.; Lee, M.J., *et al.* The versatile functions of sox9 in development, stem cells, and human diseases. *Genes Dis* **2014**, *1*, 149-161.

68. Wagner, T.; Wirth, J.; Meyer, J.; Zabel, B.; Held, M.; Zimmer, J.; Pasantes, J.; Bricarelli, F.D.; Keutel, J.; Hustert, E., *et al.* Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the sry-related gene sox9. *Cell* **1994**, *79*, 1111-1120.

69. Mansour, S.; Hall, C.M.; Pembrey, M.E.; Young, I.D. A clinical and genetic study of campomelic dysplasia. *J Med Genet* **1995**, *32*, 415-420.

70. Mansour, S.; Offiah, A.C.; McDowall, S.; Sim, P.; Tolmie, J.; Hall, C. The phenotype of survivors of campomelic dysplasia. *J Med Genet* **2002**, *39*, 597-602.

71. Rosenfeld, C.S. Brain sexual differentiation and requirement of sry: Why or why not? *Front Neurosci* **2017**, *11*, 632.

72. Wang, X.; Li, X.; Wang, T.; Wu, S.P.; Jeong, J.W.; Kim, T.H.; Young, S.L.; Lessey, B.A.; Lanz, R.B.; Lydon, J.P., *et al.* Sox17 regulates uterine epithelial-stromal cross-talk acting via a distal enhancer upstream of ihh. *Nat Commun* **2018**, *9*, 4421.

73. Foster, J.W.; Graves, J.A. An sry-related sequence on the marsupial x chromosome: Implications for the evolution of the mammalian testis-determining gene. *Proc Natl Acad Sci U S A* **1994**, *91*, 1927-1931.

74. Woods, K.S.; Cundall, M.; Turton, J.; Rizotti, K.; Mehta, A.; Palmer, R.; Wong, J.; Chong, W.K.; Al-Zyoud, M.; El-Ali, M., *et al.* Over- and underdosage of sox3 is associated with infundibular hypoplasia and hypopituitarism. *Am J Hum Genet* **2005**, *76*, 833-849.

75. Shih, Y.H.; Kuo, C.L.; Hirst, C.S.; Dee, C.T.; Liu, Y.R.; Laghari, Z.A.; Scotting, P.J. Soxb1 transcription factors restrict organizer gene expression by repressing multiple events downstream of wnt signalling. *Development* **2010**, *137*, 2671-2681.

76. Gaudet, J.; Mango, S.E. Regulation of organogenesis by the caenorhabditis elegans foxa protein pha-4. *Science* **2002**, *295*, 821-825.

77. Cunha, P.M.; Sandmann, T.; Gustafson, E.H.; Ciglar, L.; Eichenlaub, M.P.; Furlong, E.E. Combinatorial binding leads to diverse regulatory responses: Lmd is a tissue-specific modulator of mef2 activity. *PLoS Genet* **2010**, *6*, e1001014.

78. Sandmann, T.; Jensen, L.J.; Jakobsen, J.S.; Karzynski, M.M.; Eichenlaub, M.P.; Bork, P.; Furlong, E.E. A temporal map of transcription factor activity: Mef2 directly regulates target genes at all stages of muscle development. *Dev Cell* **2006**, *10*, 797-807.

79. Jakobsen, J.S.; Braun, M.; Astorga, J.; Gustafson, E.H.; Sandmann, T.; Karzynski, M.; Carlsson, P.; Furlong, E.E. Temporal chip-on-chip reveals biniou as a universal regulator of the visceral muscle transcriptional network. *Genes & development* **2007**, *21*, 2448-2460.

80. Sandmann, T.; Girardot, C.; Brehme, M.; Tongprasit, W.; Stolc, V.; Furlong, E.E. A core transcriptional network for early mesoderm development in drosophila melanogaster. *Genes & development* **2007**, *21*, 436-449.

81. Wilczynski, B.; Furlong, E.E. Dynamic crm occupancy reflects a temporal map of developmental progression. *Mol Syst Biol* **2010**, *6*, 383.

82. Masel, J.; Siegal, M.L. Robustness: Mechanisms and consequences. *Trends in genetics : TIG* **2009**, *25*, 395-403.

83. Sauvageot, C.M.; Stiles, C.D. Molecular mechanisms controlling cortical gliogenesis. *Curr Opin Neurobiol* **2002**, *12*, 244-249.

84. Morshead, C.M.; Reynolds, B.A.; Craig, C.G.; McBurney, M.W.; Staines, W.A.; Morassutti, D.; Weiss, S.; van der Kooy, D. Neural stem cells in the adult mammalian forebrain: A relatively quiescent subpopulation of subependymal cells. *Neuron* **1994**, *13*, 1071-1082.

85. Weiss, S.; Reynolds, B.A.; Vescovi, A.L.; Morshead, C.; Craig, C.G.; van der Kooy, D. Is there a neural stem cell in the mammalian forebrain? *Trends Neurosci* **1996**, *19*, 387-393.

86. Morshead, C.M. Adult neural stem cells: Attempting to solve the identity crisis. *Dev Neurosci* **2004**, *26*, 93-100.

87. Morshead, C.M.; van der Kooy, D. Disguising adult neural stem cells. *Curr Opin Neurobiol* **2004**, *14*, 125-131.

88. Fukuda, S.; Kato, F.; Tozuka, Y.; Yamaguchi, M.; Miyamoto, Y.; Hisatsune, T. Two distinct subpopulations of nestin-positive cells in adult mouse dentate gyrus. *J Neurosci* **2003**, *23*, 9357-9366.

89. Collignon, J.; Sockanathan, S.; Hacker, A.; Cohen-Tannoudji, M.; Norris, D.; Rastan, S.; Stevanovic, M.; Goodfellow, P.N.; Lovell-Badge, R. A comparison of the properties of sox-3 with sry and two related genes, sox-1 and sox-2. *Development* **1996**, *122*, 509-520.

90. Solomon, N.M.; Ross, S.A.; Forrest, S.M.; Thomas, P.Q.; Morgan, T.; Belsky, J.L.; Hol, F.A.; Karnes, P.S.; Hopwood, N.J.; Myers, S.E., *et al.* Array comparative genomic hybridisation analysis of boys with x-linked hypopituitarism identifies a 3.9 mb duplicated critical region at xq27 containing sox3. *J Med Genet* **2007**, *44*, e75.

91. Brunelli, S.; Silva Casey, E.; Bell, D.; Harland, R.; Lovell-Badge, R. Expression of sox3 throughout the developing central nervous system is dependent on the combined action of discrete, evolutionarily conserved regulatory elements. *Genesis* **2003**, *36*, 12-24.

92. Rizzoti, K.; Brunelli, S.; Carmignac, D.; Thomas, P.Q.; Robinson, I.C.; Lovell-Badge, R. Sox3 is required during the formation of the hypothalamo-pituitary axis. *Nature genetics* **2004**, *36*, 247-255.

93. Rogers, N.; Cheah, P.S.; Szarek, E.; Banerjee, K.; Schwartz, J.; Thomas, P. Expression of the murine transcription factor sox3 during embryonic and adult neurogenesis. *Gene expression patterns : GEP* **2013**, *13*, 240-248.

94. Cheah, P.S.; Thomas, P.Q. Sox3 expression in the glial system of the developing and adult mouse cerebellum. *Springerplus* **2015**, *4*, 400.

