The role of Olig2 regulation in ventral neural tube patterning

Katherine Exelby

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Division of Biosciences
University College London

February 23, 2020
I, Katherine Exelby, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.
Abstract

In the developing spinal cord, a ventral to dorsal gradient of Sonic Hedgehog (Shh) protein directs spatial organisation of neural progenitor domains that subsequently differentiate to molecularly distinct neurons. Graded Shh signalling is interpreted by a gene regulatory network (GRN) that restricts expression of different transcription factors to progenitor domains delimited by sharp boundaries. This GRN, comprising of four key transcription factors, has been modelled in a deterministic mathematical model. To explore how this GRN ensures robustness in patterning, the interactions between nodes of the network required further investigation.

In this study, we investigate interactions at a particular node, Olig2, expressed in the ventral neural tube in the motor neuron progenitor (pMN) domain. We focused on the -33 kb enhancer of Olig2 that has binding sites for all major inputs in the GRN. CRISPR/Cas9-mediated excision of this enhancer in vitro severely disrupted Olig2 expression in mouse ES cells directly differentiated towards spinal cord progenitors. Remaining Olig2 expression highlighted a second regulatory region, +75 kb, that became accessible at later stages. Deletion of both -33 and +75 kb enhancers led to a complete loss of Olig2 expression in vitro.

Excision of the -33 kb enhancer in vivo resulted in embryos having reduced Olig2 expression in a smaller pMN domain which was more severe when combined with deletion of the +75 kb enhancer. In addition to reduced Olig2 expression, a loss of precision was observed at the p3/pMN boundary in the developing neural tube. In collaboration with Edgar Herrera Delgado, we modelled this loss of precision in silico and have been able to uncover a mechanism encoded within GRN structure that drives precision in patterning even in the presence of noisy gene expression. Through analyses of Olig2 enhancers both in vitro and in vivo, we demonstrate how enhancers contribute to robustness in gene regulatory networks driving patterning.
Impact statement

Accurately reproducing robust spatial organisation of gene expression domains during development continues to be a key focus in developmental biology research. Using the ventral neural tube as a model, this study provides insight into the underlying mechanisms encoded within gene regulatory networks that dictate sharp boundaries between different expression domains. This “precision by design” highlights the capacity of transcriptional circuits to contribute to robust tissue patterning and identifies a mechanism that might be exploited in other biological settings requiring precise responses from groups of cells.

Understanding the mechanisms by which multiple inputs at an enhancer dictate gene expression at different times and locations throughout development is of broad interest to developmental biologists. Through analyses of different anterior-posterior regulation of a gene key to the development of motor neurons and oligodendrocytes, we provide insight into the differences in epigenetic control of Olig2 under different conditions.

This work also contributes towards the advance of regenerative medicine and tissue engineering. The invention of techniques to generate induced pluripotent stem (iPS) cells has allowed the potential for cell replacement therapies. This requires the directed differentiation of pluripotent cells. A detailed understanding of the mechanisms underlying developmental decision making will enable the design of differentiation protocols that generate functional cell types with high purity. Additionally, generating these differentiated cell types from patient-derived iPS cells allows the investigation of the cellular biology underlying their disease.
Acknowledgements

I would like to thank James, my supervisor, for giving me the opportunity to carry out this project. Over the past four years he has supported and encouraged me to develop my ideas and skills. As important as a good supervisor, is good colleagues, and I could not have asked for a better than the Briscoe Lab. I would like to thank all of the Briscoe Lab for being such great colleagues. Additionally, this work would not have been possible without contributions from other lab members. The ATAC-sequencing data presented in this work was carried out by Vicki Metzis. The cell line which formed the basis for all subsequent mutations was developed by Andreas Sagner and the mathematical modelling of boundary sharpness was developed by Edgar Herrera Delgado.

As well as assistance from members of the lab, I have been lucky to have access to excellent facilities at the Francis Crick Institute. In particular, the BRF, Light Microscopy and Flow Cytometry. I would like to especially mention Ashley Vanderplank, Animal House Research Officer, who has been a brilliant to work with and I also want to take this opportunity to acknowledge all the mice used throughout this project.

I would like to thank Vicki Metzis for all of her support throughout my PhD. Edgar Herrera Delgado for his everlasting patience whilst explaining dynamical systems to me. Andreas Sagner and Tom Watson for their assistance with flow cytometry analysis. Despina Stamanaki and Manuela Melchionda for always keeping the lab running smoothly and finally Teresa Rayon, Joaquina Delas Vives and Marta Moris for their moral support.

I would like to thank my Mum and Dad for all of their support, this PhD would not have been possible without them. I would also like to thank the rest
of my family and all of my friends. Throughout my PhD I have had wonderful housemates (Hannah, Alex, Sophie, Kess, Liv and last but not least, my cat, Elsa) who have always been there for me after a hard day at work and I will never forget the wonderful times at our house in Holloway. Finally, I would like to thank Eliot for being the most supportive, loving partner.
1 Introduction

1.1 The spinal cord

1.1.1 Structure of the spinal cord

1.1.2 Neurogenesis

1.1.3 Gliogenesis

1.2 Patterning of the neural tube

1.2.1 A gradient of Sonic Hedgehog patterns the ventral neural tube

1.2.2 Sonic Hedgehog gradient is relayed by a family of Gli proteins

1.2.3 Time of Sonic Hedgehog exposure directs ventral cell fates

1.2.4 Adapting dynamics of Shh signalling

1.2.5 Cross-repressive transcription factors specify progenitor domain boundaries

1.3 Gene Regulatory Networks in Patterning

1.3.1 Cis-regulatory elements

1.3.2 Construction of GRNs

1.3.3 Interpretation of morphogen gradients

1.3.4 Cross-repressive transcription factors

1.3.5 Ubiquitously expressed transcription factors

1.3.6 Mathematical model for ventral neural tube patterning

1.4 In vitro generation of neural progenitors

1.5 Project aims
## 2 Materials and Methods

### 2.1 CRISPR/Cas9 targeting
2.1.1 Generating CRISPR/Cas9 plasmids

### 2.2 Tissue Culture Techniques
- 2.2.1 MEFs and Feeders
- 2.2.2 Culture of ES cells
- 2.2.3 Differentiation of ES cells to spinal cord and hindbrain progenitors
- 2.2.4 Targeting CRISPR/Cas9 to mouse ES cells

### 2.3 Analysis of cells
- 2.3.1 Genotyping
- 2.3.2 Immunofluorescence
- 2.3.3 Flow Cytometry
- 2.3.4 Quantitative PCR (qPCR)

### 2.4 Mouse Procedures

### 2.5 Embryo processing
- 2.5.1 Genotyping embryos
- 2.5.2 Sectioning embryos
- 2.5.3 Immunohistochemistry on embryo sections
- 2.5.4 Immunohistochemistry on whole embryos
- 2.5.5 Imaging embryos and sections

## 3 Regulation of *Olig2* in vitro

### 3.1 Introduction
3.1.1 *Olig2* regulation in the ventral spinal cord
3.1.2 Bioinformatic analyses of -33 kb enhancer of *Olig2*

### 3.2 Generation of -33 kb enhancer knockout mouse ES cell lines
- 3.2.1 Targeting approach to knockout -33 kb enhancer
- 3.2.2 Screening clones for knockout deletions
- 3.2.3 Analysis of heterozygous and homozygous enhancer knockout cell lines
- 3.2.4 *Olig2* expression at day 5
3.2.5 Expression of Olig1 .......................... 63
3.2.6 Efficiency of cell differentiation .................. 64
3.2.7 Alternative cell fates ............................. 65

3.3 Motor neuron differentiation in O2e33 cells .............. 66

3.4 Bioinformatic analysis of the -33 kb enhancer ............ 68
3.4.1 Targeting of the Sox2 binding site .................... 68
3.4.2 Sox2 binding site is required for correct onset of Olig2
expression in vitro .................................. 69

3.5 Alternative regulation of Olig2 in vitro ................. 70
3.5.1 Analysis of Olig2 locus for alternative enhancers ....... 70
3.5.2 Alternative Olig2 enhancer at +75 kb .................... 71
3.5.3 Accessibility of +75 kb enhancer across differentiation
protocol .............................................. 73
3.5.4 Differential usage of Olig2 enhancers along the A-P axis 74
3.5.5 Generation of O2e75 and O2e33-75 cells ................. 76
3.5.6 Analysis of O2e75 and O2e33-75 cells ...................... 76

4 -33 kb enhancer deletion in vivo .......................... 79
4.1 O2e33 Mouse line generation .......................... 79
4.2 O2e33 patterning phenotype ............................ 80
4.3 Onset of Olig2 in O2e33 embryos ......................... 82
4.4 Motor neuron differentiation in O2e33 embryos .......... 84
4.5 Oligodendrocytes in O2e33 embryos ....................... 88
4.6 Anterior-Posterior Olig2 patterning ....................... 90
4.7 Viability of O2e33 mice ................................ 92

5 Boundary sharpness in O2e33 and Pax6+/− embryos ........ 93
5.1 Measuring sharpness .................................. 93
5.2 Boundary sharpness in Pax6+/− embryos ................... 97
5.3 Modelling sharpness .................................. 98

6 Regulation of Olig2 by -33 kb and +75 kb enhancer in vivo 103
6.1 O2e33-75 mouse line generation ........................ 103
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2 O2e33-75 phenotype</td>
<td>104</td>
</tr>
<tr>
<td>6.2.1 Motor neuron generation in O2e33-75 embryos</td>
<td>108</td>
</tr>
<tr>
<td>6.2.2 Oligodendrocyte generation in O2e33-75 embryos</td>
<td>111</td>
</tr>
<tr>
<td>6.3 O2e75 phenotype</td>
<td>112</td>
</tr>
<tr>
<td>7 Discussion</td>
<td>114</td>
</tr>
<tr>
<td>7.1 Gene regulatory networks</td>
<td>115</td>
</tr>
<tr>
<td>7.1.1 Enhancer logic in gene regulatory networks</td>
<td>117</td>
</tr>
<tr>
<td>7.1.2 Multiple enhancers combine to drive gene expression</td>
<td>119</td>
</tr>
<tr>
<td>7.1.3 Enhancers ensure robustness in patterning</td>
<td>121</td>
</tr>
<tr>
<td>7.1.4 Structure of GRNs encodes precision in patterning in the presence of noise</td>
<td>122</td>
</tr>
<tr>
<td>7.2 Robust Olig2 expression is required during development</td>
<td>125</td>
</tr>
<tr>
<td>7.2.1 Expression of Olig2</td>
<td>125</td>
</tr>
<tr>
<td>7.2.2 Differentiation of motor neurons is affected by Olig2 expression levels</td>
<td>127</td>
</tr>
<tr>
<td>7.2.3 Differentiation of oligodendrocytes is affected in O2e33 and O2e33-75 embryos</td>
<td>128</td>
</tr>
<tr>
<td>7.2.4 Viability of O2e33 mice</td>
<td>129</td>
</tr>
<tr>
<td>7.3 Concluding Remarks</td>
<td>130</td>
</tr>
</tbody>
</table>

