Identification and Validation of SOXB1 Bound Developmental Enhancers

THE UNIVERSITY of ADELAIDE

Ella Paulina Thomson

B.Sc. (Biomedical Science), Honours (Genetics)

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

July 2019

Table Of Contents

This page intentionally left blank.

Abstract

Enhancers are regions of non-coding DNA bound by transcription factors that influence gene expression, and are essential for the precise regulation of embryonic development. Many enhancers have been identified through reporter assays and bioinformatic techniques, but these are unable to show the functional contribution of the enhancer to endogenous expression. The SOXB1 proteins, expressed within the neural progenitor and spermatogonial stem cell populations are important TFs, the absence of which leads to developmental defects in both humans and mice. NES is an intermediate filament protein, and is thought to be regulated via a SOXB1 bound enhancer, commonly used in transgenic mice models to direct expression to NPCs. Through CRISPR mediated deletion of the *Nes* enhancer we show it is active from 9.5dpc in the CNS, and is responsible for up to 70% of endogenous *Nes* expression. Further, we identify possible *trans* activity of the enhancer, a new field of mammalian enhancer research. SOX3 ChIP-Seq identified a region upstream of the Wnt-receptor gene, *Fzd3* that appeared to be an enhancer. Subsequent deletion via CRISPR confirmed enhancer activity in the CNS. Combined with LacZ reporter mouse models we show the enhancer directs expression specifically to the floor plate, an important region for axon guidance for which FZD3 is required. SOX3s role within the postnatal testes is largely unknown. Through ChIP-Seq we identified putative enhancers bound by SOX3, and present evidence that it is important in the regulation of the complex chromatin reorganisation that occurs during spermatogenesis. Together, this data presents new insights into the role of SOXB1s in development as well as highlighting the importance of functional validation of putative enhancers which can be achieved through CRISPR.

Statement of Authorship

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 also 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. I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Ella Paulina Thomson

Acknowledgements

Firstly, I would like to thank Paul Thomas for being a great supervisor. I have thoroughly enjoyed my time in the lab, and your help and guidance over the years has been wonderful.

The members of the Thomas lab and MLS made coming into the lab everyday so enjoyable. Thanks to Murray Whitelaw for being a great co-supervisor. James, thank you for always being there to work through any problem. Sandie, thank you for all the microinjection sessions and animal work help. Dale, thank you for answering my many questions and all your help with the ChIP. Adi and Chee, thank you for all your work on the SOX projects.

To everyone at morning tea and afternoon drinks – James, Dale, Sandie, Mel, Dan, Louise, Emily, Chan, Connor, Stef, Ruby, Chee and Adi - thanks for all the chats, scientific and not!

Mel, Ruby and Louise – I couldn't have asked for better friends to share this experience with, I hope you know how much your friendship means to me.

Deb and Dad, thank you for your unconditional love and support always.

Hayden, thank you for absolutely everything– you're the best.

Mum, I hope you would have been proud.

This page intentionally left blank.

Chapter One

Introduction

1 Enhancers

In order to tightly control gene expression throughout development, regulatory elements are essential. Previously only thought of as 'junk DNA, the non-coding genome was thought to be unimportant when compared to the 2% of DNA that encoded proteins (Ohno 1972). Within the last few decades it has become apparent that the non-coding genome is home to many regulatory elements, such as silencers, non-coding RNA, and enhancers (ENCODE Project Consortium 2012). Enhancers are cis-regulatory elements comprised of regions of DNA, usually 50-1500 base pairs (bp) in length, to which specific transcription factors (TF) bind, allowing interaction with the target gene promoter, influencing transcription to allow for both spatial and temporal control of gene expression.

1.1 Transcription Factor Binding

Enhancers are comprised of varying numbers of transcription factor binding sites (TFBS), which are recognised by an individual TF or TF family. Highly specific TF binding allows enhancers to precisely control gene expression; for example a gene only required in the developing brain will not be activated by a brain specific TF in a developing limb (Shlyueva et al. 2014). Figure 1, (Krijger & de Laat 2016) demonstrates the various ways in which each enhancer can control target genes. There can either be many enhancers for a single tissue, which would allow for it to be activated (or repressed) at different timepoints, or various enhancers which will be activated by tissue-specific TFs for expression within multiple body regions.

The exact mechanism of how TFs are able to interact with DNA to regulate expression is a complex process which is still not completely understood. Firstly, for a TF to recognise and bind a DNA motif, it must contain a DNA binding domain. There are many different protein binding DNA domains found across the transcription factor families; these include homeodomains (HD), helix-turn-helix (HTH) domains and high mobility group box (HMG) domains. The focus of this thesis, the SOX family of TFs contain the HMG binding domain, and this will be discussed in more detail in 2.1.

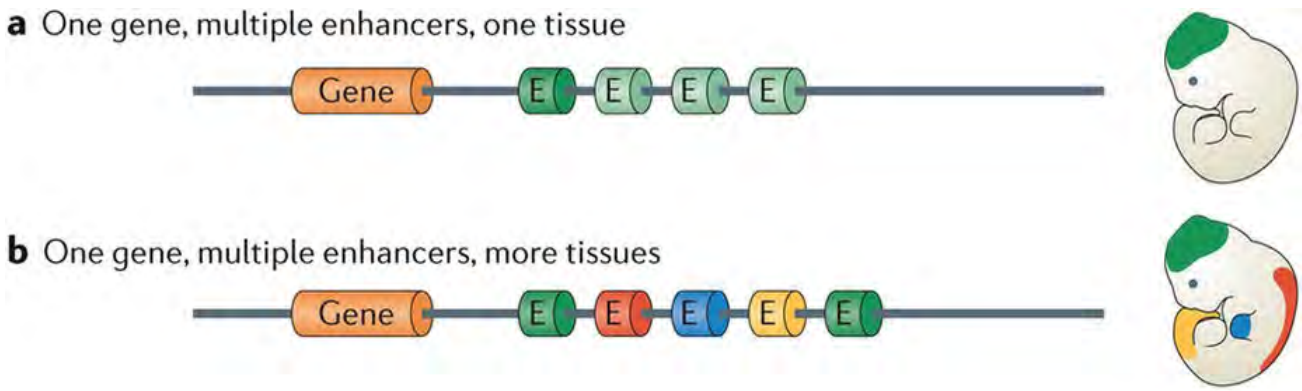


Figure 1 (Krijger & de Laat 2016)

Various types of enhancers exist within the genome. A. A single gene may contain a single or many distinct enhancers to control expression within one tissue type or expression region. Or, as shown in (B) a gene may have multiple enhancers which are individually able to direct expression to target tissue, for example one enhancer may direct expression to a limb, while another may direct expression to the brain.

The DNA binding domain of each TF will recognise a specific DNA motif, usually a 8-20 bp region to which the TF directly binds (Halford & Marko 2004). Whilst these motifs can be found repeated throughout the genome, it has been estimated by the ENCODE consortium that the vast majority of DNA that matches these specific motifs will never be bound by a TF (Yesudhas et al. 2017). The environment in which these DNA motifs are within is crucial, as will be discussed in section 1.7 where the epigenetic modifications and chromatin accessibility is vitally important.

There are three main mechanisms proposed for how a TF finds its correct motif; sliding, hopping, and intersegmental transfer. The model *lac* repressor TF is an example of a sliding mechanism. Molecular dynamics simulations have shown the TF 'sliding' along the DNA searching for the specific binding sites within the major grooves of DNA (Hammar et al. 2012).

The hopping mechanism is where the protein 'hops' between DNA motifs, governed by electrostatic interactions. It is thought that while this is slower that siding, the binding affinity between the protein and DNA is stronger (Yesudhas et al. 2017). In the intersegmental transfer mechanism model, the TF is bound at one DNA motif, and to move to the next the intervening DNA will loop to the second site, followed by a second binding event (Yesudhas et al. 2017; Hippel & Berg 1989)

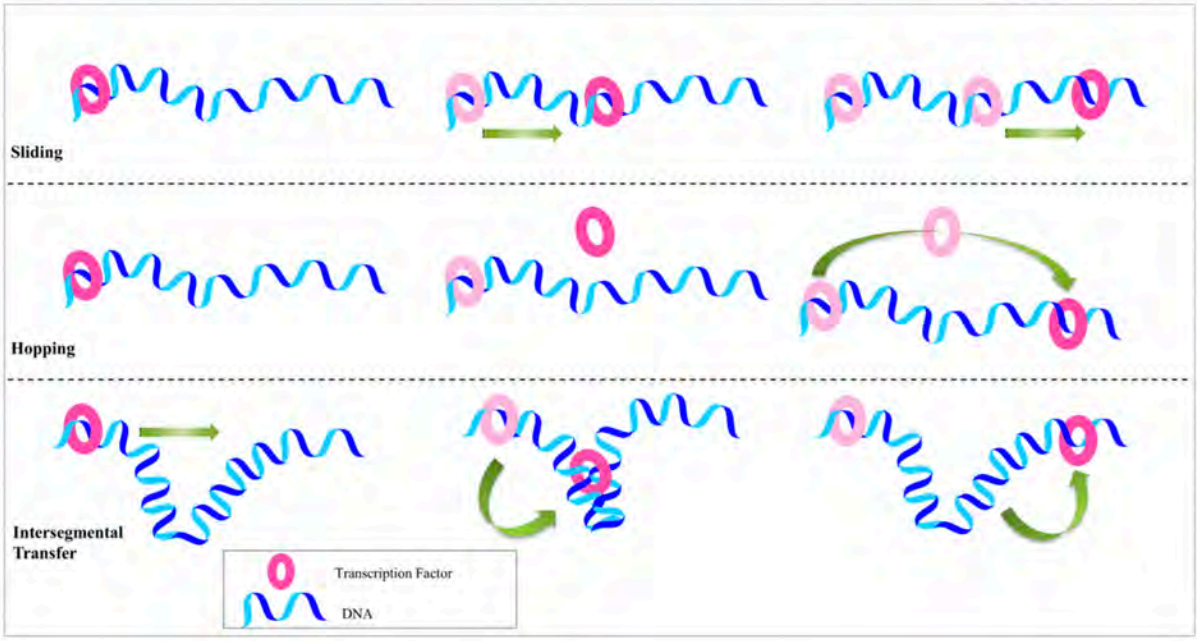


Figure 2 (Yesudhas et al. 2017)

The three protein:DNA recognition mechanisms. The pink ring symbolises the transcription factor on the DNA, in the first diagram (sliding), the TF 'slides' along the DNA and moves between the bases without dissociating until it reaches a TFBS to which it binds. The second (hopping) shows the TF continually disassociates and reassociates with the DNA until bound. The third (intersegmental transfer) is where the TF is moved along via the DNA moving, through bending or looping mechanisms.

1.2 Models of Enhancer Activity

Like TF DNA binding, there are multiple mechanisms/models for enhancer activity. There are three proposed models, two of which are well characterised, and a third that has been recently proposed. The enhanceosome model, billboard model and TF collective model will all be discussed below.

1.2.1 The Enhanceosome Model

The enhanceosome model is defined by a requirement for co-operative binding by all TFs to the enhancer for activity to occur. Co-operative TF binding relies on the architectural remodelling of the chromatin environment. The archetypal enhanceosome is that of the interferon-β gene, activated upon viral infection in mammalian cells, which establishes the antiviral state (Falvo et al. 2000). A 55 bp enhancer upstream of the promoter contains DNA binding motifs for many individual TFs, all of which need to bind co-operatively (Merika & Thanos 2001; Panne et al. 2007). Once all TFs have bound, the nucleosomes flanking the enhancers are acetylated, initiating chromatin remodelling. This allows access for the TATA-binding protein to the TATA box, a DNA binding motif, and RNA polymerase to the promoter, initiating transcription (Agalioti et al. 2000). If any of the individual TFs are unable to bind, or are not present, the interferon-β anti-viral pathway will not be activated. These types of enhancers are more susceptible to single base pair changes or small deletions; indeed, in the IF-β enhancer almost every base is bound by a TF (Agalioti et al. 2000).

1.2.2 The Billboard/Information Display Model

Whilst the enhanceosome model appears to work well for 'all-or-nothing' responses, such as those involved in immunity,the billboard/information display model is much more forgiving to modifications, and is more often found at developmental genes, such as those within this thesis (Kulkarni & Arnosti 2003; Struhl 2001). While these enhancers are also comprised of different TFBS, expression is not reliant upon co-operative binding, but rather each TF can control expression in isolation. This allows for a much greater flexibility of expression, and is also thought to have an inbuilt redundancy, as individual TFs can each control expression of the target without having to rely on the entire suite of TFs to be bound at each motif. Unsurprisingly, billboard enhancers are more forgiving to sequence changes. An example of a partially redundant developmental enhancer was described by Gao and Finkelstein (1998) who studied the *otd* expression within *Drosophila*. Two separate enhancer regions were identified that controlled *otd* expression. Each was able to produce the correct patterning on its own, however not at the same intensities of wildtype. When both regions are present, wildtype levels of expression are seen- indicating redundancy between the two regions bound by different TFs.

1.2.3 The TF Collective Model

The third and most recently proposed model of enhancer activity is the TF collective model. Like the billboard model, it has a greater amount of flexibility in regards to sequence changes, and activity is not abolished by loss of co-operative binding (Spitz & Furlong 2012). The differences between the two models lies in the combination of both protein:DNA, and protein:protein interactions occurring at the enhancer, and not all TFs need to be present or bound for full activity to occur (Junion et al. 2012; Khoueiry et al. 2017). For example, if a particular TFs binding site is lost, the protein may still be able to be recruited to the active enhancer if the other members of the collective are present, through protein:protein interactions (Uhl et al. 2016). Because of this lack on constraint of TFBS conservation and specificity, these enhancers are not always highly conserved and have increased the difficulty in predicting enhancer function and importance through DNA sequence alone.

An example of a TF collective enhancer is found within the *Drosophila* distal-less regulatory element, which controls two developmental processes; abdominal repression and thoracic activation of gene expression in leg precursor cells. Uhl et al and colleagues (2016) generated several mutations which altered the TFBS of the enhancer and showed that abdominal repression was extremely robust in regards to sequence changes, although abdominal activation was less so. To demonstrate the TF collective ability for protein:protein interactions, a HOX factor (one of the TFs) containing a DNA-binding domain mutation was still able to induce the thoracic repression without binding directly to the DNA, consistent with the transcription factor collective model.

1.3 Enhancer-Promoter Interactions

For an enhancer to impact target gene expression, it is assumed that it must first come into contact with its cognate promoter. When the enhancer is located closely, and proximal to its target gene it can be simple to see how this can occur. However, many enhancers are not located close to their target promoter, and are often found intronically, intergenically, within other genes, and have even been discovered up to 1 Mb away from their target promoter (Mora et al. 2016; Pennacchio et al. 2013; Lettice et al. 2003). TF binding to the enhancer region results in the recruitment of various co-factors and other proteins, enabling the enhancer to regulate the expression of the gene of interest. However, once the TF has bound to the enhancer, it is not clear how it interacts with the promoter. Whilst there are various hypotheses for how this interaction may occur, including DNA looping, tracking, and chromatin decompaction, or a combination these, there may still be ways in which they are interacting that are yet to be identified. The current hypotheses for TF:enhancer complex interaction with promoter regions are discussed below.

1.3.1 Looping

The most widespread hypothesis for how these interactions are made is through looping of the intervening regions of DNA. This allows the enhancer and promoter to meet, aided by various other scaffolding proteins. One of the first lines of evidence supporting the looping hypothesis came from studies of the β-globin gene, and its regulation via the locus control region (LCR) enhancer (Tolhuis et al. 2002). This study used 3C technology (Dekker et al. 2002), which utilises formaldehyde to cross-link proteins to nearby DNA. The frequency of interactions between the DNA regions can then be analysed. This method was able to show the interaction of enhancer DNA with the transcribed globin genes in erythroid cells, where the enhancer was active, but not with the repressed globin genes (Tolhuis et al. 2002).

Aside from the physical interaction between the enhancer and promoter, other factors also appear to be essential to drive enhancer-activated gene expression. One of these, RNA polymerase, must be present for transcription to occur as evidenced by Vernimmen et al (2007). However, the point has been raised as to whether the physical co-localisation of enhancers and promoters is the underlying mechanism of enhancer activated expression, or if it is simply the consequence of RNA Pol II interacting with both enhancer and promoter (Bulger & Groudine 2010).

1.3.2 Tracking Theory

For close range or proximal enhancers, the tracking theory of enhancer action is more prominent. In this model, once the DNA is bound by the TF, the intervening chromatin is modified in a way to bring the enhancer and promoter into close physical contact without looping out the DNA (Benabdallah & Bickmore 2015; Engel et al. 2008; Hatzis & Talianidis 2002).

This model of enhancer activity has not been shown to be active in any distal or long-range enhancers however, so whilst it may be occurring in these closely linked enhancer promoter interactions, it is most likely not the only way of regulating enhancer activity within the genome.

1.3.3 Chromatin Decompaction Theory

The third model of enhancer activity, chromatin decompaction theory, occurs when the chromatin between the enhancer and promoter undergoes decompaction, rather than bringing the two elements into closer physical contact. Benabdallah et al. (2017) studied the murine sonic hedgehog (*Shh)* gene and its distant neural enhancers approximately 100 and 780 kb away. By using 3D-FISH probes at the enhancer and promoter, they analysed the distances between the fluorescent signals, theorising that if looping were to occur, the two signals would become closer together upon activation. However, at two of the four brain enhancers, SBE4 and SBE6, these distances actually increased, while the others remained the same. The enhancers that increased in distance were also those that showed chromatin methylation consistent with active enhancers - suggesting that the decompaction of the chromatin was involved in the enhancer promoter interaction. It is not yet known if this form of enhancer action is occurring at other loci, but it will be interesting to see if the chromatin decompaction is a widespread way in which distal enhancers can regulate their target gene expression.

1.4 Super enhancers

Enhancers can be as small as a few TF binding motifs over 50bp, to a collection of many motifs that are bound by a variety of TFs over a large distance. Large regions of DNA with a high concentration of TF binding controlling highly expressed genes are called 'super-enhancers'. Super enhancers are often associated with higher than normal amount of chromatin modification (reviewed in section 1.7) as well as binding of BRD4, and the multiprotein complex Mediator, both important transcriptional coactivators (Pott & Lieb 2015; Olley et al. 2018).

Research by Whyte et. al. (2013) identified the first group of super-enhancers at key cell identity genes. They were able to show that the 'master' transcription factors, OCT4, SOX2 and NANOG bind to enhancers within embryonic stem cells that are then able to control much of the gene expression changes and cell fate decisions of these early cells.

The role of super-enhancers in the many aspects of development became apparent within patients with Cornelia de Lange-like syndrome. This congenital syndrome is unsurprisingly similar to Cornelia de Lange syndrome, which presents as a global developmental delay via generalised de-regulation of developmental genes affecting many body systems (Kline et al. 2007). It is most often associated with mutations in genes important for the cohesin complex. Cohesin is an essential protein complex, with its main role being mediating the separation of sister chromatids during cell division. It has also been shown to be important for DNA damage repair pathways, as well as regulating gene expression (Peters et al. 2008). Mutations in genes involved in the cohesion complex, such as NIPBL can result in the loss of cohesin dependent chromatin looping important for transcription (Boyle et al. 2015; Boudaoud et al. 2017, Deardorff 2007). A recent gene to be identified as causing Cornelia de Lang-like syndrome is BRD4 (Olley et al. 2018). BRD4 is not

part of the cohesin complex, but is important for 'reading' the acetylation status of histone associated lysine, and is found at high levels at many super-enhancers (Hnisz et al. 2013). Analysis of BRD4 and NIPBL through ChIP demonstrated they bound at similar locations at super enhancers. In the BRD4 mutants however, this binding was reduced, indicating it was unable to regulate the super-enhancers when mutated (Olley et al. 2018). The authors propose that CdL-like syndrome and CdL syndrome are a single disorder of generalised super-enhancer dysfunction.

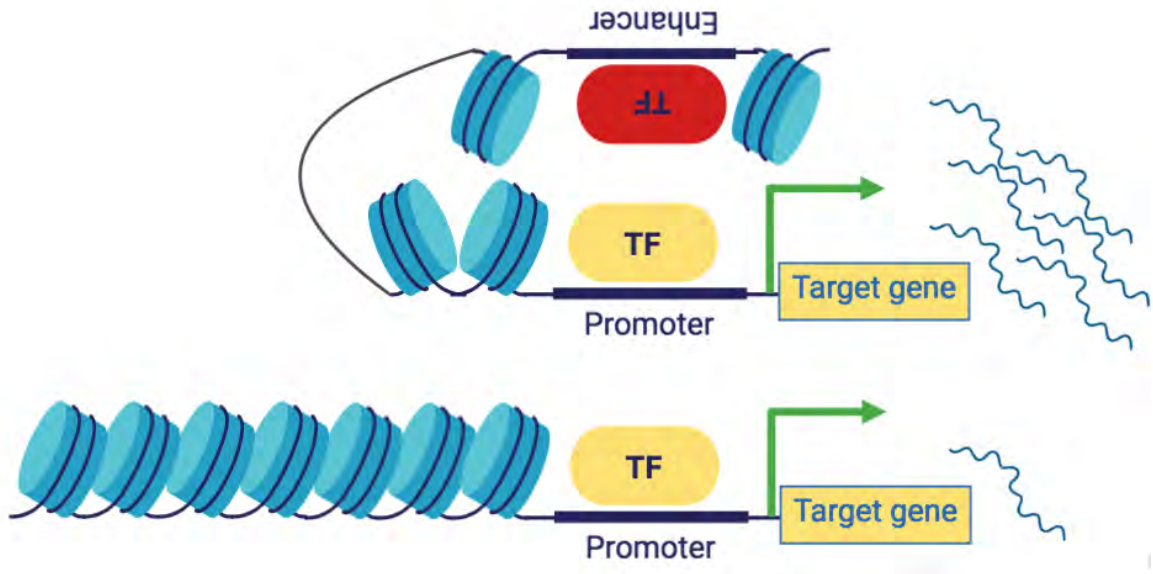
1.5 *Trans* interactions

Usually, mammalian enhancers are studied for their ability to regulate the expression of genes in *cis*. In *Drosophila* however, the presence of *trans* enhancer-promoter interactions have been well characterised, and this phenomenon has been termed transvection. The first reported gene to undergo this process was *Ubx*, whereby regulatory regions located on a non-protein producing copy of *Ubx* (*Ubx1*) could still influence the expression of the functional *Ubx* gene, found on a separate chromosome (Micol et al. 1990; Lewis et al. 1954). Since this discovery, many more genes in Drosophila have been shown to have both *tran*s and *cis* interactions between enhancers and promoters, suggesting this may be a commonly used mechanism, at least within flies (Sipos & Gyurkovics 2005; Mahmoudi et al. 2002; Mellert & Truman 2012; Kennison & Southworth 2002). While these studies are able to show partial genetic complementation, they were unable to quantify the levels of *cis* and *trans* interactions occurring at the loci, often due to the complexity of regulatory sequences. In an effort to understand this, a set of experiments were performed using fluorescent reporter alleles with complete enhancer-less and promoter-less constructs of the eye-specific enhancer, GMR (Bateman et al. 2012). Interestingly, when only *cis* interactions could occur, the Drosophila GMR enhancer could drive expression of GFP to the eye-discs as expected. However, when only *trans* interactions were allowed to occur, this expression became uneven and variegated suggesting it was not acting in the same manner as the *cis*-enhancer due to reduced interactions. Through a series of genetic crosses, the authors were able to show that the promoters in *cis* and *trans* will both compete for the activity of an enhancer, but expression is *cis*-biased possibly due to the proximity of the enhancer and promoter. It is not known whether these results are generally applicable to all enhancer-promoter interactions but they do provide an elegant experimental approach to test *cis* and *trans* interactions (Bateman et al. 2012).

Recently, a study of the IgH super-enhancer has shown its capability for *trans* interactions within mouse models (Le Noir et al. 2017). The IgH super enhancer is a well characterised region, with many individual enhancer elements, high levels of enhancer RNA (eRNA) transcription (reviewed in section 1.6) and high levels of active chromatin marks (Le Noir et al. 2017; Pinaud et al. 2011). The authors generated transgenic mice carrying either one copy of the IgH regulatory region at the endogenous location, or three copies on another chromosome. Through analysis with 3D FISH the interactions between the transgene and endogenous Igh alleles could be visualised, and demonstrated that *trans* interactions can occur independently of chromosomal context.

Although some mammalian systems have now been shown to interact in *trans* (Alvarez-Dominguez et al. 2017; Le Noir et al. 2017), this has not yet been shown to be a common feature like in *Drosophila*. In fact, a study within mammalian immune cells has proposed that there is no evidence of trans interactions during their development, and that their lineage identity is purely driven by *cis* interactions (Johanson et al. 2018). As further studies are performed in mammalian systems, it will be interesting to see how pervasive the *trans* interactions are, and if they are restricted to certain tissues or cell types.

A



B

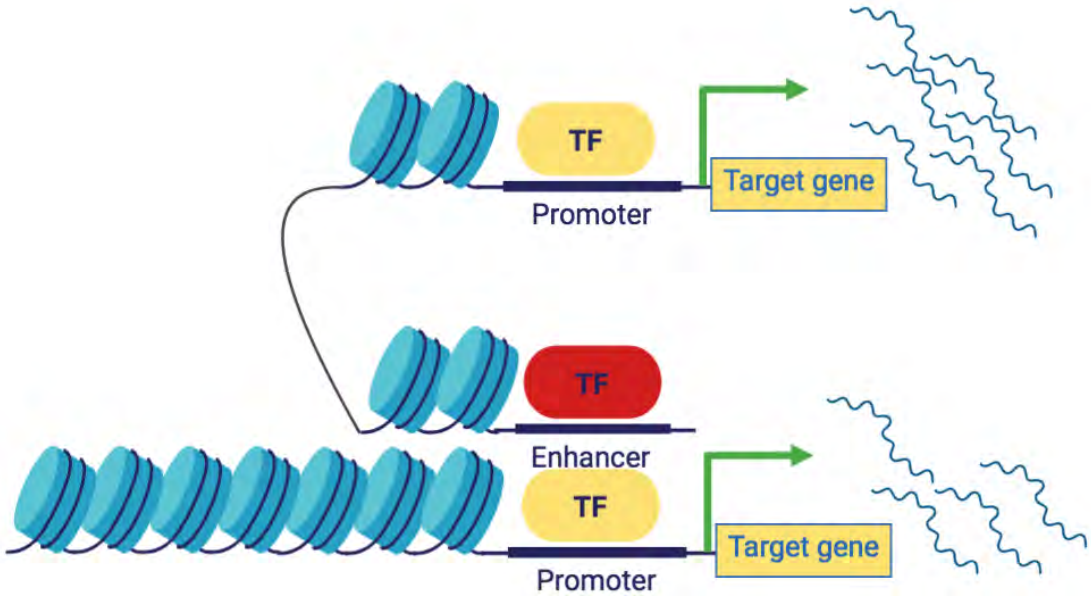


Figure 3 Proposed model of *Trans* interactions

Trans interactions have not been well characterized in mammalian systems. The enhancer promoter interactions are usually thought to occur in *cis* as shown in (A) whereby the enhancer one allele interacts with the promoter, influencing transcription (top allele). If the enhancer is lost (bottom allele), the promoter may be able to generate basal levels of expression, but this does not reach wildtype. In (B), the theory of *trans* interactions states that the TF bound enhancer of one allele (top) can interact with the enhancer-lacking allele (bottom) to influence expression, increasing transcriptional output.

1.6 E-RNAs

The β-globin enhancer is one of the most well studied enhancer regions, and in 1990 RNAs transcribed at this enhancer were identified (Collis et al. 1990; Tuan et al. 1992). Transcripts have since been noted at many enhancers and are termed enhancer RNAs (eRNAs). During the FANTOM and ENCODE projects it was confirmed that RNA synthesis occurs at almost all active enhancers sites (Lizio et al. 2015; ENCODE Project Consortium 2012; Feingold et al. 2004). There is currently no clear answer as to whether eRNAs are a functional component of an enhancer, or simply an artefact due to the high concentration of RNA Pol II which is present at active enhancers. These transcripts are usually quite unstable, and are not spliced like a protein coding or functional RNA (Smith & Shilatifard 2014). It has been proposed that eRNAs are important in keeping the enhancer within an active and open chromatin state to allow for TF binding at the enhancer, or to facilitate enhancer promoter looping mechanisms (Liu 2017, Li et al. 2013).

1.7 Enhancer States & Their Chromatin Signatures

When DNA is not being actively transcribed, it is usually tightly bound by nucleosomes. Nucleosomes are formed of four core histones (H2A, H2B, H3, H4) around which the DNA is "wrapped" (Kornberg 1977). Nucleosomes are then bound by linker histones, condensing the DNA further and making it inaccessible to most TFs (Zhou et al. 1998). However, TFs need to have a way of accessing bound chromatin to enable enhancer activation. The histone variants H2AZ, and H3.3 have been found to be enriched at enhancer regions, and these are thought to be less stable, allowing for easier displacement of nucleosomes, allowing TF binding to occur (Jin & Felsenfeld 2007). Indeed, it has been shown that H2AZ knock-down in mouse ES cells leads to the mis-regulation of developmental and pluripotency genes, possibly due to the inability of TF binding to occur allowing tissue specific expression (Creyghton et al. 2008).

Adams and Workman et al. (1995) were able to show that whilst a single TF is often insufficient to displace nucleosomes to allow accessible chromatin, the co-operation of a group of TF is able to overcome this threshold, allowing DNA binding and gene activation. However, another way in which the chromatin can become accessible is through the action of a pioneer TF. Pioneer TFs have the ability to bind nucleosome condensed DNA before the enhancer is active, creating accessible chromatin for future TFs to bind when necessary. An example of this is the binding of the FOXA TFs to liver specific enhancer regions. Here it has been shown to displace the linker proteins between the nucleosomes, allowing for tissue specific TFs to bind and activate transcription during development (Chaya et al. 2001; Iwafuchi-Doi et al. 2016).

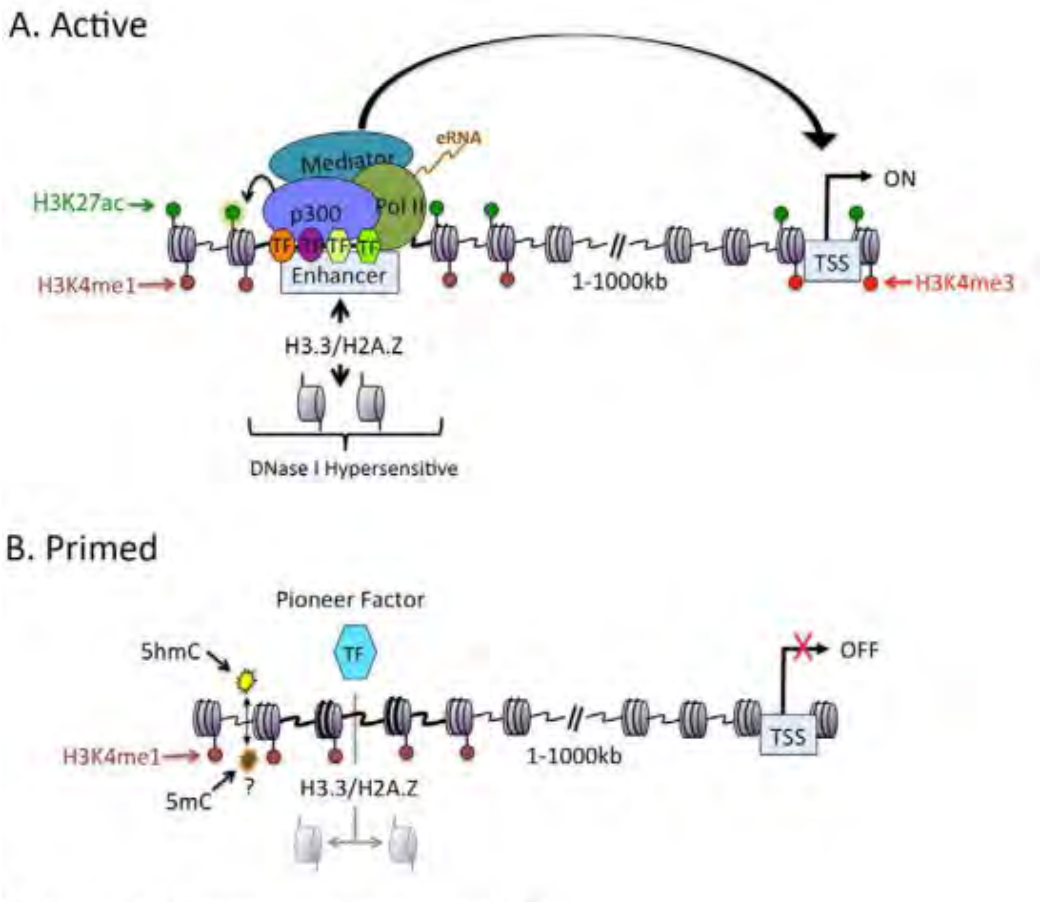


Figure 4 modified from (Calo & Wysocka 2013) Epigenetic modification at different stages of enhancer activity. At active enhancers (A) there is high levels of H3K27ac, as well as the mediator complex, RNA polymerase and p300 binding; these often produce e-RNAs. Primed enhancers (B) are characterized by H3K4me1, as well as the presence of pioneer transcription factors which are able to bind.

Histone modifications have become one of the most common ways to identify enhancers, and to determine their current states within tissues and at various time points. The H3K4me1 mark is one of the most commonly assessed modifications, and was found to be associated with enhancers in a pioneering study by Heintzman (2007) as part of the large ENCODE project. Whilst H3K4me1 marks are found at both active and primed enhancers, they have also been identified at promoters and covering broad 5' regions of actively transcribed genes, so this should be used with other data to identify the enhancer state (Calo & Wysocka 2013). It is thought that H3k4me1 methylation can help to promote the incorporation of the H2AZ nucleosomes which, as previously discussed, aid in chromatin remodelling and allow access by specific TFs. This mark is lost after enhancer decommissioning (Calo & Wysocka 2013).

Active enhancers are those that have accessible chromatin, are bound by H3K4me1, and have specific TF binding leading to gene expression. These enhancers are enriched for the acetylation mark H3K27ac, which is able to differentiate between active and poised enhancers unlike H3K4me1 (Creyghton et al. 2010). The ability for enhancers to activate transcription is dependent upon their ability to bind co-factors, such as p300. These are histone acetyltransferases able to produce epigenetic modifications thought to destabilise the nucleosomes, allowing the surrounding chromatin flexibility to generate enhancer promoter interactions (Calo & Wysocka 2013). It has also been shown that acetylation of Histone H3 (H3K64ac and H3K122ac) are also markers of a new class of active enhancers that lacks H3K27ac (Pradeepa et al. 2016). These histone modifications are thought to function by modifying the chromatin structure to destabilise the nucleosome and stimulate transcription (Pradeepa 2017). Identification of these histone marks will help to identify different types of enhancers, and by continuing to expand our

knowledge of histone marks more enhancers will be able to be found via these methods.

Table 1 modified from Mora et al (2016) Histone variants and their associated enhancer states

Histone modification/histone variant	Enzymes	Main observation
H3K4me1/2	KMT2C/2D (MLL3/4); KMT7 (SET7/9)	Active, intermediate and poised enhancers
H3K9ac	KAT2A/B (GCN5/PCAF)	Active enhancers
H3K14ac	KAT2A/B (Gcn5/PCAF); KAT3A/3B (p300/CBP); KAT6A (MYST3)	Active enhancers
H3K27ac	KAT3A/3B (p300/CBP)	Active enhancers
H3K36me3	KMT3A (SET2)	Active enhancers
H3K56ac	KAT3A/3B (p300/CBP)	Active enhancers
H4K16ac	KAT8 (MOF)	Active enhancers
H3K9me3	KMT1E (SETDB1)	Poised enhancers
H3K27me3	KMT6 (EZH2)	Poised enhancers
H2A.Z/H2A.Zac	KAT5 (TIP60)	Poised and active enhancers
H3.3	–	Poised and active enhancers

1.8 Identifying Enhancers

Before undertaking functional analyses of enhancers and their activity on gene regulation, they must first be identified. One of the most common ways in which this is performed is through ChIP-Seq experiments (Johnson et al. 2007). This involves an antibody specific to the transcription factor being studied added to the DNA to bind to the protein wherever it may be within the genome. These regions of DNA that have been bound by the antibody are cross-linked through incorporation of formaldehyde, and digested into smaller regions of DNA which can be sequenced. The sequenced regions can then be mapped to the genome to identify where the TF of interest was initially bound. Bioinformatic tools can then be used to identify common sequences within the identified regions, which can be compared to known motifs of TFs through various algorithms such as MEME-ChIP (Machanick & Bailey 2011) and RSAT peak motifs (Thomas-Chollier et al. 2012). To further validate putative binding sites, the peak data can be overlapped with other data to help increase the probability of finding active or biologically significant regions (Bailey et al. 2013; Bailey & Machanick 2012). As mentioned previously, there are different histone modifications associated with varying enhancer states, and by using ChIP-seq datasets of these modifications within a particular tissue or cell type, the enhancer state can be inferred. Other TFs that are known to bind simultaneously, or as partner factors can also be used.

It is generally thought that the more conserved region of DNA is, the higher the likelihood it will be biologically significant. Through both PhyloP and PhastCons algorithms (Siepel et al. 2005), it is possible to assign each DNA bound region a score which determines it conservation across placental mammals. By using these scores in enhancer identification studies, it helps to narrow down often very large (>10,000) numbers of sequences to only those that are highly conserved. It is

important to note that not all enhancers or biologically meaningful regions will be highly conserved or vice versa, however this is a useful tool when using large datasets.

Whilst ChIP-seq will identify where a TF of interest is binding, it will not give any information as to which promoter it is regulating. Often, it is assumed that this will be the nearest gene if intergenic, or if intronic the gene in which it is located. While this often true, there appears to be many cases where genes are not regulated by their closest enhancers. It has been estimated that only 7% of enhancers that loop to contact a promoter are located proximally to the target gene (Sanyal et al. 2012). By assessing the chromatin loops formed when the enhancer interacts with other regions of DNA, a more comprehensive picture of regulation can be determined, and this is achieved by chromatin conformation capture technology.

Chromatin conformation capture technology allows the quantification of interacting DNA regions within specific cells or tissue types (Gavrilov et al. 2009). This is accomplished through cross-linking of proteins with formaldehyde, followed by restriction enzyme digestion, and subsequent random ligation of the fragments. Interactions between cross-linked fragments are favoured over non-cross-linked, and these can then be quantified through PCR. It is important to note that through random chance many fragments will interact, so these must be able to reach statistical significance before any results can be inferred. The original incarnation of this method was 3C, in which two interacting regions can be assessed, such as a putative enhancer and its cognate promoter (Dekker et al. 2002). Further technologies have been built upon this method including 4C (one known candidate region, and unknown targets) (Zhao et al. 2006), 5C (all interacting regions within a certain region) (Dostie et al. 2006), and Hi-C (all interacting regions genome-wide) (van Berkum et al. 2010).

Incorporating this method with ChIP-Seq has also been done, leading to ChIA-Pet (chromatin immunoprecipitation with paired end tagging), which is able to identify long range and interchromosomal interactions between chromatin regions. An example of this was used in a study by McAninch (2014) in which an existing RNAPol II dataset was overlapped with a SOX3 ChIP-Seq dataset to determine interacting regions (Zhang et al. 2013). This was able to produce an interaction map identifying a region within the *Tex14* gene which shows interactions with 263 other promoters and enhancers, implicating SOX3 within a large transcriptional hub with both *cis* and *trans* regulation of target genes.

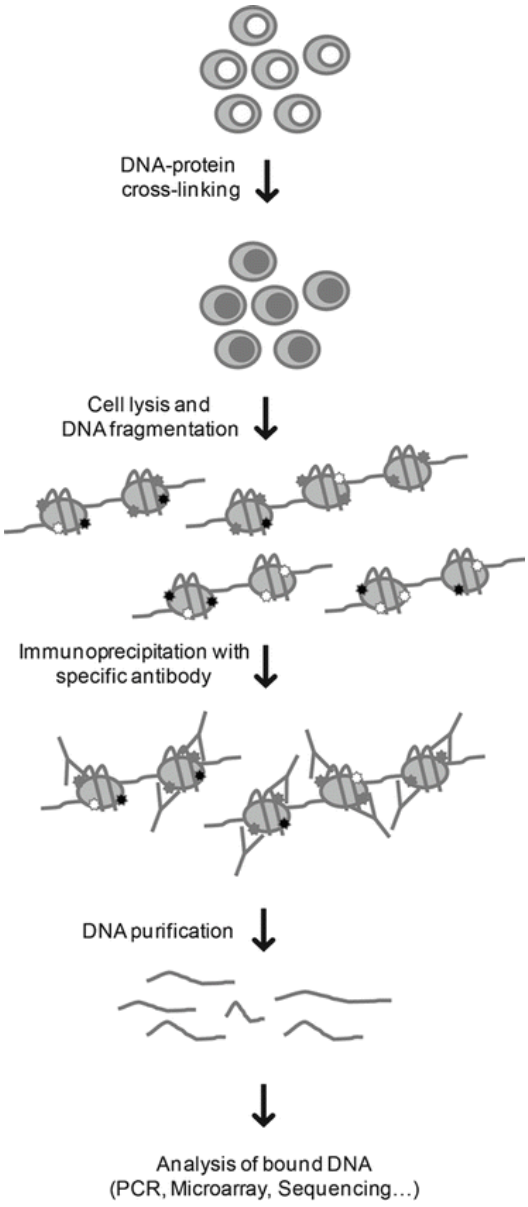
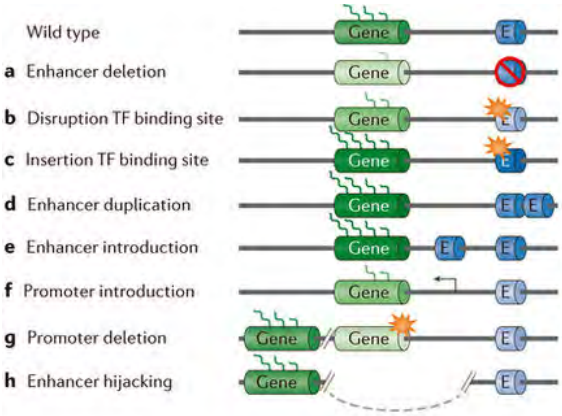


Figure 4 modified from Rodriguez-Ubreva (2013). The process of ChIP-Seq. DNA bound the protein being studied is cross-linked with formaldehyde. The cells are then lysed and DNA I fragmented into smaller pieces, before an antibody corresponding to the protein of interest is added. The DNA that is bound by an antibody is the purified, and these smaller fragments of DNA can be sequenced and mapped to the genome to determine where the protein is binding the DNA.

1.9 Enhancers and Disease

The importance of understanding enhancers, not just for controlling gene expression, but as regions of disease causing mutations, is an important area of research. The identification of these is often hampered, as many genome wide association studies (GWAS) and genome sequencing is often only performed on exomes (Gibson 2018). Although these regions are statistically much more likely to contain a disease-causing allele, and only looking at the exome reduces the cost significantly, cis-coding regions that contribute to disease will be missed when using this approach.

The introduction of large-scale datasets such as the Gnomad database is helping to identify single nucleotide polymorphisms (SNPSs) and allelic variants that are present within the healthy population, through the combination of whole genome and exome sequencing (Lek et al. 2016, Karczewski 2019). This data can be used to analyse when mutations are found within a disease patient, whether this is correlated with the disease or if it is just natural variation. There are currently over 100,000 and exome sequences, and 15,000 whole genome sequences, and this number is ever-increasing as more datasets are added regularly.



Disease	Affected gene	Enhancer	Refs
β-Thalassaemia	β-globin genes	LCR	3,4
HPE	*SHH*	SBE2	36
PDD2	*SHH*	ZRS	35,122
Lung adenocarcinoma	*MYC*	3′~450 kb SE	124
T-ALL	*TAL1*	NA	121
α-Thalassaemia	α-Globin genes	α-Globin enhancers	46
α-Thalassaemia*	α-Globin genes/*NME4*	α-Globin enhancers	44
Burkitt lymphoma	*MYC*	IgH enhancer	116,117

Nature Reviews | Molecular Cell Biology

Figure 5 (Krijger & de Laat 2016)

There are many ways in which enhancers can cause disease. If an enhancer has been deleted, gene expression may not occur, or if an individual binding site for a TF is mutated or lost, expression may be lost or reduced leading to disease. Enhancers can also be duplicated, or potentially introduced leading to disruption in gene expression. Enhancer hijacking has also been seen in some cancers whereby genomic rearrangements lead to enhancers activating oncogenic genes.

In response to this large amount of data being generated, and in an attempt to annotate human disease enhancers, the Human Enhancer Disease Database (HEDD) has recently been built and made freely available (Wang et al. 2017). This database incorporates information from enhancer studies (ENCODE, FANTOM5), disease studies (DisGeNET, MalaCards), GWAS studies, histone modifications (UCSC) and TF Binding Sites (UCSC).

As enhancers are responsible for such a large proportion of the finely controlled gene regulation during embryonic development, it is not surprising that the loss or mutation of these can lead to disease. However, it can be difficult to link enhancer mutations to diseases for a number of reasons. One of these is that many sequencing efforts or GWAS studies only focus on the exome, and oftentimes an enhancer does not control its nearest neighbouring gene.

One congenital embryonic disorder that has been associated with a particular group of enhancers is holoprosencephaly. Holoprosencephaly is a brain malformation where the forebrain midline fails to undergo cleavage to form the bilateral cerebral hemispheres (Golden 1999). This disorder affects 1 in 250 foetuses (Matsunaga & Shiota 1977), but only 1 in 8000 live births (Leoncini et al. 2008), as it is often fatal before birth.

Sonic hedgehog (*Shh*) is an essential gene which among other functions has been shown to be important in forming the ventral midline structures in the developing central nervous system (Chiang et al. 1996). When *Shh* levels are reduced to 50% in humans, holoprosencephaly is evident, however mouse models are less sensitive to this mutation, requiring homozygous loss of the gene to elicit the same effect (Roessler et al. 1996; Chiang et al. 1996). Three enhancer regions were identified in the 35Kb upstream of *Shh* (Epstein et al. 1999) and this was expanded to a total of six discrete enhancers within 1 Mb of the promoter (Jeong et al. 2006). Through

enhancer trap assays utilising BAC clones, the upstream genomic DNA was assessed for reporter activity in mouse embryo forebrains. The discovery of these regions shows that there is complex and often redundant regulation of *Shh* from long-range and close-range enhancers, which when displaced can cause expression defects leading to holoprosencephaly. This is based on the evidence of two patients with chromosomal rearrangements at 7q36, where *Shh* is shown to map (Lettice et al. 2002; Albuisson et al. 2011). This rearrangement was not found to be within any coding region, so it was hypothesised that these rearrangements have interfered with the enhancers leading to disease.

1.10 Validating Enhancers

Within the last two decades, our ability to generate bioinformatic data has greatly expanded. Whilst this has increased our knowledge about enhancers, and has improved our ability to identify them, there is a wealth of putative enhancers and regulatory regions that have not yet been functionally analysed. The technology available to assess the biological significance of putative enhancers has not yet caught up with the amount of data being produced. However, recent advances in genome editing technology including CRISPR in animal models has proved to be an exciting prospect for functional analysis.

1.10.1 Traditional Techniques

Historically, the most popular way to assess the functionality of enhancers was to use reporter assays. By placing the enhancer in question upstream of a minimal promoter and a reporter, such as luciferase, it can be transfected into a cell line that expresses the TFs thought to bind, and a read-out of activity can be assessed through quantification of the luciferase gene expression. This is done either episomally, or by integrating the reporter gene construct chromosomally. It has recently been found through a high throughput assay examining over 2000 liver specific putative enhancers, that for the most consistent and biologically reproducible data, chromosomal integration gives the best results (Inoue et al. 2017).

Even better than using a cell line reporter assay, is the use of LacZ reporter assays within developing mouse embryos. This technique was first pioneered by Mansour et al. (1990) to show the expression of *int-2* was recapitulated by LacZ in the developing embryo. This uses the same general principle as the cell line reporter assay, but the plasmid construct is randomly integrated the chromosome of a single-cell zygote by microinjection. This ensures that all cells contain a copy of the 'test enhancer' and that this will be bound by the endogenous TFs to produce a similar result to that of the endogenous gene expression. After transient embryo dissection, a β-galactosidase reaction is performed on the embryos, and all regions in which the LacZ is expressed will develop a deep blue pigment. The VISTA enhancer browser is an example of a large-scale effort to experimentally validate thousands of putative enhancers through these LacZ reporter assays in mice (Visel et al. 2007), and this has since expanded into assessing cardiac specific enhancers. This is still a widely used technique that can produce high confidence results showing if the region is sufficient for expression. However, this needs to be performed in multiple lines due to the random integration of the transgene into the

genome which can produce differing results depending on the chromatin accessibility of the region.

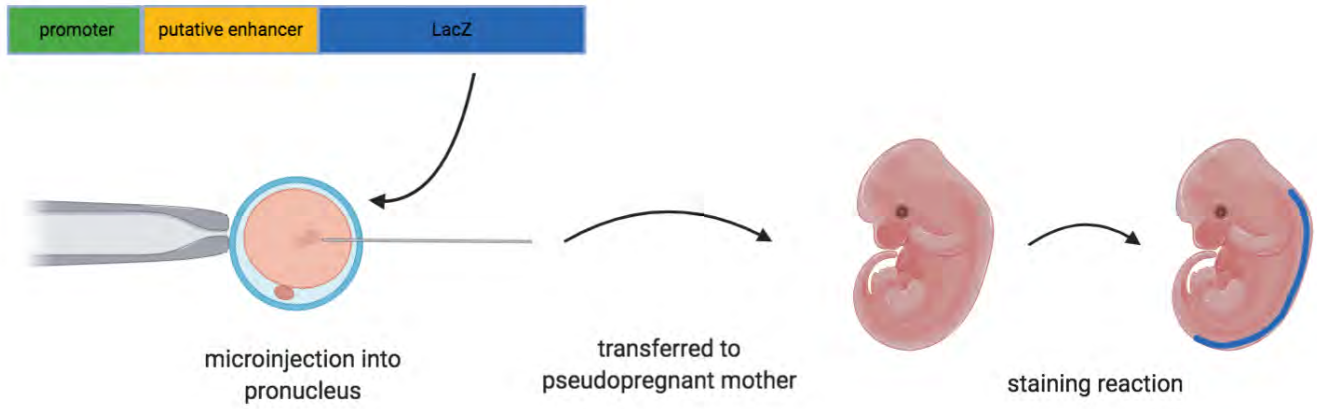


Figure 6

Generation of LacZ transgenic mice for enhancer analysis. To test the activity of a putative enhancer it can be cloned upstream of the LacZ gene, and driven by a promoter. The DNA is microinjected into a fertilized zygote where it can randomly incorporate into the genome. The injected zygotes are transferred into a pseudopregnant foster mother and allowed to develop. Embryos are dissected at a timepoint when the enhancer is thought to be active, and are stained with an X-Gal solution. If the enhancer is active, the LacZ will be expressed, and will stain the embryo blue in the tissue it is expressed. In the diagram, the blue staining can be seen in the neural tube. The enhancer's activity can be insinuated from the location of the staining.

1.10.2 CRISPR/Cas9 Technology

Over the last 5 years, the field of genetic modification has grown considerably, largely due to the development of CRISPR/Cas9 technology. First discovered as a bacterial defence system in 1998 (van Belkum et al.), it wasn't until 2013 that this system was adapted for use in mammalian systems to create specific mutations which has revolutionized genome editing (Cong et al. 2013). For a comprehensive review of CRISPR and its use in mouse models, see Appendix 1 for the publication entitled 'CRISPR Genome Editing in Mice'.

1.10.2.1 Bacterial Defence System

The use of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) to manipulate genomes has been developed based on the knowledge of the innate immunity system of the same name present in most prokaryotes. This system was discovered by van Belkum and colleagues(1998), when the CRISPR array of small DNA sequences located between repeat elements was found to correlate with regions of DNA in bacterial pathogens such as phage or virus DNA (Bolotin et al. 2005). The finding that these regions of DNA, known as protospacers could be created after challenge with a pathogen, and then later be used to protect against further infection (Barrangou et al. 2007) was an important piece of knowledge that allowed the CRISPR system to be repurposed for genetic editing applications.

1.10.2.2 Use in mammalian systems

A significant advance in CRISPR technology occurred when Cong and colleagues (2013) were able to design a system incorporating the elements of the type II CRISPR system into an easily modifiable plasmid. Instead of two separate transcripts, the pre-crRNA and tracrRNA were combined into a single transcript, known as a single guide (sgRNA). This RNA molecule includes a 20 bp targeting sequence, which is able to direct Cas9 nuclease to the genomic site for mutagenesis.

Whilst many regions of the genome can be targeted, the largest restriction with this system is the protospacer adjacent motif (PAM; corresponding to NGG for SpCas9), that must be at the 3' end of the sequence, and is essential for the initial binding of Cas9 nuclease. Sternberg et al. (2014) showed this by analysing the binding of Cas9 to target sequences, demonstrating the nuclease binds preferentially to PAM sequences within the genome, and rapidly dissociates from those not containing it. Only once the Cas9 has bound to the PAM is the DNA able to interrogate the gRNA for sequence complementarity (Sternberg et al. 2014).

Once the double strand break has been generated at the target locus, various methods are available to modify the DNA. For the introduction of insertions or deletions, the non-homologous end joining repair mechanism native to the cell can be exploited (Ran et al. 2013). This will generally result in small deletions and is often employed to create frameshift mutations within a gene to disrupt function.

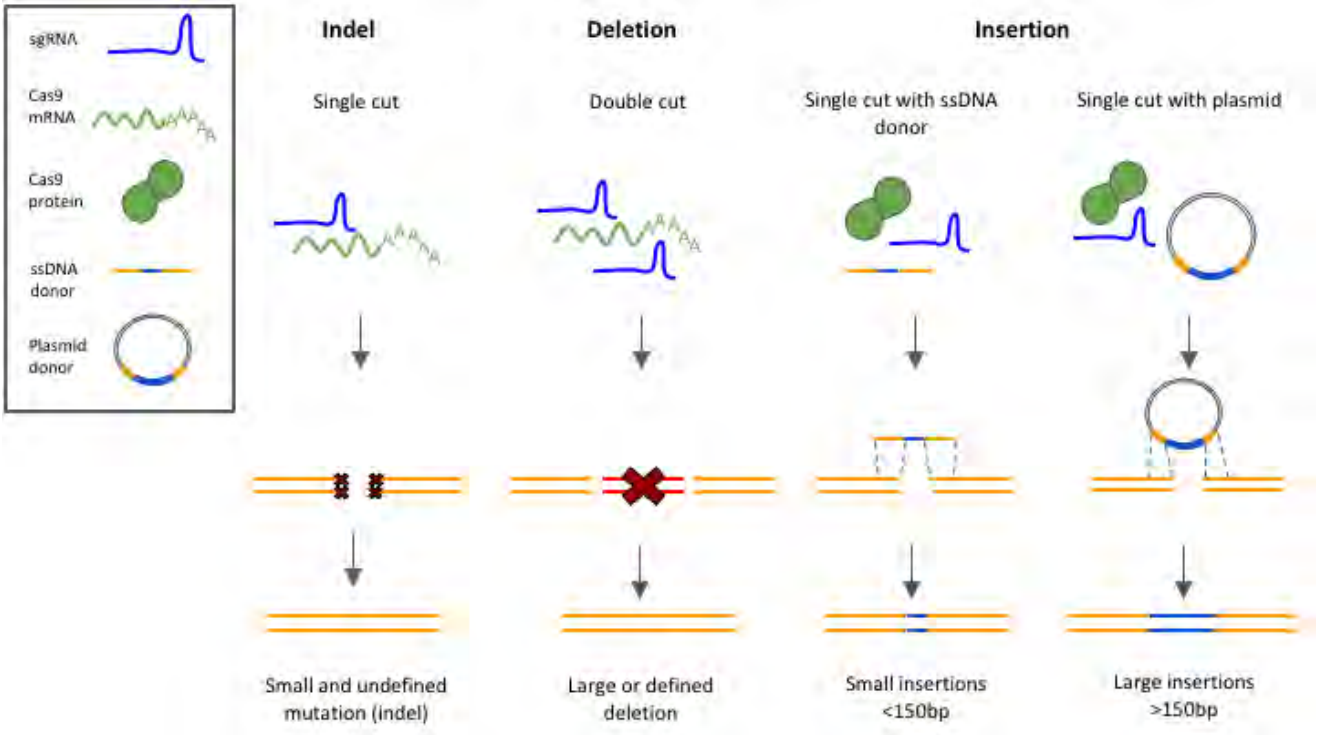


Figure 7 CRISPR CAS9 to introduce mutations in DNA (modified from Thomson et al. 2018)

The single guide RNA, and Cas9 protein or mRNA can generate a variety of mutations. The simplest is an 'indel' where a double strand break is generated and repaired by the cell leaving only a small mutation, of wither an insertion or more commonly deletion. To generate a more defined deletion, two single guide RNAs can be combined with Cas9 to generate two double strand breaks which remove the intervening sequence. This is then repaired by NHEJ at the cut sites. Insertion can be generated through incorporation of a ssDNA repair template with homologous regions flanking the new mutation to be inserted. Plasmid based methods can also be used to insert larger regions of DNA.

1.10.2.3 Further Enhancements of the System

Since the series of landmark papers outlining the use of CRISPR for mouse model modification, further refinements of both the endonuclease enzymes and the targeting strategies have emerged producing a vast 'toolbox' for researchers to utilise.

If off-target effects are a major concern, endonucleases that have higher specificity for on-target cutting have been developed, such Cas9-HF1 (Kleinstiver et al. 2016). Another way to achieve this is to use an endonuclease that does not cause a double strand break, but instead 'nicks' the DNA by cutting only a single strand, known as a 'nickase' Cas9-n (Ran et al. 2013). When two of these are used close by, they are able to increase the specificity of the target whilst minimising the potential for off-target effects as the 'nick' can be easily repaired by the cell.

Cas9 from different bacterial species such as *Staphylococcus aureus* (Thakore et al. 2018; Lee et al. 2016), *Niesseria meningiditis* and *Streptococcus thermophilus* (Müller et al. 2016; Kleinstiver et al. 2015) have been shown to have different PAM sequences, enabling a greater range of the genome to be targeted for editing. Addition of cytosine or adenine deaminase to a catalytically inactivated Cas9 protein (dCas9) have been also been produced (Gaudelli et al. 2017; Komor et al. 2016). Instead of introducing double strand breaks these are able to generate targeted SNPs, changing individual bases; either A-G or C-T. These are likely to be extremely beneficial for gene therapy targets in the future, however the off-target implications of these has not been studied in depth enough to know their specificity or usefulness yet.

1.10.2.4 CRISPR for enhancer analysis

CRISPR has become an extremely useful tool for enhancer analysis and validation. Whilst reporter assays can show when and where a putative enhancer can direct expression, this does not mean that in its native context this will occur. Many factors such as chromatin accessibility, and the number of transgene inserts can affect this expression, whereas simply removing the enhancer from the genome will give a better *in vivo* readout of its functionality.

There are two main ways in which this CRISPR can be utilised for enhancer analysis. The first is to simply remove the putative binding sites. This is best achieved via two sgRNAs, which will often remove intervening sequence without generating larger unwanted deletions. Alternatively, the non-homologous end joining repair system can be exploited and used to insert a sequence containing a modified version of the region, to which the TF of choice will be unable to bind. In results chapters 2 and 3, CRISPR was used to generate mutations of putative enhancers and is discussed in more detail.

2 SOX Proteins

One of the many transcription factor families that are active during embryonic development is the SOX family of transcription factors (Kiefer 2007). These transcription factors rely on binding to enhancer sequences throughout the genome to control many cell fate decisions that are important to development of the foetus, and postnatal life.

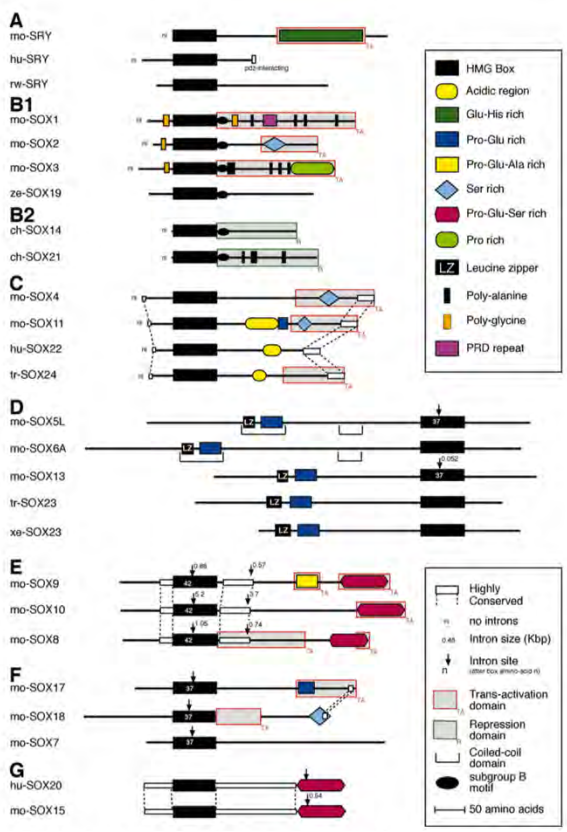


Figure 8 Modified from (Bowles et al. 2000)

The SOX Subgroups. All SOX proteins contain the conserved HMG Box, and are then sub-classified into 8 separate groups (A-G) based on their sequence structure and functional domains as outlined.

2.1 The SOX Family

SOX (**S**ry-Related HMG B**ox**) proteins form a large 20 member protein family, with varied roles within embryonic and postnatal development in mammals, with homologues in other species such as *Drosophila*, *Danio rerio* (zebrafish) and *Xenopus laevis* (frog) (Crémazy et al. 2000; Okuda et al. 2010). They encode a set of transcription factors which are categorised based on their homology to each other, and all contain a HMG Box DNA binding domain, first characterised in the founding member *Sry* (Sinclair et al. 1990). Many of these factors have been implicated in a wide variety of diseases, however a lot of their transcriptional targets are still unknown.

2.1.1 The HMG Box

The first SOX protein to be discovered was encoded by the mouse *Sry* gene (Gubbay et al. 1990; Sinclair et al. 1990) found to be the elusive testis determining factor located on the Y chromosome, and is responsible for differentiation of the gonad towards a male fate (Koopman et al. 1991). This is a single exon gene, with the only known functional domain being the HMG Box, involved in DNA binding, protein interactions and nuclear import (Wilson & Koopman 2002). HMG Box domains are present within a large variety of proteins across species, but the SOX group represents the largest cohort of these in humans and mouse (Štros et al. 2007).

Transcription factors act by their ability to bind DNA, leading to either activation or repression of target genes. In the SOX proteins, this is the function of the HMG Box, whose structure is shown to be comprised of three alpha-helices and a hydrophobic core in an inverted L-shape (van Houte et al. 1995). This is the most conserved region amongst the SOX factors, and all share a very similar consensus DNA binding motif (5′-(A/T)(A/T)CAA(A/T)G-3′) (Kamachi et al. 1999; Mertin et al. 1999). Unlike most TFs, the SOX HMG Box binds to the minor groove of the DNA, rather than the major groove (van de Wetering & Clevers 1992). It has been proposed that by binding to this region, and generating a 70-85 degree bend in the DNA, the protein is able to come into close contact to other TFs, and possibly could play an architectural role in multiprotein enhancer complexes or chromatin accessibility (Ferarri et al. 1992; Wegner 1999).

2.1.2 Subgroups and Homology

SOX proteins are classified into 9 distinct subgroups, with every member sharing at least 50% homology to the HMG Box of the original SOX member, *Sry*. Sub classification is then allocated based on homology to other SOX proteins, with those having at least 85% homology to each other forming a subgroup; these often also contain similar structural motifs and functional properties (Sarkar & Hochedlinger 2013; Chew & Gallo 2009).

The SOX subgroups, whilst varied in their roles are broadly defined as developmental transcription factors. They can be both activating and repressing, sometimes with different members showing distinctly opposing functions. The SOXB1 factors (which will be the main focus of this thesis) are active during early embryonic neural development, and their main role is to inhibit neurogenesis by maintaining stem cells within a progenitor state (Bylund et al. 2003), while the SOXB2 member SOX21 actively promotes neurogenesis by counteracting the activity of SOX1, 2 and 3 (Sandberg et al. 2005).

2.2 SOXB1 Subgroup

The SOXB1 subgroup is made up of three proteins, SOX1, SOX2 and SOX3. SOX3 was discovered soon after *Sry*, and the high level of similarity along with its location on the X chromosome has led to the hypothesis that SOX3 was the original SOX member, from which *Sry* is thought to have evolved (Waters et al. 2007; Nagai 2001). Indeed, ectopic expression of SOX3 in the developing female gonad can induce XX male sex reversal in mice and probably humans (Sutton et al. 2011). Along with the highly similar SOX1 and SOX2, these genes are responsible for the regulation of neurogenesis during early development, as well as having their own tissue specific functions in other aspects of embryogenesis.

While the function of these proteins is known, the exact mechanisms by which they perform their function is an active area of research. These TFs are thought to control the expression of a multitude of genes, but exactly what these genes are and when this control is needed is unknown. This question is the main focus of this thesis, understanding where in the genome SOXB1s are binding and exerting their effect upon target genes within the developing neural system, and during postnatal spermatogenesis.

2.2.1 SOXB1 within Neural Tissues

The SOXB1s are mainly expressed in the central nervous system and have very similar patterns of expression. Wood & Episkopou (1999) published a comprehensive comparison of SoxB1 expression from pre-gastrulation embryos to the early somite stage (Figure 6). *Sox2* is expressed earlier than either *Sox1* or *Sox3*, at 2.5 dpc within the morula, and at 3.5 dpc can be detected within the inner cell mass of the blastocyst. After implantation, expression continues throughout the epiblast and becomes restricted to the presumptive neuroectoderm during gastrulation. By 9.5 dpc, *Sox2* is expressed throughout the developing brain and neural tube. *Sox3* is the second SoxB1 gene to be expressed and is first detected immediately after implantation at 5.5 dpc in the epiblast. At 7.0 dpc, expression is restricted to the anterior neural plate. By 9.5 dpc, expression is detected throughout the neuroaxis (Uchikawa et al. 2011; Wood & Episkopou 1999). *Sox1* is the last of the family to be expressed. The first mRNA is detected at low levels in the neural plate ectoderm during the late head fold stage (Wood & Episkopou 1999), and expression continues in all neuroepithelial cells of the anteroposterior axis (8-8.5dpc), which forms the neural tube at 9.5 dpc.

The overarching function of the SOXB1 subgroup of proteins is to maintain neural progenitor cells (NPCs) of the CNS within a stem-like state, while also inhibiting the differentiation of these into other cell types. This was shown by Bylund et. al. (2003) using electroporated chick embryos whereby the presence of SOX1, SOX2 and SOX3 countered the activity of proneural proteins, and the ability to differentiate was reliant upon suppression of these SOXB1 proteins (Bylund et al. 2003). Inhibition of SOX2 however, leads to their early exit from the cells cycle, resulting in early differentiation (Graham et al. 2003). Rescue of this phenotype was achieved with co-expression of SOX1, indicating the functional similarities and overlapping roles of the SOXB1 factors in the NPCs.

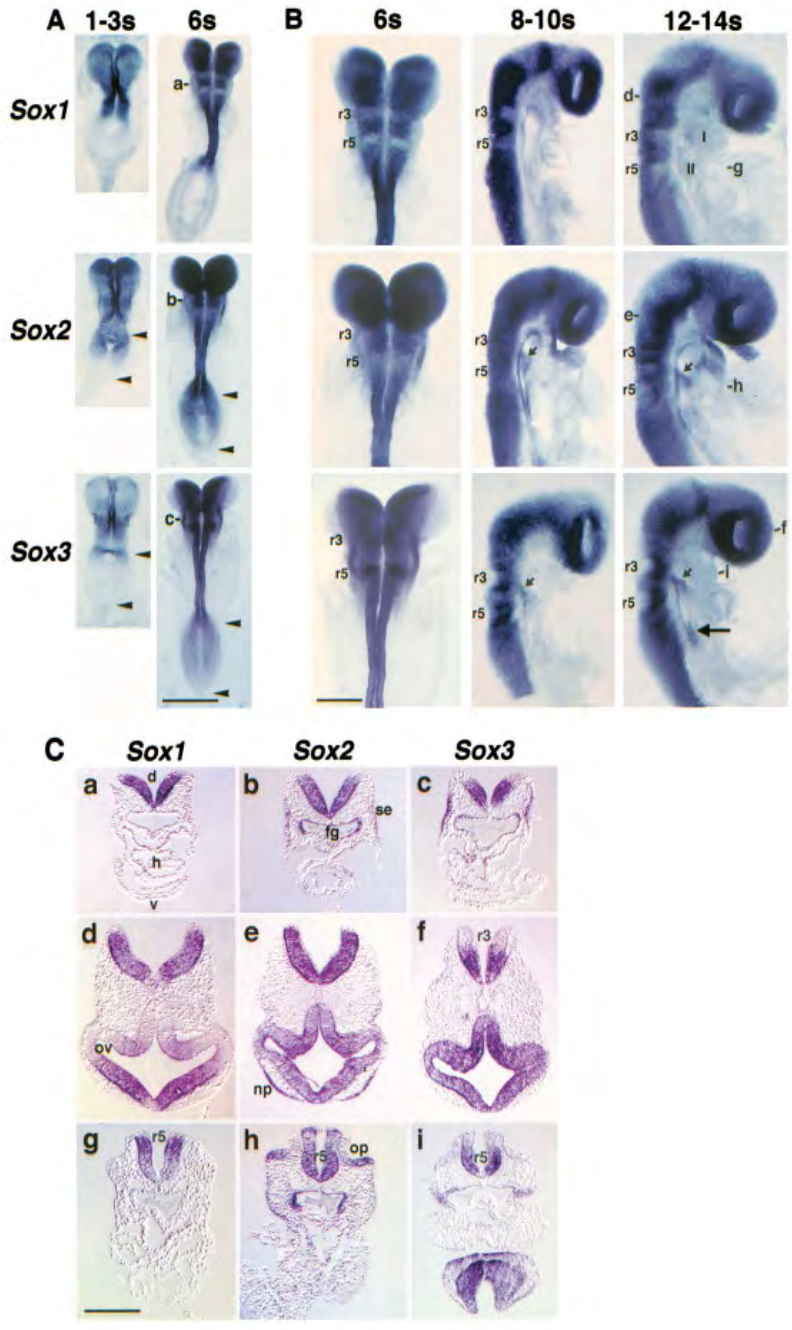


Figure 9 (Wood & Episkopou 1999)

Whole mount in situ hybridization of SOX1, 2 and 3 in mouse embryos. A. Dorsal view of embryos at the 1-3 and 6 somite stage for each of the TFs (SOX1, SOX2 and SOX3), with primitive streak shown by arrowheads. B. The hindbrain of the 6, 8-10 and 12-14 somite embryos , s; somite, r; rhombomere, I; first branchial arch, II; second branchial arch. C. Transverse sections of embryos. R; rhombomere, d; dorsal, v; ventral, h;heart, fg; foregut, se; surface ectoderm, ov; optic vesicle, np; nasal placode, op; optic placode.

2.2.2 SOXB1s in Testes Development

SRY and SOX9 are essential for the progression of the undifferentiated gonads to into the male testes (Koopman et al. 1991; Sekido & Lovell-Badge 2008; Barrinuevo 2006). While the SOXB1 members are not required for sex determination, SOX3 is essential for proper postnatal development of the testes and is the only SOXB1 protein expressed at these stages.

Spermatogenesis is the developmental process that male germ cells undergo to form mature sperm within the testes. The cells transition from spermatogonia to primary and secondary spermatocytes, then spermatids to eventually create spermatozoa able to fertilise the female oocyte (Griswold 2016). SOX3 is expressed within the early stages of this process, found within the early spermatogonia, As, Apr and Al (Raverot et al. 2005; Rizzoti et al. 2004). As will be discussed in section 2.3.2, SOX3 null mice are unable to properly undergo spermatogenesis, leading to an accumulation of spermatocytes with reduced spermatocytes (Laronda & Jameson 2011). It has been proposed that SOX3 is essential for spermatogenesis in a developmental pathway that also involves the transcription factor *Neurog3* (Raverot et al. 2005). This link is further explored in Chapter 4, where a ChIP-Seq experiment was performed on mouse postnatal testes.