95. Holmberg, J.; Hansson, E.; Malewicz, M.; Sandberg, M.; Perlmann, T.; Lendahl, U.; Muhr, J. Soxb1 transcription factors and notch signaling use distinct mechanisms to regulate proneural gene function and neural progenitor differentiation. *Development* **2008**, *135*, 1843-1851.

96. Hoffmann, S.A.; Hos, D.; Kuspert, M.; Lang, R.A.; Lovell-Badge, R.; Wegner, M.; Reiprich, S. Stem cell factor sox2 and its close relative sox3 have differentiation functions in oligodendrocytes. *Development* **2014**, *141*, 39-50.

97. Amoutzias, G.D.; Robertson, D.L.; Van de Peer, Y.; Oliver, S.G. Choose your partners: Dimerization in eukaryotic transcription factors. *Trends Biochem Sci* **2008**, *33*, 220-229.

98. Desvergne, B.; Michalik, L.; Wahli, W. Transcriptional regulation of metabolism. *Physiol Rev* **2006**, *86*, 465-514.

99. Bundschuh, R.; Hayot, F.; Jayaprakash, C. The role of dimerization in noise reduction of simple genetic networks. *J Theor Biol* **2003**, *220*, 261-269.

100. Palena, C.M.; Gonzalez, D.H.; Chan, R.L. A monomer-dimer equilibrium modulates the interaction of the sunflower homeodomain leucine-zipper protein hahb-4 with DNA. *Biochem J* **1999**, *341 (Pt 1)*, 81-87.

101. Wilson, M.; Koopman, P. Matching sox: Partner proteins and co-factors of the sox family of transcriptional regulators. *Curr Opin Genet Dev* **2002**, *12*, 441-446.

102. Ambrosetti, D.C.; Basilico, C.; Dailey, L. Synergistic activation of the fibroblast growth factor 4 enhancer by sox2 and oct-3 depends on protein-protein

interactions facilitated by a specific spatial arrangement of factor binding sites. *Mol Cell Biol* **1997**, *17*, 6321-6329.

103. Melnikova, I.N.; Lin, H.R.; Blanchette, A.R.; Gardner, P.D. Synergistic transcriptional activation by sox10 and sp1 family members. *Neuropharmacology* **2000**, *39*, 2615-2623.

104. Yuan, H.; Corbi, N.; Basilico, C.; Dailey, L. Developmental-specific activity of the fgf-4 enhancer requires the synergistic action of sox2 and oct-3. *Genes & development* **1995**, *9*, 2635-2645.

105. Kamachi, Y.; Uchikawa, M.; Tanouchi, A.; Sekido, R.; Kondoh, H. Pax6 and sox2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes & development* **2001**, *15*, 1272-1286.

106. Tanaka, S.; Kamachi, Y.; Tanouchi, A.; Hamada, H.; Jing, N.; Kondoh, H. Interplay of sox and pou factors in regulation of the nestin gene in neural primordial cells. *Mol Cell Biol* **2004**, *24*, 8834-8846.

107. Weiss, J.; Meeks, J.J.; Hurley, L.; Raverot, G.; Frassetto, A.; Jameson, J.L. Sox3 is required for gonadal function, but not sex determination, in males and females. *Mol Cell Biol* **2003**, *23*, 8084-8091.

108. Raverot, G.; Weiss, J.; Park, S.Y.; Hurley, L.; Jameson, J.L. Sox3 expression in undifferentiated spermatogonia is required for the progression of spermatogenesis. *Developmental biology* **2005**, *283*, 215-225.

109. Laronda, M.M.; Jameson, J.L. Sox3 functions in a cell-autonomous manner to regulate spermatogonial differentiation in mice. *Endocrinology* **2011**, *152*, 1606-1615.

110. Doetschman, T. Influence of genetic background on genetically engineered mouse phenotypes. *Methods Mol Biol* **2009**, *530*, 423-433.

111. Uwanogho, D.; Rex, M.; Cartwright, E.J.; Pearl, G.; Healy, C.; Scotting, P.J.; Sharpe, P.T. Embryonic expression of the chicken sox2, sox3 and sox11 genes suggests an interactive role in neuronal development. *Mech Dev* **1995**, *49*, 23-36.

112. Zappone, M.V.; Galli, R.; Catena, R.; Meani, N.; De Biasi, S.; Mattei, E.; Tiveron, C.; Vescovi, A.L.; Lovell-Badge, R.; Ottolenghi, S.*, et al.* Sox2 regulatory sequences direct expression of a (beta)-geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in cns stem cells. *Development* **2000**, *127*, 2367-2382.

113. Avilion, A.A.; Nicolis, S.K.; Pevny, L.H.; Perez, L.; Vivian, N.; Lovell-Badge, R. Multipotent cell lineages in early mouse development depend on sox2 function. *Genes & development* **2003**, *17*, 126-140.

114. Aubert, J.; Stavridis, M.P.; Tweedie, S.; O'Reilly, M.; Vierlinger, K.; Li, M.; Ghazal, P.; Pratt, T.; Mason, J.O.; Roy, D., *et al.* Screening for mammalian neural genes via fluorescence-activated cell sorter purification of neural precursors from sox1-gfp knock-in mice. *Proc Natl Acad Sci U S A* **2003**, *100 Suppl 1*, 11836-11841.

115. Nishiguchi, S.; Wood, H.; Kondoh, H.; Lovell-Badge, R.; Episkopou, V. Sox1 directly regulates the gamma-crystallin genes and is essential for lens development in mice. *Genes & development* **1998**, *12*, 776-781.

116. Rizzoti, K.; Lovell-Badge, R. Sox3 activity during pharyngeal segmentation is required for craniofacial morphogenesis. *Development* **2007**, *134*, 3437-3448.

117. Ferri, A.L.; Cavallaro, M.; Braida, D.; Di Cristofano, A.; Canta, A.; Vezzani, A.; Ottolenghi, S.; Pandolfi, P.P.; Sala, M.; DeBiasi, S., *et al.* Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development* **2004**, *131*, 3805-3819.

118. Masui, S.; Nakatake, Y.; Toyooka, Y.; Shimosato, D.; Yagi, R.; Takahashi, K.; Okochi, H.; Okuda, A.; Matoba, R.; Sharov, A.A., *et al.* Pluripotency governed by sox2 via regulation of oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* **2007**, *9*, 625-635.

119. Niwa, H.; Nakamura, A.; Urata, M.; Shirae-Kurabayashi, M.; Kuraku, S.; Russell, S.; Ohtsuka, S. The evolutionally-conserved function of group b1 sox family members confers the unique role of sox2 in mouse es cells. *BMC Evol Biol* **2016**, *16*, 173.

120. Adikusuma, F.; Pederick, D.; McAninch, D.; Hughes, J.; Thomas, P. Functional equivalence of the sox2 and sox3 transcription factors in the developing mouse brain and testes. *Genetics* **2017**, *206*, 1495-1503.

121. Solomon, N.M.; Nouri, S.; Warne, G.L.; Lagerstrom-Fermer, M.; Forrest, S.M.; Thomas, P.Q. Increased gene dosage at xq26-q27 is associated with x-linked hypopituitarism. *Genomics* **2002**, *79*, 553-559.