### Appendices

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A JASPAR analysis of -33 kb enhancer</td>
<td>131</td>
</tr>
<tr>
<td>B -33 kb enhancer conservation in humans</td>
<td>133</td>
</tr>
<tr>
<td>C Olig2 -33 kb locus</td>
<td>134</td>
</tr>
<tr>
<td>D Olig2 +75 kb locus</td>
<td>137</td>
</tr>
<tr>
<td>E FiJi macro to obtain immunofluorescence intensity data from selected nuclei</td>
<td>139</td>
</tr>
</tbody>
</table>

Bibliography | 140 |
List of Figures

1.1 Dorsal-ventral axis of the spinal cord . . . . . . . . . . . . 17
1.2 Patterning of the neural tube . . . . . . . . . . . . . . . . . 21
1.3 Neural tube patterning occurs between E8.5 to E10.5 . . . . 25
1.4 Cross-repressive transcription factor interactions specify progenitor domain boundaries . . . . . . . . . . . . . . . . . 26
1.5 ATAC-seq to identify open regions of chromatin . . . . . . . 30
1.6 Multiple inputs direct transcription factor network output . . 35
1.7 Mathematical model for ventral neural tube patterning by Shh . 38
1.8 In vitro differentiation of mouse ES cells to neural progenitors . 42
3.1 The -33 kb enhancer of Olig2 is accessible during in vitro differentiation . . . . . . . . . . . . . . . . . . . . . . . . . . . . 60
3.2 Locations of CRISPR guides to remove -33 kb enhancer . . . 61
3.3 Removal of -33 kb enhancer affects Olig2 expression in vitro . 63
3.4 Olig2 expression is delayed in O2e33 cells . . . . . . . . . . . 64
3.5 Olig1 expression in vitro . . . . . . . . . . . . . . . . . . . . 65
3.6 Neural progenitor mRNA and immunofluorescence analysis of O2e33 cells . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66
3.7 Motor neuron differentiation in O2e33 cells . . . . . . . . . . 68
3.8 CRISPR deletion of Sox2 site in -33 kb enhancer . . . . . . . 69
3.9 ATAC-seq data for Olig2 locus in showing predicted enhancer at +75 kb . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 72
3.10 ATAC-seq data for Olig2 locus in hindbrain and spinal cord differentiation conditions . . . . . . . . . . . . . . . . . . . . . 74
3.11 Quantative PCR of hindbrain and spinal cord markers . . . . 75
3.12 Flow cytometry comparing O2e33 and DOK1-1 in hindbrain and spinal cord conditions ........................................ 75

3.13 Flow cytometry analysis of mKate2 fluorescence levels comparing DOK1-1 (black), O2e33 (red), O2e75 (blue) and O2e33-75 (orange) differentiated under spinal cord conditions. Dashed line at highlights extend of fluorescence intensity in DOK1-1 cells compared to other lines. ........................................ 77

3.14 Flow cytometry analysis comparing mKate2 fluorescence in DOK1-1, O2e33, O2e75 and O2e33-75 in spinal cord and hindbrain conditions from days 5-7 of differentiation. ................. 78

4.1 Removal of -33 kb enhancer disrupts Olig2 expression in the pMN domain ......................................................... 80

4.2 O2e33 embryos have reduced Olig2 expression .................. 81

4.3 Olig2 expression occurs later in O2e33 mutants .................. 83

4.4 Onset of Olig2 expression is delayed in O2e33 mutants .......... 83

4.5 Olig2 expression does not recover in vivo ......................... 84

4.6 Reduced number of motor neurons in O2e33 embryos at E9.5 . 86

4.7 Reduced number of motor neurons in O2e33 embryos at E13.5 . 87

4.8 Olig2 expression at E14.5 ............................................. 89

4.9 Olig2 expression at E16.5 ............................................. 90

4.10 Hindbrain expression of Olig2 in wild type and O2e33 embryos . 91

5.1 Method to quantify single nuclei in neural tube sections ....... 94

5.2 Measuring sharpness of domain boundaries ...................... 95

5.3 Boundary sharpness in Pax6^-/- embryos .......................... 97

5.4 Boundary sharpness and position Pax6^-/- embryos ............... 98

5.5 Modelling boundary sharpness ...................................... 100

6.1 Reduction of Olig2 in O2e33-75 embryos ........................ 104

6.2 Loss of Olig2 in O2e33-75 embryos at hindbrain levels .......... 105

6.3 Reduced expression of Olig2 in O2e33-75 embryos ............... 106

6.4 Loss of Olig2 in O2e33-75 embryos after E9.5 .................... 107
List of Figures

6.5 O2e33-75 embryos have increased V2 interneuron differentiation 108
6.6 O2e33-75 embryos have reduced number of motor neurons . . . 109
6.7 O2e33-75 embryos have reduced number of motor neurons . . . 110
6.8 O2e33-75 embryos display reduced numbers of oligodendrocytes 111
6.9 O2e75 embryos do not display a neural tube patterning phenotype113

A.1 Results from JASPAR TF-PWM database for -33kb enhancer . 132
B.1 -33kb enhancer conservation in humans . . . . . . . . . . . . . . 133
C.1 Olig2 -33kb locus . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 136
D.1 Olig2 +75kb locus . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 138
E.1 FiJi macro to collect data from selected nuclei . . . . . . . . . 139
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATAC</td>
<td>Assay for Transposase-Accessible Chromatin</td>
</tr>
<tr>
<td>BMP</td>
<td>Bone Morphogenic Protein</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin immunoprecipitation</td>
</tr>
<tr>
<td>CLE</td>
<td>Caudal Lateral Epiblast</td>
</tr>
<tr>
<td>CNS</td>
<td>Central Nervous System</td>
</tr>
<tr>
<td>CRISPR</td>
<td>Clustered Regulatory Interspaced Short Palindromic Repeats</td>
</tr>
<tr>
<td>E</td>
<td>Embryonic day (or days post coitum)</td>
</tr>
<tr>
<td>FGF</td>
<td>Fibroblast Growth Factor</td>
</tr>
<tr>
<td>GRN</td>
<td>Gene Regulatory Network</td>
</tr>
<tr>
<td>Irx</td>
<td>Iroquois homeobox protein</td>
</tr>
<tr>
<td>Islet</td>
<td>Insulin Gene Enhancer Protein</td>
</tr>
<tr>
<td>Lhx</td>
<td>LIM Homeobox</td>
</tr>
<tr>
<td>MN</td>
<td>Motor neuron</td>
</tr>
<tr>
<td>Nkx</td>
<td>NKK-Homeodomain factor</td>
</tr>
<tr>
<td>NMP</td>
<td>Neuromesodermal Progenitor</td>
</tr>
<tr>
<td>ODE</td>
<td>Ordinary Differential Equation</td>
</tr>
<tr>
<td>Olig</td>
<td>Oligodendrocyte Transcription Factor</td>
</tr>
<tr>
<td>OPC</td>
<td>Oligodendrocyte Progenitor Cell</td>
</tr>
<tr>
<td>p</td>
<td>progenitor</td>
</tr>
<tr>
<td>Pax</td>
<td>Paired Box</td>
</tr>
<tr>
<td>PFA</td>
<td>Paraformaldehyde</td>
</tr>
<tr>
<td>RA</td>
<td>Retinoic acid</td>
</tr>
<tr>
<td>Shh</td>
<td>Sonic Hedgehog</td>
</tr>
<tr>
<td>Sox</td>
<td>SRY-box</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>Tubb3</td>
<td>Tubulin Beta 3 Class III</td>
</tr>
<tr>
<td>Wnt</td>
<td>Wnt Family Member</td>
</tr>
</tbody>
</table>
List of Tables

2.1 CRISPR guides ........................................... 46
2.2 CRISPR deletions ....................................... 46
2.3 Primers for qPCR ........................................ 52
2.4 Embryo fixation time in 4% PFA at 4°C ............... 55
2.5 Antibodies ............................................... 55
3.1 In vitro -33 kb CRISPR deletions ....................... 61
3.2 Sox2 CRISPR deletion ................................... 69
3.3 In vitro +75 kb CRISPR deletions ..................... 76
6.1 In vivo CRISPR deletions ............................... 104
Chapter 1

Introduction

1.1 The spinal cord

The complex structure of your spinal cord is the reason you can pick up a piece of paper or move across a room. It is the extension of your brain running along your body that coordinates not only voluntary actions such as walking but also involuntary movements of the diaphragm and gut. The spinal cord consists of millions of neurons running in a bundle along your spine like high speed internet but with a diameter of just over a centimetre. While constantly receiving messages from the environment and your brain, the spinal cord is also coordinating outputs in fractions of seconds - really high speed internet.

Think of a dancer, a concert pianist or even me typing this now, humans have the ability to control our bodies with absolute precision due to our spinal cords. The full workings of this system still remains to be deciphered. We still cannot repair the spinal cord when it is damaged and we still do not know exactly how changes in action potential along neurons are integrated to direct intricate movements.

But how does such a complex system develop? The axis from our head to tail and front to back are determined early on in development and these axes establish where different neurons will arise. Where neurons are born, at what time and in what number are vital to the complex assembling of the spinal cord during development.
1.1. The spinal cord

Figure 1.1: Dorsal-ventral axis of the spinal cord
At the dorsal side (towards the back of the body) of the spinal cord, sensory neurons, whose cell bodies are located in dorsal root ganglia, project axons from the body and to interneurons relaying information from the environment to the CNS. Multiple types of interneurons relay information throughout the spinal cord and to the brain. Motor neurons are located ventrally and project their axons out of the spinal cord to muscles in the body. They receive input from interneurons relaying messages from the brain or sensory neurons.