2.3 Specific Targets

The SOXB1 proteins appear to have many overlapping transcriptional targets throughout development, and all bind to the same or highly similar DNA motif through their HMG box. As mentioned previously, knockout or mutant models of each SOXB1 member still produces phenotypic effects indicating that they cannot fully compensate for each other and still maintain specific targets.

2.3.1 SOX2

Mouse Models

As SOX2 is the first of the group to be expressed, it has many specific targets. As such, SOX2 knockout mouse models cause early embryonic lethality (Avilion et al. 2003) that is unable to be rescued by either SOX1 or SOX3, as their expression is restricted until later time points. It has been proposed that SOX2 is able to act as a pioneer TF in the early inner cell mass from which the embryo is derived (Zaret & Carroll 2011; Iwafuchi-Doi et al. 2016).

By generating conditional knockout models of SOX2 which restrict expression to ~5% of wild type levels within the developing brain from E14.5, preimplantation lethality can be avoided and specific SOX2 functions can be analysed in the developing CNS (Miyagi et al. 2008). Although these mice still die shortly after birth, it was shown that there is an essential role of SOX2 in generating the appropriate number of neural progenitor cells within the developing brain. It was noted that SOX3 levels were increased in these mice, presumably to compensate for the loss
A recent study by Hagey et al (2018) undertook a large scale analysis of SOX2 binding within various tissue and cell contexts, incorporating both RNA-Seq and ChIP-Seq analysis. By looking at the cortex, spinal cord, stomach and lung/esophagus they could show that SOX2 is able to regulate both tissue-specific and more global core transcriptional programs. SOX2 was shown to bind highly similar motifs within all contexts, although enrichment for known co-factors was also detected alongside.

Human mutations

Human mutations in SOX2 have been identified, with phenotypic consequences mostly being restricted to the eye, with defects such as anophthalmia (absence of one or both eyes) and microphthalmia (significantly reduced eye size) (Driggers et al. 1999; Fantes et al. 2003; Faivre et al. 2006), and these are often coupled with delayed motor and general development. A study of 120 patients with congenital eye abnormalities identified 12 mutations within *Sox2* (Bakrania et al. 2007). These mutations included missense, frameshift and whole gene deletions of *Sox2*, indicating that SOX2 haploinsufficiency is a cause of these eye defects. Whilst this correlation had been identified, it was not until a study in 2006 (Taranova et al.) where a mouse model that generated different gene-dosages of SOX2 to assess the effects on development of the eye that this was confirmed. When a SOX2 conditional deletion was introduced, the NPCs within the retina lost the ability to divide and differentiate, whereas hypomorphic alleles led to aberrant differentiation resulting in variable phenotypes as seen in human mutations.

2.3.2 SOX3

Mouse Models

Knockout mouse models of SOX3 do not have as severe a phenotype as those of SOX2, generating mice with relatively mild defects. Rizotti and colleagues (2004) generated SOX3 null mice, and the resulting mutation led to a variety of defects including craniofacial abnormalities, as well as fertility defects. They focused however on the pituitary, and were able to show that SOX3 is essential for development of the hypothalamic-pituitary axis and is not compensated for by other members of the SOX family.

Spermatogenic defects have also been investigated in SOX3 null mice, which show that whilst *Sox3* is evolutionarily very close to *Sry*, it is not important or required for sex determination in mammals but is needed for proper gonadal function (Weiss et al. 2003). Through analysis of the postnatal testes, it was shown that loss of SOX3 generates testes with greatly reduced weight, Sertoli cell vacuolization and loss of germ cells leading to sterility. This indicated that SOX3 is an important factor involved in spermatogenesis, the maturation of germ cells into functional sperm, and that it is specifically expressed within the As, Apr and Al spermatogonia (Raverot et al. 2005). Within this study, a functional link was proposed between SOX3 and NEUROG3, another TF expressed within both neural and endocrine progenitors, and the same spermatogonia subtypes as SOX3. Within results Chapter 3 of this thesis, a functional link between NEUROG3 and SOX3 in the progression of spermatogenesis is proposed.

A SOX3 specific target within the spinal cord is the homeobox gene *Dbx1* (Rogers et al. 2014). Intriguingly, this is one of the few SOX3 bound genes to be identified that is unable to be rescued by other SOX family members. This research suggests that while most genes can be compensated for, there may be inherent differences

between each of the SOXB1 proteins and their activity which cannot always be overcome.

Human Mutations

Phenotypes such as hypopituitarism, mental retardation and spermatogenic defects are the most common disorders of *Sox3* mutation in humans. Polyalanine tract expansions within SOX3 have been identified as the causative mutation in a family with mental retardation and growth hormone deficiencies (Hamel et al. 1996) (Laumonnier et al. 2002). The association of the polyalanine tract expansions with hypopituitarism was further investigated in a chimeric mouse model by Hughes et al. (2013) where it was found that the expanded polyalanine tracts generate very low levels of SOX3 protein, not sufficient for proper development of the hypothalamus.

Sex reversal has also been seen in patients carrying mutations which affect the regulation of SOX3. Analysis by Sutton et al. (2011) of patients that were 46, XX but phenotypically male, showed chromosomal rearrangements and duplications of SOX3 which were shown to alter the regulation of expression. As mentioned previously, when SOX3 is ectopically expressed within the developing gonad of mice, it can act in a similar way to SRY, and activate *Sox9* leading to phenotypically male embryos. It is thought that these patient mutations are leading to modification of the regulation of SOX3 and allowing it to act as an SRY surrogate, producing testes and upregulating 'maleness' pathways (Sutton et al. 2011).

2.3.3 SOX1

Mouse Models

Like SOX3, the loss of SOX1 is not as severe as that of SOX2, presumably because of the functional redundancy between the other members of the SOXB1 family. However, SOX1 is specifically expressed within certain regions which are unable to be compensated for. By generating SOX1 null mouse lines, Nishiguchi and colleagues (1998) were able to show the importance of SOX1 for proper development of the lens fiber within the eye. Whilst SOX2 is expressed within the sensory placode of which develops into the lens, it is replaced by SOX1 prior to the expression of **γ**-crystallin genes . SOX1 has been shown to partner with PAX6 to bind to regulate the suite of **γ**-crystallin genes leading to elongation of the lens fiber cells critical for lens development (Nishiguchi et al. 1998). These mice lacking SOX1, whilst only showing mild phenotypic effects of the eye; small eyes with opaque lenses, also showed signs of epileptic seizures. Whilst these mice develop brains without major malformations, there is a loss of neurons within the ventral striatum present from birth (Malas et al. 2003). The loss of these neurons disrupts the olfactory circuit, shown by abnormal electrophysiological recordings which can lead to spontaneous seizures of the forebrain.

Human Mutations

Whilst research in mouse shows the effects of SOX1 loss in a model system, human mutations in *Sox1* have not yet been well explored. There are however patients that present with large deletions on chromosome 13, which encompass among other genes, *Sox1* (Reinstein et al. 2016). The smallest of these deletions, 1.3Mb, has been found in 2 brothers who present with epilepsy and mild intellectual disability as well as genitourinary tract defects (Orsini et al. 2018). SOX1, along with ARHGEF7 have been implicated in this phenotype due to their associations in studies of epilepsy.

2.4 Functional Redundancy

Highlighting the importance of the SOX factors, knockout mouse models of SOX2 (the first to be expressed) exhibit early embryonic lethality (Avilion et al. 2003). Knockout models of SOX3 and SOX1 however, only show mild phenotypic effects, such as growth hormone deficiencies and microphthalmia, respectively (Rizzoti et al. 2004; Nishiguchi et al. 1998). Given that the effects seen in SOX1 and SOX3 knockout and mutant models are so mild when compared to SOX2 knockouts it was proposed that these three proteins are functionally redundant (Miyagi et al. 2009). Each of the proteins are very similar and each are capable of binding to the same TFBS within areas where they are all expressed, such as the neural progenitor cells. As *Sox2* is the first of the group to be expressed, neither SOX1 or SOX3 is able to compensate for this loss and bind to the important SOX2 transcriptional targets leading to early lethality. Once *Sox1* and *Sox3* are expressed however, SOX2 protein is already present and capable of binding to each of the others transcriptional targets, reducing the phenotypic effects seen.

By utilising CRISPR technology, Adikusuma (2017) replaced the coding sequence of SOX3 with SOX2 and assessed its effect within the postnatal testes, a SOX3 specific region. This demonstrated that this gene swap was able to functionally compensate for the loss of SOX3 under the regulation of its endogenous promoter region.

2.5 DNA Binding Partners

While TFs are able to regulate expression of genes through binding to the DNA, an extra level of regulation can be imposed through partner factor binding. Throughout the 20 members of the SOX family, the DNA binding motif is highly similar- so how the individual TFs are able to bind to their own motif was not well understood. Kamachi et al (2000) has proposed a model of SOX binding, whereby an individual SOX TF alone is insufficient to activate gene expression and must be co-operatively binding with a partner to elicit an effect. Binding partners have not yet been identified for every SOX protein at every enhancer, but this has been seen at some loci.

One of the most well studied partner factors of the SOXB1 proteins is the POU (Pit-Oct-Unc) family. These are a group of transcription factors that contain the DNA binding POU domain, and are important for regulating cellular identify during development and are expressed within the developing CNS along with many SOX proteins (Latchman 1999). It has been estimated that SOX2 and Oct4 are able to bind together at thousands of enhancers within various tissue contexts, most prominently within neural stem/progenitor cells (Nishimoto et al. 1999; Tomioka et al. 2002). The importance of the regulation of many of these genes led to the theory that the two families may have co-evolved as cooperative regulators of many cell fate decision (Wilson & Koopman 2002).

2.6 Identification of SOXB1-bound enhancers

To identify enhancers bound by the SOXB1 proteins within the central nervous system, a ChIP-Seq experiment was performed on cultured mouse neural progenitor cells with a SOX3 antibody (McAninch & Thomas 2014). From this dataset over 8000 regions were identified as being bound by SOX3, so further ChIP data from SOX2 binding and p300 in similar cell types was used to identify a smaller subset of regions bound by each of these, that could be identified as putative enhancers. As well as evidence that the SOXB1 proteins bound to these regions of DNA, it also gave further insights into how and what SOXB1 is regulating and its specific roles within the NPCs. GO terms were generated from the binding regions which suggests that the SOXB1s are involved in various aspects of neural tube and brain development. Further evidence for SOXB1 and Oct protein partner binding was also shown when motifs were identified that incorporated the TFBS of both SOX and POU. This study has informed the work of both Chapters 2 and 3 of this thesis.

2.7 Project Aims

All of the SOXB1 proteins are expressed within the neuroprogenitor cells of the CNS, while SOX3 is also found within the postnatal testes. The transcriptional targets of these proteins have not yet been fully elucidated, however previous work such as that by McAninch and Thomas (2014) described previously generated a list of putative enhancers bound by SOXB1s. The aims of this thesis have been to both identify and investigate putative enhancers bound by SOXB1 through both ChIP-Seq data analysis as well as the generation of CRISPR mouse models to investigate gene expression and phenotypic effects.

A previously described enhancer of the neurofilament gene *Nes* is shown to contain SOXB1 and POU TFBS, and Chapter 2 of this thesis examines the contribution of this enhancer to overall expression as well as the possibility that it is capable of *trans* interactions.

FZD3 (Frizzled 3) is an essential protein that is important for brain development. In the SOXB1 ChIP-Seq analysis a region of high conservation was identified with a *Fzd3* intron that appeared to be bound by SOXB1 (McAninch & Thomas 2014). In Chapter 3, we deleted this putative enhancer and assess its role in *Fzd3* expression though CRISPR modification and reporter mice.

The final results chapter focuses on SOX3's role within the postnatal mouse testes. A ChIP-Seq experiment was performed and the chapter analyses the results of the bound regions. This shows a potential role for SOX3 in the regulation of histone replacement in sperm as well as providing evidence that strengthens the link between NEUROG3 and SOX3 during spermatogenesis.

Chapter Two

The Nestin neural enhancer is essential for normal levels of endogenous Nestin in neuroprogenitors but is not required for embryo development

Summary

This manuscript outlines the functional analysis of the SOXB1 bound enhancer region that controls expression of the neurofilament gene *Nes*. This enhancer has previously been identified as a CNS specific enhancer which can be used to direct transgene expression to neuroprogenitors, however the endogenous functionality of this region has not been assessed.

By using CRISPR to generate deletions of the SOXB1 binding motifs we have been able to study the contribution of this enhancer to Nes expression throughout embryogenesis. We can show that this enhancer is active from 9.5dpc in the head, and at 10.5 dpc is responsible for up to 70% of Nes expression.

We also generated 2 independent Nes null lines using CRISPR. Through breeding with enhancer deleted mice we show promising evidence of ***trans*** enhancer-promoter interactions, a relatively new field of mammalian enhancer research.

Through these experiments we have shown the Nes enhancer is essential for proper levels of Nes within the CNS, but its loss does not overtly affect embryonic development.

Statement of Authorship

Title of Paper	The nestin neural enhancer is essential for normal levels of endogenous nestin in neuroprogenitors but is not required for embryo development
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style
Publication Details	

Principal Author

Name of Principal Author (Candidate)	Ella Thomson
Contribution to the Paper	Designed and performed experiments, analysed results, wrote 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 5/7/19

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 to 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	Ruby Dawson
Contribution to the Paper	Generated initial mouse model and performed experiments leading to figure 1
Signature	Date 5/7/19

Name of Co-Author	Chee He H'ng
Contribution to the Paper	Performed experiments leading to figure 2 and 4
Signature	Date 4/7/19

Name of Co-Author	Fatwa Adikusuma
Contribution to the Paper	Designed and performed experiments leading to figure 2 and 4
Signature	Date 4/7/19
Name of Co-Author	Sandra Pilz
Contribution to the Paper	Performed all embryo microinjections to generate mouse models
Signature	Date 4/7/19
Name of Co-Author	Paul Thomas
Contribution to the Paper	Conceived and designed study, analysed results, evaluated and edited manuscript.
Signature	Date 4/7/19.

The Nestin neural enhancer is essential for normal levels of endogenous Nestin in neuroprogenitors but is not required for embryo development

Ella Thomson, Ruby Dawson, Chee Ho H'ng, Fatwa Adikusuma, Sandra Piltz and Paul Q Thomas

Abstract

Enhancers are vitally important during embryonic development to control the spatial and temporal expression of genes. Recently, large scale genome projects have identified a vast number of putative developmental regulatory elements. However the proportion of these that have been functionally assessed is relatively low. While enhancers have traditionally been studied using reporter assays, this approach does not characterize their contribution to endogenous gene expression. We have studied the murine Nestin (*Nes)* intron 2 enhancer, which is widely used to direct exogenous gene expression within neural progenitor cells in cultured cells and in vivo. We generated CRISPR deletions of the enhancer region in mice assessed their impact on *Nes* expression during embryonic development. Loss of the *Nes* neural enhancer significantly reduced *Nes* expression in the developing CNS by as much as 82%. By assessing NES protein localization, we also show that this enhancer region contains repressor element(s) that inhibit *Nes* expression within the vasculature. Previous reports have stated that *Nes* is an essential gene, and its loss causes embryonic lethality. We also generated 2 independent *Nes* null lines, and show that both develop without any obvious phenotypic effects. Finally, through crossing of null and enhancer deletion mice we provide evidence of trans-chromosomal interaction of the *Nes* enhancer and promoter.

Introduction

Embryonic development requires precise coordinated expression of thousands of genes across space and time. Regulatory elements such as enhancers have a critical role in coordinating spatio-temporal gene expression during embryogenesis. Enhancers are typically located within introns and intergenic regions and comprise DNA motifs that can be bound by transcription factors (TF). TF binding promotes interaction of the enhancer with the target promoter via DNA looping. This process, which involves cohesins and the mediator complex (Kagey et al. 2010), allows TF-associated co-activators to engage the transcriptional machinery and stimulate RNA Pol II-mediated transcription of the target gene. While enhancers are generally regarded to function as *cis*-acting elements, recent evidence suggests that some enhancers can act in *trans* to influence expression of their target gene on the homologous chromosome. *Trans* enhancer-promoter interaction in *Drosophila*, termed transvection, is relatively well characterised and has recently been visualised within developing embryos (Lim et al. 2018). Few examples of *trans* interactions have been reported in vertebrates, although a recent analysis at the IGH super-enhancer indicates that *trans*-enhancer activity can occur in mammals (Le Noir et al. 2017).

The Nestin gene (*Nes*) encodes an intermediate filament protein and is widely expressed during embryonic development including progenitor cells throughout the neuroaxis (Lendahl et al. 1990; Lothian & Lendahl 1997). Differing reports of NES functionality have been published, with research in 2011(Mohseni et al. 2011) suggesting *Nes* is not essential for development of the central nervous system, in contrast to an earlier paper (Park et al. 2010) indicating that loss of the gene results in embryonic lethality. The reasons for this disparity are unclear given the similarity of the experimental design used by both groups. The *Nes* neural enhancer

(Zimmerman et al. 1994) is a highly conserved element located in intron 2 and is commonly used to activate exogenous gene expression in neural progenitor cells in vivo and in vitro (Dubois et al. 2006; Trumpp et al. 1999; Petersen et al. 2002). In vitro and transgenic data indicate the transcription factors belonging to the SOX and POU families bind the *Nes* enhancer and function synergistically to control the *Nes* expression in the CNS progenitors (Tanaka et al. 2004). Consistent with these data, ChIP-seq experiments have identified robust binding of endogenous SOX3 protein at the *Nes* enhancer in cultured neuroprogenitor cells (McAninch & Thomas 2014).

Traditionally, enhancers have been identified and characterized using transgenic reporter assays (Kvon 2015). This has proven to be a useful approach to determine the contribution of specific enhancer elements to the spatiotemporal expression of its cognate gene. However this strategy is incapable of recapitulating the endogenous genomic and chromatin environment in which the enhancer is usually located. The emergence of CRISPR gene editing technology (Cong et al. 2013) enables rapid and efficient deletion of enhancer sequences in vivo and provides the endogenous environment is maintained and allows for a better understanding of both enhancer activity and contribution to gene expression. This is an important advancement as the number of putative enhancers identified via bioinformatic and TF binding studies continues to grow, while functional studies are lagging.

Despite widespread use of the *Nes* neural enhancer, the contribution of this enhancer to *Nes* expression during development has not been studied, nor have the effects of removing the enhancer on the developing CNS. Here we show that CRISPR-mediated deletion of the *Nes* enhancer results in a significant reduction in mRNA expression as well as altered protein levels within the developing mouse central nervous system. Using CRISPR/Cas9, we also generate two NES loss of

function mouse lines and show that NES KO mice are viable. Finally, we present evidence that the *Nes* enhancer is able to function in *trans*.

Materials and Methods

Mouse Generation

CRISPR guides were designed using the crispr.mit.edu tool to determine off-target scores. Guide RNA sequences (enhancer deletion-TTTGCGGTCTGAAAAGGATT, AGAATCGGCCTCCCTCTCCG, nestin null lines - GGAGCTCAATCGACGCCTGG, GCACAGGAGACCCTACTAAA) were annealed and ligated into px330 (Addgene) after digestion with Bbs1 (NEB) using Rapid Ligation Kit (ThermoFisher Scientific), and transformed into *E. Coli* using standard protocols. Plasmid was extracted from positive colonies using a Midi-Prep kit (Qiagen). Primers were designed to incorporate T7 promoter sequence and tracR sequence, and PCR was performed on plasmid DNA with Phusion High Fidelity PCR Kit (NEB). PCR products were converted to RNA using the T7 RNA Transcription Kit (NEB) and purified with RNEasy Kit (Qiagen) to generate sgRNA. Cas9 mRNA was synthesised from the XhoI (NEB) digested plasmid (Addgene) using the Mmessage Mmachine T7 Ultra Transcription Kit (ThermoFisher).

BL6/2J females were superovulated with Pregnant Mare Serum Gonadotropin (PMSG) and human Chorionic Gonadotropin (hCG) prior to mating with BL6 males for zygote harvesting. Single cell zygotes were collected on the day of microinjection, and treated with hyaluronidase to remove surrounding cumulus cells. Cytoplasmic injection was performed with CRISPR reagents (50ng/uL Cas9 mRNA, 100ng/uL sgRNA) before transfer into pseudopregnant CD1 females.

Genomic DNA was extracted from 3 week old tail tips and ear clippings using KAPA Mouse DNA Extraction Kit (KAPA Biosystems) or High Pure PCR Template Kit (Roche).

Founder mice were genotyped using FailSafe PCR Kit (EpiCentre) and run on a 12% polyacrylamide gel for heteroduplex assay. The genotype of the founder mice was confirmed via Sanger sequencing after BigDye Terminator v3.1 (Applied Biosystems) PCR reaction using reverse primer.

Regular colony and embryo genotyping was performed with primers flanking deleted sequence (enhancer deletion line F-GCCCCAGTCAGTCTTCTGAG R-GCCACTGCAGGATCACTCTT, nes null FS F1 – CTGCTGAGCTGGGATGATGC F2 – AGCTCAATCGACGCCTGGA R- GCATTCTTCTCCGCCTCGA, nes null BD F-CTGCTGAGCTGGGATGATGC R- CTGCTGAGCTGGGATGATGC) using 2G Fast MasterMix (KAPA), or Buffer J (EpiCentre) with Taq Polymerase (Roche).

All mouse breeding and experimental work was performed at the University of Adelaide in accordance with relevant ethics approvals (S-201-2013 and S-173-2015).

Tissue Preparation

Heterozygous (wt/-255) males and females were set up as timed matings for embryo collection. Females were humanely killed via cervical dislocation and embryos removed and stored in cold 1x PBS until dissected. Tails were removed and kept at -20°C. Heads were removed and flash frozen on dry ice and kept at -80°C for RNA extraction, or kept o/n in 4% paraformaldehyde in PBS, washed 3x in PBS and cryoprotected overnight in 30% sucrose before flash freezing in OCT and stored at -20°C for immunohistochemical analysis.

Immunohistochemistry

Trunks were sectioned at 16um on a cryostat (Leica CM1900) and slides washed 3x 10mins in PBT (1xPBS, 0.25% Triton-X), blocked for 30min in BS (1x PBS, 0.25%

Triton-X, 10% Horse Serum), then stained o/n with 200uL primary antibody diluted in BS and kept in humidified chamber 4°C. Primary antibody washed off with 3x 10mins PBS. 200 uL of secondary antibody diluted in Blocking Solution was added to the slides and kept in a dark humidified chamber for 4hrs at room temperature. The secondary antibody was removed with 3x 10min washes in PBS, slides were dried and set with Prolong Gold Antifade + DAPI (Molecular Probes) and coverslip was applied. Slides kept overnight in the dark before image acquisition Nikon Eclipse Ti Microscope using ND2 Elements software. Images were modified for colour, brightness and contrast using Adobe Photoshop v7 (Adobe Systems). Antibodies used, Anti-SOX3 (R&D Systems, AF2568), Anti-Nestin (Abcam AB82375), Anti-CD31 (BD Pharmigen 550274). Secondary antibodies, Donkey anti-Goat-Cy3 (Jackson ImmunoResearch), Donkey anti-Rat-Cy5 (Jackson ImmunoResearch), Donkey anti-Rabbit-488 (Jackson ImmunoResearch).

In situ Hybridization

In situ probes were designed to target exon 4 of the *Nes* gene. Primers corresponding to the region were used to amplify the DNA from wildtype mouse DNA and incorporate a T7 promoter at the 5' end. The DNA was converted to RNA using the T7 IVT Kit (NEB), followed by DNase I (NEB) treatment and purification with RNEasy kit (Qiagen).

Embryo trunks were sectioned at 16um on a cryostat (Leica CM1900) and stored at -20°. Prior to in situ hybridisation, slides were defrosted for 1hr at room temperature. The RNA in situ probe was denatured at 72°C for 2 minutes and kept on ice. 100ul hybridisation buffer containing 1uL diluted riboprobe/slide added to slides and kept in humidified chamber containing formamide overnight at 65°C. Slides were washed 3x 30 mins at 65°C in Wash Buffer (50% Formamide, 5% 20x SSC), then 3x

30mins washes in MABT (Maleic Acid Buffer + 0.1% Tween-20) at RT. Slides were blocked with 300uL Blocking Solution (Blocking Reagent, Sheep Serum, MABT) and kept in humidified chamber at RT for 2 hrs. 75uL of anti-DIG antibody diluted in Blocking Solution added to slides and kept o/n at RT in humidified chamber. anti-DIG antibody was washed off with 4x 20min washes in MABT, then wash 2x 10mins in Alkaline Phosphatase Staining Buffer (4M NaCl, 1M MgCl2, 1M Tris pH 9.5). Slides were then stained with 95uL staining solution (NBT, BCIP, Alkaline Phosphatase Staining Buffer), coverslipped, and kept in the dark at RT overnight. Staining solution was removed by washing 3x 5mins in PBS, and fixed with 300uL 4% PFA added to slides and incubated for 1hr in sealed contained. Fixative was washed off with 3x 10min PBS washes, and 50uL Mowiol added to each slide for mounting with coverslip. Slides were analysed using brightfield microscopy on Nikon Eclipse Ti Microscope using ND2 Elements software (Nikon).

qRT-PCR

RNA was extracted from flash frozen embryo heads by Trizol. Briefly, heads were homogenised in 500uL Trizol, 100uL chloroform was added to mixture and centrifuged at 6000xg for 30mins. The aqueous layer was removed and equal amount of 70% EtOH added. Then placed in RNEasy spin column and centrifuged at 13000 rpm for 1 minute. Column was washed with 2x Buffer RLT (Qiagen), and purified RNA eluted in 30uL of RNAse free H20, and stored at -20C. RNA samples converted to cDNA using AB Systems High Capacity RNA to cDNA Kit. SYBR Fast standard protocols used for qPCR, samples run in quadruplicate. *B-actin* (F-CTGCCTGA CGGCCAGG, R- GATTCCATACCCAAGAAGGAAGG) used to normalise cDNA levels across samples, and *Nes* primers used to measure expression levels across timepoints and samples (F-GCTTCTCTTGGCTTTCCTGA; R-AGAGAAGGATGTTGGGCTGA). Prism software was used for the statistical analysis

of qPCR data. Unpaired t-tests were performed to determine if wildtype *Nes* expression was significantly different from enhancer-deleted lines at each timepoint.

Results

Generating a *Nes* Enhancer Deletion Mouse Model

To investigate the role of the *Nes* enhancer in directing endogenous expression in vivo, we generated an enhancer deletion mouse model using CRISPR-Cas9 mutagenesis. Two gRNAs flanking the *Nes* enhancer were microinjected into mouse zygotes with Cas9 mRNA. 21 founder mice were generated with a range of deletions that partially or completely deleted the *Nes* enhancer. We selected a single founder animal containing both 255bp and 208bp deletion alleles that encompassed all SOXB1 binding sites identified in the ChIP-Seq analysis for these experiments (Figure 1A). Independent lines were generated for each deletion (hereafter referred to as -255 and -208). The -255 line was used for subsequent analysis of the enhancer deletion. Heterozygous and homozygous pups and embryos were generated at expected ratios indicating that viability was not compromised by the deletion mutation (Figure 1B) No morphological abnormalities were identified in either line indicating that the enhancer deletion did not overtly impact development.

Nes mRNA expression is reduced in enhancer deletion mice

To determine the impact of enhancer deletion on *Nes* expression, qPCR was performed on -255 homozygous whole embryos (8.5 dpc) and embryonic heads (9.5 dpc-15.5 dpc), as seen in Figure 2A. No significant difference in *Nes* expression was detected in mutant embryos at 8.5 dpc. However, from 9.5 dpc significantly reduced levels of *Nes* mRNA were detected in the embryonic cranium. Notably, the greatest reduction in Nes expression was detected at 10.5 dpc, with mutant embryos expressing just 18% of Nes mRNA compared with wild type controls. From 11.5 dpc, a gradual increase in expression was detected in mutants which by 15.5 dpc had recovered to 60% of wild type expression. A reduction in *Nes* expression was also observed in -208 homozygotes at 11.5 dpc (Supplementary Fig. 1).

Next, we determined the spatial impact of enhancer deletion on *Nes* expression in the developing CNS (Figure 2B). For this experiment we analysed the spinal cord at 11.5 dpc as *Nes* is robustly expressed in a steriotypical pattern throughout the trunk at this stage due to the abundance of neural progenitors (Dahlstrand et al. 1995). In situ hybridization was performed on the trunk sections of wild type (wt) and homozygous enhancer deletion (-255) embryos. As expected, expression of *Nes* was detected throughout the spinal cord, with the highest levels confined to the lateral regions and the floor plate. In contrast, the spinal cord of enhancer-deleted embryos was virtually devoid of *Nes* mRNA, except for restricted expression in the floor plate and the lateral regions. Notably, lateral expression in the mesoderm was not noticeably diminished in mutant embryos, consistent with the neural-specific activity of the *Nes* enhancer in transgenic mice. Thus, deletion of the *Nes* neural enhancer results in a striking reduction in the level and extent of *Nes* expression.

Protein Expression within Enhancer Deleted Mice

As mRNA expression was significantly reduced in both the embryonic head (qRT-PCR analysis) and neural tube (in situ hybridization), we performed protein expression analysis in these regions to determine whether NES was similarly reduced. Both head and trunk transverse sections were taken from wildtype, heterozygous and homozygous embryos and co-stained with anti-NES and anti-SOX3 antibodies (Figure 3). We theorised that as the enhancer is controlled by the SOXB1 proteins binding to the region, that we would see little to no NES expression throughout the SOX3 expressing zones of the neural tube and brain.

The wildtype brain sections show the telencephalon is densely stained for Nestin, showing a long filamentous structure without nuclear staining (Figure 3A). The SOX3 is shown to be overlapping the NES in most regions. There is a small area along the telencephalon that is NES positive and SOX3 negative. In the heterozygous sections, the NES reactivity is lower, but still shows a similar pattern to wildtype and no apparent difference in expression are seen within the SOX3 expressing zones. In the homozygous enhancer deletion however, there are obvious differences in the staining pattern of the NES protein, as it is duller throughput the telencephalon, but shows regions of high reactivity that appear to be surrounding the vasculature. This effect is not seen in either wildtype or heterozygous samples.

This experiment was repeated using neural tube sections, and similar results were seen (Figure 3B). The wildtype embryos show NES forming a smooth filamentous structure from the lateral edges, towards the midline. Whilst the SOX3 is confined to the ventricular zone with no expression seen in the intermediate zones or the floor plate. Interestingly, the base of the neural tube in the -255 embryos shows a region of intense NES reactivity in a non-SOX3 expressing zone that is not seen in

wildtype samples. The NES reactive structures/stripe phenotype seen in the brain is recapitulated again in the neural tube, with these seen throughout the whole neural tube, and not just the SOX3 expressing regions. The somites, lateral to the neural tube show NES staining in all 3 genotypes, with no apparent differences in intensity or pattern. It is important to note that these are not SOX3 positive zones, and can be used as an internal control to compare neural tube NES staining intensity.

Ectopic Nestin expression in vasculature of enhancer deleted embryos

Whilst analysing -255 embryos for NES protein expression, we noted specific staining in discrete structures within the neural tube and cortex that appeared to be the developing vasculature. Notably, this signal was not present in wildtype or heterozygous embryos. To further investigate this finding, we co-stained 10.5 dpc embryo heads with CD31 (endothelial cell marker) and NES antibodies (Figure 4A). Images captured using an inverted fluorescence microscope indicated colocalisation of NES and CD31 in -255/-255 embryos but not in WT controls. Additional analysis using confocal microscopy revealed widespread expression of NES in endothelial cells lining the developing vasculature of -255/-255 embryos. In contrast, NES expression was rarely detected in wild type endothelial cells. Together, these data indicate that deletion of the *Nes* neural enhancer induces ectopic expression in endothelial cells.

Nes is not required for CNS development

It has previously been reported that deletion of *Nes* causes extensive cell death in the developing CNS and embryonic lethality at approximately 8.5 dpc (Park et al 2010). Given that -255/-255 mutants do not exhibit overt developmental defects, it appears that the level of *Nes* in these enhancer-deleted embryos exceeds the threshold required for normal development. We were therefore interested in assessing whether further reduction of *Nes* levels in -255/KO compound heterozygous embryos would compromise CNS development. To generate *Nes* knockout mice, we employed a dual gRNA deletion strategy (Figure 5). The proximal gRNA targeted exon 1 immediately downstream of the start codon and the distal gRNA cut immediately upstream of the stop codon in exon 4. The rationale for this approach was that null alleles could be generated via frameshifting indels at the proximal cut site or from deletion of the ~8.7 kb intervening sequence between the proximal and distal cut sites. This approach also provided the necessary alleles for the *trans* enhancer interaction experiment (see below). PCR genotyping indicated that four of the six founder animals contained at least one large deletion allele. Sanger sequencing confirmed that the founder used for subsequent breeding carried the expected 8672 bp deletion that encompassed almost all of the coding region and introns of the *Nes* gene, including the neural enhancer in intron 2.

This null allele, *Nes* g.54_4518/p.L19_V1506del/p.L19fsX, termed BD, encodes only the first 18AA of the NES open reading frame, and a frameshift causes the last 30AA of exon 4 to be incorrect. This founder also carried an 8bp frameshifting deletion, g.50_57del/p.R17fsX75, termed FS, at the proximal cut site that terminated the protein after 13 amino acids. Breeding colonies for each mutation were generated. Surprisingly, BD and FS homozygous mice were viable and did not exhibit any overt phenotype or developmental defects. Compound heterozygous

FS/BD mice were also phenotypically normal. To confirm that NES protein was not generated from the FS allele, we stained FS/FS embryonic brain sections with anti-NESTIN antibody. In contrast to WT control tissue, no discernible expression was detected in mutant tissue. We therefore conclude that *Nes* is not required for CNS development or viability.

Trans Interactions of the *Nes* neural Enhancer

While enhancers are generally considered to be *cis*-regulatory elements, previous studies have provided evidence for interchromosomal *trans* interaction between enhancers and their cognate promoters (Bateman et al. 2012). To investigate possible interchromosomal activity of the *Nes* neural enhancer *in vivo*, we used qPCR to assay allele-specific expression in a series of compound heterozygous embryos. For this experiment, we exploited the presence of the *Nes* enhancer in the FS allele but not the BD allele. Thus, the difference in *Nes* expression from the WT allele in FS/WT and BD/WT embryos will reflect trans activity of the (FS) Nes enhancer. Similarly, the difference in Nes expression from the -255 allele in FS/-255 and BD/-255 embryos will reflect trans activity of the (FS) Nestin enhancer. No discernible signal was generated from FS/BD embryos indicating that *Nes* mRNA is not generated from either null allele (presumably due to nonsense mediated decay for the FS allele). -255/-255 *Nes* expression was 23% of WT expression, consistent with Fig. 2A. Comparison of FS/-255 and BD/-255 expression revealed a significant increase in the former (17% vs 11%; $p < 0.01$). Similarly, WT/FS *Nes* expression was higher than WT/BD, however this did not reach significance (64% vs 49%; $p < 0.07$). Together, these data suggest that the *Nes* enhancer can function in *trans*.

Discussion

While enhancers are routinely used to drive spatio-temporally restricted expression of heterologous genes, their functional role in coordinating cognate gene expression remains poorly understood. Using CRISPR/Cas9 technology, we show that deletion of the *Nes* neural enhancer has a profound impact on endogenous *Nes* expression in the developing nervous system. Our data also indicate that this region also contains a repressor element that inhibits expression in endothelial cells, underlining the ability of deletion analysis to identify both positive negative regulatory interactions.

Nes is expressed within the incipient neural progenitor cells during early embryogenesis and is maintained during expansion of this cell population. Upon differentiation, *Nes* is downregulated and is replaced by other members of the intermediate filament family (Michalczyk & Ziman 2005). At 8.5 dpc, *Nes* expression is not significantly different in -255 homozygous embryos indicating that the neural enhancer is not functionally required for initiation of *Nes* expression. Given that there is robust SOXB1 expression in neuroprogenitors at this stage, it appears that putative binding of these factors to the *Nes* enhancer at this stage is not required for expression, but may nevertheless play a role in maintaining the locus in an "open for business" conformation (Bergsland et al. 2011). From 9.5 dpc, *Nes* expression is significantly lower in -255 homozygous embryos-indeed, at 9.5 dpc, the neural enhancer is required for approximately 80% of the *Nes* expression within the head. The activity of the enhancer remains functionally significant until at least 15.5dpc, although the differential between -255 and WT expression becomes less pronounced, suggesting that other neural enhancer(s) have increasingly important roles as the nervous system develops. It is interesting to compare our data with other recently published examples of developmental enhancer deletion. In the vast

majority of examples, deletion of a single conserved enhancer element has no impact on cognate gene expression (Osterwalder et al. 2018;Cadiz-Rivera et al. 2014). The Nestin enhancer therefore appears to be unusual in having such a profound impact on *Nes* expression. The mechanism that underpins this unusually high activity remains to be determined.

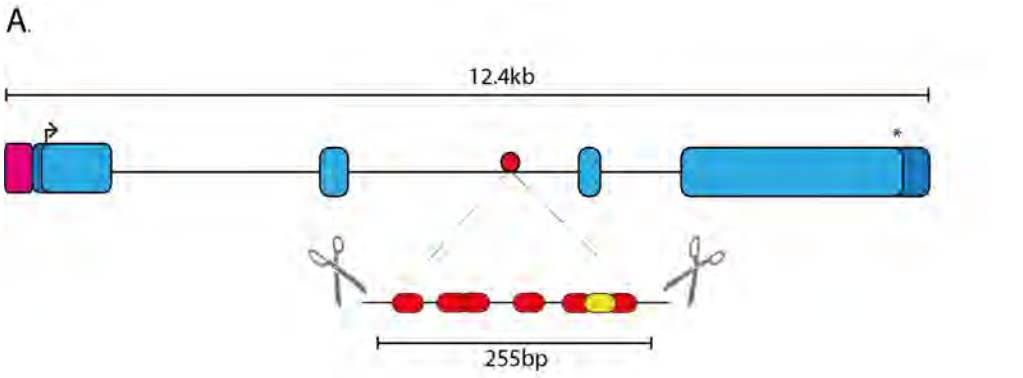
The expression of *Nes* mRNA throughout the neural tube is considerably affected in mutants lacking the 255bp enhancer, as seen in both *in situ* hybridization and qRT-PCR experiments. However, when protein expression is analysed via immunohistochemistry for these same mutants, there is not the same significant loss as would be expected. It is usually predicted when there is a large decrease in mRNA, the resulting protein expression would also be compromised. Decreased protein reactivity is seen in the -255 embryos, however the staining is still present throughout the neural tube where the mRNA is not visualized. This is possibly due to very low levels of Nestin expression within these cells, undetectable through *in situ* hybridization. It is also expected that *Nes* expression is controlled by other transcription factors other than SOX or POU proteins that are expressed within non-SOX/POU regions.

An unexpected finding of this study was that deletion of the *Nes* enhancer resulted in ectopic expression within the vasculature. Through confocal microscopy, NES was shown to co-localise with the endothelial cell marker CD31. Previously, NES has been reported to be expressed within the vasculature of different tissues such as developing kidneys, and also shown to be upregulated within vasculature following focal cerebral ischemia (Suzuki et al. 2010; Wagner et al. 2006; Shin et al. 2013), indicating a role in development and repair. Whilst the mechanism is unclear, it appears that the -255 deleted region also contains a repressor element that prevents NES expression in developing vasculature. While further studies are

required to determine the protein-sequence interaction(s) that mediate this repressor activity, it is worth noting that unmasking of repressor elements cannot be achieved using traditional enhancer activity assays such as transgenic reporter analysis.

Within the literature there are conflicting reports as to whether *Nes* is an essential gene in mice (Park et al. 2010; Mohseni et al. 2011; Dickinson et al. 2016). Through generation of two independent CRISPR KO mouse lines, we have shown that NES null mice are viable and do not exhibit overt deleterious phenotypes, consistent with two previous reports (Mohseni et al. 2011). In contrast, the NES null mice reported by Park et al (2010) exhibit embryonic lethality. The reason for this inconsistency remains unclear. Although not explored in this study, mild phenotypes such as impaired motor coordination (Mohseni et al. 2011) in KO mice suggest that NES function cannot be entirely replaced by other members of the IF family.

It is often assumed that all enhancers are cis-acting. However, it remains unclear whether some enhancers can also function in *trans* to activate cognate target gene(s). Trans enhancer interactions or transvection is well characterised in *Drosophila*, (Lim et al. 2018; Mellert & Truman 2012) but has rarely been observed in mammalian cells (Le Noir et al. 2017). Utilising a genetic approach, we provide evidence that the Nestin enhancer can undergo functional *trans* interactions in vivo. While the effect is relatively weak, these data raise the possibility that transvection of developmental enhancers may be more common than is currently recognised. Further investigation using chromatin capture technology would be beneficial in characterising these putative trans interactions.



B. Pup Genotypes

	wt/wt	-255/-255	-255/wt	% male
Observed	16%(11)	28%(19)	56%(38)	47%(32)
Expected	25%(17)	25%(17)	50%(34)	50%(34)

Chi-squared score: 2.824
p-value: 0.2437

C. Embryo Genotypes

	wt/wt	-255/-255	-255/wt
Observed	21%(33)	30%(46)	49%(75)
Expected	25%(39)	25%(39)	50%(77)

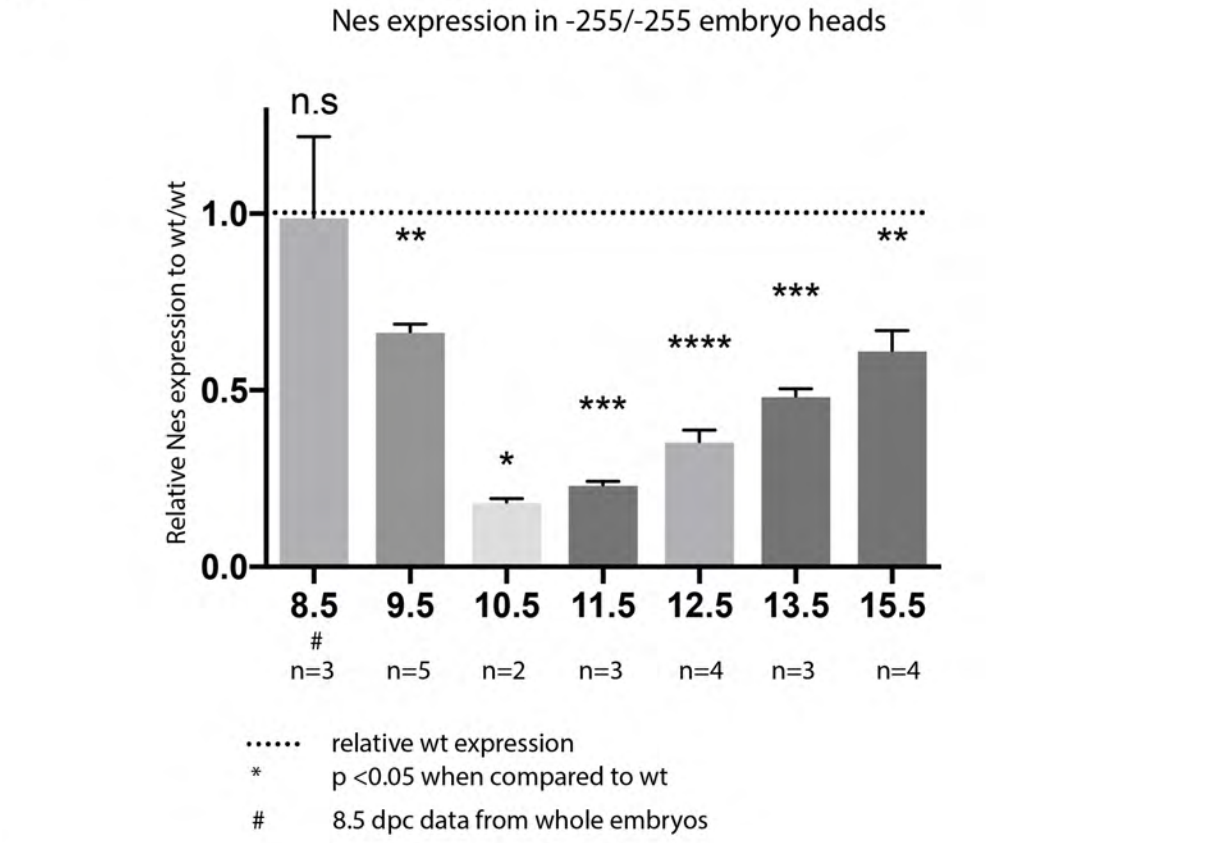
Chi-squared score: 2.299
p-value: 0.3168

Figure 1

Generation of Nes Enhancer deletion (-255/-255) mouse line

A . Guide RNAs (scissors) were designed to flank the six SOXB1 sites (red) and the POU site (yellow) within the second intron of the Nestin gene. Arrow indicates start codon, asterisk indicates stop codon, pink square denotes promoter region and dark blue, the 5' and 3' UTRs. This generated a 255bp deletion encompassing all SOXB1 sites identified by ChIP-seq. Observed/Expected tables of both live pups born (B) and transient embryos (C) show there is no embryonic lethality or sex bias evident in wildtype, heterozygous or homozygous animals.

A.



B.

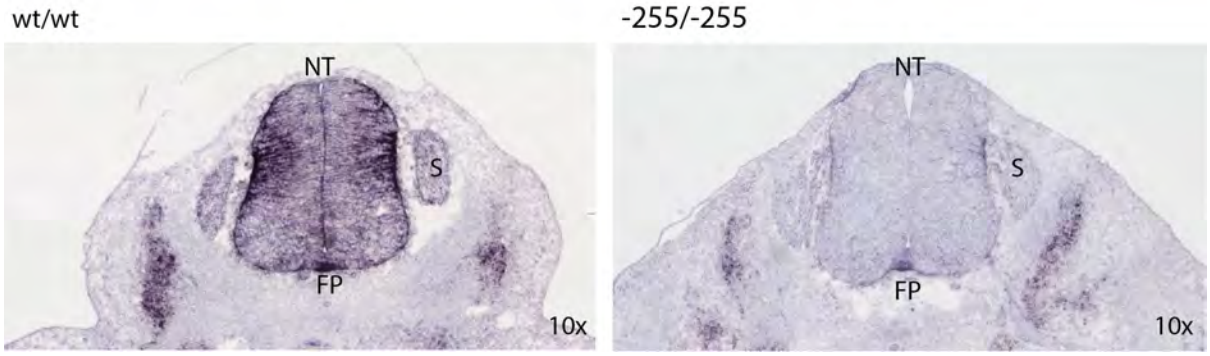


Figure 2

Loss of Nestin mRNA during embryonic development

A . Analysis of embryonic heads from aged 8.5 to 15.5 by qRT-PCR. All values are normalized to wildtype samples of the same developmental stage. Due to size constraints, whole 8.5 dpc embryos were used rather than embryonic heads. From 9.5dpc the level of Nes decreases in the -255/-255 embryos, and progressively increases from 11.5dpc. * indicates p-value <0.05, ** indicates p-value <0.01, ***

indicates p-value <0.001, **** indicates p-value <0.0001. Error bars represent the standard deviation of the mean. B. *In Situ* hybridization of Nes mRNA in an 11.5dpc neural tube section shows large amounts Nes mRNA throughout the neural tube in the wildtype sample. In the -255/-255 section, the majority of the staining is lost, although somite staining outside the neural tube remains as well as a small region at the base of the neural tube. NT denotes neural tube, FP denotes floor plate and S denotes somite.

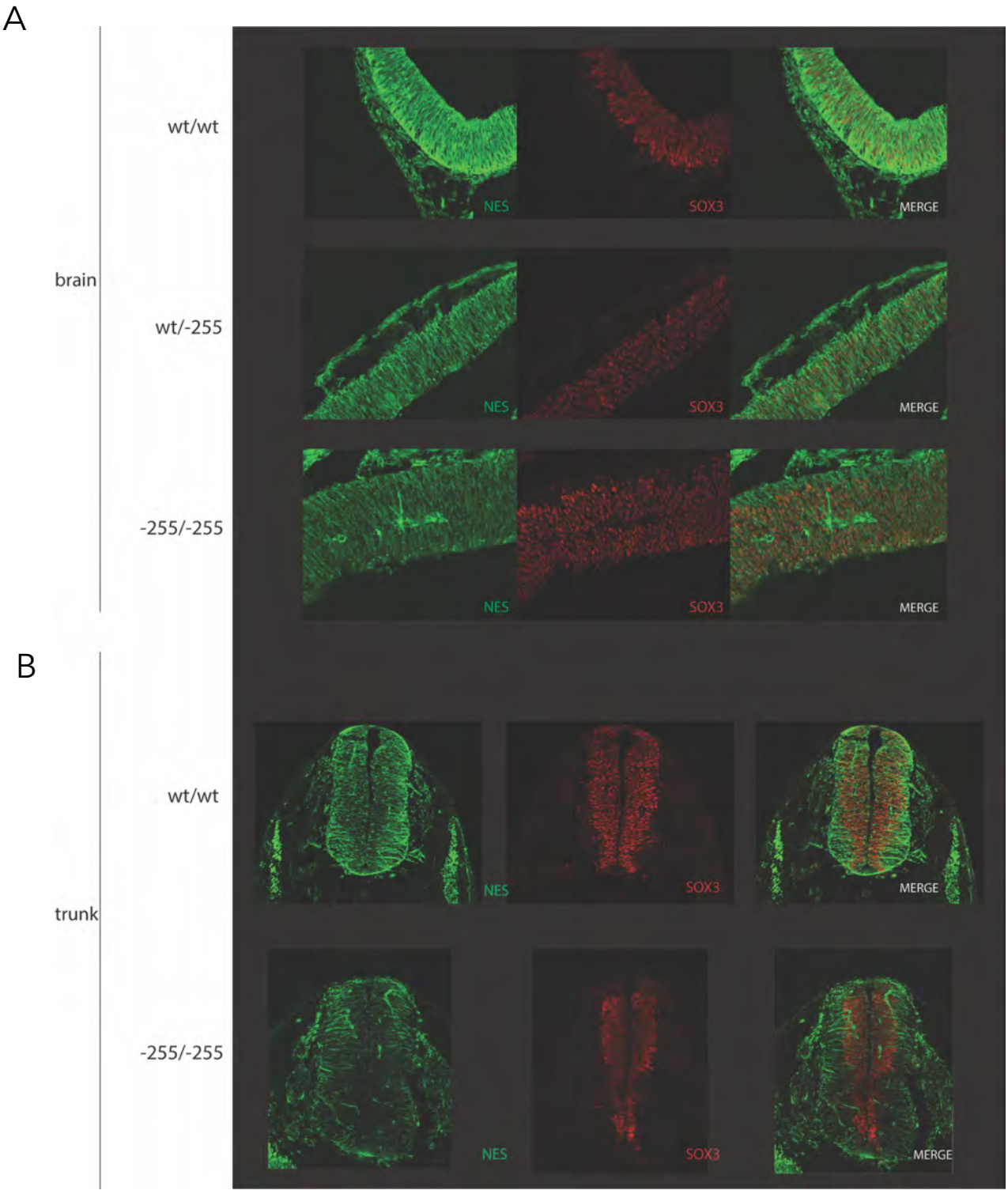


Figure 3

Immunohistochemical analysis of brain and trunk sections

Wildtype (wt/wt), heterozygous (-255/wt) and homozygous (-255/-255) transverse cortex (A) and trunk (B) sections labelled with anti-SOX3 and anti-NESTIN antibodies. The Nestin reactivity is decreased in both wt/-255 and -255/-255 sections, while the SOX3 remains consistent across genotypes.

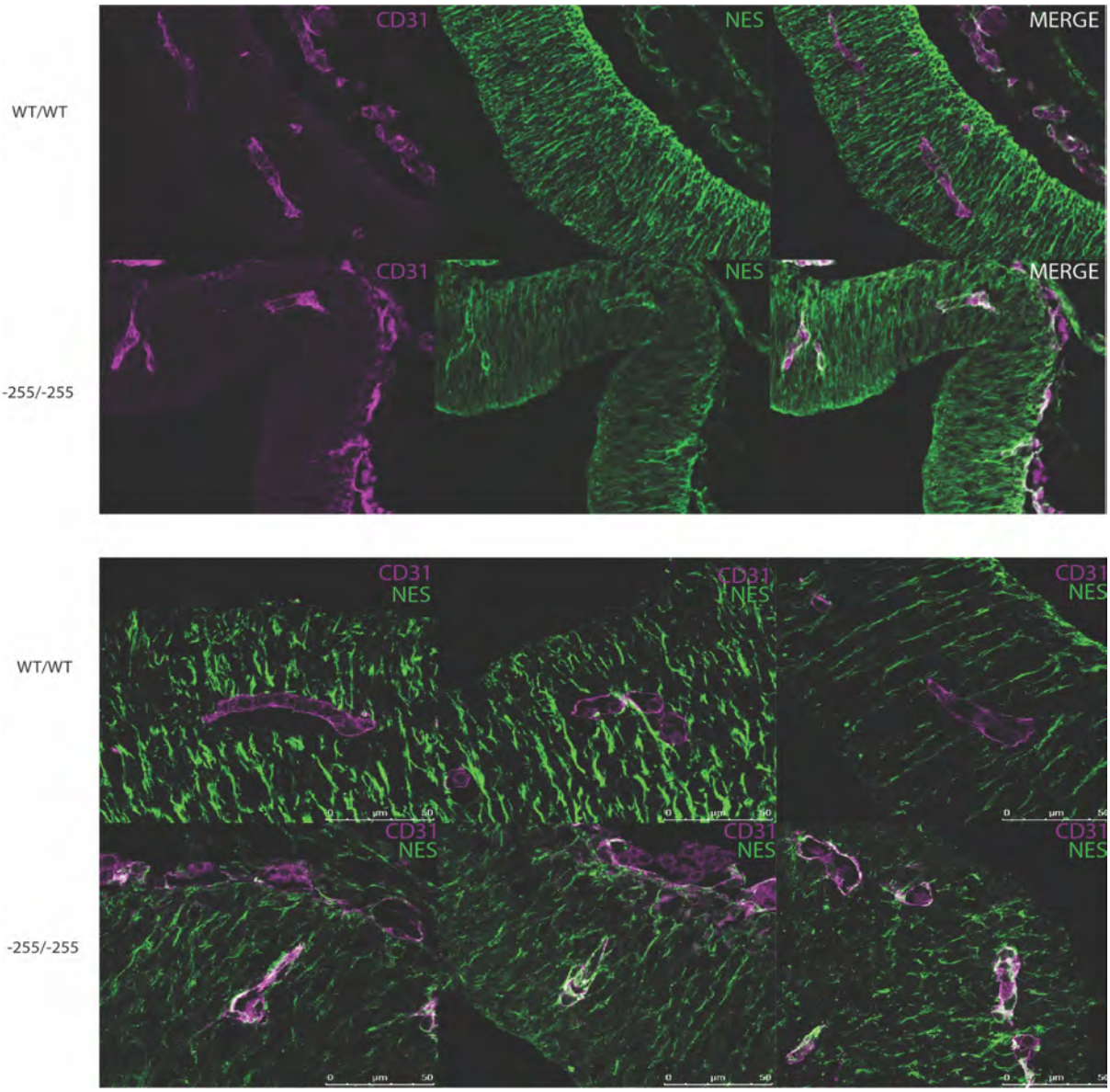


Figure 4

Immunohistochemical analysis of Nestin reactivity within vasculature of 10.5 dpc cortex sections

Confocal microscopy of wt/wt and -255/-255 10.5 dpc cortex sections labelled with NESTIN and CD31 to mark epithelial cells of the vasculature. Co-localisation (white) of Nestin and CD31 is apparent in the -255/-255 sections, and not seen in wildtype sections.

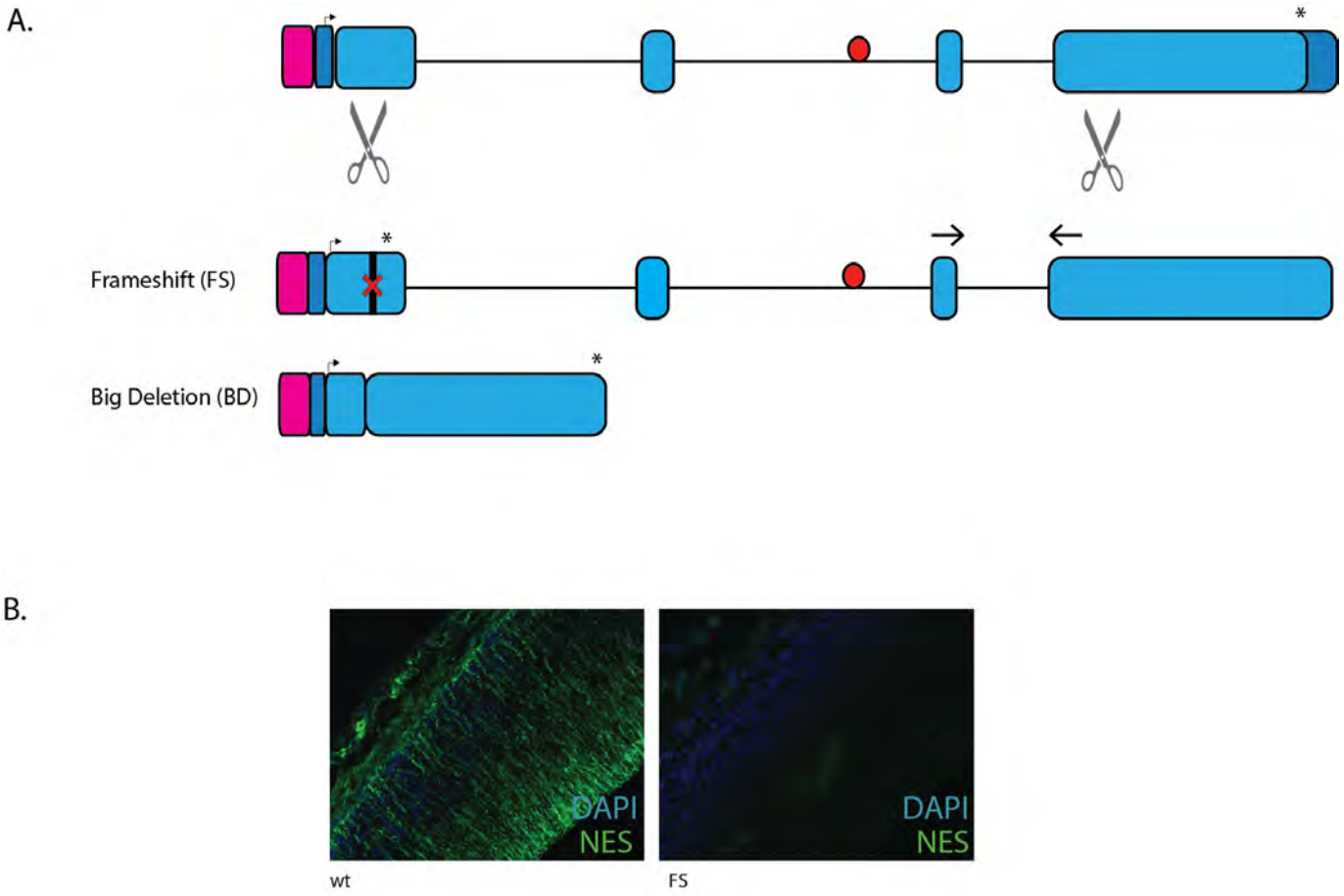


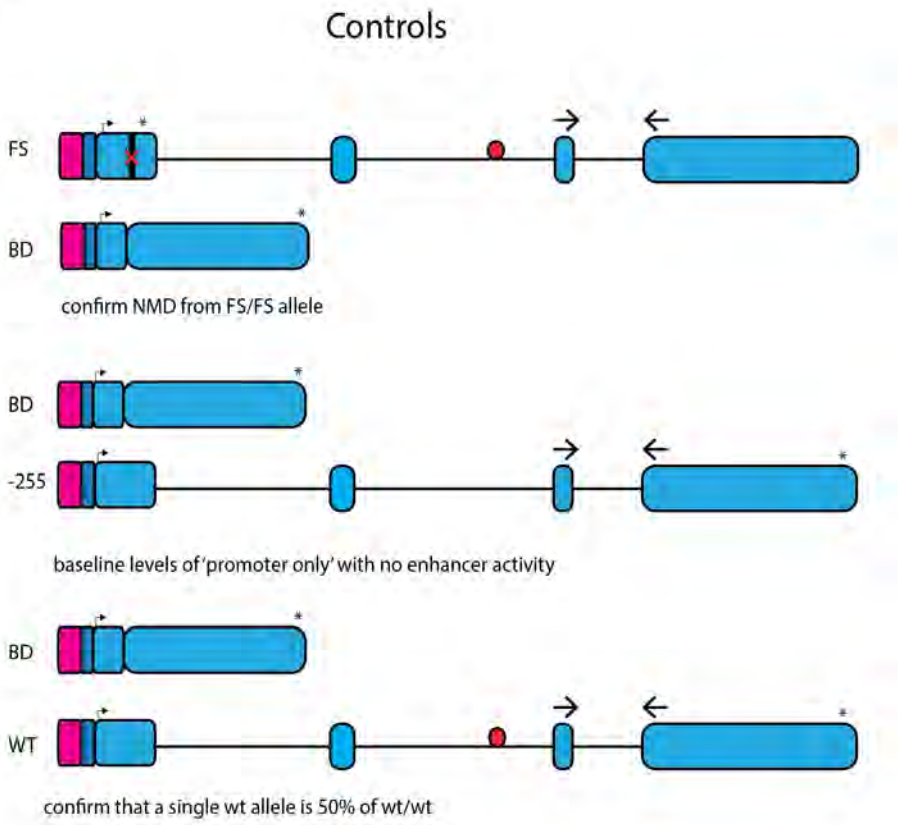
Figure 5

Generation of Nes null lines

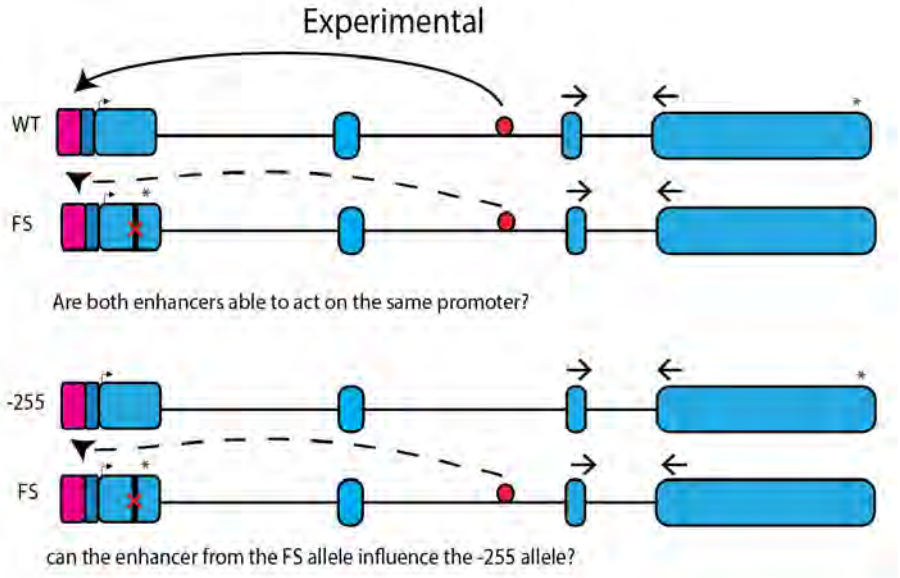
A. CRISPR guide sequences (scissors) designed to cut within exon 1 and exon 4 of the *Nes* gene. The FS allele generated a frameshift mutation at codon 50, while the BD allele removed the 8.7 kb of intervening sequence. qRT-PCR primers indicated by the arrows amplify the FS, wt and -255 alleles. Pink box indicates the promoter, dark blue is 5' and 3' UTR, pale blue is coding regions, red circle is *Nes* enhancer, arrow is transcriptional start site and asterix is the stop codon.

B. Immunohistochemical staining with NES antibody on 'frameshift' embryonic cortex shows no discernable protein product.

A.


Controls
FS
BD
confirm NMD from FS/FS allele
BD
-255
baseline levels of 'promoter only' with no enhancer activity
BD
WT
confirm that a single wt allele is 50% of wt/wt

B.


Experimental
WT
FS
Are both enhancers able to act on the same promoter?
-255
FS
can the enhancer from the FS allele influence the -255 allele?

C.

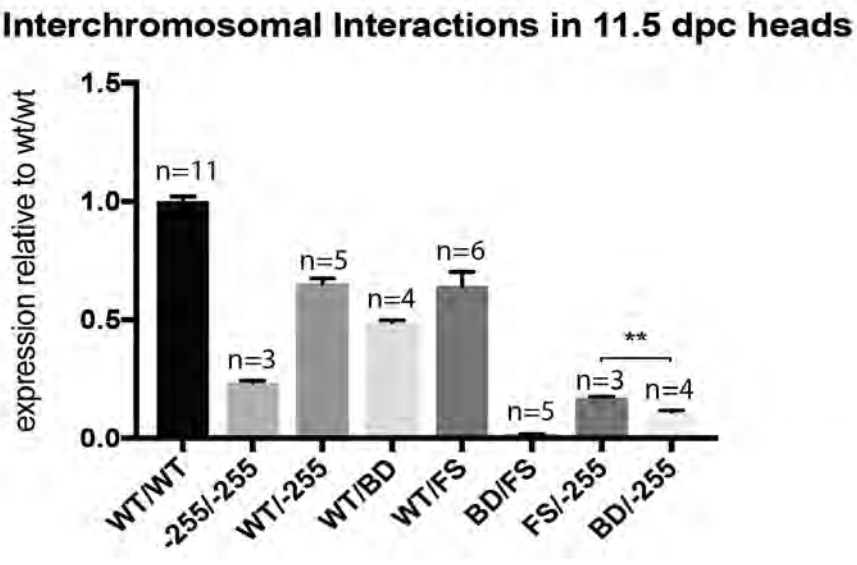

Interchromosomal Interactions in 11.5 dpc heads
expression relative to wt/wt
1.5
1.0
0.5
0.0
n=11
n=3
n=5
n=4
n=6
n=5
n=3
n=4
**
WT/WT
-255/-255
WT/-255
WT/BD
WT/FS
BD/FS
FS/-255
BD/-255

Figure 6.

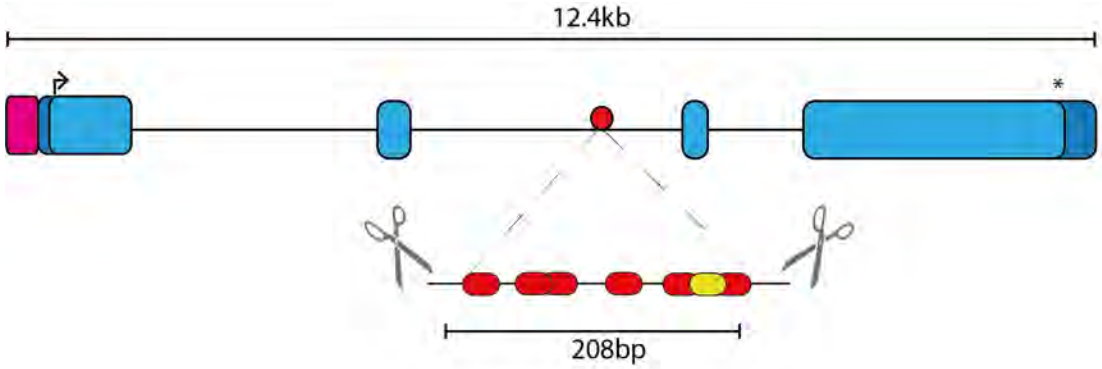
Interchromosomal Interactions of the Nestin enhancer and promoter

A . Control crosses to determine *trans* interactions. The first will determine if nonsense mediated decay occurs in the 'big deletion' (BD) allele. The second determines the baseline level of 'promoter only' activity when no enhancer is present. The third will confirm is a single allele produces exactly half of the total wildtype mRNA. qRT-PCR primers indicated by the arrows amplify the FS, wt and -255 alleles. Pink box indicates promoter, dark blue is 5' and 3' UTR, pale blue is coding regions, red circle is *Nes* enhancer, arrow is transcriptional start site and asterix is the stop codon.

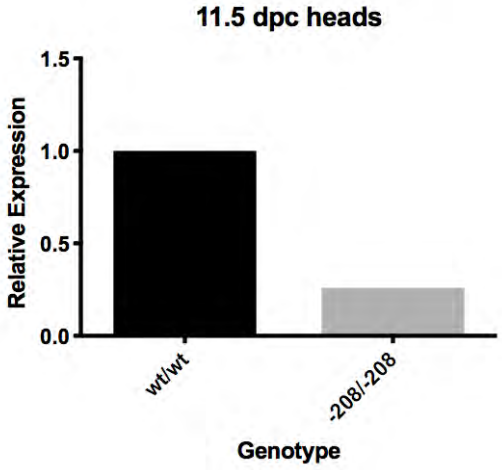
B. Experimental workflow to determine *trans* interactions. The first will determine whether both copies of the *Nes* enhancer are capable of influencing only one functional promoter. The second will determine if an enhancer on one allele can compensate for the loss of the enhancer on another allele. qRT-PCR primers indicated by the arrows amplify the FS, wt and -255 alleles. Pink box indicates promoter, dark blue is 5' and 3' UTR, pale blue is coding regions, red circle is *Nes* enhancer, arrow is transcriptional start site and asterix is the stop codon.

C. The qPCR results of the above experimental crosses and embryo analyses on 11.5 dpc heads. By using various mating of FS, BD, wt and -255 alleles embryonic heads were analysed for changed in Nestin gene expression. The BD/-255 produces significantly less Nestin mRNA that the FS/-255, indicating the presence of a single enhancer on one chromosome can interact with the promoter of another. Unpaired t-test between FS/-255 and BD/-255 show p-value of 0.0011, other t-test results not shown for clarity. Error bars represent the standard deviation of the mean.

A



B



Supplementary Figure 1

The -208 Nestin enhancer deletion line shows a reduction in *Nes* expression within the head similar to that of the -255 line. These preliminary findings were performed with n=2 embryos.

References

Bergsland, M. et al., 2011. Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev*, 25(23), pp.2453–2464.

Cadiz-Rivera, B. et al., 2014. The Chromatin "Landscape" of a Murine Adult **β**-Globin Gene Is Unaffected by Deletion of Either the Gene Promoter or a Downstream Enhancer M. Fernandez-Zapico, ed. *Plos One*, 9(5), p.e92947.

Cong, L. et al., 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, 339(6121), pp.819–823.

Dahlstrand, J., Lardelli, M. & Lendahl, U., 1995. Nestin mRNA expression correlates with the central nervous system progenitor cell state in many, but not all, regions of developing central nervous system. *Developmental Brain Research*, 84(1), pp.109–129.

Dickinson, M.E. et al., 2016. High-throughput discovery of novel developmental phenotypes. *Nature*, 537(7621), pp.508–514.

Dubois, N.C. et al., 2006. Nestin-Cre transgenic mouse line Nes-Cre1 mediates highly efficient Cre/loxP mediated recombination in the nervous system, kidney, and somite-derived tissues. *Genesis*, 44(8), pp.355–360.

Kagey, M.H. et al., 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 467(7314), pp.430–435.

Kvon, E.Z., 2015. Using transgenic reporter assays to functionally characterize enhancers in animals. *Genomics*, 106(3), pp.185–192.

Le Noir, S. et al., 2017. The IgH locus 3' cis-regulatory super-enhancer co-opts AID for allelic transvection. *Oncotarget*, 8(8), pp.12929–12940.

Lendahl, U., Zimmerman, L.B. & McKay, R.D., 1990. CNS stem cells express a new class of intermediate filament protein. *Cell*, 60(4), pp.585–595.

Lim, B. et al., 2018. Visualization of Transvection in Living Drosophila Embryos. *Molecular Cell*, 70(2), pp.287–296.e6.

Lothian, C. & Lendahl, U., 1997. An evolutionarily conserved region in the second intron of the human nestin gene directs gene expression to CNS progenitor cells and to early neural crest cells. *European Journal of Neuroscience*, 9(3), pp.452–462.

McAninch, D. & Thomas, P., 2014. Identification of highly conserved putative developmental enhancers bound by SOX3 in neural progenitors using ChIP-seq. *PloS ONE*.

Mellert, D.J. & Truman, J.W., 2012. Transvection is common throughout the Drosophila genome. *Genetics*, 191(4), pp.1129–1141.

Michalczyk, K. & Ziman, M., 2005. Nestin structure and predicted function in cellular cytoskeletal organisation. *Histology and Histopathology*, 20(2), pp.665–671.