122. Arya, V.B.; Chawla, G.; Nambisan, A.K.R.; Muhi-Iddin, N.; Vamvakiti, E.; Ajzensztejn, M.; Hulse, T.; Ferreira Pinto, C.; Lahiri, N.; Bint, S., *et al.* Xq27.1 duplication encompassing sox3: Variable phenotype and smallest duplication associated with hypopituitarism to date - a large case series of unrelated patients and a literature review. *Horm Res Paediatr* **2019**, 1-8.

123. Haines, B.; Hughes, J.; Corbett, M.; Shaw, M.; Innes, J.; Patel, L.; Gecz, J.; Clayton-Smith, J.; Thomas, P. Interchromosomal insertional translocation at xq26.3 alters sox3 expression in an individual with xx male sex reversal. *J Clin Endocrinol Metab* **2015**, *100*, E815-820.

124. Moalem, S.; Babul-Hirji, R.; Stavropolous, D.J.; Wherrett, D.; Bagli, D.J.; Thomas, P.; Chitayat, D. Xx male sex reversal with genital abnormalities

associated with a de novo sox3 gene duplication. *Am J Med Genet A* **2012**, *158A*, 1759-1764.

125. Sutton, E.; Hughes, J.; White, S.; Sekido, R.; Tan, J.; Arboleda, V.; Rogers, N.; Knower, K.; Rowley, L.; Eyre, H., *et al.* Identification of sox3 as an xx male sex reversal gene in mice and humans. *J Clin Invest* **2011**, *121*, 328-341.

126. Tasic, V.; Mitrotti, A.; Riepe, F.G.; Kulle, A.E.; Laban, N.; Polenakovic, M.; Plaseska-Karanfilska, D.; Sanna-Cherchi, S.; Kostovski, M.; Gucev, Z. Duplication of the sox3 gene in an sry-negative 46,xx male with associated congenital anomalies of kidneys and the urinary tract: Case report and review of the literature. *Balkan J Med Genet* **2019**, *22*, 81-88.

127. Wilhelm, D.; Yang, J.X.; Thomas, P. Mammalian sex determination and gonad development. *Current topics in developmental biology* **2013**, *106*, 89-121.

128. Burkitt Wright, E.M.; Perveen, R.; Clayton, P.E.; Hall, C.M.; Costa, T.; Procter, A.M.; Giblin, C.A.; Donnai, D.; Black, G.C. X-linked isolated growth hormone deficiency: Expanding the phenotypic spectrum of sox3 polyalanine tract expansions. *Clin Dysmorphol* **2009**, *18*, 218-221.

129. Laumonnier, F.; Ronce, N.; Hamel, B.C.; Thomas, P.; Lespinasse, J.; Raynaud, M.; Paringaux, C.; Van Bokhoven, H.; Kalscheuer, V.; Fryns, J.P., *et al.* Transcription factor sox3 is involved in x-linked mental retardation with growth hormone deficiency. *Am J Hum Genet* **2002**, *71*, 1450-1455.

130. Hughes, J.; Piltz, S.; Rogers, N.; McAninch, D.; Rowley, L.; Thomas, P. Mechanistic insight into the pathology of polyalanine expansion disorders revealed by a mouse model for x linked hypopituitarism. *PLoS Genet* **2013**, *9*, e1003290.

131. Wong, J.; Farlie, P.; Holbert, S.; Lockhart, P.; Thomas, P.Q. Polyalanine expansion mutations in the x-linked hypopituitarism gene sox3 result in aggresome formation and impaired transactivation. *Front Biosci* **2007**, *12*, 2085-2095.

132. Rousseau, F.; Heitz, D.; Biancalana, V.; Blumenfeld, S.; Kretz, C.; Boue, J.; Tommerup, N.; Van Der Hagen, C.; DeLozier-Blanchet, C.; Croquette, M.F., *et al.* Direct diagnosis by DNA analysis of the fragile x syndrome of mental retardation. *N Engl J Med* **1991**, *325*, 1673-1681.

133. Stevanovic, M.; Lovell-Badge, R.; Collignon, J.; Goodfellow, P.N. Sox3 is an x-linked gene related to sry. *Hum Mol Genet* **1993**, *2*, 2013-2018.

134. Vencesla, A.; Barcelo, M.J.; Baena, M.; Quintana, M.; Baiget, M.; Tizzano, E.F. Marker and real-time quantitative analyses to confirm hemophilia b carrier diagnosis of a complete deletion of the f9 gene. *Haematologica* **2007**, *92*, 1583-1584.

135. Alatzoglou, K.S.; Azriyanti, A.; Rogers, N.; Ryan, F.; Curry, N.; Noakes, C.; Bignell, P.; Hall, G.W.; Littooij, A.S.; Saunders, D., *et al.* Sox3 deletion in mouse and human is associated with persistence of the craniopharyngeal canal. *J Clin Endocrinol Metab* **2014**, *99*, E2702-2708.

136. Helle, J.R.; Baroy, T.; Misceo, D.; Braaten, O.; Fannemel, M.; Frengen, E. Hyperphagia, mild developmental delay but apparently no structural brain anomalies in a boy without sox3 expression. *Am J Med Genet A* **2013**, *161A*, 1137-1142.

137. Tegelenbosch, R.A.; de Rooij, D.G. A quantitative study of spermatogonial multiplication and stem cell renewal in the c3h/101 f1 hybrid mouse. *Mutat Res* **1993**, *290*, 193-200.

138. Culty, M. Gonocytes, the forgotten cells of the germ cell lineage. *Birth Defects Res C Embryo Today* **2009**, *87*, 1-26.

139. Dutta, S.; Sengupta, P. Men and mice: Relating their ages. *Life Sci* **2016**, *152*, 244-248.

140. Nishimune, Y.; Aizawa, S.; Komatsu, T. Testicular germ cell differentiation in vivo. *Fertil Steril* **1978**, *29*, 95-102.

141. Shinohara, T.; Orwig, K.E.; Avarbock, M.R.; Brinster, R.L. Spermatogonial stem cell enrichment by multiparameter selection of mouse testis cells. *Proc Natl Acad Sci U S A* **2000**, *97*, 8346-8351.

142. de Rooij, D.G.; Grootegoed, J.A. Spermatogonial stem cells. *Current opinion in cell biology* **1998**, *10*, 694-701.

143. de Rooij, D.G.; Russell, L.D. All you wanted to know about spermatogonia but were afraid to ask. *J Androl* **2000**, *21*, 776-798.

144. Dym, M.; Fawcett, D.W. Further observations on the numbers of spermatogonia, spermatocytes, and spermatids connected by intercellular bridges in the mammalian testis. *Biol Reprod* **1971**, *4*, 195-215.

145. O'Donnell, L. Mechanisms of spermiogenesis and spermiation and how they are disturbed. *Spermatogenesis* **2014**, *4*, e979623.

146. Fayomi, A.P.; Orwig, K.E. Spermatogonial stem cells and spermatogenesis in mice, monkeys and men. *Stem Cell Res* **2018**, *29*, 207-214.

147. Verver, D.E.; van Pelt, A.M.; Repping, S.; Hamer, G. Role for rodent smc6 in pericentromeric heterochromatin domains during spermatogonial differentiation and meiosis. *Cell Death Dis* **2013**, *4*, e749.

148. de Rooij, D.G.; de Boer, P. Specific arrests of spermatogenesis in genetically modified and mutant mice. *Cytogenet Genome Res* **2003**, *103*, 267-276.