1.1.1 Structure of the spinal cord

Early patterning events of the future spinal cord, the neural tube, lay the foundations of where different populations of neurons will arise. During initial patterning of the neural tube, both rostral-caudal (RC) and dorsal-ventral (DV), the head to tail and back to belly, axes are established by inductive signals. The RC axis is divided into four main sections, the forebrain, midbrain, hindbrain and spinal cord. The spinal cord is further divided into regions determined by the vertebrae that enclose it; cervical, thoracic, lumbar and sacral. In addition, regions are further divided. Within the cervical and thoracic regions identified as the brachial domain between vertebrae C5 to T1 different “columns” of motor neurons located on the ventral side of the spinal cord are generated across these domains with lateral column (LMC) neurons arising from the brachial domain directing axons into the limbs and autonomic motor neurons (MNs) generated from the remainder of the thoracic region projecting axons to muscles of the body wall (Hollyday et al. 1977, Landmesser 1978).

At the dorsal side of the spinal cord, input from sensory neurons in the pe-
1.1. The spinal cord

Ripheal nervous system relay information on touch, temperature, pain and position from the skin and musculoskeletal structures in the body via mechanoreceptors, thermoreceptors, nociceptors and proprioceptors respectively. Sensory neuron cell bodies are located in dorsal root ganglia adjacent to the spinal cord. They are derived from neural crest cells around the same time neurons in the spinal cord are differentiating (Anderson 1997). Signals from sensory neurons are relayed across the spinal cord and to the brain via neural circuits generated by multiple populations of interneurons. Equally, interneuron circuits send information from the brain to motor neurons initiating muscle contraction and therefore movement.

Encoded within our CNS are behaviours that depend on neuronal circuit’s called central pattern generators, which can be initiated by the brain but then do not rely on further coordination (Matsuoka 1987). For example, when we walk, our brain initiates the act of walking but does not direct individual neurons involved in movement and timing of muscle contraction, this is instead undertaken by central pattern generators located within the spinal cord. General principles for rhythmic motor pattern behaviours have largely been explored in smaller circuits in invertebrates as well as in fish and frogs (Marder & Bucher 2001, Pearson 1993).

1.1.2 Neurogenesis

The first neurons are born in the developing spinal cord between week 3.5 and 4 of human development and from E9.0 in mice. The period of neurogenesis in the spinal cord is relatively brief, occurring over just 3 days in mice until approximately E12.0. Different populations of neurons arise from defined populations of progenitors in the neural tube identified by the different transcription factors they express (Jessell 2000). Differentiation of neural progenitors to neurons in the spinal cord initially occurs ventrally before spreading more dorsally, starting with the motor neuron progenitor domain (pMN). Progenitors in the neural tube continue to symmetrically divide but the number of progenitors is offset by the rate of differentiation to neurons. These proliferative progenitor cells are located centrally and during differentiation they
migrate towards the periphery of the spinal cord. Neurons continue differentiating from these progenitors until E12.5 in mice. Progenitor domains along the dorsal-ventral axis differentiate at different rates resulting in changes in progenitor domain size during this period (Kicheva et al. 2014). It has been demonstrated that the particular time as well as the location neurons are born determines their identity and role in the spinal cord (Delile et al. 2019).

Once neurons have migrated to the periphery of the spinal cord, they begin to further arrange themselves into discrete populations. Throughout this period, post mitotic neurons further diversify through expression of different transcription factors. These different factors give pools of neurons their physiological signalling properties and direct their correct positioning and grouping within the spinal cord (Price et al. 2002, Osseward & Pfaff 2019). After the generation of neurons, the progenitor population within the neural tube undergoes a neural-to-glial switch and progenitors begin to give rise to glia.

1.1.3 Gliogenesis

Following neurogenesis, progenitors in the spinal cord give rise to glial cells. In mice, gliogenesis in the spinal cord begins at around E12.5 and can be identified by expression of transcription factors including Sox9 and NFIA within progenitors (Kang et al. 2012, Deneen et al. 2006). Glia are non-neuronal cells in the nervous system that support neurons by maintaining homeostasis, generating myelin and altogether providing protection and support for neurons. They make up approximately half of the cells in the central nervous system and are equally as important as neurons in CNS function. Three main classes of glia, astrocytes, oligodendrocytes and microglia, are responsible for maintaining homeostasis, myelination and immune response in the CNS. Oligodendrocytes myelinate axons of neurons the CNS and astrocytes have a range of functions including maintaining homeostasis in the CNS through their star-like projections attaching to neurons and blood vessels. These projections allow astrocytes to provide neurons with essential nutrients from the blood and maintain neuron survival (Zuchero & Barres 2015). Astrocytes and oligodendrocytes differentiate from neuroepithelial progenitor cells within the neural
tube following neurogenesis and similar to neurogenesis in the spinal cord, the
dorsal-ventral position and timing of differentiation of progenitors directs glial
subtype specification (Hochstim et al. 2008). In particular, oligodendrocytes
are generated from the Olig2 expressing progenitors in pMN domain (Pringle
et al. 1996, Poncet et al. 1996, Orentas et al. 1999). Microglia are the immune
cells of the brain, they patrol the CNS and respond to inflammation or disease
(Aguzzi et al. 2013). They are different in origin to astrocytes and oligoden-
drocytes as they are derived from the mesodermal haematopoetic stem cell
lineage in the yolk sac at E7.5 (Ginhoux & Prinz 2015).

The correct generation of neurons and glia in the spinal cord rely on pro-
genitors within the neural tube being assigned the correct identity early on in
development. Neural progenitor identity is governed by expression of different
transcription factors that direct cells to differentiate into distinct subsets of
interneurons and motor neurons at defined locations of the neural tube. These
domains of progenitors are arrayed in stripes across the dorsal-ventral axis of
the neural tube (Jessell 2000). This striped pattern of different transcription
factor expression domains is generated by cells interpreting their position in
the system from positional cues emitted from both dorsal and ventral poles
of the neural tube. Disruption to patterning at this early stage has down-
stream effects on both neurogenesis and gliogenesis as the correct number of
neurons and glia is an extension of the correct number of progenitors within
each domain.

## 1.2  Patterning of the neural tube

The neural tube is patterned across the dorsal ventral axis by two poles of
secreted signals, Sonic Hedgehog (Shh) emitted from the notochord and floor
plate and BMPs and WNTs from the roof plate (Fig. 1.1). These in addition
to retinoic acid (RA) signalling from the adjacent somitic mesoderm are re-
quired for patterning (Le Dréau & Martí 2012). These two poles of signal,
Shh ventrally and BMP/Wnt dorsally, generate morphogen gradients across
the neural tube patterning cells into 11 distinct progenitor domains (p3, pMN,
p2-p0, d1-6) (Alaynick et al. 2011). As the neural tube grows, progenitor
cells are able to retain their knowledge of previous position from interpretation of the two anti-parallel morphogen gradients of Shh and BMP allowing neural tube patterning to be maintained during growth (Kicheva et al. 2014, Zagorski et al. 2017). The patterned neural progenitor domains of the neural tube are characterised by expression of a distinct code of TFs. Cross-repressive interactions between different transcription factors delimit expression of specific TFs to individual domains of progenitors separated by sharp boundaries (Briscoe & Ericson 2001). These domains of progenitors in turn, differentiate into molecularly distinct subsets of neurons that populate the spinal cord.

Figure 1.2: Patterning of the neural tube

The neural tube is patterned by opposing morphogen gradients of BMP/Wnt secreted dorsally from the roof plate (RP) and Shh secreted ventrally from the notochord (NC) and floor plate (FP). Shh signalling generates anti-parallel gradients of repressor and activator forms of its conserved transcriptional effector protein, Gli. These gradients result in the formation of 11 distinct progenitor domains p3, pMN, p2-p0 in the ventral half and pd1-6 in the dorsal half. These progenitor domains give rise to molecularly distinct neurons V3, MN, V2-V0 and dl1-dl6 respectively.
1.2. Patterning of the neural tube

1.2.1 A gradient of Sonic Hedgehog patterns the ventral neural tube

Shh secreted from the ventral pole of the neural tube generates a gradient across the developing tissue. Through this gradient, Shh acts as a morphogen to pattern neural progenitors into distinct domains. Ventral to the neural tube sits the notochord, a rod of axial mesodermal cells that is vital for generation of the ventral most cell types (Placzek et al. 1991, Straaten et al. 1988). Shh secreted from the notochord in turn directs the floor plate to express Shh. Over time this generates a gradient of Shh protein across the ventral neural tube (Echelard et al. 1993, Riddle et al. 1993, Roelink et al. 1995, Placzek 1995). The floor plate continues to secrete Shh even after the notochord regresses and this maintains a gradient of Shh protein. This gradient of Shh dictates patterning of the ventral cell types that interpret their position depending on concentration of Shh. This has been demonstrated in vitro where different concentrations of Shh are able to direct differentiation of five different ventral cell types with Shh concentration reflecting the position of the cell type compared to Shh source in vivo (Ericson et al. 1997, Roelink et al. 1995). If Shh signalling is blocked, the most ventral cell types of the neural tube fail to generate and more dorsal fates are present in the ventral half of the neural tube (Chiang et al. 1996, Ericson et al. 1996).

1.2.2 Sonic Hedgehog gradient is relayed by a family of Gli proteins

Transmembrane proteins Patched1 (Ptch1) and Smoothened (Smo) mediate the intracellular transmission of Shh signal (Ingham 2001). Shh binds to Ptch1 which relieves its inhibition of Smo, enabling intracellular signal transduction. Smo has been previously shown to be required for transduction of graded Shh input in receiving cells (Wiigerde 2002, Hynes et al. 2000). Small molecule agonists that directly inhibit or bind and activate Smo are able to recapitulate the outcome of graded Shh input on receptive cells causing expression of ventral neural tube markers in vitro (Dessaud et al. 2007). In turn, Smo regulates the activity of a family of downstream bi-functional transcriptional effector proteins termed Gli proteins. This highly conserved family of Gli proteins
1.2. **Patterning of the neural tube**

can act as either repressors (GliR) or activators (GliA). In the ventral neural tube, the Shh gradient is relayed by both GliA and GliR to regulate expression target genes (Matise & Joyner 1999).

Three Gli proteins contribute to the expression of target genes in the neural tube, Gli1-3 (Jacob & Briscoe 2003, Vokes et al. 2007). Of these three Gli proteins, Gli3 predominantly acts as a repressor and Gli2 contributes mostly as an activator (Ding et al. 1998, Matise et al. 1998, Persson et al. 2002). In cells exposed to low levels or no Shh, GliR is present, this is due to the partial proteolytical cleaving of Gli3 to Gli3R and the degradation of Gli2 (Pan et al. 2006). It has been demonstrated that Gli3R plays an important role in the correct positioning and restriction of dorsal boundaries of the ventral domains (Persson et al. 2002, Oosterveen et al. 2012). One demonstration of this is in Gli3 mutant embryos that displayed dorsally expanded expression of some ventral neural tube markers. Interestingly, these boundary shifts could be rescued by addition of Gli3 that was only able to be in its repressor form (Persson et al. 2002).