Mohseni, P. et al., 2011. Nestin is not essential for development of the CNS but required for dispersion of acetylcholine receptor clusters at the area of neuromuscular junctions. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(32), pp.11547–11552.

Osterwalder, M. et al., 2018. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature*, 554(7691), pp.239–243.

Park, D. et al., 2010. Nestin is required for the proper self-renewal of neural stem cells. *Stem Cells*, 28(12), pp.2162–2171.

Petersen, P.H. et al., 2002. Progenitor cell maintenance requires numb and numblike during mouse neurogenesis. *Nature*, 419(6910), pp.929–934.

Shin, Y.-J. et al., 2013. Characterization of nestin expression and vessel association in the ischemic core following focal cerebral ischemia in rats. *Cell and tissue research*, 351(3), pp.383–395.

Suzuki, S. et al., 2010. The neural stem/progenitor cell marker nestin is expressed in proliferative endothelial cells, but not in mature vasculature. *Journal of Histochemistry & Cytochemistry*, 58(8), pp.721–730.

Tanaka, S. et al., 2004. Interplay of SOX and POU factors in regulation of the Nestin gene in neural primordial cells. *Molecular and Cellular Biology*, 24(20), pp.8834–8846.

Trumpp, A. et al., 1999. Cre-mediated gene inactivation demonstrates that FGF8 is required for cell survival and patterning of the first branchial arch. *Genes & Development*, 13(23), pp.3136–3148.

Wagner, N. et al., 2006. Intermediate filament protein nestin is expressed in developing kidney and heart and might be regulated by the Wilms' tumor suppressor Wt1. *American Journal of Physiology-Regulatory Integrative and Comparative Physiology*, 291(3), pp.R779–87.

Zimmerman, L. et al., 1994. Independent Regulatory Elements in the Nestin Gene Direct Transgene Expression to Neural Stem-Cells or Muscle Precursors. *Neuron*, 12(1), pp.11–24.

Chapter Three

Identification and *in vivo* validation of a Frizzled 3 neuroprogenitor enhancer bound by SOXB1 proteins

Summary

This manuscript outlines the identification and validation of a SOXB1 bound enhancer controlling the expression of *Fzd3* within neuroprogenitors of the CNS. Previous work within the laboratory identified a region bound by SOX3 and the co-activator protein p300 located within an intron of the Wnt-receptor Fzd3.

To assess if the putative enhancer was functional, mouse models were generated using CRISPR that deleted the SOXB1 motif as well as 573bp of surrounding sequence. Through RNA expression analysis, we are able to show that this region influences *Fzd3* expression within the developing neural tube.

We also used traditional enhancer validation techniques to further understand the functionality of the putative enhancer. Through generation of LacZ transgenic embryos we could show the enhancer drives expression towards the floor plate of the neural tube, an important region required for axon development.

Fzd3 is an essential gene. Homozygous loss leads to death soon after birth, however heterozygotes appear phenotypically normal. We generated our own *Fzd3* null line, and used this in combination with the enhancer deletion allele to show that only 40% of WT *Fzd3* expression is sufficient for survival.

Through these experiments we have provided the first evidence of an enhancer controlling *Fzd3* expression within the CNS, as well as the first functional interactions between *Fzd3* and the SOXB1 proteins.

Statement of Authorship

Title of Paper	Identification, and in vivo validation of a Frizzled 3 neuroprogenitor enhancer bound by SOXB1 proteins
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style
Publication Details	

Principal Author

Name of Principal Author (Candidate)	Ella Thomson
Contribution to the Paper	Designed and performed experiments, analysed results, wrote 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 4/7/19

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 to 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	Fatwa Adikusuma
Contribution to the Paper	Designed study, generated enhancer deletion mouse model and performed experiments
Signature	Date 4/7/19

Name of Co-Author	Chee Ho H'ng
Contribution to the Paper	Generated Fzd3 knockout mouse model and performed experiments
Signature	Date 4/7/19

Name of Co-Author	Sandra Piltz
Contribution to the Paper	Performed all microinjection experiments to generate mouse models
Signature	Date 4/5/19
Name of Co-Author	Paul Thomas
Contribution to the Paper	Conceived and designed study, analysed results, evaluated and edited manuscript.
Signature	Date 4/7/19

Abstract

Enhancers are essential to direct gene expression during neural development. SOXB1 proteins (SOX1, SOX2 and SOX3) are important transcriptional regulators within neural progenitor cells, which act upon enhancers to maintain a stem-like state. While the transcriptional targets of these proteins are being elucidated through chromatin immunoprecipitation experiments such as ChIP-Seq, the vast majority of putative targets have not been validated. Here we identify a novel SOXB1 target in the Wnt protein receptor gene, Frizzled 3 (*Fzd3*). Located intronically, this 573bp region is highly conserved, and bound by both SOX2, SOX3 and the enhancer co-activator p300 in neuroprogenitor cells. Using CRISPR, we generated mouse models lacking either the individual SOXB1 TFBS or the entire enhancer region. This shows that loss of an individual motif only causes very minimal reduction in *Fzd3*, whereas the larger deletion visibly reduced the *Fzd3* levels within both the floorplate and intermediate zone of the developing neural tube. Enhancer activity was assessed via generation of a LacZ reporter mouse carrying the putative 573bp enhancer. This showed enhancer activity in the floor plate, but interestingly not the intermediate zone. This raises an important issue in enhancer validation, whereby the endogenous chromatin and epigenetic landscape is needed to fully understand how the enhancer is acting upon target gene expression. Further to this, we generated an independent FZD3 KO mouse, known to cause perinatal lethality. This was used to create a compound heterozygote model (-573/KO), producing approximately 40% of wt *Fzd3* mRNA. These mice do not appear to have any phenotypic effects, suggesting only a small proportion of the wildtype FZD3 is needed during development. Through these experiments we provide the first functional link between the SOXB1 proteins and the regulation of *Fzd3* during neurogenesis.

Introduction

Neural development is a tightly controlled process that requires precise spatial and temporal activity of hundreds of genes. One of the ways in which this achieved is through the regulatory elements such as enhancers. Putative enhancers are increasingly being identified through chromatin immunoprecipitation approaches such as ChIP-Seq, which provide genome-wide assessment of TF binding (Mundade et al. 2014; Visel et al. 2009). Typically, these studies yield hundreds, if not thousands of sites but do not provide information about which interactions are biologically significant. For this reason, putative enhancers identified using these techniques must be functionally validated. Traditionally, enhancer testing was performed using transgenic mice carrying a reporter gene (e.g. LacZ) linked to a putative enhancer region (Kvon 2015). This is a useful approach and can identify the embryonic stage and regions where the enhancer is sufficient for expression. However, this approach does not functionally replicate the genomic or epigenomic context of the putative enhancer *in vivo*, and as such may yield limited or even misleading data. Recently, with the advent of new and simpler genome editing technologies such as CRISPR, fast and precise deletion of genomic regions in mouse zygotes is now possible (Cong et al, 2013; Mali et al. 2013). By removing the putative enhancer from its native context, its contribution to gene expression and the phenotypic effect of its loss can be tested *in vivo*, and can be analysed.

The SOX (Sry-related High Mobility Group Box) family of TFs are expressed in virtually all developing tissues and in most cases are essential for normal embryonic development. The SOXB1 proteins (SOX1, SOX2 and SOX3) are a closely-related sub-group of SOX factors that function as transcriptional activators to maintain the undifferentiated state of neural stem/progenitor cell (Bergsland et al. 2011; Sarkar & Hochedlinger 2013). Previously published SOX3, SOX2 and p300 co-activator ChIP-

seq experiments have identified a region within the second intron of Frizzled 3 (*Fzd3*) as a putative developmental enhancer in neuroprogenitor cells (McAninch & Thomas 2014). Fzd3 encodes a Wnt protein receptor that is widely expressed in the developing nervous system including the neuroprogenitors throughout the CNS (Wang et al. 2002; Sala et al. 2000). It has roles within the central nervous system, such as axon growth and guidance, and is also essential for the formation of many major fibre tracts within the brain (Wang et al. 2002; Wang et al. 2006). Homozygous loss of *Fzd3* leads to death shortly after birth, and the inability of the neural tube to fold. These mice also present with curly tails, an indicator of neural tube maldevelopment (Wang et al. 2002).

We hypothesised that SOXB1/p300 intronic binding demarcates a novel *Fzd3* neurodevelopmental enhancer. Using CRISPR, we generated a mouse model lacking the putative enhancer and showed that this region is essential for complete *Fzd3* expression during CNS development. Further, we show that only 40% of *Fzd3* expression is required for normal brain development. This work establishes a direct link between *Fzd3* regulation and the SOXB1 proteins and demonstrates the utility of CRISPR mutagenesis for functional assessment of putative enhancers in vivo.

Materials and Methods

Generation of Enhancer Deletion and Knockout Mouse

Guides for CRISPR experiments were designed on the crispr.mit.edu web tool. Enhancer Deletion guide (TTAGCAAGGGTGTGAAAAG) Knockout guides (1- AGTTATAATGTAAAGGGCCG, 2- GCTCACTCTTACAACACTAC). Guides were cloned into the PX330 plasmid (Addgene) and the tracR sequence and T7 promoter were added via PCR (TTAATACGACTCACTATAGCTTAGCAAGGGTGTGAAAAG GTTTTAGAGCTAGAAATAGC). PCR products were transcribed to RNA using the T7 RNA transcription kit (NEB) and purified with RNEasy Kit (QIAGEN) to generate sgRNA. Cas9 mRNA was synthesis from XhoI digested Cas9 plasmid (Addgene) using the Mmessage Mmachine T7 Ultra Transcription Kit.

C57BL6/J females were superovulated with Pregnant Mare Serum Gonatropin (PMSG) and human Chorionic Gonatropin (hCG) prior to mating with C57BL6/J males for zygote harvesting. Single cell zygotes were collected on the day of microinjection, and treated with hyaluronidase to remove surrounding cumulus cells. Cytoplasmic injection performed with CRISPR reagents (50ng/uL Cas9 mRNA, 100ng/uL sgRNA) before transfer into pseudopregnant CD1 females.

Regular colony and embryo genotyping performed with primers flanking deleted sequence (enhancer deletion line F- AGGCTGTTCCACATTGGTTC, R- CATCTGCATAAACCCACACTC). Knockout genotyping performed using 3 primers to distinguish between alleles (F- AGCCCAGTGTTAGAGTATAGCCAG, M- TCCTAGCCCTTCCACCCTATG, R- CTGCCTCATCTTCCCAAATGC) using KAPA 2G Fast MasterMix, KAPA, or EpiCentre Buffer J with Roche Taq Polymerase.

All mouse breeding and experimental work was performed at the University of Adelaide in accordance with relevant ethics approvals (S-201-2013 and S-173-2015).

Tissue Preparation

Pregnant females were culled via cervical dislocation and embryos harvested from 10-12dpc. Tails were removed for genotyping. If tissue was required for qRT-PCR analysis, it was immediately snap frozen on dry ice and stored at -80°C until RNA extraction. If tissue was required for staining, it was kept overnight in 4% PFA, washed 3 times in PBS and then stored in 30% sucrose as a cryoprotectant before flash freezing in OCT for storage at -20°C.

qRT-PCR analysis

RNA was extracted from flash frozen embryo heads using Trizol, according to manufacturer's instructions. Briefly, heads were homogenised in 500uL Trizol, 100uL chloroform added to mixture and centrifuged at 6000xg for 30mins. The aqueous layer was removed and equal amount of 70% EtOH added. The solution was then placed in RNEasy spin column and centrifuged at 13000 rpm for 1 minute. Column washed with 2x Buffer RLT (Qiagen), and purified RNA eluted in 30uL of RNAse free H20, and stored at -20C. RNA samples converted to cDNA using AB Systems High Capacity RNA to cDNA Kit. SYBR Fast standard protocols used for qPCR with samples run in quadruplicate. *B-actin* (F-CTGCCTGACGGCCAGG, R-GATTCCATACCCAAGAAGGAAGG) was used to normalise cDNA levels across samples, and *Fzd3* primers used to assess *Fzd3* levels. For *Fzd3* knockout samples, exon 3 primers were used F-GACATGCTTTGAATGGGCCAG, R-CAAAGTCAGGTTCCTGGAGCAC).

In situ hybridisation

Probes for in situ hybridisation were designed to target the 5' end of the Fzd3 gene primers (F- GACATGCTTTGAATGGGCCAG, R- CACATGGCACCAGCATGAACC) Primers corresponding to the region were used to amplify the DNA from wildtype mouse DNA and incorporate a T7 promoter at the 5' end. The DNA was converted to RNA using the T7 In Vitro Transcription Kit (NEB), followed by DNase I (NEB) treatment and purification with RNEasy kit (Qiagen).

Embryo trunks sectioned at 16um on a cryostat (Leica CM1900) and stored at -20°. Prior to in situ hybridisation, slides were defrosted for 1hr at RT. The RNA in situ probe was denatured at 72°C for 2 minutes and kept on ice. 100ul hybridisation buffer containing 1uL diluted riboprobe/slide was added to the slides and kept in a humidified chamber containing formamide overnight at 65°C. Slides were washed 3x 30 mins at 65°C in Wash Buffer (50% formamide, 5% 20x SSC), then 330 minute washes in MABT (Maleic Acid Buffer + 0.1% Tween-20) at room temperature. Slides were blocked with 300uL Blocking Solution (Blocking Reagent, Sheep Serum, MABT) and kept in a humidified chamber at room temperature for 2 hours. 75uL of anti-DIG antibody was diluted in Blocking Solution and added to slides and kept overnight at room temperature in a humidified chamber. Antibody removed with 4 20 minute washes in MABT, then wash 2 10 minutes washes in Alkaline Phosphatase Staining Buffer (4M NaCl, 1M MgCl2, 1M Tris pH 9.5). Slides were then stained with 95uL staining solution (NBT, BCIP, Alkaline Phosphatase Staining Buffer), coverslipped, and kept in the dark at room temperature overnight. Staining removed by washing 3times for5mins in PBS, and fixed with 300uL 4% PFA added to slides and incubated for 1 hour in a sealed contained. Fixative was washed off with 3 10min PBS washes, and 50uL mowiol added to each slide for mounting with a

coverslip. Slides were analysed using brightfield microscopy on Nikon Eclipse Ti Microscope using ND2 Elements software.

LacZ Reporter Mouse Generation

A vector containing the Hsp68 minimal promoter followed by the LacZ gene (hsp-LacZ) was digested with PstI and XhoI. Primers were designed to amplify the Fzd3 enhancer fragment and incorporate PstI and XhoI restrictions sites (F-TAAGCACTGCAGCACTGTCCCCTTTTCACACCC, R – ACGAATCTCGAGGC ATGAGCAGAGAATGTGGAC). The PCR product was digested and ligated into the digested vector, generating the Fzd3-Hsp68-LacZ vector. Prior to injection, the plasmid digested with XhoI and NotI to release fragment. The fragment was gel extracted and purified with QIAGEN gel extraction kit. CRISPR components were also included in the injection, to generate a cut-site at the *Rosa26* locus (guide RNA - ACTCCAGTCTTTCTAGAAGA) to help increase the chance of integration at a cut-site.

The injection mix (4ng/uL Fzd3-Hsp-LacZ, 25ng/uL Rosa26sgRNA, 0.05ug/uL Cas9 protein) was injected into single cell zygotes and transferred at the 1 cell stage to pseudopregnant foster mothers. Where foster mothers were not available immediately post injection, injected embryos were cryopreserved until females available for transfer.

Embryos were dissected at 11.5dpc and immediately fixed in 4% PFA for 45 mins, and washed 3 times in PBS. Embryos were stained in 1mL staining solution (2mM MgCl2, 5mM K3Fe(CN)6, 5mM K4Fe(CN)6, 0.01% sodium deoxycholate, 10% igepal, 1mg/mL X-gal) at 37C overnight in the dark. Embryos were imaged using a dissecting microscope and post-fixed again in 4% PFA overnight and cryopreserved in 30% sucrose until set in OCT and sectioned on a cryostat (Leica CM1900).

Immunohistochemistry

Trunks and heads were sectioned at 16um on a cryostat (Leica CM1900) and slides washed 3 times for 10mins in PBT (1xPBS, 0.25% Triton-X), blocked for 30min in BS (1x PBS, 0.25% Triton-X, 10% Horse Serum), and stained overnight with 200uL primary antibodydiluted in BS and kept in humidified chamber 4°C. Primary antibody removed with 3 10 minute PBS washes. 200 uL of secondary antibody was diluted in BS added to slides and kept in dark humidified chamber for 4hrs at room temperature. Secondary antibody removed with 310 minute washes in PBS. Slides were dried and set with Prolong Gold Antifade + DAPI (Molecular Probes) and coverslip applied. Slides kept overnight in the dark before image acquisition Nikon Eclipse Ti Microscope using ND2 Elements software. Images modified for colour, brightness and contrast using Adobe Photoshop v7 (Adobe Systems). Primary antibodies used; anti-SOX2 (Abcam, ab15830), anti-Neurofilament-H (Abcam, ab4680).

Results

Generating a putative Fzd3 enhancer deletion mouse model

Through overlay of SOX3, SOX2 and p300 ChIP-Seq data from mouse neuroprogenitor cells, we identified a putative highly conserved enhancer in the second intron of the *Fzd3* gene containing a single SOXB1 binding motif (McAninch & Thomas 2014). To investigate the role of the putative enhancer in vivo, we sought to generate mouse lines lacking the enhancer sequence. Using a CRISPR/Cas9 gRNA that binds immediately upstream of the SOXB1 binding site, we generated two mouse lines containing an 8bp deletion (-8) (Supplementary Figure 1) and a 573bp (-573) deletion (Figure 1). For the majority of experimental analysis, the -573 line was used as this deleted the SOXB1 motif as well as downstream DNA, whereas the -8bp deletion only partially interrupts the SOXB1 motif. All mice generated from both the -573 and -8 colonies appeared phenotypically normal with no obvious morphological defects.

Fzd3 mRNA Expression is altered in Enhancer Deleted Mice

To assess the impact of putative enhancer deletion on endogenous *Fzd* expression, we initially performed qRT-PCR of wild type (WT) and homozygous -573 (-573/-573) embryos (Figure 2). We compared *Fzd3* expression in embryonic heads at ages 10.5 dpc, 11.5 dpc and 12.5 dpc, when FZD3 and SOXB1 proteins are expressed within the neural progenitor population (Wang et al. 2016; Pevny & Placzek 2005). At each time point there is a significant reduction (20-40%) in *Fzd3* expression, indicating that the enhancer is active during early brain development. A slight but significant reduction in *Fzd3* expression was also detected in -8/-8 embryo heads at 11.5 dpc (Supplementary Figure 1).

Next, we assessed *Fzd3* expression in the 11.5 dpc neural tube, where neuroprogenitors undergo stereotypical positional-dependent differentiation with concomitant downregulation of *Fzd3*. In WT embryos, *Fzd3* was expressed throughout the midline and ventricular zone, continuing into the floor plate, and was absent from the intermediate zones. This correlates almost perfectly to SOX2 protein expression (Figure 2). In contrast, *Fzd3* expression in -573 homozygotes was greatly reduced in the floor plate and ventricular zone and in the latter was retained only in the lateral edges. This is particularly obvious when compared with the SOX2 staining (which is unchanged between WT and -573-deletion embryos). Together these data suggest that the -573 deletion encompasses a *Fzd3* neuroprogenitor enhancer.

The -573bp region drives floor plate reporter expression in transgenic embryos

To determine if the *Fzd3* enhancer is sufficient for expression in neuroprogenitor cells in vivo, we generated 11.5dpc transgenic embryos containing the -573 element linked to a minimal promoter driving LacZ. Whole mount staining revealed that the 7 transient transgenic embryos were positive for LacZ reporter activity (Figure 3). Whilst some variation was observed, staining was consistently detected in the midline region throughout the extent of the neuroaxis, as well as the eye ad midbrain. Section analysis revealed robust LacZ activity in the floor plate in all 6 embryos that were sectioned (Supplementary Figure 2). Expression in the neural tube progenitors was generally absent apart from one embryo with patchy staining. These data suggest that the -573 enhancer can function in isolation to drive expression in the floor plate but not in more dorsal neuroprogenitors.

Fzd3 Knockout

It has been shown previously that *Fzd3* null mice have significant neurodevelopmental defects and die shortly after birth while heterozygotes are apparently unaffected. Given the (~30%) reduction in *Fzd3* expression in -573/-573 embryos, it is possible that neural development in -573/null compound heterozygotes will be compromised due to *Fzd3* levels not exceeding a critical threshold. To test this hypothesis, we first sought to generate a *Fzd3* null (KO) allele. We identified 2 high-scoring gRNAs targeting intronic sequences flanking exon 3, deletion of which is predicted to cause an early frameshifting mutation (Figure 4). Of the 24 founder pups born from zygotic injections, 11 died soon after birth with 9 having a curly tail. This phenotype is indicative of neural tube defects and has been observed previously in *Fzd3* null mice (Wang et al. 2002). PCR genotyping of the dead pups revealed large deletions of exon 3 and the surrounding genomic DNA (Supplementary Figure 3). The 13 surviving pups also showed a variety of deletions indicating the CRISPR guides cut at each site and possibly removed exon 3, but each could amplify at least one exon 3 allele with specific primers. A breeding colony was established from a founder harbouring a 1261bp deletion that encompassed exon 3 and flanking intronic regions (referred to hereafter as KO mice). As expected, KO/KO mice had curly tails and died perinatally. 18.5 dpc null (KO/KO) embryos had severe fibre tracts abnormalities as previously described in the literature (Wang et al. 2002) (Supplementary Figure 4)

Compound Het Morphology

Next, we crossed KO and -573 mice to generate -573/KO compound heterozygous 11.5 embryos and assessed *Fzd3* expression. Initial qRT-PCR analysis indicated that the KO allele was not subject to nonsense mediated decay. To avoid amplification of the KO allele, we used primers binding to exon 3. Comparison of the KO/WT and KO/-573 11.5 heads revealed a significant 20% decrease in the latter (Figure 5). Despite their reduced Fzd3 expression, 573/KO pups were born at the expected frequency, did not exhibit curly tails and had a normal physical appearance. Fibre tract development in 573/KO embryos also appeared normal (Supplementary Figure 4). Thus, a *Fzd3* expression level of ~40% is sufficient to rescue the neurodevelopmental defects that occur in FZD3 null pups, however comprehensive analysis of other *Fzd* family members was not assessed in these mutants.

Discussion

Bioinformatic and chromatin capture studies have estimated that there are hundreds of thousands of enhancers scattered throughout mammalian genomes (ENCODE Project Consortium 2012). However, functional assessment of enhancer activity in the vast majority of these predicted elements is lacking. Using CRISPR, we have identified and validated a novel enhancer that is essential for normal levels of endogenous *Fzd3* expression in neural progenitor cells in vivo.

Although the *Fzd3* enhancer region is bound by SOXB1 proteins (McAninch & Thomas 2014) deletion of the single canonical SOXB1 binding site resulted in only a small reduction in *Fzd3* expression. This suggests that SOXB1 binding is essential for normal *Fzd3* expression, but only makes a modest contribution to overall expression levels in neuroprogenitor cells. As deletion of the 573bp enhancer region resulted in a much greater (30%) reduction in *Fzd3* expression, it appears that binding of additional transcription factors is a feature of this element, consistent with the idea that enhancers are made up many different transcription factor binding sites, and are rarely composed of just a single motif. Bioinformatic analyses of the region indicates the motifs known to be bound by SP1, IRF3, NR2E3 and PRDM4 are present, however these would need to be assessed by ChIP or immunoprecipitation to confirm binding.

Traditional enhancer tests utilize transgenic reporter assays, which permit visualisation of a reporter in regions where the enhancer is active (Kvon 2015; Visel et al. 2007). However, one of the disadvantages of this approach is that it does not recapitulate the enhancer activity in its in vivo context where the epigenetic landscape and interaction with neighbouring sequences may be crucial. Further, random integration of the enhancer construct can lead to spurious expression due

to the influence of flanking sequences, described as a "position effect" (Guy et al. 1997). Thus, although rarely performed, comparing the impact of enhancer deletion and transgenic reporter studies provides valuable insight into the physiological role of a particular enhancer. Our experiments revealed some interesting differences in the activity of the *Fzd3* enhancer in each of these assays. Enhancer deletion showed that the 573 region was required for complete *Fzd3* expression in the floor plate and intermediate zone of the neural tube. In contrast, the reporter assay showed consistent robust LacZ expression in the floor plate but not the intermediate zone. This suggests that the 573 region is sufficient for the floor plate expression but may interact with other enhancers/regions to drive intermediate zone expression. This data suggest that this enhancer acts as part of a 'billboard' model – whereby many smaller regions (or individual TFBS) make up the larger enhancer (Arnosti & Kulkarni 2005; Kulkarni & Arnosti 2003).

A previous study has reported that *Fzd3* null homozygous pups die shortly after birth with fibre tract abnormalities and curly tails, an indicator of defective neural tube development (Wang et al. 2002; Peeters et al. 1998). We generated an independent *Fzd3* KO line and observed the same phenotype, with heterozygous animals appearing normal as previously reported (Wang et al. 2002). It was hypothesised that the *Fzd3* phenotype may become apparent at expression levels between 0-50%. To test this we generated -573/KO mice in which the endogenous *Fzd3* expression was approximately 40% of WT levels. However, these mice appeared normal and did not display curly tails or fibre tract abnormalities. Whilst the *Fzd3* compound heterozygote did not show a phenotype, given the ease with which CRISPR deletion mutations can be generated in mice, this will likely be a useful approach to determine the developmental consequences of enhancer mutation. Although these mice produced no observable phenotype, this is not an unexpected consequence. It has recently been reported that many enhancers are

redundant, with combinatorial deletions required to induce morphological defects (Osterwalder et al. 2018) and reporter assay-positive enhancer regions often producing no phenotype or expression change (Cunningham et al. 2018).

In conclusion, we have identified a novel intronic *Fzd3* enhancer, likely activated by SOXB1 transcription factors in neuroprogenitors that is responsible for up to ~30% of *Fzd3* mRNA levels. It is highly likely that other TFs bind either within this region, or working in complex with this enhancer to regulate the expression of *Fzd3* during development. The 573 element may also be a useful tool to drive transgene expression within the floorplate, especially given that this area functions as a signalling centre for neural tube patterning (Yu et al. 2013; Ribes et al. 2010). Given the increasing rate at which bioinformatic studies are identifying putative enhancer elements, the approach we describe herein provides a valuable template for their functional validation.

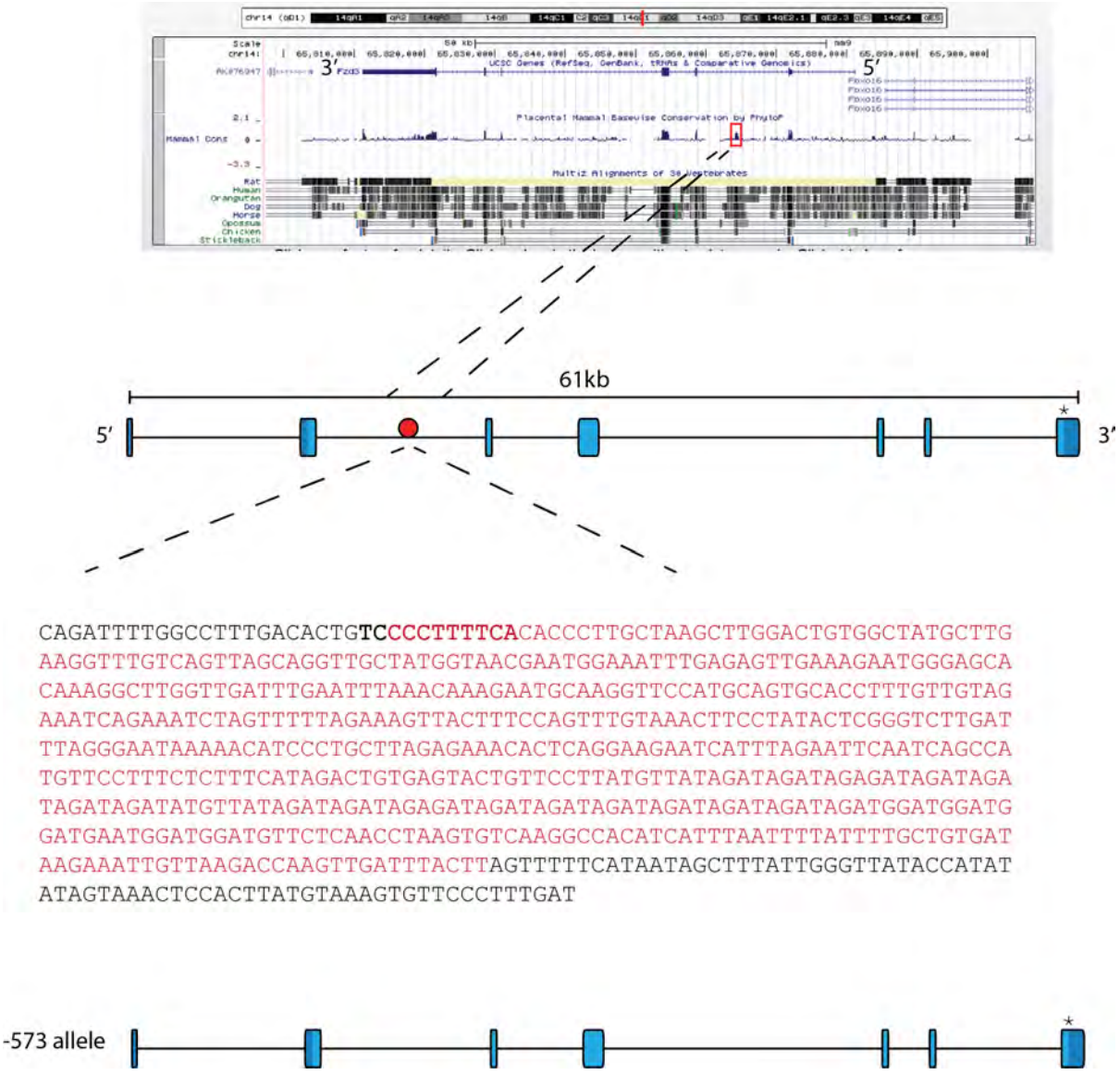
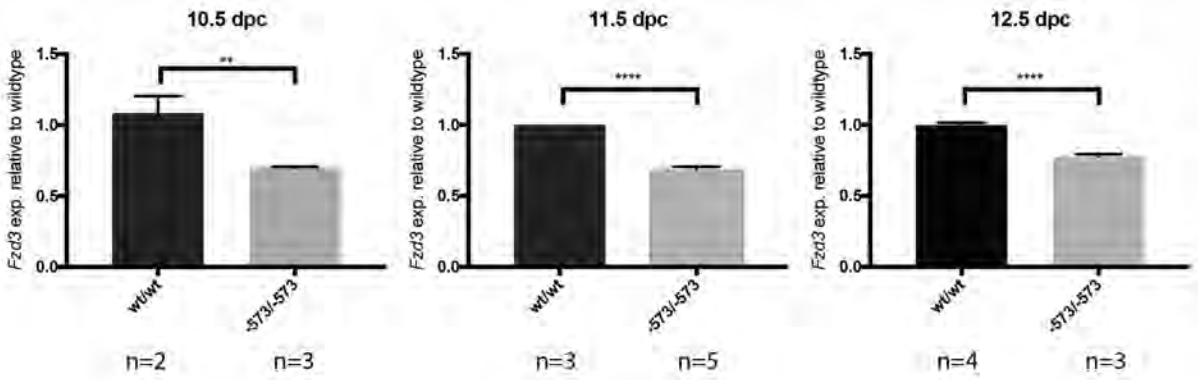


Figure 1

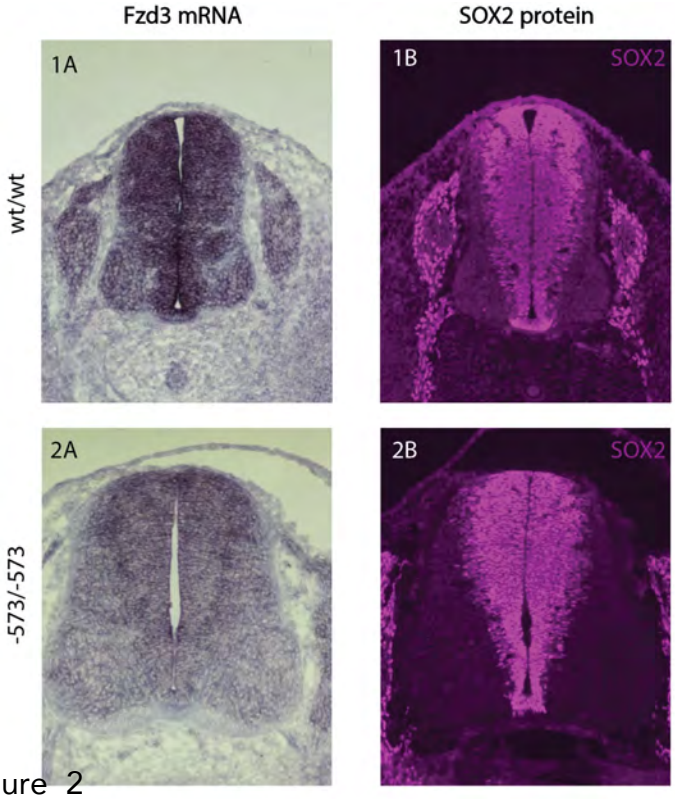
Generation of the -573 Fzd3 Enhancer Deletion Line

A SOXB1 putative enhancer was identified in a region of high conservation within intron 2 of the mouse Fzd3 gene (red box and circle). This region was mutated using CRISPR and generated a 573bp deletion (red text) incorporating the majority of the proposed TFBS (bold red text). The resultant allele does not contain the putative enhancer (deletion indicated by red cross) and is referred to as the -573 line.

A



B



C.C

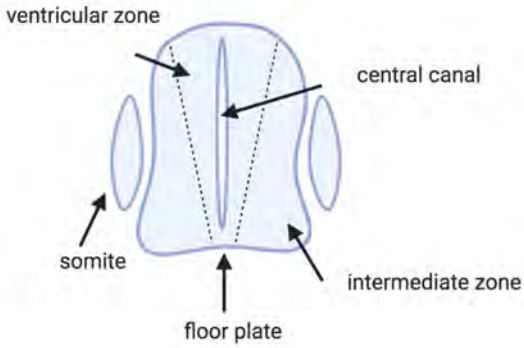


Figure 2

Analysis of mRNA expression in embryonic -573 lines

A. qRT-PCR was performed on embryonic heads to assess the levels of Fzd3 expression of -573 embryos compared to wildtype. Expression of Fzd3 is reduced at each time point. Students t-test performed, 10.5 dpc samples show p-value of 0.0088, 11.5dpc samples show p-value of <0.0001 and 12.5 dpc sample show p-value of <0.0001. Error bars indicate the standard error from the mean B. Expression patterns of Fzd3 mRNA were assessed using in situ hybridization on 11.5dpc trunk sections to visualize the neural tube (1A and 2A). These were compared to SOX2 protein(1B and 2B) The Fzd3 can be seen to be reduced throughout the ventricular zones and the floor plate of the neural tube, consistent with regions of SOX2 staining. C. A schematic of the neural tube outlining the regions present within the sections.

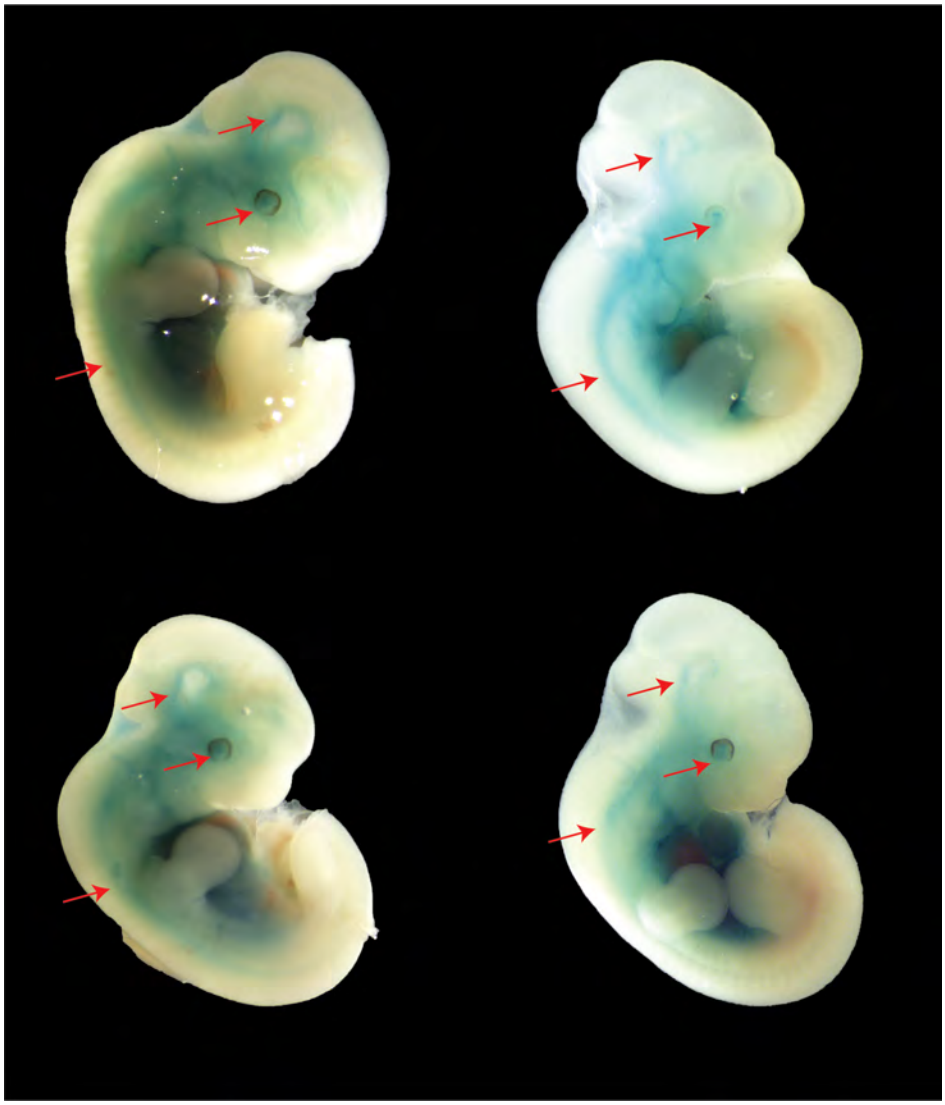


Figure 3

LacZ staining of -573 enhancer transgenic mice

Representative images of the 7 LacZ positively stained embryos, out of 7 embryos analysed. Staining can be visualized in the midbrain, eye and neural tube regions depicted by the red arrows.

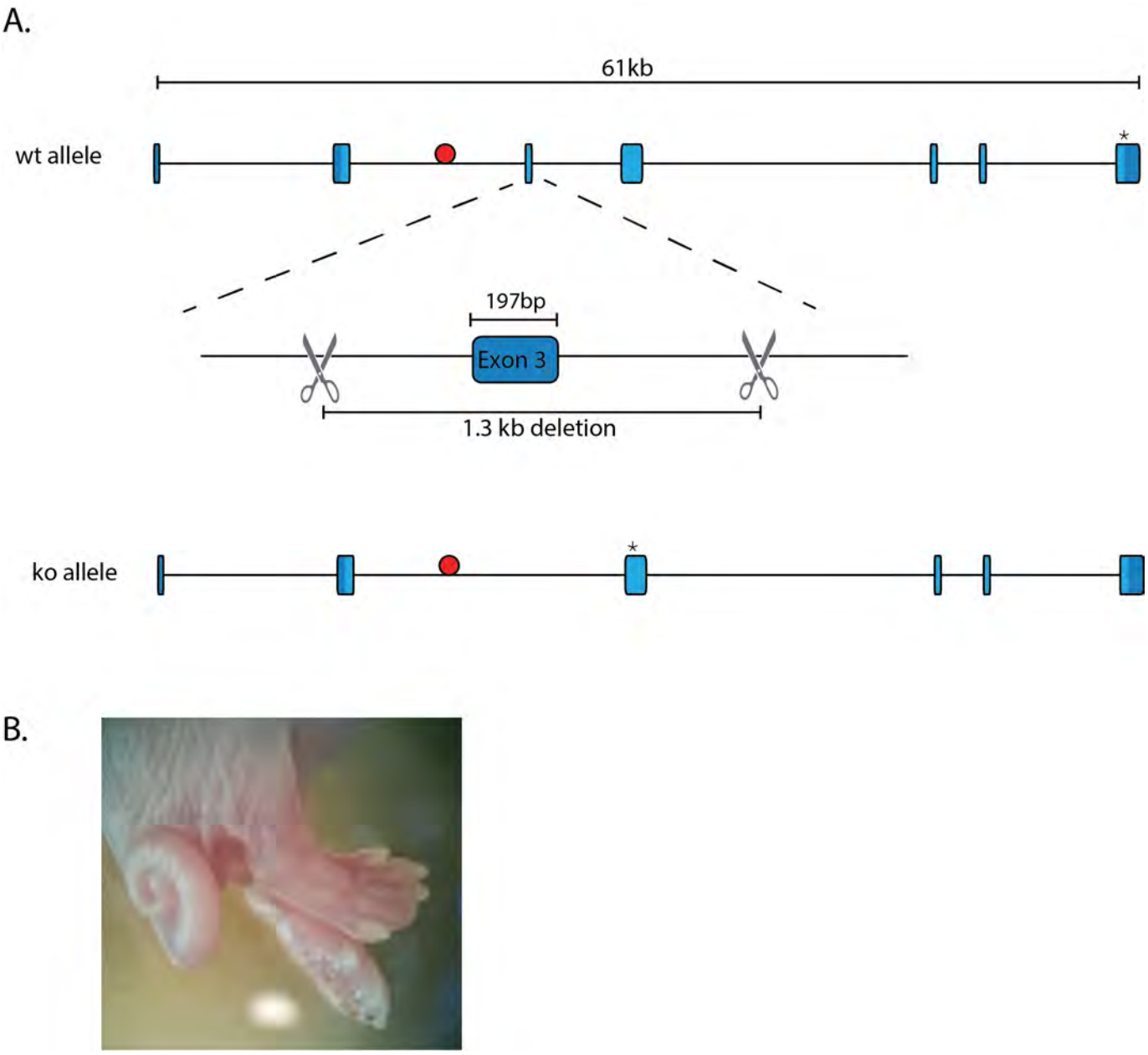


Figure 4

Generation of *Fzd3* knockout mouse line

A. To generate *Fzd3* knockout mice, exon 3 was targeted as it would produce a frameshift mutation if deleted. CRISPR guides were designed to flank exon 3 and this generated a large deletion encompassing the region. Dark blue boxes indicates 5' and 3' UTR, light blue represent coding exons, and the red circle indicates the *Fzd3* putative enhancer. The asterisk represents the stop codon, within the wildtype allele this is in the last coding exon, while in the new knockout allele this occurs within exon 3 due to a frameshift mutation. B. Representative image of the 'curly tail' phenotype associated with *Fzd3* knockout mutations.

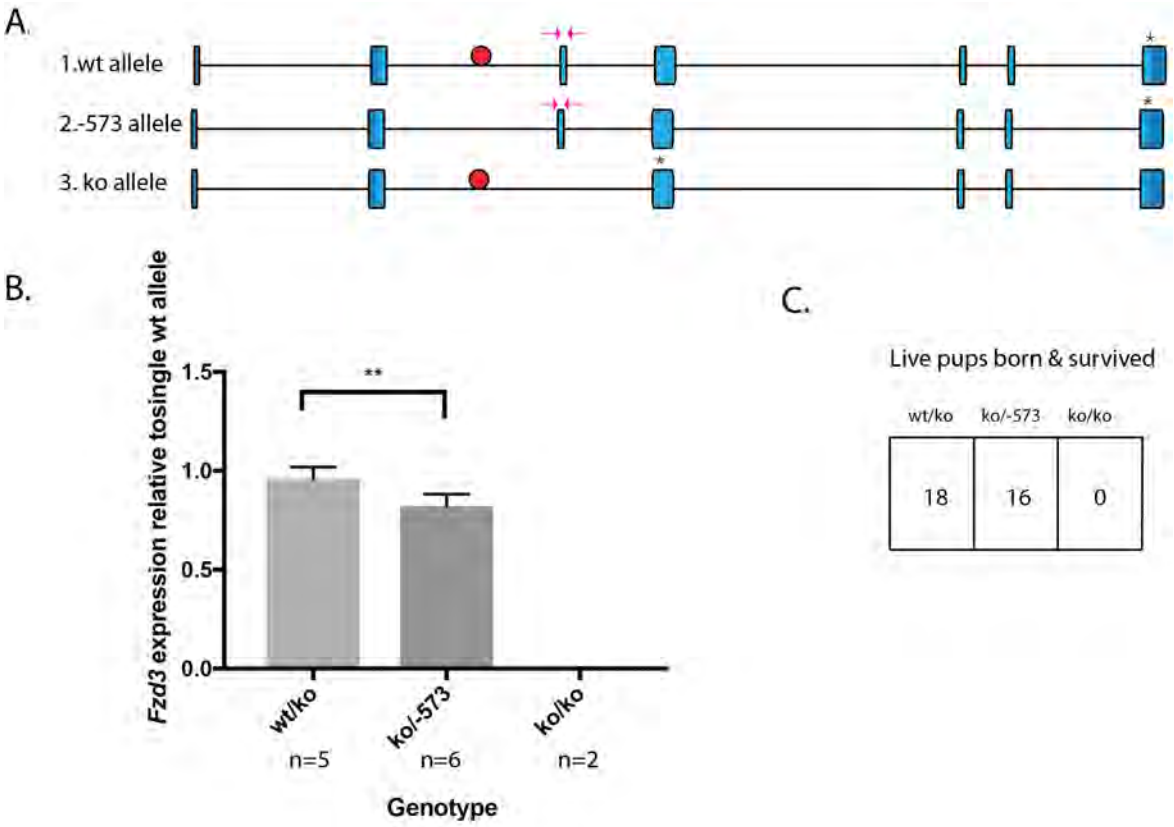


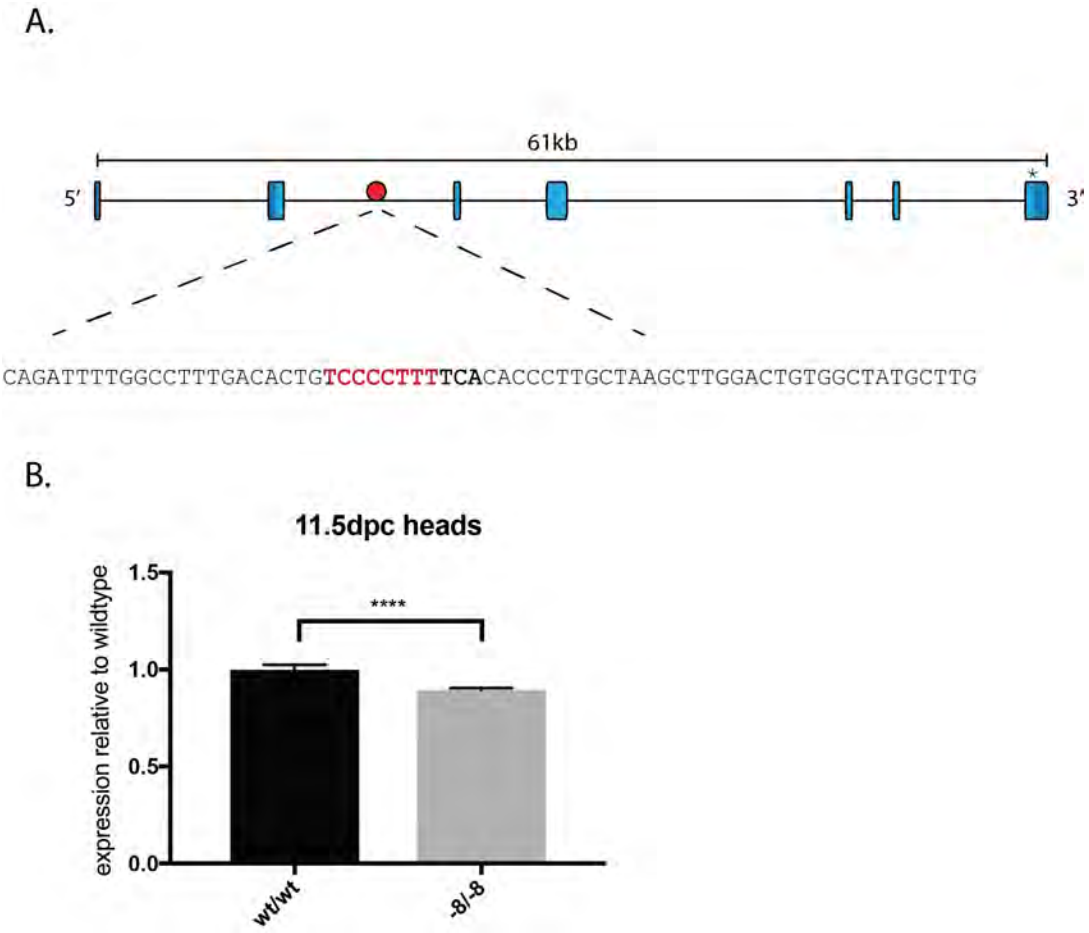
Figure 5

Analysis of compound heterozygote (-573/ko) embryos

A. Diagrams showing each of the alleles used for the crosses. Dark blue boxes indicate the 5' and 3' UTR, light blue represent coding exons, and the red circle indicates the *Fzd3* putative enhancer. The asterisk represents the stop codon 1. The wildtype allele contains both the -573 enhancer (red circle) as well as exon 3 producing wildtype levels of *Fzd3.* 2. The -573 allele does not contain the enhancer, but all exons are present. 3. The ko allele contains the enhancer but no exon 3 of the *Fzd3* gene. The primers used for mRNA analysis amplify exon 3 to avoid amplifying product from the wt allele which does not undergo nonsense mediated decay.

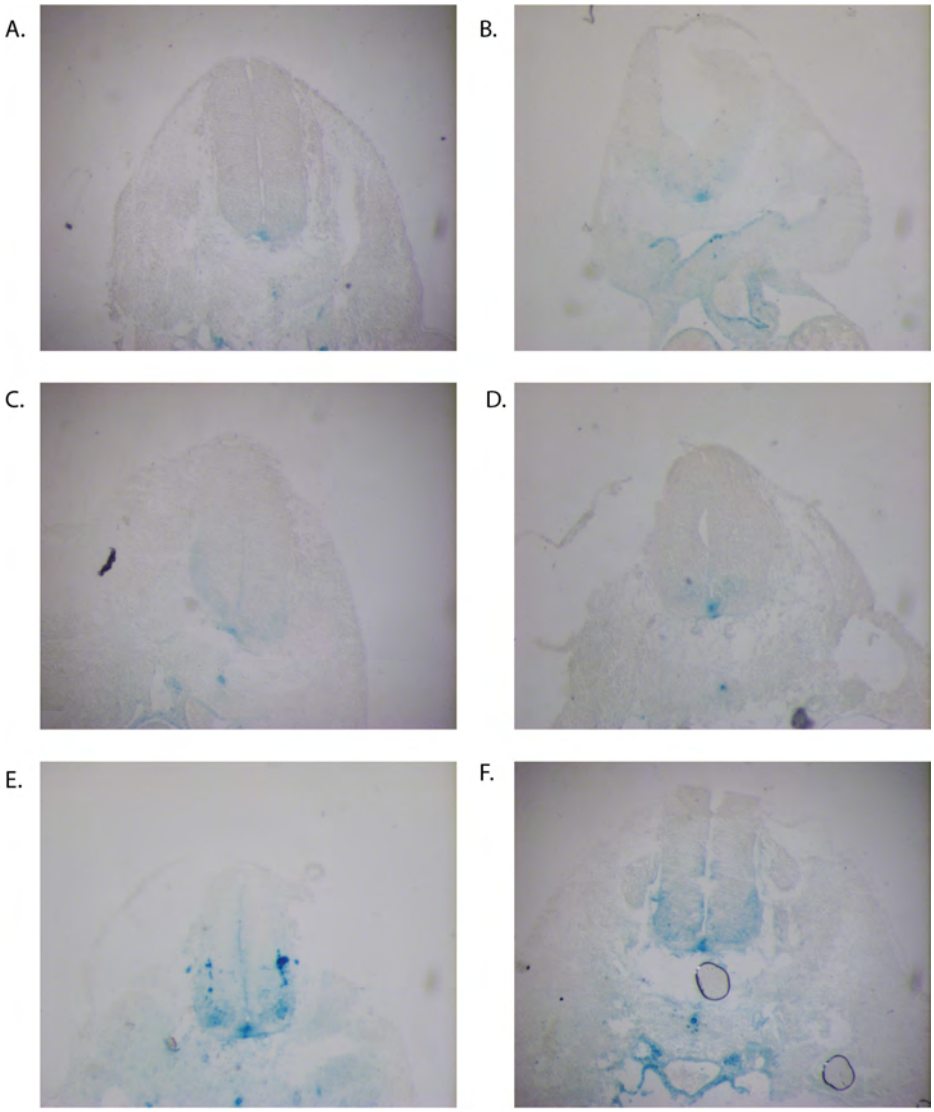
B. The expression levels of *Fzd3* were compared between a wt/ko and ko/-573 allele. This will only measure expression of a single allele that is not ko as the primers do not amplify this. The compound heterozygote (-573/ko) show a 20% reduction in expression from the -573 allele consistent with previous results. Unpaired t-test between wt/ko and ko/-573 samples generates a p value of 0.0038. Error bars represent the standard error of the mean.

C. Table showing the numbers of live born pups that are able to survive from intercrosses. *Fzd3* knockout pups (ko/ko) die shortly after birth, but compound heterozygote (-573/ko) and *Fzd3* heterozygous (wt/ko) pups appear healthy and phenotypically normal.



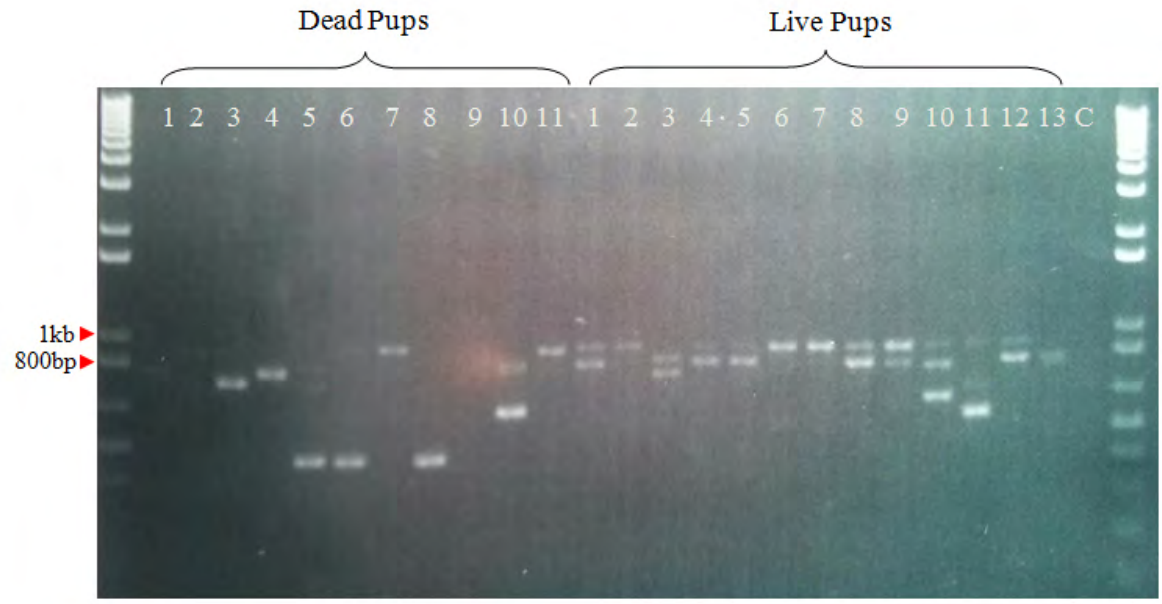
Supplementary Figure 1

Generation and analysis of the Fzd3 -8 line. A shows the *Fzd3* putative enhancer region. The red text has been deleted, while the bold black text is the remaining portion of the SOXB1 consensus motif. B shows qRT-PCR analysis of 11.5dpc heads measuring *Fzd3* mRNA levels.



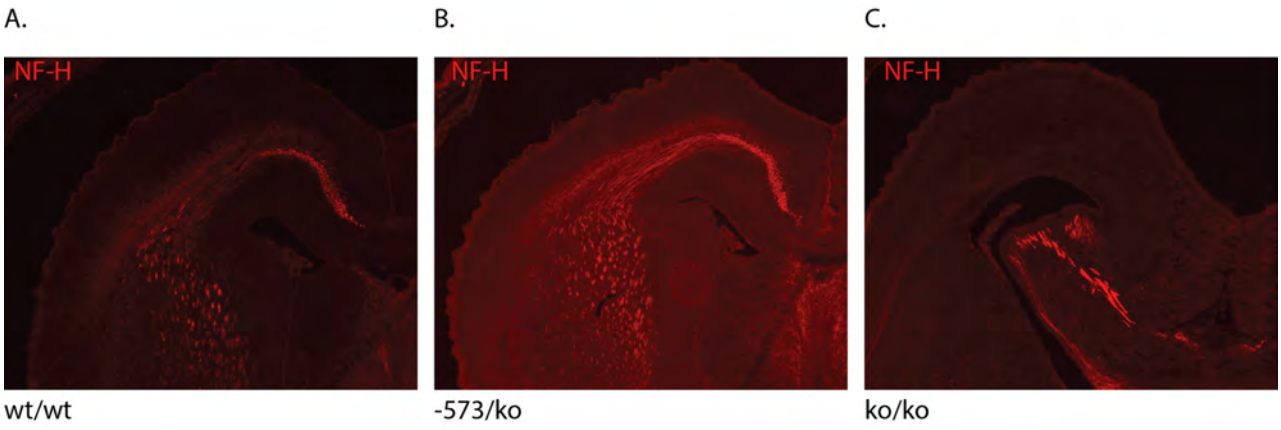
Supplementary Figure 2

6 transgenic *Fzd3* enhancer embryo trunks were sectioned after whole mount staining to determine the location of the neural tube staining (E-F) The seventh embryo was not sectioned due to damage. Consistent expression can be seen within the floor plate of each of the embryos.



Supplementary figure 3

Genotyping Gel of F1 generation of *Fzd3* KO pups. The band at 900bp indicates the presence of a wildtype allele. The pups in lanes 1-11 were found dead soon after birth, and genotyping indicates most of them do not have a wildtype allele and are homozygous for a embryonic lethal mutation. The pups that survived 'live pups' almost all show a band at approximately 900bp indicating the presence of a wildtype allele.



Supplementary Figure 4

Neurofilament staining of 18.5 dpc *Fzd3* ko and enhancer deletion mutations shows that the loss of fibre tract phenotype within mice lacking *Fzd3* does not occur in compound heterozygotes with approximately 40% of wildtype expression.

References

Arnosti, D.N. & Kulkarni, M.M., 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *Journal of Cellular Biochemistry*, 94(5), pp.890–898.

Bergsland, M. et al., 2011. Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev*, 25(23), pp.2453–2464.

Cong, L. et al., 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, 339(6121), pp.819–823.

Cunningham, T.J. et al., 2018. Genomic Knockout of Two Presumed Forelimb Tbx5 Enhancers Reveals They Are Nonessential for Limb Development. *Cell Reports*, 23(11), pp.3146–3151.

ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), pp.57–74.

Guy, L.G., Kothary, R. & Wall, L., 1997. Position effects in mice carrying a lacZ transgene in cis with the beta-globin LCR can be explained by a graded model. *Nucleic Acids Research*, 25(21), pp.4400–4407.

Kulkarni, M.M. & Arnosti, D.N., 2003. Information display by transcriptional enhancers. *Development*, 130(26), pp.6569–6575.

Kvon, E.Z., 2015. Using transgenic reporter assays to functionally characterize enhancers in animals. *Genomics*, 106(3), pp.185–192.

Mali, P., Esvelt, K.M. & Church, G.M., 2013. Cas9 as a versatile tool for engineering biology. *Nature Methods*, 10(10), pp.957–963.

McAninch, D. & Thomas, P., 2014. Identification of highly conserved putative developmental enhancers bound by SOX3 in neural progenitors using ChIP-seq. *PloS ONE*.

Mundade, R. et al., 2014. Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell Cycle*, 13(18), pp.2847–2852.

Osterwalder, M. et al., 2018. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature*, 554(7691), pp.239–243.

Peeters, M.C. et al., 1998. Role of differential cell proliferation in the tail bud in aberrant mouse neurulation. *Developmental dynamics : an official publication of the American Association of Anatomists*, 211(4), pp.382–389.

Pevny, L. & Placzek, M., 2005. SOX genes and neural progenitor identity. *Current Opinion in Neurobiology*, 15(1), pp.7–13.

Ribes, V. et al., 2010. Distinct Sonic Hedgehog signaling dynamics specify floor plate and ventral neuronal progenitors in the vertebrate neural tube. *Genes Dev*, 24(11), pp.1186–1200.

Sala, C.F. et al., 2000. Identification, gene structure, and expression of human frizzled-3 (FZD3). *Biochemical and Biophysical Research Communications*, 273(1), pp.27–34.

Sarkar, A. & Hochedlinger, K., 2013. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell Stem Cell*, 12(1), pp.15–30.

Visel, A. et al., 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, 457(7231), pp.854–858.

Visel, A. et al., 2007. VISTA Enhancer Browser - a database of tissue-specific human enhancers. *Nucleic Acids Research*, 35, pp.D88–D92.

Wang, W. et al., 2016. Feedback regulation of apical progenitor fate by immature neurons through Wnt7-Celsr3-Fzd3 signalling. *Nature Communications*, 7(1), p.10936.

Wang, Y. et al., 2006. Axonal growth and guidance defects in Frizzled3 knock-out mice: a comparison of diffusion tensor magnetic resonance imaging, neurofilament staining, and genetically directed cell labeling. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26(2), pp.355–364.

Wang, Y. et al., 2002. Frizzled-3 is required for the development of major fiber tracts in the rostral CNS. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 22(19), pp.8563–8573.

Yu, K., McGlynn, S. & Matise, M.P., 2013. Floor plate-derived sonic hedgehog regulates glial and ependymal cell fates in the developing spinal cord. *Development (Cambridge, England)*, 140(7), pp.1594–1604.

Chapter 4

Identification of SOX3 bound putative enhancers within the postnatal mouse testes using ChIP-Seq

Statement

The SOX3 ChIP-Seq experiment in this chapter was performed by Dale McAninch.

Introduction

SOX3 is the only member of the SOXB1 family of transcription factors that is expressed within the postnatal testes. Located within only the germ cells (spermatogonial stem cells and early spermatocytes) it is thought to play a role in spermatogenesis, although the mechanisms of this have not yet been fully elucidated (J. Weiss et al. 2003).

Due to the similarities of SOX3 and the sex determining gene *Sry*, widely thought to have evolved from *Sox3* (Graves 1998), previous research sought to determine if *Sox3* also plays a role in early testis or ovary development (Weiss et al. 2003). It has since been shown that SOX3 is not required for sex determination, as SOX3 null males appear phenotypically normal during gonad differentiation and throughout embryogenesis (Weiss et al. 2003; Raverot et al. 2005). Postnatally however, SOX3 null mice have impaired spermatogenesis and lack mature sperm, affecting fertility (Raverot et al. 2005).

After the bipotential gonad has its male fate determined by *Sry* expression, *Sox9* is activated, leading to the downregulation of female specific genes and upregulation of genes that drive testes development (Sekido & Lovell-Badge 2008). The primordial germ cells become encased within Sertoli cell precursors and form gonocytes which, after rapid proliferation, eventually halt at G0/G1 phase until birth, when they begin to again undergo mitosis/meiosis after 2/3 days (Griswold 2016). It is at this stage when the *Sox3* null testes phenotype becomes apparent (J. Weiss et al. 2003). After birth, a subset of the gonocytes within the testicular cords will transform into the spermatogonial stem cell population, whilst the remaining gonocytes will undergo apoptosis. This transformation is essential to allow a constant supply of Spermatogonial Stem Cells (SSCs) throughout the fertile period,

as these provide cells to enter the spermatogenic cell cycle eventually undergoing spermatogenesis and becoming mature sperm. During this process there is a tremendous amount of chromatin remodeling occurring, whereby the histones which bind the DNA are replaced by testes specific histones, and subsequently protamines, another form of chromatin condensing protein (Rathke et al. 2014). Various knockout studies performed in mice have shown that altered regulation of the histone and protamine genes causes defects in spermatogenesis leading to infertility; indicating that this is an essential process that must be undertaken for viable sperm production (Ueda et al. 2017).

Histones consist of the 4 core histone (H2A, H2B, H3, H4) proteins along with a linker histone protein (H1). The genes for each of these proteins are found in clusters throughout the genome, with many variants of each found. In mice, there are 3 main clusters found on chromosome 13 (51 Hist1 genes), chromosome 3 (6 Hist2 genes) and chromosome 11 (3 Hist3 genes) (Marzluff et al. 2002). Other important histone genes, such as H2afx are not found in clusters but located throughout the genome as isolated genes. The regulation of these genes is largely cell cycle dependent due to the need of large amounts of histone expression during replication phases (Mei 2016). Histones and testes specific histone variants are an essential component of effective gene regulation during spermatogenesis, without which lead to fertility and sperm defects (Hoghoughi et al. 2017).

In this chapter we have analysed a SOX3 binding in mouse postnatal testes using ChIP-Seq. We identify almost 800 putative SOXB1 binding regions, including those close to *Neurog3, H3t* and *Th2b*, thought to be important genes regulating sperm development. Further, we show SOX3 binding near a large number of histone genes, suggesting it may have a regulatory role in chromatin and nuclear compaction in both the testes and CNS.

Materials and Methods

Animal Work

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.

Chromatin Immunoprecipitation

Testes from postnatal day 7 wildtype 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). DNA was recovered by reversing crosslinks, and purified by PCR clean-up 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 on the Illumina HiSeq producing 50 bp single end reads. A control sample (without SOX3 antibody) was run as input for background control.

Peak Calling

As published in McAninch & Thomas 2014, Bowtie (Langmead 2010) was used to align reads to the mouse genome (mm9). Peaks were called for each biological replicate using MACS (Feng 2011), 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.

Gene Ontology

Gene ontology was performed using the GREAT webtool {McLean:2010iq) using 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.

de novo Motif Analysis

The RSAT Peak Motifs program (Thomas-Chollier et al. 2012) was used for de novo motif analysis for each group of peak sequences analysed. Default parameters were used and the JASPAR core non-redundant vertebrates was used to compare to known motifs.

FIMO (Find Individual Motif Occurrences), part of MEME-Suite was used to confirm peak regions containing the motifs identified previously (Grant et al. 2011; Bailey et al. 2015).

Conservation Analysis

PhastCons scores were generated for each peak sequence (ref). These were then ranked, and sorted into high, moderate and low conservation based on their mean 0 score (Siepel et al. 2005).

Published ChIP-Seq data

SOX3 ChIP-Seq data from NPCs was obtained from McAninch (2014b). Microarray data of SOX3 knockout testes obtained from Adikusuma (2017)

Results

Genomic location of SOX3 bound regions

To identify the regions of the genome bound by SOX3 within the postnatal mouse testes, the three independent ChIP-Seq experiments were performed using a previously validated SOX3 specific antibody (McAninch & Thomas 2014). Postnatal day 7 testes were used, as SOX3 is known to be expressed at this time, however the effects of germ cell loss are not yet evident (Raverot et al. 2005). Independent replicate datasets were overlaid and overlapping regions identified from the peak sequences. Across the three experiments, 2209, 1991 and 2017 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 downstream analysis.

To determine the genomic context of SOX3-bound regions, HOMER ChIP-Seq analysis was performed (Figure 1). This revealed that 50% of the sites were intergenic (388/778), while 31% are intronic (243/778), followed by 11% within the promoter region (86/778), 2% transcriptional start and stop sites (18/778) and 2% within the UTR (14/778)

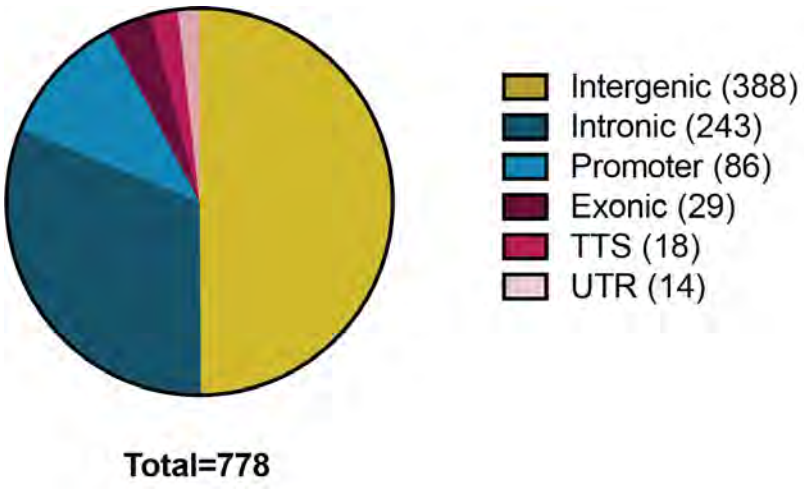


Figure 1

The genomic context of all peak regions from SOX3 peaks identified within at least 2 or more ChIP-Seq experiments. This data shows the majority of peaks were present within intergenic and intronic regions, followed by promoter, exons, transcriptional start and stop sites (TSS) and untranslated regions (UTR).

GO Terms associated with peak regions

To determine if there is any known functional associations of these peak regions, Gene Ontology (GO) Term analysis was performed using GREAT (Genomic Regions Enrichment of Annotations Tool). This tool interprets the functional significance of given non-coding genomic regions through analysis of the annotations of nearby and neighboring genes (McLean et al. 2010). As seen in Figure 2, the top 5 terms produced were "negative regulation of megakaryocyte differentiation", "nucleosome assembly", "nucleosome organization", "chromatin assembly" and "chromatin assembly or disassembly". From this it can be seen that there is a highly significant enrichment for functional association with nucleosomes and chromatin environment. Looking further into what regions are responsible for these terms, it is interesting to see that the majority of the peaks are located close to various histone, and histone regulation genes such as *Brd2*, *Setd2* and *H2afx*. The full list of region and gene associations for each of the GO Terms produced is shown in Appendix 2.

To determine if the high level of 'chromatin' GO terms were masking the presence of less represented functional associations the data was reanalyzed. All histone associated peak regions were removed from the dataset, and this was assessed via GREAT. Interestingly, no significant GO terms were found from this data, suggesting that the histone and chromatin related gene regulation is the key role of SOX3 within the testes.

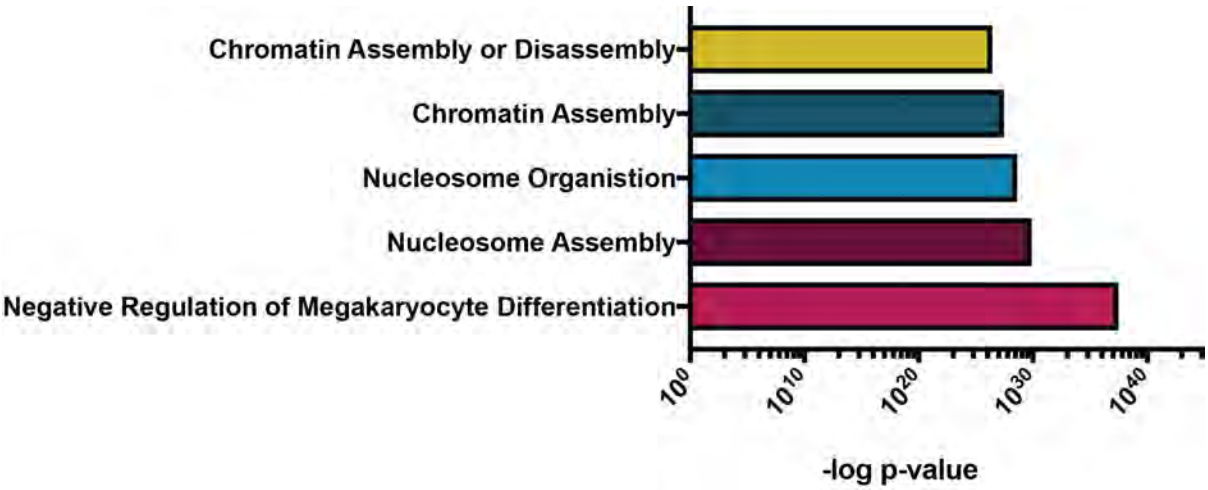


Figure 2

GREAT Analysis identified the most highly represented significant gene ontology terms associated with the 778 peak regions. The most highly significant term is 'negative regulation of megakaryocyte differentiation'

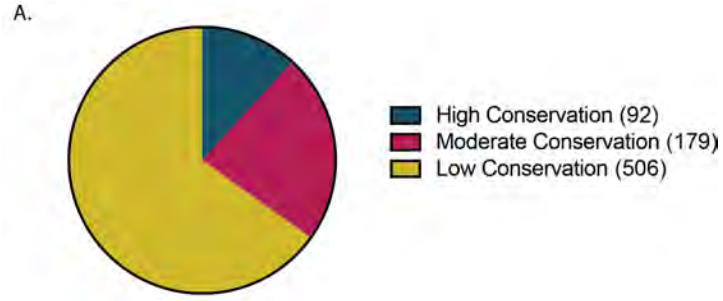
Conservation of peak regions

A feature of many known enhancers is their high level of evolutionary sequence conservation. We assessed the conservation of SOX3 bound regions using phastCons, which generates a conservation score for input sequences based on the conservation between 30 placental mammals (< 0.1 low conservation, 0.1- 0.5 moderate conservation, and >0.5 high conservation) (Figure 3A). The majority of the sequences were lowly conserved (65%), with 23% moderately conserved, whilst 12% were highly conserved.