149. Aloisio, G.M.; Nakada, Y.; Saatcioglu, H.D.; Pena, C.G.; Baker, M.D.; Tarnawa, E.D.; Mukherjee, J.; Manjunath, H.; Bugde, A.; Sengupta, A.L., *et al.* Pax7 expression defines germline stem cells in the adult testis. *J Clin Invest* **2014**, *124*, 3929-3944.

150. Grisanti, L.; Falciatori, I.; Grasso, M.; Dovere, L.; Fera, S.; Muciaccia, B.; Fuso, A.; Berno, V.; Boitani, C.; Stefanini, M., *et al.* Identification of spermatogonial stem cell subsets by morphological analysis and prospective isolation. *Stem Cells* **2009**, *27*, 3043-3052.

151. Oatley, M.J.; Kaucher, A.V.; Racicot, K.E.; Oatley, J.M. Inhibitor of DNA binding 4 is expressed selectively by single spermatogonia in the male germline and regulates the self-renewal of spermatogonial stem cells in mice. *Biol Reprod* **2011**, *85*, 347-356.

152. Yoshida, S. Elucidating the identity and behavior of spermatogenic stem cells in the mouse testis. *Reproduction* **2012**, *144*, 293-302.

153. Hara, K.; Nakagawa, T.; Enomoto, H.; Suzuki, M.; Yamamoto, M.; Simons, B.D.; Yoshida, S. Mouse spermatogenic stem cells continually interconvert between equipotent singly isolated and syncytial states. *Cell stem cell* **2014**, *14*, 658-672.

154. Buaas, F.W.; Kirsh, A.L.; Sharma, M.; McLean, D.J.; Morris, J.L.; Griswold, M.D.; de Rooij, D.G.; Braun, R.E. Plzf is required in adult male germ cells for stem cell self-renewal. *Nature genetics* **2004**, *36*, 647-652.

155. Costoya, J.A.; Hobbs, R.M.; Barna, M.; Cattoretti, G.; Manova, K.; Sukhwani, M.; Orwig, K.E.; Wolgemuth, D.J.; Pandolfi, P.P. Essential role of plzf in maintenance of spermatogonial stem cells. *Nature genetics* **2004**, *36*, 653-659.

156. Schrans-Stassen, B.H.; van de Kant, H.J.; de Rooij, D.G.; van Pelt, A.M. Differential expression of c-kit in mouse undifferentiated and differentiating type a spermatogonia. *Endocrinology* **1999**, *140*, 5894-5900.

157. Tokuda, M.; Kadokawa, Y.; Kurahashi, H.; Marunouchi, T. Cdh1 is a specific marker for undifferentiated spermatogonia in mouse testes. *Biol Reprod* **2007**, *76*, 130-141.

158. Hofmann, M.C.; Braydich-Stolle, L.; Dym, M. Isolation of male germ-line stem cells; influence of gdnf. *Developmental biology* **2005**, *279*, 114-124.

159. Nakagawa, T.; Sharma, M.; Nabeshima, Y.; Braun, R.E.; Yoshida, S. Functional hierarchy and reversibility within the murine spermatogenic stem cell compartment. *Science* **2010**, *328*, 62-67.

160. Yoshida, S.; Takakura, A.; Ohbo, K.; Abe, K.; Wakabayashi, J.; Yamamoto, M.; Suda, T.; Nabeshima, Y. Neurogenin3 delineates the earliest stages of spermatogenesis in the mouse testis. *Developmental biology* **2004**, *269*, 447-458.

161. Nakagawa, T.; Nabeshima, Y.; Yoshida, S. Functional identification of the actual and potential stem cell compartments in mouse spermatogenesis. *Dev Cell* **2007**, *12*, 195-206.

162. Ikami, K.; Tokue, M.; Sugimoto, R.; Noda, C.; Kobayashi, S.; Hara, K.; Yoshida, S. Hierarchical differentiation competence in response to retinoic acid ensures stem cell maintenance during mouse spermatogenesis. *Development* **2015**, *142*, 1582-1592.

163. Hogarth, C.A.; Arnold, S.; Kent, T.; Mitchell, D.; Isoherranen, N.; Griswold, M.D. Processive pulses of retinoic acid propel asynchronous and continuous murine sperm production. *Biol Reprod* **2015**, *92*, 37.

164. Kamachi, Y.; Kondoh, H. Sox proteins: Regulators of cell fate specification and differentiation. *Development* **2013**, *140*, 4129-4144.

165. Reiprich, S.; Wegner, M. From cns stem cells to neurons and glia: Sox for everyone. *Cell Tissue Res* **2015**, *359*, 111-124.

166. Favaro, R.; Valotta, M.; Ferri, A.L.; Latorre, E.; Mariani, J.; Giachino, C.; Lancini, C.; Tosetti, V.; Ottolenghi, S.; Taylor, V.*, et al.* Hippocampal development and neural stem cell maintenance require sox2-dependent regulation of shh. *Nature neuroscience* **2009**, *12*, 1248-1256.

167. Shirakawa, T.; Yaman-Deveci, R.; Tomizawa, S.; Kamizato, Y.; Nakajima, K.; Sone, H.; Sato, Y.; Sharif, J.; Yamashita, A.; Takada-Horisawa, Y.*, et al.* An epigenetic switch is crucial for spermatogonia to exit the undifferentiated state toward a kit-positive identity. *Development* **2013**, *140*, 3565-3576.

168. Simons, B.D.; Clevers, H. Strategies for homeostatic stem cell self-renewal in adult tissues. *Cell* **2011**, *145*, 851-862.

169. La, H.M.; Hobbs, R.M. Mechanisms regulating mammalian spermatogenesis and fertility recovery following germ cell depletion. *Cell Mol Life Sci* **2019**, *76*, 4071-4102.

170. Sharma, M.; Srivastava, A.; Fairfield, H.E.; Bergstrom, D.; Flynn, W.F.; Braun, R.E. Identification of eomes-expressing spermatogonial stem cells and their regulation by plzf. *Elife* **2019**, *8*.

171. La, H.M.; Makela, J.A.; Chan, A.L.; Rossello, F.J.; Nefzger, C.M.; Legrand, J.M.D.; De Seram, M.; Polo, J.M.; Hobbs, R.M. Identification of dynamic undifferentiated cell states within the male germline. *Nat Commun* **2018**, *9*, 2819.

172. Gely-Pernot, A.; Raverdeau, M.; Celebi, C.; Dennefeld, C.; Feret, B.; Klopfenstein, M.; Yoshida, S.; Ghyselinck, N.B.; Mark, M. Spermatogonia differentiation requires retinoic acid receptor gamma. *Endocrinology* **2012**, *153*, 438-449.

173. Hobbs, R.M.; La, H.M.; Mäkelä, J.A.; Kobayashi, T.; Noda, T.; Pandolfi, P.P. Distinct germline progenitor subsets defined through tsc2-mtorc1 signaling. *EMBO Rep* **2015**, *In press*.

174. Sada, A.; Suzuki, A.; Suzuki, H.; Saga, Y. The rna-binding protein nanos2 is required to maintain murine spermatogonial stem cells. *Science* **2009**, *325*, 1394-1398.

175. Suzuki, H.; Ahn, H.W.; Chu, T.; Bowden, W.; Gassei, K.; Orwig, K.; Rajkovic, A. Sohlh1 and sohlh2 coordinate spermatogonial differentiation. *Developmental biology* **2012**, *361*, 301-312.