At ventral positions, high levels of Shh prevent Gli3 from being proteolytically cleaved to its repressor form. In agreement with this, in Shh null embryos, genetic removal of Gli3 leads to a partial recovery of ventral cell types pMN, p2 and p1 demonstrating that one of the main functions of Gli3 is to repress ventral fates (Persson et al. 2002). Although these embryos show partial recovery, in order to generate cells with p3 identity, GliA is required (Persson et al. 2002). In the ventralmost domains, p3 and floor plate, accumulation of GliA via Gli2 and Gli1 drive gene expression (Lee et al. 1997). Gli2 knockout mice do not generate a floor plate and most V3 interneurons (Ding et al. 1998, Matise et al. 1998). Gli1 has been shown to only act as an activator, however, genetic knockout of this gene does not affect neural tube development (Park et al. 2000). However, knockout of Gli1 in addition to Gli2 results in a more severe phenotype than Gli2 single knockout regarding the induction of ventralmost TFs suggesting that different Gli proteins may compensate for each other (Ding et al. 1998). These studies together demonstrate that different levels of
1.2. Patternning of the neural tube

repressive and active Gli activity are involved in regulating expression of Shh target genes in the ventral neural tube.

1.2.3 Time of Sonic Hedgehog exposure directs ventral cell fates

As well as Shh concentration directing cell fate, the length of time exposed to Shh is important with the ventral-most domains only formed at the highest concentrations as well as the longest exposure time (Briscoe et al. 2000, Ericson et al. 1996, Dessaud et al. 2007, Ribes & Briscoe 2009). In Shh/Gli3 double knockouts pMN, p2 and p1 fates are generated, implying that absolute levels of Shh alone do not determine cell fate (Persson et al. 2002). Therefore the system patterning the ventral neural tube requires not only the concentration of Shh but also the time exposed directs cell fate.

The order in which ventral cell fates are specified also provides evidence that timing of Shh exposure is important for cell fate transitions. Sequential induction of Olig2 and Nkx2.2 is observed in vivo with Olig2 being induced initially in cells at the ventral midline that over time transition to Nkx2.2-expressing p3 progenitors (Jeong & McMahon 2005, Stamataki et al. 2005). Lineage tracing studies showed that it was indeed the same cells that transition through more ventral progenitor cell fates over time due to Shh exposure (Dessaud et al. 2007). Taken together, these studies show that length of Shh exposure and concentration determine where in the neural tube different fates will be specified.

1.2.4 Adapting dynamics of Shh signalling

Analysis of a fluorescent reporter of Gli activity in the neural tube, constructed using 8 concatemerised Gli binding sites driving GFP expression, showed that over time there are changes in the level of Shh activity (Balaskas et al. 2012). Gli activity reaches a peak of activity at around E9.0 before then decreasing. However, this change in Shh activity is not linked to a change in Shh protein gradient which remains consistent over this time period (Chamberlain et al. 2008). This adaptation of cells to Shh signalling over time means that ventral fates generated after the longest period of Shh exposure are generated at a lower concentration of Shh than more dorsal, earlier specified domains that
1.2. Patterning of the neural tube

Figure 1.3: Neural tube patterning occurs between E8.5 to E10.5

Cells at the ventral midline transition from Pax6 to Olig2 to Nkx2.2 expression over time. Immunofluorescence of brachial embryo sections at different stages stained with antibodies against Pax6 (blue), Olig2 (red) and Nkx2.2 (green). Sections generated, stained and imaged using methods described (Chapter 2). Scale bar is 50µm.

had a shorter exposure to a higher concentration of Shh signalling, confirming that absolute levels of Shh activity do not dictate ventral progenitor fates.

Through computational analysis of the Shh signalling pathway three mechanisms have been proposed for the dynamic change in Gli activity over time (Cohen et al. 2015, Junker et al. 2014). It has been shown that in response to Gli activity, the inhibitory receptor Ptch1 is upregulated (Dessaud et al. 2007). Biochemical experiments have shown differential stability of Gli isoforms in response to Shh signalling (Wang et al. 2010, Wen et al. 2010, Humke et al. 2010) and finally, transcriptional down-regulation of both Gli2 and Gli3 as well as other components of the pathway contribute to reduced activity over time (Lee et al. 1997, Peterson et al. 2012, Junker et al. 2014, Ribes et al. 2010).

1.2.5 Cross-repressive transcription factors specify progenitor domain boundaries

To clearly define different progenitor domains, expression of different transcription factors need to limited to specific locations. This is achieved through cross-repressive interactions between classes of transcription factors at domains boundaries allowing sharp boundaries to be generated. In early work, transcription factors in the ventral neural tube were subdivided into two classes
1.2. Patterning of the neural tube

Figure 1.4: Cross-repressive transcription factor interactions specify progenitor domain boundaries

A Expression of different transcription factors give progenitors in different domains distinct identities that will determine their subsequent differentiation to molecularly distinct neurons.

B Genetic perturbations knocking out different transcription factors alters neural tube patterning.

C Transcription factors Olig2 and Pax6 are expressed in pMN and p2 progenitors. Expression of these two transcription factors is required to limit the dorsal boundary of the p3 domain containing Nkx2.2-expressing progenitors. Embryo ages is equivalent to e9.0. Scale bar 50µm.

Figure adapted from both (Dessaud et al. 2008, Balaskas et al. 2012)

Based on their regulation by Shh (Briscoe et al. 2000). Class I transcription factors are constitutively expressed in the neural tube and Shh signalling represses their expression in the ventral-most domains whereas Class II transcription factors are dependent on Shh signalling for their expression. These differential responses to certain levels of GliA and GliR restrict transcription factors to certain dorsal-ventral regions within the neural tube.

Pairs of Class I and Class II transcription factors cross-repress each other preventing expression in incorrect domains leading to the formation of sharp boundaries of expression. One example of these pairs is Class II gene \textit{Nkx2.2},
expressed in the p3 domain, and Class I gene, *Pax6*. This pair of transcription factors share a domain boundary, between p3 and pMN domains. Forced expression of *Nkx2.2* dorsally is able to remove *Pax6* expression and genetic knockout of *Pax6* results in a dorsal expansion of *Nkx2.2* (Briscoe et al. 2000). When both Olig2 and *Pax6* are knocked out, the Nkx2.2 domain is expanded even further dorsally provided further confirmation that location of ventral gene expression is not solely determined by Shh concentration (Balaskas et al. 2012). A similar cross-repressive relationship between Class I and II genes is observed between *Nkx6.1* and *Dbx2* (Briscoe et al. 2000).

In addition to cross-repressive relationships between Class I and II transcription factors, repressive interactions between Class II factors also exist. A cross-repressive interaction between *Olig2* and *Nkx2.2* contributes to the formation of a sharp boundary between the p3 and pMN domain (Fig.1.4)(Sun et al. 2001, Novitch et al. 2001). The pMN/p2 domain boundary is maintained by cross-repressive interactions between *Olig2* and *Irx3* (Novitch et al. 2001).

The combination of interactions between TFs in the developing neural tube forms a gene regulatory network (GRN) that can interpret Shh signalling over time resulting in the expression of an identifying code of transcription factors in sharply separated domains.

1.3 Gene Regulatory Networks in Patterning

Gene Regulatory Networks (GRNs) provide a framework to understand regulatory systems that control developmental events. A GRN is a logic map that shows details of inputs at cis-regulatory modules for different transcription factors. GRNs are a tool to visualise specific spatiotemporal expression of a single gene within a much larger network (Davidson 2002). These regulatory maps vary in size with some including up to 100 genes assembled in smaller sub-circuits connected together through multiple interactions (Oliveri et al. 2008). The ability to assemble these systems allows those who use them to take a step back from individual interactions and focus on the system as a whole. Graphical representations of network circuitry in the form of a topological model display the genes involved (nodes) and how output of one gene
inputs onto others in the system (edges). These edges represent the direct functional interactions between TFs and their cis-regulatory DNA, enhancers. Determining the structure of a GRN relies on identification of regulatory inputs at each node as well as the logic linking nodes to determine output.

GRNs facilitate a rapid response to signalling in embryonic patterning. Initial boundaries between different neural progenitor identities in the ventral neural tube form in the first 12 hours of Shh exposure (Briscoe & Ericson 2001, Jeong & McMahon 2005). During development some GRNs can generate an even faster spacial response to patterning signals. One example is in the transcriptional GRNs present in the pre-gastrulation mouse embryo downstream of Nodal, Wnt and BMP signals required to determine axis specification (Rossant & Tam 2004). Nodal is rapidly restricted to the future posterior side of the embryo where the primitive streak will form (Zhou et al. 1993, Conlon et al. 1994, Collignon et al. 1996). One transcription factor expressed downstream of Nodal and Smad2/3 signalling that promotes anterior visceral endoderm (AVE) formation, is Eomes (Ciruna & Rossant 1999). Eomes is expressed in proximal posterior cells of the E6.5 embryo and is directs epithelial-to-mesenchymal transition of future mesoderm populations (Arnold et al. 2008). Correct expression of Eomes in the AVE is required for its migration at the onset of gastrulation (Arnold et al. 2008, Ciruna & Rossant 1999). Although a complete GRN has not been assembled for rapid restriction of Nodal, investigation into the regulation of Eomes in response to Nodal has been carried out through analysis of a +8 kb enhancer (Simon et al. 2017). This work has shown that deletion of this enhancer leads to variably penetrant defects in axis formation (Simon et al. 2017). Functional analysis of the cis-regulatory elements for transcription factors included in these developmental events is key to explore and describe the GRNs directing these events.

1.3.1 Cis-regulatory elements

To assemble a specific GRN, interactions between different nodes of a network need to be determined. The nodes, transcription factors, of a GRN interact via binding to cis-regulatory elements, enhancers. When bound by the activating
transcription factors, an enhancer is thought to loop to link with the promoter of a certain gene, this recruits DNA polymerase and initiates gene expression (Benabdallah & Bickmore 2015). How an enhancer interprets the binding of multiple elements to direct a single gene output is still not well understood. Many aspects including multiple enhancers directing expression of a single gene, chromatin landscape and redundancy in enhancer usage make challenging to explore enhancer logic (Spitz & Furlong 2012). Multiple techniques exist to discern locations of cis-regulatory elements of different genes.