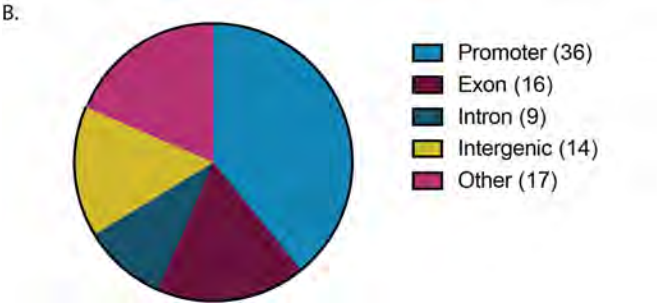
To determine the genomic location of these highly conserved regions, we performed HOMER analysis on this subset. This revealed that the majority of these lie within the promoter region (defined as 1kb upstream of start site) or overlap an exon. While these still may be functional enhancers, it is likely that these regions are highly conserved due to the fact they are closer to coding regions; as these are generally more conserved than an intron. The genomic regions and their nearest genes are outlined in Appendix 2.

Interestingly, among this group of highly conserved peaks, a notable region was found within an intergenic regions (Fig 3B) whose neighboring gene is known to be developmentally important in the progenitor cells of the testes, *Neurogenin3* (Kaucher et al. 2012). This peak region appears to be part of a much larger region of conservation suggesting the peak we have identified could be part of a larger overall enhancer region (Figure 3C).

A.



B.



C.

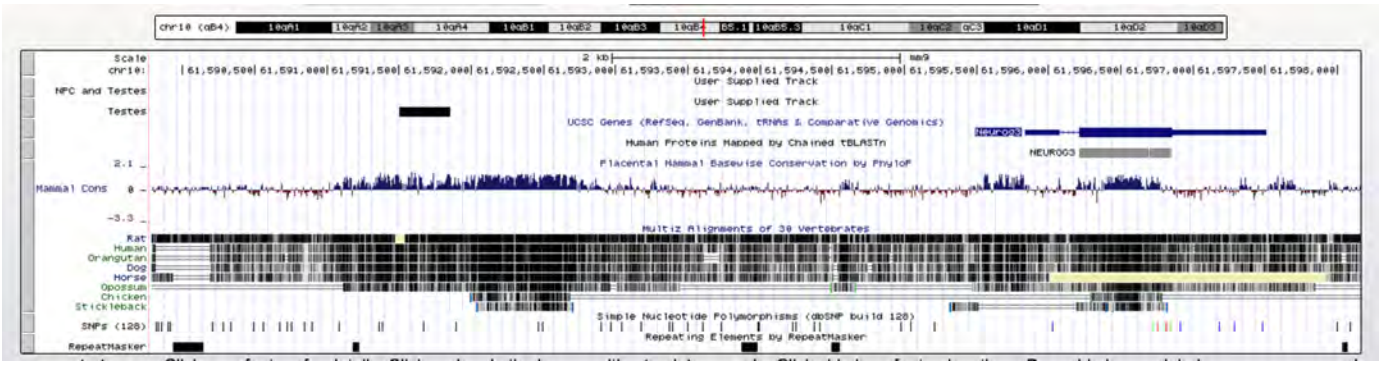


Figure 3

Conservation analysis of the peak regions as assessed by PhastCons scores. A shows the conservation distribution of all 778 testes peak regions, most regions show low conservation. B shows the genomic location of the 12% of peaks from 3A that are highly conserved. C shows the SOX3 ChIP-Seq binding peak in the 'testes' track is present within a region of high conservation (blue, mammal cons track) and is located adjacent to the Neurog3 gene in mouse. Image is taken from the UCSC mm9 genome browser.

Identification of SOX Motifs

The 20 different SOX proteins have been shown to bind to a consensus SOX core motif, ($G_1A_2C_3A_4A_5A_6G_7$) (Hou et al. 2017). We wanted to determine if the sequences within our peak regions contained this core motif or variations of it. To identify putative SOX3 binding motifs, we performed *de novo* motif analysis using RSAT peak motifs.

We identified 3 similar "SOX" motifs that were enriched within the peak sequences at similar frequencies (Figure 4A-C). Motif 1 and 2 are highly similar, but differ at the core position 6, corresponding to A9 in Motif 1, and T9 in Motif 2. The third motif appears to be similar to the 'core' sequence, however does not show a preference for the first G at position 1 (core).

As these motifs vary, each of the groups were assessed via GREAT to determine if the different binding motifs may be indicative of function. However, for each of the three groups, as well as when combined, did not produce any significant GO Terms. The binding locations and nearest genes of the peak sequences containing a SOX motif can be found in Appendix 2.

Importantly, the distribution of each motif is highest at the center of the peak sequence, suggesting these are *bona fide* binding sequences. Each of these motifs is present within different peaks (motif 1-124, motif 2-94, motif 3-93), although due to the nature of position-specific frequency matrix, some peak sequences contain more than one of the SOX motifs. From this, there are 204 individual peaks that contain a SOX like motif (Figure 4D). For downstream analysis of putative binding partners or co-regulators, each of the peak sequences that contained at least one of the three SOX-like motifs was used.

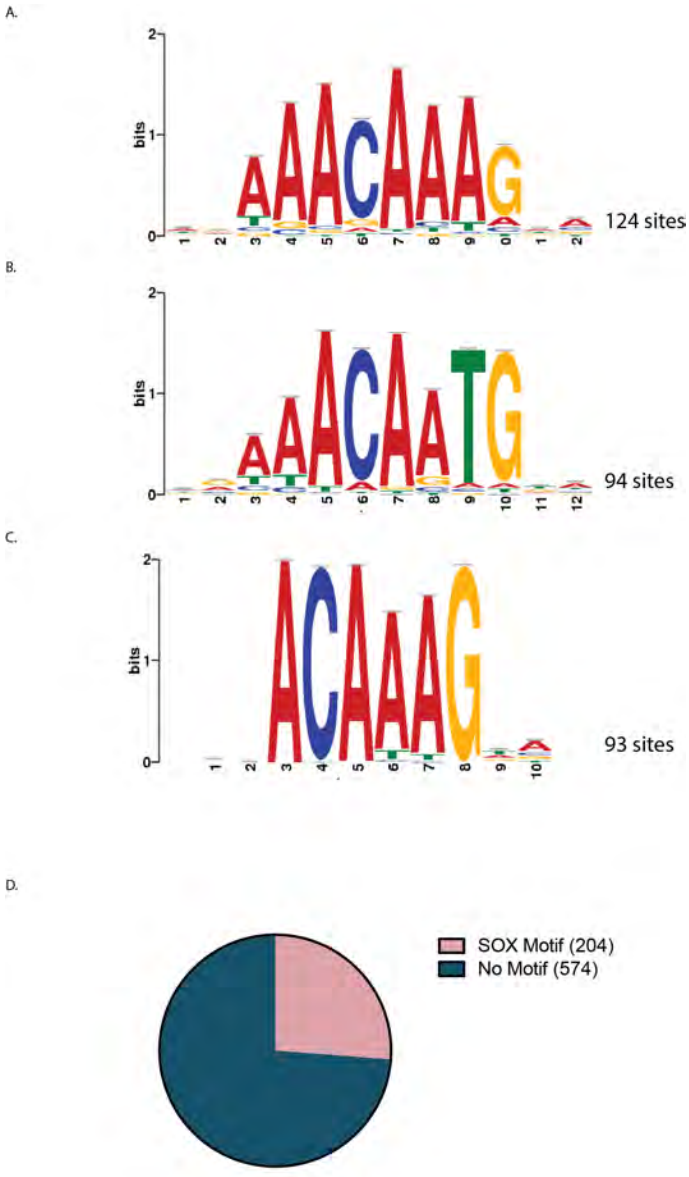


Figure 4

The three SOX-like motifs identified via RSAT Peak motifs are highly similar although differ in core position 6, corresponding to position 9 in A and B, and 7 in C. The proportion of peak sequences containing one of the shown SOX motifs is shown in D, with 26% of peaks showing a SOX motif, and 74% not.

Identification of putative co-regulator proteins

The motif identification program, RSAT Peak Motifs identified a further 8 possible motifs present within the set of sequences. These include motifs shown to be similar to Ahr/Arnt, SP, and HES families. SOX proteins have been shown to co-activate expression of target genes with varying partner factors, dependent upon the tissue context. However none of those identified in this search have been shown previously to be SOX-protein partners. Interestingly, the most common partner factor of SOXB1 proteins, POU family members, were not identified in any *de novo* analyses, or through manual searches using the POU motif.

One motif that warranted further investigation was the SP1/SP2 like motif (CGCCTCC/CCGCCGC) as seen in Figure 5A (SP Motif) and 5B (identified motif in dataset). SP transcription factors are expressed throughout many tissues and cell types, including spermatogonia, and been shown to induce expression of histone genes (Wilkerson 2002). We show 27 regions in which both SP and SOX3 motif were identified, and whilst likely not a partner protein due to considerable variation in motif location, it may be a putative co-regulator present at the same genomic regions.

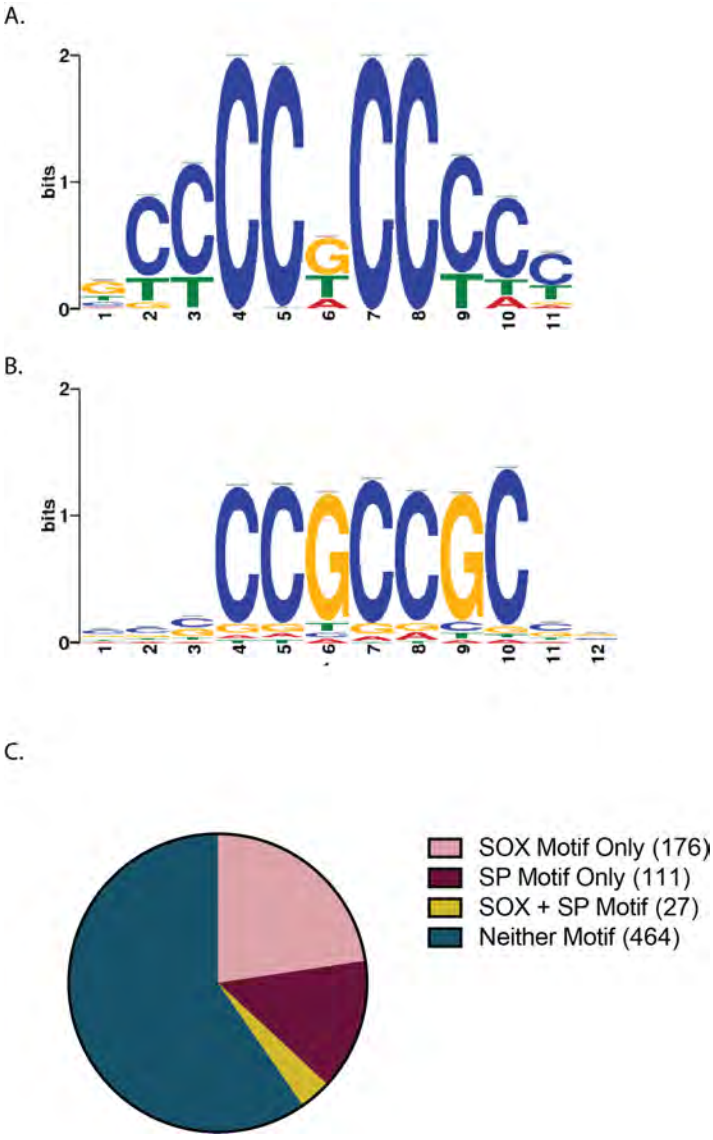


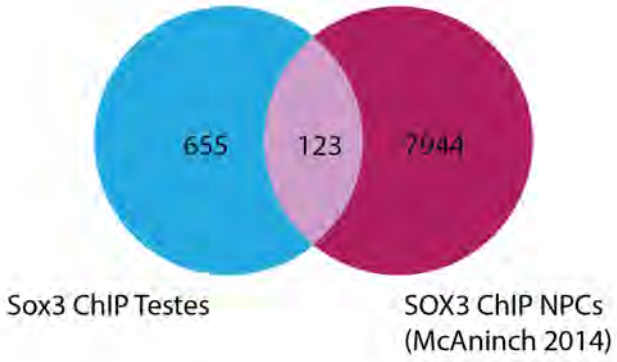
Figure 5

De novo motif analysis of the peak regions identified a SP like motif. A shows the Transfac known SP1 motif while B shows the motif identified within the peak sequences. The proportion of peaks containing both SOX and SP motifs is shown in C, with the majority of peaks containing neither motif, while 3.5% contain both.

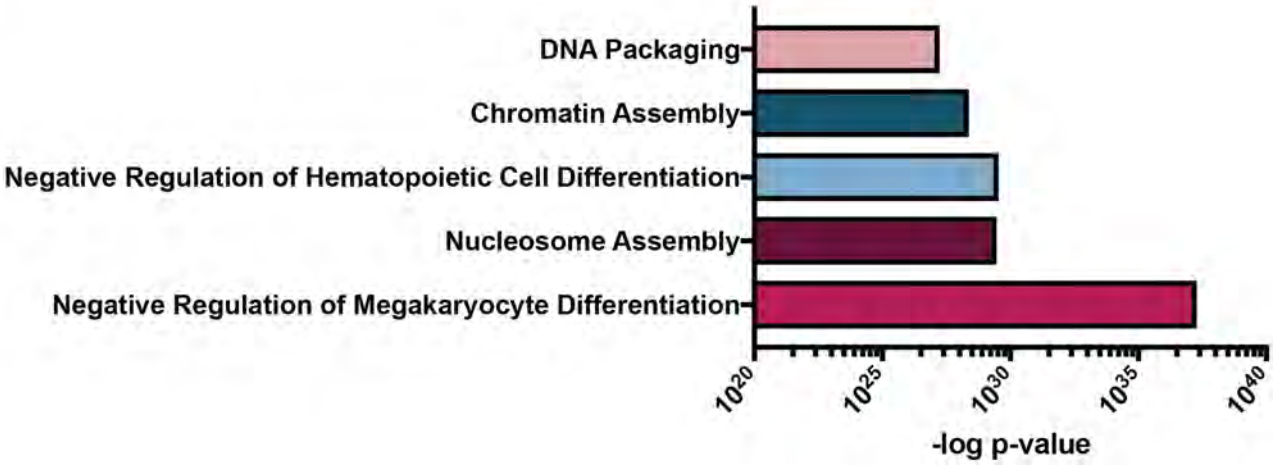
SOX3 bound regions present in both NPCs and Testes

Previous ChIP-Seq analysis of mouse neural progenitor cells (NPCs) using the same SOX antibody (McAninch & Thomas 2014b) identified 8064 regions bound by SOX3. Comparison of the NPC and testes sequences identified a subset of 123 regions bound by SOX3 in both tissues (Figure 6A). This suggests some overlap of SOX3 function in these distinct cellular contexts. GO Terms were generated for the common subset of SOX3-bound sequences to assess what functions may be consistent. Significant enrichment for terms involving DNA packaging, nucleosome and chromatin assembly, as well as the negative regulation of both megakaryocyte and hematopoietic cell differentiation were identified (Figure 6B). These terms are similar to those produced when all 778 testes peaks were assessed. A large proportion of the peaks located within clusters near histone genes were also shown to be overlapping in the NPC data (Figure 6 C-D). This suggests a potential common function of SOX3 at these regions, and may be what accounts for the similarity of GO terms between the two sets of peaks.

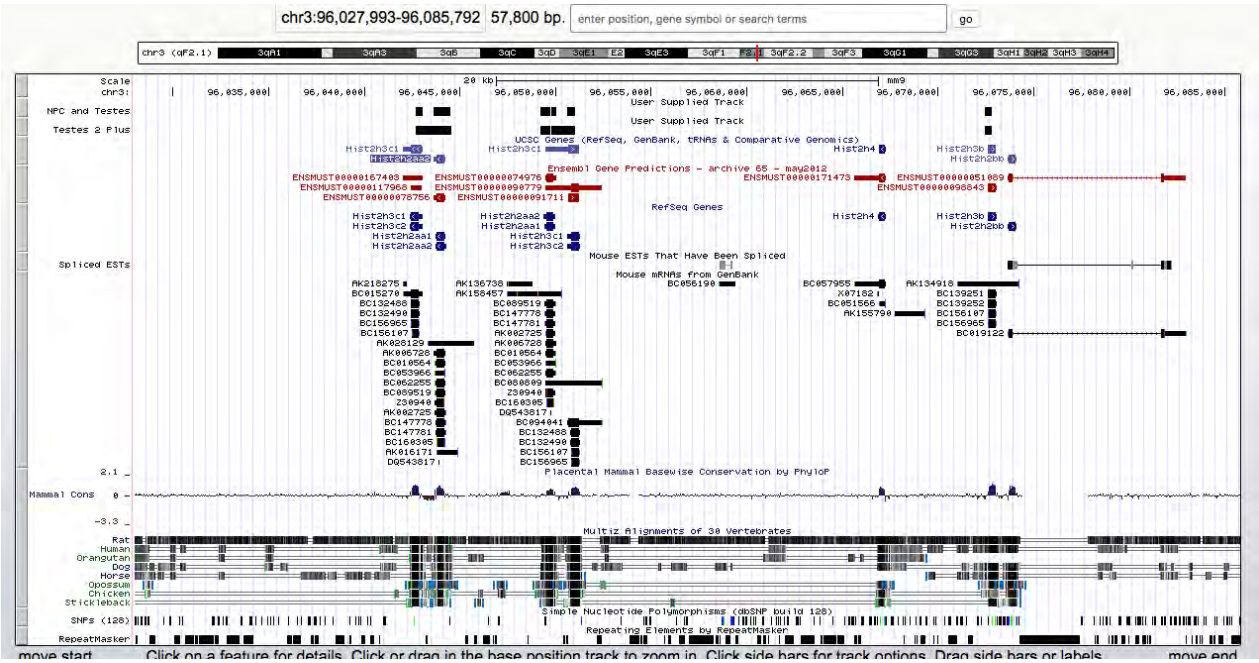
A.


655
123
7944
Sox3 ChIP Testes
SOX3 ChIP NPCs
(McAninch 2014)

B.


DNA Packaging
Chromatin Assembly
Negative Regulation of Hematopoietic Cell Differentiation
Nucleosome Assembly
Negative Regulation of Megakaryocyte Differentiation
10^20
10^25
10^30
10^35
10^40
-log p-value

C.


chr3:96,027,993-96,085,792
57,800 bp.
enter position, gene symbol or search terms
go
User Supplied Track
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
Ensembl Gene Predictions - archive 65 - may2012
RefSeq Genes
Mouse ESTs That Have Been Spliced
Mouse mRNAs from GenBank
Placental Mammal Basewise Conservation by PhyloP
Multiz Alignments of 30 Vertebrates
Simple Nucleotide Polymorphisms (dbSNP build 128)
Repeating Elements by RepeatMasker
Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels

D.

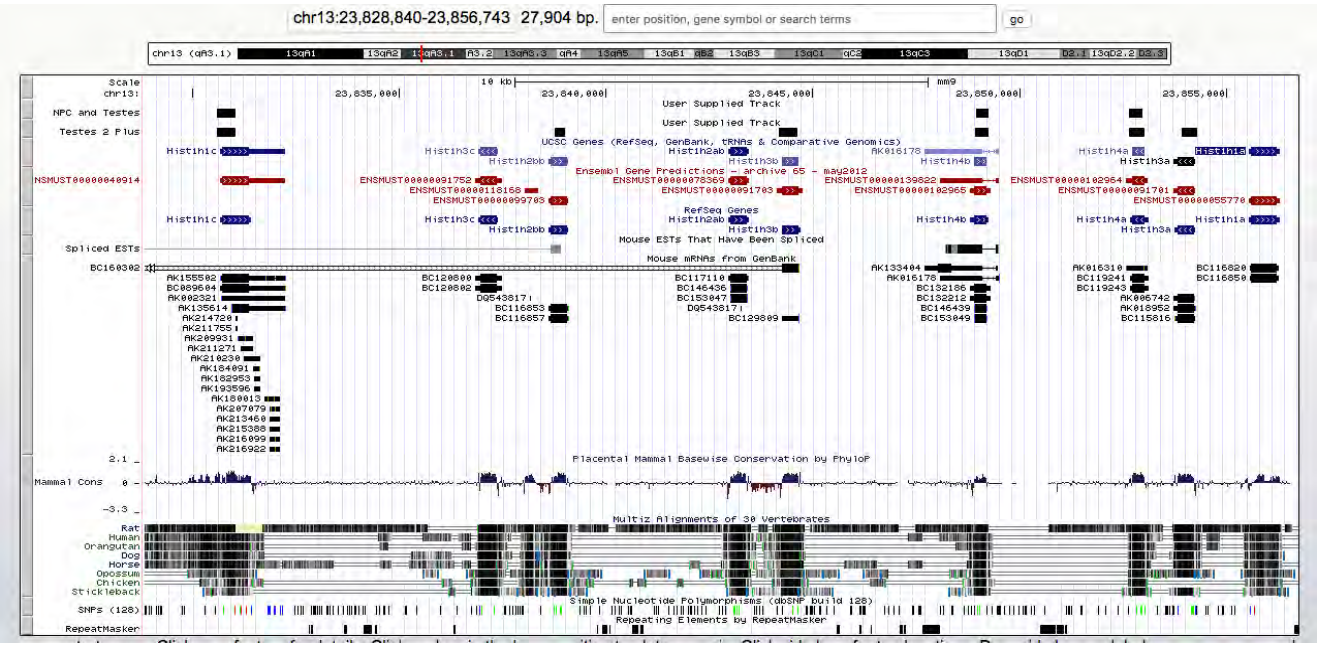


Figure 6

Comparison of the bound regions between the NPC and Testes cell types identified 123 genomic regions that are bound within both ChIP-Seq data sets (A). The GO terms of this subset were assessed via GREAT, with the most significant enriched terms being shown in B. UCSC browser images of the histone gene clusters on chromosome 3 and 13 show the overlapping peak regions (C and D). The track 'NPC and Testes' only identifies peaks present in both datasets, while Testes 2 Plus identifies peaks present within at least 2 of 3 ChIP samples from the testes. Other tracks from the browser are also show, such as gene annotations, spliced ESTs and mammalian conservation.

Motif identification in peaks present in both NPCs and Testes

As binding motifs often differ depending upon partner factor interaction and/or cellular context, we wanted to determine whether a specific motif was enriched in the common testes/NPC ChIP sequences. Analysis of the 123 peak subset identified a SOX-like motif, with preference for an A or C at 'core' position 6 (Figure 7). This is most similar to the SOX-like Motif 1 (Figure 4A). Again, similar to the complete testes dataset, an SP1-like motif was identified, albeit in a much larger proportion of peak regions (60%). As these are mostly not present within the same regions it is difficult to assign them as partner proteins and the motifs may be present as coincidence, or as individual components of a larger enhancer.

A.

B.

C.

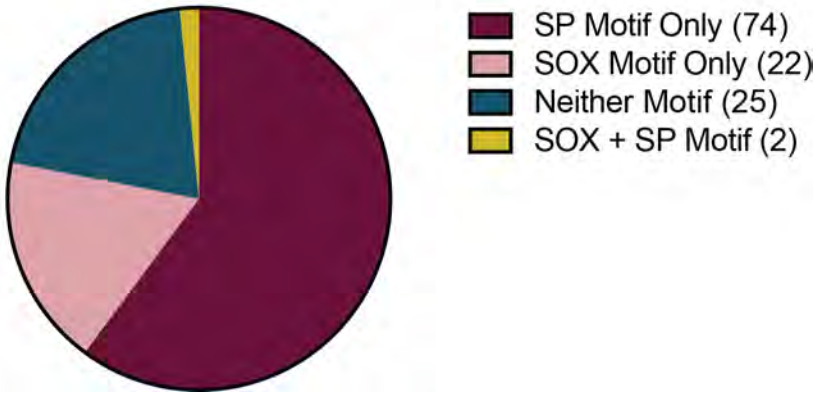


Figure 7

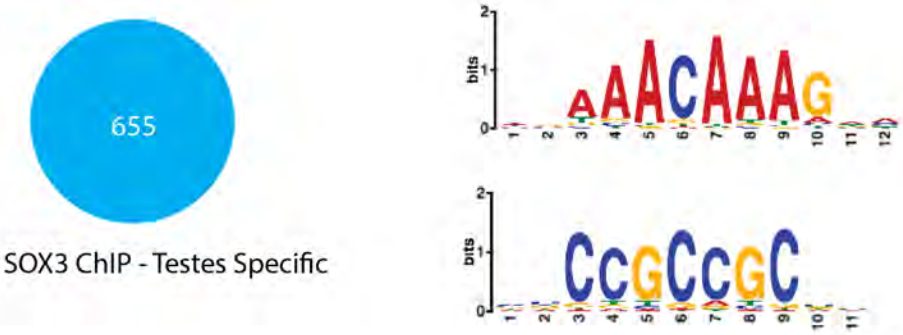
Motif analysis of overlapping regions between NPC and Testes ChIP-Seq data identified a SOX-like motif (A) as well an SP1 like motif (B). C shows the overlap between SOX and SP1 motifs within the 123 peak sequences, with the majority of sequences containing only an SP motif, and only 2 sequences containing both motifs.

Peaks only present within Testes

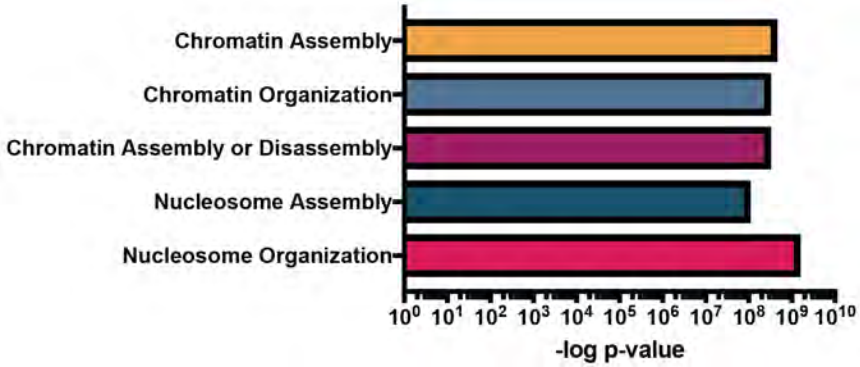
We were interested to determine if there was a different subset of motifs or functions associated with testes-specific peak regions. Within this group of 655 peaks, we still identified a SOX-like motif and an Sp1 like motif – however these do differ slightly from those found in both the NPC overlap data and the complete 778 region dataset. The SOX Motif is most similar to Motif 1 and 3 from the original dataset, maintaining an A at position 9 (core position 6) rather than a T. The GO Terms generated are highly focused on chromatin and nucleosome organisation and assembly (Figure 8B). Interestingly, the 'negative regulation of megakaryocyte differentiation' was not enriched in this dataset unlike the NPC/Testes overlap and compete testes dataset, suggesting this may have been associated with regions also active within a neural context. Again, many histone genes were associated with these GO Terms as can be seen in Appendix 2.

Two peak regions of interest were located nearby to testes specific histone genes, *Hist1h2b* and *H3t* (GM12260) (located within another gene, *Trim17*). These peak regions can be seen in 8C and 8D, and are specific to only the testes dataset, and not found within the NPC data.

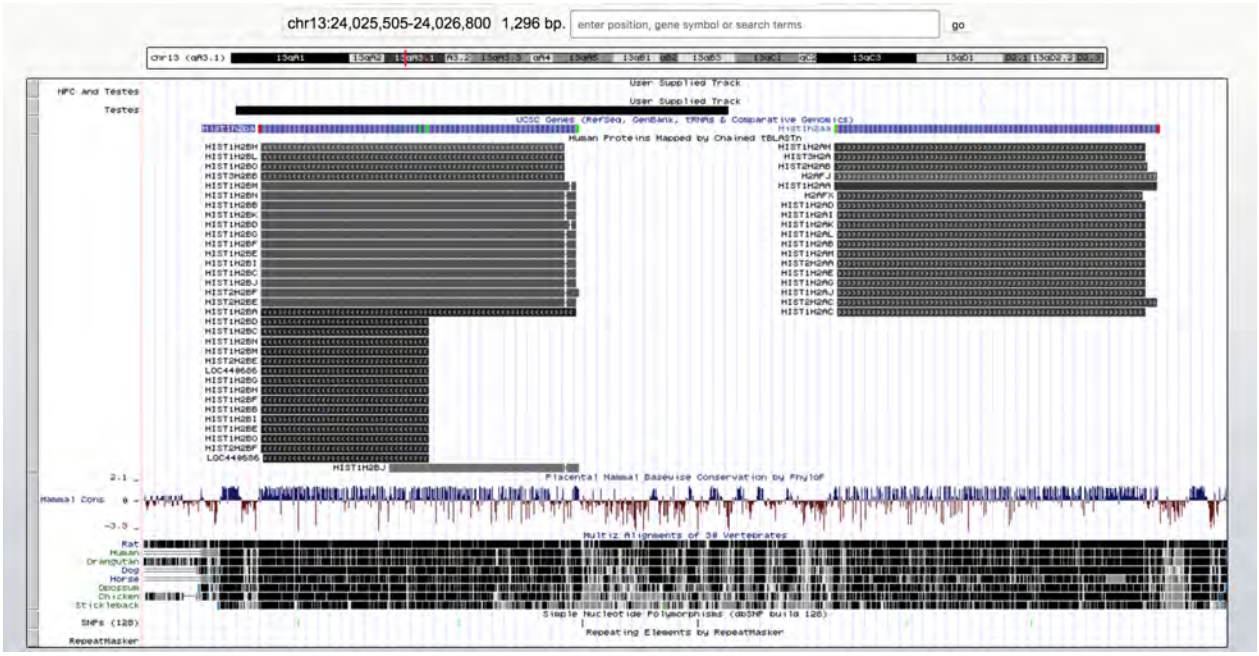
A.


655
SOX3 ChIP - Testes Specific
bits

B.


Chromatin Assembly
Chromatin Organization
Chromatin Assembly or Disassembly
Nucleosome Assembly
Nucleosome Organization
-log p-value

C.



D.

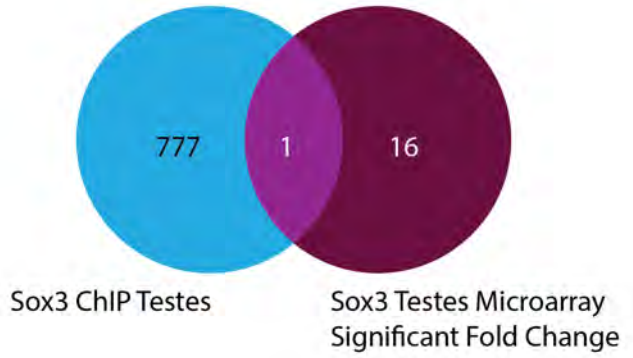
Figure 8

Analysis of the 655 testes-specific bound peaks identified both a SOX like motif and SP1 like motif (A). The significantly enriched GO Terms generated using GREAT show chromatin organisation and assembly like terms, with chromatin organisation giving the highest score. C and D show UCSC browser images of regions only found within the SOX3 testes ChIP-Seq located near testes specific histone genes (C- Hist1h2aa and Hist1h2b, D- H3t). The tracks 'testes' indicates the peak is present in at least 2 of 3 ChIP samples. In D, the testes specific histone gene H3t is indicated in blue by GM12260 as it was not given an official title in mm9.

Regions present in ChIP-Seq and Microarray

A microarray performed on postnatal mouse testes of both wildtype and *Sox3* null mice has previously been performed to elucidate SOX3 target genes (Adikusuma et al. 2017). From this data, only 18 genes were found to be significantly up or downregulated in response to the lack of SOX3 (Figure 9). To identify putative SOX3 direct target genes, we compared these regions to the ChIP-Seq data, and found that only 1 nearest-neighbour peak sequence appeared to have a dysregulated gene in response to SOX3, neurogenin3 (*Neurog3*, also known as *Ngn3*). In the microarray, *Neurog3* is shown to be downregulated 2.9 fold, with only 0.35 gene expression compared to wildtype. The peak sequence for *Neurog3* can be seen approximately 3kb upstream of the TSS, and is within a larger region of conservation, potentially a large regulatory region.

A.



B.

Peak Region	Nearest Gene	Conservation Score	Microarray Fold Change	SOX Motif
chr10: 61591506-61591858	Neurog3/NGN3	0.65	-2.9	GGAAACAAAGTA

C.

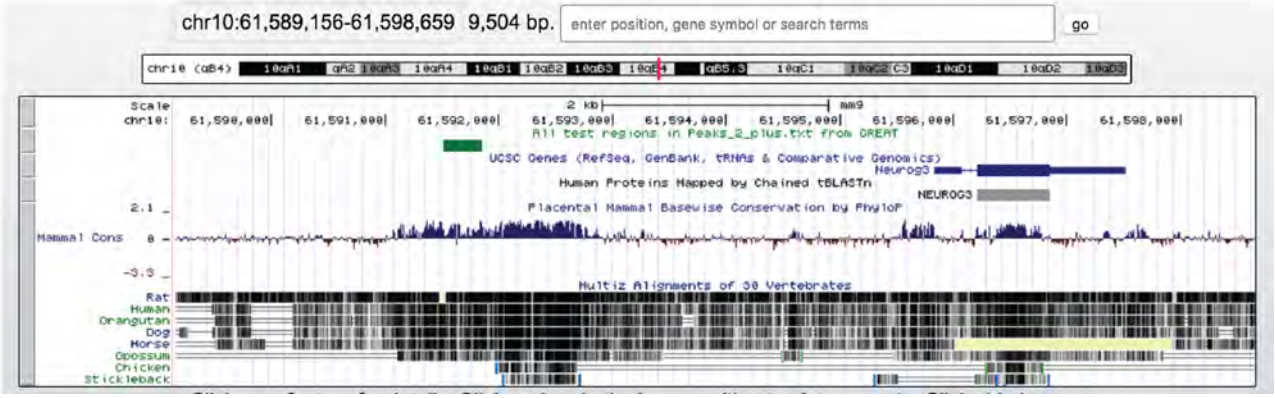


Figure 9

Overlap of the testes ChIP-seq peaks nearest genes with microarray data from SOX3 null mice only produced one gene, *Neurog3* (A). B shows the details of this peak regions, including genomic location in mm9, conservation score and SOX motif. C shows the UCSC browser image, with the peak sequence seen in green overlapping a large region of conservation (mammalian conservation track) upstream of the *Neurog3* gene.

Discussion

SOX3 is expressed within the postnatal testes, and whilst it is known that loss of this transcription factor leads to spermatogenesis defects and sterility, its transcriptional targets are currently unknown (Raverot et al. 2005; K. M. Weiss et al. 2003). Through 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 regulator regions that may be important for the initiation and progression of spermatogenesis, as well as the regulation of histone gene expression.

The highest scoring TF motifs identified by de novo analysis were consistent with known SOX binding motifs, namely the core SOX motif ($G_1A_2C_3A_4A_5A_6G_7$) (Hou et al. 2017). From the three closely-related motifs that were identified, the most obvious difference is the strong preference for a T at core position 6 rather than A (Motif 2). However it is worth noting that SOXB1 motifs have been identified with either A or T at this position, and it has previously been shown that SRY, which is closely related to SOX3, favors the T at this position rather than A (Harley 1994). Previous ChIP-Seq analysis of SOXB1 binding within neural progenitors identified this position as being equally favoured as either A or T (McAninch & Thomas 2014). Interestingly, each of these motif variants were enriched in specific subsets of peaks. However, functional GO terms were not associated with each of these groups, so the significance of this remains to be determined. Another difference between the identified motifs (AAACAAAG, AAACAATG, ACAAAG) and the core motif (GACAAAG) is the lack of a G at position 1. In all motifs identified there was either an A at position 1, or no strong preference. This is not surprising as many SOX factors have shown preferences for A at this position, including SOX18, SOX9 and SOX4 (Hosking et al. 1995; Kamachi & Kondoh 2013).

Together these data may indicate that a subset of SOX3 binding that occurs within the testes, preferentially binds to the AACAATG motif. While within the CNS, and targets that are common between testes and CNS SOX3 binds with the other SOXB1 factors to AACAAAG. However, until further analysis can be performed on these regions it is difficult to assign a functional reason for this preference.

SOX TFs do not appear to require strict binding motifs and have been shown to bind to transcriptional targets of other family members when ectopically expressed. For example, SOX3 can bind to the SRY-target enhancer upstream of SOX9 (TESCO) (Sekido & Lovell-Badge 2008a) when ectopically expressed in Sertoli cells (Sutton et al. 2011), and SOX2 is able to almost fully compensate for SOX3 when ectopically expressed in the testes using the endogenous *Sox3* promoter (Adikusuma et al. 2017). It is thought that while all SOX proteins have a core motif that they need to bind, the presence of different flanking nucleotides is what allows the specificity of binding by each protein (Mertin et al. 1999).

SOX proteins are often shown to bind co-operatively at target genes with a partner protein (Kamachi et al. 2000). These partner proteins are thought to help increase the specificity of TF binding, as often more than one SOX factor is present in a tissue type. None of the known SOXB1 partner binding proteins such as the POU or PAX family members were enriched in motif searches of the testes dataset. SOX3 has been shown to bind co-operatively with OCT4/POU in the nervous system (Wilson & Koopman 2002; Tanaka et al. 2004; Ng et al. 2012), and this was observed within a SOXB1 ChIP-Seq dataset of mouse NPCs (McAninch & Thomas 2014b). Although OCT4 is known to be expressed within the gonocytes and spermatogonial stem cells along with SOX3, we did not see any evidence of these two TFs acting together at this stage and tissue to regulate gene expression.

In identifying the SOX3 motif within the peak regions, a SP1 motif was also found within a large proportion of these regions. SP1 is a zinc finger transcription factor with a large array of functions including apoptosis, cell growth and chromatin remodeling (Tan & Khachigian 2009; Cakouros et al. 2001). Sp1 is expressed throughout the early stages of spermatogenesis, decreasing at the pachytene stages (Panigrahi et al. 2012), while SOX3 is expressed throughout these same regions and timepoints (Raverot et al. 2005). In our analysis, we show that 27 peaks regions contain both a SOX3 motif and an SP1 motif, which may represent a subpopulation of genes that are regulated by both SOX3 and SP1/SP2 TFs. Whilst SP1 has not previously been shown to partner with SOX3, it has been implicated as a binding partner with other SOX proteins, such as those from the SOXE subgroup, as well as SRY, which is thought to have evolved from SOX3 ((Wissmüller et al. 2006; Nagai 2001). This subgroup of peak regions may represent a group of genes containing enhancers bound by both SOX3 and SP1. Whilst there is little evidence that these are partner factors, they may regulate the same genes either as part of larger enhancer complexes, or could each control an aspect of the spatial or temporal expression of gene by acting separately at individual enhancer regions.

Although only 204 of the 778 peaks were identified as containing at least one of the three SOX motifs, this may not be a complete representation of SOX3 binding. There is often flexibility in TF binding motifs, which can be influenced by partner factor binding. It is possible that there are some motifs which were not found via the motif searching algorithms due to low occurrence within our datasets. It is also possible that some of the peak regions generated are not bound by SOX3, but are a consequence of the looping interactions generated between enhancers and other TFs or promoters. Protein:Protein, rather than Protein:DNA interactions can be missed in ChIP based methods due to the lack of DNA binding. If this is occurring,

there would not be a SOX3 motif found as it would be binding a protein not identified in the ChIP-Seq. Chromatin capture experiments or ChIP combined with mass spectrometry could be applied to these regions, and help determine any partner factors that may bind at the same targets with SOX3, potentially regulating gene expression.

The peaks identified from the testes ChIP-seq are mostly found within intergenic and intronic regions, consistent with other ChIP-seq analyses (Ohba et al. 2015; C.-F. Liu & Lefebvre 2015; McAninch & Thomas 2014b). This suggests a regulatory role such as enhancer binding. Conservation analysis via PhastCons scoring showed that 12% of bound regions had high conservation, indicative of their importance within the mammalian genome. When only the highest conservation peaks are assessed, the genomic location of these changes significantly, with a higher percentage of exon and promoter associated regions now seen. The subset of exon-bound regions within the 'high conservation' group could be due to the generally much higher level of conservation of exonic regions compared to non-coding regions, and the high percentage of promoter bound regions may also be a symptom of this, albeit more functional. Recent evidence does however suggest that exons can function as tissue-specific enhancers (Birnbaum et al. 2012).

SOX3 is present within both the CNS and the testes, and it is unknown id they share common targets. Using a ChIP-Seq dataset of SOX3 binding in NPC we could identify if binding regions and TF motifs were shared. The 15% of testes peak regions also identified as bound in NPCs, show GO terms again suggestive of nucleosome and chromatin regulation and assembly. This indicates that SOX3 may be an important regulator of histone expression throughout tissues types, and further experiments to investigate if this occurs within other tissues or at different developmental timepoints would be beneficial. Although the main GO terms

produced from the peak data appear to be due to the proximity of binding closely to histone and histone related genes, the 3 most common SOX motifs identified in the initial searches are not found in these regions.

During the progression of spermatogonia through spermatogenesis there is complete reorganisation of the chromatin structure within the nucleus (Bošković & Torres-Padilla 2013). Whilst almost all cells use histones to "bundle" their DNA, sperm cells are unique in that histones are replaced by protamines during maturation. In our analysis of SOX3 binding, we saw many peaks at the hist1 and hist2 clusters, as well as other key histone genes, *H2afx* and *Hist1h2ba*. Testes-specific peaks were detected near histone 3 genes, although not in the same cluster-like formation seen at hist1 and hist2 regions. These genes are responsible for the majority of the GO terms associated with the peak dataset due to their importance in regulation of chromatin and nucleosome organisation. The vast majority of these did not however contain identifiable SOX binding site motifs within the peak sequences, and it may be that SOX is binding with an unidentified partner at these regions. Further investigation into the expression regulation of these genes within each of the stages of spermatogenesis would again be beneficial to try and understand the relationship between SOX3 and regulation of histone genes.

The binding of SOX3 close to many histone and chromatin-associated genes suggests that it could be playing a role in the regulation of histone expression throughout spermatogenesis; potentially during the essential histone to protamine transition that must occur for viable sperm. The lack of any significant differences in expression of these genes in SOX3 knockout testes suggests that although it may be involved, there are most likely redundancy pathway present, or the putative enhancers are able to compensate for the loss of a single factor. Functional studies

removing the potential enhancers in vivo and assessing both changes in gene expression as well as phenotypic analysis of the testes and sperm may help to determine if SOX3 is involved in these pathways.

One of the most interesting regions identified within the ChIP-seq is a binding peak between the *Hist1h2aa* and *Hist1h2ba* genes, which encode TH2A and TH2B, and have been shown to utilize the same regulatory region {Huh:1991uz}. These are testis-specific histone variants that replace the canonical core histone proteins H2A and H2B within the early spermatocytes (Montellier et al. 2013). Depletion of Th2B in spermatocytes does not seem detrimental as compensatory mechanisms are induced allowing for an upregulation of the canonical H2b histone, allowing spermatogenesis to be completed. However, loss of both Th2b and Th2a which are controlled by the same regulatory region between the two genes does cause defects in spermatogenesis, leading to infertile males (Shinagawa et al. 2015). Another testes specific histone variant, H3t has been shown to be essential for the entry into spermatogenesis (Ueda et al. 2017). Whilst this gene is not an "officially-named" gene within the mm9 database to which out ChIP-Seq data was mapped, a binding peak is located within the same region as the predicted gene, GM12260-001 which is located within an intron of *Trim17*. The h3t histone variant is shown to be incorporated into the genome prior to spermatogenesis, displacing the canonical H3 variant. This replacement, although only 3 amino acids different to H3 is thought to generate a more flexible and open chromatin structure, necessary for the progression into meiotic recombination (Ueda et al. 2017). Similar to SOX3 knockout mouse models, loss of *H3t* leads to reduced testes weight and a loss of germ cells (Ueda et al. 2017). Further analysis of the SOX3 binding at this region is warranted, and could show 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.

While SOX3 has not yet been shown to be involved in the regulation of these testes specific histone genes, the finding of SOX3 binding within the known promoter region, and overlapping an important gene is worth further investigation. Combined with the evidence of SOX3 binding near or at many other histone genes suggests it plays a role in their regulation. Although TH2B and TH2A do not replace the canonical histones until the early spermatocyte stage, and SOX3 is thought to be expressed within the earlier spermatogonia, it may be that SOX3 is acting as a pioneer transcription factor or is pre-binding genes prior to their expression, as other members of the SOXB1 family have been shown to do within neural contexts (Hagey et al. 2018; Bergsland et al. 2011; Klum et al. 2018). We would expect that if SOX3 was important in directly regulating either TH2A or TH2B, that this would be reflected in the microarray data of SOX3 KO testes, of which only 18 genes were shown to be differentially expressed. H3t was not identified as a functional gene when the microarray experiments were performed, and as such was not measured. Further experiments such as single cell RNA Seq of each stage of spermatogenesis would give a much better overall view of the changes in expression of these genes, when they are activated, and how long they are expressed for. If this was compared to a *Sox3* null mouse model the nuanced changes in expression may be apparent and able to compared more closely.

Another gene that may be regulated by SOX3, is *Neurog3,* which encodes a basic Helix Loop Helix transcription factor, and known mostly for its role in the differentiation of endocrine cells and NPCs (Gradwohl et al. 2000; Pelling et al. 2011). NEUROG3 mirrors the expression pattern of SOX3 during early spermatogenesis, being found in As, Apr and Aal spermatogonia (Yoshida et al. 2004; Yoshida et al. 2006), and has been shown that its expression is abolished in a *Sox3* null mouse line (Raverot et al. 2005). 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 (Kan et al. 2004), led Raverot and colleagues (2005) to propose a functional link between SOX3 and NEUROG3 in the differentiation of spermatogonia. In the microarray data of *Sox3* null testes (Adikusuma et al. 2017), its expression was reduced to 50% of normal levels, indicating a strong reliance on SOX3, whether through direct enhancer binding or a signaling pathway involving SOX3. We propose that the highly conserved binding site upstream of *Neurog3* is an enhancer for NEUROG3 expression within SSCs and the early spermatogonial cell types; however functional analysis of this enhancer would need to occur before this can be fully validated. When looking at the putative enhancer region within the UCSC genome browser, it is evident there is approximately 1kb of highly conserved DNA surrounded by less conserved regions upstream of NEUROG3. As the SOX3 peak only covers a few hundred bp of this, it is highly likely that there is a number of other TF binding sites present that are working together to control expression. Determining if this region is an enhancer could first be examined through use of a reporter assay. By using the entire conserved region to drive expression of a LacZ reporter and assessing expression of this within the testes could show if this region is active and if it can drive expression to the expected regions. Further ChIP experiments with different TFs would be interesting to show what other factors are binding close to this peak, to see what factors are important in NEUROG3 expression.

In summary, through ChIP-Seq analysis within the postnatal mouse testes, we have identified almost 800 regions that are bound by SOX3, including three putative enhancers implicated in spermatogenesis. 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 role for SOX3 in regulation of histone genes, and their involvement in chromatin and nucleosome assembly and organisation. To

show functional evidence of co-regulation or partner protein activity with SP1, ChIP combined with Mass Spectrometry would be a logical next step. Other experiments such as proximity ligation assays could provide visual confirmation of interaction between the two transcription factors. However this does not prove functional interaction but could show binding at the same locations. Single cell RNA Seq data on each stage of spermatogenesis, from spermatogonia to spermatocytes would be an invaluable resource for both this study and others to decipher gene expression changes during the complex progression of sperm development. This data would allow us to correlate the SOX3 binding with active or inactive genes, and help to determine if SOX3 is acting as a transcriptional activator, repressor or is pre-binding to enhancers to allow other TFs access.

Chapter 5

General Discussion and Future Directions

Nestin

Widely used in transgenic mouse model research, the Nestin promoter and enhancer driven Cre recombinase, Nes-Cre, is known to direct expression of target genes to the neural stem and progenitor cells within the CNS ((Liang et al. 2012; Isaka et al. 1999; Trumpp et al. 1999; Petersen et al. 2002; Dubois et al. 2006). First discovered in 1994 in rat as a CNS-specific enhancer of expression (Zimmerman et al. 1994), further research on the region has identified a smaller 30bp section comprised of SOX and POU binding motifs that are essential for reporter gene expression (Tanaka et al. 2004). Although reporter assays are able to show the CNS-specific activity of this enhancer, they are unable to demonstrate the contribution to endogenous Nestin gene expression during development.

In Chapter 2, functional analysis of the SOXB1 bound Nestin enhancer was undertaken. From this, we wanted to ascertain the endogenous functionality of the enhancer, and whether it was required and/or essential for *Nes* expression within the neural progenitor cell population of the developing mouse CNS. Initially, we removed the enhancer using CRISPR and assessed two independent lines with slightly different deletions, both of which showed significant reduction in *Nes* expression via qPCR at 12.5dpc. Further investigation into the -255 line showed this reduction in expression was present from 9.5 until 15.5dpc, with the most dramatic change being a 70% reduction in *Nes* expression within the head.

Our results confirmed previous assumptions that this enhancer is required for full expression of NES within the developing CNS. While spatial expression had been tested through LacZ reporters, the endogenous temporal expression of the enhancer has not previously been analysed (Tanaka et al. 2004).

In our analysis of the *Nes* enhancer, the earliest timepoint analysed for *Nes* expression within the head was 8.5dpc. At this timepoint, all 3 SOXB1s as well as POU proteins are present within the CNS in the NPCs (Uchikawa et al. 2011; Wood & Episkopou 1999). Gel supershift assays using OCT1, BRN1 and BRN2 antibodies indicate that OCT1 binds the *Nes* enhancer at 8.5dpc, followed by BRN1 and BRN2 (Z. Jin et al. 2009). We were able to show that loss of this enhancer does not affect *Nes* mRNA expression at this early stage. However, from 9.5dpc *Nes* expression is reduced within the head, suggesting it is only active from this developmental stage onwards. This could be due to the requirement of other TFs to bind to the region. Indeed, binding sites for SF1 and NKX2.1 as well as hormone responsive elements have been identified within the enhancer, although the activity of these have not been tested *in vivo* (Pelizzoli et al. 2008; Z. G. Jin et al. 2009). There is large variation across enhancers in the number of TFs needed to bind for effective regulation of gene expression, making it difficult to know which or how many TFBS are functional at each enhancer. Further bioinformatic analysis along with integration of ChIP data could help to unravel which TFBS present are functional and required for *Nes* enhancer activity. In comparison with the endogenous enhancer activity, transgenic mouse models utilising the *Nes* enhancer to drive Cre recombinase, recombination begins to be seen between 8 and 10 dpc within the ectoderm, mesenchyme, somites and mesonephric cord (Dubois et al. 2006). This indicates that the activity may be regulated by the chromatin structure, which may be inaccessible endogenously, but easily bound within an open conformation transgene element.

It is possible that the reduction in *Nes* expression when the enhancer region is removed is simply due to the change in the local chromatin structure, and not due to the removal of SOX and POU sites. The architecture of the chromatin environment is essential for proper gene regulation, due to the dynamic nature of

chromatin and the need to form topologically associated domains (Kragesteen et al. 2018; Spielmann & Mundlos 2013). As a way to overcome this issue in future experiments, point mutations that alter nucleotides critical for TF binding would be a more useful approach. In these more refined experiments, individual SOX and POU motifs that contribute to enhancer activity could be identified. This strategy could also be used to determine if this enhancer requires all TF to be bound for full activity, or if each individual TF is capable of a small contribution to overall expression. Enhancer deletion is becoming a widely used approach to assess the *in vivo* activity of these regulatory regions (Szafranski et al. 2017; Johnson et al. 2018; Dickel et al. 2018), aided by the continually improving CRISPR technologies which mean specific and precise mutation generation is becoming easier and more accessible. The newest form of genome editing, base editing, is now possible whereby individual adenine residues can be converted to guanine, and cytosine to thymine by dCas9 molecules fused to adenosine and cytidine deaminases, respectively, without inducing double strand breaks (Komor et al. 2016; Komor et al. 2017; Gaudelli et al. 2017). This is an attractive strategy for modifying individual TF motifs without interrupting the 3D architecture, however recent reports have highlighted possible increases in off-target effects using this strategy (Zuo et al. 2019) although these effects can potentially be mitigated in an established mouse line through breeding.

Nestin was first discovered as a marker of neural stem cells, hence where its name is derived; **Ne**uroepithelial **St**em cells (*Nes*) (Lendahl et al. 1990). Nestin is an intermediate filament protein, although its exact function within the cell and during development are not completely understood. It is expressed in many cell types, mostly those during periods of active proliferation (Park et al. 2010; Wiese et al. 2004), including neural precursor cells, muscle progenitors, newly formed blood vessels and some cancers. In 2010, Park et al published data indicating that loss of

Nestin caused embryonic lethality due to its role in the self-renewal of NSCs, rather than its role as a structural protein within these cells' cytoskeleton. We then generated two independent NES null lines using CRISPR, an entire coding sequence deletion, and a frameshift mutation. Contrary to Park et al (2010) we showed that NES null mice and embryos are viable, fertile and show no obvious phenotypic effects. Other studies have also shown similar results with no defects (Austin et al. 2004), or with minor phenotypic defects such as impaired motor co-ordination (Mohseni et al. 2011).

Comprehensive phenotyping was not performed on our two NES null lines. If this was performed, it may show if similar motor coordination or subtle CNS defects are present and consistent with previous models. In the most recent knockout model by (Mohseni et al. 2011) and colleagues it was found that the number of acetylcholine receptor clusters and nerve length are significantly increased in mice lacking NES. These would be interesting to assess in both our knockout model of Nestin, as well as the enhancer deficient mice to ascertain if any of these phenotypes could be due to the loss of Nestin specifically within neural precursors. It is possible that the variation in phenotypes seen across these studies could be due to the use of different targeting vectors and recombination cassettes. Other factors such as differing mouse strains can lead to vastly different outcomes (Montagutelli 2000), however in this case, our experiments as well as those mentioned all used C57BL/6 mice. This highlights the usefulness of CRISPR in generating 'clean' deletions, eliminating any potential side effects that may be caused due to selection cassettes left in the genome.

During our analyses of Nestin expression, it could be seen that animals heterozygous for the enhancer deletion showed consistently higher expression than expected. We were able to provide evidence that this enhancer is acting in both *cis*

and *trans* to control gene expression. Generally, mammalian enhancers are viewed as *cis*-acting. In contrast, the phenomena of *trans*-activation of gene expression has been described for many decades in *Drosophila* models, where it is known as transvection (Lewis 1954). Our results are consistent with recent findings that some super enhancers such as those required for red blood cell development and IgH expression are capable of interchromosomal interactions between enhancer and promoter (Alvarez-Dominguez et al. 2017; Le Noir et al. 2017). These experiments are often difficult to perform in a wildtype mouse system as the mRNA expressed from each allele appears the same in qPCR based experiments. One way of overcoming this issue is to use two different strains of mice that contain SNPs between coding sequences, so individual alleles can be distinguished by sequence variation. These would need to bred with a enhancer deleted mouse line, and the resultant mRNA from each allele can then be analysed.

In *Drosophila*, Lim and colleagues (2018) have demonstrated real time visualisation of *trans* enhancer-promoter interaction within the developing embryo (Lim et al. 2018). They found that for the developmental enhancer *Snail*, *gypsy* insulators at each chromosome were critically important for the *trans* interactions to occur. This is thought to be due to their effect on stabilising the interactions between the two chromosomes (Kravchenko et al. 2005). In vertebrates, the only known insulator bound protein is the transcription factor CTCF, which is known to be critical for regulating the 3D structure of chromatin (Bell et al. 1999; Kim et al. 2015). Future experiments to understand the interchromosomal interactions of the *Nes* enhancer could focus on identification of putative insulator sequences. ChIP experiments assessing CTCF binding within NPCs at timepoints where the *Nes* enhancer is known to be active may show binding of CTCF flanking the enhancer. Previous genome-wide studies of CTCF binding have identified over 800 000 CTCF binding sites within the mouse genome, although use of the online tool CTCFBSDB 2.0

does not recognise any experimentally verified CTCF binding near the *Nes* gene (Ziebarth et al. 2013). This does not however take into account cell and tissue types. If found through ChIP binding, further experiments removing the CTCF binding regions and assessing *trans* enhancer activity could show its effect on expression from each allele.

NES is known to be expressed within the endothelial cells of the blood vessels within various tissues, and can also be used as a marker of neovascularisation (Suzuki et al. 2010; Mokrý et al. 2008). We observed that when the *Nes* enhancer was deleted, ectopic expression of NES was seen within the vasculature of the neural tube. Whilst there is no identified enhancer responsible for this vascular expression, we proposed that by deleting the enhancer region other important regulatory elements may have been disrupted. As the expression was ectopic, rather downregulated, we reasoned a repressive element may be present within the 255 bp enhancer. Repressive elements have been identified for other genes, such as GAP-43 where a 30bp region inhibits expression in non-neuronal cell types (Weber & Skene 1997), as well as the RE1/NRSE repressor that is found at a group of neural specific genes. There is not currently any evidence as to what other TF may be bound at the *Nes* enhancer, and future ChIP data would be needed to confirm any putative TF motifs identified. This result highlights the importance of endogenous enhancer analysis, as this effect would not be detected only using reporter assays or mouse models. As CRISPR deletions become the standard method for enhancer analysis, it is possible that more previously unknown repressive elements may be identified.

Frizzled 3

FZD3 is a Wnt receptor protein involved in the Wnt/b-catenin canonical signaling pathway that controls many developmental programs (Sala et al. 2000). There has not previously been any link between the SOXB1 transcription factors and the regulation of *Fzd3*, apart from the fact that they are expressed within the same regions of the CNS during embryonic development. Whilst the SOXB1 proteins maintain the neural progenitor cells within a stem-like state, it appears that *Fzd3* along with other proteins is important for the cell fate decisions of NPCs through activation of Notch pathways (Wang et al. 2016).

ChIP-Seq studies of the SOXB1 factors in NPCs identified the putative *Fzd3* enhancer within intron 2 of *Fzd3* (McAninch & Thomas 2014). CRISPR mediated deletions of the SOXB1 motif and the larger putative enhancer enabled us to assess endogenous activity during embryonic development of the CNS. We detected a small but significant decrease in *Fzd3* expression when only the SOXB1 motif was removed, and a more striking effect of a 30% expression reduction when 573bp of putative enhancer was removed. No previous enhancers or regulatory regions of *Fzd3* have been identified within the literature. By generating two different deletions we were able to show that the SOXB1 motif is responsible for a small but significant proportion of this expression. The larger enhancer deletion (-573) provides evidence that it is highly likely other TFs are binding and controlling expression of *Fzd3*. To determine what these other TFs are, luciferase reporter assays using TFs known to be involved upstream of Fzd3 signaling pathways could identify initial targets. Alternatively, bioinformatic approaches such as integration of other ChIP datasets or motif searching software could be used.

During our analyses, we were interested in the control of *Fzd3* within the NPC population via SOXB1 proteins. However, *Fzd3* is expressed within a variety of tissue and cell types including the thymus, testis, skin and hair follicles. (Romanowska et al. 2009; Papatheodorou et al. 2018). By utilising the *Fzd3* enhancer deletion model, the expression of *Fzd3* can be measured in these other tissues to define the spatial and temporal activity of the enhancer. This could also help to identify any other TFs that may be regulating *Fzd3* within these other contexts, through analysis of TFs known to be expressed in regions where *Fzd3* expression is modulated.

Although removal of the 573bp putative enhancer showed a marked decrease in *Fzd3* mRNA expression, no obvious phenotype was observed. FZD3 is an essential protein, without which mice die soon after birth due to brain and neural tube defects (Wang et al. 2002; Wang et al. 2006; Peeters et al. 1998), although heterozygotes develop normally. We generated a *Fzd3* null line, which showed phenotypes consistent with previous reports, with pups displaying a curly tail and dying soon after birth. We used this line to generate compound heterozygote mice to determine the level of *Fzd3* expression sufficient for survival, and showed that only 40% of wildtype was sufficient. Although we could not detect any phenotypes with the compound heterozygote line, this is a useful approach to determine the levels of expression required for development of embryonic lethal genes. It is often difficult to generate mouse models with expression levels outside of 100%, 50% or 0% expression, and this method can be applied to many genes with known enhancers.

The lack of phenotype in the compound heterozygote *Fzd3* mouse was not entirely unexpected, as genetic and functional redundancy has previously been shown between FZD3 and FZD6 (Dong et al. 2018). These proteins are closely related in

Frizzled protein evolution (Huang & Klein 2004), and FZD3, along with FZD6 is known to be important in the closure of the neural tube. Mice lacking both these proteins die within minutes of being born due the failure of this essential process in CNS development, a more severe phenotype than seen in FZD3 or FZD6 single knockouts (Y. Wang et al. 2006; Stuebner et al. 2009). It may be that the SOXB1s are important in regulating the expression of *Fzd3* within the neural tube for this closure to occur, but the phenotype could masked due to the functional redundancy with FZD6. FZD6 knockout mice are viable, showing hair patterning defects (Guo et al. 2004) but no CNS phenotype, presumably due to the presence of FZD3 within these areas. By generating mice lacking FZD6, and crossing these with mice homozygous for the *Fzd3* enhancer deletion, possible effects of the *Fzd3* enhancer loss may become more apparent and be able to be studied in more detail. This functional redundancy has also been observed within the control of planar cells polarity of hair development as well as midbrain development which are potential further avenues to explore with *Fzd6* null, *Fzd3* enhancer deleted mice (Dong et al. 2018; Stuebner et al. 2009; Wang et al. 2006).

As previously mentioned, throughout our experiments we were able to show reduction in *Fzd3* mRNA, however no phenotype could be seen even when expression was reduced to 40% of wildtype levels. It is possible that this RNA loss is not recapitulated as a loss in protein expression. These two measures are not always proportional (Liu et al. 2016), and without a suitable antibody we are unable to visualize the effect on FZD3 protein. To overcome this, an endogenously tagged FZD3 mouse line with a HA or FLAG could be generated via CRISPR. This could then be used to again generate enhancer deletions, or for more refined targeted mutations of specific TFBS within the enhancer using CRISPR HDR-repair strategies. The effect of enhancer mutation or heterozygous loss of *Fzd3* could then be visualised during development. This would give a clearer understanding of the role

of the regulatory regions on expression and how much these contribute to the overall expression of FZD3.

We have shown that loss of the putative enhancer of *Fzd3* can reduce mRNA expression within embryonic heads. We proposed that this was occurring in the NPCs under the control of SOXB1 factors, however the enhancer may be active elsewhere due to the influence of other TFs. We generated a LacZ reporter line using the 573bp enhancer region and assessed where this could drive expression within 12.5 dpc embryos. Expression could be seen in whole mount embryos throughout the neural tube, and specifically, within the floor plate. Similarly, *in situ* hybridization on wildtype mice results were able to identify expression of *Fzd3* throughout the neural tube, within both the midline, ventricular zone and the floor plate. Previous studies have assessed *Fzd3* expression via RNA *in situ* hybridization, and these show similar results to ours with expression throughout the neural tube and brain at 12.5dpc (Visel et al. 2004). Deletion of the *Fzd3* enhancer led to loss of expression mostly within the floor plate of the neural tube, with only minimal losses in the ventricular zone and no changes in the midline structures. This is in agreement with the LacZ reporter which consistently drove expression to the floor plate, where SOX2 is known to be expressed. During CNS development, neurons must extend their axons to generate neural circuits. Signaling molecules Netrin and Shh produced within the floor plate guide axons to extend, mediated by the growth cone on the elongating tip of the axon (Kennedy et al. 1994; Kennedy et al. 2006; Charron et al. 2003). Once at the floor plate, the axons cross the midline, followed by post-crossing rostral turning which is mediated by Wnt signaling and FZD3 (Onishi & Zou 2017; Stoeckli 2018). Our LacZ reporter and enhancer deletion lines show evidence that the putative enhancer region is required for full *Fzd3* expression within the floor plate. As previously mentioned, although we did not see any phenotype in the mice lacking the enhancer, we did not perform in depth

experiments assessing axon growth and guidance. Future experiments could be used to assess the guidance of axons along the rostro-caudal axis.

The *Fzd3* enhancer driven LacZ reporter mouse that was produced could be a useful tool for studying CNS development, particularly axon guidance. As axons are guided towards the floor plate due to the signaling gradients produced from this region, being able to specifically control expression within the floor plate could be beneficial. Using our enhancer region to drive a *Cre*-cassette could allow for conditional mutants to be made to study specific interactions within the neural tube without affecting other body systems.

Genome Editing to study Enhancers

Throughout this thesis, we described the deletion and subsequent phenotypic evaluation of two enhancers thought to be controlled by SOXB1 proteins. These enhancers, one of which was previously known, the *Nes* intron 2 enhancer, and the previously undescribed *Fzd3* enhancer were both identified within a SOX3 ChIP-Seq experiment of mouse neural progenitor cells (McAninch & Thomas 2014).

Using ChIP-Seq combined with other bioinformatic data such as level of DNA conservation across species, DNase hypersensitivity and methylation status is a common way to identify putative enhancers. However, these do not always guarantee that region identified will be functional within the genome, or if the degree to which it is functional can be identified.

For the two enhancers that were studied in this thesis, deletion of each showed no obvious phenotypic differences, despite previous evidence that each of the genes are essential for survival (Park et al. 2010; Y. Wang et al. 2006). Further analysis did show an appreciable reduction in expression of both *Fzd3* and *Nes* during embryogenesis, however this did not appear to have an effect on overall CNS development. These results reflect a trend in enhancer deletion studies, where even highly conserved enhancers of developmental genes appear to have very little to no effect on overall development. One of the first reports of ultraconserved elements not being as essential as previously predicted was published in 2007. Ultraconserved elements upstream of the developmentally important gene, *Arx* were deleted and both individual and combinatorial deletion produced viable mice with no obvious phenotypes. However, in 2018 when these kind of experiments were reproduced, it was shown that although mice could survive, deficits were seen in both body weight and neuron density; subtle defects in a laboratory setting but

that may reduce overall evolutionary fitness (Dickel et al. 2018). Our results are consistent with these reports that enhancers are only responsible for small amounts of expression, and that these are possibly additive with other enhancers or regulatory regions to control endogenous gene expression. Future experiments for enhancers of both *Nes* and *Fzd3* could focus on combinatorial enhancer deletions to determine the precise effects of each enhancer, and to understand if enhancer redundancy is a factor. There is some evidence that shows enhancers may be redundant, and indeed this was the original theory for the *Arx* enhancers, however the evidence of combinatorial enhancer deletions leading to additive effects phenotypically disputes this idea. Other reports, such as those by Oosterwalder (2018) showed a high level of functional redundancy between tissue-specific limb, brain and heart enhancers. It will be interesting in the future to determine if this is locus dependent, or if all enhancers are responsible for a certain subset of expression with no redundancy between regulatory regions.

Reporter assays, such as LacZ transgenic mice have been used for decades to assess the spatiotemporal activity of putative enhancers. Until the development of CRISPR/Cas genome editing, this was seen as the best evidence for enhancer functionality. We generated a LacZ transgenic mouse using the 573 bp *Fzd3* enhancer which consistently showed expression throughout the neural tube and specifically within the floor plate. While this correlated well with our previous result of *Fzd3* mRNA loss in the floor plate when the enhancer was deleted *in vivo*, it did not completely recapitulate the expression pattern. As CRISPR editing of enhancers is becoming more facile and accessible, these two methods should be used in conjunction. Consistent with this is a recent report of two putative *Tbx5* enhancers thought to be required for limb development with substantial evidence from LacZ reporter mice showing expression in all expected regions (Cunningham et al. 2018). Upon enhancer deletion however, no phenotype was seen, indicating either these

enhancers are non-functional endogenously, or there is a possible redundancy effect with another unknown enhancer. Another possibility is that small reductions in expression were present but these were not enough to cause a phenotype; however this was not analysed within the study.

Our experiments assessing the endogenous enhancer activity of *Nes* and *Fzd3* enhancers used CRISPR to delete regions of DNA. This enabled us to show that each of the enhancers contain functional DNA that when lost, affects expression of the target gene. These kinds of enhancer deletion experiment are similar to those found within the recent literature (Cunningham et al. 2018; Dickel et al. 2018; Meyer et al. 2019), however like these, they do not identify the exact TFBS that are essential for target gene expression. As discussed previously, utilising HDR repair templates with CRISPR/Cas9 to modify the individual SOXB1 binding sites will be beneficial to understanding the function of the enhancer more. However, for the *Fzd3* enhancer, deletion of single identified SOXB1 site yielded only very minimal changes in expression, indicating there are more functional regions throughout the 573bp deleted enhancer. Bioinformatic analysis of this region identifies many putative TFBS which without further evidence is difficult to assign any functionality to. As more ChIP-Seq datasets become available these should be used to identify any overlapping binding peaks which could identify putative regulators of *Fzd3* along with SOXB1.

SOX3 function within postnatal testes

As well as being an important transcription factor during central nervous system development, SOX3 is also required for proper gonadal function (Weiss et al. 2003). Throughout Chapter 4 of this thesis, the putative transcriptional targets of SOX3 within the postnatal testes was explored, and these results identified a number of testes-specific histone genes as well previously hypothesised interactors.

The SOX3 ChIP-Seq on postnatal testes revealed a group of almost 800 putative binding targets. As no other ChIP-Seq had been performed on the testes, or within any other non-neural contexts, the extent to which SOX3 is binding in the testes was unknown. However, there were significantly less targets than within the neural progenitors where almost 8000 targets were identified (McAninch & Thomas 2014). These were identified through cultured neural progenitor cells, rather than tissue such as the experiments in this thesis and could account for the higher rate of protein binding as it is not within an endogenous context.

Of the SOX3 binding targets identified within the testes, a large proportion with functional gene ontology annotations were consistent with nuclear organisation and chromatin regulation. The majority of genes proximal to these regions were either core histone genes clusters, or histone associated genes, or histone variants. Previous studies have not focused on SOX3 as a regulator of histone genes, although many of these genes are located close to peaks within the neural progenitor ChIP-Seq (McAninch & Thomas 2014), suggesting these may be a conserved binding site within different tissue contexts. Previous evidence implicating SOX3 in histone regulation comes from the sea urchin, where modification of a SOX motif upstream of H2A.Z was shown to decrease its expression (Hajdu et al. 2016). Although within our ChIP-Seq a SOX3 binding peak

is not observed near H2AZ, this adds evidence that the SOX proteins may be regulating histone variant expression. Analysis of other SOX ChIP-Seq datasets, such as SOX3 within NPCs (McAninch & Thomas 2014) and SOX2 within cortex, spinal cord, and lung (Hagey et al. 2018) may identify a binding peak near the H2AZ gene or other histone variants. If these are consistently identified within other datasets, the TFBS could be modified or deleted via CRISPR and the effect histone expression and gene regulation could be assessed.