176. Zhou, Z.; Shirakawa, T.; Ohbo, K.; Sada, A.; Wu, Q.; Hasegawa, K.; Saba, R.; Saga, Y. Rna binding protein nanos2 organizes post-transcriptional buffering system to retain primitive state of mouse spermatogonial stem cells. *Dev Cell* **2015**, *34*, 96-107.

177. Busada, J.T.; Chappell, V.A.; Niedenberger, B.A.; Kaye, E.P.; Keiper, B.D.; Hogarth, C.A.; Geyer, C.B. Retinoic acid regulates kit translation during spermatogonial differentiation in the mouse. *Developmental biology* **2015**, *397*, 140-149.

178. Phillips, B.T.; Gassei, K.; Orwig, K.E. Spermatogonial stem cell regulation and spermatogenesis. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **2010**, *365*, 1663-1678.

179. Beumer, T.L.; Roepers-Gajadien, H.L.; Gademan, I.S.; Kal, H.B.; de Rooij, D.G. Involvement of the d-type cyclins in germ cell proliferation and differentiation in the mouse. *Biol Reprod* **2000**, *63*, 1893-1898.

180. Hobbs, R.M.; Seandel, M.; Falciatori, I.; Rafii, S.; Pandolfi, P.P. Plzf regulates germline progenitor self-renewal by opposing mtorc1. *Cell* **2010**, *142*, 468-479.

181. Hobbs, R.M.; Fagoonee, S.; Papa, A.; Webster, K.; Altruda, F.; Nishinakamura, R.; Chai, L.; Pandolfi, P.P. Functional antagonism between sall4 and plzf defines germline progenitors. *Cell stem cell* **2012**, *10*, 284-298.

182. Nagano, M.C.; Yeh, J.R. The identity and fate decision control of spermatogonial stem cells: Where is the point of no return? *Current topics in developmental biology* **2013**, *102*, 61-95.

183. Chakraborty, P.; Buaas, F.W.; Sharma, M.; Snyder, E.; de Rooij, D.G.; Braun, R.E. Lin28a marks the spermatogonial progenitor population and regulates its cyclic expansion. *Stem Cells* **2014**, *32*, 860-873.

184. Chan, F.; Oatley, M.J.; Kaucher, A.V.; Yang, Q.E.; Bieberich, C.J.; Shashikant, C.S.; Oatley, J.M. Functional and molecular features of the id4+ germline stem cell population in mouse testes. *Genes & development* **2014**, *28*, 1351-1362.

185. Chan, A.L.; La, H.M.; Legrand, J.M.D.; Makela, J.A.; Eichenlaub, M.; De Seram, M.; Ramialison, M.; Hobbs, R.M. Germline stem cell activity is sustained by sall4-dependent silencing of distinct tumor suppressor genes. *Stem Cell Reports* **2017**, *9*, 956-971.

186. de Rooij, D.G. Stem cells in the testis. *International journal of experimental pathology* **1998**, *79*, 67-80.

187. Sada, A.; Hasegawa, K.; Pin, P.H.; Saga, Y. Nanos2 acts downstream of glial cell line-derived neurotrophic factor signaling to suppress differentiation of spermatogonial stem cells. *Stem Cells* **2012**, *30*, 280-291.

188. McAninch, D.; Thomas, P. Identification of highly conserved putative developmental enhancers bound by sox3 in neural progenitors using chip-seq. *PLoS One* **2014**, *9*, e113361.

189. Rogers, N.; McAninch, D.; Thomas, P. Dbx1 is a direct target of sox3 in the spinal cord. *PLoS One* **2014**, *9*, e95356.

190. Yoshida, S.; Sukeno, M.; Nakagawa, T.; Ohbo, K.; Nagamatsu, G.; Suda, T.; Nabeshima, Y. The first round of mouse spermatogenesis is a distinctive program that lacks the self-renewing spermatogonia stage. *Development* **2006**, *133*, 1495-1505.

191. Kaucher, A.V.; Oatley, M.J.; Oatley, J.M. Neurog3 is a critical downstream effector for stat3-regulated differentiation of mammalian stem and progenitor spermatogonia. *Biol Reprod* **2012**, *86*, 164, 161-111.

192. Koubi, M.; Poplineau, M.; Vernerey, J.; N'Guyen, L.; Tiberi, G.; Garciaz, S.; El-Kaoutari, A.; Maqbool, M.A.; Andrau, J.C.; Guillouf, C., *et al.* Regulation of the positive transcriptional effect of plzf through a non-canonical ezh2 activity. *Nucleic Acids Res* **2018**, *46*, 3339-3350.

193. Filipponi, D.; Hobbs, R.M.; Ottolenghi, S.; Rossi, P.; Jannini, E.A.; Pandolfi, P.P.; Dolci, S. Repression of kit expression by plzf in germ cells. *Mol Cell Biol* **2007**, *27*, 6770-6781.

194. Song, H.W.; Wilkinson, M.F. Transcriptional control of spermatogonial maintenance and differentiation. *Semin Cell Dev Biol* **2014**, *30*, 14-26.

195. Rathke, C.; Baarends, W.M.; Awe, S.; Renkawitz-Pohl, R. Chromatin dynamics during spermiogenesis. *Biochim Biophys Acta* **2014**, *1839*, 155-168.

196. Hoghoughi, N.; Barral, S.; Vargas, A.; Rousseaux, S.; Khochbin, S. Histone variants: Essential actors in male genome programming. *J Biochem* **2018**, *163*, 97-103.

197. Ueda, J.; Harada, A.; Urahama, T.; Machida, S.; Maehara, K.; Hada, M.; Makino, Y.; Nogami, J.; Horikoshi, N.; Osakabe, A., *et al.* Testis-specific histone variant h3t gene is essential for entry into spermatogenesis. *Cell Rep* **2017**, *18*, 593-600.

198. McLean, C.Y.; Bristor, D.; Hiller, M.; Clarke, S.L.; Schaar, B.T.; Lowe, C.B.; Wenger, A.M.; Bejerano, G. Great improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **2010**, *28*, 495-501.

199. Buske, F.A.; Boden, M.; Bauer, D.C.; Bailey, T.L. Assigning roles to DNA regulatory motifs using comparative genomics. *Bioinformatics* **2010**, *26*, 860-866.

200. Kulakovskiy, I.V.; Vorontsov, I.E.; Yevshin, I.S.; Sharipov, R.N.; Fedorova, A.D.; Rumynskiy, E.I.; Medvedeva, Y.A.; Magana-Mora, A.; Bajic, V.B.; Papatsenko, D.A.*, et al.* Hocomoco: Towards a complete collection of transcription factor binding models for human and mouse via large-scale chip-seq analysis. *Nucleic Acids Res* **2018**, *46*, D252-D259.

201. Lovelace, D.L.; Gao, Z.; Mutoji, K.; Song, Y.C.; Ruan, J.; Hermann, B.P. The regulatory repertoire of plzf and sall4 in undifferentiated spermatogonia. *Development* **2016**, *143*, 1893-1906.

202. Sakaki-Yumoto, M.; Kobayashi, C.; Sato, A.; Fujimura, S.; Matsumoto, Y.; Takasato, M.; Kodama, T.; Aburatani, H.; Asashima, M.; Yoshida, N.*, et al.* The murine homolog of sall4, a causative gene in okihiro syndrome, is essential for embryonic stem cell proliferation, and cooperates with sall1 in anorectal, heart, brain and kidney development. *Development* **2006**, *133*, 3005-3013.