Enhancers have previously been identified by assessment of highly conserved areas in the genome (Sandelin et al. 2004). More recently, extensive genome-wide studies that use multiple techniques to map where particular proteins interact with the genome have annotated thousands of enhancer regions in a variety of cell types and organisms (Sakabe et al. 2012). ATAC-seq, a technique to map open regions of chromatin (Fig. 1.5), has enabled researchers to study the commissioning and decommissioning of enhancers throughout cellular differentiation without needing to know specific binding factors (Buenrostro et al. 2013, Cusanovich et al. 2015). This assay is extremely useful to explore global changes in genome accessibility in response to different regulatory signals. The number of accessible regions ATAC-seq identifies can make it a complex task to assign regions to particular genes they direct, particularly as it has been shown that enhancers can direct expression from up to 1 mB away (Lettice 2003). In addition, multiple enhancers have been shown to direct expression of a single gene making regulation of a gene complex to interpret (Krivega & Dean 2012).

When trying to identify a gene specific enhancer, previous knowledge of transcription factors that directly influence gene expression can be used. Chromatin immunoprecipitation sequencing, (ChIP-seq), uses antibodies against known proteins to identify regions in the genome where they bind (Visel et al. 2009). This technique has also widely been used to annotate binding locations for both active and repressive chromatin markers which identifies enhancers or repressors of gene activity (Johnson et al. 2007, Park 2009). These analyses
1.3. Gene Regulatory Networks in Patterning

**Figure 1.5: ATAC-seq to identify open regions of chromatin**

A hyperactive form of Tn5 transposase simultaneously cleaves and adds adapters to nucleosome-free regions of DNA, priming them for sequencing. Following sequencing, fragmented DNA sequences are aligned back to the genome. Multiple reads across certain regions highlight areas of open chromatin in the genome.

can indicate whether a predicted enhancer has appropriate factors bound for it to drive expression of the gene of interest.

An alternative approach to link enhancer regions to genes they regulate is to explore the chromatin structure. Multiple techniques have been developed that capture the 3D organisation of chromatin (Dekker 2002, Rowley & Corces 2018). These data shows the frequency of interactions between different regions of DNA across the genome. These approaches have also highlighted isolated regions in which the majority of enhancer-gene interactions occur called Topologically Associated Domains (TADs) (Dixon et al. 2012, Nora et al. 2012). The importance of TAD boundaries and their requirement to prevent interactions occurring from regulatory regions in one TAD to genes in a second is still being investigated. It had previously been proposed that these boundaries are required for correct gene expression (Ruiz-Velasco & Zaugg 2017). However, recent work analysing the HoxD locus has suggested that these boundaries may not be as important as previously believed (Rodríguez-Carballo et al. 2019).
1.3. Gene Regulatory Networks in Patterning

Equally, work that uses highly disrupted Drosophila balancer chromosomes has shown that in this case only a small number of genes were affected by chromosome reorganisation (Ghavi-Helm et al. 2019).

One approach to explore transcription factor binding at enhancers is to analyse these regions at the DNA sequence level. Each transcription factor typically identifies a collection of similar DNA sequences and these binding site motifs can be represented in position weight matrix models (PWMs) (Stormo 2013). The DNA sequence can scanned for binding motifs of different factors. This can be used to predict the binding location and binding affinity of a known protein within an enhancer region. However, there are caveats to using this approach. Individual TFs can recognise many different motifs making it challenging to identify sites and even if sites are located using this approach. Additionally, it is important to consider that bound transcription factors does not always equal gene expression. Finally, DNA structure, chromatin markers, binding co-factors and sequence either side of binding motifs may influence TF binding (Inukai et al. 2017).

Once predicted, enhancer reporters can be used to assess where activity of the enhancer manifests and if this coincides to the predicted gene expression. These are often constructed by using enhancer DNA linked to a minimal promoter to drive expression of LacZ or a fluorescent protein (Banerji et al. 1981, Shlyueva et al. 2014). This approach has been widely used to analyse the logic of transcription factor binding at an enhancer and often includes mutating, reorganising or adjusting the affinity or number of binding sites (Oosterveen et al. 2012, Peterson et al. 2012, Crocker et al. 2017). Perturbations to enhancers studied under synthetic or in vitro conditions have shed light on gene regulatory logic but often when these are transferred to in vivo models phenotypes are absent or very mild compared to their in vitro counterparts. This occurs even when enhancers are highly conserved i.e. “super enhancers” (Dickel et al. 2018, Osterwalder et al. 2018). The phenotypes from in vivo enhancer knockout studies might be explained by the fact that genes can have multiple enhancers that coordinate expression (Levine 2010). Crucially these enhancers
may only act when a primary enhancer is affected therefore acting as a “shadow enhancer” (Barolo 2012). These enhancers may not ordinarily be necessary to drive gene expression but when a primary enhancer is affected, these enhancer will ensure close to normal expression is maintained (Hong et al. 2008, Perry et al. 2010). Due to these observations, shadow enhancers have been proposed to provide robustness to developmental regulatory networks (Cannavò et al. 2016).

Using a combination of these techniques, analysis of interactions between a network of genes can be uncovered. Once these interactions have been uncovered, assembly of different interactions between nodes can be assembled into a GRN.

1.3.2 Construction of GRNs

One of the first constructed patterning GRNs is one dictating early patterning of the sea urchin. This network describes the sequential regulatory changes that occur during the early specification of different cell types in the sea urchin embryo (Davidson 2002, Oliveri et al. 2008). The structure of this networks is based upon perturbations, quantative expression data as well as knowledge and direct testing of cis-regulatory inputs at different nodes. The sea urchin GRN has been extensively built upon to the extent that it now contains over 100 transcription factors and extends much further through development (Peter & Davidson 2010).

GRNs constructed and visualised in this way are useful to identify GRN hierarchy, the points at which interactions occur between subcircuits and the collective output of the network in terms of spatial and temporal gene expression. It has been most valuable for the exploration of genetic control of different cell fate processes, for example, epithelial to mesenchymal transition (EMT) for which there are 5 separate subnetworks directing different aspects of the EMT process (Saunders & McClay 2014). One example of GRNs assembled in a similar manner include early heart specification in mice. This research demonstrated the two stable regulatory states representing both the first and second heart field (Herrmann et al. 2012).
Although models constructed in this way are useful to identify outputs of multiple linked subcircuits, in order to understand dynamical behaviour individual subcircuits need to be analysed in isolation. This is because in order for a network to be considered in dynamical terms, it needs to be small enough so parameters values determining output can be screened. With each additional node considered, the number of parameters quickly becomes unsuitable for screening different values (Hecker et al. 2009). Ordinary differential equation (ODE) models can be used to investigate dynamical behaviour of small (less than 10 genes) GRNs. These models are a system of equations used to describe the changes of output of gene product concentration in response to change in the inputs (Karlebach & Shamir 2008).

1.3.3 Interpretation of morphogen gradients

During embryonic patterning, graded signals are vital to provide positional information in developing tissues. Multiple methods of interpreting morphogen gradients exist in order to correctly relay positional information (Rogers & Schier 2011). It has become evident that multiple levels of interpretation of graded signals results in specific downstream patterning and often these can be integrated in GRNs (Cotterell & Sharpe 2010).

It had previously been proposed that the structure of enhancers of downstream target genes dictates their expression domain in response to morphogen gradients. A proposed enhancer-directed method of morphogen interpretation is the affinity-threshold model (Driever et al. 1989). The model proposes that binding sites at enhancers for genes expressed in response to a graded input have varying binding affinity which determines their sensitivity to morphogen input and these different affinities produce responses of different genes along the gradient. Enhancers for genes expressed closest to the source have a small number of low affinity sites and those activated furthest away have a large number of high affinity sites. This model was proposed during analysis of an enhancer for one of the four gap genes, *hunchback* (Hb), expressed in response to an anterior-posterior gradient of maternal Bicoid protein. Hb is expressed at more posterior positions in the Drosophila blastoderm and its enhancer
contains multiple high affinity Bicoid binding sites (Driever et al. 1989).

Further analyses of enhancers downstream of the maternal Bicoid gradient suggested that the affinity-threshold model was not able to account entirely for expression of downstream patterning genes. Computational analyses on a number of Bicoid enhancers in this system indicated that there was a poor correlation between strength of binding affinity and the AP expression limits of a gene (Ochoa-Espinosa et al. 2005). Differential binding affinity has also been ruled out for interpretation of Shh in the ventral neural tube. Previously, ChIP-seq analysis identified Gli binding sites at enhancers of ventral neural tube genes (Oosterveen et al. 2012, 2013, Peterson et al. 2012, Vokes et al. 2007). As observed for Bicoid binding sites (Ochoa-Espinosa et al. 2005), binding affinity analysis of these Gli binding sites also has the opposite correlation to the proposed threshold-affinity model. Instead, genes expressed closest to the source of Shh were shown to have the highest affinity binding sites (Oosterveen et al. 2012). These data suggest alternative methods of signal integration at enhancers is required for correct gene expression and cannot be relied simply on binding sensitivity of enhancers to morphogen signalling.

In order to determine sharp boundaries of gene expression, a GRN needs to be able to translate small differences from a graded input into all or nothing changes in gene expression. To achieve this, interactions between nodes of a network generate regions of bistability.

1.3.4 Cross-repressive transcription factors

One feature of cross-repressive interactions between transcription factors is the formation of bistable switches between fates in neighbouring domains (Ferrell 2002). This results in on/off gene expression toggle-switch between cells with either side of an expression domain boundary. Bistable switches are a mechanism to convert a graded input, i.e. from a morphogen, to a sharp transition between neighbouring fates (Saka & Smith 2007, Cotterell & Sharpe 2010). The presence of repressive interactions such as these are often seen between nodes in GRNs (Peter & Davidson 2010).

As seen for the ventral neural tube, pairs of transcription factors cross-
1.3. Gene Regulatory Networks in Patterning

Figure 1.6: Multiple inputs direct transcription factor network output

Schematic of an enhancer driving ventral neural tube transcription factor expression. Inputs at a TF enhancer include morphogen signalling (GliA/R); ubiquitously expressed TFs (Sox2) and other TFs expressed. Understanding of these interactions enables us to construct topological models of GRNs driving patterning.

repress each other to limit expression to specific domains (Briscoe et al. 2000, Ericson et al. 1997). Through ChIP-seq and bioinformatic analyses, binding sites for ventral neural tube TFs have been located at enhancers of other ventral neural tube genes (Kutejova et al. 2016). Bistable switches between Olig2 and Nkx2.2 as well as between Olig2 and Irx3 contribute to sharp boundaries of gene expression between the p3, pMN and p2 domains (Balaskas et al. 2012, Perez-Carrasco et al. 2016).