Putative enhancers identified within the ChIP-Seq were located near two testes-specific histone variants, TH2B and H3t. These histone variants are incorporated into the sperm DNA at different stages. H3t is expressed is from P8 as the differentiating spermatogonia emerge, while expression of TH2B is initiated at P10 (Ueda et al. 2017). SOX3 expression is seen from P7 within the prospermatogonia, however defects in development are not observed until P10, when germ cell loss is apparent. As previously discussed, the remodelling of chromatin within sperm is an essential process whereby histones are first replaced by transition histones to protect sensitive regions such as the telomeres and centromeres (Govin et al. 2007; Gill-Sharma et al. 2012). Protamines then replace these to form the chromatin into a toroidal conformation modifying its epigenetic state (Hud et al. 1993; Mudrak et al. 2005) . The underlying mechanisms of histone displacement are not fully understood, and the finding of SOX3 binding near these genes within the postnatal testes may help to uncover regulatory regions involved. To further analyse the link between SOX3 and H3t prior to *in vivo* deletion, protein expression analysis of the SOX3 knockout mouse could prove valuable. By performing Western blots at each stage of spermatogenesis from P7 through to P14, it could be determined if the loss of SOX3 that leads to infertility involves a reduction in H3t or TH2B. If this is occurring, it could be hypothesised that SOX3 loss leading to spermatogenic

defects is partially due to defects within chromatin structure, which possibly affects many aspects of gene regulation in the differentiating spermatogonia.

During analysis of the ChIP-Seq dataset, a microarray performed on *Sox3* null testes was used to determine if any of the putative SOX3 target genes were affected by loss of SOX3. The microarray revealed only 17 genes were dysregulated in response to SOX3 loss, and *Neurog3* was the only gene implicated with both the microarray and ChIP data. This data is in agreement with previous work showing *Neurog3* expression to be dependent upon SOX3 within the testes (Raverot et al. 2005). These genes are thought to interact within the same pathway however a functional link between the two has not yet been proposed. Previous enhancers of *Neurog3* have been proposed within the large region of conservation where we identified SOX3 binding. STAT3 bound approximately 4kb upstream of *Neurog3* identified via a ChIP on undifferentiated spermatogonia, and was also able to activate expression of a luciferase reporter of the 4kb upstream of *Neurog3* (Kaucher et al. 2012). These data indicate the SOX3 binding peak we identified is part of an enhancer, and further experiments to identify is SOX3 is active within it are warranted. This could be initially determined using luciferase reporter assays analysing SOX3 and STAT3 activation via the putative enhancer region.

Our SOX3 ChIP-Seq was performed on whole P7 testes to obtain a genome-wide view of SOX3 binding. To refine these experiments, and determine where SOX3 is binding within specific cell types, ChIP could be performed on individual subpopulations of spermatogonial stages. Previous work has identified ways to individually sort spermatogonial stem cells from whole testis using fluorescence and magnetic activated cell sorting (FACS and MACS) (Valli et al. 2014). Unfortunately, there is not currently a way to sort the As, Apr, and Aal spermatogonia as they are difficult to differentiate between (Phillips et al. 2010). If a technique is developed, it

would be interesting to perform ChIP experiments on each developmental stage to determine which genes SOX3 is potentially regulating across spermatogonial development.

For each of the putative SOX3 enhancers identified within the ChIP-Seq experiments, no functional analysis could be performed due to time constraints. This is an important next step that will define whether these are true regulatory regions active within the postnatal testes. The best way to assess this will be to generate *in vivo* deletions using CRISPR and assess the impact on target gene expression as well as identifying if there are any spermatogenic defects present. For the *Neurog3* enhancer, mutation of the SOX3 TFBS as well as the larger enhancer region will show the endogenous contribution of both SOX3, and the larger enhancer.

No SOX motif could be identified for many of the histone associated genes, including H3t and TH2B. It is possible that SOX3 is instead binding to another protein, itself which is bound to the DNA. If this is occurring, no TFBS motif would be identified via ChIP. A way to confirm this could be to perform ChIP combined with Mass Spectrometry (Bensaddek & Lamond 2016). This allows the identification of protein:protein interactions occurring with the protein of interest (ie SOX3) which can be performed within a specific contexts (postnatal testes tissue). SOX3 is known to partner with other factors to regulate gene expression, such as POU at the *Nes* enhancer (Tanaka et al. 2004), however protein:protein interactions at enhancers have not yet been shown.

If a putative partner factor is identified throughout the further analysis, proximity ligation (PLA) assays will be a useful tool. This an immunohistochemistry based assay for tissue samples which relies on two proteins being in close proximity. When both

are bound by antibodies conjugated to oligonucleotides, they are able to connect, forming circular DNA. This DNA can then be amplified by a DNA polymerase and subsequently visualised by fluorescence signals (Bellucci et al. 2014). At many of the SOX3 bound DNA regions an SP motif was also identified. Although these were not a consistent distance apart, or connected, as would be expected for partner protein binding, it is possible that they both act at some of the same enhancers to control expression of target genes. SP proteins are expressed throughout many body systems to regulate gene expression, but they are also important regulators during sperm development (Lania et al. 1997). Gene set enrichment analysis performed on different stages of spermatogenesis identified SP1 as being one of the key regulators in spermatogonial stem cells, spermatocytes and spermatids (Zhu et al. 2016). From this, it is predicted that of the SP1 is the most likely SP factor to be bound to the SP motif. By using the PLA approach on testes tissue with SOX3 and SP1 antibodies it is possible that some regions may show an interaction between the two proteins. The downside of this assay is that the specific locations this is occurring is not able to be identified, however this would be informative for any further studies.

Although we identified putative enhancers bound by SOX3 from the ChIP-Seq data, most of the proposed target genes did not appear dysregulated when overlapped with SOX3 null microarray data. This may be due to the different timepoints being analysed, or due to subtle expression differences that may only be seen in a single cell type, such as the SSCs which only account for 0.03% of the germ cells in rodent testes (Tegelenbosch & de Rooij 1993; Phillips et al. 2010). The recent generation of a single cell RNA-Seq dataset encompassing each stage and cell type of spermatogenesis will be a useful resource for future ChIP studies (Chen et al. 2018). Extending this to RNA-Seq of SOX null testes would show the impact of SOX3 at a much greater level of detail than we currently have. Using this data in companion

with the SOX3 bound regions in the testes will help to decipher functional targets before generating *in vivo* enhancer deletions.

Concluding Remarks

This thesis focused on the SOXB1 proteins throughout different developmental stages, within embryonic CNS development as well as the postnatal development of the testes.

Chapter 2 showed the functional analysis and endogenous function of the previously described *Nes* enhancer during embryogenesis. We were able to show this is able to contribute for up to 70% of *Nes* expression when active in the CNS. The third chapter both identified and validated an enhancer of the Wnt-receptor protein *Fzd3*. Through both CRISPR mouse models and transgenic LacZ models we were able to identify its activity within the neural tube of the developing embryo. The final chapter of this thesis focussed on the role of SOX3 within the postnatal testes. Through bioinformatic analysis of a SOX3 ChIP-Seq dataset we have been able to identify putative enhancers thought to be involved in the progression of spermatogenesis. We also provide evidence that SOX3 is involved in the complex chromatin organisation via regulation the of histone genes.

CRISPR modification of the genome was an important aspect of these studies. As the bioinformatic identification of regulatory elements grows, it is important these are functionally validated. We have shown methods on how to do this, and incorporated with traditional techniques, this will help to elucidate the role of many enhancers within the genome.

Appendix One

CRISPR Genome Editing in Mice

Ella Thomson, Ruby Dawson, James Hughes & Paul Q Thomas

Book Chapter

Genome Editing and Engineering, From TALENs, ZFNs and CRISPRs to Molecular Surgery

Edited by Krishnarao Appasani

Published by Cambridge Press 2018

11 CRISPR Genome Editing in Mice

Ella Paulina Thomson,[1] Ruby Emily Dawson,[1] James Nicholas Hughes and Paul Quinton Thomas

11.1 Introduction

Mice are one of the most widely used model organisms for dissecting mammalian gene function and human disease modeling. Targeted mutagenesis in mice has traditionally been performed using "gene targeting," in which mutations of interest are initially generated in embryonic stem (ES) cells by homologous recombination, which are then used to produce mutant mouse lines (Griep *et al.*, 2011). While this technique has been widely used, it is lengthy (1–2 years are often required to produce a mutant), labor-intensive, expensive and prone to failure. More recently, genome editing technologies such as zinc-finger nucleases (ZFNs) (Carbery *et al.*, 2010) and transcription activator-like effector nucleases (TALENs) (Davies *et al.*, 2013) have been successfully used to generate mutant mice via zygotic injection, although high expense and relatively cumbersome production methods have limited their use.

A paradigm shift in the field of genome editing occurred in 2013 when a group of landmark papers were published demonstrating that the clustered regularly interspaced short palindromic repeats (CRISPR) bacterial defense mechanism could be used to introduce specific mutations at precise locations within eukaryotic cells (Cong *et al.*, 2013; Jinek *et al.*, 2012; Mali *et al.*, 2013). This sparked enormous interest from the scientific community, and soon many laboratories around the world were using this system to modify a diverse variety of organisms, including plants (Xing *et al.*, 2014), yeast (DiCarlo *et al.*, 2013), zebrafish (Hwang *et al.*, 2013), fruit flies (Basset *et al.*, 2013) and monkeys (Niu *et al.*, 2014), among others (Onuma *et al.*, 2016; Wang *et al.*, 2016). This chapter focuses on the development and application of CRISPR mutagenesis in mice.

Over the past few years, CRISPR has been used extensively by our research group, allowing the generation of a variety of mutations in mice, including floxed alleles, null alleles, enhancer deletions, epitope FLAG-TAG insertions and promoter modifications. Based on our experience with this technique, we also provide insights into which strategies are the most efficient, as well as highlighting significant, but less well-known issues that can arise.

11.2 System Components and Activity

Establishing the mechanism by which the CRISPR system cleaves specific sequences in bacteria provided the groundwork for its rapid development as a genome editing tool in eukaryotes (Bolotin *et al.*, 2005; Brouns *et al.*, 2008; Cong *et al.*, 2013; Jinek *et al.*, 2012; Mali *et al.* 2013; Sorek *et al.*, 2013). The endonuclease in the CRISPR system is Cas9, which can be directed to bind and cleave specific genomic target sites through complementary base pairing of its single-guide RNA (sgRNA) to the DNA (Brouns *et al.*, 2008). By modifying the sequence of the sgRNA, Cas9 can be directed to precise locations within the genome, causing a double-strand break (DSB) in the DNA, the repair of which can be used to produce a variety of mutation types (Cong *et al.*, 2013). The flexibility and practicability of this approach provides a significant advantage over other genome editing platforms and has contributed to the massive uptake of CRISPR across the research community.

Despite its utility, it is important to note that the CRISPR–Cas9 system cannot be used to target every sequence in the genome due to the absolute requirement for a protospacer adjacent motif (PAM) sequence immediately 3′ to the targeting region of the sgRNA (Sternberg *et al.*, 2014). For the most widely used Cas9 variant (from *Streptococcus pyogenes*), the PAM corresponds to an NGG sequence, which on average will occur every 8 bp across the genome. Structural studies indicate that Cas9 binds directly to the PAM followed by DNA strand separation and "sampling" of the neighboring sequence for complementarity by the sgRNA (Sternberg *et al.*, 2014). Accordingly, mutation of the PAM will completely abolish cleavage by the Cas9, regardless of whether or not the sgRNA is a perfect match to the adjacent sequence (Jinek *et al.*, 2012; Sternberg *et al.*, 2014).

The mechanism by which Cas9 cleaves DNA is reasonably well understood and is covered in some detail in other chapters. Briefly, Cas9 has two active nuclease sites: the RuvC domain at the amino terminus and the HNH domain in the central region (Nishimasu *et al.*, 2014). Both of these active sites generate site-specific nicks on opposing DNA strands to create a blunt-ended double-strand break 3–4 nucleotides upstream of the PAM site (Jinek *et al.*, 2012). Repair of the DSB by the cell's endogenous repair machinery can result in the generation of alleles with the desired functional consequence (e.g. loss of function). Unlike traditional gene targeting, CRISPR mutagenesis permits directed modification of the zygotic genome through direct injection of sgRNA(s) designed to target the location of interest along with Cas9 mRNA or protein. Surviving zygotes are then transferred into pseudopregnant females for development to term when they can be genotyped to identify mutations of interest. Thus, remarkably, CRISPR mutagenesis permits mutant founder mice to be generated within three weeks (summarized in Figure 11.1)

11.3 Repair Mechanisms

CRISPR mutagenesis, like all forms of genome editing, relies on exploitation of the cell's endogenous repair mechanisms that ordinarily operate to protect the cell

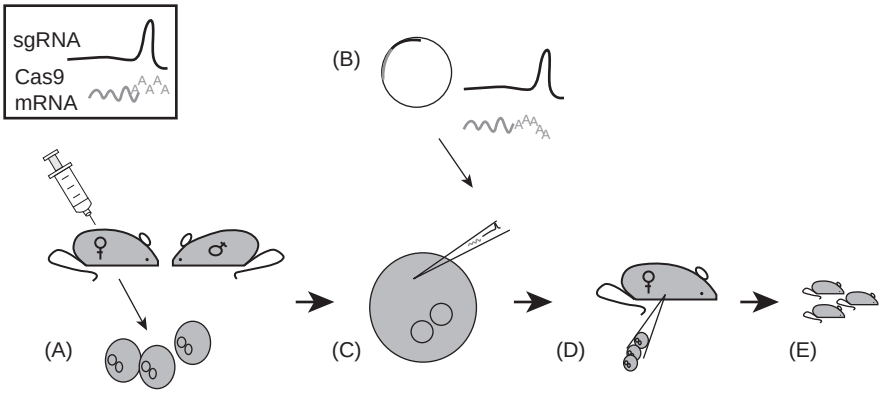


Figure 11.1 The process of using CRISPR in mice. (A) Female mice are superovulated by the injection of hormones, PMSG and HCG, and bred with males to obtain 20–30 zygotes per female. Cas9 mRNA and sgRNA are prepared through *in vitro* transcription from a plasmid template (B). CRISPR reagents are then microinjected into the cytoplasm or pronucleus of zygotes (C). Injected zygotes are transferred to pseudopregnant females that have been bred with vasectomized males (D). After three weeks gestation, mutant founder mice are born (E).

against DNA damage from sources such as replication error, fragile sites and ionizing radiation (Mehta and Haber, 2014). These repair mechanisms include canonical non-homologous end joining (cNHEJ), microhomology-mediated end joining (MMEJ), single-strand annealing (SSA) and homology-directed repair (HDR). Utilization of a particular repair mechanism is determined by a number of factors, such as cell cycle stage and flanking sequence, and, depending on the desired outcome, each of these can be exploited to generate specific types of mutations in mice, as outlined below and summarized in Figure 11.2 (Deriano and Roth, 2013).

11.3.1 Non-homologous End Joining

While cNHEJ repair is sensitive, rapid and precise, it can also result in mutations and can therefore be harnessed in CRISPR mutagenesis. This repair mechanism ligates and rejoins two blunt DNA ends without the use of a homologous template and with limited resection or restoration of nucleotides around the break (Davis and Chen, 2013; Dueva and Iliakis, 2013; Lieber, 1999; Matsuzaki *et al.*, 2012). The commonly held belief among users of CRISPR technology that NHEJ is highly error-prone is unlikely to be true (Bétermier *et al.*, 2014). This misconception arises in part because of the difficulty in measuring the fidelity of repair and is exacerbated by the fact that creation of DSBs and their repair will inevitably end only when a mistake is made. For example, in the context of CRISPR mutagenesis, precise repair will result in restoration of the target sequence, meaning that sequential rounds of cutting and repair will continue until misrepair destroys either the sgRNA-binding sequence or the PAM (Bétermier *et al.*, 2014). For this reason, CRISPR mutagenesis in the absence of a repair template invariably results in small insertions or deletions of nucleotides (indels) at the cut site. This property has been used extensively to generate frame-shifting loss-of-function mutations (Carbery *et al.*, 2010; Shen *et al.*, 2013).

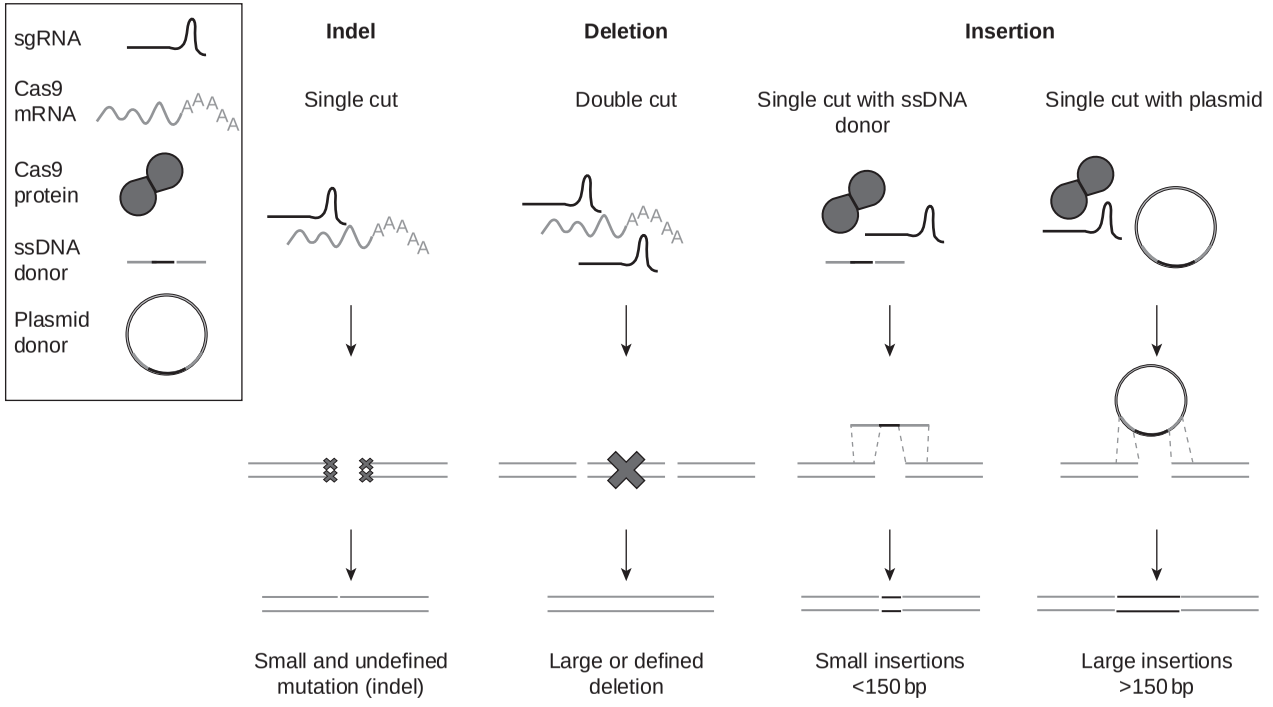


Figure 11.2 CRISPR can be used to generate a range of mutations in mice. Depending on the desired mutation, different CRISPR reagents are used to create indels and defined deletions through NHEJ or insertions through HDR. (A black-and-white version of this figure will appear in some formats. For the color version, please refer to the plate section.)

11.3.2 MMEJ and SSA (Alternative NHEJ)

Microhomology-mediated end joining (MMEJ) and single-strand annealing (SSA), which are subsets of the alternative NHEJ pathway (aNHEJ), are inherently error-prone, unlike cNHEJ (Dueva and Illakis, 2013; Simsek and Jasin 2010). Using enzymes distinct from cNHEJ, aNHEJ pathways repair DSBs by catalyzing the joining of DNA ends after resection, revealing small regions of homology on the two strands (Chiruvella *et al.* 2013; Crespan *et al.* 2012). Base pairing of the homologous regions is followed by trimming, extension and ligation to complete the repair, which has the effect of generating defined small deletions to the sequence (Crespan *et al.*, 2012; McVey and Lee, 2008; Roth and Wilson, 1986). SSA involves the annealing of larger repeat sequences, whereas MMEJ involves much shorter sequences of homology (5–25 bp; Decottignies, 2013). Because MMEJ and SSA rely on complementary base pairing of two homologous regions, the resultant deletions are predictable, allowing them to be used to our advantage in CRISPR genome editing.

11.3.3 Homology-directed Repair

Homology-directed repair (HDR) operates in competition with NHEJ to repair DSBs. The main difference between these mechanisms is that HDR relies on a repair template that is used to restore resected sequences at the break site and therefore usually results in precise repair (Jasin and Rothstein, 2013). Cells can undergo HDR when a homologous sequence of DNA is present, usually the homologous chromosome or sister chromatid (Deriano and Roth, 2013; Heyer *et al.*, 2010; Mao *et al.*, 2009). Importantly, exogenous DNA that is introduced to the cell can be copied into the genome to introduce specific alterations using this repair, by incorporating flanking regions of homology, which is a method that has historically been used in traditional gene-targeting methods. However the efficiency of this approach is very low, as millions of ES cells are typically needed in order to select for these rare recombination events (Hall *et al.*, 2009). However, when Cas9 is used to create the DSB, a remarkable increase in efficiency of template incorporation is seen, making it possible to directly modify mouse zygotes and avoid the use of ES cells entirely (Yang *et al.*, 2014).

It should be noted that all of the discussed repair mechanisms work in competition with each other, and greater understanding of these pathways should lead to an improved ability to control the type of repair used and hence the efficiency of a desired mutational outcome.

11.4 Harnessing Endogenous Repair Mechanism for Desired Mutational Outcomes

11.4.1 Single-cut Gene Disruption

The simplest way to utilize CRISPR in model organisms is to generate non-functional copies of genes following indel mutations caused by imprecise NHEJ, MMEJ or SSA. When the mutation causes a frameshift, this can result in a null allele due to loss of functional protein domain(s) (Isken and Maquat, 2007). Being able to

create null mice in an efficient and reliable way is especially useful when none are commercially available on the background of choice (Barbaric *et al.*, 2007).

In 2013, Shen *et al.* published the first mutations to disrupt gene function in mice using a single-guide CRISPR strategy when they targeted a GFP transgene. They were successful in detecting mutations in 2 out of 12 founder mice. Since this study, many examples of single-cut indel mutants have been published, with mutations being found in almost all alleles (Qin *et al.*, 2015; Yasue *et al.*, 2014). In addition, Wang and colleagues (2013) and subsequent studies have demonstrated that this approach can be multiplexed to simultaneously generate mutations in multiple genes using a single zygotic injection. The utility of this approach was nicely demonstrated in a recent study of acute myeloid leukemia, in which different combinations of genes were inactivated simultaneously to allow for more accurate disease modeling (Heckl *et al.*, 2014). The use of CRISPR to generate null mice is now routine in research and is already becoming a leading genome editing technique (Singh *et al.*, 2015).

In our experience, generating single-cut null (frameshift) mice has been relatively simple. We use a strategy targeting an early exon to ensure incorporation of a stop codon soon after the translation start site. Although we have been successful in generating null mouse lines using the single-cut method, one of the disadvantages of this approach is the generation of in-frame deletions that occur in approximately one-third of mutated alleles, which can allow the resultant protein to retain significant function. In an attempt to overcome this problem, we have since moved away from this approach to a double-cut method, which also has the added benefit of easy genotyping (discussed in detail below).

11.4.2 Genotyping Single-cut Mutants

Given that single-cut mutations are generally small insertions or small deletions, they must be screened using a method that does not rely on discrimination based on mutation size. Assays that involve heteroduplex formation can be used to detect small changes to the sequence. Amplification around the mutated region will generate two or more PCR products with at least one base mismatch, which, after melting and annealing, will form DNA heteroduplexes. These heteroduplexes can either be visualized on a polyacrylamide gel, as they migrate much more slowly than homoduplexes, or they can be detected through restriction fragment length polymorphism (RFLP) in a T7E1 assay, where heteroduplexes are cleaved to form two distinct products (Shen *et al.*, 2013). These mutation screening techniques allow specific samples to be selected for sequence analysis.

Sequencing chromatograms from founder mice frequently contain overlapping distinct sequence reads from the cleavage site due to the presence of different mutations on multiple alleles. These can sometimes be difficult to unravel, particularly if more than two alleles are present as a result of Cas9 endonuclease activity post-zygotically. However, if necessary, mutant alleles can be discriminated by cloning of individual sequences. When the size difference between the two products is sufficient, the easiest method is to gel purify the individual alleles

and sequence separately. Importantly, we have found that standard PCR screening using products of 200–400 bp will often miss alleles that result from larger deletions due to loss of the primer binding sequences. For this reason, it is important to out-cross founders before starting any experimental work to ensure that only the allele(s) of interest is present in the mutant line.

11.4.3 Double-cut Deletions

While frameshift mutations are very useful and relatively easy to make, the size and functional impact of deletions generated by a single cut can be difficult to predict. To overcome this issue, a double-cut strategy can be employed to generate deletions across a predetermined genomic region. The double-cut approach uses two sgRNAs flanking the interval of interest, and, after generation of the DSBs, rejoining of the blunt ends through NHEJ results in deletion of the intervening sequence (Fujii *et al.*, 2013; Mali *et al.*, 2013). Mice with specific deletions up to 10 kb were first reported by Fujii *et al.* (2013) and since, even larger deletions (65 kb) have been generated using this approach (Zhang *et al.*, 2015). Our group has adopted a double-cut strategy to delete enhancer regions, exons (around 200 bp) and whole genes (up to 9 kb). Exon deletions are now our favored approach to generate non-functional alleles by targeting cuts flanking a relatively short exon early in the transcript that contains a number of nucleotides not divisible by three. This approach can also be extended to other untranslated regions of the genome such as intronic and intergenic regions, promoters and UTRs.

11.4.4 Genotyping Double-cut Mutants

When there is a larger deletion or insertion being made, it is relatively easy to screen for the desired mutation using a PCR-based method, where a size difference will be apparent on an agarose gel. However, it is important to remember that whilst this is an efficient process, the DBS is not always at the precise predicted location and therefore can be slightly larger than intended due to NHEJ-mediated imperfect repair. For this reason, it is recommended to sequence all potential founders to determine the exact mutation.

The importance of screening for regions of homology or microhomology when designing the sgRNAs was emphasized to us when we observed a high frequency of identical small deletions in multiple mice generated from cleavage in a region of microhomology. This phenomenon has been reported previously in the literature in the context of CRISPR (Wang *et al.*, 2013) and indeed other gene editing reagents, such as TALENs (Tesson *et al.*, 2011). Microhomology can be used to predict precise deletion events, which, in theory, could be used to the advantage of the researcher (Bae *et al.*, 2014). However, these mechanisms can also result in a nasty surprise after mice have been generated, if repeat sequences are not identified during sgRNA design. Widely used current design tools do not always take this into account, but there are online tools available which can be used to screen guides for microhomology independently (Bae *et al.*, 2014; Haeussler *et al.*, 2016).

11.4.5 Point Mutations and Small Insertions

By using HDR-mediated CRISPR mutagenesis, it is now possible to introduce specific point mutations and small insertions such as epitope tags (e.g. HA-FLAG) relatively cheaply and rapidly compared to traditional methods (Wang *et al.*, 2013). These mutations are most easily achieved using a single-stranded oligonucleotide (ssDNA) donor that contains regions of homology (50–80 bp) to the endogenous locus flanking the donor sequence to be inserted at the cut site (Lee and Lloyd, 2014; Nakagawa *et al.*, 2016; Wang *et al.*, 2015; Yang *et al.*, 2014). There is ongoing optimization of the HDR-mediated CRISPR method to increase its low efficiency, including changes to homology arm length and silencing or inhibition of genes involved in NHEJ (Chu *et al.*, 2015; Maruyama *et al.*, 2015; Nakagawa *et al.*, 2016; Vartak and Raghavan, 2015). Substitution of a single base can also be used to generate silent mutations, which can destroy the relevant PAM site to prevent recutting after donor sequence insertion. From a practical standpoint we have found it useful to either destroy or introduce a restriction site as a silent mutation to aid in genotyping of founders.

Although we have been successful in generating 12 mouse lines with a variety of HDR-mediated insertions, in all cases NHEJ repair predominated, resulting in a relatively low efficiency of the desired insertion. In an attempt to increase the frequency of NHEJ repair events, we trialled co-injection of the DNA ligase IV inhibitor SCR7, which was previously shown to repress NHEJ and increase HDR efficiency (Chu *et al.*, 2015; Maruyama *et al.*, 2015; Vartak and Raghavan, 2015), although it did not increase insertion rates in our hands. Recently, it has been hypothesized in the literature that low HDR efficiency may be due to the donor oligonucleotides being degraded by the cell prior to the Cas9 mRNA being translated into functional protein (Ménoret *et al.*, 2015). We have found that injection of Cas9 protein rather than the mRNA directly into the zygote increases insertion efficiency as previously shown in the literature (Ménoret *et al.*, 2015). For the same reason, using phosphorothioate-modified oligos has also proven to be a useful approach to promote HDR insertion (Renaud *et al.*, 2016).

11.4.6 Generating Reporter and Floxed Conditional Mouse Lines

Using CRISPR to create more complex mutations, such as reporter gene insertions and floxed alleles, is possible, but generally much less efficient than the deletions and HDR-mediated small insertions described above. However, these modifications are becoming easier to generate using emerging CRISPR technologies, although strategies for longer or more complex insertions such as reporter genes (e.g. EGFP and LacZ) or two LoxP sequences are still being optimized. Whilst single-stranded oligonucleotides are an inexpensive and accessible option to insert small sequences into the mouse genome, often a much larger insert is required (Mali *et al.*, 2013; Yang *et al.*, 2013; 2014). Inserting longer sequences can be achieved using a plasmid donor template containing the insertion sequence flanked by large homology arms greater than 800 bp (Zhu *et al.*, 2015). This approach is similar in concept to targeting vectors that have been used in ES cells for over 30 years, although there is no requirement for a drug selection

cassette for zygotic injections. To generate a large insertion, sgRNAs targeting the genomic region between the homology arms are injected along with the plasmid donor template (in linearized or circular form). The long homology arms provided by the donor template facilitate repair of the DSB via homologous recombination, thereby resulting in targeted insertion of the intervening sequence. It is important to note that the efficiency of this strategy is relatively low (8–10%) (Nakagawa *et al.*, 2016; Yang *et al.*, 2013). We have used this approach to generate targeted insertions of 750 bp–4 kb at three loci and, consistent with published reports, our insertion rate is relatively low (2.5–12.5%).

Floxed conditional alleles are probably the most challenging type of mutation to generate. Several alternative approaches have been used successfully, although all remain relatively inefficient. The first published approach used an oligonucleotide HDR repair strategy to simultaneously insert LoxP sequences either side of an exon of interest (Bishop *et al.*, 2016; Yang *et al.*, 2013). Using this strategy, we recently generated a floxed allele containing LoxP sites spaced 2 kb apart through co-injection of gRNAs and single-stranded oligonucleotides repair templates with 60 bp homology arms. Of the 11 founder pups born, only one had both the LoxP sequences inserted on the same allele, while the majority of the cut sites generated indels or single LoxP insertions and one case of deletion of the 2 kb region between the cuts. This relatively low efficiency is consistent with reports from the literature (Maruyama *et al.*, 2015; Nakagawa *et al.*, 2016; Yang *et al.*, 2013). The alternative approach utilizes a dsDNA (plasmid) repair template (identical in sequence to the intended floxed allele) that includes long homology arms (typically >1 kb). The floxed allele is generated by sgRNA-mediated cutting at one or both of the LoxP insertion sites, followed by homologous recombination repair. This approach has been used to insert LoxP sites separated by 71–600 bp albeit with low efficiency (2.7–16%) (Maruyama *et al.*, 2015; Yang *et al.*, 2013). A dual sgRNA Cas9-nickase approach has also been successfully used but with similarly low efficiency (2.7%) (Ran *et al.*, 2013; see below).

11.5 The CRISPR Toolbox

Since 2013, when the first mutant mouse lines were generated using the CRISPR system (Shen *et al.*, 2013; Wang *et al.*, 2013), the number of tools available to researchers has rapidly expanded to include modified versions of the wild-type *S. pyogenes* Cas9 (wt *Sp*Cas9) with altered PAM specificity, Cas9 proteins from a range of bacterial species and new programmable endonuclease enzymes (summarized in Table 11.1). This section focuses on the ever-increasing "CRISPR Toolbox" and some of the ways these new enzymes can be implemented to enable increased flexibility of target site selection and reduce off-target cutting.

11.5.1 Off-target Reduction

One of the enduring concerns of CRISPR mutagenesis is the occurrence of off-target endonuclease activity. Whilst breeding is a simple way to eliminate these from a mutant line, it can be time-consuming and expensive to backcross to be

Table 11.1 A summary of endonucleases that can be used in CRISPR genome editing

Name	Species	Engineered property	Endonuclease	PAM	Cells	Mice	Reference
SpCas9	*S. pyogenes*	N/A	Cas9	NGG	Yes	Yes	Cong *et al.*, 2013; Shen *et al.*, 2013
SpCas9n	*S. pyogenes*	Off-target minimization	Cas9	NGG	Yes	Yes	Ran *et al.*, 2013
SpCas9 VQR	*S. pyogenes*	PAM variant	Cas9	NGAN/NGNG	Yes	No	Kleinstiver *et al.*, 2015
SpCas9 EQR	*S. pyogenes*	PAM variant	Cas9	NGAG	Yes	No	Kleinstiver *et al.*, 2015
SpCas9 VRER	*S. pyogenes*	PAM variant	Cas9	NGCG	Yes	No	Kleinstiver *et al.*, 2015
eSpCas9	*S. pyogenes*	High fidelity	Cas9	NGG	Yes	No	Slaymaker *et al.*, 2016
SpCas9-HF1	*S. pyogenes*	High fidelity	Cas9	NGG	Yes	No	Kleinstiver *et al.*, 2016
SaCas9	*S. aureus*	N/A	Cas9	NGRRT/NGRRN	Yes	Yes	Ran *et al.*, 2013
KKH SaCas9	*S. aureus*	PAM variant	Cas9	NNNRRT	Yes	No	Kleinstiver *et al.*, 2015
AsCpf1	*Acidaminococcus* sp.	N/A	Cpf1	TTN	Yes	Yes	Zetsche *et al.*, 2015; Kim *et al.*, 2016
LbCpf1	*Lachnospiraceae*	N/A	Cpf1	TTN	Yes	Yes	Zetsche *et al.*, 2015; Kim *et al.*, 2016
dCas9-Fok1	*S. pyogenes*	Off-target minimization	FokI	NGG	Yes	Yes	Guilinger *et al.*, 2014; Hara *et al.*, 2015
NmCas9	*N. meningitidis*	PAM variant	Cas9	NNNNGATT	Yes	No	Lee *et al.*, 2016
StCas9	*S. thermophilus*	PAM variant	Cas9	NNAGAAW	Yes	No	Müller *et al.*, 2016

sure all off-targets have been segregated away. Targeted resequencing of potential off-target sites in founder mice indicates that sgRNA binding sites that differ by three or more nucleotides from the on-target sequence are very unlikely to be mutated (Haeussler *et al.*, 2016). Various online tools have been developed to predict where the most likely off-target cuts will occur and guide sequences can often be designed to avoid those within coding sequences (Haeussler *et al.*, 2016). In cases where there is limited flexibility in sgRNA targeting (e.g. generating a point mutation via HDR-mediated repair), potentially problematic off-target sgRNAs can now be paired with modified endonucleases that are able to minimize this occurrence (discussed below).

One of the first published solutions to off-target mutagenesis was the "dual nickase" strategy (Ran *et al.*, 2013). This approach uses a modified Cas9 protein containing an inactivating mutation in the RuvC domain that functions as a nickase. To create a mutation at the site of interest, a pair of sgRNAs (spaced by <100 bp) are used to nick opposite strands to create a dsDNA with a large 5′ overhang (Ran *et al.*, 2013). Through repair of this staggered break it is possible to generate indels or localized targeted modifications such as point mutations via HDR using a ssDNA repair template. Off-target nicks generated by a single sgRNA/Cas9 complex will be repaired efficiently by the high fidelity base excision repair method (Dianov and Hübscher, 2013), greatly minimizing the chances of unwanted mutations elsewhere in the genome. The dual nickase approach has been shown to be efficient in both cell lines and mouse models without any apparent reduction in on-target DNA cutting (Cheng *et al.*, 2014; Ran *et al.*, 2013).

Off-target cleavage has also been addressed by fusing an endonuclease "dead" Cas9 protein (dCas9) with a dimer-dependent FokI endonuclease domain to create fCas9. The FokI endonuclease domain has been used extensively in zinc-finger nuclease and TALEN genome editing, where pairs of DNA-binding proteins bind to adjacent sequences to enable generation of a double-strand break at the target site. A similar principle is used in the context of fCas9; two sgRNAs are designed to target regions 15–20 bp apart, allowing the two FokI modules to create the DSB in the intervening "spacer" region (Guilinger *et al.*, 2014). In human cell lines this approach was shown to have 140-fold higher specificity for target sites compared to wt *Sp*Cas9 (Guilinger *et al.*, 2014), and this has also been shown to be an effective strategy for use in mouse models, with similar efficiencies to Cas9 being reported (Hara *et al.*, 2015). However, a disadvantage of this approach is that regions that are able to be targeted via this method occur on average once every 34 bp, compared to every 8 bp for a single sgRNA strategy with wt *Sp*Cas9 (Guilinger *et al.*, 2014).

Structure-guided rational engineering has also been used to increase the fidelity of Cas9. Two different versions of Cas9 recently published were eSpCas9 (Slaymaker *et al.*, 2016) and SpCas9 HF-1 (Kleinstiver *et al.*, 2016). It was shown that these mutants are able to disproportionately decrease off-target binding with respect to on-target binding and hence improve specificity without compromising on-target activity in human cell lines, although this has not yet been tested in a mouse model.

11.5.2 PAM Variants

PAM sites for wt SpCas9 (NGG) are relatively common throughout genomes, although there are regions where it can be difficult to find a suitable sgRNA, such as A-T-rich regions, or in regions of highly repetitive sequence. Identification of Cas9 proteins that bind to different PAM sequences will therefore provide increased flexibility in target site identification.

To generate SpCas9 enzymes that recognize alternative PAM sequences, Kleinstiver *et al*. (2015) used a directed evolution approach targeting residues within the PAM-recognizing region of the Cas9 enzyme. This approach yielded a number of Cas9 mutants with altered PAM specificities: SpCas9 VQR, EQR and VRER, which recognize NGAN/NGNG, NGAG and NGCG, respectively. As with other Cas9 mutants, these have been tested in cell lines, but have not yet been validated in mouse zygotes.

CRISPR/Cas9 defense systems are widely used in bacteria, providing considerable scope for the development of new genome editing platforms. Inspired by the success of CRISPR/Cas9 from *Streptococcus pyogenes*, Cas9 orthologs from other bacterial species including *Staphylococcus aureus, Niessera meningiditis* (Lee *et al.*, 2016) and *Streptococcus thermophilus* (Kleinstiver *et al.*, 2015; Müller *et al.*, 2016) have recently been published, each of which utilizes a different PAM sequence. So far, only the SaCas9 has been shown to work successfully in mice, although the other strains are also likely to prove effective agents for generation of mutant mice.

11.5.3 Non-Cas9 Endonuclease

Whilst Cas9 is the pioneer endonuclease used in mammalian CRISPR applications, a newly recognized endonuclease known as Cpf1 has similar properties, and can also be harnessed for genome engineering (Zetsche *et al.*, 2015). By analyzing the CRISPR systems in hundreds of bacteria to determine if other enzymes are also capable of cutting DNA, Cpf1 was found in the *Acidaminococcus* and *Lachnospiraceae* families, and has been shown to be successful at creating mutations in human cell lines and mouse models (Hur *et al.*, 2016; Kim *et al.*, 2016).

Although they both work as part of a CRISPR bacterial defense system, Cas9 and Cpf1 have different modes of action. Specifically, Cpf1 cleaves DNA asymmetrically, such that rather than forming blunt ends, sticky ends are formed, creating a 4 or 5 bp 5′ overhang (Zetsche *et al.*, 2015). In addition, Cpf1 makes a staggered cut, 23 bp upstream from the PAM on the non-targeted strand and 18 bp on the targeted strand, rather than 3 bp upstream of the PAM site (Zetsche *et al.*, 2015). Due to the distance between the PAM site and the staggered DSB, repeated cleavage of the repaired allele can readily occur if the gRNA binding or PAM sequence is not altered. While this is tolerable for simple deletion strategies, it is essential that donor fragment integration includes a PAM-destroying silent mutation to eliminate the possibility of recutting after correct insertion.

In 2016 it was reported that the argonaute proteins from the bacterial species *Natronobacterium gregoryi* (NgAgo) could be used as ssDNA-guided genome editing agents in human cell lines (Gao *et al.*, 2016). The NgAgo protein was reported to

function without the requirement for a PAM sequence, thereby bypassing one of the key limitations of existing CRISPR genome editing platforms. However, whilst this technology appeared to function efficiently in cell lines, replication studies have not yet been published. Indeed, concerns have been expressed about reproducibility of the published data. We recently compared the mutagenesis efficiency of wt SpCas9 and NgAgo in mouse zygotes (n = 29) and mouse ES cells and no evidence of mutations were detected in the NgAgo samples. Thus, the suitability of this platform for routine genome editing remains highly questionable.

11.6 Conclusion

The advent of CRISPR technology for genome editing has greatly improved the ease and affordability of creating mouse models for biological research. By harnessing the cell's endogenous repair mechanisms through this technology, researchers can now reliably and efficiently generate a range of mutations for a plethora of experimental applications. As the interest and use of CRISPR continues to escalate, it seems likely that new editing platforms and techniques will continue to be developed that will circumvent existing limitations, and inevitably increase the scope of research applications.

Acknowledgements

We would like to thank Louise Robertson and the members of the Thomas Laboratory for their help and willingness to share unpublished data.

NOTE

1. The authors contributed equally to the work.

REFERENCES

Bae S, Kweon J, Kim HS, Kim J-S. 2014. Microhomology-based choice of Cas9 nuclease target sites. *Nat Methods* **11**: 705–706.

Barbaric I, Miller G, Dear TN. 2007. Appearances can be deceiving: phenotypes of knockout mice. *Brief Funct Genomic Proteomic* **6**: 91–103.

Bassett AR, Tibbit C, Ponting CP, Liu J-L. 2013. Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. *Cell Rep* **4**: 220–228.

Bétermier M, Bertrand P, Lopez BS. 2014. Is non-homologous end-joining really an inherently error-prone process? *PLoS Genet* **10**(1): e1004086.

Bishop KA, Harrington A, Kouranova E, *et al*. 2016. CRISPR/Cas9-mediated insertion of loxP sites in the mouse dock7 gene provides an effective alternative to use of targeted embryonic stem cells. *G3 (Bethesda)* **6**: 2051–2061.

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**: 2551–2561.

Brouns SJJ, Jore, MM, Lundgren M, *et al*. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**: 960–964.

Carbery ID, Ji D, Harrington A, *et al.* 2010. Targeted genome modification in mice using zinc-finger nucleases. *Genetics* **186**: 451–459.

Cheng R, Peng J, Yan Y, *et al.* 2014. Efficient gene editing in adult mouse livers via adenoviral delivery of CRISPR/Cas9. *FEBS Lett* **588**: 3954–3958.

Chiruvella KK, Liang Z, Wilson TE. 2013. Repair of double-strand breaks by end joining. *Cold Spring Harb Perspect Biol* **5**(5): a012757.

Chu VT, Weber T, Wefers B, *et al.* 2015. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol* **33**: 543–548.

Cong L, Ran FA, Cox D, *et al.* 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819–823.

Crespan E, Czabany T, Maga G, Hübscher U. 2012. Microhomology-mediated DNA strand annealing and elongation by human DNA polymerases λ and β on normal and repetitive DNA sequences. *Nucleic Acids Res* **40**(12): 5577–5590.

Davies B, Davies G, Preece C, *et al.* 2013. Site specific mutation of the Zic2 locus by microinjection of TALEN mRNA in mouse CD1, C3H and C57BL/6J oocytes. *PLoS One* **8**: e60216.

Davis AJ, Chen DJ, 2013. DNA double strand break repair via non-homologous end-joining. *Transl Cancer Res* **2**: 130–143.

Decottignies A. 2013. Alternative end-joining mechanisms: a historical perspective. *Front Genet* **4**: 48.

Deriano L, Roth DB. 2013. Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. *Annu Rev Genet* **47**: 433–455.

Dianov GL, Hübscher, U. 2013. Mammalian base excision repair: the forgotten archangel. *Nucleic Acids Res* **41**: 3483–3490.

DiCarlo JE, Norville JE, Mali P, *et al.* 2013. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res* **41**: 4336–4343.

Dueva R, Iliakis G. 2013. Alternative pathways of non-homologous end joining (NHEJ) in genomic instability and cancer. *Transl Cancer Res* **2**: 163–177.

Fujii W, Kawasaki K, Sugiura K, Naito K. 2013. Efficient generation of large-scale genome-modified mice using gRNA and CAS9 endonuclease. *Nucleic Acids Res* **41**: e187.

Gao P, Yang H, Rajashankar KR, Huang Z, Patel DJ. 2016. Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Res* **26**: 901–913.

Griep AE, John MC, Ikeda S, Ikeda A. 2011. Gene targeting in the mouse. *Methods Mol Biol* **770**: 293–312.

Guilinger JP, Thompson DB, Liu DR. 2014. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol* **32**: 577–582.

Haeussler M, Schönig K, Eckert H. *et al.* 2016. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* **17**: 148.

Hall B, Limaye A, Kulkarni AB. 2009. Overview: generation of gene knockout mice. *Curr Protoc Cell Biol* Chapter: Unit-19.1217.

Hara S, Tamano M, Yamashita S, *et al.* 2015. Generation of mutant mice via the CRISPR/Cas9 system using FokI-dCas9. *Sci Rep* **5**: 11221.

Heckl D, Kowalczyk MS, Yudovich D, *et al.* 2014. Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nat Biotechnol* **32**: 941–946.

Heyer W-D, Ehmsen KT, Liu J. 2010. Regulation of homologous recombination in eukaryotes. *Annu Rev Genet* **44**: 113–139.

Hur JK, Kim K, Been KW, *et al.* 2016. Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins. *Nat Biotechnol* **34**(8): 807–808.

Hwang WY, Fu Y, Reyon D, *et al.* 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* **31**: 227–229.

Isken O, Maquat LE. 2007. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev* **21**: 1833–3856.

Jasin M, Rothstein R. 2013. Repair of strand breaks by homologous recombination. *Cold Spring Harb Perspect Biol* **5**: a012740.

Jinek M, Chylinski K, Fonfara I, *et al.* 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816–821.

Kim D, Kim J, Hur JK, *et al.* 2016. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat Biotechnol* **34**: 863–868.

Kleinstiver BP, Pattanayak V, Prew MS, *et al.* 2016. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**: 490–495.

Kleinstiver BP, Prew MS, Tsai SQ, *et al.* 2015. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**: 481–485.

Lee AY-F, Lloyd KCK. 2014. Conditional targeting of Ispd using paired Cas9 nickase and a single DNA template in mice. *FEBS Open Bio* **4**: 637–642.

Lee CM, Cradick TJ, Bao G. 2016. The *Neisseria meningitidis* CRISPR-Cas9 system enables specific genome editing in mammalian cells. *Mol Ther J Am Soc Gene Ther* **24**: 645–654.

Lieber MR. 1999. The biochemistry and biological significance of nonhomologous DNA end joining: an essential repair process in multicellular eukaryotes. *Genes Cells Devoted Mol Cell Mech* **4**: 77–85.

Mali P, Yang L, Esvelt KM, *et al.* 2013. RNA-guided human genome engineering via Cas9. *Science* **339**: 823–826.

Mao Z, Jiang Y, Liu X, Seluanov A, Gorbunova V. 2009. DNA repair by homologous recombination, but not by nonhomologous end joining, is elevated in breast cancer cells. *Neoplasia N Y N* **11**: 683–691.

Maruyama T, Dougan SK, Truttmann MC, *et al.* 2015. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol* **33**: 538–542.

Matsuzaki K, Terasawa M, Iwasaki D, Higashide M, Shinohara M. 2012. Cyclin-dependent kinase-dependent phosphorylation of Lif1 and Sae2 controls imprecise nonhomologous end joining accompanied by double-strand break resection. *Genes Cells Devoted Mol Cell Mech* **17**: 473–493.

McVey M, Lee SE. 2008. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet* **24**: 529–538.

Mehta A, Haber JE. 2014. Sources of DNA double-strand breaks and models of recombinational DNA repair. *Cold Spring Harb Perspect Biol* **6**: a016428.

Ménoret S, De Cian A, Tesson L, *et al.* 2015. Homology-directed repair in rodent zygotes using Cas9 and TALEN engineered proteins. *Sci Rep* **5**: 14410.

Müller M, Lee CM, Gasiunas G, *et al.* 2016. *Streptococcus thermophilus* CRISPR-Cas9 systems enable specific editing of the human genome. *Mol Ther J Am Soc Gene Ther* **24**: 636–644.

Nakagawa Y, Sakuma T, Nishimichi N, *et al.* 2016. Ultra-superovulation for the CRISPR-Cas9-mediated production of gene-knockout, single-amino-acid-substituted, and floxed mice. *Biol Open* **5**(8): 1142–1148.

Nishimasu H, Ran FA, Hsu PD, *et al.* 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**: 935–949.

Niu Y, Shen B, Cui Y, *et al.* 2014. Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell* **156**: 836–843.

Onuma A, Fujii W, Sugiura K, Naito K. 2016. Efficient mutagenesis by CRISPR/Cas system during meiotic maturation of porcine oocytes. *J Reprod Dev* **63**(1): 45–50.

Qin W, Dion SL, Kutny PM, *et al.* 2015. Efficient CRISPR/Cas9-mediated genome editing in mice by zygote electroporation of nuclease. *Genetics* **200**: 423–430.

Ran FA, Hsu PD, Lin C-Y, *et al.* 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**: 1380–1389.

Renaud J-B, Boix C, Charpentier M, *et al.* 2016. Improved genome editing efficiency and flexibility using modified oligonucleotides with TALEN and CRISPR-Cas9 nucleases. *Cell Rep* **14**: 2263–2272.

Roth DB, Wilson JH. 1986. Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. *Mol Cell Biol* **6**: 4295–4304.

Shen B, Zhang J, Wu H, *et al.* 2013. Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Res* **23**: 720–723.

Simsek D, Jasin M. 2010. Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. *Nat Struct Mol Biol* **17**: 410–416.

Singh P, Schimenti JC, Bolcun-Filas E. 2015. A mouse geneticist's practical guide to CRISPR applications. *Genetics* **199**: 1–15.

Slaymaker IM, Gao L, Zetsche B, *et al.* 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**: 84–88.

Sorek R, Lawrence CM, Wiedenheft B. 2013. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu Rev Biochem* **82**: 237–266.

Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**: 62–67.

Tesson L, Usal C, Ménoret S, *et al.* 2011. Knockout rats generated by embryo microinjection of TALENs. *Nat Biotechnol* **29**: 695–696.

Vartak SV, Raghavan SC. 2015. Inhibition of nonhomologous end joining to increase the specificity of CRISPR/Cas9 genome editing. *FEBS J* **282**: 4289–4294.

Wang B, Li K, Wang A, *et al.* 2015. Highly efficient CRISPR/HDR-mediated knock-in for mouse embryonic stem cells and zygotes. *BioTechniques* **59**: 201–202, 204, 206–208.

Wang H, Yang H, Shivalila CS, *et al.* 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**: 910–918.

Wang X, Cai B, Zhou J, *et al.* 2016. Disruption of FGF5 in cashmere goats using CRISPR/Cas9 results in more secondary hair follicles and longer fibers. *PLoS One* **11**: e0164640.

Xing H-L, Dong L, Wang Z-P, *et al.* 2014. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol* **14**: 327.

Yang H, Wang H, Jaenisch R. 2014. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nat Protoc* **9**: 1956–1968.

Yang H, Wang H, Shivalila CS, *et al.* 2013. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas mediated genome engineering. *Cell* **154**: 1370–1379.

Yasue A, Mitsui SN, Watanabe T, *et al.* 2014. Highly efficient targeted mutagenesis in one-cell mouse embryos mediated by the TALEN and CRISPR/Cas systems. *Sci Rep* **4**: 5705.

Zetsche B, Gootenberg JS, Abudayyeh OO, *et al.* 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**: 759–771.

Zhang L, Jia R, Palange NJ, *et al.* 2015. Large genomic fragment deletions and insertions in mouse using CRISPR/Cas9. *PLoS One* **10**: e0120396.

Zhu Z, Verma N, González F, Shi Z-D, Huangfu D. 2015. A CRISPR/Cas-mediated selection-free knockin strategy in human embryonic stem cells. *Stem Cell Rep* **4**: 1103–1111.

Appendix Two

Data from SOX3 ChIP-Seq experiments

1 Peak regions identified in 2 or nore ChIP samples, and their nearest gene

2 Top 5 GO Terms associated with the ChIP peak regions in 2 or more samples

- 2.1 Negative regulation of Megakaryocyte Differentiation
- 2.2 Nucleosome Assembly
- 2.3 Nucleosome Organisation
- 2.4 Chromatin Assembly
- 2.5 Chromatin Assembly or Disassembly

3 Regions identified as containing individual SOX motifs

- 3.1 arACAAAGwa
- 3.2 drAAACAATGka
- 3.3 awACAAAGwa

4 Regions containing atleast 1 SOX motif and their nearest genes

5 Overlapping peak regions between NPC and Testes SOX3 ChIP datasets

6 Peak regions only present within Testes ChIP (Testes Specific)

7 Top 5 GO Terms associated with the Testes Specific peaks

- 7.1 Nucleosome Organisation
- 7.2 Nucleosome Assembly
- 7.3 Chromatin Assembly or Dissassembly
- 7.4 Chromatin Assembly
- 7.5 Chromatin Organisation

1. Testes Peak Regions from 2 or More Samples

chromosome	start	stop	nearest gene (distance to TSS)
chr7	3000020	3000235	Gm7353 (+111822)
chr7	4999282	4999547	Zfp580 (-3719)
chr7	4999598	4999859	Zfp580 (-3405)
chr7	5543821	5543962	Vmn1r62 (-82337), Vmn1r60 (-47161)
chr7	6147041	6147413	Galp (-1465)
chr7	11787109	11787324	Zscan4d (-35720), Zscan4e (+108814)
chr7	16994559	16994776	Tmem160 (-43460), Zc3h4 (+8438)
chr7	27031275	27031464	Vmn1r184 (-20480), Cyp2b9 (+72940)
chr7	29827407	29827755	Map4k1 (+59708), Ryr1 (+82589)
chr7	29958939	29959245	Ggn (+3853), Psmd8 (+6600)
chr7	31306647	31307327	Arhgap33 (+13092), Prodh2 (+28346)
chr7	31307390	31308185	Arhgap33 (+12291), Prodh2 (+29147)
chr7	39231479	39233273	Pop4 (-176009), Gm5591 (+80836)
chr7	39352222	39353766	Gm5591 (-39782)
chr7	50122401	50122572	Gm2381 (+117)
chr7	50509190	50509325	EU599041 (+30313), Zfp715 (+57598)
chr7	52964901	52965056	Rpl18 (-5848), Dbp (+4521)
chr7	52965206	52965362	Rpl18 (-5543), Dbp (+4826)
chr7	52965417	52966222	Rpl18 (-5007), Dbp (+5362)
chr7	72747499	72747658	Tjp1 (-231454), Tarsl2 (-42205)
chr7	72750212	72750303	Tjp1 (-234133), Tarsl2 (-39526)
chr7	87853701	87853957	Crtc3 (-20066), Iqgap1 (+94388)
chr7	88903178	88903348	Fam103a1 (-4548)
chr7	90974226	90974809	1700026D08Rik (-31128), Mesdc1 (+58333)
chr7	91561750	91562035	Arnt2 (-3424)
chr7	95678989	95679332	Rab38 (+100378), Tmem135 (+808136)
chr7	96802485	96802569	A230065N10Rik (-20893), Ccdc81 (+249601)
chr7	107520727	107520893	Ppme1 (-404), C2cd3 (+67)
chr7	108458149	108458626	Atg16l2 (-7786), Stard10 (-7212)
chr7	108789576	108789755	Art2b (-59991), Clpb (-22481)
chr7	115820499	115820711	Olfr512 (-36229), Olfr510 (+9673)
chr7	117817849	117818128	Ampd3 (-93731), Adm (+46814)
chr7	118226353	118226835	Eif4g2 (-50)
chr7	120444291	120444522	Btbd10 (+68446), Arntl (+93428)
chr7	121290447	121290617	Rras2 (-29237), Copb1 (+107662)
chr7	122814005	122814257	Gm6816 (-742891), Sox6 (+324469)
chr7	124444366	124444948	Xylt1 (-79836), Nucb2 (+796769)
chr7	130559517	130559817	Aqp8 (-46141), Lcmt1 (+38171)
chr7	133056871	133057003	Gsg1l (+168988), D430042O09Rik (+205508)
chr7	134171142	134171368	Maz (-1262)
chr7	138559852	138560095	Hmx3 (-126407), Acadsb (+5859)
chr7	141150079	141150413	Fank1 (+181702), Adam12 (+266534)
chr7	148151421	148151685	Ifitm1 (-2386)
chr6	3150680	3151774	Samd9l (+198344)
chr6	8115351	8115691	Col28a1 (+10873), C1galt1 (+320679)

chr6	12106365	12106464	Gm6578 (-46832), Thsd7a (+592838)
chr6	17741408	17741763	St7 (+97554), Wnt2 (+238999)
chr6	17741787	17742080	St7 (+97902), Wnt2 (+238651)
chr6	22299685	22299898	Fam3c (+6374), Wnt16 (+61562)
chr6	30677243	30677546	Cep41 (-33646), Mest (-10668)
chr6	42445265	42445364	Olfr456 (-8102), Fam115e (+102056)
chr6	44742921	44743299	Cntnap2 (-266950)
chr6	47091238	47091832	Cul1 (-312784)
chr6	47110170	47110764	Cul1 (-293852)
chr6	50146919	50147262	Dfna5 (+66770), Mpp6 (+86851)
chr6	51419242	51419535	Hnrnpa2b1 (-1961), Cbx3 (-970)
chr6	51419690	51420373	Hnrnpa2b1 (-2604), Cbx3 (-327)
chr6	51936401	51936542	Skap2 (+26076), Snx10 (+462572)
chr6	54927446	54927856	Nod1 (-5045), Ggct (+15293)
chr6	54927977	54928346	Nod1 (-5556), Ggct (+14782)
chr6	58767579	58767663	Herc3 (-16073), Abcg2 (+220956)
chr6	58775563	58775687	Herc3 (-8069), Abcg2 (+228960)
chr6	66579484	66579697	Vmn1r33 (-13535), Vmn1r34 (+8155)
chr6	68200903	68201280	Tacstd2 (-715276), Igkv4-71 (+992347)
chr6	79075351	79075546	Lrrtm4 (-893422), Reg3g (-656583)
chr6	83417694	83418110	Tet3 (-26233), Dguok (+39061)
chr6	83530301	83530768	Stambp (-8037), Clec4f (+75646)
chr6	87931409	87931718	Gm5577 (-113), H1fx (-88)
chr6	88444425	88445476	Sec61a1 (+23948), Ruvbl1 (+29548)
chr6	89950379	89950539	Vmn1r47 (-21423), Vmn1r46 (+24294)
chr6	92356089	92356462	Trh (-161632), Prickle2 (+161084)
chr6	103599005	103599368	Cntn6 (-843729), Chl1 (+138317)
chr6	116201111	116202043	Zfand4 (-12663), Fam21 (+43526)
chr6	121600081	121600259	Mug1 (-188389), A2m (+13979)
chr6	126489214	126489669	Kcna5 (-3869)
chr6	132413674	132413881	Prh1 (-106049), Gm8882 (-99626)
chr6	143101173	143101323	C2cd5 (-52621), Etnk1 (-14502)
chr6	145881183	145881640	Sspn (-1230)
chr6	146525985	146526167	Fgfr1op2 (-305), Asun (+281)
chr6	147501623	147501803	Far2 (-494225), Ccdc91 (+77320)
chr5	11220304	11220562	4933402N22Rik (-374523), Gm5861 (+37361)
chr5	11220573	11220751	4933402N22Rik (-374294), Gm5861 (+37590)
chr5	11400416	11400736	4933402N22Rik (-194380), Gm5861 (+217504)
chr5	11829495	11829872	Sema3d (-553482), 4933402N22Rik (+234728)
chr5	11881170	11881428	Sema3d (-501867), 4933402N22Rik (+286343)
chr5	14918464	14922081	Gm9758 (-5374), Speer4e (+18156)
chr5	16889535	16889710	Speer4f (-92293), Gm3495 (+10597)
chr5	16974433	16975269	Speer4f (-7065), Gm3495 (+95825)
chr5	28661168	28661314	Evc2 (-9068497), Rbm33 (+17512), Shh (+132400)
chr5	54893762	54893923	Stim2 (+504187)
chr5	66330250	66330572	Chrna9 (-27952), Rhoh (+75603)
chr5	66375177	66375388	Chrna9 (+16920), 9130230L23Rik (+20241)
chr5	72438606	72438857	Commd8 (+120690), Gabrb1 (+347561)

chr5	76651154	76651299	Tmem165 (+38322), Clock (+82590)
chr5	76657257	76657453	Tmem165 (+44450), Clock (+76462)
chr5	78845888	78845995	NONE
chr5	102235572	102235858	Cds1 (+41566), Wdfy3 (+263225)
chr5	105855709	105855904	Lrrc8c (-92600), Lrrc8b (+11013)
chr5	108295958	108296386	Evi5 (+7954), Ube2d2b (+36900)
chr5	109111463	109111867	Fgfrl1 (-11583), Slc26a1 (-7077)
chr5	109176965	109177370	Rnf212 (+1299), Fgfrl1 (+53920)
chr5	109231240	109231381	Tmed11 (-6929), Vmn2r8 (+6462)
chr5	109722239	109722366	Vmn2r14 (-68662), Vmn2r15 (+4272)
chr5	113763814	113764285	Sgsm1 (-24244), Aym1 (-22264)
chr5	115939382	115939805	Sirt4 (-4860)
chr5	115940244	115940408	Pxn (-16491), Sirt4 (-5592)
chr5	123453058	123453399	Kdm2b (-14121), Orai1 (-11854)
chr5	123960959	123961289	B3gnt4 (+655)
chr5	123961314	123961799	B3gnt4 (+1088), Diablo (+12628)
chr5	129661847	129662082	Sfswap (-345141), Gpr133 (+59340)
chr5	129760523	129760777	Sfswap (-246456), Gpr133 (+158025)
chr5	135003226	135003818	Gtf2ird1 (-70941), Clip2 (+24782)
chr5	137540337	137540712	Ap1s1 (-18520), Serpine1 (+7617)
chr5	140277688	140278199	Mafk (+10477), Tmem184a (+12293)
chr5	143455156	143455542	Tnrc18 (+123717), Slc29a4 (+277294)
chr5	143579353	143579517	Tnrc18 (-369)
chr5	143579622	143579759	Tnrc18 (-625)
chr5	148910391	148910631	Mtus2 (+141615), Slc7a1 (+300969)
chr5	149864457	149864875	Hmgb1 (-50)
chr4	3429671	3429955	Vmn1r3 (-317308), Tmem68 (+72187)
chr4	3540699	3541766	Lyn (-64029), Tgs1 (+39211)
chr4	8154834	8155634	Car8 (+10954)
chr4	8619372	8619621	Clvs1 (-576967), Chd7 (+1429)
chr4	9133821	9134179	Clvs1 (-62464), Chd7 (+515932)
chr4	11324250	11324666	1110037F02Rik (-88647), Esrp1 (-10528)
chr4	17379647	17379762	Mmp16 (-400900)
chr4	37908025	37908169	NONE
chr4	41811878	41812712	Ccl27a (-91280), Gm20938 (-24489)
chr4	41925351	41926345	Gm2564 (+17616), Gm20938 (+89064)
chr4	41963771	41963931	Gm13304 (-21689), Gm2564 (-20387)
chr4	42023793	42024677	Gm13306 (-528)
chr4	42269110	42269783	Gm13305 (-230121), Il11ra2 (-69813)
chr4	42353908	42354512	Il11ra2 (+14950), Gm13298 (+187497)
chr4	42414654	42414814	Il11ra2 (+75474), Gm13298 (+126973)
chr4	42452240	42453234	Gm13298 (+88970), Il11ra2 (+113477)
chr4	42565486	42566480	4930578G10Rik (-183094), Gm13298 (-24276)
chr4	42603910	42604070	4930578G10Rik (-145087), Gm13298 (-62283)
chr4	42663761	42664645	Gm13298 (-122496), 4930578G10Rik (-84874)
chr4	42748677	42749163	4930578G10Rik (-157)
chr4	46596519	46596890	Trim14 (-47692), Coro2a (+18369)
chr4	50709855	50710061	Grin3a (-851537), Cylc2 (-519592)

chr4	51825350	51825590	Smc2 (-626645), Cylc2 (+595920)
chr4	80355629	80355771	Tyrp1 (-124434)
chr4	81597155	81597635	Mpdz (-508684), Nfib (+553817)
chr4	83269166	83269399	Ccdc171 (+97834)
chr4	101763903	101764474	Pde4b (-163419), Gm12789 (+104958)
chr4	107831578	107831728	Scp2 (-40516), Echdc2 (-6389)
chr4	108952605	108952986	Eps15 (-77)
chr4	114943768	114943896	Cyp4x1 (-137250), Cyp4a12a (-27819)
chr4	114986056	114986184	Cyp4a12b (-98109), Cyp4a12a (+14469)
chr4	115718594	115718991	Faah (-28262), Nsun4 (+7192)
chr4	115893683	115893927	Pik3r3 (-461)
chr4	116862455	116862794	Kif2c (-7381), Gm1661 (+61897)
chr4	116875066	116875292	Kif2c (-19935), Gm1661 (+49343)
chr4	116883412	116883551	Kif2c (-28238), Gm1661 (+41040)
chr4	117559595	117559777	B4galt2 (-3594), Atp6v0b (+248)
chr4	118110022	118110315	Cdc20 (-212)
chr4	118219655	118221372	Tmem125 (-4181)
chr4	120709969	120710060	Col9a2 (-1975)
chr4	123369561	123369835	Macf1 (-8095), Ndufs5 (+25747)
chr4	124617836	124618218	Cdca8 (-3910)
chr4	125568410	125568787	Csf3r (-133195), Grik3 (+400655)
chr4	131301601	131301835	Ptpru (+92428), Matn1 (+801418)
chr4	131673819	131674044	Epb4.1 (-42725), Oprd1 (+26469)
chr4	131825733	131826483	Taf12 (-4182)
chr4	132894275	132894409	Wdtc1 (+897)
chr4	132994761	132995022	Fam46b (-41155), Slc9a1 (+69271)
chr4	134024169	134024391	Stmn1 (+45)
chr4	141102627	141102930	Spen (-8267), Fblim1 (+53060)
chr4	146316224	146316405	Gm13150 (-29770), Zfp600 (+191224)
chr4	149185046	149185351	Slc25a33 (-36813), Spsb1 (+143953)
chr4	149612437	149612864	Rere (-43104), Eno1 (+1590)
chr4	153394885	153395163	Smim1 (+5237), Lrrc47 (+9184)
chr4	153527347	153527613	Tprgl (+7295), Wrap73 (+10999)
chr4	154231384	154231576	Mmel1 (-14258), Ttc34 (+1171)
chr4	154306287	154306408	Tnfrsf14 (-4162)
chr3	5860243	5860873	Pex2 (-284319), 1700008P02Rik (+759885)
chr3	22026150	22026308	Tbl1xr1 (+50655)
chr3	23209200	23209294	NONE
chr3	31260640	31260800	Kcnmb2 (-540709), Slc7a14 (-51420)
chr3	52059270	52059807	Maml3 (-150726), Foxo1 (-12720)
chr3	58469472	58469815	Fam194a (-28515), Siah2 (+26678)
chr3	58469948	58470228	Fam194a (-28959), Siah2 (+26234)
chr3	61949655	61949811	B430305J03Rik (-779860), Arhgef26 (-192533)
chr3	64230774	64230952	Vmn2r4 (-11645), Vmn2r5 (+82794)
chr3	68096244	68096863	Il12a (-398012), Schip1 (+227690)
chr3	69490627	69490976	B3galnt1 (-88019), Nmd3 (-35105)
chr3	77059606	77059705	NONE
chr3	93518698	93518941	Tdpoz4 (-81500), S100a10 (+159818)

chr3	95276043	95276623	Ctsk (-26875), Arnt (+38023)
chr3	96024557	96024756	Hist2h2be (-387), Hist2h2ac (+146), Hist2h2ab (+869)
chr3	96042667	96043048	Hist2h3c2 (+192)
chr3	96043066	96043922	Hist2h3c2 (-444), Hist2h2aa2 (+722)
chr3	96043950	96044533	Hist2h3c2 (-1192), Hist2h2aa2 (-26)
chr3	96049190	96049726	Hist2h3c1 (-1216), Hist2h2aa1 (-3)
chr3	96049754	96050602	Hist2h3c1 (-496), Hist2h2aa1 (+717)
chr3	96050624	96051005	Hist2h3c1 (+141)
chr3	96072452	96072813	Hist2h2bb (-1042), Hist2h3b (+56)
chr3	96132007	96132599	Fcgr1 (-34411), BC107364 (+123926)
chr3	96163525	96164181	Fcgr1 (-65961), BC107364 (+92376)
chr3	96177802	96178458	Fcgr1 (-80238), BC107364 (+78099)
chr3	96226594	96226840	Fcgr1 (-128825), BC107364 (+29512)
chr3	96254032	96254704	Fcgr1 (-156476), BC107364 (+1861)
chr3	96263776	96264448	Hfe2 (-64996), BC107364 (-7883)
chr3	96290119	96291035	Hfe2 (-38531), BC107364 (-34348)
chr3	96292179	96293188	BC107364 (-36455), Hfe2 (-36424)
chr3	103897340	103897866	Phtf1 (+125570), Magi3 (+126423)
chr3	106243488	106243658	BC051070 (-34473), 2010016I18Rik (+42782)
chr3	109061732	109061903	Vav3 (-81753), Slc25a24 (+135751)
chr3	115609465	115610475	Dph5 (+18869), Slc30a7 (+100354)
chr3	124418625	124418909	Ndst4 (-688242), 1700003H04Rik (-134758)
chr3	142973853	142974253	Pkn2 (-429085), Lmo4 (+891228)
chr3	144408324	144408651	Sh3glb1 (-25201), Clca1 (+15317)
chr3	144467409	144467757	Clca1 (-43778), Clca2 (+14875)
chr3	146005710	146006075	Ssx2ip (-61713), Lpar3 (+121968)
chr3	152165618	152166329	Zzz3 (+107007), Ak5 (+165130)
chr3	152996314	152997030	St6galnac5 (-351462), St6galnac3 (+391354)
chr3	154085430	154086146	Tyw3 (+174271), Slc44a5 (+449388)
chr3	154136600	154137010	Tyw3 (+123254), Slc44a5 (+500405)
chr3	159218929	159219028	Rpe65 (-43218), Depdc1a (+60582)
chr2	3356844	3357148	Dclre1c (+15574), Suv39h2 (+35307)
chr2	16187859	16187954	Plxdc2 (-90024), Gm13318 (+224154)
chr2	18461040	18461299	Dnajc1 (-146713), Commd3 (-132880)
chr2	20196979	20197177	Etl4 (-244340), Otud1 (+617389)
chr2	20821416	20821899	Arhgap21 (+67689), Etl4 (+380240)
chr2	22443076	22443924	Gad2 (-34225), Myo3a (+294370)
chr2	25035800	25036694	Nrarp (-31)
chr2	28980310	28981010	Setx (+148)
chr2	34611014	34611361	Gapvd1 (-25021), Hspa5 (-16424)
chr2	51198320	51198829	Tas2r134 (-284456), Rnd3 (-193944)
chr2	57246848	57247325	Galnt5 (-603208), Gpd2 (+156379)
chr2	57481881	57482153	Galnt5 (-368278), Gpd2 (+391309)
chr2	80302072	80302264	Frzb (-14386), Nckap1 (+119369)
chr2	83373876	83374011	Zc3h15 (-110648), Fsip2 (+590153)
chr2	98502287	98503234	Gm10801 (+367)
chr2	98503710	98504293	Gm10801 (+1608), Gm10800 (+3456)
chr2	98505008	98505437	Gm10800 (+2235), Gm10801 (+2829)