203. Liu, C.F.; Lefebvre, V. The transcription factors sox9 and sox5/sox6 cooperate genome-wide through super-enhancers to drive chondrogenesis. *Nucleic Acids Res* **2015**, *43*, 8183-8203.

204. Ohba, S.; He, X.; Hojo, H.; McMahon, A.P. Distinct transcriptional programs underlie sox9 regulation of the mammalian chondrocyte. *Cell Rep* **2015**, *12*, 229-243.

205. Siggers, T.; Gordan, R. Protein-DNA binding: Complexities and multi-protein codes. *Nucleic Acids Res* **2014**, *42*, 2099-2111.

206. Han, J.; Zhang, Z.; Wang, K. 3c and 3c-based techniques: The powerful tools for spatial genome organization deciphering. *Mol Cytogenet* **2018**, *11*, 21.

207. Ma, W.; Noble, W.S.; Bailey, T.L. Motif-based analysis of large nucleotide data sets using meme-chip. *Nat Protoc* **2014**, *9*, 1428-1450.

208. Ng, C.K.; Li, N.X.; Chee, S.; Prabhakar, S.; Kolatkar, P.R.; Jauch, R. Deciphering the sox-oct partner code by quantitative cooperativity measurements. *Nucleic Acids Res* **2012**, *40*, 4933-4941.

209. Li, R.; Vannitamby, A.; Zhang, J.G.; Fehmel, E.L.; Southwell, B.R.; Hutson, J.M. Oct4-gfp expression during transformation of gonocytes into spermatogonial stem cells in the perinatal mouse testis. *J Pediatr Surg* **2015**, *50*, 2084-2089.

210. Dann, C.T.; Alvarado, A.L.; Molyneux, L.A.; Denard, B.S.; Garbers, D.L.; Porteus, M.H. Spermatogonial stem cell self-renewal requires oct4, a factor downregulated during retinoic acid-induced differentiation. *Stem Cells* **2008**, *26*, 2928-2937.

211. Boskovic, A.; Torres-Padilla, M.E. How mammals pack their sperm: A variant matter. *Genes & development* **2013**, *27*, 1635-1639.

212. Huh, N.E.; Hwang, I.W.; Lim, K.; You, K.H.; Chae, C.B. Presence of a bi-directional s phase-specific transcription regulatory element in the promoter shared by testis-specific th2a and th2b histone genes. *Nucleic Acids Res* **1991**, *19*, 93-98.

213. Montellier, E.; Boussouar, F.; Rousseaux, S.; Zhang, K.; Buchou, T.; Fenaille, F.; Shiota, H.; Debernardi, A.; Hery, P.; Curtet, S.*, et al.* Chromatin-to-nucleoprotamine transition is controlled by the histone h2b variant th2b. *Genes & development* **2013**, *27*, 1680-1692.

214. Shinagawa, T.; Huynh, L.M.; Takagi, T.; Tsukamoto, D.; Tomaru, C.; Kwak, H.G.; Dohmae, N.; Noguchi, J.; Ishii, S. Disruption of th2a and th2b genes causes defects in spermatogenesis. *Development* **2015**, *142*, 1287-1292.

215. Singh, R.; Bassett, E.; Chakravarti, A.; Parthun, M.R. Replication-dependent histone isoforms: A new source of complexity in chromatin structure and function. *Nucleic Acids Res* **2018**, *46*, 8665-8678.

216. Langmead, B. Aligning short sequencing reads with bowtie. *Curr Protoc Bioinformatics* **2010**, *Chapter 11*, Unit 11 17.

217. Feng, J.; Liu, T.; Zhang, Y. Using macs to identify peaks from chip-seq data. *Curr Protoc Bioinformatics* **2011**, *Chapter 2*, Unit 2 14.

218. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. Weblogo: A sequence logo generator. *Genome Res* **2004**, *14*, 1188-1190.

219. Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The meme suite. *Nucleic Acids Res* **2015**, *43*, W39-49.

220. Grant, C.E.; Bailey, T.L.; Noble, W.S. Fimo: Scanning for occurrences of a given motif. *Bioinformatics* **2011**, *27*, 1017-1018.

221. Siepel, A.; Bejerano, G.; Pedersen, J.S.; Hinrichs, A.S.; Hou, M.; Rosenbloom, K.; Clawson, H.; Spieth, J.; Hillier, L.W.; Richards, S.*, et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **2005**, *15*, 1034-1050.

222. Oatley, J.M.; Oatley, M.J.; Avarbock, M.R.; Tobias, J.W.; Brinster, R.L. Colony stimulating factor 1 is an extrinsic stimulator of mouse spermatogonial stem cell self-renewal. *Development* **2009**, *136*, 1191-1199.

223. Davidson, S.; Starkey, A.; MacKenzie, A. Evidence of uneven selective pressure on different subsets of the conserved human genome; implications for the significance of intronic and intergenic DNA. *BMC Genomics* **2009**, *10*, 614.

224. Swanson, W.J.; Vacquier, V.D. The rapid evolution of reproductive proteins. *Nat Rev Genet* **2002**, *3*, 137-144.

225. Panhuis, T.M.; Clark, N.L.; Swanson, W.J. Rapid evolution of reproductive proteins in abalone and drosophila. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **2006**, *361*, 261-268.

226. Wilburn, D.B.; Swanson, W.J. From molecules to mating: Rapid evolution and biochemical studies of reproductive proteins. *J Proteomics* **2016**, *135*, 12-25.

227. Kvon, E.Z. Using transgenic reporter assays to functionally characterize enhancers in animals. *Genomics* **2015**, *106*, 185-192.

228. Zhang, G.; Shi, J.; Zhu, S.; Lan, Y.; Xu, L.; Yuan, H.; Liao, G.; Liu, X.; Zhang, Y.; Xiao, Y.*, et al.* Diseaseenhancer: A resource of human disease-associated enhancer catalog. *Nucleic Acids Res* **2018**, *46*, D78-D84.

229. Herz, H.M. Enhancer deregulation in cancer and other diseases. *Bioessays* **2016**, *38*, 1003-1015.

230. Carullo, N.V.N.; Day, J.J. Genomic enhancers in brain health and disease. *Genes (Basel)* **2019**, *10*.

231. Chang, Y.K.; Srivastava, Y.; Hu, C.; Joyce, A.; Yang, X.; Zuo, Z.; Havranek, J.J.; Stormo, G.D.; Jauch, R. Quantitative profiling of selective sox/pou pairing on hundreds of sequences in parallel by coop-seq. *Nucleic Acids Res* **2017**, *45*, 832-845.

232. Kim, B.R.; Coyaud, E.; Laurent, E.M.N.; St-Germain, J.; Van de Laar, E.; Tsao, M.S.; Raught, B.; Moghal, N. Identification of the sox2 interactome by bioid reveals ep300 as a mediator of sox2-dependent squamous differentiation and lung squamous cell carcinoma growth. *Mol Cell Proteomics* **2017**, *16*, 1864-1888.