Within the Drosophila Gap gene network, cross-repressive pairs of genes generate regions of bistability that translate as sharp boundaries between expression domains defining future segments of the body plan (Sanchez & Thiery 2001). The structure of these cross-repressive interactions a GRN that interprets the Bicoid morphogen gradient (Jaeger 2011). Mathematical modelling has been able to recapitulate this GRN in silico (Ashyraliyev et al. 2009, Verd et al. 2017).

In these systems, the cross-repressive interactions generate a genetic toggle switch, but how the relationship responds to changes in gene expression is vital for determining function. Patterning during early patterning events in the embryo is a highly dynamic process therefore dynamic processes within cells need to be considered when constructing GRNs. For example, different degradation or production rates of proteins can have a significant effect on bistable switch behaviour (Verd et al. 2017, Perez-Carrasco et al. 2016).
1.3.5 Ubiquitously expressed transcription factors

In both the neural tube and Drosophila blastoderm, morphogen gradient is just one of the inputs into TF expression. The majority of enhancers of A-P patterning gap genes in Drosophila contain binding sites for the ubiquitously expressed transcription factor Zelda. Zelda is expressed throughout the entire blastoderm embryo and has been shown to bind to many different developmental enhancers at this stage. Experimental data has shown that Zelda is required for correct gap gene expression and altering binding of Zelda to these enhancers disrupts patterning (Kanodia et al. 2012). More recently, synthetic network construction in the Drosophila blastoderm was able to show the importance of Zelda binding sites in patterning enhancers (Crocker et al. 2017). In the ventral neural tube, uniformly expressed SoxB1 family of TFs (Sox1-3) can be thought of as playing a similar role to Zelda in enabling expression of neural tube TFs (Bergsland et al. 2011, Oosterveen et al. 2012, Peterson et al. 2012).

ChIP-seq analysis has shown Sox2 binding at enhancers of neural tube TFs (Peterson et al. 2012, Nishi et al. 2015, Kutejova et al. 2016). Experimental manipulation of these sites has demonstrated the importance of Sox2 binding for correct gene expression in the ventral neural tube. In particular, removal of Sox2 sites in Gli-bound enhancers for Nkx6.1 and Nkx2.2 removes reporter expression in chick neural tube electroporations (Peterson et al. 2012). In addition, 4 Gli binding sites from ventral neural tube genes Olig2 and Nkx6.1 failed to drive reporter expression in vivo implying that positive input from SoxB1 binding is required alongside morphogen input to responding genes (Oosterveen et al. 2012). These data demonstrated, as had been the case for Zelda in Drosophila, that input of uniformly expressed TFs to cis-regulatory regions are required to enable expression of patterning TFs.
1.3.6 Mathematical model for ventral neural tube patterning

The signals involved in patterning the ventral neural tube together generate a complex system. The combination of GliA and GliR inputs, the adaptation of Shh signalling over time and cross-repressive TF interactions mean that investigations into patterning can quickly become difficult to understand. To further our understanding of how the ventral neural tube is patterned our knowledge on all of these different interactions needs to be assembled. This can be achieved by constructing a GRN that encompasses these signalling interactions that can then be mathematically modelled. Having a mathematical model will enable us to test perturbations that may be difficult to attempt or analyse in vivo.

Using our knowledge of different transcription factor relationships, the influence of Sox2 and the complicated outcomes of differential Gli binding at enhancers it has been possible to build upon existing topologies of the ventral neural tube GRN and explore mathematical modelling approaches to understand patterning (Oosterveen et al. 2012, Peterson et al. 2012, Balaskas et al. 2012). Although many different TF interactions occur during the patterning of the ventral neural tube, the GRN interpreting Shh signalling can be reduced to a subnetwork including only four of these transcription factors: Pax6, Olig2, Nkx2.2 and Irx3 (PONI) (Balaskas et al. 2012, Cohen et al. 2014). The lab developed a mathematical model recapitulating ventral neural tube patterning using this four node network with a graded input from Shh signalling (Cohen et al. 2014). The following describes how this model was constructed.

In this model, expression of a given gene is determined by the probability of polymerase bound at its promoter. The probability of bound polymerase is determined from the probability of binding of the other network factors: morphogen (GliA/R), uniformly expressed and domain specific TFs in the network.

Most Shh target genes only have one Gli binding site (Peterson et al. 2012), so only one GBS is considered in the model. Due to the bifunctional forms of Gli, this site can bind with equal affinity to both GliA and GliR.
1.3. Gene Regulatory Networks in Patterning

Figure 1.7: Mathematical model for ventral neural tube patterning by Shh

A Transcription factor expression in the three ventralmost domains of the neural tube (p3, pMN, p2). Expression pattern of the four genes used to model ventral neural tube patterning, Pax6, Olig2, Nkx2.2 and Irx3. Olig2-expressing progenitors in the pMN will differentiate into motor neurons (MNs).

B Nkx2.2, Olig2, Irx3 and Pax6 form a cross-repressive network with the regulatory links indicated in the diagram. Pax6 and Irx3 are expressed in the absence of any Shh signal. Nkx2.2 and Olig2 are regulated by Gli activity.

C Temporal dynamics of the model at a position that will have ultimately have Nkx2.2 expression. As Shh signal increases, Irx3 and Pax6 are quickly downregulated at this position. Over time, Shh signalling causes cells to up-regulate Olig2 then ultimately transition to Nkx2.2, the final steady state for this neural tube position.

D The sets of target patterns for wild type and six different mutants are shown in the top row. The x-axis represents the ventral-dorsal position in the neural tube extending away from the morphogen source. The y-axis represents protein concentration of the four transcription factors.

*Figure adapted from (Cohen et al. 2014)*

Input of cross-repressive TF interactions must also be considered in modelling ventral neural tube patterning. Input at each node includes two binding sites that can be occupied by other TFs from the network. These two sites work independently to repress probability of gene expression with each TF is acting as a strong repressor. In addition to Gli and alternative TF input, pan-neural input from Sox2 has shown to be important for expression of target genes in the ventral neural tube (Peterson et al. 2012). To keep the model in its
simplest form, uniformly expressed TFs are accounted for as part of the basal input of gene expression. For construction of the model, equations for the above probabilities of gene expression can be constructed and used to form an ordinary differential equation (ODE) network that takes into account the production and degradation of different TFs across a GliA/GliR gradient.

To start, Bayesian approximation methodology (Toni et al. 2009, Turner & Van Zandt 2012), a method to estimate posterior distributions of parameters for simulation-based models, was used to identify parameter sets that could recapitulate wild type patterning as well as the sequential onset of Olig2 then Nkx2.2 as well as the step-wise repression of Pax6 in the pMN domain. Once these parameter sets were uncovered, they were further reduced to include those that could also recapitulate patterning outcomes observed in mutant phenotypes (Briscoe et al. 2000, Novitch et al. 2001, Persson et al. 2002). This provided a smaller parameter space to explore the model.

From the model, strengths of different repressive TF interactions could be inferred, for example, Nkx2.2 acts as a stronger repressor than Olig2 or Pax6. Equally, manipulations of Gli binding affinity aligned with outcomes modifying GBS in enhancer reporters (Oosterveen et al. 2012). Including, the counter intuitive result of expansion of Olig gene expression when Gli binding affinity is reduced, even though Olig2 is activated by Shh signalling. The reason for this observation is accounted for by the single GBS in the model. The site binds both GliA and GliR therefore reducing the binding affinity, reduces the GliR binding in the region where the dorsal limit of the Olig2 domain usually is. This results in an expansion of the domain as GliR is no longer repressing enough to restrict it to its wild type position.

The model provides a useful tool to explore how alterations at different cis-regulatory elements for TFs in the ventral neural tube may affect patterning. However, there are some limitations to the model. The gradients of GliA/GliR are set for the timescale of the model which does not take into account the dynamic change in Gli activity over time (Cohen et al. 2015). Equally, in order to properly explore how boundaries between expression of different TFs form,
bistable switches need to be explore under conditions including noise. This would account for fluctuations in gene expression over time observed in gene expression processes including transcription and translation and would enable us to explore how this GRN is constructed to ensure robust patterning in the presence of noise (Elowitz 2002).

1.4 \textit{In vitro} generation of neural progenitors

Our understanding of the developing nervous system has been influenced by different techniques used to investigate development. A key method used across multiple areas of development is the use of \textit{in vitro} models of cell differentiation. Embryonic stem cells have the ability to differentiate along multiple pathways and much work has been done to establish reproducible \textit{in vitro} protocols that recapitulate known differentiation patterns.

\textit{In silico} methods of predicting patterning and investigating cis-regulatory interactions between TFs of the neural tube is one approach to generate predictions that can then be tested experimentally. Previous work analysing enhancer regions of neural tube TFs used reporter assays to investigate differential binding (Oosterveen et al. 2012, Peterson et al. 2012). One caveat of this approach is that this read out of expression is unable to then feedback into the transcription factor network so we are unable to test how these alterations would affect patterning as a whole. We therefore wanted to target enhancers within their loci. Prior to targeting directly in mice, it is often more efficient to test in a more accessible system. In our case we decided to use \textit{in vitro} models of differentiation to investigate enhancers.

Differentiation of ESC to neural progenitors is a promising \textit{in vitro} approach to model neural development as it allows easier examination of signalling pathways outside of the early embryo. Recently a method that recapitulates the developmental trajectory of spinal cord has been developed (Gouti et al. 2014, Sagner et al. 2018). To generate \textit{in vitro} progenitors with a spinal cord identity it is important to consider the steps in development that generate these cell types. Initially, inductive signals pattern the neural plate along both rostral-caudal (RC) and dorsal-ventral (DV) axes. Along the RC axis, the
main subdivision of the CNS are established: forebrain, midbrain, hindbrain and spinal cord. The neural plate undergoes complex folding by generating furrows and hinges to form a tube. The resulting structure at spinal cord levels is the neural tube with overlaying epidermis and delaminating neural crest cells in between. It was initially believed that the entire CNS developed from the neural plate (Nieuwkoop & Nigtevecht 1954) but it has been shown that in addition to the neural plate, a second population of cells also contributes to the formation of the spinal cord. Lineage tracing experiments demonstrated that spinal cord progenitors share a common ancestor with paraxial mesoderm cells (Tzouanacou et al. 2009). This population of cells, termed neuromesodermal progenitors (NMPs), are located in the caudal lateral epiblast and can be identified by the co-expression of Brachyury and Sox2 (Henrique et al. 2015). As the embryo undergoes axis elongation, this population is retained posteriorly and over time contributes to the neural tube and paraxial mesoderm. FGF and Wnt signalling induce Cdx transcription factors and the cause the sequential activation of Hox gene clusters. The Hox code expressed by NMPs when they exit the NMP-state is retained by either paraxial mesoderm or neural tube tissue at that AP position. In vitro, spinal cord progenitors differentiated via an NMP state express more posterior Hox genes and also have a different epigenetic landscape due to their earlier expression of Cdx2 (Metzis et al. 2018, Gouti et al. 2014). These data provide insights into the differing epigenetic control of spinal cord identity depending on AP position of progenitors. One reason this is an important to consider when investigating the patterning of neural progenitors in the developing spinal cord is because, depending on AP position along the axis, different columns of motor neurons are specified (Jessell 2000).