chr2	98506338	98507542	Gm10800 (+518)
chr2	99785651	99785785	NONE
chr2	105308425	105308582	Pax6 (-200549), Rcn1 (-69028)
chr2	111925966	111926040	Olfr1313 (-13263), Olfr1314 (+6853)
chr2	117059330	117059501	Fam98b (-16059), Spred1 (+112306)
chr2	121962909	121963298	Spg11 (-18982), B2m (-10318)
chr2	124986392	124986659	Dut (-86400), Slc12a1 (+8285)
chr2	125443221	125443497	Fbn1 (-111238), Cep152 (+7490)
chr2	131233744	131234170	Smox (-83722), Rnf24 (-55340)
chr2	142290456	142290537	Kif16b (+436770)
chr2	146335546	146336272	Ralgapa2 (+1831), Insm1 (+288252)
chr2	162688176	162688554	Ptprt (-201482), Srsf6 (-68899)
chr2	166065850	166066134	Sulf2 (-84829), Prex1 (+473340)
chr2	167702233	167702451	Ptpn1 (-55485), Cebpb (+187927)
chr2	167702530	167702802	Ptpn1 (-55161), Cebpb (+188251)
chr2	170945712	170946046	Dok5 (+388572), Cbln4 (+923087)
chr1	6638065	6638194	St18 (-82002), Fam150a (+288593)
chr1	8400880	8401020	Sntg1 (+59651)
chr1	12634806	12634963	Sulf1 (-73741)
chr1	16647040	16647245	Tceb1 (-197)
chr1	23067721	23067887	Rims1 (-464382), 4933415F23Rik (+41288)
chr1	24618332	24623038	Gm10222 (-1146)
chr1	30819114	30819315	Gm9898 (-42409), Phf3 (+100886)
chr1	33951484	33951619	Zfp451 (-80112), Bend6 (+13109)
chr1	33951622	33951859	Zfp451 (-80301), Bend6 (+12920)
chr1	35326119	35326392	Hs6st1 (-798989), Plekhb2 (+419419)
chr1	36425881	36426371	Kansl3 (-1267)
chr1	38461097	38461289	Rev1 (-274686), Aff3 (+135494)
chr1	39576302	39576542	Tbc1d8 (-40830), Cnot11 (-15415)
chr1	46910029	46910937	Slc39a10 (+408)
chr1	48519013	48519161	C230029F24Rik (-782373)
chr1	49014660	49014787	C230029F24Rik (-286736)
chr1	57513605	57513800	Spats2l (-318002), Tyw5 (-49758)
chr1	66863661	66863845	Kansl1l (+367)
chr1	69851276	69851438	Ikzf2 (-118823), Spag16 (-22205)
chr1	72283383	72283807	Mreg (-24714), Pecr (+47293)
chr1	72290738	72291074	Mreg (-32025), Pecr (+39982)
chr1	72301572	72301870	Mreg (-42840), Pecr (+29167)
chr1	83996431	83996610	Sphkap (-591746), Pid1 (+284699)
chr1	84836064	84836238	Trip12 (-272), Fbxo36 (-265)
chr1	88383859	88384257	Ptma (-39253), 1700019O17Rik (+61112)
chr1	88422999	88423418	Ptma (-102)
chr1	88617661	88617989	Dis3l2 (+17446), Alppl2 (+368678)
chr1	91355897	91356040	D130058E05Rik (-470840), Agap1 (+4583)
chr1	95841863	95842057	Gm6086 (-45126), Gm9994 (+17596)
chr1	99816002	99816317	Ppip5k2 (-149172), Pam (+176038)
chr1	105755854	105756011	Cdh20 (-909173)
chr1	115518268	115518479	NONE

chr1	120355482	120355898	2900060B14Rik (-366)
chr1	130932071	130932291	Cxcr4 (-443314), Thsd7b (-237740)
chr1	133716876	133717034	Slc41a1 (-7599), Pm20d1 (+22997)
chr1	134318052	134318382	Rbbp5 (-55727), Dstyk (+4187)
chr1	136457209	136457415	Kdm5b (+564)
chr1	136457440	136457676	Kdm5b (+810)
chr1	136457717	136457907	Syt2 (-148065), Kdm5b (+1064)
chr1	136920453	136920843	Ube2t (+61506), Lgr6 (+81205)
chr1	138233122	138233315	Gpr25 (-75769), Camsap2 (+9462)
chr1	138235598	138235739	Gpr25 (-78219), Camsap2 (+7012)
chr1	141954325	141954737	Gm4788 (-276715), Cfh (+125457)
chr1	142869679	142869779	Kcnt2 (+726934)
chr1	159482429	159482704	Sec16b (+45708), Fam5b (+803824)
chr1	159661262	159661466	Sec16b (+224505), Fam5b (+625027)
chr1	162275331	162275404	Cacybp (-132460), Rabgap1l (+447701)
chr1	163027312	163027488	Cenpl (+26502), Klhl20 (+34242)
chr1	169270302	169270622	Aldh9a1 (-9660), Tmco1 (+31661)
chr1	169328474	169328599	Mgst3 (-4565)
chr1	169600816	169601057	Lmx1a (-18752), Rxrg (+72422)
chr1	172572750	172572955	Olfml2b (-1810)
chr1	172958057	172958306	Fcgr4 (+9131), Fcgr3 (+31352)
chr1	173001780	173005313	1700009P17Rik (-48245), Fcgr3 (-14013)
chr1	173012987	173014985	1700009P17Rik (-37806), Fcgr3 (-24452)
chr1	173432897	173433394	Alyref2 (-463)
chr1	173808683	173808857	Gm10521 (-17025), Cd84 (+38942)
chr1	174196993	174197081	Atp1a4 (-8492), Atp1a2 (+31158)
chr1	178744251	178744519	Sdccag8 (-406)
chr1	180250830	180251576	Cox20 (+1942), Hnrnpu (+16725)
chr1	180251592	180252051	Cox20 (+2561), Hnrnpu (+16106)
chr1	180252156	180252389	Cox20 (+3012), Hnrnpu (+15655)
chr1	180252951	180253209	Cox20 (+3819), Hnrnpu (+14848)
chr1	180266685	180267253	Hnrnpu (+959)
chr1	183634010	183634190	Ccdc121 (-192518), Dnahc14 (-37067)
chr1	186854896	186855388	Rab3gap2 (-172905), Mark1 (-31714)
chr1	192053952	192054194	Prox1 (-59514), Rps6kc1 (+681576)
chr1	194061545	194061791	Kcnh1 (+47002), Hhat (+535739)
chr1	197190299	197190481	Cr2 (-187481)
chrY	2869581	2872285	Gm10352 (+480543)
chr9	2999998	3002205	Gm10722 (+180)
chr9	3002257	3003490	Gm10720 (-12780), Gm10722 (+1952)
chr9	3003894	3004721	Gm10720 (-11346), Gm10722 (+3386)
chr9	3008887	3009792	Gm10720 (-6314), Gm10722 (+8418)
chr9	3011305	3013245	Gm10720 (-3379)
chr9	3013318	3014219	Gm10720 (-1885)
chr9	3015054	3015859	Gm10720 (-197)
chr9	3016808	3017613	Gm10718 (-6336), Gm10720 (+1557)
chr9	3017722	3018549	Gm10718 (-5411), Gm10720 (+2482)
chr9	3023697	3024501	Gm10718 (+552)

chr9	3024619	3028825	Gm10718 (+3175), 4930433N12Rik (+173060)
chr9	3029353	3030835	Gm10718 (+6547), 4930433N12Rik (+169688)
chr9	3030999	3033810	Gm10718 (+8858), 4930433N12Rik (+167377)
chr9	3035278	3036770	Gm10718 (+12477), 4930433N12Rik (+163758)
chr9	3036840	3037472	Gm10718 (+13609), 4930433N12Rik (+162626)
chr9	3037542	3038423	Gm10718 (+14436), 4930433N12Rik (+161799)
chr9	9927813	9927959	Arhgap42 (-688873), Cntn5 (+976889)
chr9	10564871	10565170	Cntn5 (+339754)
chr9	13338640	13339469	Maml2 (-85379), Phxr4 (+104474)
chr9	19029130	19029345	Olfr843 (+24669), Olfr836 (+103837)
chr9	25059849	25060024	Sept7 (-232)
chr9	28182826	28182970	Opcml (+584366)
chr9	35112756	35113275	Cdon (-116166), Rpusd4 (+37566)
chr9	37980371	37980508	Olfr891 (+7966), Olfr890 (+29718)
chr9	41056175	41056341	Ubash3b (-90113), Sorl1 (+876122)
chr9	41780734	41780824	Ubash3b (-814634), Sorl1 (+151601)
chr9	44142700	44142969	H2afx (+37)
chr9	44689301	44689523	Mll1 (-33)
chr9	45064717	45064951	Tmprss4 (-52659), Il10ra (+12398)
chr9	45443886	45444033	Cep164 (+192814), Dscaml1 (+205584)
chr9	49183419	49183597	Drd2 (+34776), Ankk1 (+51618)
chr9	56071355	56071791	Tspan3 (-62977), C230081A13Rik (+194284)
chr9	58538904	58539073	2410076I21Rik (+50377), Nptn (+108890)
chr9	58743557	58744359	Hcn4 (+72639), Neo1 (+140290)
chr9	59505862	59506188	Gramd2 (-49419), Pkm (+1642)
chr9	65044185	65044584	Parp16 (-18569), Igdcc3 (+55424)
chr9	65049332	65049789	Parp16 (-13393), Igdcc3 (+60600)
chr9	67419497	67419783	C2cd4b (-187604), Tln2 (-12130)
chr9	70942604	70942770	Lipc (-160072), Aqp9 (+68409)
chr9	71908716	71909086	Cgnl1 (-289492), Tcf12 (+50725)
chr9	75471052	75471301	Lysmd2 (-2362)
chr9	82868925	82869191	Phip (+38)
chr9	82869297	82869515	Phip (-310)
chr9	96435662	96435871	Rnf7 (-56673), Rasa2 (+96269)
chr9	99341156	99341459	Mras (-4171)
chr9	100119352	100119755	Sox14 (-342891), Il20rb (+267653)
chr9	100119779	100119975	Sox14 (-343214), Il20rb (+267330)
chr9	108465913	108466233	Impdh2 (+3299), Ndufaf3 (+3600)
chr9	108466612	108467145	Ndufaf3 (+2794), Impdh2 (+4105)
chr9	110154884	110154975	Elp6 (-52766), Cspg5 (+8643)
chr9	110428162	110428353	Setd2 (-6843), Kif9 (+48760)
chr9	110428485	110429550	Setd2 (-6083), Kif9 (+49520)
chr9	110989678	110990024	Lrrfip2 (-30764), Ccrl2 (-30077)
chr9	112911176	112911399	Arpp21 (-776283), Pdcd6ip (+706074)
chr9	115695410	115695590	Stt3b (-475961), Gadl1 (-122182)
chr9	115695612	115695770	Stt3b (-476152), Gadl1 (-121991)
chr9	118647931	118648304	Ctdspl (-187453), Itga9 (+132310)
chr9	120058537	120058711	Mobp (-236)

chr9	123370836	123371309	Limd1 (-16746), Lars2 (+95015)
chr8	4944835	4945023	Shcbp1 (-165395), Slc10a2 (+160303)
chr8	10676257	10676741	3930402G23Rik (+251958), Myo16 (+403927)
chr8	12651761	12651915	Tubgcp3 (+20410), Spaca7 (+78789)
chr8	15519762	15520194	Myom2 (+462325)
chr8	25462270	25462920	1810011O10Rik (+86823), A730045E13Rik (+178219)
chr8	31438279	31438743	Dusp26 (-761432)
chr8	48678169	48678670	Rwdd4a (+59402), Ing2 (+82092)
chr8	48799209	48799365	Cdkn2aip (-2)
chr8	53422254	53422419	NONE
chr8	62368986	62369151	Gm10283 (+612926), BC030500 (+978517)
chr8	68498956	68499055	March1 (+357082), Tma16 (+511412)
chr8	73222346	73223170	Jund (+1120), Gm11175 (+1221)
chr8	82479259	82479547	Hhip (+102502), Anapc10 (+243684)
chr8	83262886	83263555	Smarca5 (+175)
chr8	85884272	85884590	Clgn (-29346), 4933434I20Rik (+12061)
chr8	89571694	89572030	N4bp1 (-162705), Cbln1 (+424646)
chr8	93860665	93860961	Fto (+23382), Irx3 (+464740)
chr8	97794284	97794571	Kifc3 (-127988), Cngb1 (+13653)
chr8	107828141	107828923	Elmo3 (-969)
chr8	110220901	110221068	Psmd7 (-108603)
chr8	112302324	112302543	Ap1g1 (-20)
chr8	119567477	119567793	Gcsh (-50198), Pkd1l2 (+38714)
chr8	122124325	122124484	Taf1c (+4717), Dnaaf1 (+25270)
chr8	122125009	122125167	Taf1c (+4034), Dnaaf1 (+25953)
chr8	122126044	122126895	Taf1c (+2652), Dnaaf1 (+27335)
chr8	124494912	124495249	Gm22 (-298388), Banp (+20637)
chr8	125801296	125801869	Vps9d1 (-23335), Fanca (+40893)
chr8	127528426	127529090	Egln1 (-55604), Tsnax (-8139)
chr13	3026151	3026232	Gdi2 (-511129)
chr13	3179280	3179495	Gdi2 (-357933)
chr13	6134585	6134854	Pitrm1 (-412683), Klf6 (+273985)
chr13	9833500	9833762	Zmynd11 (-69225), Chrm3 (+44698)
chr13	12366712	12366965	Mtr (-16572), Actn2 (+66188)
chr13	16464769	16464850	Inhba (+361126)
chr13	20876498	20876763	Aoah (-9357), Elmo1 (+694143)
chr13	21809473	21809969	Hist1h2bm (-4192), Hist1h3h (+224)
chr13	21813489	21813714	Hist1h2bm (-311)
chr13	21826842	21827336	Hist1h4j (+186)
chr13	21842001	21842479	Hist1h4k (+134)
chr13	21846168	21846343	Hist1h4k (-3882), Hist1h2ak (-480)
chr13	21874791	21875299	Hist1h2bp (-4285), Hist1h1b (-2551), Hist1h3i (+221)
chr13	21879560	21879759	Hist1h3i (-4394), Hist1h2an (-573), Hist1h2bp (+330)
chr13	21901526	21901863	Hist1h4m (-1920), Hist1h2bq (+358)
chr13	21901911	21902528	Hist1h4m (-1395), Hist1h2bq (-167)
chr13	21903666	21904052	Hist1h2bq (-1806), Hist1h4m (+244)
chr13	21923627	21924238	Hist1h2br (-1694), Hist1h4n (+132)
chr13	21925147	21925768	Hist1h4n (-1393), Hist1h2br (-169)

chr13	21925816	21926153	Hist1h4n (-1920), Hist1h2br (+358)
chr13	22127932	22128107	Hist1h2ah (-508), Hist1h2bk (+330)
chr13	22132822	22133297	Hist1h2bj (-2023), Hist1h4i (+161)
chr13	22135028	22135487	Hist1h4i (-2037), Hist1h2ag (-440), Hist1h2bj (+175)
chr13	23622999	23623365	Hist1h3g (-4109), Hist1h2af (-2593), Hist1h4h (+263)
chr13	23643135	23643389	Hist1h1d (-3639), Hist1h4f (+255)
chr13	23646975	23647201	Hist1h4f (-3571), Hist1h1d (+187)
chr13	23667674	23668022	Hist1h2ae (-4759), Hist1h2bf (-1783), Hist1h3d (+216)
chr13	23714034	23714302	Hist1h2be (-1175), Hist1h1e (+203)
chr13	23776051	23776229	Hist1h2ac (-323), Hist1h2bc (+72)
chr13	23789960	23790459	Hist1h4c (+113)
chr13	23830593	23831031	Hist1h1c (+136)
chr13	23838768	23839012	Hist1h2ab (-4067), Hist1h3c (-1419), Hist1h2bb (+287)
chr13	23844179	23844628	Hist1h4b (-4402), Hist1h3b (+268)
chr13	23848933	23849239	Hist1h4b (+280)
chr13	23852658	23853012	Hist1h1a (-2702), Hist1h4a (+264)
chr13	23853920	23854301	Hist1h1a (-1426), Hist1h4a (-1012), Hist1h3a (+144)
chr13	24025614	24026204	Hist1h2aa (-422), Hist1h2ba (+116)
chr13	24633416	24633684	Fam65b (-96967), Cmah (+214261)
chr13	42190662	42190816	Edn1 (-205900), Hivep1 (+43349)
chr13	44964885	44965059	Dtnbp1 (+132537), Jarid2 (+138780)
chr13	47046203	47046525	Kif13a (-21277), Nhlrc1 (+63855)
chr13	49494901	49495368	Ippk (-21545), Bicd2 (+58217)
chr13	51297951	51298500	Hist1h2al (+208)
chr13	51941709	51941925	Gadd45g (-230)
chr13	51941968	51942164	Gadd45g (+19)
chr13	52846109	52846573	Syk (+167799), Auh (+178689)
chr13	56199754	56200260	Pitx1 (-267221), H2afy (+36985)
chr13	58502820	58503374	Rmi1 (-861)
chr13	58503773	58503940	Rmi1 (-101), Hnrnpk (+847)
chr13	61052440	61052524	Tpbpa (-9186), Ctsj (+54795)
chr13	62460597	62460717	Gm10260 (-141328), Gm5141 (+12639)
chr13	65552554	65553189	Gm10139 (-60742), Zfp369 (+172750)
chr13	65967440	65968021	Gm10772 (-39856), Gm10139 (+354117)
chr13	66394389	66396325	Vmn2r-ps104 (+60964), 2610044O15Rik8 (+256118)
chr13	67299065	67299259	Zfp455 (+3720), Zfp458 (+70842)
chr13	67474005	67474254	Zfp953 (-12589), Zfp456 (+2616)
chr13	67497929	67498144	Zfp456 (-21291), Zfp429 (+2666)
chr13	68096842	68096930	BC048507 (+132612), Mtrr (+624112)
chr13	69733647	69733870	Papd7 (-61042), Srd5a1 (+16561)
chr13	70231949	70232253	Med10 (+283341), BC018507 (+544411)
chr13	91778066	91778279	Acot12 (-102944), Ssbp2 (+177518)
chr13	91778284	91778432	Acot12 (-102759), Ssbp2 (+177703)
chr13	92094003	92094177	Ckmt2 (-77600), Rasgrf2 (+807359)
chr13	97342467	97342837	Col4a3bp (+29962), Hmgcr (+98239)
chr13	101432401	101432811	Ccdc125 (-6830), Taf9 (+11308)
chr12	3036912	3037000	Rab10 (+273013)
chr12	4786211	4786418	0610009D07Rik (-38099), Pfn4 (+10214)

chr12	5503370	5503761	2810032G03Rik (+85119)
chr12	9432167	9432451	Osr1 (-148939), Ttc32 (+395506)
chr12	12018484	12018858	Fam49a (-250274), Rad51ap2 (+555786)
chr12	12018866	12019112	Fam49a (-249956), Rad51ap2 (+556104)
chr12	13143011	13143206	Mycn (-194389), Ddx1 (+112910)
chr12	17539673	17540043	Odc1 (-11821), Nol10 (+184573)
chr12	19226266	19226526	Gm5784 (-167269), 5730507C01Rik (+704852)
chr12	20828925	20829099	1700030C10Rik (-7424), Zfp125 (+77716)
chr12	22990713	22991030	NONE
chr12	35958601	35958755	Snx13 (+226817), Ahr (+261032)
chr12	37392265	37392572	Meox2 (-442708), D630036H23Rik (-283651)
chr12	38091134	38091230	Dgkb (-516110), Agmo (+122954)
chr12	50712141	50712236	Gm9804 (+209925)
chr12	55797359	55798061	Eapp (-842)
chr12	55810882	55811993	Eapp (-14570), Snx6 (+85252)
chr12	55814849	55815664	Eapp (-18389), Snx6 (+81433)
chr12	55819723	55820200	Eapp (-23094), Snx6 (+76728)
chr12	55820236	55820781	Eapp (-23641), Snx6 (+76181)
chr12	55830120	55831178	Eapp (-33781), Snx6 (+66041)
chr12	55835205	55835592	Eapp (-38531), Snx6 (+61291)
chr12	55835629	55836060	Eapp (-38977), Snx6 (+60845)
chr12	55838910	55839950	Eapp (-42562), Snx6 (+57260)
chr12	59261099	59261197	Foxa1 (-614040), Sstr1 (-51611)
chr12	62477546	62477700	Lrfn5 (-147312)
chr12	71978715	71978914	Actr10 (-60029), Frmd6 (+52314)
chr12	79949073	79949263	Atp6v1d (+13457), Mpp5 (+99234)
chr12	92828267	92828427	Gtf2a1 (+102)
chr12	93036365	93036569	Ston2 (-11591), Sel1l (+51130)
chr12	98299606	98300037	NONE
chr12	103079306	103079937	Fbln5 (-22357), Trip11 (+71855)
chr12	111934120	111934354	Hsp90aa1 (+368)
chr12	112341390	112341756	Traf3 (-63101), Rcor1 (+63748)
chr12	112579306	112579480	Cdc42bpb (+36536), Amn (+70071)
chr12	113088150	113088677	Ppp1r13b (+22412), Zfyve21 (+36033)
chr12	113338413	113338652	BC048943 (-119571), Aspg (-6361)
chr11	3724658	3725260	Osbp2 (+38947), Morc2a (+175462)
chr11	3971581	3971778	Sec14l3 (+6836), Mtfp1 (+23768)
chr11	11397628	11398506	4930415F15Rik (+8798), 4930512M02Rik (+129760)
chr11	14977824	14977948	NONE
chr11	22170490	22170823	Otx1 (-269039), Ehbp1 (+15184)
chr11	23624548	23624891	Pex13 (-58761), Rel (+46250)
chr11	25000048	25000203	5730522E02Rik (+670444)
chr11	32076000	32076214	Il9r (+24129), Nsg2 (+175644)
chr11	40613689	40613785	Ccng1 (-44924)
chr11	50191163	50191410	Hnrnph1 (+16)
chr11	52099825	52100178	Tcf7 (-3929)
chr11	58775172	58775820	Trim17 (-1808)
chr11	60187409	60187676	Lrrc48 (+20712), Atpaf2 (+43010)

chr11	62364915	62365180	Ubb (+375)
chr11	62365648	62366351	Trpv2 (-22040), Ubb (+1327)
chr11	69571960	69572257	Polr2a (+30)
chr11	70729707	70729900	Nup88 (+53660), Rabep1 (+71524)
chr11	73312282	73312397	Olfr381 (-7361), Olfr382 (+18359)
chr11	76148608	76148883	Glod4 (-91519), Nxn (+63896)
chr11	76191172	76191993	Glod4 (-134356), Nxn (+21059)
chr11	76191995	76192164	Glod4 (-134853), Nxn (+20562)
chr11	79398713	79398991	Evi2a (-54762), Rab11fip4 (-5862)
chr11	79687340	79687428	Utp6 (+88508), Rab11fip4 (+282670)
chr11	82841057	82841310	Slfn2 (-37430), Slfn8 (-6872)
chr11	83084948	83085270	Pex12 (+27370), Slfn3 (+80277)
chr11	86887233	86887808	Gdpd1 (+43)
chr11	87236084	87236908	Gm11492 (-143659), Tex14 (+17929)
chr11	87239984	87240835	Gm11492 (-139745), Tex14 (+21843)
chr11	87256352	87257209	Gm11492 (-123374), Tex14 (+38214)
chr11	87261721	87262640	Gm11492 (-117974), Tex14 (+43614)
chr11	87275408	87276317	Gm11492 (-104292), Tex14 (+57296)
chr11	87284520	87285299	Gm11492 (-95245), Tex14 (+66343)
chr11	95737123	95737400	Abi3 (-33539), B4galnt2 (+38943)
chr11	95893927	95894111	Snf8 (-2180)
chr11	98724900	98725256	Wipf2 (+126)
chr11	101466103	101466410	Rdm1 (-22999), Tmem106a (+22701)
chr11	101488622	101488966	Rdm1 (-462)
chr11	101502809	101503485	Arl4d (-23708), Rdm1 (+13891)
chr11	101506771	101507524	Arl4d (-19707), Rdm1 (+17892)
chr11	101519404	101519849	Arl4d (-7228), Rdm1 (+30371)
chr11	103268409	103268520	Arhgap27 (-46251), Plekhm1 (+5513)
chr11	105301298	105301520	March10 (+16640), Mrc2 (+147449)
chr11	106586494	106586742	Tex2 (-112374), Pecam1 (+25324)
chr11	108872866	108873417	Axin2 (+91479), E030025P04Rik (+132541)
chr11	115885512	115885926	Galk1 (-11686), H3f3b (+3557)
chr11	116928412	116928569	Sec14l1 (-47995), Mgat5b (+148314)
chr11	119320596	119321173	Endov (-31776), Rnf213 (+66471)
chr11	119720614	119720939	Chmp6 (-54347), Rptor (+256314)
chr11	120092832	120092980	Bahcc1 (-1355)
chr11	121518509	121519059	Zfp750 (-138137), B3gntl1 (+15683)
chr10	3831543	3831933	Rgs17 (-592417), Gm10945 (+246958)
chr10	4930547	4930938	Syne1 (+126216), Esr1 (+803777)
chr10	9467362	9467556	Samd5 (-72453), Stxbp5 (+153418)
chr10	11710100	11710246	Gm9797 (+381152), Utrn (+871360)
chr10	12273996	12274342	Utrn (+307364), Gm9797 (+945148)
chr10	17037210	17037294	Cited2 (-405782)
chr10	17044382	17044530	Cited2 (-398578)
chr10	21000123	21000287	Myb (-119415), Hbs1l (-15580)
chr10	26647120	26647204	Arhgap18 (+154879), Lama2 (+689586)
chr10	26655097	26655181	Arhgap18 (+162856), Lama2 (+681609)
chr10	35943971	35944136	Amd2 (-512357), Hs3st5 (-282566)

chr10	37782321	37782472	Lama4 (-902924), 5930403N24Rik (+923033)
chr10	39691073	39691718	G630090E17Rik (-11437), BC021785 (+15049)
chr10	39978217	39978605	Gtf3c6 (-940)
chr10	49977195	49977294	Grik2 (-468673), Ascc3 (-335230)
chr10	59378436	59378712	Ddit4 (+35944), Dnajb12 (+36251)
chr10	61591506	61591858	Neurog3 (-4156)
chr10	76766491	76766846	Pofut2 (+44624), Adarb1 (+114346)
chr10	77152943	77153141	Gm10272 (-16347), Ube2g2 (+67961)
chr10	80827990	80829989	4930404N11Rik (-429)
chr10	81167324	81168170	Zfp873 (-353415), Ankrd24 (+76367)
chr10	81326463	81327279	Zfp873 (-194291), Ankrd24 (+235491)
chr10	81459541	81460166	Zfp873 (-61308), Ankrd24 (+368474)
chr10	89139305	89139721	Slc17a8 (-55515), Scyl2 (+9517)
chr10	110654809	110655409	Bbs10 (-80626), Osbpl8 (+53251)
chr10	115632868	115633040	Ptprb (-105476), 4933416C03Rik (-81981)
chr17	4626143	4626590	Nox3 (-930106), Arid1b (-368707)
chr17	5011512	5011895	Arid1b (+16630), Tmem242 (+428556)
chr17	10380494	10380692	Pabpc6 (-518024), Qk (+131633)
chr17	13743457	13746115	Mllt4 (-152769), Smok4a (+30464)
chr17	13911136	13911387	Gm7168 (-174118), Mllt4 (+13707)
chr17	14847569	14847759	Thbs2 (-16395), Wdr27 (+223994)
chr17	15098409	15099907	Phf10 (-921)
chr17	15115289	15116168	Gm3448 (-108), Gm3435 (-100)
chr17	15129992	15131491	Gm3435 (+14913), Gm3417 (+16406)
chr17	15146650	15147540	9030025P20Rik (-105), Gm3417 (+53)
chr17	15161360	15162859	9030025P20Rik (+14910), Tcte3 (+16413)
chr17	15178018	15178623	2210404J11Rik (-284), Tcte3 (+202)
chr17	15344966	15345538	2210404J11Rik (+166647), Dll1 (+167688)
chr17	15950608	15950912	Rgmb (+12790), Chd1 (+108829)
chr17	25896831	25897067	Narfl (-13772), Msln (-5677)
chr17	27298995	27299442	Mnf1 (-28358), Ip6k3 (+5490)
chr17	33811718	33812117	Hnrnpm (+11887), Pram1 (+36917)
chr17	34256094	34256550	Brd2 (+2370), H2-Oa (+27037)
chr17	35618079	35618232	Pou5f1 (-24826), H2-Q10 (+11122)
chr17	39979856	39985845	Gm7148 (-269450)
chr17	43892632	43892898	Rcan2 (-46035), Cyp39a1 (+88391)
chr17	56339395	56339517	D17Wsu104e (-16113), Dpp9 (+18856)
chr17	60653649	60653794	NONE
chr17	65987738	65987998	Vapa (-24973), Txndc2 (+3676)
chr17	66713720	66713971	Soga2 (+85244), Ddx11 (+240986)
chr17	69054628	69055037	Epb4.1l3 (-451317), L3mbtl4 (+431696)
chr17	69816419	69816669	Dlgap1 (-957590), A330050F15Rik (+27878)
chr17	71834308	71834443	Smchd1 (-9693), Ndc80 (+41821)
chr17	71834471	71834609	Smchd1 (-9857), Ndc80 (+41657)
chr17	75509211	75509527	Rasgrp3 (-355509), Ltbp1 (+104461)
chr17	80606236	80606439	Ttc39d (-462), Srsf7 (-318)
chr17	80606473	80606939	Srsf7 (-686), Ttc39d (-94)
chr17	84581105	84581330	Haao (-335088), Zfp36l2 (+6069)

chr17	86824334	86824518	Gm10309 (+80146), Prkce (+257301)
chr17	87452962	87453212	Gm5499 (-24524), Cript (+28197)
chr17	87608182	87608477	Mcfd2 (+56945), Socs5 (+101311)
chr17	89174552	89174875	Lhcgr (+16602), Gtf2a1l (+106714)
chr16	3174030	3174196	Olfr161 (-418285)
chr16	4301021	4301518	Crebbp (-87866), Adcy9 (+118538)
chr16	9241776	9241912	1810013L24Rik (+411651), Grin2a (+753170)
chr16	11015717	11015936	Snn (-50564), Litaf (-22613)
chr16	11143902	11144496	Txndc11 (-9456), Zc3h7a (+32287)
chr16	17088054	17088368	Ypel1 (+18422), Ppil2 (+23114)
chr16	17088394	17089580	Ypel1 (+19198), Ppil2 (+22338)
chr16	17090102	17090259	Ypel1 (+20392), Ppil2 (+21144)
chr16	17090584	17090999	Ppil2 (+20533), Ypel1 (+21003)
chr16	19925935	19926054	A930003A15Rik (-42005), Klhl6 (+57115)
chr16	20129532	20130245	Yeats2 (-11247), Klhl24 (+32262)
chr16	22672855	22673155	Dgkg (-15759), Crygs (+138478)
chr16	31162663	31162843	Xxylt1 (-81235), Acap2 (+38324)
chr16	34671860	34672520	Ccdc14 (-18512), Ropn1 (+20916)
chr16	36991075	36991339	Fbxo40 (-654)
chr16	42569884	42570037	Zbtb20 (-677436), Gap43 (-229197)
chr16	56096558	56096901	Impg2 (-107721), Senp7 (+21208)
chr16	57391146	57391885	Filip1l (+38310), Cmss1 (+215426)
chr16	61238108	61238289	Epha6 (-633125)
chr16	67714057	67714263	Cadm2 (-93092)
chr16	68274642	68274795	Cadm2 (-653651)
chr16	70076415	70076548	Gbe1 (-237712), Speer2 (-212493)
chr16	72258386	72258542	Robo1 (-404930)
chr16	75294903	75295094	Robo2 (-883765), Lipi (+291302)
chr16	80759616	80759772	Ncam2 (-441248)
chr16	84647990	84648749	Mrpl39 (+87617)
chr16	84975160	84975506	Atp5j (-139463), App (+198615)
chr16	94057510	94057989	Cldn14 (-48668), Sim2 (-28311)
chr16	95979203	95979519	Ets2 (+55679), Psmg1 (+233206)
chr16	95995796	95996097	Ets2 (+72265), Psmg1 (+216620)
chr15	5099576	5099919	Ttc33 (-35812), Prkaa1 (+5887)
chr15	7527614	7527751	Gdnf (-233328), Egflam (-179332)
chr15	11259966	11260124	Tars (+69368), Adamts12 (+265500)
chr15	25143986	25144140	9230109A22Rik (-65149), Basp1 (+199456)
chr15	26188739	26188813	March11 (-50043), Zfp622 (+274655)
chr15	30510534	30510630	Dap (-643558), Ctnnd2 (+408234)
chr15	36131941	36132320	Spag1 (+23008), Rnf19a (+80771)
chr15	38212163	38212414	Klf10 (+18173), Odf1 (+63331)
chr15	44622528	44622623	Sybu (-2967)
chr15	76128460	76128765	Smpd5 (+3749), Oplah (+8918)
chr15	76128804	76129053	Smpd5 (+4065), Oplah (+8602)
chr15	79376866	79377092	Ddx17 (+192)
chr15	79560750	79561176	Sun2 (+12003), Gtpbp1 (+39643)
chr15	79561216	79561436	Sun2 (+11640), Gtpbp1 (+40006)

chr15	82725119	82725439	Cyp2d26 (-100604), Tcf20 (+17285)
chr15	86025175	86025777	Tbc1d22a (-19401), Cerk (-8905)
chr15	87797353	87797527	Zdhhc25 (-633305), Fam19a5 (+422536)
chr15	88568772	88569185	Brd1 (-4316)
chr15	92626642	92626733	Pdzrn4 (+399342), Gxylt1 (+478904)
chr15	92628957	92629048	Pdzrn4 (+401657), Gxylt1 (+476589)
chr15	98364436	98364632	Kansl2 (+118)
chr15	101724635	101724846	Krt76 (-1390)
chr14	4814465	4814595	Gm9602 (-64846), Gm3159 (+313946)
chr14	11353424	11353519	Fhit (+641077)
chr14	12935863	12935995	3830406C13Rik (-180794), Ptprg (+549883)
chr14	14855953	14856561	Atxn7 (+11252), Psmd6 (+97241)
chr14	37719530	37719737	Ccser2 (+62316), 4930474N05Rik (+811489)
chr14	46052730	46052914	Gnpnat1 (-44345), Fermt2 (+96971)
chr14	48705251	48705540	Peli2 (-35148), Ktn1 (+421965)
chr14	48705553	48705674	Peli2 (-34930), Ktn1 (+422183)
chr14	49739614	49739845	Exoc5 (-53402), Naa30 (-52543)
chr14	54880915	54881135	Abhd4 (+2118), Olfr49 (+21575)
chr14	57315836	57316043	Mphosph8 (+28855), Pspc1 (+81213)
chr14	57430145	57430484	Zmym5 (+238)
chr14	57505231	57505594	Zmym2 (-1218)
chr14	57505618	57505797	Zmym2 (-923)
chr14	58458772	58459036	Mrp63 (+13828), Zdhhc20 (+50113)
chr14	59029643	59029822	Rpl13-ps3 (-482610), Fgf9 (+338210)
chr14	60064554	60064873	Setdb2 (-4993), Cab39l (+4896)
chr14	61964386	61964739	Ebpl (+14719), Arl11 (+35973)
chr14	63379777	63380091	Ints6 (+15)
chr14	69799171	69799572	Nkx3-1 (-9323), Nkx2-6 (+9739)
chr14	104736093	104736324	Ednrb (-492819), Pou4f1 (+130999)
chr14	105115012	105115391	Rnf219 (-193319), Rbm26 (+461338)
chr14	118206572	118206838	Dct (+244763), Gpc6 (+881952)
chr14	120737186	120737476	Rap2a (-140335), Mbnl2 (+62402)
chr14	121483899	121484504	Farp1 (+49416), Stk24 (+294354)
chr19	3767959	3768195	Suv420h1 (+656)
chr19	4078363	4078607	BC021614 (-19190), Cabp2 (-5034)
chr19	5803685	5803874	Scyl1 (-32379), Frmd8 (+71494)
chr19	6005129	6005459	Slc22a20 (-19151), Capn1 (+10531)
chr19	6363806	6363988	Sf1 (+207)
chr19	6996550	6997247	Esrra (-601)
chr19	7015773	7016395	Bad (-267), Gpr137 (-115)
chr19	9441752	9441926	Pcna-ps2 (+83970), Stxbp3b (+191899)
chr19	9633435	9633601	Stxbp3b (+220)
chr19	10379302	10379563	Dagla (-66)
chr19	21581779	21582122	Gm3443 (-48213), Gda (-34789)
chr19	29703448	29703901	Ermp1 (+19230), C030046E11Rik (+106903)
chr19	34812509	34812659	Slc16a12 (+9195), Ifit1 (+97223)
chr19	44076755	44076987	Cpn1 (-15825), Cyp2c44 (+26866)
chr19	46654640	46655055	Wbp1l (-18756), Sfxn2 (+6993)

chr19	55776609	55776938	Tcf7l2 (-39600), Vti1a (+385934)
chr19	57565272	57565511	Atrnl1 (-120132), Trub1 (+37996)
chr19	58592967	58593225	Pnlip (-151759), Ccdc172 (+6604)
chr18	3004771	3005630	Vmn1r238 (+118264)
chr18	6543924	6544060	4921524L21Rik (-59639), Epc1 (-53138)
chr18	6544599	6544711	4921524L21Rik (-58976), Epc1 (-53801)
chr18	10151061	10151227	Usp14 (-121005), Rock1 (+30646)
chr18	10547960	10549005	Esco1 (+61867), Greb1l (+223306)
chr18	10569683	10569963	Esco1 (+40527), Greb1l (+244646)
chr18	12990093	12990353	Cabyr (+90390), Osbpl1a (+110127)
chr18	13201410	13201953	Hrh4 (+36183), Zfp521 (+929560)
chr18	14806362	14806473	Ss18 (+35005), Gm5160 (+224259)
chr18	15592764	15593169	Aqp4 (-31466), Chst9 (+283588)
chr18	18051314	18051426	NONE
chr18	24306003	24306426	Galnt1 (-57630), Ino80c (-25761)
chr18	24826363	24826870	Mocos (+14425), Gm9955 (+41232)
chr18	34602257	34602558	Pkd2l2 (+33331), Fam13b (+64069)
chr18	35803792	35803919	Slc23a1 (-16958), Mzb1 (+5165)
chr18	35815163	35815453	Prob1 (-467)
chr18	35815494	35815670	Prob1 (-741)
chr18	49829559	49829753	Dtwd2 (+85599)
chr18	51835086	51835240	Gm4950 (+190372), Prr16 (+557771)
chr18	57845035	57845163	Slc12a2 (-193233), 1700011I03Rik (+151665)
chr18	61541009	61541432	Ppargc1b (+18833), Pde6a (+161085)
chr18	66145885	66146203	Cplx4 (-16212), Lman1 (+16247)
chr18	75040100	75040260	Lipg (+80737), Acaa2 (+101329)
chr18	75040666	75040986	Lipg (+80091), Acaa2 (+101975)
chr18	76613456	76613637	Skor2 (-481597), Smad2 (+212313)
chr18	81797204	81797484	Sall3 (-614027), Galr1 (+778825)
chr18	83740362	83740551	Tshz1 (+515339), Zfp516 (+660186)
chr18	84588162	84588554	Zfp407 (+170538), Zadh2 (+330808)
chr18	87314713	87314808	Gm5096 (-610986), Cbln2 (+434259)
chrX	3302194	3302325	Gm14345 (-108408)
chrX	3376735	3376894	Gm14345 (-33853)
chrX	3512215	3512374	Gm14351 (-48975), Gm3701 (+155142)
chrX	3873121	3873280	Gm3701 (-205764), Gm14347 (-33810)
chrX	4041451	4041610	Gm3763 (-556635), Gm10922 (-45958)
chrX	25615314	25615470	Gm5168 (-137225), Gm2012 (+43089)
chrX	30762714	30762936	Gm21637 (-60425), Gm2799 (+164429)
chrX	31786255	31786414	Gm2927 (-45925), Gm2913 (-33869)
chrX	58586282	58586381	Ldoc1 (-376457), Cdr1 (-147599)
chrX	59903462	59903624	4931400O07Rik (+182377)
chrX	105879298	105879495	Gm732 (-735622), Brwd3 (+150290)
chrX	113580034	113580205	H2afb2-ps (-214667), Cpxcr1 (+17567)
chrX	115952722	115952852	Tgif2lx1 (-358441)
chrX	120113639	120114527	Nap1l3 (-603089), 3110007F17Rik (-103085)
chrX	120405249	120406500	Srsx (-69785), Vmn2r121 (+843644)
chrX	120634626	120635877	Srsx (-299162), Vmn2r121 (+614267)

chrX	120970533	120970966	Srsx (-634660), Vmn2r121 (+278769)
chrX	121577239	121578865	Vmn2r121 (-328533)
chrX	121706644	121708339	Vmn2r121 (-457973)
chrX	122458051	122459744	4932411N23Rik (+891718)
chrX	130684776	130684881	Nox1 (-38436), Xkrx (+11638)
chrX	136596335	136596619	E230019M04Rik (-941), Nup62cl (+630)
chrX	137431184	137431272	Tex13 (-83256), Vsig1 (-10919)
chrX	150514181	150514469	Gm15140 (+40903), Spin2 (+247489)
chrX	150565454	150565742	Gm15140 (-10370), 4930524N10Rik (+212337)
chrX	150587185	150587473	Gm15140 (-32101), 4930524N10Rik (+190606)
chrX	157555276	157555507	Cdkl5 (-122758), Scml2 (-45600)
chrX	159793558	159793691	S100g (-391094), Grpr (+193984)
chrX	166438853	166447041	4933400A11Rik (-225380)

2.1 Testes Peaks GO Term Associations

Negative Regulation of Megakaryocyte Differentiation

nearest gene	distance to TSS
Hist1h4a	-1012, +264
Hist1h4b	-4402, +280
Hist1h4c	113
Hist1h4f	-3571, +255
Hist1h4h	263
Hist1h4i	-2037, +161
Hist1h4j	186
Hist1h4k	-3382, +134
Hist1h4m	-1920, -1394, +244
Hist1h4n	-1920, -1393, +132

2.2 Testes Peaks GO Term Associations

Nucleosome Assembly

nearest gene	distance to TSS
Smarca5	175
Brd2	2370
H1fx	-88
H2afb2-ps	-214667
H2afx	37
H2afy	36985
H3f3b	3557
Hist1h1a	-2702, -1426
Hist1h1b	-2551
Hist1h1c	136
Hist1h1d	-3639, +187
Hist1h1e	203
Hist1h2aa	-422
Hist1h2ab	-4067
Hist1h2ac	-323
Hist1h2ae	-4759
Hist1h2af	-2593
Hist1h2ag	-440
Hist1h2ah	-508
Hist1h2ak	-480
Hist1h2al	208
Hist1h2an	-573
Hist1h2ba	116
Hist1h2bb	287
Hist1h2bc	72
Hist1h2be	-1175
Hist1h2bf	-1783
Hist1h2bj	-2023, +175
Hist1h2bk	330
Hist1h2bm	-4192, -311
Hist1h2bp	-4285, +330
Hist1h2bq	-1806, -167, +358
Hist1h2br	-1694, -169, +358
Hist1h3a	144
Hist1h3b	268
Hist1h3c	-1419
Hist1h3d	216
Hist1h3g	-4109
Hist1h3h	224
Hist1h3i	-4394, +221
Hist1h4a	-1012, +264
Hist1h4b	-4402, +280
Hist1h4c	113
Hist1h4f	-3571, +255
Hist1h4h	263
Hist1h4i	-2037, +161
Hist1h4j	186
Hist1h4k	-3882, +134
Hist1h4m	-1920, -1395, +244
Hist1h4n	-1920, -1393, +132
Hist2h2aa1	-3, +717
Hist2h2aa2	-26, +722

Hist2h2ab	869
Hist2h2ac	146
Hist2h2bb	-1042
Hist2h2be	-387
Hist2h3b	56
Hist2h3c1	-1216, -496, +141
Hist2h3c2	-1192, -444, +192
Nap1l3	-603089

2.3 Testes Peaks GO Term Associations

Nucleosome Organisation

nearest gene	distance to TSS
Brd2	2370
H1fx	-88
H2afb2-ps	-214667
H2afx	37
H2afy	36985
H3f3b	3557
Hist1h1a	-2702, -1426
Hist1h1b	-2551
Hist1h1c	136
Hist1h1d	-3639, +187
Hist1h1e	203
Hist1h2aa	-422
Hist1h2ab	-4067
Hist1h2ac	-323
Hist1h2ae	-4759
Hist1h2af	-2593
Hist1h2ag	-440
Hist1h2ah	-508
Hist1h2ak	-480
Hist1h2al	208
Hist1h2an	-573
Hist1h2ba	116
Hist1h2bb	287
Hist1h2bc	72
Hist1h2be	-1175
Hist1h2bf	-1783
Hist1h2bj	-2023, +175
Hist1h2bk	330
Hist1h2bm	-4192, -311
Hist1h2bp	-4285, +330
Hist1h2bq	-1806, -167, +358
Hist1h2br	-1694, -169, +358
Hist1h3a	144
Hist1h3b	268
Hist1h3c	-1419
Hist1h3d	216
Hist1h3g	-4109
Hist1h3h	224
Hist1h3i	-4394, +221
Hist1h4a	-1012, +264
Hist1h4b	-4402, +280
Hist1h4c	113
Hist1h4f	-3571, +255
Hist1h4h	263
Hist1h4i	-2037, +161
Hist1h4j	186
Hist1h4k	-3882, +134
Hist1h4m	-1920, -1395, +244
Hist1h4n	-1920, -1393, +132
Hist2h2aa1	-3, +717
Hist2h2aa2	-26, +722
Hist2h2ab	869

Hist2h2ac	146
Hist2h2bb	-1042
Hist2h2be	-387
Hist2h3b	56
Hist2h3c1	-1216, -496, +141
Hist2h3c2	-1192, -444, +192
Nap1l3	-603089
Ptma	-39253, -102
Setd2	-6843, -6083
Smarca5	175

2.4 Testes Peaks GO Term Associations

Chromatin Assembly

nearest gene	distance to TSS
Brd2	2370
H1fx	-88
H2afb2-ps	-214667
H2afx	37
H2afy	36985
H3f3b	3557
Hist1h1a	-2702, -1426
Hist1h1b	-2551
Hist1h1c	136
Hist1h1d	-3639, +187
Hist1h1e	203
Hist1h2aa	-422
Hist1h2ab	-4067
Hist1h2ac	-323
Hist1h2ae	-4759
Hist1h2af	-2593
Hist1h2ag	-440
Hist1h2ah	-508
Hist1h2ak	-480
Hist1h2al	208
Hist1h2an	-573
Hist1h2ba	116
Hist1h2bb	287
Hist1h2bc	72
Hist1h2be	-1175
Hist1h2bf	-1783
Hist1h2bj	-2023, +175
Hist1h2bk	330
Hist1h2bm	-4192, -311
Hist1h2bp	-4285, +330
Hist1h2bq	-1806, -167, +358
Hist1h2br	-1694, -169, +358
Hist1h3a	144
Hist1h3b	268
Hist1h3c	-1419
Hist1h3d	216
Hist1h3g	-4109
Hist1h3h	224
Hist1h3i	-4394, +221
Hist1h4a	-1012, +264
Hist1h4b	-4402, +280
Hist1h4c	113
Hist1h4f	-3571, +255
Hist1h4h	263
Hist1h4i	-2037, +161
Hist1h4j	186
Hist1h4k	-3882, +134
Hist1h4m	-1920, -1395, +244
Hist1h4n	-1920, -1393, +132
Hist2h2aa1	-3, +717
Hist2h2aa2	-26, +722
Hist2h2ab	869

Hist2h2ac	146
Hist2h2bb	-1042
Hist2h2be	-387
Hist2h3b	56
Hist2h3c1	-1216, -496, +141
Hist2h3c2	-1192, -444, +192
Nap1l3	-603089
Smarca5	175

2.5 Testes Peaks GO Term Associations

Chromatin Assembly or Dissassembly

nearest gene	distance to TSS
Brd2	2370
Chd1	108829
H1fx	-88
H2afb2-ps	-214667
H2afx	37
H2afy	36985
H3f3b	3557
Hist1h1a	-2702, -1426
Hist1h1b	-2551
Hist1h1c	136
Hist1h1d	-3639, +187
Hist1h1e	203
Hist1h2aa	-422
Hist1h2ab	-4067
Hist1h2ac	-323
Hist1h2ae	-4759
Hist1h2af	-2593
Hist1h2ag	-440
Hist1h2ah	-508
Hist1h2ak	-480
Hist1h2al	208
Hist1h2an	-573
Hist1h2ba	116
Hist1h2bb	287
Hist1h2bc	72
Hist1h2be	-1175
Hist1h2bf	-1783
Hist1h2bj	-2023, +175
Hist1h2bk	330
Hist1h2bm	-4192, -311
Hist1h2bp	-4285, -311
Hist1h2bq	-1806, +330
Hist1h2br	-1694, -167, +330
Hist1h3a	+144, -169, +330
Hist1h3b	268
Hist1h3c	-1419
Hist1h3d	216
Hist1h3g	-4109
Hist1h3h	224
Hist1h3i	-4394, +221
Hist1h4a	-1012, +264
Hist1h4b	-4402, +280
Hist1h4c	113
Hist1h4f	-3571, +255
Hist1h4h	263
Hist1h4i	-2037, +161
Hist1h4j	186
Hist1h4k	-3882, +134
Hist1h4m	-1920, +244, -1395
Hist1h4n	-1920, +132, -1393
Hist2h2aa1	-3, +717
Hist2h2aa2	-26, +722

Hist2h2ab	869
Hist2h2ac	146
Hist2h2bb	-1042
Hist2h2be	-387
Hist2h3b	56
Hist2h3c1	-1216, +141, -496
Hist2h3c2	-1192, +192, -444
Nap1l3	-603089
Smarca5	175
Suv39h2	35307

3 SOX Motif Regions

3.1 arACAAAGwa

chr	start	stop	p-value	matched_sequence
chr19	34812509	34812659	1.02E-07	AGAAACAAAGAA
chr9	59505862	59506188	1.02E-07	AGAAACAAAGAA
chr13	12366712	12366965	1.85E-07	AGAAACAAAGGA
chr4	101763903	101764474	4.10E-07	AGAAACAAAGTA
chr12	5503370	5503761	4.10E-07	AGAAACAAAGTA
chr2	121962909	121963298	6.97E-07	TGAAACAAAGAA
chrX	122458051	122459744	7.80E-07	AGAAACAAAGAC
chrX	121706644	121708339	7.80E-07	AGAAACAAAGAC
chrX	121577239	121578865	7.80E-07	AGAAACAAAGAC
chr12	113338413	113338652	1.16E-06	GGAAACAAAGGA
chr10	61591506	61591858	1.66E-06	GGAAACAAAGTA
chr17	13911136	13911387	1.66E-06	GGAAACAAAGTA
chr4	125568410	125568787	1.97E-06	tgaaacaaagta
chr4	3429671	3429955	2.35E-06	AGAAACAAAGTC
chr9	120058537	120058711	4.99E-06	TGAAACAAAGAG
chr1	192053952	192054194	6.99E-06	CCAAACAAAGGA
chr9	120058537	120058711	7.60E-06	GAAAACAAAGGC
chr3	93518698	93518941	7.85E-06	AGAAACAAAGTT
chr16	95979203	95979519	8.33E-06	TAAAACAAAGAG
chr17	13911136	13911387	9.64E-06	TCAAACAAAGAC
chr11	76148608	76148883	9.64E-06	GGAAACAAAGAT
chr1	136920453	136920843	1.31E-05	aataacaaagaa
chr9	41056175	41056341	1.34E-05	AGTAACAAAGTA
chr3	144408324	144408651	1.34E-05	AGTAACAAAGTA
chr3	144467409	144467757	1.34E-05	AGTAACAAAGTA
chr3	144408324	144408651	1.34E-05	AGTAACAAAGTA
chr3	144467409	144467757	1.34E-05	AGTAACAAAGTA
chr3	144408324	144408651	1.34E-05	AGTAACAAAGTA
chr3	144408324	144408651	1.34E-05	AGTAACAAAGTA
chr3	144408324	144408651	1.34E-05	AGTAACAAAGTA
chr3	144467409	144467757	1.34E-05	AGTAACAAAGTA
chr3	144408324	144408651	1.34E-05	AGTAACAAAGTA
chr3	144467409	144467757	1.34E-05	AGTAACAAAGTA
chr3	144408324	144408651	1.34E-05	AGTAACAAAGTA
chr3	144408324	144408651	1.34E-05	AGTAACAAAGTA
chr3	144467409	144467757	1.34E-05	AGTAACAAAGTA
chr11	121518509	121519059	1.49E-05	AGAAACAAAAAA
chr19	34812509	34812659	1.49E-05	AGAAACAAAAAA
chr10	89139305	89139721	1.49E-05	AGAAACAAAAAA
chr6	83530301	83530768	1.49E-05	AGAAACAAAAAA
chr12	5503370	5503761	1.70E-05	AGGAACAAAGAA
chr11	11397628	11398506	1.87E-05	GGTAACAAAGGA
chr11	87275408	87276317	1.91E-05	AGAAACAAACAA
chr11	87239984	87240835	1.94E-05	AAAAACAAAAAA
chr5	14918464	14922081	1.94E-05	AAAAACAAAAAA
chrX	122458051	122459744	1.94E-05	aaaaacaaaaaa
chrX	121706644	121708339	1.94E-05	aaaaacaaaaaa
chr16	95979203	95979519	2.02E-05	AGAAACAAAATA
chr4	125568410	125568787	2.02E-05	agaaacaaaata
chr14	120737186	120737476	2.07E-05	TCAAACAAAGCC
chr18	35803792	35803919	2.17E-05	AGAAAGAAAGAA
chr18	35803792	35803919	2.17E-05	AGAAAGAAAGAA

chr18	35803792	35803919	2.17E-05	AGAAAGAAAGAA
chr18	35803792	35803919	2.17E-05	AGAAAGAAAGAA
chr18	35803792	35803919	2.17E-05	AGAAAGAAAGAA
chr18	35803792	35803919	2.17E-05	AGAAAGAAAGAA
chr18	35803792	35803919	2.17E-05	AGAAAGAAAGAA
chr18	35803792	35803919	2.17E-05	AGAAAGAAAGAA
chr2	34611014	34611361	2.20E-05	TGTAACAAAGTA
chrX	166438853	166447041	2.32E-05	CCAAACAAAGCC
chrX	166438853	166447041	2.32E-05	CCAAACAAAGCC
chrX	166438853	166447041	2.32E-05	CCAAACAAAGCC
chrX	166438853	166447041	2.32E-05	CCAAACAAAGCC
chr11	121518509	121519059	2.45E-05	AGAAACAAAAAC
chr1	173012987	173014985	2.47E-05	AAAAACAAACAA
chr9	110428485	110429550	2.47E-05	aaaaacaaacaa
chr9	110428485	110429550	2.47E-05	AAAAACAAACAA
chrX	122458051	122459744	2.47E-05	aaaaacaaacaa
chrX	121706644	121708339	2.47E-05	aaaaacaaacaa
chr12	20828925	20829099	2.51E-05	ACAAACAAAAAA
chrY	2869581	2872285	2.51E-05	acaaacaaaaaa
chr2	146335546	146336272	2.51E-05	GGTAACAAAGAC
chr9	110428485	110429550	2.51E-05	acaaacaaaaAA
chr9	110428485	110429550	2.51E-05	ACAAACAAAAAA
chr11	87275408	87276317	2.51E-05	ACAAACAAAAAA
chrX	122458051	122459744	2.51E-05	acaaacaaaaaa
chrX	121706644	121708339	2.51E-05	acaaacaaaaaa
chr17	15950608	15950912	2.57E-05	TGCAACAAAGAA
chr15	25143986	25144140	2.60E-05	gctaacaaagaa
chr9	65049332	65049789	2.61E-05	AAAAACAAAATA
chr11	70729707	70729900	2.71E-05	AAGAACAAAGGA
chr12	112341390	112341756	2.71E-05	AAGAACAAAGGA
chr12	112341390	112341756	2.75E-05	TGGAACAAAGAA
chr11	69571960	69572257	2.77E-05	GGAAACAAAAGA
chr9	25059849	25060024	2.83E-05	aaaaagaaagaa
chr17	39979856	39985845	2.83E-05	aaaaagaaagaa
chr13	12366712	12366965	2.86E-05	AGAAACAATGTA
chr3	96290119	96291035	2.86E-05	AGAAACAATGTA
chr3	96292179	96293188	2.86E-05	AGAAACAATGTA
chr12	20828925	20829099	2.94E-05	GGAAACAAACAA
chr1	133716876	133717034	2.96E-05	AGAAACAAAAAG
chr17	15344966	15345538	2.96E-05	AGAAACAAAAAG
chr7	118226353	118226835	3.00E-05	gaaaacaaaaaa
chr11	11397628	11398506	3.00E-05	GGCAACAAAGGA
chr5	54893762	54893923	3.20E-05	agaaacatagaa
chr14	54880915	54881135	3.26E-05	AGCAACAAAGAG
chr4	154231384	154231576	3.26E-05	AGCAACAAAGAG
chrX	120405249	120406500	3.35E-05	AAAAACAAAAAC
chrX	120634626	120635877	3.35E-05	AAAAACAAAAAC
chr1	180252156	180252389	3.35E-05	AAAAACAAAAAC
chrX	122458051	122459744	3.35E-05	aaaaacaaaaac
chrX	121706644	121708339	3.35E-05	aaaaacaaaaac
chr12	20828925	20829099	3.39E-05	ACAAACAAACAA
chrY	2869581	2872285	3.39E-05	acaaacaaacaa
chrY	2869581	2872285	3.39E-05	acaaacaaacaa
chrY	2869581	2872285	3.39E-05	acaaacaaacaa
chrY	2869581	2872285	3.39E-05	acaaacaaacaa
chr15	5099576	5099919	3.39E-05	ACAAACAAACAA

chr15	5099576	5099919	3.39E-05	ACAAACAAACAA
chr15	5099576	5099919	3.39E-05	ACAAACAAACAA
chr1	173012987	173014985	3.39E-05	ACAAACAAACAA
chr1	173012987	173014985	3.39E-05	ACAAACAAACAA
chr1	173012987	173014985	3.39E-05	ACAAACAAACAA
chr11	11397628	11398506	3.39E-05	acaaacaaacaa
chr11	11397628	11398506	3.39E-05	acaaacaaacaa
chr11	11397628	11398506	3.39E-05	acaaacaaacaa
chr11	101466103	101466410	3.48E-05	TGAGACAAAGAA
chr11	87239984	87240835	3.48E-05	TGAGACAAAGAA
chr7	108458149	108458626	3.53E-05	AGACACAAAGTA
chr7	108458149	108458626	3.53E-05	AGACACAAAGTA
chr7	108458149	108458626	3.53E-05	AGACACAAAGTA
chr7	108458149	108458626	3.53E-05	AGACACAAAGTA
chr7	108458149	108458626	3.53E-05	AGACACAAAGTA
chr7	108458149	108458626	3.53E-05	AGACACAAAGTA
chr13	21809473	21809969	3.53E-05	AAAAACACAGAA
chr9	25059849	25060024	3.56E-05	agaaaaaaagaa
chrX	30762714	30762936	3.56E-05	agaaaaaaagaa
chr15	38212163	38212414	3.56E-05	AGAAAAAAAGAA
chr17	80606473	80606939	3.56E-05	AGAAAAAAAGAA
chr17	15950608	15950912	3.59E-05	TGAAACAATGAA
chr5	113763814	113764285	3.62E-05	AGGAACAAAGGC
chrX	122458051	122459744	3.62E-05	AAAAAGAAAGGA
chrX	121706644	121708339	3.62E-05	AAAAAGAAAGGA
chrX	121577239	121578865	3.62E-05	AAAAAGAAAGGA
chr13	12366712	12366965	3.68E-05	AGGAACAAAGCA
chr15	101724635	101724846	3.70E-05	TGCAACAAAGTA
chr4	123369561	123369835	3.70E-05	TGCAACAAAGTA
chr17	75509211	75509527	3.70E-05	TGCAACAAAGTA
chr18	3004771	3005630	3.86E-05	AGAAAGAAAGAC
chr12	93036365	93036569	3.86E-05	TAGAACAAAGAA
chr13	52846109	52846573	3.94E-05	AGTAACAAAGGT
chr19	44076755	44076987	3.97E-05	aaagacaaagta
chr19	44076755	44076987	3.97E-05	AAAGACAAAGTA
chr1	192053952	192054194	4.03E-05	AGAAACAGAGAA
chr5	11220304	11220562	4.08E-05	ggaaacaaaaac
chr5	11881170	11881428	4.08E-05	ggaaacaaaaac
chr8	85884272	85884590	4.15E-05	AAAAAGAAAGTA
chr8	127528426	127529090	4.18E-05	GGAGACAAAGGA
chr8	127528426	127529090	4.18E-05	GGAGACAAAGGA
chr11	86887233	86887808	4.24E-05	AGCAACAAAGGG
chr17	15098409	15099907	4.24E-05	AAAAACAAAAAG
chr17	15161360	15162859	4.24E-05	AAAAACAAAAAG
chr17	15129992	15131491	4.24E-05	AAAAACAAAAAG
chr18	84588162	84588554	4.40E-05	AAAAACAAAACA
chr16	84647990	84648749	4.40E-05	AAAAACAAAACA
chr1	180252156	180252389	4.40E-05	AAAAACAAAACA
chr17	15344966	15345538	4.40E-05	AAAAACAAAACA
chr5	123453058	123453399	4.47E-05	CAGAACAAAGAA
chr17	87452962	87453212	4.49E-05	AGAAATAAAGAA
chr7	5543821	5543962	4.49E-05	AGAAATAAAGAA
chrX	120634626	120635877	4.49E-05	AGAAATAAAGAA
chrX	120405249	120406500	4.49E-05	AGAAATAAAGAA
chr4	17379647	17379762	4.57E-05	AGAAAAAAAGGA
chr4	125568410	125568787	4.63E-05	aaaaacatagaa

chr9	41056175	41056341	4.65E-05	AGAGACAAAGAG
chr16	84647990	84648749	4.69E-05	ataaacaaaCAA
chr14	69799171	69799572	4.74E-05	GCCAACAAAGAA
chr16	84647990	84648749	4.83E-05	acaaaCAAAAAC
chr11	87239984	87240835	4.86E-05	ACACACAAAGAA
chr18	35803792	35803919	4.90E-05	AGAAAGAAAGAG
chr16	84647990	84648749	4.96E-05	CAAAACAAAAGA
chr9	120058537	120058711	4.96E-05	GGGAACAAAGAC
chr15	88568772	88569185	4.96E-05	CAAAACAAAAGA
chr14	120737186	120737476	4.96E-05	tttaacaAAGGA
chr13	66394389	66396325	5.03E-05	tagaacaaagga
chr7	108458149	108458626	5.11E-05	ACAAACAATGGA
chr13	66394389	66396325	5.15E-05	ataaacaaaata
chr9	25059849	25060024	5.20E-05	aaaaaaaaagaa
chr13	9833500	9833762	5.20E-05	AAAAAAAAAGAA
chr7	16994559	16994776	5.20E-05	AAAAAAAAAGAA
chrX	30762714	30762936	5.20E-05	aaaaaaaaagaa
chrY	2869581	2872285	5.20E-05	aaaaaaaaagaa
chr17	39979856	39985845	5.20E-05	aaaaaaaaagaa
chr17	39979856	39985845	5.20E-05	aaaaaaaaagaa
chr11	101466103	101466410	5.25E-05	ACAAAGAAAGGA
chr16	34671860	34672520	5.25E-05	AGAAACAAAAAT
chr9	75471052	75471301	5.27E-05	AAGAACAAAGGC
chr9	110989678	110990024	5.27E-05	AAGAACAAAGGC
chr12	5503370	5503761	5.32E-05	AGAAACAGAGGA
chr13	21842001	21842479	5.39E-05	AAAGACAAAGAC
chr3	96290119	96291035	5.39E-05	CAAAACAAACAA
chr2	98506338	98507542	5.39E-05	TAAAAGAAAGAA
chr12	5503370	5503761	5.39E-05	AAGAACAAAGCA
chr13	65967440	65968021	5.39E-05	TAAAAGAAAGAA
chr17	15344966	15345538	5.39E-05	CAAAACAAACAA
chr13	65552554	65553189	5.39E-05	taaaagaaagaa
chr3	96292179	96293188	5.39E-05	CAAAACAAACAA
chr17	14847569	14847759	5.47E-05	TGAAACAATGTA
chr12	55835205	55835592	5.47E-05	ACCAACAAAGAc
chr12	55814849	55815664	5.47E-05	ACCAACAAAGAC
chr3	58469472	58469815	5.51E-05	AAAAACAATGAC
chr7	5543821	5543962	5.66E-05	ATAAAGAAAGAA
chr11	121518509	121519059	5.66E-05	ATAAAGAAAGAA
chr11	32076000	32076214	5.66E-05	aaaaacaaaagg
chr9	120058537	120058711	5.73E-05	TGAAACAAAAAG
chr7	39352222	39353766	5.85E-05	AGCAACAAAGAT
chr7	39231479	39233273	5.85E-05	AGCAACAAAGAT
chr9	118647931	118648304	5.88E-05	TGTAACAAAGAT
chr8	83262886	83263555	6.01E-05	GGCAACAAAGGC
chr12	5503370	5503761	6.07E-05	ACAAAGAAAGTA
chrY	2869581	2872285	6.21E-05	ACAAACAAAAAG
chr3	58469948	58470228	6.29E-05	GGAAACAAAATC
chr16	34671860	34672520	6.29E-05	GCCAACAAAGGA
chr4	132994761	132995022	6.69E-05	AAAAATAAAGAA
chr4	149612437	149612864	6.69E-05	TAAGACAAAGGA
chr17	87452962	87453212	6.73E-05	AATAACAAAGTT
chr15	5099576	5099919	6.96E-05	ACAAACAAACAC
chr7	31306647	31307327	7.04E-05	ACCAACAAAGAG
chr11	32076000	32076214	7.12E-05	tctaacaaagag
chr9	59505862	59506188	7.12E-05	AAAGACAAAGGC

chr1	173012987	173014985	7.16E-05	ACAAACAAATAA
chr3	68096244	68096863	7.47E-05	AAAAGCAAAGAA
chr13	51941709	51941925	7.50E-05	AAAAAGAAAGGC
chr1	194061545	194061791	7.50E-05	AAAAAGAAAGGC
chr3	23209200	23209294	7.65E-05	tgaaacaaaagg
chr13	51941709	51941925	7.65E-05	ACAAAAAAAGAA
chr11	87239984	87240835	7.65E-05	ACAAAAAAAGAA
chr2	98505008	98505437	7.70E-05	gaaaagaaagta
chr2	121962909	121963298	7.81E-05	TTTAACAAAGAC
chr11	11397628	11398506	7.92E-05	GGCAACAAAGGG
chr11	11397628	11398506	7.92E-05	ccaaacaaacaa
chr13	51941709	51941925	8.00E-05	AAAAAAAAAGTA
chr9	82869297	82869515	8.00E-05	GGAAACAAATGA
chr7	108458149	108458626	8.00E-05	TCCAACAAAGTA
chr5	14918464	14922081	8.14E-05	cgaaacaaaatc
chr1	169328474	169328599	8.49E-05	TAAAACAAAAAG
chr4	123369561	123369835	8.49E-05	GGGAACAAAGGG
chr15	88568772	88569185	8.49E-05	GGAAAAAAAGGA
chr3	96290119	96291035	8.89E-05	AGAAACTAAGAA
chr11	32076000	32076214	8.89E-05	aaaaataaagga
chr3	96292179	96293188	8.89E-05	AGAAACTAAGAA
chr5	14918464	14922081	8.95E-05	AAACACAAAGAG
chr16	36991075	36991339	9.08E-05	TGTAACAAagtt
chr7	120444291	120444522	9.17E-05	TGAAAAAAAGGA
chr7	120444291	120444522	9.17E-05	TGAAAAAAAGGA
chr7	120444291	120444522	9.17E-05	TGAAAAAAAGGA
chr7	120444291	120444522	9.17E-05	TGAAAAAAAGGA
chr7	120444291	120444522	9.17E-05	TGAAAAAAAGGA
chr12	19226266	19226526	9.30E-05	ATAAACAAAACA
chr13	56199754	56200260	9.41E-05	AAAAACAATGGG
chr13	56199754	56200260	9.41E-05	AAAAACAATGGG
chr13	56199754	56200260	9.41E-05	AAAAACAATGGG
chrY	2869581	2872285	9.41E-05	GCAAACACAGAA
chr16	34671860	34672520	9.49E-05	AGAAAAAAAGAG
chr3	52059270	52059807	9.49E-05	ttaaacaaacaa
chr17	39979856	39985845	9.58E-05	gaaaaaaaagaa
chr18	3004771	3005630	9.69E-05	AAAAAGAAAGGG
chr17	14847569	14847759	9.93E-05	TGAAACAATGCA