233. Roux, K.J.; Kim, D.I.; Raida, M.; Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* **2012**, *196*, 801-810.

234. Schoenfelder, S.; Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet* **2019**, *20*, 437-455.

235. Petrascheck, M.; Escher, D.; Mahmoudi, T.; Verrijzer, C.P.; Schaffner, W.; Barberis, A. DNA looping induced by a transcriptional enhancer in vivo. *Nucleic Acids Res* **2005**, *33*, 3743-3750.

236. Auerbach, R.K.; Euskirchen, G.; Rozowsky, J.; Lamarre-Vincent, N.; Moqtaderi, Z.; Lefrancois, P.; Struhl, K.; Gerstein, M.; Snyder, M. Mapping accessible chromatin regions using sono-seq. *Proc Natl Acad Sci U S A* **2009**, *106*, 14926-14931.

237. Teytelman, L.; Thurtle, D.M.; Rine, J.; van Oudenaarden, A. Highly expressed loci are vulnerable to misleading chip localization of multiple unrelated proteins. *Proc Natl Acad Sci U S A* **2013**, *110*, 18602-18607.

238. Worsley Hunt, R.; Wasserman, W.W. Non-targeted transcription factors motifs are a systemic component of chip-seq datasets. *Genome Biol* **2014**, *15*, 412.

239. Kubota, H.; Brinster, R.L. Culture of rodent spermatogonial stem cells, male germline stem cells of the postnatal animal. *Methods Cell Biol* **2008**, *86*, 59-84.

240. Opel, M.; Lando, D.; Bonilla, C.; Trewick, S.C.; Boukaba, A.; Walfridsson, J.; Cauwood, J.; Werler, P.J.; Carr, A.M.; Kouzarides, T.*, et al.* Genome-wide studies of histone demethylation catalysed by the fission yeast homologues of mammalian lsd1. *PLoS One* **2007**, *2*, e386.

241. Fischle, W.; Tseng, B.S.; Dormann, H.L.; Ueberheide, B.M.; Garcia, B.A.; Shabanowitz, J.; Hunt, D.F.; Funabiki, H.; Allis, C.D. Regulation of hp1-chromatin binding by histone h3 methylation and phosphorylation. *Nature* **2005**, *438*, 1116-1122.

242. Ruthenburg, A.J.; Allis, C.D.; Wysocka, J. Methylation of lysine 4 on histone h3: Intricacy of writing and reading a single epigenetic mark. *Mol Cell* **2007**, *25*, 15-30.

243. Santos-Rosa, H.; Schneider, R.; Bannister, A.J.; Sherriff, J.; Bernstein, B.E.; Emre, N.C.; Schreiber, S.L.; Mellor, J.; Kouzarides, T. Active genes are tri-methylated at k4 of histone h3. *Nature* **2002**, *419*, 407-411.

244. Zentner, G.E.; Tesar, P.J.; Scacheri, P.C. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res* **2011**, *21*, 1273-1283.

245. Rada-Iglesias, A.; Bajpai, R.; Swigut, T.; Brugmann, S.A.; Flynn, R.A.; Wysocka, J. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **2011**, *470*, 279-283.

246. Calo, E.; Wysocka, J. Modification of enhancer chromatin: What, how, and why? *Mol Cell* **2013**, *49*, 825-837.

247. Creyghton, M.P.; Cheng, A.W.; Welstead, G.G.; Kooistra, T.; Carey, B.W.; Steine, E.J.; Hanna, J.; Lodato, M.A.; Frampton, G.M.; Sharp, P.A.*, et al.* Histone h3k27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **2010**, *107*, 21931-21936.

248. Buenrostro, J.D.; Wu, B.; Chang, H.Y.; Greenleaf, W.J. Atac-seq: A method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol* **2015**, *109*, 21 29 21-29.

249. Lieberman-Aiden, E.; van Berkum, N.L.; Williams, L.; Imakaev, M.; Ragoczy, T.; Telling, A.; Amit, I.; Lajoie, B.R.; Sabo, P.J.; Dorschner, M.O.*, et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **2009**, *326*, 289-293.

250. Hakim, O.; Misteli, T. Snapshot: Chromosome confirmation capture. *Cell* **2012**, *148*, 1068 e1061-1062.

251. Garbuzov, A.; Pech, M.F.; Hasegawa, K.; Sukhwani, M.; Zhang, R.J.; Orwig, K.E.; Artandi, S.E. Purification of gfralpha1+ and gfralpha1- spermatogonial stem cells reveals a niche-dependent mechanism for fate determination. *Stem Cell Reports* **2018**, *10*, 553-567.

252. Overton, P.M.; Meadows, L.A.; Urban, J.; Russell, S. Evidence for differential and redundant function of the sox genes dichaete and soxn during cns development in drosophila. *Development* **2002**, *129*, 4219-4228.

253. Wegner, M. From head to toes: The multiple facets of sox proteins. *Nucleic Acids Res* **1999**, *27*, 1409-1420.

254. Gu, B.; Posfai, E.; Rossant, J. Efficient generation of targeted large insertions by microinjection into two-cell-stage mouse embryos. *Nat Biotechnol* **2018**, *36*, 632-637.

255. Raymond, C.S.; Murphy, M.W.; O'Sullivan, M.G.; Bardwell, V.J.; Zarkower, D. Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. *Genes & development* **2000**, *14*, 2587-2595.

256. De Grandi, A.; Calvari, V.; Bertini, V.; Bulfone, A.; Peverali, G.; Camerino, G.; Borsani, G.; Guioli, S. The expression pattern of a mouse doublesex-related gene is consistent with a role in gonadal differentiation. *Mech Dev* **2000**, *90*, 323-326.

257. Lindeman, R.E.; Gearhart, M.D.; Minkina, A.; Krentz, A.D.; Bardwell, V.J.; Zarkower, D. Sexual cell-fate reprogramming in the ovary by dmrt1. *Curr Biol* **2015**, *25*, 764-771.

258. Zhang, T.; Oatley, J.; Bardwell, V.J.; Zarkower, D. Dmrt1 is required for mouse spermatogonial stem cell maintenance and replenishment. *PLoS Genet* **2016**, *12*, e1006293.

259. Green, C.D.; Ma, Q.; Manske, G.L.; Shami, A.N.; Zheng, X.; Marini, S.; Moritz, L.; Sultan, C.; Gurczynski, S.J.; Moore, B.B.*, et al.* A comprehensive roadmap of murine spermatogenesis defined by single-cell rna-seq. *Dev Cell* **2018**, *46*, 651-667 e610.

260. Diepenbruck, M.; Waldmeier, L.; Ivanek, R.; Berninger, P.; Arnold, P.; van Nimwegen, E.; Christofori, G. Tead2 expression levels control the subcellular distribution of yap and taz, zyxin expression and epithelial-mesenchymal transition. *J Cell Sci* **2014**, *127*, 1523-1536.

261. Sawada, A.; Kiyonari, H.; Ukita, K.; Nishioka, N.; Imuta, Y.; Sasaki, H. Redundant roles of tead1 and tead2 in notochord development and the regulation of cell proliferation and survival. *Mol Cell Biol* **2008**, *28*, 3177-3189.

262. Gu, P.; Goodwin, B.; Chung, A.C.; Xu, X.; Wheeler, D.A.; Price, R.R.; Galardi, C.; Peng, L.; Latour, A.M.; Koller, B.H.*, et al.* Orphan nuclear receptor lrh-1 is required to maintain oct4 expression at the epiblast stage of embryonic development. *Mol Cell Biol* **2005**, *25*, 3492-3505.