At the start of my PhD the lab had recently published a protocol for generating posterior spinal cord progenitors (Gouti et al. 2014). Prior to this, spinal cord neural progenitor differentiation protocols largely generated neural progenitors with a cervical identity, expressing Hox1-5 genes, but not more posterior Hox6-9 paralogs that determine brachial and thoracic fates (Liu et al.
1.4. *In vitro generation of neural progenitors*

Figure 1.8: *In vitro differentiation of mouse ES cells to neural progenitors with different anterior-posterior fates*

A. Diagram of E9.5 mouse embryo highlighting anterior, posterior and spinal cord regions.

B. Differentiation of mouse ES cells over 5 days directed towards anterior, hindbrain and posterior neural tube fates. The addition of SAG (Smoothened agonist), generated progenitors with a ventral neural tube fate. RA (retinoic acid).

C. The addition of Wnt signalling between days 2 and 3, results in cells with neuromesodermal progenitor identity at day 3. The addition of RA/SAG is then able to transition cells to a spinal cord neural progenitor fate.

2001, Wichterle et al. 2002). This protocol was able to generate spinal cord progenitors by first differentiating cells via an NMP state. To achieve this, mouse ES cells are plated as single cells in media containing FGF for 48 hours; this initiates cells to differentiate from stem cells towards a “epiblast-like” state. Cells are then exposed to a 24 hour pulse of Wnt signalling, mimicking the posterior Wnt signalling cells are exposed to in the posterior part of the developing embryo. This Wnt pulse posteriorises the cells transitioning them to NMPs. Addition of retinoic acid (RA) and Shh signalling (SAG - a Smoothened agonist) to NMPs direct the cells to a neural fate and ventral fate respectively. At day 4 of the differentiation, cells express transcription factors expressed by posterior neural tube progenitors, *Sox2*, *Pax6*, and *Irx3*. At day 5 a large number of *Olig2*-expressing cells (~70%) are generated followed by the appearance of *Nkx2.2*-expressing cells at day 6 (Sagner et al. 2018). Addition of RA and SAG without a pulse of Wnt signalling at day 3 generates cells with a more anterior fate that express hindbrain markers at day 5 i.e. *Phox2b* (Gouti et al. 2014). The ability to generate cells with different anterior-posterior identities allows us to explore differential regulation across
this axis even though the patterning remains the same.

1.5 Project aims

Before the start of this project, our understanding of the contribution of cis-regulatory elements within ventral neural tube GRN to patterning was only beginning to be explored. Recent data had shed light on Gli and Sox2 binding at enhancers within the ventral neural tube and results from direct testing of these sites provided the knowledge to build a GRN that could be mathematically modelled in silico (Oosterveen et al. 2012, Peterson et al. 2012, Cohen et al. 2014). Using this model, interesting predictions had been drawn on how altering the edges of the network would influence patterning downstream of Shh signalling. Previous analysis on cis-regulatory elements directing gene expression in this network had largely been as a readout of enhancer reporter constructs so did not indicate patterning outcomes if these changes were made to expression of that gene. In order to test patterning predictions, we needed to analyse ventral tube GRN enhancers in their own loci before testing specific predictions from the model. The system to test these predictions needed to facilitate manipulation of DNA and also easy analysis of changes in gene expression. In the lab we used an in vitro differentiation protocol that directs mouse ES cells towards a ventral spinal cord progenitor fate to explore enhancers of neural tube patterning genes (Gouti et al. 2014). A recent generation of ATAC-seq data in the lab from these in vitro generated progenitors allowed us to explore genome accessibility surrounding patterning target genes in our in vitro differentiation protocol (Metzis et al. 2018). This confirmed accessibility of previously identified enhancers that could be explored.

• Aim 1 - Explore genetic regulation of ventral neural tube patterning genes

To analyse regulation of ventral neural tube genes, we used our in vitro differentiation protocol of mouse ES cells to spinal cord progenitors (Gouti et al. 2014). With published ChIP-seq, functional dissection using enhancer reporters and supportive evidence from ATAC-seq generated in the lab, the -33 kb enhancer of Olig2 provided a good candidate to explore (Oosterveen et al. 2012, Pe-
1.5. Project aims

terson et al. 2012, Kutejova et al. 2016, Metzis et al. 2018). We wanted to understand how this enhancer, with inputs from Sox2, Gli and Nkx2.2, regulated gene expression *in vitro*. To answer this question, we initially needed to test the whether *Olig2* expression is being regulated by this enhancer in our *in vitro* system. Once this was assessed, further analysis could be undertaken on enhancer integration of inputs within a GRN. To assess the requirement of this enhancer, we decided to excise the -33 kb enhancer DNA within an Olig2 reporter cell line (Sagner & Briscoe 2017).

- **Aim 2 - To understand regulation of Olig2 in the ventral neural tube**

We hypothesised that if we removed *Olig2* expression via CRISPR-mediated deletion of enhancer DNA, we would be able to dissect the different enhancer regions that contribute to *Olig2* expression and also if particular binding sites were important for expression. Ultimately, we would be able to generate a single synthetic enhancer driving *Olig2* expression within its locus to analyse how multiple GRN inputs, from morphogen gradient, cross-repressive TFs and uniformly expressed TFs, drive target gene expression.

- **Aim 3 - To understand how disruption of enhancers at GRN genes affects patterning**

We wanted to assess whether we could relay our understanding of GRN enhancers *in vitro* to alterations in patterning outcome by the GRN downstream of Shh. Using the previously generated model (Cohen et al. 2014), we generated *in silico* predictions from alterations to the enhancer of *Olig2*. By targeting *Olig2* enhancer regions *in vivo* we were able to explore how disruption of cis-regulatory regions would affect patterning. Changes in the resulting neural tube pattern enabled us to explore, using mathematical modelling, how enhancers ensure robustness in GRN patterning and contribute to sharpness of boundaries between bordering domains of gene expression.
Chapter 2

Materials and Methods

2.1 CRISPR/Cas9 targeting

For CRISPR/Cas9 targeting ZiFit online tool (http://zifit.partners.org/) was used to scan the target sequence region and select single guide RNAs (sgRNA) that had the lowest number of potential off target sites.

2.1.1 Generating CRISPR/Cas9 plasmids

In order to clone sgRNAs into the CRISPR/Cas9 plasmid, oligos were ordered as follows 5'(Phosphate)CACC(Target sequence (sense)) and 5'(Phosphate)AAAC(Target sequence (anti-sense)). To clone the guide into the targeting plasmid, px459 V2.0 (Zhang et al. 2012), oligos were annealed by heating to 95°C in T4 ligation buffer then allowing to cool to room temperature. The px459 V2.0 was digested with BbsI which due to two neighbouring digestion sites results in overhangs that the phosphorylated sgRNA DNA sequence can clone easily into. This was carried out in a thermocycling reaction containing BbsI, ATP, DTT, T4 Ligase and Tango buffer incubated for 6 cycles of 37°C for 5 minutes, 21°C for 5 minutes. 1/10 of the reaction mixture was transformed into competent cells and plated to puromycin plates to obtain isolated colonies. Colonies were selected, grown up and plasmid DNA was extracted using Qiagen Mini Prep kit as per manufacturer’s instructions. DNA was then sequenced using U6 promoter sequencing primer to ensure sgRNA sequence had been inserted.
2.2. Tissue Culture Techniques

All tissue culture was performed in a Class II Laminar flow hood using aseptic technique. Unless otherwise stated all centrifugation steps are at 1,000 rpm (approx. 0.3 xg), for 4 minutes on a bench top centrifuge. Cells were incubated in tissue culture incubators set at 37°C with 5% CO₂.

2.2.1 MEFs and Feeders

Mouse Embryonic Fibroblasts (MEFs) were expanded and mitotically inactivated to used as feeders for culture of mouse ES cells. MEFs had previously been collected from e12.5 mouse embryos and frozen down in vials. One vial was thawed to 3 10cm plate in MEF medium: (Dulbecco’s Modified Eagle Medium (DMEM); 15% FBS; 1X Glutamine; 1X Penicillin/Streptomycin (all Gibco)). Once cells reached confluency, cells were dissociated using 0.25% Trypsin/EDTA (Gibco) and split 1:4. This was repeated twice more before confluent cells were mitotically inactivated using mitomycin C. Mitomyin C

Table 2.1: CRISPR guides

<table>
<thead>
<tr>
<th>Name</th>
<th>System</th>
<th>Deletion (bp)</th>
<th>Location (Chr16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O2e33 (62)</td>
<td>Cells</td>
<td>3270</td>
<td>91191287 to 91194556</td>
</tr>
<tr>
<td>O2e33 (52)</td>
<td>Cells</td>
<td>3269</td>
<td>91191291 to 91194559</td>
</tr>
<tr>
<td>O2e33 (60)</td>
<td>Cells</td>
<td>3272</td>
<td>91191288 to 91194559</td>
</tr>
<tr>
<td>O2e33 (68)</td>
<td>Cells</td>
<td>3270</td>
<td>91191287 to 91194556</td>
</tr>
<tr>
<td>O2e33 (23)</td>
<td>Cells</td>
<td>999, 1007</td>
<td>91192464 to 91193462</td>
</tr>
<tr>
<td>O2e33 (17)</td>
<td>Cells</td>
<td>1007</td>
<td>91192459 to 91193465</td>
</tr>
<tr>
<td>O2e33 (Sox2)</td>
<td>Cells</td>
<td>13</td>
<td>91192839 to 91192848</td>
</tr>
<tr>
<td>O2e33-75 (+75) (1-14)</td>
<td>Cells</td>
<td>1352</td>
<td>91300820 to 91302185</td>
</tr>
<tr>
<td>O2e75 (2-10)</td>
<td>Cells</td>
<td>1250</td>
<td>91300841 to 91302104</td>
</tr>
<tr>
<td>O2e33</td>
<td>Mice</td>
<td>3259</td>
<td>91191295 to 91194570</td>
</tr>
<tr>
<td>O2e33-75 (+75)</td>
<td>Mice</td>
<td>1250</td>
<td>91300841 to 91302104</td>
</tr>
<tr>
<td>O2e75</td>
<td>Mice</td>
<td>1252</td>
<td>91300842 to 91302107</td>
</tr>
</tbody>
</table>