3 SOX Motif Regions

3.2 drAAACAATGka

chr	start	stop	p-value	matched_sequence
chr13	12366712	12366965	1.02E-07	AGAAACAATGTA
chr3	96292179	96293188	1.02E-07	AGAAACAATGTA
chr3	96290119	96291035	1.02E-07	AGAAACAATGTA
chr17	14847569	14847759	4.38E-07	TGAAACAATGTA
chr17	15950608	15950912	1.76E-06	TGAAACAATGAA
chr7	108458149	108458626	3.04E-06	ACAAACAATGGA
chr17	14847569	14847759	3.79E-06	TGAAACAATGCA
chr13	56199754	56200260	4.84E-06	AAAAACAATGGG
chr13	56199754	56200260	4.84E-06	AAAAACAATGGG
chr16	84975160	84975506	4.84E-06	AAAAACAATGTT
chr13	56199754	56200260	4.84E-06	AAAAACAATGGG
chr12	112341390	112341756	5.62E-06	GAAAACAATGTT
chr3	144408324	144408651	7.22E-06	AGTAACAATGTA
chr3	58469472	58469815	9.19E-06	AAAAACAATGAC
chr3	69490627	69490976	9.19E-06	CCAAACAATGGA
chr11	22170490	22170823	1.03E-05	agcaacaatgta
chr11	22170490	22170823	1.03E-05	agcaacaatgta
chr11	22170490	22170823	1.03E-05	agcaacaatgta
chr19	57565272	57565511	1.03E-05	AGCAACAATGTA
chr11	22170490	22170823	1.03E-05	agcaacaatgta
chr11	22170490	22170823	1.03E-05	agcaacaatgta
chr11	22170490	22170823	1.03E-05	agcaacaatgta
chr11	22170490	22170823	1.03E-05	agcaacaatgta
chr11	22170490	22170823	1.03E-05	agcaacaatgta
chr11	22170490	22170823	1.03E-05	agcaACAATGTA
chr15	87797353	87797527	1.05E-05	GAAAACAATGAC
chr14	121483899	121484504	1.09E-05	GCAAACAATGGC
chr14	121483899	121484504	1.09E-05	GCAAACAATGGC
chr14	121483899	121484504	1.09E-05	GCAAACAATGGC
chr14	121483899	121484504	1.09E-05	GCAAACAATGGC
chr14	121483899	121484504	1.09E-05	GCAAACAATGGC
chr14	121483899	121484504	1.09E-05	GCAAACAATGGC
chr14	121483899	121484504	1.09E-05	GCAAACAATGGC
chr14	121483899	121484504	1.09E-05	GCAAACAATGGC
chr14	121483899	121484504	1.09E-05	GCAAACAATGGC
chr3	152996314	152997030	1.18E-05	AGATACAATGTA
chr3	152996314	152997030	1.18E-05	AGATACAATGTA
chr19	44076755	44076987	1.18E-05	AGATACAATGTA
chr3	152996314	152997030	1.18E-05	AGATACAATGTA
chr19	44076755	44076987	1.18E-05	AGATACAATGTA
chr3	152996314	152997030	1.18E-05	AGATACAATGTA
chr3	152996314	152997030	1.18E-05	AGATACAATGTA
chr3	152996314	152997030	1.18E-05	AGATACAATGTA
chr3	152996314	152997030	1.18E-05	AGATACAATGTA
chr3	152996314	152997030	1.18E-05	AGATACAATGTA
chr4	125568410	125568787	1.25E-05	CAAAACAATGTT
chr7	91561750	91562035	1.31E-05	AGATACAATGGA
chr7	91561750	91562035	1.31E-05	AGATACAATGGA
chr7	91561750	91562035	1.31E-05	AGATACAATGGA
chr7	91561750	91562035	1.31E-05	AGATACAATGGA
chr7	91561750	91562035	1.31E-05	AGATACAATGGA
chr9	49183419	49183597	1.33E-05	GGATACAATGTA

chr4	101763903	101764474	1.33E-05	AGAAACAGTGTA
chr4	101763903	101764474	1.33E-05	AGAAACAGTGTA
chr4	101763903	101764474	1.33E-05	AGAAACAGTGTA
chr4	101763903	101764474	1.33E-05	AGAAACAGTGTA
chr4	101763903	101764474	1.33E-05	AGAAACAGTGTA
chr4	101763903	101764474	1.33E-05	AGAAACAGTGTA
chr4	101763903	101764474	1.33E-05	AGAAACAGTGTA
chr16	95979203	95979519	1.37E-05	TGCAACAATGTA
chr16	95979203	95979519	1.49E-05	TGCAACAATGGA
chr8	119567477	119567793	1.52E-05	TGATACAATGTA
chr3	68096244	68096863	1.52E-05	TGATACAATGTA
chr8	31438279	31438743	1.75E-05	AACAACAATGGA
chr2	57246848	57247325	1.94E-05	AGTAACAATGGC
chr2	57246848	57247325	1.94E-05	AGTAACAATGGC
chr2	57246848	57247325	1.94E-05	AGTAACAATGGC
chr2	57246848	57247325	1.94E-05	AGTAACAATGGC
chr4	8154834	8155634	1.97E-05	gacaacaatgga
chr8	31438279	31438743	2.00E-05	AGTAACAATGGG
chr11	108872866	108873417	2.05E-05	TGGAACAATGTA
chr7	108458149	108458626	2.11E-05	AGACACAATGTA
chr13	52846109	52846573	2.21E-05	GGTAACAATGGC
chr8	127528426	127529090	2.27E-05	TGGAACAATGGA
chr10	89139305	89139721	2.27E-05	TGGAACAATGGA
chr8	127528426	127529090	2.27E-05	TGGAACAATGGA
chr10	9467362	9467556	2.36E-05	CAAAACAATGAT
chrY	2869581	2872285	2.40E-05	GGACACAATGTA
chr19	57565272	57565511	2.45E-05	CCAAACAATGGT
chr9	110989678	110990024	2.47E-05	GGATACAATGAA
chr17	15950608	15950912	2.51E-05	AGAAACAGTGAA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA

chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr6	116201111	116202043	2.52E-05	GAAAACAGTGGA
chr16	84975160	84975506	2.63E-05	AGCAACAATGTT
chr16	84975160	84975506	2.63E-05	AGCAACAATGTT
chr4	123369561	123369835	2.68E-05	agatacaatgtg
chr18	81797204	81797484	2.69E-05	AGAAACACTGTA
chr18	81797204	81797484	2.69E-05	AGAAACACTGTA
chr18	81797204	81797484	2.69E-05	AGAAACACTGTA
chr16	95979203	95979519	2.72E-05	TGTAACAATGCA
chr15	87797353	87797527	2.89E-05	GGCAACAATGTT
chr17	33811718	33812117	2.89E-05	GGCAACAATGTT
chr7	91561750	91562035	2.89E-05	AGATACAATGTT
chr4	3429671	3429955	2.97E-05	GAGAACAATGGA
chr6	83530301	83530768	3.17E-05	AAACACAATGTA
chr16	36991075	36991339	3.22E-05	GGATACAATGTT
chr19	21581779	21582122	3.36E-05	TGCAACAATGTT
chr15	82725119	82725439	3.40E-05	GATAACAATGTT
chr3	146005710	146006075	3.40E-05	TGATACAATGTG
chr9	19029130	19029345	3.74E-05	GTAAACAATGAT
chr1	192053952	192054194	3.79E-05	TGGAACAATGAA
chr1	88617661	88617989	3.85E-05	TATAACAATGGC
chr1	88617661	88617989	3.85E-05	TATAACAATGGC
chr1	88617661	88617989	3.85E-05	TATAACAATGGC
chr1	88617661	88617989	3.85E-05	TATAACAATGGC
chr17	15950608	15950912	3.85E-05	TACAACAATGAA
chr1	88617661	88617989	3.85E-05	TATAACAATGGC
chr1	88617661	88617989	3.85E-05	TATAACAATGGC
chr1	88617661	88617989	3.85E-05	TATAACAATGGC
chr17	84581105	84581330	3.98E-05	AGGAACAATGGG
chr15	88568772	88569185	3.98E-05	TTAAACAATGAG
chr5	28661168	28661314	4.04E-05	AACAACAATGTT
chr5	28661168	28661314	4.04E-05	AACAACAATGTT
chr5	28661168	28661314	4.04E-05	AACAACAATGTT
chr5	28661168	28661314	4.04E-05	AACAACAATGTT
chr11	87261721	87262640	4.04E-05	AACAACAATGTT
chr1	88422999	88423418	4.17E-05	AGGAACAATGCA
chr8	31438279	31438743	4.18E-05	AACAACAATGCA
chr12	19226266	19226526	4.29E-05	TGAAACAGTGTT
chr4	114943768	114943896	4.37E-05	ggacacaatgaa
chr4	114986056	114986184	4.37E-05	GGACACAATGAA
chr1	72290738	72291074	4.41E-05	GGGAACAATGTT
chr16	84975160	84975506	4.43E-05	AACAACAATGGT
chr5	72438606	72438857	4.49E-05	GAATACAATGTG
chr16	84975160	84975506	4.55E-05	AAGAACAATGAA
chr16	22672855	22673155	4.57E-05	GGAAAAAATGGA
chr3	124418625	124418909	4.82E-05	AGCAACAATGAT
chr3	124418625	124418909	4.82E-05	AGCAACAATGAT
chr3	124418625	124418909	4.82E-05	AGCAACAATGAT
chr3	124418625	124418909	4.82E-05	AGCAACAATGAT
chr3	124418625	124418909	4.82E-05	AGCAACAATGAT
chr18	76613456	76613637	4.87E-05	AGATACAATGAG
chr5	14918464	14922081	4.91E-05	AGAAACACTGAA

chr12	20828925	20829099	4.94E-05	GAAAACACTGGA
chr4	101763903	101764474	5.05E-05	AGAAACAAAGTA
chr12	5503370	5503761	5.05E-05	AGAAACAAAGTA
chr10	12273996	12274342	5.46E-05	ACACACAATGTA
chr15	11259966	11260124	5.53E-05	ATATACAATGTA
chr13	12366712	12366965	5.53E-05	AGAAACAAAGGA
chr10	61591506	61591858	5.59E-05	GGAAACAAAGTA
chr17	13911136	13911387	5.59E-05	GGAAACAAAGTA
chr8	25462270	25462920	5.59E-05	GAAAACAGTGTT
chr1	136920453	136920843	5.69E-05	CAGAACAATGGA
chr17	87452962	87453212	5.69E-05	CAGAACAATGGA
chr10	4930547	4930938	5.71E-05	GGACACAATGTT
chrY	2869581	2872285	5.71E-05	GGACACAATGGG
chr14	54880915	54881135	5.84E-05	AAGAACAATGGC
chr11	87236084	87236908	5.84E-05	AAGAACAATGGC
chr18	81797204	81797484	5.87E-05	AGAAACACTGTG
chr12	113338413	113338652	6.13E-05	GGAAACAAAGGA
chr11	121518509	121519059	6.32E-05	TTCAACAATGTA
chr4	125568410	125568787	6.39E-05	tgaaacaaagta
chr17	15950608	15950912	6.60E-05	TCCAACAATGAA
chr3	96163525	96164181	6.64E-05	TCTAACAATGTT
chr3	96177802	96178458	6.64E-05	TCTAACAATGTT
chr3	146005710	146006075	6.70E-05	GCACACAATGGA
chr3	93518698	93518941	6.86E-05	ggaaacaatata
chr13	70231949	70232253	7.21E-05	TGAAACATTGTA
chr2	22443076	22443924	7.35E-05	TGAAACACTGTC
chr1	24618332	24623038	7.35E-05	TGAAACACTGTC
chr9	41056175	41056341	7.40E-05	CGGAACAATGTG
chr9	41056175	41056341	7.75E-05	GTTAACAATGAA
chr5	11829495	11829872	7.75E-05	TTATACAATGGA
chr5	11400416	11400736	7.75E-05	TTATACAATGGA
chr2	162688176	162688554	7.79E-05	agctacaatgta
chr2	131233744	131234170	7.84E-05	AGAAACAATTTA
chr9	110989678	110990024	8.77E-05	TTTAACAATGAA
chr1	141954325	141954737	8.90E-05	AATTACAATGTA
chr19	34812509	34812659	8.95E-05	AGAAACAAAGAA
chr9	59505862	59506188	8.95E-05	AGAAACAAAGAA
chr6	83417694	83418110	9.02E-05	GCATACAATGGT
chr13	52846109	52846573	9.21E-05	GAAAACATTGTA
chr14	118206572	118206838	9.26E-05	AGAAACAAGGGA
chr3	68096244	68096863	9.42E-05	AAAATCAATGGA
chr5	143579353	143579517	9.68E-05	GGGAACAATGCG
chr18	84588162	84588554	9.73E-05	TGCTACAATGTA
chr5	16974433	16975269	9.73E-05	TGCTACAATGTA
chr5	14918464	14922081	9.73E-05	tgctacaatgta

3 SOX Motif Regions

3.3 awACAAAGwa

chr	start	stop	p-value	matched_sequence
chr19	44076755	44076987	2.69E-06	agacaaagta
chr19	44076755	44076987	2.69E-06	AGACAAAGTA
chr7	108458149	108458626	2.69E-06	ACACAAAGTA
chr7	108458149	108458626	2.69E-06	ACACAAAGTA
chr7	108458149	108458626	2.69E-06	ACACAAAGTA
chr7	108458149	108458626	2.69E-06	ACACAAAGTA
chr7	108458149	108458626	2.69E-06	ACACAAAGTA
chr7	108458149	108458626	2.69E-06	ACACAAAGTA
chr4	101763903	101764474	4.34E-06	AAACAAAGTA
chr10	61591506	61591858	4.34E-06	AAACAAAGTA
chr12	5503370	5503761	4.34E-06	AAACAAAGTA
chr17	13911136	13911387	4.34E-06	AAACAAAGTA
chr4	125568410	125568787	4.34E-06	aaacaaagta
chr17	10380494	10380692	5.98E-06	ATACAAAGTA
chr10	80827990	80829989	8.19E-06	CCACAAAGTA
chr15	101724635	101724846	9.54E-06	CAACAAAGTA
chr4	123369561	123369835	9.54E-06	CAACAAAGTA
chr7	108458149	108458626	9.54E-06	CAACAAAGTA
chr17	75509211	75509527	9.54E-06	CAACAAAGTA
chr18	84588162	84588554	1.09E-05	CTACAAAGTA
chr14	121483899	121484504	1.09E-05	CTACAAAGTA
chr14	121483899	121484504	1.09E-05	CTACAAAGTA
chr14	121483899	121484504	1.09E-05	CTACAAAGTA
chr14	121483899	121484504	1.09E-05	CTACAAAGTA
chr14	121483899	121484504	1.09E-05	CTACAAAGTA
chr14	121483899	121484504	1.09E-05	CTACAAAGTA
chr11	101466103	101466410	1.36E-05	AGACAAAGAA
chr4	8154834	8155634	1.36E-05	agacaaagaa
chrX	113580034	113580205	1.36E-05	ACACAAAGAA
chr4	8154834	8155634	1.36E-05	agacaaagaa
chr4	8154834	8155634	1.36E-05	agacaaagaa
chr4	8154834	8155634	1.36E-05	agacaaagaa
chr4	8154834	8155634	1.36E-05	agacaaagaa
chr11	87239984	87240835	1.36E-05	AGACAAAGAA
chr4	8154834	8155634	1.36E-05	agacaaagaa
chr4	8154834	8155634	1.36E-05	agacaaagaa
chr4	8154834	8155634	1.36E-05	agacaaagaa
chr4	8154834	8155634	1.36E-05	agacaaagaa
chr11	87239984	87240835	1.36E-05	ACACAAAGAA
chr11	87275408	87276317	1.36E-05	ACACAAAGAA
chr19	34812509	34812659	1.52E-05	AAACAAAGAA
chr9	59505862	59506188	1.52E-05	AAACAAAGAA
chr2	121962909	121963298	1.52E-05	AAACAAAGAA
chrX	30762714	30762936	1.69E-05	atacaaagaa
chr19	57565272	57565511	1.69E-05	ATACAAAGAA
chr6	145881183	145881640	1.91E-05	GGACAAAGTA
chr8	127528426	127529090	2.40E-05	AGACAAAGGA
chr8	127528426	127529090	2.40E-05	AGACAAAGGA
chr4	149612437	149612864	2.40E-05	AGACAAAGGA
chr1	136457440	136457676	2.40E-05	ACACAAAGGA
chr1	192053952	192054194	2.53E-05	AAACAAAGGA
chr12	113338413	113338652	2.53E-05	AAACAAAGGA

chr13	12366712	12366965	2.53E-05	AAACAAAGGA
chr15	5099576	5099919	2.89E-05	CCACAAAGAA
chrX	120634626	120635877	2.89E-05	CCACAAAGAA
chrX	120405249	120406500	2.89E-05	CCACAAAGAA
chr14	69799171	69799572	3.02E-05	CAACAAAGAA
chr17	15950608	15950912	3.02E-05	CAACAAAGAA
chr9	75471052	75471301	3.16E-05	CTACAAAGAA
chr3	52059270	52059807	3.43E-05	TCACAAAGTA
chr15	101724635	101724846	3.43E-05	tcacaaagta
chr6	83417694	83418110	3.43E-05	TGACAAAGTA
chrX	120405249	120406500	3.43E-05	TCACAAAGTA
chrX	120634626	120635877	3.43E-05	TCACAAAGTA
chr13	66394389	66396325	3.43E-05	TCACAAAGTA
chr3	144408324	144408651	3.59E-05	TAACAAAGTA
chr3	144467409	144467757	3.59E-05	TAACAAAGTA
chr3	144408324	144408651	3.59E-05	TAACAAAGTA
chr3	144467409	144467757	3.59E-05	TAACAAAGTA
chr2	34611014	34611361	3.59E-05	TAACAAAGTA
chr9	41056175	41056341	3.59E-05	TAACAAAGTA
chr3	144408324	144408651	3.59E-05	TAACAAAGTA
chr3	144408324	144408651	3.59E-05	TAACAAAGTA
chr3	144408324	144408651	3.59E-05	TAACAAAGTA
chr3	144467409	144467757	3.59E-05	TAACAAAGTA
chr3	144408324	144408651	3.59E-05	TAACAAAGTA
chr3	144467409	144467757	3.59E-05	TAACAAAGTA
chr3	144408324	144408651	3.59E-05	TAACAAAGTA
chr3	144408324	144408651	3.59E-05	TAACAAAGTA
chr3	144467409	144467757	3.59E-05	TAACAAAGTA
chr11	105301298	105301520	3.76E-05	TTACAAAGTA
chr9	59505862	59506188	3.76E-05	TTACAAAGTA
chr16	84647990	84648749	3.76E-05	TTACAAAGTA
chr4	101763903	101764474	3.98E-05	GGACAAAGAA
chr11	11397628	11398506	3.98E-05	GGACAAAGAA
chr7	108789576	108789755	3.98E-05	GGACAAAGAA
chr16	84647990	84648749	3.98E-05	GCACAAAGAA
chr5	14918464	14922081	3.98E-05	gcacaaagaa
chr17	80606473	80606939	3.98E-05	GCACAAAGAA
chr6	83417694	83418110	3.98E-05	GGACAAAGAA
chr11	76191172	76191993	3.98E-05	GCACAAAGAA
chr5	123453058	123453399	4.11E-05	GAACAAAGAA
chr12	93036365	93036569	4.11E-05	GAACAAAGAA
chr12	112341390	112341756	4.11E-05	GAACAAAGAA
chr12	5503370	5503761	4.11E-05	GAACAAAGAA
chr11	76148608	76148883	4.43E-05	ccacaaagga
chr11	76191172	76191993	4.43E-05	ccacaaagga
chr11	76148608	76148883	4.43E-05	CCACAAAGGA
chr14	57315836	57316043	4.43E-05	CCACAAAGGA
chr11	11397628	11398506	4.54E-05	CAACAAAGGA
chr16	34671860	34672520	4.54E-05	CAACAAAGGA
chr4	153394885	153395163	4.65E-05	CTACAAAGGA
chr2	80302072	80302264	4.65E-05	CTACAAAGGA
chr5	66330250	66330572	4.83E-05	GGACAAAGGA
chr5	108295958	108296386	4.83E-05	ggacaaagga
chr11	76191172	76191993	4.83E-05	gcacaaagga
chr11	76191172	76191993	4.83E-05	GCACAAAGGA
chr11	70729707	70729900	4.94E-05	GAACAAAGGA

chr12	112341390	112341756	4.94E-05	GAACAAAGGA
chr13	66394389	66396325	4.94E-05	gaacaaagga
chr16	72258386	72258542	5.05E-05	GTACAAAGGA
chr12	112341390	112341756	5.05E-05	GTACAAAGGA
chr11	76191172	76191993	5.05E-05	GTACAAAGGA
chrX	113580034	113580205	5.32E-05	TGACAAAGAA
chrX	122458051	122459744	5.32E-05	TGACAAAGAA
chrX	121706644	121708339	5.32E-05	TGACAAAGAA
chrX	121577239	121578865	5.32E-05	TGACAAAGAA
chr14	54880915	54881135	5.32E-05	TGACAAAGAA
chr11	108872866	108873417	5.32E-05	tcacaaagaa
chr1	136920453	136920843	5.49E-05	taacaaagaa
chr15	25143986	25144140	5.49E-05	taacaaagaa
chr2	3356844	3357148	5.87E-05	ACACAAAGTC
chr2	3356844	3357148	5.87E-05	ACACAAAGTC
chr2	3356844	3357148	5.87E-05	ACACAAAGTC
chr2	3356844	3357148	5.87E-05	ACACAAAGTC
chr2	3356844	3357148	5.87E-05	ACACAAAGTC
chr2	3356844	3357148	5.87E-05	ACACAAAGTC
chr4	3429671	3429955	6.01E-05	AAACAAAGTC
chr13	24025614	24026204	6.36E-05	TCACAAAGGA
chr14	120737186	120737476	6.50E-05	taacaAAGGA
chr11	11397628	11398506	6.50E-05	TAACAAAGGA
chr17	71834308	71834443	6.85E-05	acacaaagtg
chr15	38212163	38212414	6.85E-05	AGACAAAGTG
chr9	75471052	75471301	6.85E-05	AGACAAAGTG
chr12	55838910	55839950	7.30E-05	CCACAAAGTC
chr19	29703448	29703901	7.30E-05	CCACAAAGTC
chr19	29703448	29703901	7.30E-05	CCACAAAGTC
chr12	55810882	55811993	7.30E-05	CCACAAAGTC
chr13	21842001	21842479	7.75E-05	AGACAAAGAC
chr18	35815494	35815670	7.75E-05	AGACAAAGAC
chr6	68200903	68201280	7.75E-05	agacaaagac
chr4	46596519	46596890	7.75E-05	agacaaagac
chrX	122458051	122459744	7.88E-05	AAACAAAGAC
chrX	121706644	121708339	7.88E-05	AAACAAAGAC
chrX	121577239	121578865	7.88E-05	AAACAAAGAC
chr17	13911136	13911387	7.88E-05	AAACAAAGAC
chr9	115695410	115695590	8.02E-05	ATACAAAGAC
chr12	55819723	55820200	8.64E-05	AGACAAAGCA
chr2	146335546	146336272	8.64E-05	ACACAAAGCA
chr12	55830120	55831178	8.64E-05	AGACAAAGCA
chr9	59505862	59506188	9.09E-05	AGACAAAGGC
chr9	120058537	120058711	9.38E-05	AAACAAAGGC
chr7	39231479	39233273	9.61E-05	CAACAAAGTG
chr3	103897340	103897866	9.72E-05	ctacaaagtg
chr15	82725119	82725439	9.72E-05	ctacaaagtg
chr17	13911136	13911387	9.72E-05	CTACAAAGTG
chr18	3004771	3005630	9.72E-05	CTACAAAGTG
chr18	81797204	81797484	9.94E-05	AGACAAAGAG
chr9	120058537	120058711	9.94E-05	AGACAAAGAG
chr9	41056175	41056341	9.94E-05	AGACAAAGAG
chr4	154231384	154231576	9.94E-05	AGACAAAGAG
chr11	87239984	87240835	9.94E-05	ACACAAAGAG
chr5	14918464	14922081	9.94E-05	ACACAAAGAG

4 Regions containing 1 or more SOX Motif

chr	start	stop	nearest gene (distance to TSS)
chr19	34812509	34812659	Slc16a12 (+9195), Ifit1 (+97223)
chr9	59505862	59506188	Gramd2 (-49419), Pkm (+1642)
chr13	12366712	12366965	Mtr (-16572), Actn2 (+66188)
chr4	101763903	101764474	Pde4b (-163419), Gm12789 (+104958)
chr12	5503370	5503761	2810032G03Rik (+85119)
chr2	121962909	121963298	Spg11 (-18982), B2m (-10318)
chrX	122458051	122459744	4932411N23Rik (+891718)
chrX	121706644	121708339	Vmn2r121 (-457973)
chrX	121577239	121578865	Vmn2r121 (-328533)
chr12	113338413	113338652	BC048943 (-119571), Aspg (-6361)
chr10	61591506	61591858	Neurog3 (-4156)
chr17	13911136	13911387	Gm7168 (-174118), Mllt4 (+13707)
chr4	125568410	125568787	Csf3r (-133195), Grik3 (+400655)
chr4	3429671	3429955	Vmn1r3 (-317308), Tmem68 (+72187)
chr9	120058537	120058711	Mobp (-236)
chr1	192053952	192054194	Prox1 (-59514), Rps6kc1 (+681576)
chr3	93518698	93518941	Tdpoz4 (-81500), S100a10 (+159818)
chr16	95979203	95979519	Ets2 (+55679), Psmg1 (+233206)
chr11	76148608	76148883	Glod4 (-91519), Nxn (+63896)
chr1	136920453	136920843	Ube2t (+61506), Lgr6 (+81205)
chr9	41056175	41056341	Ubash3b (-90113), Sorl1 (+876122)
chr3	144408324	144408651	Sh3glb1 (-25201), Clca1 (+15317)
chr3	144467409	144467757	Clca1 (-43778), Clca2 (+14875)
chr11	121518509	121519059	Zfp750 (-138137), B3gntl1 (+15683)
chr10	89139305	89139721	Slc17a8 (-55515), Scyl2 (+9517)
chr6	83530301	83530768	Stambp (-8037), Clec4f (+75646)
chr11	11397628	11398506	4930415F15Rik (+8798), 4930512M02Rik (+129760)
chr11	87275408	87276317	Gm11492 (-104292), Tex14 (+57296)
chr11	87239984	87240835	Gm11492 (-139745), Tex14 (+21843)
chr5	14918464	14922081	Gm9758 (-5374), Speer4e (+18156)
chr14	120737186	120737476	Rap2a (-140335), Mbnl2 (+62402)
chr18	35803792	35803919	Slc23a1 (-16958), Mzb1 (+5165)
chr2	34611014	34611361	Gapvd1 (-25021), Hspa5 (-16424)
chrX	166438853	166447041	4933400A11Rik (-225380)
chr1	173012987	173014985	1700009P17Rik (-37806), Fcgr3 (-24452)
chr9	110428485	110429550	Setd2 (-6083), Kif9 (+49520)
chr12	20828925	20829099	1700030C10Rik (-7424), Zfp125 (+77716)
chrY	2869581	2872285	Gm10352 (+480543)
chr2	146335546	146336272	Ralgapa2 (+1831), Insm1 (+288252)
chr17	15950608	15950912	Rgmb (+12790), Chd1 (+108829)
chr15	25143986	25144140	9230109A22Rik (-65149), Basp1 (+199456)
chr9	65049332	65049789	Parp16 (-13393), Igdcc3 (+60600)
chr11	70729707	70729900	Nup88 (+53660), Rabep1 (+71524)
chr12	112341390	112341756	Traf3 (-63101), Rcor1 (+63748)
chr11	69571960	69572257	Polr2a (+30)
chr9	25059849	25060024	Sept7 (-232)
chr17	39979856	39985845	Gm7148 (-269450)
chr3	96290119	96291035	Hfe2 (-38531), BC107364 (-34348)
chr3	96292179	96293188	BC107364 (-36455), Hfe2 (-36424)
chr1	133716876	133717034	Slc41a1 (-7599), Pm20d1 (+22997)
chr17	15344966	15345538	2210404J11Rik (+166647), Dll1 (+167688)
chr7	118226353	118226835	Eif4g2 (-50)
chr5	54893762	54893923	Stim2 (+504187)
chr14	54880915	54881135	Abhd4 (+2118), Olfr49 (+21575)

chr4	154231384	154231576	Mmel1 (-14258), Ttc34 (+1171)
chrX	120405249	120406500	Srsx (-69785), Vmn2r121 (+843644)
chrX	120634626	120635877	Srsx (-299162), Vmn2r121 (+614267)
chr1	180252156	180252389	Cox20 (+3012), Hnrnpu (+15655)
chr15	5099576	5099919	Ttc33 (-35812), Prkaa1 (+5887)
chr11	101466103	101466410	Rdm1 (-22999), Tmem106a (+22701)
chr7	108458149	108458626	Atg16l2 (-7786), Stard10 (-7212)
chr13	21809473	21809969	Hist1h2bm (-4192), Hist1h3h (+224)
chrX	30762714	30762936	Gm21637 (-60425), Gm2799 (+164429)
chr15	38212163	38212414	Klf10 (+18173), Odf1 (+63331)
chr17	80606473	80606939	Srsf7 (-686), Ttc39d (-94)
chr5	113763814	113764285	Sgsm1 (-24244), Aym1 (-22264)
chr15	101724635	101724846	Krt76 (-1390)
chr4	123369561	123369835	Macf1 (-8095), Ndufs5 (+25747)
chr17	75509211	75509527	Rasgrp3 (-355509), Ltbp1 (+104461)
chr18	3004771	3005630	Vmn1r238 (+118264)
chr12	93036365	93036569	Ston2 (-11591), Sel1l (+51130)
chr13	52846109	52846573	Syk (+167799), Auh (+178689)
chr19	44076755	44076987	Cpn1 (-15825), Cyp2c44 (+26866)
chr5	11220304	11220562	4933402N22Rik (-374523), Gm5861 (+37361)
chr5	11881170	11881428	Sema3d (-501867), 4933402N22Rik (+286343)
chr8	85884272	85884590	Clgn (-29346), 4933434I20Rik (+12061)
chr8	127528426	127529090	Egln1 (-55604), Tsnax (-8139)
chr11	86887233	86887808	Gdpd1 (+43)
chr17	15098409	15099907	Phf10 (-921)
chr17	15161360	15162859	9030025P20Rik (+14910), Tcte3 (+16413)
chr17	15129992	15131491	Gm3435 (+14913), Gm3417 (+16406)
chr18	84588162	84588554	Zfp407 (+170538), Zadh2 (+330808)
chr16	84647990	84648749	Mrpl39 (+87617)
chr5	123453058	123453399	Kdm2b (-14121), Orai1 (-11854)
chr17	87452962	87453212	Gm5499 (-24524), Cript (+28197)
chr7	5543821	5543962	Vmn1r62 (-82337), Vmn1r60 (-47161)
chr4	17379647	17379762	Mmp16 (-400900)
chr14	69799171	69799572	Nkx3-1 (-9323), Nkx2-6 (+9739)
chr15	88568772	88569185	Brd1 (-4316)
chr13	66394389	66396325	Vmn2r-ps104 (+60964), 2610044O15Rik8 (+256118)
chr13	9833500	9833762	Zmynd11 (-69225), Chrm3 (+44698)
chr7	16994559	16994776	Tmem160 (-43460), Zc3h4 (+8438)
chr16	34671860	34672520	Ccdc14 (-18512), Ropn1 (+20916)
chr9	75471052	75471301	Lysmd2 (-2362)
chr9	110989678	110990024	Lrrfip2 (-30764), Ccrl2 (-30077)
chr13	21842001	21842479	Hist1h4k (+134)
chr2	98506338	98507542	Gm10800 (+518)
chr13	65967440	65968021	Gm10772 (-39856), Gm10139 (+354117)
chr13	65552554	65553189	Gm10139 (-60742), Zfp369 (+172750)
chr17	14847569	14847759	Thbs2 (-16395), Wdr27 (+223994)
chr12	55835205	55835592	Eapp (-38531), Snx6 (+61291)
chr12	55814849	55815664	Eapp (-18389), Snx6 (+81433)
chr3	58469472	58469815	Fam194a (-28515), Siah2 (+26678)
chr11	32076000	32076214	Il9r (+24129), Nsg2 (+175644)
chr7	39352222	39353766	Gm5591 (-39782)
chr7	39231479	39233273	Pop4 (-176009), Gm5591 (+80836)
chr9	118647931	118648304	Ctdspl (-187453), Itga9 (+132310)
chr8	83262886	83263555	Smarca5 (+175)
chr3	58469948	58470228	Fam194a (-28959), Siah2 (+26234)
chr4	132994761	132995022	Fam46b (-41155), Slc9a1 (+69271)

chr4	149612437	149612864	Rere (-43104), Eno1 (+1590)
chr7	31306647	31307327	Arhgap33 (+13092), Prodh2 (+28346)
chr3	68096244	68096863	Il12a (-398012), Schip1 (+227690)
chr13	51941709	51941925	Gadd45g (-230)
chr1	194061545	194061791	Kcnh1 (+47002), Hhat (+535739)
chr3	23209200	23209294	NONE
chr2	98505008	98505437	Gm10800 (+2235), Gm10801 (+2829)
chr9	82869297	82869515	Phip (-310)
chr1	169328474	169328599	Mgst3 (-4565)
chr16	36991075	36991339	Fbxo40 (-654)
chr7	120444291	120444522	Btbd10 (+68446), Arntl (+93428)
chr12	19226266	19226526	Gm5784 (-167269), 5730507C01Rik (+704852)
chr13	56199754	56200260	Pitx1 (-267221), H2afy (+36985)
chr3	52059270	52059807	Maml3 (-150726), Foxo1 (-12720)
chr16	84975160	84975506	Atp5j (-139463), App (+198615)
chr3	69490627	69490976	B3galnt1 (-88019), Nmd3 (-35105)
chr11	22170490	22170823	Otx1 (-269039), Ehbp1 (+15184)
chr19	57565272	57565511	Atrnl1 (-120132), Trub1 (+37996)
chr15	87797353	87797527	Zdhhc25 (-633305), Fam19a5 (+422536)
chr14	121483899	121484504	Farp1 (+49416), Stk24 (+294354)
chr3	152996314	152997030	St6galnac5 (-351462), St6galnac3 (+391354)
chr7	91561750	91562035	Arnt2 (-3424)
chr9	49183419	49183597	Drd2 (+34776), Ankk1 (+51618)
chr8	119567477	119567793	Gcsh (-50198), Pkd1l2 (+38714)
chr8	31438279	31438743	Dusp26 (-761432)
chr2	57246848	57247325	Galnt5 (-603208), Gpd2 (+156379)
chr4	8154834	8155634	Car8 (+10954)
chr11	108872866	108873417	Axin2 (+91479), E030025P04Rik (+132541)
chr10	9467362	9467556	Samd5 (-72453), Stxbp5 (+153418)
chr6	116201111	116202043	Zfand4 (-12663), Fam21 (+43526)
chr18	81797204	81797484	Sall3 (-614027), Galr1 (+778825)
chr17	33811718	33812117	Hnrnpm (+11887), Pram1 (+36917)
chr19	21581779	21582122	Gm3443 (-48213), Gda (-34789)
chr15	82725119	82725439	Cyp2d26 (-100604), Tcf20 (+17285)
chr3	146005710	146006075	Ssx2ip (-61713), Lpar3 (+121968)
chr9	19029130	19029345	Olfr843 (+24669), Olfr836 (+103837)
chr1	88617661	88617989	Dis3l2 (+17446), Alppl2 (+368678)
chr17	84581105	84581330	Haao (-335088), Zfp36l2 (+6069)
chr5	28661168	28661314	Evc2 (-9068497), Rbm33 (+17512), Shh (+132400)
chr11	87261721	87262640	Gm11492 (-117974), Tex14 (+43614)
chr1	88422999	88423418	Ptma (-102)
chr4	114943768	114943896	Cyp4x1 (-137250), Cyp4a12a (-27819)
chr4	114986056	114986184	Cyp4a12b (-98109), Cyp4a12a (+14469)
chr1	72290738	72291074	Mreg (-32025), Pecr (+39982)
chr5	72438606	72438857	Commd8 (+120690), Gabrb1 (+347561)
chr16	22672855	22673155	Dgkg (-15759), Crygs (+138478)
chr3	124418625	124418909	Ndst4 (-688242), 1700003H04Rik (-134758)
chr18	76613456	76613637	Skor2 (-481597), Smad2 (+212313)
chr10	12273996	12274342	Utrn (+307364), Gm9797 (+945148)
chr15	11259966	11260124	Tars (+69368), Adamts12 (+265500)
chr8	25462270	25462920	1810011O10Rik (+86823), A730045E13Rik (+178219)
chr10	4930547	4930938	Syne1 (+126216), Esr1 (+803777)
chr11	87236084	87236908	Gm11492 (-143659), Tex14 (+17929)
chr3	96163525	96164181	Fcgr1 (-65961), BC107364 (+92376)
chr3	96177802	96178458	Fcgr1 (-80238), BC107364 (+78099)
chr13	70231949	70232253	Med10 (+283341), BC018507 (+544411)

chr2	22443076	22443924	Gad2 (-34225), Myo3a (+294370)
chr1	24618332	24623038	Gm10222 (-1146)
chr5	11829495	11829872	Sema3d (-553482), 4933402N22Rik (+234728)
chr5	11400416	11400736	4933402N22Rik (-194380), Gm5861 (+217504)
chr2	162688176	162688554	Ptprt (-201482), Srsf6 (-68899)
chr2	131233744	131234170	Smox (-83722), Rnf24 (-55340)
chr1	141954325	141954737	Gm4788 (-276715), Cfh (+125457)
chr6	83417694	83418110	Tet3 (-26233), Dguok (+39061)
chr14	118206572	118206838	Dct (+244763), Gpc6 (+881952)
chr5	143579353	143579517	Tnrc18 (-369)
chr5	16974433	16975269	Speer4f (-7065), Gm3495 (+95825)
chr17	10380494	10380692	Pabpc6 (-518024), Qk (+131633)
chr10	80827990	80829989	4930404N11Rik (-429)
chrX	113580034	113580205	H2afb2-ps (-214667), Cpxcr1 (+17567)
chr6	145881183	145881640	Sspn (-1230)
chr1	136457440	136457676	Kdm5b (+810)
chr11	105301298	105301520	March10 (+16640), Mrc2 (+147449)
chr7	108789576	108789755	Art2b (-59991), Clpb (-22481)
chr11	76191172	76191993	Glod4 (-134356), Nxn (+21059)
chr14	57315836	57316043	Mphosph8 (+28855), Pspc1 (+81213)
chr4	153394885	153395163	Smim1 (+5237), Lrrc47 (+9184)
chr2	80302072	80302264	Frzb (-14386), Nckap1 (+119369)
chr5	66330250	66330572	Chrna9 (-27952), Rhoh (+75603)
chr5	108295958	108296386	Evi5 (+7954), Ube2d2b (+36900)
chr16	72258386	72258542	Robo1 (-404930)
chr2	3356844	3357148	Dclre1c (+15574), Suv39h2 (+35307)
chr13	24025614	24026204	Hist1h2aa (-422), Hist1h2ba (+116)
chr17	71834308	71834443	Smchd1 (-9693), Ndc80 (+41821)
chr12	55838910	55839950	Eapp (-42562), Snx6 (+57260)
chr19	29703448	29703901	Ermp1 (+19230), C030046E11Rik (+106903)
chr12	55810882	55811993	Eapp (-14570), Snx6 (+85252)
chr18	35815494	35815670	Prob1 (-741)
chr6	68200903	68201280	Tacstd2 (-715276), Igkv4-71 (+992347)
chr4	46596519	46596890	Trim14 (-47692), Coro2a (+18369)
chr9	115695410	115695590	Stt3b (-475961), Gadl1 (-122182)
chr12	55819723	55820200	Eapp (-23094), Snx6 (+76728)
chr12	55830120	55831178	Eapp (-33781), Snx6 (+66041)
chr3	103897340	103897866	Phtf1 (+125570), Magi3 (+126423)

5 Overlapping Peak Regions of NPC and Testes ChIP Datasets

chr	start	stop	nearest gene (distance to TSS)
chr1	12634809	12634963	Sulf1 (-73,740)
chr1	24618373	24623025	Gm10222 (-1,160)
chr1	36425881	36426233	Kansl3 (-1,198)
chr1	39576302	39576463	Tbc1d8 (-40,791), Cnot11 (-15,454)
chr1	46910029	46910937	Slc39a10 (+408)
chr1	72283446	72283800	Mreg (-24,742), Pecr (+47,265)
chr1	72290752	72291047	Mreg (-32,019), Pecr (+39,988)
chr1	72301572	72301858	Mreg (-42,834), Pecr (+29,173)
chr1	120355509	120355898	2900060B14Rik (-380)
chr1	172958057	172958254	Fcgr4 (+9,105), Fcgr3 (+31,378)
chr1	173432897	173433394	Alyref2 (-463)
chr1	180250830	180251028	Cox20 (+1,668), Hnrnpu (+16,999)
chr1	180251135	180251576	Cox20 (+2,095), Hnrnpu (+16,572)
chr10	12273996	12274252	Utrn (+307,409), Gm9797 (+945,103)
chr10	39978217	39978605	Gtf3c6 (-940)
chr10	81167521	81167907	Zfp873 (-353,448), Ankrd24 (+76,334)
chr10	81326630	81327016	Zfp873 (-194,339), Ankrd24 (+235,443)
chr10	81459738	81460166	Zfp873 (-61,210), Ankrd24 (+368,572)
chr11	22170496	22170759	Otx1 (-269,010), Ehbp1 (+15,213)
chr11	52099833	52100032	Tcf7 (-3,860)
chr11	76191172	76191993	Glod4 (-134,356), Nxn (+21,059)
chr11	83084989	83085258	Pex12 (+27,355), Slfn3 (+80,292)
chr11	86887233	86887609	Gdpd1 (+143)
chr11	87236084	87236908	Gm11492 (-143,659), Tex14 (+17,929)
chr11	87239984	87240724	Gm11492 (-139,801), Tex14 (+21,787)
chr11	87256379	87257209	Gm11492 (-123,361), Tex14 (+38,227)
chr11	87261777	87262640	Gm11492 (-117,946), Tex14 (+43,642)
chr11	87275408	87276302	Gm11492 (-104,300), Tex14 (+57,288)
chr11	87284520	87285299	Gm11492 (-95,245), Tex14 (+66,343)
chr11	95893927	95894077	Snf8 (-2,197)
chr11	101466103	101466410	Rdm1 (-22,999), Tmem106a (+22,701)
chr11	101488657	101488966	Rdm1 (-444)
chr11	101502809	101503485	Arl4d (-23,708), Rdm1 (+13,891)
chr11	101506785	101507524	Arl4d (-19,700), Rdm1 (+17,899)
chr11	101519481	101519810	Arl4d (-7,209), Rdm1 (+30,390)
chr11	106586499	106586648	Tex2 (-112,330), Pecam1 (+25,368)
chr12	55797359	55798055	Eapp (-839)
chr12	55810882	55811864	Eapp (-14,505), Snx6 (+85,317)
chr12	55814849	55815664	Eapp (-18,389), Snx6 (+81,433)
chr12	55819723	55820200	Eapp (-23,094), Snx6 (+76,728)
chr12	55820236	55820781	Eapp (-23,641), Snx6 (+76,181)
chr12	55830120	55831178	Eapp (-33,781), Snx6 (+66,041)
chr12	55835205	55835589	Eapp (-38,529), Snx6 (+61,293)
chr12	55835629	55836060	Eapp (-38,977), Snx6 (+60,845)
chr12	55839039	55839950	Eapp (-42,627), Snx6 (+57,195)
chr12	98299628	98300037	NONE
chr12	111934191	111934354	Hsp90aa1 (+332)
chr13	21813490	21813709	Hist1h2bm (-313)
chr13	21826870	21827022	Hist1h4j (+43)
chr13	21827037	21827336	Hist1h4j (+284)
chr13	21842001	21842463	Hist1h4k (+142)
chr13	21901526	21901858	Hist1h4m (-1,923), Hist1h2bq (+361)
chr13	21923627	21924238	Hist1h2br (-1,694), Hist1h4n (+132)
chr13	21925185	21925768	Hist1h4n (-1,412), Hist1h2br (-150)

chr13	21925821	21926153	Hist1h4n (-1,922), Hist1h2br (+360)
chr13	22132895	22133107	Hist1h2bj (-2,082), Hist1h4i (+220)
chr13	22135090	22135471	Hist1h4i (-2,060), Hist1h2ag (-463), Hist1h2bj (+198)
chr13	23623016	23623365	Hist1h3g (-4,100), Hist1h2af (-2,584), Hist1h4h (+272)
chr13	23643141	23643381	Hist1h1d (-3,640), Hist1h4f (+256)
chr13	23789960	23790459	Hist1h4c (+113)
chr13	23830593	23831031	Hist1h1c (+136)
chr13	23848946	23849239	Hist1h4b (+287)
chr13	23852658	23852981	Hist1h1a (-2,717), Hist1h4a (+279)
chr13	51297963	51298394	Hist1h2al (+255)
chr14	121483899	121484504	Farp1 (+49,416), Stk24 (+294,354)
chr15	88568772	88569183	Brd1 (-4,315)
chr16	11143902	11144450	Txndc11 (-9,433), Zc3h7a (+32,310)
chr16	20129532	20130245	Yeats2 (-11,247), Klhl24 (+32,262)
chr16	57391146	57391854	Filip1l (+38,294), Cmss1 (+215,442)
chr17	4626224	4626585	Nox3 (-930,144), Arid1b (-368,669)
chr17	15098574	15099277	Phf10 (-689)
chr17	15099292	15099907	Phf10 (-1,363)
chr17	15115491	15116168	Gm3448 (-209), Gm3435 (+1)
chr17	15130157	15131491	Gm3435 (+14,995), Gm3417 (+16,324)
chr17	15146863	15147540	Gm3417 (-54), 9030025P20Rik (+2)
chr17	15161525	15162859	9030025P20Rik (+14,992), Tcte3 (+16,331)
chr17	15344966	15345538	2210404J11Rik (+166,647), Dll1 (+167,688)
chr17	39979865	39985845	Gm7148 (-269,446)
chr17	80606527	80606939	Srsf7 (-713), Ttc39d (-67)
chr18	10151061	10151227	Usp14 (-121,005), Rock1 (+30,646)
chr19	4078426	4078532	BC021614 (-19,184), Cabp2 (-5,040)
chr19	6996550	6996897	Esrra (-426)
chr19	7015795	7016395	Bad (-256), Gpr137 (-126)
chr19	10379354	10379563	Dagla (-92)
chr19	21581807	21582122	Gm3443 (-48,199), Gda (-34,803)
chr2	25035827	25036514	Nrarp (-107)
chr3	5860327	5860828	Pex2 (-284,339), 1700008P02Rik (+759,865)
chr3	96042667	96043048	Hist2h3c2 (+192)
chr3	96043599	96043917	Hist2h3c2 (-708), Hist2h2aa2 (+458)
chr3	96043950	96044507	Hist2h3c2 (-1,179), Hist2h2aa2 (-13)
chr3	96049212	96049726	Hist2h3c1 (-1,205), Hist2h2aa1 (+8)
chr3	96049759	96050060	Hist2h3c1 (-764), Hist2h2aa1 (+449)
chr3	96050624	96051005	Hist2h3c1 (+141)
chr3	96072452	96072813	Hist2h2bb (-1,042), Hist2h3b (+56)
chr3	96132007	96132599	Fcgr1 (-34,411), BC107364 (+123,926)
chr3	96163525	96164181	Fcgr1 (-65,961), BC107364 (+92,376)
chr3	96177802	96178458	Fcgr1 (-80,238), BC107364 (+78,099)
chr3	96226594	96226840	Fcgr1 (-128,825), BC107364 (+29,512)
chr3	96254032	96254704	Fcgr1 (-156,476), BC107364 (+1,861)
chr3	96263776	96264448	Hfe2 (-64,996), BC107364 (-7,883)
chr3	96290119	96291009	Hfe2 (-38,544), BC107364 (-34,335)
chr3	96292179	96293188	BC107364 (-36,455), Hfe2 (-36,424)
chr3	144408348	144408645	Sh3glb1 (-25,210), Clca1 (+15,308)
chr3	152165632	152166311	Zzz3 (+107,005), Ak5 (+165,132)
chr3	154136647	154137000	Tyw3 (+123,235), Slc44a5 (+500,424)
chr4	51825350	51825556	Smc2 (-626,662), Cylc2 (+595,903)
chr4	108952605	108952944	Eps15 (-98)
chr4	116862488	116862794	Kif2c (-7,397), Gm1661 (+61,881)
chr4	124617836	124618210	Cdca8 (-3,906)
chr4	131825783	131826441	Taf12 (-4,178)

chr4	134024169	134024324	Stmn1 (+12)
chr5	115939403	115939805	Sirt4 (-4,870)
chr5	149864485	149864708	Hmgb1 (+19)
chr6	3150763	3151774	Samd9l (+198,302)
chr6	50146939	50147260	Dfna5 (+66,761), Mpp6 (+86,860)
chr6	51419690	51420373	Hnrnpa2b1 (-2,604), Cbx3 (-327)
chr6	132413723	132413820	Prh1 (-106,055), Gm8882 (-99,620)
chr6	146525985	146526146	Fgfr1op2 (-315), Asun (+291)
chr8	73222346	73222949	Jund (+1,010), Gm11175 (+1,331)
chr8	83262886	83263555	Smarca5 (+175)
chr9	65044185	65044584	Parp16 (-18,569), Igdcc3 (+55,424)
chr9	65049402	65049749	Parp16 (-13,378), Igdcc3 (+60,615)
chr9	110989678	110990024	Lrrfip2 (-30,764), Ccrl2 (-30,077)

6 Peaks only present with Testes ChIP (Testes Specific)

chr	start	stop	nearest gene (distance to TSS)
chr1	6638065	6638194	St18 (-82002), Fam150a (+288593)
chr1	8400880	8401020	Sntg1 (+59651)
chr1	16647040	16647245	Tceb1 (-197)
chr1	23067721	23067887	Rims1 (-464382), 4933415F23Rik (+41288)
chr1	30819114	30819315	Gm9898 (-42409), Phf3 (+100886)
chr1	33951484	33951619	Zfp451 (-80112), Bend6 (+13109)
chr1	33951622	33951859	Zfp451 (-80301), Bend6 (+12920)
chr1	38461097	38461289	Rev1 (-274686), Aff3 (+135494)
chr1	48519013	48519161	C230029F24Rik (-782373)
chr1	49014660	49014787	C230029F24Rik (-286736)
chr1	57513605	57513800	Spats2l (-318002), Tyw5 (-49758)
chr1	66863661	66863845	Kansl1l (+367)
chr1	69851276	69851438	Ikzf2 (-118823), Spag16 (-22205)
chr1	83996431	83996610	Sphkap (-591746), Pid1 (+284699)
chr1	84836064	84836238	Trip12 (-272), Fbxo36 (-265)
chr1	88383859	88384257	Ptma (-39253), 1700019O17Rik (+61112)
chr1	88422999	88423418	Ptma (-102)
chr1	88617661	88617989	Dis3l2 (+17446), Alppl2 (+368678)
chr1	91355897	91356040	D130058E05Rik (-470840), Agap1 (+4583)
chr1	95841863	95842057	Gm6086 (-45126), Gm9994 (+17596)
chr1	99816002	99816317	Ppip5k2 (-149172), Pam (+176038)
chr1	105755854	105756011	Cdh20 (-909173)
chr1	115518268	115518479	NONE
chr1	130932071	130932291	Cxcr4 (-443314), Thsd7b (-237740)
chr1	133716876	133717034	Slc41a1 (-7599), Pm20d1 (+22997)
chr1	134318052	134318382	Rbbp5 (-55727), Dstyk (+4187)
chr1	136457209	136457415	Kdm5b (+564)
chr1	136457440	136457676	Kdm5b (+810)
chr1	136457717	136457907	Syt2 (-148065), Kdm5b (+1064)
chr1	136920453	136920843	Ube2t (+61506), Lgr6 (+81205)
chr1	138233122	138233315	Gpr25 (-75769), Camsap2 (+9462)
chr1	138235598	138235739	Gpr25 (-78219), Camsap2 (+7012)
chr1	141954325	141954737	Gm4788 (-276715), Cfh (+125457)
chr1	142869679	142869779	Kcnt2 (+726934)
chr1	159482429	159482704	Sec16b (+45708), Fam5b (+803824)
chr1	159661262	159661466	Sec16b (+224505), Fam5b (+625027)
chr1	162275331	162275404	Cacybp (-132460), Rabgap1l (+447701)
chr1	163027312	163027488	Cenpl (+26502), Klhl20 (+34242)
chr1	169270302	169270622	Aldh9a1 (-9660), Tmco1 (+31661)
chr1	169328474	169328599	Mgst3 (-4565)
chr1	169600816	169601057	Lmx1a (-18752), Rxrg (+72422)
chr1	172572750	172572955	Olfml2b (-1810)
chr1	173001780	173005313	1700009P17Rik (-48245), Fcgr3 (-14013)
chr1	173012987	173014985	1700009P17Rik (-37806), Fcgr3 (-24452)
chr1	173808683	173808857	Gm10521 (-17025), Cd84 (+38942)
chr1	174196993	174197081	Atp1a4 (-8492), Atp1a2 (+31158)

chr1	178744251	178744519	Sdccag8 (-406)
chr1	180251592	180252051	Cox20 (+2561), Hnrnpu (+16106)
chr1	180252156	180252389	Cox20 (+3012), Hnrnpu (+15655)
chr1	180252951	180253209	Cox20 (+3819), Hnrnpu (+14848)
chr1	180266685	180267253	Hnrnpu (+959)
chr1	183634010	183634190	Ccdc121 (-192518), Dnahc14 (-37067)
chr1	186854896	186855388	Rab3gap2 (-172905), Mark1 (-31714)
chr1	192053952	192054194	Prox1 (-59514), Rps6kc1 (+681576)
chr1	194061545	194061791	Kcnh1 (+47002), Hhat (+535739)
chr1	197190299	197190481	Cr2 (-187481)
chr10	3831543	3831933	Rgs17 (-592417), Gm10945 (+246958)
chr10	4930547	4930938	Syne1 (+126216), Esr1 (+803777)
chr10	9467362	9467556	Samd5 (-72453), Stxbp5 (+153418)
chr10	11710100	11710246	Gm9797 (+381152), Utrn (+871360)
chr10	17037210	17037294	Cited2 (-405782)
chr10	17044382	17044530	Cited2 (-398578)
chr10	21000123	21000287	Myb (-119415), Hbs1l (-15580)
chr10	26647120	26647204	Arhgap18 (+154879), Lama2 (+689586)
chr10	26655097	26655181	Arhgap18 (+162856), Lama2 (+681609)
chr10	35943971	35944136	Amd2 (-512357), Hs3st5 (-282566)
chr10	37782321	37782472	Lama4 (-902924), 5930403N24Rik (+923033)
chr10	39691073	39691718	G630090E17Rik (-11437), BC021785 (+15049)
chr10	49977195	49977294	Grik2 (-468673), Ascc3 (-335230)
chr10	59378436	59378712	Ddit4 (+35944), Dnajb12 (+36251)
chr10	61591506	61591858	Neurog3 (-4156)
chr10	76766491	76766846	Pofut2 (+44624), Adarb1 (+114346)
chr10	77152943	77153141	Gm10272 (-16347), Ube2g2 (+67961)
chr10	80827990	80829989	4930404N11Rik (-429)
chr10	89139305	89139721	Slc17a8 (-55515), Scyl2 (+9517)
chr10	110654809	110655409	Bbs10 (-80626), Osbpl8 (+53251)
chr10	115632868	115633040	Ptprb (-105476), 4933416C03Rik (-81981)
chr11	3724658	3725260	Osbp2 (+38947), Morc2a (+175462)
chr11	3971581	3971778	Sec14l3 (+6836), Mtfp1 (+23768)
chr11	11397628	11398506	4930415F15Rik (+8798), 4930512M02Rik (+129760)
chr11	14977824	14977948	NONE
chr11	23624548	23624891	Pex13 (-58761), Rel (+46250)
chr11	25000048	25000203	5730522E02Rik (+670444)
chr11	32076000	32076214	Il9r (+24129), Nsg2 (+175644)
chr11	40613689	40613785	Ccng1 (-44924)
chr11	50191163	50191410	Hnrnph1 (+16)
chr11	58775172	58775820	Trim17 (-1808)
chr11	60187409	60187676	Lrrc48 (+20712), Atpaf2 (+43010)
chr11	62364915	62365180	Ubb (+375)
chr11	62365648	62366351	Trpv2 (-22040), Ubb (+1327)
chr11	69571960	69572257	Polr2a (+30)
chr11	70729707	70729900	Nup88 (+53660), Rabep1 (+71524)
chr11	73312282	73312397	Olfr381 (-7361), Olfr382 (+18359)
chr11	76148608	76148883	Glod4 (-91519), Nxn (+63896)

chr11	76191995	76192164	Glod4 (-134853), Nxn (+20562)
chr11	79398713	79398991	Evi2a (-54762), Rab11fip4 (-5862)
chr11	79687340	79687428	Utp6 (+88508), Rab11fip4 (+282670)
chr11	82841057	82841310	Slfn2 (-37430), Slfn8 (-6872)
chr11	95737123	95737400	Abi3 (-33539), B4galnt2 (+38943)
chr11	98724900	98725256	Wipf2 (+126)
chr11	103268409	103268520	Arhgap27 (-46251), Plekhm1 (+5513)
chr11	105301298	105301520	March10 (+16640), Mrc2 (+147449)
chr11	108872866	108873417	Axin2 (+91479), E030025P04Rik (+132541)
chr11	115885512	115885926	Galk1 (-11686), H3f3b (+3557)
chr11	116928412	116928569	Sec14l1 (-47995), Mgat5b (+148314)
chr11	119320596	119321173	Endov (-31776), Rnf213 (+66471)
chr11	119720614	119720939	Chmp6 (-54347), Rptor (+256314)
chr11	120092832	120092980	Bahcc1 (-1355)
chr11	121518509	121519059	Zfp750 (-138137), B3gntl1 (+15683)
chr12	3036912	3037000	Rab10 (+273013)
chr12	4786211	4786418	0610009D07Rik (-38099), Pfn4 (+10214)
chr12	5503370	5503761	2810032G03Rik (+85119)
chr12	9432167	9432451	Osr1 (-148939), Ttc32 (+395506)
chr12	12018484	12018858	Fam49a (-250274), Rad51ap2 (+555786)
chr12	12018866	12019112	Fam49a (-249956), Rad51ap2 (+556104)
chr12	13143011	13143206	Mycn (-194389), Ddx1 (+112910)
chr12	17539673	17540043	Odc1 (-11821), Nol10 (+184573)
chr12	19226266	19226526	Gm5784 (-167269), 5730507C01Rik (+704852)
chr12	20828925	20829099	1700030C10Rik (-7424), Zfp125 (+77716)
chr12	22990713	22991030	NONE
chr12	35958601	35958755	Snx13 (+226817), Ahr (+261032)
chr12	37392265	37392572	Meox2 (-442708), D630036H23Rik (-283651)
chr12	38091134	38091230	Dgkb (-516110), Agmo (+122954)
chr12	50712141	50712236	Gm9804 (+209925)
chr12	59261099	59261197	Foxa1 (-614040), Sstr1 (-51611)
chr12	62477546	62477700	Lrfn5 (-147312)
chr12	71978715	71978914	Actr10 (-60029), Frmd6 (+52314)
chr12	79949073	79949263	Atp6v1d (+13457), Mpp5 (+99234)
chr12	92828267	92828427	Gtf2a1 (+102)
chr12	93036365	93036569	Ston2 (-11591), Sel1l (+51130)
chr12	103079306	103079937	Fbln5 (-22357), Trip11 (+71855)
chr12	112341390	112341756	Traf3 (-63101), Rcor1 (+63748)
chr12	112579306	112579480	Cdc42bpb (+36536), Amn (+70071)
chr12	113088150	113088677	Ppp1r13b (+22412), Zfyve21 (+36033)
chr12	113338413	113338652	BC048943 (-119571), Aspg (-6361)
chr13	3026151	3026232	Gdi2 (-511129)
chr13	3179280	3179495	Gdi2 (-357933)
chr13	6134585	6134854	Pitrm1 (-412683), Klf6 (+273985)
chr13	9833500	9833762	Zmynd11 (-69225), Chrm3 (+44698)
chr13	12366712	12366965	Mtr (-16572), Actn2 (+66188)
chr13	16464769	16464850	Inhba (+361126)
chr13	20876498	20876763	Aoah (-9357), Elmo1 (+694143)

chr13	21809473	21809969	Hist1h2bm (-4192), Hist1h3h (+224)
chr13	21846168	21846343	Hist1h4k (-3882), Hist1h2ak (-480)
chr13	21874791	21875299	Hist1h2bp (-4285), Hist1h1b (-2551), Hist1h3i (+221)
chr13	21879560	21879759	Hist1h3i (-4394), Hist1h2an (-573), Hist1h2bp (+330)
chr13	21901911	21902528	Hist1h4m (-1395), Hist1h2bq (-167)
chr13	23646975	23647201	Hist1h4f (-3571), Hist1h1d (+187)
chr13	23667674	23668022	Hist1h2ae (-4759), Hist1h2bf (-1783), Hist1h3d (+216)
chr13	23714034	23714302	Hist1h2be (-1175), Hist1h1e (+203)
chr13	23776051	23776229	Hist1h2ac (-323), Hist1h2bc (+72)
chr13	23838768	23839012	Hist1h2ab (-4067), Hist1h3c (-1419), Hist1h2bb (+287)
chr13	23844179	23844628	Hist1h4b (-4402), Hist1h3b (+268)
chr13	23853920	23854301	Hist1h1a (-1426), Hist1h4a (-1012), Hist1h3a (+144)
chr13	24025614	24026204	Hist1h2aa (-422), Hist1h2ba (+116)
chr13	24633416	24633684	Fam65b (-96967), Cmah (+214261)
chr13	42190662	42190816	Edn1 (-205900), Hivep1 (+43349)
chr13	44964885	44965059	Dtnbp1 (+132537), Jarid2 (+138780)
chr13	47046203	47046525	Kif13a (-21277), Nhlrc1 (+63855)
chr13	49494901	49495368	Ippk (-21545), Bicd2 (+58217)
chr13	51941709	51941925	Gadd45g (-230)
chr13	51941968	51942164	Gadd45g (+19)
chr13	52846109	52846573	Syk (+167799), Auh (+178689)
chr13	56199754	56200260	Pitx1 (-267221), H2afy (+36985)
chr13	58502820	58503374	Rmi1 (-861)
chr13	58503773	58503940	Rmi1 (-101), Hnrnpk (+847)
chr13	61052440	61052524	Tpbpa (-9186), Ctsj (+54795)
chr13	62460597	62460717	Gm10260 (-141328), Gm5141 (+12639)
chr13	65552554	65553189	Gm10139 (-60742), Zfp369 (+172750)
chr13	65967440	65968021	Gm10772 (-39856), Gm10139 (+354117)
chr13	66394389	66396325	Vmn2r-ps104 (+60964), 2610044O15Rik8 (+256118)
chr13	67299065	67299259	Zfp455 (+3720), Zfp458 (+70842)
chr13	67474005	67474254	Zfp953 (-12589), Zfp456 (+2616)
chr13	67497929	67498144	Zfp456 (-21291), Zfp429 (+2666)
chr13	68096842	68096930	BC048507 (+132612), Mtrr (+624112)
chr13	69733647	69733870	Papd7 (-61042), Srd5a1 (+16561)
chr13	70231949	70232253	Med10 (+283341), BC018507 (+544411)
chr13	91778066	91778279	Acot12 (-102944), Ssbp2 (+177518)
chr13	91778284	91778432	Acot12 (-102759), Ssbp2 (+177703)
chr13	92094003	92094177	Ckmt2 (-77600), Rasgrf2 (+807359)
chr13	97342467	97342837	Col4a3bp (+29962), Hmgcr (+98239)
chr13	101432401	101432811	Ccdc125 (-6830), Taf9 (+11308)
chr14	4814465	4814595	Gm9602 (-64846), Gm3159 (+313946)
chr14	11353424	11353519	Fhit (+641077)
chr14	12935863	12935995	3830406C13Rik (-180794), Ptprg (+549883)
chr14	14855953	14856561	Atxn7 (+11252), Psmd6 (+97241)
chr14	37719530	37719737	Ccser2 (+62316), 4930474N05Rik (+811489)
chr14	46052730	46052914	Gnpnat1 (-44345), Fermt2 (+96971)
chr14	48705251	48705540	Peli2 (-35148), Ktn1 (+421965)
chr14	48705553	48705674	Peli2 (-34930), Ktn1 (+422183)

chr14	49739614	49739845	Exoc5 (-53402), Naa30 (-52543)
chr14	54880915	54881135	Abhd4 (+2118), Olfr49 (+21575)
chr14	57315836	57316043	Mphosph8 (+28855), Pspc1 (+81213)
chr14	57430145	57430484	Zmym5 (+238)
chr14	57505231	57505594	Zmym2 (-1218)
chr14	57505618	57505797	Zmym2 (-923)
chr14	58458772	58459036	Mrp63 (+13828), Zdhhc20 (+50113)
chr14	59029643	59029822	Rpl13-ps3 (-482610), Fgf9 (+338210)
chr14	60064554	60064873	Setdb2 (-4993), Cab39l (+4896)
chr14	61964386	61964739	Ebpl (+14719), Arl11 (+35973)
chr14	63379777	63380091	Ints6 (+15)
chr14	69799171	69799572	Nkx3-1 (-9323), Nkx2-6 (+9739)
chr14	104736093	104736324	Ednrb (-492819), Pou4f1 (+130999)
chr14	105115012	105115391	Rnf219 (-193319), Rbm26 (+461338)
chr14	118206572	118206838	Dct (+244763), Gpc6 (+881952)
chr14	120737186	120737476	Rap2a (-140335), Mbnl2 (+62402)
chr15	5099576	5099919	Ttc33 (-35812), Prkaa1 (+5887)
chr15	7527614	7527751	Gdnf (-233328), Egflam (-179332)
chr15	11259966	11260124	Tars (+69368), Adamts12 (+265500)
chr15	25143986	25144140	9230109A22Rik (-65149), Basp1 (+199456)
chr15	26188739	26188813	March11 (-50043), Zfp622 (+274655)
chr15	30510534	30510630	Dap (-643558), Ctnnd2 (+408234)
chr15	36131941	36132320	Spag1 (+23008), Rnf19a (+80771)
chr15	38212163	38212414	Klf10 (+18173), Odf1 (+63331)
chr15	44622528	44622623	Sybu (-2967)
chr15	76128460	76128765	Smpd5 (+3749), Oplah (+8918)
chr15	76128804	76129053	Smpd5 (+4065), Oplah (+8602)
chr15	79376866	79377092	Ddx17 (+192)
chr15	79560750	79561176	Sun2 (+12003), Gtpbp1 (+39643)
chr15	79561216	79561436	Sun2 (+11640), Gtpbp1 (+40006)
chr15	82725119	82725439	Cyp2d26 (-100604), Tcf20 (+17285)
chr15	86025175	86025777	Tbc1d22a (-19401), Cerk (-8905)
chr15	87797353	87797527	Zdhhc25 (-633305), Fam19a5 (+422536)
chr15	92626642	92626733	Pdzrn4 (+399342), Gxylt1 (+478904)
chr15	92628957	92629048	Pdzrn4 (+401657), Gxylt1 (+476589)
chr15	98364436	98364632	Kansl2 (+118)
chr15	101724635	101724846	Krt76 (-1390)
chr16	3174030	3174196	Olfr161 (-418285)
chr16	4301021	4301518	Crebbp (-87866), Adcy9 (+118538)
chr16	9241776	9241912	1810013L24Rik (+411651), Grin2a (+753170)
chr16	11015717	11015936	Snn (-50564), Litaf (-22613)
chr16	17088054	17088368	Ypel1 (+18422), Ppil2 (+23114)
chr16	17088394	17089580	Ypel1 (+19198), Ppil2 (+22338)
chr16	17090102	17090259	Ypel1 (+20392), Ppil2 (+21144)
chr16	17090584	17090999	Ppil2 (+20533), Ypel1 (+21003)
chr16	19925935	19926054	A930003A15Rik (-42005), Klhl6 (+57115)
chr16	22672855	22673155	Dgkg (-15759), Crygs (+138478)
chr16	31162663	31162843	Xxylt1 (-81235), Acap2 (+38324)

chr16	34671860	34672520	Ccdc14 (-18512), Ropn1 (+20916)
chr16	36991075	36991339	Fbxo40 (-654)
chr16	42569884	42570037	Zbtb20 (-677436), Gap43 (-229197)
chr16	56096558	56096901	Impg2 (-107721), Senp7 (+21208)
chr16	61238108	61238289	Epha6 (-633125)
chr16	67714057	67714263	Cadm2 (-93092)
chr16	68274642	68274795	Cadm2 (-653651)
chr16	70076415	70076548	Gbe1 (-237712), Speer2 (-212493)
chr16	72258386	72258542	Robo1 (-404930)
chr16	75294903	75295094	Robo2 (-883765), Lipi (+291302)
chr16	80759616	80759772	Ncam2 (-441248)
chr16	84647990	84648749	Mrpl39 (+87617)
chr16	84975160	84975506	Atp5j (-139463), App (+198615)
chr16	94057510	94057989	Cldn14 (-48668), Sim2 (-28311)
chr16	95979203	95979519	Ets2 (+55679), Psmg1 (+233206)
chr16	95995796	95996097	Ets2 (+72265), Psmg1 (+216620)
chr17	5011512	5011895	Arid1b (+16630), Tmem242 (+428556)
chr17	10380494	10380692	Pabpc6 (-518024), Qk (+131633)
chr17	13743457	13746115	Mllt4 (-152769), Smok4a (+30464)
chr17	13911136	13911387	Gm7168 (-174118), Mllt4 (+13707)
chr17	14847569	14847759	Thbs2 (-16395), Wdr27 (+223994)
chr17	15178018	15178623	2210404J11Rik (-284), Tcte3 (+202)
chr17	15950608	15950912	Rgmb (+12790), Chd1 (+108829)
chr17	25896831	25897067	Narfl (-13772), Msln (-5677)
chr17	27298995	27299442	Mnf1 (-28358), Ip6k3 (+5490)
chr17	33811718	33812117	Hnrnpm (+11887), Pram1 (+36917)
chr17	34256094	34256550	Brd2 (+2370), H2-Oa (+27037)
chr17	35618079	35618232	Pou5f1 (-24826), H2-Q10 (+11122)
chr17	43892632	43892898	Rcan2 (-46035), Cyp39a1 (+88391)
chr17	56339395	56339517	D17Wsu104e (-16113), Dpp9 (+18856)
chr17	60653649	60653794	NONE
chr17	65987738	65987998	Vapa (-24973), Txndc2 (+3676)
chr17	66713720	66713971	Soga2 (+85244), Ddx11 (+240986)
chr17	69054628	69055037	Epb4.1l3 (-451317), L3mbtl4 (+431696)
chr17	69816419	69816669	Dlgap1 (-957590), A330050F15Rik (+27878)
chr17	71834308	71834443	Smchd1 (-9693), Ndc80 (+41821)
chr17	71834471	71834609	Smchd1 (-9857), Ndc80 (+41657)
chr17	75509211	75509527	Rasgrp3 (-355509), Ltbp1 (+104461)
chr17	80606236	80606439	Ttc39d (-462), Srsf7 (-318)
chr17	84581105	84581330	Haao (-335088), Zfp36l2 (+6069)
chr17	86824334	86824518	Gm10309 (+80146), Prkce (+257301)
chr17	87452962	87453212	Gm5499 (-24524), Cript (+28197)
chr17	87608182	87608477	Mcfd2 (+56945), Socs5 (+101311)
chr17	89174552	89174875	Lhcgr (+16602), Gtf2a1l (+106714)
chr18	3004771	3005630	Vmn1r238 (+118264)
chr18	6543924	6544060	4921524L21Rik (-59639), Epc1 (-53138)
chr18	6544599	6544711	4921524L21Rik (-58976), Epc1 (-53801)
chr18	10547960	10549005	Esco1 (+61867), Greb1l (+223306)

chr18	10569683	10569963	Esco1 (+40527), Greb1l (+244646)
chr18	12990093	12990353	Cabyr (+90390), Osbpl1a (+110127)
chr18	13201410	13201953	Hrh4 (+36183), Zfp521 (+929560)
chr18	14806362	14806473	Ss18 (+35005), Gm5160 (+224259)
chr18	15592764	15593169	Aqp4 (-31466), Chst9 (+283588)
chr18	18051314	18051426	NONE
chr18	24306003	24306426	Galnt1 (-57630), Ino80c (-25761)
chr18	24826363	24826870	Mocos (+14425), Gm9955 (+41232)
chr18	34602257	34602558	Pkd2l2 (+33331), Fam13b (+64069)
chr18	35803792	35803919	Slc23a1 (-16958), Mzb1 (+5165)
chr18	35815163	35815453	Prob1 (-467)
chr18	35815494	35815670	Prob1 (-741)
chr18	49829559	49829753	Dtwd2 (+85599)
chr18	51835086	51835240	Gm4950 (+190372), Prr16 (+557771)
chr18	57845035	57845163	Slc12a2 (-193233), 1700011I03Rik (+151665)
chr18	61541009	61541432	Ppargc1b (+18833), Pde6a (+161085)
chr18	66145885	66146203	Cplx4 (-16212), Lman1 (+16247)
chr18	75040100	75040260	Lipg (+80737), Acaa2 (+101329)
chr18	75040666	75040986	Lipg (+80091), Acaa2 (+101975)
chr18	76613456	76613637	Skor2 (-481597), Smad2 (+212313)
chr18	81797204	81797484	Sall3 (-614027), Galr1 (+778825)
chr18	83740362	83740551	Tshz1 (+515339), Zfp516 (+660186)
chr18	84588162	84588554	Zfp407 (+170538), Zadh2 (+330808)
chr18	87314713	87314808	Gm5096 (-610986), Cbln2 (+434259)
chr19	3767959	3768195	Suv420h1 (+656)
chr19	5803685	5803874	Scyl1 (-32379), Frmd8 (+71494)
chr19	6005129	6005459	Slc22a20 (-19151), Capn1 (+10531)
chr19	6363806	6363988	Sf1 (+207)
chr19	9441752	9441926	Pcna-ps2 (+83970), Stxbp3b (+191899)
chr19	9633435	9633601	Stxbp3b (+220)
chr19	29703448	29703901	Ermp1 (+19230), C030046E11Rik (+106903)
chr19	34812509	34812659	Slc16a12 (+9195), Ifit1 (+97223)
chr19	44076755	44076987	Cpn1 (-15825), Cyp2c44 (+26866)
chr19	46654640	46655055	Wbp1l (-18756), Sfxn2 (+6993)
chr19	55776609	55776938	Tcf7l2 (-39600), Vti1a (+385934)
chr19	57565272	57565511	Atrnl1 (-120132), Trub1 (+37996)
chr19	58592967	58593225	Pnlip (-151759), Ccdc172 (+6604)
chr2	3356844	3357148	Dclre1c (+15574), Suv39h2 (+35307)
chr2	16187859	16187954	Plxdc2 (-90024), Gm13318 (+224154)
chr2	18461040	18461299	Dnajc1 (-146713), Commd3 (-132880)
chr2	20196979	20197177	Etl4 (-244340), Otud1 (+617389)
chr2	20821416	20821899	Arhgap21 (+67689), Etl4 (+380240)
chr2	22443076	22443924	Gad2 (-34225), Myo3a (+294370)
chr2	28980310	28981010	Setx (+148)
chr2	34611014	34611361	Gapvd1 (-25021), Hspa5 (-16424)
chr2	51198320	51198829	Tas2r134 (-284456), Rnd3 (-193944)
chr2	57246848	57247325	Galnt5 (-603208), Gpd2 (+156379)
chr2	57481881	57482153	Galnt5 (-368278), Gpd2 (+391309)