263. Stergiopoulos, A.; Politis, P.K. Nuclear receptor nr5a2 controls neural stem cell fate decisions during development. *Nat Commun* **2016**, *7*, 12230.

264. Seitz, C.; Huang, J.; Geiselhoringer, A.L.; Galbani-Bianchi, P.; Michalek, S.; Phan, T.S.; Reinhold, C.; Dietrich, L.; Schmidt, C.; Corazza, N.*, et al.* The orphan nuclear receptor lrh-1/nr5a2 critically regulates t cell functions. *Sci Adv* **2019**, *5*, eaav9732.

265. Meinsohn, M.C.; Morin, F.; Bertolin, K.; Duggavathi, R.; Schoonjans, K.; Murphy, B.D. The orphan nuclear receptor liver homolog receptor-1 (nr5a2) regulates ovarian granulosa cell proliferation. *J Endocr Soc* **2018**, *2*, 24-41.

266. Ballow, D.; Meistrich, M.L.; Matzuk, M.; Rajkovic, A. Sohlh1 is essential for spermatogonial differentiation. *Developmental biology* **2006**, *294*, 161-167.

267. Pederick, D.T.; Richards, K.L.; Piltz, S.G.; Kumar, R.; Mincheva-Tasheva, S.; Mandelstam, S.A.; Dale, R.C.; Scheffer, I.E.; Gecz, J.; Petrou, S.*, et al.* Abnormal cell sorting underlies the unique x-linked inheritance of pcdh19 epilepsy. *Neuron* **2018**, *97*, 59-66 e55.

268. Ravassard, P.; Chatail, F.; Mallet, J.; Icard-Liepkalns, C. Relax, a novel rat bhlh transcriptional regulator transiently expressed in the ventricular proliferating zone of the developing central nervous system. *J Neurosci Res* **1997**, *48*, 146-158.

269. Cau, E.; Gradwohl, G.; Fode, C.; Guillemot, F. Mash1 activates a cascade of bhlh regulators in olfactory neuron progenitors. *Development* **1997**, *124*, 1611-1621.

270. Sommer, L.; Ma, Q.; Anderson, D.J. Neurogenins, a novel family of atonal-related bhlh transcription factors, are putative mammalian neuronal determination genes that reveal progenitor cell heterogeneity in the developing cns and pns. *Mol Cell Neurosci* **1996**, *8*, 221-241.

271. Hamilton, B.A.; Yu, B.D. Modifier genes and the plasticity of genetic networks in mice. *PLoS Genet* **2012**, *8*, e1002644.

272. Klein, A.D.; Ferreira, N.S.; Ben-Dor, S.; Duan, J.; Hardy, J.; Cox, T.M.; Merrill, A.H., Jr.; Futerman, A.H. Identification of modifier genes in a mouse model of gaucher disease. *Cell Rep* **2016**, *16*, 2546-2553.

273. Hummel, K. Difference in response of mice of different strains to human pregnancy urine. *Endocrinology* **1942**, *30*, 74-76.

274. Muhlbock, O.; Tengbergen, E.; Van Maurik, G. Differences in the gonadotrophic hormone content and the cytology in the anterior lobe of the hypophysis of male mice of various strains. *Acta Endocrinol (Copenh)* **1952**, *9*, 48-58.

275. Plant, T.M.; Marshall, G.R. The functional significance of fsh in spermatogenesis and the control of its secretion in male primates. *Endocr Rev* **2001**, *22*, 764-786.

276. Jiang, P.; Selvaraj, V.; Deng, W. Differentiation of embryonic stem cells into oligodendrocyte precursors. *J Vis Exp* **2010**.

277. Procter, A.M.; Phillips, J.A., 3rd; Cooper, D.N. The molecular genetics of growth hormone deficiency. *Hum Genet* **1998**, *103*, 255-272.

278. Raverot, G.; Lejeune, H.; Kotlar, T.; Pugeat, M.; Jameson, J.L. X-linked sex-determining region y box 3 (sox3) gene mutations are uncommon in men with idiopathic oligoazoospermic infertility. *J Clin Endocrinol Metab* **2004**, *89*, 4146-4148.

279. Harrill, A.H.; McAllister, K.A. New rodent population models may inform human health risk assessment and identification of genetic susceptibility to environmental exposures. *Environ Health Perspect* **2017**, *125*, 086002.

280. Zeise, L.; Bois, F.Y.; Chiu, W.A.; Hattis, D.; Rusyn, I.; Guyton, K.Z. Addressing human variability in next-generation human health risk assessments of environmental chemicals. *Environ Health Perspect* **2013**, *121*, 23-31.

281. Rahim, N.G.; Harismendy, O.; Topol, E.J.; Frazer, K.A. Genetic determinants of phenotypic diversity in humans. *Genome Biol* **2008**, *9*, 215.

282. Greenbaum, M.P.; Iwamori, N.; Agno, J.E.; Matzuk, M.M. Mouse tex14 is required for embryonic germ cell intercellular bridges but not female fertility. *Biol Reprod* **2009**, *80*, 449-457.

283. Greenbaum, M.P.; Yan, W.; Wu, M.H.; Lin, Y.N.; Agno, J.E.; Sharma, M.; Braun, R.E.; Rajkovic, A.; Matzuk, M.M. Tex14 is essential for intercellular bridges and fertility in male mice. *Proc Natl Acad Sci U S A* **2006**, *103*, 4982-4987.

284. Pelling, M.; Anthwal, N.; McNay, D.; Gradwohl, G.; Leiter, A.B.; Guillemot, F.; Ang, S.L. Differential requirements for neurogenin 3 in the development of pomc and npy neurons in the hypothalamus. *Developmental biology* **2011**, *349*, 406-416.

285. Gradwohl, G.; Dierich, A.; LeMeur, M.; Guillemot, F. Neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. *Proc Natl Acad Sci U S A* **2000**, *97*, 1607-1611.

286. Kan, L.; Israsena, N.; Zhang, Z.; Hu, M.; Zhao, L.R.; Jalali, A.; Sahni, V.; Kessler, J.A. Sox1 acts through multiple independent pathways to promote neurogenesis. *Developmental biology* **2004**, *269*, 580-594.

287. Gunesdogan, U.; Jackle, H.; Herzig, A. Histone supply regulates s phase timing and cell cycle progression. *Elife* **2014**, *3*, e02443.

288. Mendiratta, S.; Gatto, A.; Almouzni, G. Histone supply: Multitiered regulation ensures chromatin dynamics throughout the cell cycle. *J Cell Biol* **2019**, *218*, 39-54.

289. Marino-Ramirez, L.; Jordan, I.K.; Landsman, D. Multiple independent evolutionary solutions to core histone gene regulation. *Genome Biol* **2006**, *7*, R122.

290. Elkouris, M.; Balaskas, N.; Poulou, M.; Politis, P.K.; Panayiotou, E.; Malas, S.; Thomaidou, D.; Remboutsika, E. Sox1 maintains the undifferentiated state of cortical neural progenitor cells via the suppression of prox1-mediated cell cycle exit and neurogenesis. *Stem Cells* **2011**, *29*, 89-98.

291. Zane, L.; Chapus, F.; Pegoraro, G.; Misteli, T. Hihimap: Single-cell quantitation of histones and histone posttranslational modifications across the cell cycle by high-throughput imaging. *Mol Biol Cell* **2017**, *28*, 2290-2302.