chr2	80302072	80302264	Frzb (-14386), Nckap1 (+119369)
chr2	83373876	83374011	Zc3h15 (-110648), Fsip2 (+590153)
chr2	98502287	98503234	Gm10801 (+367)
chr2	98503710	98504293	Gm10801 (+1608), Gm10800 (+3456)
chr2	98505008	98505437	Gm10800 (+2235), Gm10801 (+2829)
chr2	98506338	98507542	Gm10800 (+518)
chr2	99785651	99785785	NONE
chr2	105308425	105308582	Pax6 (-200549), Rcn1 (-69028)
chr2	111925966	111926040	Olfr1313 (-13263), Olfr1314 (+6853)
chr2	117059330	117059501	Fam98b (-16059), Spred1 (+112306)
chr2	121962909	121963298	Spg11 (-18982), B2m (-10318)
chr2	124986392	124986659	Dut (-86400), Slc12a1 (+8285)
chr2	125443221	125443497	Fbn1 (-111238), Cep152 (+7490)
chr2	131233744	131234170	Smox (-83722), Rnf24 (-55340)
chr2	142290456	142290537	Kif16b (+436770)
chr2	146335546	146336272	Ralgapa2 (+1831), Insm1 (+288252)
chr2	162688176	162688554	Ptprt (-201482), Srsf6 (-68899)
chr2	166065850	166066134	Sulf2 (-84829), Prex1 (+473340)
chr2	167702233	167702451	Ptpn1 (-55485), Cebpb (+187927)
chr2	167702530	167702802	Ptpn1 (-55161), Cebpb (+188251)
chr2	170945712	170946046	Dok5 (+388572), Cbln4 (+923087)
chr3	22026150	22026308	Tbl1xr1 (+50655)
chr3	23209200	23209294	NONE
chr3	31260640	31260800	Kcnmb2 (-540709), Slc7a14 (-51420)
chr3	52059270	52059807	Maml3 (-150726), Foxo1 (-12720)
chr3	58469472	58469815	Fam194a (-28515), Siah2 (+26678)
chr3	58469948	58470228	Fam194a (-28959), Siah2 (+26234)
chr3	61949655	61949811	B430305J03Rik (-779860), Arhgef26 (-192533)
chr3	64230774	64230952	Vmn2r4 (-11645), Vmn2r5 (+82794)
chr3	68096244	68096863	Il12a (-398012), Schip1 (+227690)
chr3	69490627	69490976	B3galnt1 (-88019), Nmd3 (-35105)
chr3	77059606	77059705	NONE
chr3	93518698	93518941	Tdpoz4 (-81500), S100a10 (+159818)
chr3	95276043	95276623	Ctsk (-26875), Arnt (+38023)
chr3	96024557	96024756	Hist2h2be (-387), Hist2h2ac (+146), Hist2h2ab (+869)
chr3	103897340	103897866	Phtf1 (+125570), Magi3 (+126423)
chr3	106243488	106243658	BC051070 (-34473), 2010016I18Rik (+42782)
chr3	109061732	109061903	Vav3 (-81753), Slc25a24 (+135751)
chr3	115609465	115610475	Dph5 (+18869), Slc30a7 (+100354)
chr3	124418625	124418909	Ndst4 (-688242), 1700003H04Rik (-134758)
chr3	142973853	142974253	Pkn2 (-429085), Lmo4 (+891228)
chr3	144467409	144467757	Clca1 (-43778), Clca2 (+14875)
chr3	146005710	146006075	Ssx2ip (-61713), Lpar3 (+121968)
chr3	152996314	152997030	St6galnac5 (-351462), St6galnac3 (+391354)
chr3	154085430	154086146	Tyw3 (+174271), Slc44a5 (+449388)
chr3	159218929	159219028	Rpe65 (-43218), Depdc1a (+60582)
chr4	3429671	3429955	Vmn1r3 (-317308), Tmem68 (+72187)
chr4	3540699	3541766	Lyn (-64029), Tgs1 (+39211)

chr4	8154834	8155634	Car8 (+10954)
chr4	8619372	8619621	Clvs1 (-576967), Chd7 (+1429)
chr4	9133821	9134179	Clvs1 (-62464), Chd7 (+515932)
chr4	11324250	11324666	1110037F02Rik (-88647), Esrp1 (-10528)
chr4	17379647	17379762	Mmp16 (-400900)
chr4	37908025	37908169	NONE
chr4	41811878	41812712	Ccl27a (-91280), Gm20938 (-24489)
chr4	41925351	41926345	Gm2564 (+17616), Gm20938 (+89064)
chr4	41963771	41963931	Gm13304 (-21689), Gm2564 (-20387)
chr4	42023793	42024677	Gm13306 (-528)
chr4	42269110	42269783	Gm13305 (-230121), Il11ra2 (-69813)
chr4	42353908	42354512	Il11ra2 (+14950), Gm13298 (+187497)
chr4	42414654	42414814	Il11ra2 (+75474), Gm13298 (+126973)
chr4	42452240	42453234	Gm13298 (+88970), Il11ra2 (+113477)
chr4	42565486	42566480	4930578G10Rik (-183094), Gm13298 (-24276)
chr4	42603910	42604070	4930578G10Rik (-145087), Gm13298 (-62283)
chr4	42663761	42664645	Gm13298 (-122496), 4930578G10Rik (-84874)
chr4	42748677	42749163	4930578G10Rik (-157)
chr4	46596519	46596890	Trim14 (-47692), Coro2a (+18369)
chr4	50709855	50710061	Grin3a (-851537), Cylc2 (-519592)
chr4	80355629	80355771	Tyrp1 (-124434)
chr4	81597155	81597635	Mpdz (-508684), Nfib (+553817)
chr4	83269166	83269399	Ccdc171 (+97834)
chr4	101763903	101764474	Pde4b (-163419), Gm12789 (+104958)
chr4	107831578	107831728	Scp2 (-40516), Echdc2 (-6389)
chr4	114943768	114943896	Cyp4x1 (-137250), Cyp4a12a (-27819)
chr4	114986056	114986184	Cyp4a12b (-98109), Cyp4a12a (+14469)
chr4	115718594	115718991	Faah (-28262), Nsun4 (+7192)
chr4	115893683	115893927	Pik3r3 (-461)
chr4	116862455	116862794	Kif2c (-7381), Gm1661 (+61897)
chr4	116875066	116875292	Kif2c (-19935), Gm1661 (+49343)
chr4	116883412	116883551	Kif2c (-28238), Gm1661 (+41040)
chr4	117559595	117559777	B4galt2 (-3594), Atp6v0b (+248)
chr4	118110022	118110315	Cdc20 (-212)
chr4	118219655	118221372	Tmem125 (-4181)
chr4	120709969	120710060	Col9a2 (-1975)
chr4	123369561	123369835	Macf1 (-8095), Ndufs5 (+25747)
chr4	125568410	125568787	Csf3r (-133195), Grik3 (+400655)
chr4	131301601	131301835	Ptpru (+92428), Matn1 (+801418)
chr4	131673819	131674044	Epb4.1 (-42725), Oprd1 (+26469)
chr4	132894275	132894409	Wdtc1 (+897)
chr4	132994761	132995022	Fam46b (-41155), Slc9a1 (+69271)
chr4	141102627	141102930	Spen (-8267), Fblim1 (+53060)
chr4	146316224	146316405	Gm13150 (-29770), Zfp600 (+191224)
chr4	149185046	149185351	Slc25a33 (-36813), Spsb1 (+143953)
chr4	149612437	149612864	Rere (-43104), Eno1 (+1590)
chr4	153394885	153395163	Smim1 (+5237), Lrrc47 (+9184)
chr4	153527347	153527613	Tprgl (+7295), Wrap73 (+10999)

chr4	154231384	154231576	Mmel1 (-14258), Ttc34 (+1171)
chr4	154306287	154306408	Tnfrsf14 (-4162)
chr5	11220304	11220562	4933402N22Rik (-374523), Gm5861 (+37361)
chr5	11220573	11220751	4933402N22Rik (-374294), Gm5861 (+37590)
chr5	11400416	11400736	4933402N22Rik (-194380), Gm5861 (+217504)
chr5	11829495	11829872	Sema3d (-553482), 4933402N22Rik (+234728)
chr5	11881170	11881428	Sema3d (-501867), 4933402N22Rik (+286343)
chr5	14918464	14922081	Gm9758 (-5374), Speer4e (+18156)
chr5	16889535	16889710	Speer4f (-92293), Gm3495 (+10597)
chr5	16974433	16975269	Speer4f (-7065), Gm3495 (+95825)
chr5	28661168	28661314	Evc2 (-9068497), Rbm33 (+17512), Shh (+132400)
chr5	54893762	54893923	Stim2 (+504187)
chr5	66330250	66330572	Chrna9 (-27952), Rhoh (+75603)
chr5	66375177	66375388	Chrna9 (+16920), 9130230L23Rik (+20241)
chr5	72438606	72438857	Commd8 (+120690), Gabrb1 (+347561)
chr5	76651154	76651299	Tmem165 (+38322), Clock (+82590)
chr5	76657257	76657453	Tmem165 (+44450), Clock (+76462)
chr5	78845888	78845995	NONE
chr5	102235572	102235858	Cds1 (+41566), Wdfy3 (+263225)
chr5	105855709	105855904	Lrrc8c (-92600), Lrrc8b (+11013)
chr5	108295958	108296386	Evi5 (+7954), Ube2d2b (+36900)
chr5	109111463	109111867	Fgfrl1 (-11583), Slc26a1 (-7077)
chr5	109176965	109177370	Rnf212 (+1299), Fgfrl1 (+53920)
chr5	109231240	109231381	Tmed11 (-6929), Vmn2r8 (+6462)
chr5	109722239	109722366	Vmn2r14 (-68662), Vmn2r15 (+4272)
chr5	113763814	113764285	Sgsm1 (-24244), Aym1 (-22264)
chr5	115940244	115940408	Pxn (-16491), Sirt4 (-5592)
chr5	123453058	123453399	Kdm2b (-14121), Orai1 (-11854)
chr5	123960959	123961289	B3gnt4 (+655)
chr5	123961314	123961799	B3gnt4 (+1088), Diablo (+12628)
chr5	129661847	129662082	Sfswap (-345141), Gpr133 (+59340)
chr5	129760523	129760777	Sfswap (-246456), Gpr133 (+158025)
chr5	135003226	135003818	Gtf2ird1 (-70941), Clip2 (+24782)
chr5	137540337	137540712	Ap1s1 (-18520), Serpine1 (+7617)
chr5	140277688	140278199	Mafk (+10477), Tmem184a (+12293)
chr5	143455156	143455542	Tnrc18 (+123717), Slc29a4 (+277294)
chr5	143579353	143579517	Tnrc18 (-369)
chr5	143579622	143579759	Tnrc18 (-625)
chr5	148910391	148910631	Mtus2 (+141615), Slc7a1 (+300969)
chr6	8115351	8115691	Col28a1 (+10873), C1galt1 (+320679)
chr6	12106365	12106464	Gm6578 (-46832), Thsd7a (+592838)
chr6	17741408	17741763	St7 (+97554), Wnt2 (+238999)
chr6	17741787	17742080	St7 (+97902), Wnt2 (+238651)
chr6	22299685	22299898	Fam3c (+6374), Wnt16 (+61562)
chr6	30677243	30677546	Cep41 (-33646), Mest (-10668)
chr6	42445265	42445364	Olfr456 (-8102), Fam115e (+102056)
chr6	44742921	44743299	Cntnap2 (-266950)
chr6	47091238	47091832	Cul1 (-312784)

chr6	47110170	47110764	Cul1 (-293852)
chr6	51419242	51419535	Hnrnpa2b1 (-1961), Cbx3 (-970)
chr6	51936401	51936542	Skap2 (+26076), Snx10 (+462572)
chr6	54927446	54927856	Nod1 (-5045), Ggct (+15293)
chr6	54927977	54928346	Nod1 (-5556), Ggct (+14782)
chr6	58767579	58767663	Herc3 (-16073), Abcg2 (+220956)
chr6	58775563	58775687	Herc3 (-8069), Abcg2 (+228960)
chr6	66579484	66579697	Vmn1r33 (-13535), Vmn1r34 (+8155)
chr6	68200903	68201280	Tacstd2 (-715276), Igkv4-71 (+992347)
chr6	79075351	79075546	Lrrtm4 (-893422), Reg3g (-656583)
chr6	83417694	83418110	Tet3 (-26233), Dguok (+39061)
chr6	83530301	83530768	Stambp (-8037), Clec4f (+75646)
chr6	87931409	87931718	Gm5577 (-113), H1fx (-88)
chr6	88444425	88445476	Sec61a1 (+23948), Ruvbl1 (+29548)
chr6	89950379	89950539	Vmn1r47 (-21423), Vmn1r46 (+24294)
chr6	92356089	92356462	Trh (-161632), Prickle2 (+161084)
chr6	103599005	103599368	Cntn6 (-843729), Chl1 (+138317)
chr6	116201111	116202043	Zfand4 (-12663), Fam21 (+43526)
chr6	121600081	121600259	Mug1 (-188389), A2m (+13979)
chr6	126489214	126489669	Kcna5 (-3869)
chr6	143101173	143101323	C2cd5 (-52621), Etnk1 (-14502)
chr6	145881183	145881640	Sspn (-1230)
chr6	147501623	147501803	Far2 (-494225), Ccdc91 (+77320)
chr7	3000020	3000235	Gm7353 (+111822)
chr7	4999282	4999547	Zfp580 (-3719)
chr7	4999598	4999859	Zfp580 (-3405)
chr7	5543821	5543962	Vmn1r62 (-82337), Vmn1r60 (-47161)
chr7	6147041	6147413	Galp (-1465)
chr7	11787109	11787324	Zscan4d (-35720), Zscan4e (+108814)
chr7	16994559	16994776	Tmem160 (-43460), Zc3h4 (+8438)
chr7	27031275	27031464	Vmn1r184 (-20480), Cyp2b9 (+72940)
chr7	29827407	29827755	Map4k1 (+59708), Ryr1 (+82589)
chr7	29958939	29959245	Ggn (+3853), Psmd8 (+6600)
chr7	31306647	31307327	Arhgap33 (+13092), Prodh2 (+28346)
chr7	31307390	31308185	Arhgap33 (+12291), Prodh2 (+29147)
chr7	39231479	39233273	Pop4 (-176009), Gm5591 (+80836)
chr7	39352222	39353766	Gm5591 (-39782)
chr7	50122401	50122572	Gm2381 (+117)
chr7	50509190	50509325	EU599041 (+30313), Zfp715 (+57598)
chr7	52964901	52965056	Rpl18 (-5848), Dbp (+4521)
chr7	52965206	52965362	Rpl18 (-5543), Dbp (+4826)
chr7	52965417	52966222	Rpl18 (-5007), Dbp (+5362)
chr7	72747499	72747658	Tjp1 (-231454), Tarsl2 (-42205)
chr7	72750212	72750303	Tjp1 (-234133), Tarsl2 (-39526)
chr7	87853701	87853957	Crtc3 (-20066), Iqgap1 (+94388)
chr7	88903178	88903348	Fam103a1 (-4548)
chr7	90974226	90974809	1700026D08Rik (-31128), Mesdc1 (+58333)
chr7	91561750	91562035	Arnt2 (-3424)

chr7	95678989	95679332	Rab38 (+100378), Tmem135 (+808136)
chr7	96802485	96802569	A230065N10Rik (-20893), Ccdc81 (+249601)
chr7	107520727	107520893	Ppme1 (-404), C2cd3 (+67)
chr7	108458149	108458626	Atg16l2 (-7786), Stard10 (-7212)
chr7	108789576	108789755	Art2b (-59991), Clpb (-22481)
chr7	115820499	115820711	Olfr512 (-36229), Olfr510 (+9673)
chr7	117817849	117818128	Ampd3 (-93731), Adm (+46814)
chr7	118226353	118226835	Eif4g2 (-50)
chr7	120444291	120444522	Btbd10 (+68446), Arntl (+93428)
chr7	121290447	121290617	Rras2 (-29237), Copb1 (+107662)
chr7	122814005	122814257	Gm6816 (-742891), Sox6 (+324469)
chr7	124444366	124444948	Xylt1 (-79836), Nucb2 (+796769)
chr7	130559517	130559817	Aqp8 (-46141), Lcmt1 (+38171)
chr7	133056871	133057003	Gsg1l (+168988), D430042O09Rik (+205508)
chr7	134171142	134171368	Maz (-1262)
chr7	138559852	138560095	Hmx3 (-126407), Acadsb (+5859)
chr7	141150079	141150413	Fank1 (+181702), Adam12 (+266534)
chr7	148151421	148151685	Ifitm1 (-2386)
chr8	4944835	4945023	Shcbp1 (-165395), Slc10a2 (+160303)
chr8	10676257	10676741	3930402G23Rik (+251958), Myo16 (+403927)
chr8	12651761	12651915	Tubgcp3 (+20410), Spaca7 (+78789)
chr8	15519762	15520194	Myom2 (+462325)
chr8	25462270	25462920	1810011O10Rik (+86823), A730045E13Rik (+178219)
chr8	31438279	31438743	Dusp26 (-761432)
chr8	48678169	48678670	Rwdd4a (+59402), Ing2 (+82092)
chr8	48799209	48799365	Cdkn2aip (-2)
chr8	53422254	53422419	NONE
chr8	62368986	62369151	Gm10283 (+612926), BC030500 (+978517)
chr8	68498956	68499055	March1 (+357082), Tma16 (+511412)
chr8	82479259	82479547	Hhip (+102502), Anapc10 (+243684)
chr8	85884272	85884590	Clgn (-29346), 4933434I20Rik (+12061)
chr8	89571694	89572030	N4bp1 (-162705), Cbln1 (+424646)
chr8	93860665	93860961	Fto (+23382), Irx3 (+464740)
chr8	97794284	97794571	Kifc3 (-127988), Cngb1 (+13653)
chr8	107828141	107828923	Elmo3 (-969)
chr8	110220901	110221068	Psmd7 (-108603)
chr8	112302324	112302543	Ap1g1 (-20)
chr8	119567477	119567793	Gcsh (-50198), Pkd1l2 (+38714)
chr8	122124325	122124484	Taf1c (+4717), Dnaaf1 (+25270)
chr8	122125009	122125167	Taf1c (+4034), Dnaaf1 (+25953)
chr8	122126044	122126895	Taf1c (+2652), Dnaaf1 (+27335)
chr8	124494912	124495249	Gm22 (-298388), Banp (+20637)
chr8	125801296	125801869	Vps9d1 (-23335), Fanca (+40893)
chr8	127528426	127529090	Egln1 (-55604), Tsnax (-8139)
chr9	2999998	3002205	Gm10722 (+180)
chr9	3002257	3003490	Gm10720 (-12780), Gm10722 (+1952)
chr9	3003894	3004721	Gm10720 (-11346), Gm10722 (+3386)
chr9	3008887	3009792	Gm10720 (-6314), Gm10722 (+8418)

chr9	3011305	3013245	Gm10720 (-3379)
chr9	3013318	3014219	Gm10720 (-1885)
chr9	3015054	3015859	Gm10720 (-197)
chr9	3016808	3017613	Gm10718 (-6336), Gm10720 (+1557)
chr9	3017722	3018549	Gm10718 (-5411), Gm10720 (+2482)
chr9	3023697	3024501	Gm10718 (+552)
chr9	3024619	3028825	Gm10718 (+3175), 4930433N12Rik (+173060)
chr9	3029353	3030835	Gm10718 (+6547), 4930433N12Rik (+169688)
chr9	3030999	3033810	Gm10718 (+8858), 4930433N12Rik (+167377)
chr9	3035278	3036770	Gm10718 (+12477), 4930433N12Rik (+163758)
chr9	3036840	3037472	Gm10718 (+13609), 4930433N12Rik (+162626)
chr9	3037542	3038423	Gm10718 (+14436), 4930433N12Rik (+161799)
chr9	9927813	9927959	Arhgap42 (-688873), Cntn5 (+976889)
chr9	10564871	10565170	Cntn5 (+339754)
chr9	13338640	13339469	Maml2 (-85379), Phxr4 (+104474)
chr9	19029130	19029345	Olfr843 (+24669), Olfr836 (+103837)
chr9	25059849	25060024	Sept7 (-232)
chr9	28182826	28182970	Opcml (+584366)
chr9	35112756	35113275	Cdon (-116166), Rpusd4 (+37566)
chr9	37980371	37980508	Olfr891 (+7966), Olfr890 (+29718)
chr9	41056175	41056341	Ubash3b (-90113), Sorl1 (+876122)
chr9	41780734	41780824	Ubash3b (-814634), Sorl1 (+151601)
chr9	44142700	44142969	H2afx (+37)
chr9	44689301	44689523	Mll1 (-33)
chr9	45064717	45064951	Tmprss4 (-52659), Il10ra (+12398)
chr9	45443886	45444033	Cep164 (+192814), Dscaml1 (+205584)
chr9	49183419	49183597	Drd2 (+34776), Ankk1 (+51618)
chr9	56071355	56071791	Tspan3 (-62977), C230081A13Rik (+194284)
chr9	58538904	58539073	2410076I21Rik (+50377), Nptn (+108890)
chr9	58743557	58744359	Hcn4 (+72639), Neo1 (+140290)
chr9	59505862	59506188	Gramd2 (-49419), Pkm (+1642)
chr9	67419497	67419783	C2cd4b (-187604), Tln2 (-12130)
chr9	70942604	70942770	Lipc (-160072), Aqp9 (+68409)
chr9	71908716	71909086	Cgnl1 (-289492), Tcf12 (+50725)
chr9	75471052	75471301	Lysmd2 (-2362)
chr9	82868925	82869191	Phip (+38)
chr9	82869297	82869515	Phip (-310)
chr9	96435662	96435871	Rnf7 (-56673), Rasa2 (+96269)
chr9	99341156	99341459	Mras (-4171)
chr9	100119352	100119755	Sox14 (-342891), Il20rb (+267653)
chr9	100119779	100119975	Sox14 (-343214), Il20rb (+267330)
chr9	108465913	108466233	Impdh2 (+3299), Ndufaf3 (+3600)
chr9	108466612	108467145	Ndufaf3 (+2794), Impdh2 (+4105)
chr9	110154884	110154975	Elp6 (-52766), Cspg5 (+8643)
chr9	110428162	110428353	Setd2 (-6843), Kif9 (+48760)
chr9	110428485	110429550	Setd2 (-6083), Kif9 (+49520)
chr9	112911176	112911399	Arpp21 (-776283), Pdcd6ip (+706074)
chr9	115695410	115695590	Stt3b (-475961), Gadl1 (-122182)

chr9	115695612	115695770	Stt3b (-476152), Gadl1 (-121991)
chr9	118647931	118648304	Ctdspl (-187453), Itga9 (+132310)
chr9	120058537	120058711	Mobp (-236)
chr9	123370836	123371309	Limd1 (-16746), Lars2 (+95015)
chrX	3302194	3302325	Gm14345 (-108408)
chrX	3376735	3376894	Gm14345 (-33853)
chrX	3512215	3512374	Gm14351 (-48975), Gm3701 (+155142)
chrX	3873121	3873280	Gm3701 (-205764), Gm14347 (-33810)
chrX	4041451	4041610	Gm3763 (-556635), Gm10922 (-45958)
chrX	25615314	25615470	Gm5168 (-137225), Gm2012 (+43089)
chrX	30762714	30762936	Gm21637 (-60425), Gm2799 (+164429)
chrX	31786255	31786414	Gm2927 (-45925), Gm2913 (-33869)
chrX	58586282	58586381	Ldoc1 (-376457), Cdr1 (-147599)
chrX	59903462	59903624	4931400O07Rik (+182377)
chrX	105879298	105879495	Gm732 (-735622), Brwd3 (+150290)
chrX	113580034	113580205	H2afb2-ps (-214667), Cpxcr1 (+17567)
chrX	115952722	115952852	Tgif2lx1 (-358441)
chrX	120113639	120114527	Nap1l3 (-603089), 3110007F17Rik (-103085)
chrX	120405249	120406500	Srsx (-69785), Vmn2r121 (+843644)
chrX	120634626	120635877	Srsx (-299162), Vmn2r121 (+614267)
chrX	120970533	120970966	Srsx (-634660), Vmn2r121 (+278769)
chrX	121577239	121578865	Vmn2r121 (-328533)
chrX	121706644	121708339	Vmn2r121 (-457973)
chrX	122458051	122459744	4932411N23Rik (+891718)
chrX	130684776	130684881	Nox1 (-38436), Xkrx (+11638)
chrX	136596335	136596619	E230019M04Rik (-941), Nup62cl (+630)
chrX	137431184	137431272	Tex13 (-83256), Vsig1 (-10919)
chrX	150514181	150514469	Gm15140 (+40903), Spin2 (+247489)
chrX	150565454	150565742	Gm15140 (-10370), 4930524N10Rik (+212337)
chrX	150587185	150587473	Gm15140 (-32101), 4930524N10Rik (+190606)
chrX	157555276	157555507	Cdkl5 (-122758), Scml2 (-45600)
chrX	159793558	159793691	S100g (-391094), Grpr (+193984)
chrX	166438853	166447041	4933400A11Rik (-225380)
chrY	2869581	2872285	Gm10352 (+480543)

7 .1 Testes Only GO Term Associations

Nucleosome Organisation

nearest gene	distance to TSS
Brd2	2370
H1fx	-88
H2afb2-ps	-214667
H2afx	37
H2afy	36985
H3f3b	3557
Hist1h1a	-1426
Hist1h1b	-2551
Hist1h1d	187
Hist1h1e	203
Hist1h2aa	-422
Hist1h2ab	-4067
Hist1h2ac	-323
Hist1h2ae	-4759
Hist1h2ak	-480
Hist1h2an	-573
Hist1h2ba	116
Hist1h2bb	287
Hist1h2bc	72
Hist1h2be	-1175
Hist1h2bf	-1783
Hist1h2bm	-4192
Hist1h2bp	-4285, +330
Hist1h2bq	-167
Hist1h3a	144
Hist1h3b	268
Hist1h3c	-1419
Hist1h3d	216
Hist1h3h	224
Hist1h3i	-4394, +221
Hist1h4a	-1012
Hist1h4b	-4402
Hist1h4f	-3571
Hist1h4k	-3882
Hist1h4m	-1395
Hist2h2ab	869
Hist2h2ac	146
Hist2h2be	-387
Nap1l3	-603089
Ptma	-39253, -102
Setd2	-6843, -6083

7.2 Testes Peaks GO Term Associations

Nucleosome Assembly

nearest gene	distance to TSS
Brd2	2370
H1fx	-88
H2afb2-ps	-214667
H2afx	37
H2afy	36985
H3f3b	3557
Hist1h1a	-1426
Hist1h1b	-2551
Hist1h1d	187
Hist1h1e	203
Hist1h2aa	-422
Hist1h2ab	-4067
Hist1h2ac	-323
Hist1h2ae	-4759
Hist1h2ak	-480
Hist1h2an	-573
Hist1h2ba	116
Hist1h2bb	287
Hist1h2bc	72
Hist1h2be	-1175
Hist1h2bf	-1783
Hist1h2bm	-4192
Hist1h2bp	-4285, +330
Hist1h2bq	-167
Hist1h3a	144
Hist1h3b	268
Hist1h3c	-1419
Hist1h3d	216
Hist1h3h	224
Hist1h3i	-4394, +221
Hist1h4a	-1012
Hist1h4b	-4402
Hist1h4f	-3571
Hist1h4k	-3882
Hist1h4m	-1395
Hist2h2ab	869
Hist2h2ac	146
Hist2h2be	-387
Nap1l3	-603089

7.3 Testes Peaks GO Term Associations

Chromatin Assembly or Disassembly

nearest gene	distance to TSS
Brd2	2370
Chd1	108829
H1fx	-88
H2afb2-ps	-214667
H2afx	37
H2afy	36985
H3f3b	3557
Hist1h1a	-1426
Hist1h1b	-2551
Hist1h1d	187
Hist1h1e	203
Hist1h2aa	-422
Hist1h2ab	-4067
Hist1h2ac	-323
Hist1h2ae	-4759
Hist1h2ak	-480
Hist1h2an	-573
Hist1h2ba	116
Hist1h2bb	287
Hist1h2bc	72
Hist1h2be	-1175
Hist1h2bf	-1783
Hist1h2bm	-4192
Hist1h2bp	-4285, +330
Hist1h2bq	-167
Hist1h3a	144
Hist1h3b	268
Hist1h3c	-1419
Hist1h3d	216
Hist1h3h	224
Hist1h3i	-4394, +221
Hist1h4a	-1012
Hist1h4b	-4402
Hist1h4f	-3571
Hist1h4k	-3882
Hist1h4m	-1395
Hist2h2ab	869
Hist2h2ac	146
Hist2h2be	-387
Nap1l3	-603089
Suv39h2	35307

7.4 Testes Peaks GO Term Associations

Chromatin Assembly

nearest gene	distance to TSS
Banp	20637
Brd2	2370
Cbx3	-970
Chd1	108829
Chd7	1429, +515932
Clock	76462, +82590
Crebbp	-87866
Epc1	-53801, -53138
Foxa1	-614040
H1fx	-88
H2afb2-ps	-214667
H2afx	37
H2afy	36985
H3f3b	3557
Hist1h1a	-1426
Hist1h1b	-2551
Hist1h1d	187
Hist1h1e	203
Hist1h2aa	-422
Hist1h2ab	-4067
Hist1h2ac	-323
Hist1h2ae	-4759
Hist1h2ak	-480
Hist1h2an	-573
Hist1h2ba	116
Hist1h2bb	287
Hist1h2bc	72
Hist1h2be	-1175
Hist1h2bf	-1783
Hist1h2bm	-4192
Hist1h2bp	-4285, +330
Hist1h2bq	-167
Hist1h3a	144
Hist1h3b	268
Hist1h3c	-1419
Hist1h3d	216
Hist1h3h	224
Hist1h3i	-4394, +221
Hist1h4a	-1012
Hist1h4b	-4402
Hist1h4f	-3571
Hist1h4k	-3882
Hist1h4m	-1395
Hist2h2ab	869
Hist2h2ac	146
Hist2h2be	-387
Ing2	82092
Jarid2	138780
Kansl2	118
Kdm2b	-14121
Kdm5b	564, +810, +1064
L3mbtl4	431696

Mll1	-33
Nap1l3	-603089
Prkaa1	5887
Ptma	-39253, -102
Rbbp5	-55727
Rcor1	63748
Rere	-43104
Ruvbl1	29548
Setd2	-6843, -6083
Setdb2	-4993
Suv39h2	35307
Suv420h1	656
Taf9	11308
Tbl1xr1	50655
Tet3	-26233
Zmynd11	-69225

Bibliography

Adams, C.C. & Workman, J.L., 1995. Binding of Disparate Transcriptional Activators to Nucleosomal DNA Is Inherently Cooperative. *Molecular and Cellular Biology*, 15(3), pp.1405–1421.

Adikusuma, F. et al., 2017. Functional Equivalence of the SOX2 and SOX3 Transcription Factors in the Developing Mouse Brain and Testes. *Genetics*, 206(3), pp.genetics.117.202549–1503.

Agalioti, T. et al., 2000. Ordered Recruitment of Chromatin Modifying and General Transcription Factors to the IFN-β Promoter. *Cell*, 103(4), pp.667–678.

Albuisson, J. et al., 2011. Identification of two novel mutations in Shh long-range regulator associated with familial pre-axial polydactyly. *Clinical genetics*, 79(4), pp.371–377.

Alvarez-Dominguez, J.R. et al., 2017. The Super-Enhancer-Derived alncRNA-EC7/Bloodlinc Potentiates Red Blood Cell Development in trans. *Cell Reports*, 19(12), pp.2503–2514.

Austin, C.P. et al., 2004. The knockout mouse project. *Nature Genetics*, 36(9), pp.921–924.

Avilion, A.A. et al., 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & Development*, 17(1), pp.126–140.

Bailey, T. et al., 2013. Practical guidelines for the comprehensive analysis of ChIP-seq data. F. Lewitter, ed. *PLoS computational biology*, 9(11), p.e1003326.

Bailey, T.L. & Machanick, P., 2012. Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Research*, 40(17), pp.gks433–e128.

Bailey, T.L. et al., 2015. The MEME Suite. *Nucleic Acids Research*, 43(W1), pp.W39–49.

Bakrania, P. et al., 2007. SOX2 anophthalmia syndrome: 12 new cases demonstrating broader phenotype and high frequency of large gene deletions. *The British journal of ophthalmology*, 91(11), pp.1471–1476.

Barrangou, R. et al., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315(5819), pp.1709–1712.

Bateman, J.R., Johnson, J.E. & Locke, M.N., 2012. Comparing enhancer action in cis and in trans. *Genetics*, 191(4), pp.1143–1155.

Bell, A.C., West, A.G. & Felsenfeld, G., 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, 98(3), pp.387–396.

Bellucci, A. et al., 2014. The "In Situ" Proximity Ligation Assay to Probe Protein–Protein Interactions in Intact Tissues. In *Exocytosis and Endocytosis*. Methods in Molecular Biology. New York, NY: Humana Press, New York, NY, pp. 397–405.

Benabdallah, N.S. & Bickmore, W.A., 2015. Regulatory Domains and Their Mechanisms. *Cold Spring Harbor Symposia on Quantitative Biology*, 80, pp.45–51.

Benabdallah, N.S. et al., 2017. PARP mediated chromatin unfolding is coupled to long-range enhancer activation. *bioRxiv*, p.155325.

Bensaddek, D. & Lamond, A.I., 2016. Unlocking the chromatin code by deciphering protein-DNA interactions. *Molecular systems biology*, 12(11), p.887.

Bergsland, M. et al., 2011. Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev*, 25(23), pp.2453–2464.

Kornberg, R, 1977. Structure of chromatin. *Annual Reviews of Biochemistry* 46 pp.931-954

Birnbaum, R.Y. et al., 2012. Coding exons function as tissue-specific enhancers of nearby genes. *Genome Research*, 22(6), pp.1059–1068.

Bolotin, A. et al., 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology-Sgm*, 151, pp.2551–2561.

Bošković, A. & Torres-Padilla, M.-E., 2013. How mammals pack their sperm: a variant matter. *Genes Dev*, 27(15), pp.1635–1639.

Boudaoud, I. et al., 2017. Connected Gene Communities Underlie Transcriptional Changes in Cornelia de Lange Syndrome. *Genetics*, 207(1), pp.139–151.

Bowles, J., Schepers, G. & Koopman, P., 2000. Phylogeny of the SOX Family of Developmental Transcription Factors Based on Sequence and Structural Indicators. *Developmental Biology*, 227(2), pp.239–255.

Boyle, M.I. et al., 2015. Cornelia de Lange syndrome. *Clinical genetics*, 88(1), pp.1–12.

Bulger, M. & Groudine, M., 2010. Enhancers: The abundance and function of regulatory sequences beyond promoters. *Developmental Biology*, 339(2), pp.250–257.

Bylund, M. et al., 2003. Vertebrate neurogenesis is counteracted by Sox1–3 activity. *Nature Neuroscience*, 6(11), pp.1162–1168.

Cakouros, D. et al., 2001. A NF-kappa B/Sp1 region is essential for chromatin remodeling and correct transcription of a human granulocyte-macrophage colony-stimulating factor transgene. *Journal of Immunology*, 167(1), pp.302–310.

Calo, E. & Wysocka, J., 2013. Modification of Enhancer Chromatin: What, How, and Why? *Molecular Cell*, 49(5), pp.825–837.

Charron, F. et al., 2003. The morphogen sonic hedgehog is an axonal chemoattractant that collaborates with netrin-1 in midline axon guidance. *Cell*, 113(1), pp.11–23.

Chaya, D. et al., 2001. Transcription factor FoxA (HNF3) on a nucleosome at an enhancer complex in liver chromatin. *Journal of Biological Chemistry*, 276(48), pp.44385–44389.

Chen, Y. et al., 2018. Single-cell RNA-seq uncovers dynamic processes and critical regulators in mouse spermatogenesis. *Cell Research*, 28(9), pp.879–896.

Chew, L.-J. & Gallo, V., 2009. The Yin and Yang of Sox proteins: Activation and repression in development and disease. R. Bansal, W. B. Macklin, & J. de Vellis, eds. *J Neurosci Res*, 87(15), pp.3277–3287.

Chiang, C. et al., 1996. Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature*, 383(6599), pp.407–413.

Collis, P., Antoniou, M. & Grosveld, F., 1990. Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression. *EMBO J*, 9(1), pp.233–240.

Cong, L. et al., 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, 339(6121), pp.819–823.

Creyghton, M.P. et al., 2008. H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment. *Cell*, 135(4), pp.649–661.

Creyghton, M.P. et al., 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, 107(50), pp.21931–21936.

Crémazy, F., Berta, P. & Girard, F., 2000. Sox neuro, a new Drosophila Sox gene expressed in the developing central nervous system. *Mechanisms of Development*, 93(1-2), pp.215–219.

Cunningham, T.J. et al., 2018. Genomic Knockout of Two Presumed Forelimb Tbx5 Enhancers Reveals They Are Nonessential for Limb Development. *Cell Reports*, 23(11), pp.3146–3151.

Dekker, J. et al., 2002. Capturing chromosome conformation. *Science*, 295(5558), pp.1306–1311.

Dickel, D.E. et al., 2018. Ultraconserved Enhancers Are Required for Normal Development. *Cell*, 172(3), pp.491–499.e15.

Dong, B. et al., 2018. Functional redundancy of frizzled 3 and frizzled 6 in planar cell polarity control of mouse hair follicles. *Development (Cambridge, England)*, 145(19), p.dev168468.

Dostie, J. et al., 2006. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Research*, 16(10), pp.1299–1309.

Driggers, R.W. et al., 1999. Isolated bilateral anophthalmia in a girl with an apparently balanced de novo translocation: 46,XX,t(3;11)(q27;p11.2). *American journal of medical genetics*, 87(3), pp.201–202.

Dubois, N.C. et al., 2006. Nestin-Cre transgenic mouse line Nes-Cre1 mediates highly efficient Cre/loxP mediated recombination in the nervous system, kidney, and somite-derived tissues. *Genesis*, 44(8), pp.355–360.

ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), pp.57–74.

Engel, N. et al., 2008. Three-dimensional conformation at the H19/Igf2 locus supports a model of enhancer tracking. *Hum Mol Genet*, 17(19), pp.3021–3029.

Epstein, D.J., McMahon, A.P. & Joyner, A.L., 1999. Regionalization of Sonic hedgehog transcription along the anteroposterior axis of the mouse central nervous system is regulated by Hnf3-dependent and -independent mechanisms. *Development*, 126(2), pp.281–292.

Faast, R. et al., 2001. Histone variant H2A.Z is required for early mammalian development. *Current Biology*, 11(15), pp.1183–1187.

Faivre, L. et al., 2006. Recurrence of SOX2 anophthalmia syndrome with gonosomal mosaicism in a phenotypically normal mother. *American journal of medical genetics. Part A*, 140(6), pp.636–639.

Falvo, J.V. et al., 2000. Assembly of a functional beta interferon enhanceosome is dependent on ATF-2-c-jun heterodimer orientation. *Molecular and Cellular Biology*, 20(13), pp.4814–4825.

Fantes, J. et al., 2003. Mutations in SOX2 cause anophthalmia. *Nature Genetics*, 33(4), pp.461–463.

Feingold, E.A. et al., 2004. The ENCODE (ENCyclopedia of DNA elements) Project. *Science*, 306(5696), pp.636–640.

Gao, Q. & Finkelstein, R., 1998. Targeting gene expression to the head: the Drosophila orthodenticle gene is a direct target of the Bicoid morphogen. *Development*, 125(21), pp.4185–4193.

Gaudelli, N.M. et al., 2017. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, 551(7681), pp.464–471.

Gavrilov, A. et al., 2009. Chromosome conformation capture (from 3C to 5C) and its ChIP-based modification. *Methods in molecular biology (Clifton, N.J.)*, 567(Chapter 12), pp.171–188.

Gibson, G., 2018. Population genetics and GWAS: A primer. *Plos Biology*, 16(3), p.e2005485.

Gill-Sharma, M.K. et al., 2012. Putative molecular mechanism underlying sperm chromatin remodelling is regulated by reproductive hormones. *Clinical epigenetics*, 4(1), p.23.

Golden, J.A., 1999. Towards a greater understanding of the pathogenesis of holoprosencephaly. *Brain & development*, 21(8), pp.513–521.

Govin, J. et al., 2007. Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis. *Journal of Cell Biology*, 176(3), pp.283–294.

Gradwohl, G. et al., 2000. neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. *Proceedings of the National Academy of Sciences*, 97(4), pp.1607–1611.

Graham, V. et al., 2003. SOX2 Functions to Maintain Neural Progenitor Identity. *Neuron*, 39(5), pp.749–765.

Grant, C.E., Bailey, T.L. & Noble, W.S., 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics*, 27(7), pp.1017–1018.

Graves, J.A., 1998. Evolution of the mammalian Y chromosome and sex-determining genes. *The Journal of experimental zoology*, 281(5), pp.472–481.

Griswold, M.D., 2016. Spermatogenesis: The Commitment to Meiosis. *Physiological reviews*, 96(1), pp.1–17.

Gubbay, J. et al., 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature*, 346(6281), pp.245–250.

Guo, N., Hawkins, C. & Nathans, J., 2004. Frizzled6 controls hair patterning in mice. *Proceedings of the National Academy of Sciences*, 101(25), pp.9277–9281.

Hagey, D.W. et al., 2018. SOX2 regulates common and specific stem cell features in the CNS and endoderm derived organs H. Kondoh, ed. *Plos Genetics*, 14(2), p.e1007224.

Hajdu, M. et al., 2016. Transcriptional and post-transcriptional regulation of histone variant H2A.Z during sea urchin development. Development Growth & Differentiation, 58(9), pp.727–740.

Halford, S.E. & Marko, J.F., 2004. How do site-specific DNA-binding proteins find their targets? Nucleic Acids Research, 32(10), pp.3040–3052.

Hamel, B.C. et al., 1996. Familial X-linked mental retardation and isolated growth hormone deficiency: clinical and molecular findings. American journal of medical genetics, 64(1), pp.35–41.

Hammar, P. et al., 2012. The lac repressor displays facilitated diffusion in living cells. Science, 336(6088), pp.1595–1598.

Harley, V. R., Lovellbadge, R. & Goodfellow, P. N. Definition of a Consensus DNA-Binding Site for Sry. Nucleic Acids Res 22, 1500–1501 (1994).

Hatzis, P. & Talianidis, I., 2002. Dynamics of enhancer-promoter communication during differentiation-induced gene activation. Molecular Cell, 10(6), pp.1467–1477.

Heintzman, N.D. et al., 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, 39(3), pp.311–318.

Hippel, von, P.H. & Berg, O.G., 1989. Facilitated target location in biological systems. *Journal of Biological Chemistry*, 264(2), pp.675–678.

Hnisz, D. et al., 2013. Super-Enhancers in the Control of Cell Identity and Disease. *Cell*, 155(4), pp.934–947.

Hoghoughi, N. et al., 2017. Histone variants: essential actors in male genome programming. *Journal of Biochemistry*, 163(2), pp.97–103.

Hosking, B.M. et al., 1995. Trans-activation and DNA-binding properties of the transcription factor, Sox-18. *Nucleic Acids Research*, 23(14), pp.2626–2628.

Hou, L., Srivastava, Y. & Jauch, R., 2017. Molecular basis for the genome engagement by Sox proteins. *Semin Cell Dev Biol*, 63, pp.2–12.

Huang, H.-C. & Klein, P.S., 2004. The Frizzled family: receptors for multiple signal transduction pathways., 5(7), p.234.

Hud, N.V. et al., 1993. Identification of the elemental packing unit of DNA in mammalian sperm cells by atomic force microscopy. *Biochemical and Biophysical Research Communications*, 193(3), pp.1347–1354.

Hughes, J. et al., 2013. Mechanistic insight into the pathology of polyalanine expansion disorders revealed by a mouse model for X linked hypopituitarism. S. Camper, ed. *Plos Genetics*, 9(3), p.e1003290.

Inoue, F. et al., 2017. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Research*, 27(1), pp.38–52.

Isaka, F. et al., 1999. Ectopic expression of the bHLH gene Math1 disturbs neural development. *European Journal of Neuroscience*, 11(7), pp.2582–2588.

Iwafuchi-Doi, M. et al., 2016. The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. *Molecular Cell*, 62(1), pp.79–91.

Jeong, Y. et al., 2006. A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. *Development*, 133(4), pp.761–772.

Jin, C.Y. & Felsenfeld, G., 2007. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes & Development*, 21(12), pp.1519–1529.

Jin, Z. et al., 2009. Different transcription factors regulate nestin gene expression during P19 cell neural differentiation and central nervous system development. *Journal of Biological Chemistry*, 284(12), pp.8160–8173.

Jin, Z.G. et al., 2009. Different Transcription Factors Regulate nestin Gene Expression during P19 Cell Neural Differentiation and Central Nervous System Development. *Journal of Biological Chemistry*, 284(12), pp.8160–8173.

Johanson, T.M. et al., 2018. Genome-wide analysis reveals no evidence of trans chromosomal regulation of mammalian immune development. G. S. Barsh, ed. *Plos Genetics*, 14(6), p.e1007431.

Johnson, D.S. et al., 2007. Genome-Wide Mapping of in Vivo Protein-DNA Interactions. *Science*, 316(5830), pp.1497–1502.

Johnson, K.R. et al., 2018. Deletion of a Long-Range Dlx5 Enhancer Disrupts Inner Ear Development in Mice. *Genetics*, 208(3), pp.1165–1179.

Junion, G. et al., 2012. A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell*, 148(3), pp.473–486.

Kamachi, Y. & Kondoh, H., 2013. Sox proteins: regulators of cell fate specification and differentiation. *Development*, 140(20), pp.4129–4144.

Kamachi, Y., Cheah, K.S.E. & Kondoh, H., 1999. Mechanism of regulatory target selection by the SOX high-mobility-group domain proteins as revealed by comparison of SOX1/2/3 and SOX9. *Molecular and Cellular Biology*, 19(1), pp.107–120.

Kamachi, Y., Uchikawa, M. & Kondoh, H., 2000. Pairing SOX off with partners in the regulation of embryonic development. *Trends in Genetics*, 16(4), pp.182–187.

Kan, L.X. et al., 2004. Sox1 acts through multiple independent pathways to promote neurogenesis. *Developmental Biology*, 269(2), pp.580–594.

Kaucher, A.V., Oatley, M.J. & Oatley, J.M., 2012. NEUROG3 Is a Critical Downstream Effector for STAT3-Regulated Differentiation of Mammalian Stem and Progenitor Spermatogonia1. *Biology of Reproduction*, 86(5).

Kennedy, T.E. et al., 2006. Axon guidance by diffusible chemoattractants: a gradient of netrin protein in the developing spinal cord. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26(34), pp.8866–8874.

Kennedy, T.E. et al., 1994. Netrins are diffusible chemotropic factors for commissural axons in the embryonic spinal cord. *Cell*, 78(3), pp.425–435.

KENNISON, J. & SOUTHWORTH, J., 2002. Transvection in Drosophila. *Advances in Genetics*, 46, pp.399–420.

Khoueiry, P. et al., 2017. Uncoupling evolutionary changes in DNA sequence, transcription factor occupancy and enhancer activity. *eLife*, 6, p.563.

Kiefer, J.C., 2007. Back to basics: Sox genes. *Developmental Dynamics*, 236(8), pp.2356–2366.

Kim, S., Yu, N.-K. & Kaang, B.-K., 2015. CTCF as a multifunctional protein in genome regulation and gene expression. *Experimental & molecular medicine*, 47(6), pp.e166–e166.

Kleinstiver, B.P. et al., 2015. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, 523(7561), pp.481–485.

Kleinstiver, B.P. et al., 2016. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*.

Kline, A.D. et al., 2007. Cornelia de Lange syndrome: clinical review, diagnostic and scoring systems, and anticipatory guidance. *American journal of medical genetics. Part A*, 143A(12), pp.1287–1296.

Klum, S. et al., 2018. Sequentially acting SOX proteins orchestrate astrocyte- and oligodendrocyte-specific gene expression. *EMBO reports*, p.e46635.

Komor, A.C. et al., 2017. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Science advances*, 3(8), p.eaao4774.

Komor, A.C. et al., 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, 533(7603), pp.420–424.

Koopman, P. et al., 1991. Male development of chromosomally female mice transgenic for Sry. *Nature*, 351(6322), pp.117–121.

Kragesteen, B.K. et al., 2018. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. *Nature Genetics*, 50(10), pp.1463–1473.

Kravchenko, E. et al., 2005. Pairing between gypsy insulators facilitates the enhancer action in trans throughout the Drosophila genome. *Molecular and Cellular Biology*, 25(21), pp.9283–9291.

Krijger, P.H.L. & de Laat, W., 2016. Regulation of disease-associated gene expression in the 3D genome. *Nature Reviews Molecular Cell Biology*, 17(12), pp.771–782.

Kulkarni, M.M. & Arnosti, D.N., 2003. Information display by transcriptional enhancers. *Development*, 130(26), pp.6569–6575.

Lania, L., Majello, B. & de Luca, P., 1997. Transcriptional regulation by the Sp family proteins. *The international journal of biochemistry & cell biology*, 29(12), pp.1313–1323.

Laronda, M.M. & Jameson, J.L., 2011. Sox3 functions in a cell-autonomous manner to regulate spermatogonial differentiation in mice. *Endocrinology*, 152(4), pp.1606–1615.

Latchman, D.S., 1999. POU family transcription factors in the nervous system. *Journal of Cellular Physiology*, 179(2), pp.126–133.

Laumonnier, F. et al., 2002. Transcription factor SOX3 is involved in X-linked mental retardation with growth hormone deficiency. *American Journal of Human Genetics*, 71(6), pp.1450–1455.

Le Noir, S. et al., 2017. The IgH locus 3' cis-regulatory super-enhancer co-opts AID for allelic transvection. *Oncotarget*, 8(8), pp.12929–12940.

Lee, C.M., Cradick, T.J. & Bao, G., 2016. The Neisseria meningitidis CRISPR-Cas9 System Enables Specific Genome Editing in Mammalian Cells. *Molecular therapy : the journal of the American Society of Gene Therapy*, 24(3), pp.645–654.

Lek, M. et al., 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616), pp.285–291.

Lendahl, U., Zimmerman, L.B. & McKay, R.D., 1990. CNS stem cells express a new class of intermediate filament protein. *Cell*, 60(4), pp.585–595.

Leoncini, E. et al., 2008. Frequency of holoprosencephaly in the International Clearinghouse Birth Defects Surveillance Systems: searching for population variations. *Birth defects research. Part A, Clinical and molecular teratology*, 82(8), pp.585–591.

Lettice, L.A. et al., 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, 12(14), pp.1725–1735.

Lettice, L.A. et al., 2002. Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proceedings of the National Academy of Sciences*, 99(11), pp.7548–7553.

Liang, H., Hippenmeyer, S. & Ghashghaei, H.T., 2012. A Nestin-cre transgenic mouse is insufficient for recombination in early embryonic neural progenitors. *Biology open*, 1(12), pp.1200–1203.

Lim, B. et al., 2018. Visualization of Transvection in Living Drosophila Embryos. *Molecular Cell*, 70(2), pp.287–296.e6.

Liu, C.-F. & Lefebvre, V., 2015. The transcription factors SOX9 and SOX5/SOX6 cooperate genome-wide through super-enhancers to drive chondrogenesis. *Nucleic Acids Research*, 43(17), pp.8183–8203.

Liu, F., 2017. Enhancer-derived RNA: A Primer. *Genomics, proteomics & bioinformatics*, 15(3), pp.196–200.

Liu, Y., Beyer, A. & Aebersold, R., 2016. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell*, 165(3), pp.535–550.

Lizio, M. et al., 2015. Gateways to the FANTOM5 promoter level mammalian expression atlas., 16(1), p.22.

Machanick, P. & Bailey, T.L., 2011. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*, 27(12), pp.1696–1697.

Mahmoudi, T., Katsani, K.R. & Verrijzer, C.P., 2002. GAGA can mediate enhancer function in trans by linking two separate DNA molecules. *EMBO J*, 21(7), pp.1775–1781.

Malas, S. et al., 2003. Sox1-deficient mice suffer from epilepsy associated with abnormal ventral forebrain development and olfactory cortex hyperexcitability. *Neuroscience*, 119(2), pp.421–432.

Mansour, S.L. et al., 1990. Introduction of a lacZ reporter gene into the mouse int-2 locus by homologous recombination. *Proceedings of the National Academy of Sciences*, 87(19), pp.7688–7692.

Marzluff, W.F. et al., 2002. The Human and Mouse Replication-Dependent Histone Genes. *Genomics*, 80(5), pp.487–498.

Matsunaga, E. & Shiota, K., 1977. Holoprosencephaly in human embryos: Epidemiologic studies of 150 cases. *Teratology*, 16(3), pp.261–272.

McAninch, D. & Thomas, P., 2014a. Identification of highly conserved putative developmental enhancers bound by SOX3 in neural progenitors using ChIP-Seq. D. Zheng, ed. *Plos One*, 9(11), p.e113361.

McAninch, D. & Thomas, P., 2014b. Identification of highly conserved putative developmental enhancers bound by SOX3 in neural progenitors using ChIP-seq. *PloS ONE*.

McLean, C.Y. et al., 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology*, 28(5), pp.495–501.

Mellert, D.J. & Truman, J.W., 2012. Transvection is common throughout the Drosophila genome. *Genetics*, 191(4), pp.1129–1141.

Merika, M. & Thanos, D., 2001. Enhanceosomes. *Current Opinion in Genetics & Development*, 11(2), pp.205–208.

Mertin, S., McDowall, S.G. & Harley, V.R., 1999. The DNA-binding specificity of SOX9 and other SOX proteins. *Nucleic Acids Research*, 27(5), pp.1359–1364.

Meyer, M.B. et al., 2019. Targeted genomic deletions identify diverse enhancer functions and generate a kidney-specific, endocrine-deficient Cyp27b1 pseudo-null mouse. *The Journal of biological chemistry*, 294(24), pp.9518–9535.

Micol, J.L., Castelli-Gair, J.E. & García-Bellido, A., 1990. Genetic analysis of transvection effects involving cis-regulatory elements of the Drosophila Ultrabithorax gene. *Genetics*, 126(2), pp.365–373.

Miyagi, S. et al., 2008. Consequence of the loss of Sox2 in the developing brain of the mouse. *FEBS Letters*, 582(18), pp.2811–2815.

Miyagi, S., Kato, H. & Okuda, A., 2009. Role of SoxB1 transcription factors in development. *Cellular and Molecular Life Sciences*, 66(23), pp.3675–3684.

Mohseni, P. et al., 2011. Nestin is not essential for development of the CNS but required for dispersion of acetylcholine receptor clusters at the area of neuromuscular junctions. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(32), pp.11547–11552.

Mokrý, J. et al., 2008. Expression of intermediate filament nestin in blood vessels of neural and non-neural tissues. *Acta medica (Hradec Kralove)*, 51(3), pp.173–179.

MONTAGUTELLI, X., 2000. Effect of the Genetic Background on the Phenotype of Mouse Mutations. *J Am Soc Nephrol*, 11(suppl 2), pp.S101–S105.

Montellier, E. et al., 2013. Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B. *Genes Dev*, 27(15), pp.1680–1692.

Mora, A. et al., 2016. In the loop: promoter-enhancer interactions and bioinformatics. *Briefings in bioinformatics*, 17(6), pp.980–995.

Mudrak, O., Tomilin, N. & Zalensky, A., 2005. Chromosome architecture in the decondensing human sperm nucleus. *Journal of Cell Science*, 118(Pt 19), pp.4541–4550.

Müller, M. et al., 2016. Streptococcus thermophilus CRISPR-Cas9 Systems Enable Specific Editing of the Human Genome. *Molecular therapy : the journal of the American Society of Gene Therapy*, 24(3), pp.636–644.

Nagai, K., 2001. Molecular evolution of Sry and Sox gene. *Gene*, 270(1-2), pp.161–169.

Nat, B.L.A.1954, *The theory and applications of a new method of detecting chromosomal rearrangement in Drosophila melanogaster*,

Ng, C.K.L. et al., 2012. Deciphering the Sox-Oct partner code by quantitative cooperativity measurements. *Nucleic Acids Research*, 40(11), pp.4933–4941.

Nishiguchi, S. et al., 1998. Sox1 directly regulates the gamma-crystallin genes and is essential for lens development in mice. *Genes & Development*, 12(6), pp.776–781.

Nishimoto, M. et al., 1999. The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Molecular and Cellular Biology*, 19(8), pp.5453–5465.

Ohba, S. et al., 2015. Distinct Transcriptional Programs Underlie Sox9 Regulation of the Mammalian Chondrocyte. *Cell Reports*, 12(2), pp.229–243.

Ohno, S., 1972. Simplicity of mammalian regulatory systems. *Developmental Biology*, 27(1), pp.131–136.

Okuda, Y. et al., 2010. B1 SOX coordinate cell specification with patterning and morphogenesis in the early zebrafish embryo. V. van Heyningen, ed. *Plos Genetics*, 6(5), p.e1000936.

Olley, G. et al., 2018. BRD4 interacts with NIPBL and BRD4 is mutated in a Cornelia de Lange-like syndrome. *Nature Genetics*, 50(3), pp.329–332.

Onishi, K. & Zou, Y., 2017. Sonic Hedgehog switches on Wnt/planar cell polarity signaling in commissural axon growth cones by reducing levels of Shisa2. *eLife*, 6, p.708.

Orsini, A. et al., 2018. Generalized epilepsy and mild intellectual disability associated with 13q34 deletion: A potential role for SOX1 and ARHGEF7. *Seizure*, 59, pp.38–40.

Osterwalder, M. et al., 2018. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature*, 554(7691), pp.239–243.

Panigrahi, S.K., Vasileva, A. & Wolgemuth, D.J., 2012. Sp1 transcription factor and GATA1 cis-acting elements modulate testis-specific expression of mouse cyclin A1. M. Miozzo, ed. *Plos One*, 7(10), p.e47862.

Panne, D., Maniatis, T. & Harrison, S.C., 2007. An atomic model of the interferon-beta enhanceosome. *Cell*, 129(6), pp.1111–1123.

Papatheodorou, I. et al., 2018. Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Research*, 46(D1), pp.D246–D251.

Park, D. et al., 2010. Nestin is required for the proper self-renewal of neural stem cells. *Stem Cells*, 28(12), pp.2162–2171.

Peeters, M.C. et al., 1998. Role of differential cell proliferation in the tail bud in aberrant mouse neurulation. *Developmental dynamics : an official publication of the American Association of Anatomists*, 211(4), pp.382–389.

Pelizzoli, R. et al., 2008. TTF-1/NKX2.1 up-regulates the in vivo transcription of nestin. *International Journal of Developmental Biology*, 52(1), pp.55–62.

Pelling, M. et al., 2011. Differential requirements for neurogenin 3 in the development of POMC and NPY neurons in the hypothalamus. *Dev Biol*, 349(2), pp.406–416.

Pennacchio, L.A. et al., 2013. Enhancers: five essential questions. *Nature Reviews Genetics*, 14(4), pp.288–295.

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

Petersen, P.H. et al., 2002. Progenitor cell maintenance requires numb and numblike during mouse neurogenesis. *Nature*, 419(6910), pp.929–934.

Phillips, B.T., Gassei, K. & Orwig, K.E., 2010. Spermatogonial stem cell regulation and spermatogenesis. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1546), pp.1663–1678.

Pinaud, E. et al., 2011. The IgH locus 3' regulatory region: pulling the strings from behind. *Advances in immunology*, 110, pp.27–70.

Pott, S. & Lieb, J.D., 2015. What are super-enhancers? *Nature Genetics*, 47(1), pp.8–12.

Pradeepa, M.M., 2017. Causal role of histone acetylations in enhancer function. *Transcription*, 8(1), pp.40–47.

Pradeepa, M.M. et al., 2016. Histone H3 globular domain acetylation identifies a new class of enhancers. *Nature Genetics*, 48(6), pp.681–686.

Ran, F.A. et al., 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 154(6), pp.1380–1389.

Rathke, C. et al., 2014. Chromatin dynamics during spermiogenesis. *Biochimica et biophysica acta*, 1839(3), pp.155–168.

Raverot, G. et al., 2005. Sox3 expression in undifferentiated spermatogonia is required for the progression of spermatogenesis. *Developmental Biology*, 283(1), pp.215–225.

Reinstein, E. et al., 2016. Terminal microdeletions of 13q34 chromosome region in patients with intellectual disability: Delineation of an emerging new microdeletion syndrome. *Molecular genetics and metabolism*, 118(1), pp.60–63.

Rizzoti, K. et al., 2004. SOX3 is required during the formation of the hypothalamo-pituitary axis. *Nature Genetics*, 36(3), pp.247–255.

Roessler, E. et al., 1996. Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. *Nature Genetics*, 14(3), pp.357–360.

Rogers, N., McAninch, D. & Thomas, P., 2014. Dbx1 Is a Direct Target of SOX3 in the Spinal Cord. *Plos One*, 9(4).

Romanowska, M. et al., 2009. Wnt5a exhibits layer-specific expression in adult skin, is upregulated in psoriasis, and synergizes with type 1 interferon. E. Didier, ed. *Plos One*, 4(4), p.e5354.

Sala, C.F. et al., 2000. Identification, gene structure, and expression of human frizzled-3 (FZD3). *Biochemical and Biophysical Research Communications*, 273(1), pp.27–34.

Sandberg, M., Kallstrom, M. & Muhr, J., 2005. Sox21 promotes the progression of vertebrate neurogenesis. *Nature Neuroscience*, 8(8), pp.995–1001.

Sanyal, A. et al., 2012. The long-range interaction landscape of gene promoters. *Nature*, 489(7414), pp.109–113.

Sarkar, A. & Hochedlinger, K., 2013. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell Stem Cell*, 12(1), pp.15–30.

Sekido, R. & Lovell-Badge, R., 2008a. Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature*, 453(7197), pp.930–934.

Sekido, R. & Lovell-Badge, R., 2008b. Sex determination involves synergistic action of SRY anSF1 on a specific Sox9 enhancer. *Nature*, 453(7197), pp.930–934.

Shinagawa, T. et al., 2015. Disruption of Th2a and Th2b genes causes defects in spermatogenesis. *Development*, 142(7), pp.1287–1292.

Shlyueva, D., Stampfel, G. & Stark, A., 2014. Transcriptional enhancers: from properties to genome-wide predictions. *Nature Reviews Genetics*, 15(4), pp.272–286.

Siepel, A. et al., 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, 15(8), pp.1034–1050.

Sinclair, A.H. et al., 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, 346(6281), pp.240–244.

Sipos, L. & Gyurkovics, H., 2005. Long-distance interactions between enhancers and promoters. *The FEBS Journal*, 272(13), pp.3253–3259.

Smith, E. & Shilatifard, A., 2014. Enhancer biology and enhanceropathies. *Nature Structural and Molecular Biology*, 21(3), pp.210–219.

Spielmann, M. & Mundlos, S., 2013. Structural variations, the regulatory landscape of the genome and their alteration in human disease. *Bioessays*, 35(6), pp.533–543.

Spitz, F. & Furlong, E.E.M., 2012. Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics*, 13(9), pp.613–626.

Sternberg, S.H. et al., 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507(7490), pp.62–67.

Stoeckli, E.T., 2018. Understanding axon guidance: are we nearly there yet? *Development (Cambridge, England)*, 145(10), p.dev151415.

Struhl, K., 2001. Gene regulation. A paradigm for precision. *Science*, 293(5532), pp.1054–1055.

Stuebner, S. et al., 2009. Fzd3and Fzd6deficiency results in a severe midbrain morphogenesis defect. *Developmental Dynamics*, 239(1), pp.NA–NA.

Sutton, E. et al., 2011. Identification of SOX3 as an XX male sex reversal gene in mice and humans. *Journal of Clinical Investigation*, 121(1), pp.328–341.

Suzuki, S. et al., 2010. The neural stem/progenitor cell marker nestin is expressed in proliferative endothelial cells, but not in mature vasculature. *Journal of Histochemistry & Cytochemistry*, 58(8), pp.721–730.

Szafranski, P. et al., 2017. CRISPR/Cas9-mediated deletion of lncRNA Gm26878 in the distant Foxf1 enhancer region. *Mammalian genome : official journal of the International Mammalian Genome Society*, 28(7-8), pp.275–282.

Štros, M., Launholt, D. & Grasser, K.D., 2007. The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cellular and Molecular Life Sciences*, 64(19-20), pp.2590–2606.

Tan, N.Y. & Khachigian, L.M., 2009. Sp1 phosphorylation and its regulation of gene transcription. *Molecular and Cellular Biology*, 29(10), pp.2483–2488.

Tanaka, S. et al., 2004. Interplay of SOX and POU factors in regulation of the Nestin gene in neural primordial cells. *Molecular and Cellular Biology*, 24(20), pp.8834–8846.

Taranova, O.V. et al., 2006. SOX2 is a dose-dependent regulator of retinal neural progenitor competence. *Genes & Development*, 20(9), pp.1187–1202.

Tegelenbosch, R.A. & de Rooij, D.G., 1993. A quantitative study of spermatogonial multiplication and stem cell renewal in the C3H/101 F1 hybrid mouse. *Mutation research*, 290(2), pp.193–200.

Thakore, P.I. et al., 2018. RNA-guided transcriptional silencing in vivo with S. aureus CRISPR-Cas9 repressors. *Nature Communications*, 9(1), p.1674.

Thomas-Chollier, M. et al., 2012. RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. *Nucleic Acids Research*, 40(4), pp.e31–e31.

Tolhuis, B. et al., 2002. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Molecular Cell*, 10(6), pp.1453–1465.

Tomioka, M. et al., 2002. Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. *Nucleic Acids Research*, 30(14), pp.3202–3213.

Trumpp, A. et al., 1999. Cre-mediated gene inactivation demonstrates that FGF8 is required for cell survival and patterning of the first branchial arch. *Genes & Development*, 13(23), pp.3136–3148.

Tuan, D., Kong, S. & Hu, K., 1992. Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proceedings of the National Academy of Sciences*, 89(23), pp.11219–11223.

Uchikawa, M. et al., 2011. B1 and B2 Sox gene expression during neural plate development in chicken and mouse embryos: Universal versus species-dependent features. *Development Growth & Differentiation*, 53(6), pp.761–771.

Ueda, J. et al., 2017. Testis-Specific Histone Variant H3t Gene Is Essential for Entry into Spermatogenesis. *Cell Reports*, 18(3), pp.593–600.

Uhl, J.D., Zandvakili, A. & Gebelein, B., 2016. A Hox Transcription Factor Collective Binds a Highly Conserved Distal-less cis-Regulatory Module to Generate Robust Transcriptional Outcomes. A. Stathopoulos, ed. *Plos Genetics*, 12(4), p.e1005981.

Valli, H. et al., 2014. Fluorescence- and magnetic-activated cell sorting strategies to isolate and enrich human spermatogonial stem cells. *Fertility and sterility*, 102(2), pp.566–580.e7.

van Belkum, A. et al., 1998. Short-sequence DNA repeats in prokaryotic genomes. *Microbiology and Molecular Biology Reviews*, 62(2), pp.275—.

van Berkum, N.L. et al., 2010. Hi-C: A Method to Study the Three-dimensional Architecture of Genomes. *Journal of Visualized Experiments*, (39).

van de Wetering, M. & Clevers, H., 1992. Sequence-specific interaction of the HMG box proteins TCF-1 and SRY occurs within the minor groove of a Watson-Crick double helix. *EMBO J*, 11(8).

van Houte, L.P. et al., 1995. Solution structure of the sequence-specific HMG box of the lymphocyte transcriptional activator Sox-4. *Journal of Biological Chemistry*, 270(51), pp.30516–30524.

Vernimmen, D. et al., 2007. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J*, 26(8), pp.2041–2051.

Visel, A. et al., 2007. VISTA Enhancer Browser - a database of tissue-specific human enhancers. *Nucleic Acids Research*, 35, pp.D88–D92.

Visel, A., Thaller, C. & Eichele, G., 2004. GenePaint.org: an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Research*, 32(Database issue), pp.D552–6.

Wang, W. et al., 2016. Delivery of Cas9 Protein into Mouse Zygotes through a Series of Electroporation Dramatically Increases the Efficiency of Model Creation. *Journal of Genetics and Genomics*, 43(5), pp.319–327.

Wang, Y. et al., 2006. Axonal growth and guidance defects in Frizzled3 knock-out mice: a comparison of diffusion tensor magnetic resonance imaging, neurofilament staining, and genetically directed cell labeling. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26(2), pp.355–364.

Wang, Y. et al., 2002. Frizzled-3 is required for the development of major fiber tracts in the rostral CNS. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 22(19), pp.8563–8573.

Wang, Z. et al., 2017. HEDD: Human Enhancer Disease Database. *Nucleic Acids Research*, 46(D1), pp.D113–D120.

Waters, P.D., Wallis, M.C. & Marshall Graves, J.A., 2007. Mammalian sex--Origin and evolution of the Y chromosome and SRY. *Seminars in Cell & Developmental Biology*, 18(3), pp.389–400.

Weber, J.R. & Skene, J.H., 1997. Identification of a novel repressive element that contributes to neuron-specific gene expression. *Journal of Neuroscience*, 17(20), pp.7583–7593.

Wegner, M., 1999. From head to toes: the multiple facets of Sox proteins. *Nucleic Acids Research*, 27(6), pp.1409–1420.

Weiss, J. et al., 2003. Sox3 is required for gonadal function, but not sex determination, in males and females. *Molecular and Cellular Biology*, 23(22), pp.8084–8091.

Weiss, K.M. et al., 2003. Sox3 Is Required for Gonadal Function, but Not Sex Determination, in Males and Females. *Molecular and Cellular Biology*, 23(22).

Whyte, W.A. et al., 2013. Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell*, 153(2), pp.307–319.

Wiese, C. et al., 2004. Nestin expression--a property of multi-lineage progenitor cells? *Cellular and Molecular Life Sciences*, 61(19-20), pp.2510–2522.

Wilson, M. & Koopman, P., 2002. Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. *Current Opinion in Genetics & Development*, 12(4), pp.441–446.

Wissmüller, S. et al., 2006. The high-mobility-group domain of Sox proteins interacts with DNA-binding domains of many transcription factors. *Nucleic Acids Research*, 34(6), pp.1735–1744.

Wood, H.B. & Episkopou, V., 1999. Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages. *Mechanisms of Development*, 86(1-2), pp.197–201.

Yesudhas, D. et al., 2017. Proteins Recognizing DNA: Structural Uniqueness and Versatility of DNA-Binding Domains in Stem Cell Transcription Factors. *Genes*, 8(8), p.192.

Yoshida, S. et al., 2004. Neurogenin3 delineates the earliest stages of spermatogenesis in the mouse testis. *Developmental Biology*, 269(2), pp.447–458.

Yoshida, S. et al., 2006. The first round of mouse spermatogenesis is a distinctive program that lacks the self-renewing spermatogonia stage. *Development*, 133(8), pp.1495–1505.

Zaret, K.S. & Carroll, J.S., 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes & Development*, 25(21), pp.2227–2241.

Zhang, Y. et al., 2013. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature*, 504(7479), pp.306–310.

Zhao, Z. et al., 2006. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genetics*, 38(11), pp.1341–1347.

Zhou, Y.B. et al., 1998. Position and orientation of the globular domain of linker histone H5 on the nucleosome. *Nature*, 395(6700), pp.402–405.

Zhu, Z. et al., 2016. Dynamics of the Transcriptome during Human Spermatogenesis: Predicting the Potential Key Genes Regulating Male Gametes Generation. *Scientific Reports*, 6(1), p.19069.

Ziebarth, J.D., Bhattacharya, A. & Cui, Y., 2013. CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. *Nucleic Acids Research*, 41(Database issue), pp.D188–94.

Zimmerman, L. et al., 1994. Independent Regulatory Elements in the Nestin Gene Direct Transgene Expression to Neural Stem-Cells or Muscle Precursors (Vol 12, Pg 11, 1994). *Neuron*, 12(6), pp.1389–1389.

Zuo, E. et al., 2019. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science*, 364(6437), pp.289–292.