Table 2.2: CRISPR deletions

<table>
<thead>
<tr>
<th>Name</th>
<th>System</th>
<th>Location (Chr16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O2e33</td>
<td>Cells</td>
<td>91191292 to 91191308</td>
</tr>
<tr>
<td>O2e33 (Right)</td>
<td>Cells</td>
<td>91194549 to 91194565</td>
</tr>
<tr>
<td>O2e33 (Left)</td>
<td>Cells</td>
<td>91192460 to 91192476</td>
</tr>
<tr>
<td>O2e33 (Right)</td>
<td>Cells</td>
<td>91193448 to 91193464</td>
</tr>
<tr>
<td>O2e33 (Left)</td>
<td>Cells</td>
<td>91192826 to 91192842</td>
</tr>
<tr>
<td>O2e33 (Right)</td>
<td>Cells</td>
<td>91300839 to 91300855</td>
</tr>
<tr>
<td>O2e33 (Left)</td>
<td>Cells</td>
<td>91302103 to 91302119</td>
</tr>
</tbody>
</table>
(Sigma) was dissolved in PBS (2 mg in 2ml) using a needle and syringe. Mitomycin solution was diluted in 18 mL of MEF medium. This stock solution was diluted 1:10 in MEF medium and was added to confluent 10cm plates of MEFs. Cells were incubated for 2.5-3 hours before mitomycin C containing medium was removed. Cells were washed with PBS and dissociated using 0.25% Trypsin/EDTA, centrifuged and resuspended in MEF medium to a concentration of 5x10^6 cells per mL. Cells were frozen in cryovials containing 2.5x10^6 cells with 0.5mL FBS + 20% DMSO (Sigma). Cells were frozen for further use (Feeders). When required, one vial of feeders was thawed and used to cover approximately 80cm² of tissue culture dishes (1x10 cm; 3x60mm; 1x multi-well plate). Vials of feeders were thawed by placing in a waterbath at 37°C. Once thawed, the cell suspension was transferred to a 15mL Falcon tube containing 5mL of prewarmed MEF medium. Following centrifugation, medium was removed and cells were resuspended. Cells were seeded into tissue culture dishes containing MEF medium. Prior to addition of cells, tissue culture dishes had been coated for at least 30 minutes at 37°C with 0.1% gelatin (Sigma). Cells were fed with MEF medium 24 hours after thawing. After 48 hours, feeders were sufficiently spread and ready to support ES cell growth.

2.2.2 Culture of ES cells

Mouse embryonic stem cell line (DOK1-1) containing a fluorescent reporter cotranslated with Olig2 (Olig2::T2A-mKate2) (Sagner et al. 2018) was used for all experiments. Mouse ES cells were maintained on feeders under pluripotent ES cell conditions. ES cells were thawed and placed in ES medium (KODMEM + 10% FBS, 1X Glutamine, 1X Pen/Strep, 0.1mM 2-mercaptoethanol) with added 1000 U/mL ESGRO Leukaemia Inhibitory Factor (Merck Millipore) (ES + LIF). The number of ES cells frozen per vial was dependent on the dish size they would be thawed onto, for example, 1.5x10^6 for subsequent thawing to a 60mm dish. ES cells were fed every 24 hours with new ES + LIF. 48 hours after thawing, ES cells were split. On the day of splitting, ES cells were fed with new ES + LIF at least two hours before. Cells were washed with PBS before addition of 0.05% Trypsin/EDTA for approximately
2.2. Tissue Culture Techniques  

2 minutes. Cells were collected from the tissue culture plate using ES medium and transferred to a 15 mL Falcon tube containing 8 mL ES medium. After centrifugation, cells were resuspended in 1 mL of ES medium and counted using Nexcelom Biocience Auto 2000 Cell Viability Counter. $0.5 \times 10^6$ cells were plated per 60mm tissue culture plate. ES cells were frozen in the same manner as Feeders.

2.2.3 Differentiation of ES cells to spinal cord and hindbrain progenitors

Cells were differentiated in accordance to a previously published protocol from the lab (Gouti et al. 2014). Before starting the differentiation, CellBIND dishes (Corning) were incubated with 0.1% gelatin at 37°C for 24 hours. In order to plate ES cells for differentiation, feeders were first removed. ES cells were dissociated using 0.05% Trypsin/EDTA and panned in ES medium on gelatin treated culture plates one size larger than their growth dish for example, cells harvested from a 60mm dish would be panned in a 10cm dish. After 15 minutes medium and cell solution was transferred to a new gelatinised plate for a further 15 minutes. During this process, the larger feeder cells would attach to the plate leaving mostly ES cells in the medium. ES cells were collected, centrifuged and ES medium was removed. Cell were resuspended and washed in 10mL PBS before a second centrifugation. Cells were resuspended in 1mL N2B27 (Advanced DMEM/F12 and Neurobasal medium (1: 1), supplemented with 1X N2, 1X B27, 1X Glutamine, 1X Pen/Strep, 0.1 mM 2-mercaptoethanol (all Gibco), 40 µg/mL BSA (Sigma) then counted. 50,000 cells were plated on 35mm CellBIND dishes (Corning) containing 1.5 mL N2B27 + 10 ng/mL FGF-2 (day (D) 0). For spinal cord differentiation, at D2 the media was replaced with N2B27 + 10 ng/mL FGF-2 + 5 µM CHIR99021 (Axon). For hindbrain differentiation, at D2 media was replaced with N2B27 + 10 ng/mL FGF-2. At D3 for both hindbrain and spinal cord differentiation, media was replaced with N2B27 + 100 nM RA (Sigma) and 500 nM SAG (Calbiochem). This was replaced every 24 hours for the remainder of the experiment.
2.2. Tissue Culture Techniques

2.2.4 Targeting CRISPR/Cas9 to mouse ES cells

To prepare cells for nucleofection of CRISPR/Cas9 targeting plasmids, ES cells were adapted to 2i+LIF conditions (Ying et al. 2008). Feeders were removed from ES cells though panning as described above and were plated on CellBIND dishes that had been incubated with 0.1% gelatin overnight at 37°C. Cells were cultured in 2i+LIF medium (N2B27 + CHIR99021 + PD184352 (Axon) + LIF). ES cells were passaged 3 times prior to nucleofection. 2x10^6 cells were resuspended in 100 µL mouse ES cell Nucleofector reagent (Lonza) with CRISPR/Cas9 plasmid DNA. A total of 2 µg of plasmid DNA was targeted to cells using Nucleofector II electroporator (Amava), program A-023 (mouse ES cells). Targeted cells were plated in 2i+LIF media across two gelatinised CellBIND 10 cm dishes (10% and 90%). 24 hours after seeding the cells were exposed to puromycin (1.5 ng/mL) in 2i+LIF for 48 hours before continued culture in 2i+LIF. Colonies were allowed to grow for a further 5 days before picking. Colonies were picked using a p2 pipette and dissociated in a round-bottomed 96-well plate in 20 µL 0.25% Trypsin/EDTA. A higher concentration of trypsin was needed to dissociate compact single colonies compared to normal dissociation of monolayer ES cells. Dissociated colonies were split equally between one well of a 96-well plate coated with feeders and a 48-well CellBIND plate coated in 0.1% gelatin in ES+LIF medium. Cells on feeders were grown to confluency then split across two 96-well plates coated with feeders. Once these two plates reached confluency, cells were dissociated using 0.05% Trypsin/EDTA and resuspended in ES+LIF + 40% FBS + 10% DMSO and stored at -80°C. Cells plated in the 48-well plate were grown to confluency then dissociated using 0.05% Trypsin/EDTA. Cells centrifuged, resuspended and washed in PBS then centrifuged again. PBS was removed and the cell pellet was stored at -20°C before DNA extraction. DNA was extracted using the PureLink Genomic DNA kit (Invitrogen).
2.3 Analysis of cells

2.3.1 Genotyping

DNA was extracted from cells using PureLink Genomic DNA kit (Invitrogen). Polymerase Chain Reaction (PCR) was carried out using KAPA Taq Ready Mix (Kapa Biosystems) in a T100 Thermocycler (Bio-Rad Laboratories). Genotyping primers were designed using Primer3 online platform. Primer annealing temperatures from Primer3 platform were used. Resulting PCR products were analysed by gel electrophoresis using agarose gel (1-2% in 1x TAE (20mM TRIS acetate, 1mM Na$_2$EDTA.2H$_2$O). For sequencing, PCR products were purified using QIAquick gel extraction kit (Qiagen) as per manufacturer’s protocol. Purified PCR product was sent for sequencing using Eurofins. Once received, sequences were analysed using LaserGene software.

2.3.2 Immunofluorescence

Samples were washed in PBS and fixed in 4% paraformaldehyde for 15 min at 4°C then washed in PBS then 0.1% Triton PBS (PBST). Samples were stored in 2 mL PBS at 4°C prior to antibody staining. Primary antibodies diluted in blocking solution (1% BSA in PBST) were applied overnight at 4°C. The following day, dishes were washed 3 x 5 minute in PBST before secondary antibodies diluted in PBST were added for 1 hour at room temperature. All secondary antibodies were raised in donkey and conjugated to either Alexa488, Alexa568, Alexa647 (Abcam). Secondary antibodies were removed with 3 x 5 min washes with PBST and one wash containing PBST and DAPI. PBST was removed and 30μL of Prolong Gold (Invitrogen) was added to the dishes which were then covered with circular glass coverslips (thickness 1.5). Images were collected on a Zeiss Imager.Z2 microscope fitted with an Apotome.2 structured illumination module using a 20x magnification lens. For structured illumination 5 phase images were acquired and a z-projection was made using 8 sections. Images were processed using FiJi software.

2.3.3 Flow Cytometry

Cells were dissociated using 0.05% Trypsin and collected in ES media. Cells were then washed in PBS and resuspended in PBS containing live-cell Cal-
2.4 Mouse Procedures

Cyanine Violet dye (Life Technologies) diluted as per manufacturer’s instructions. Cells were transferred to flow cytometry tube (Falcon) via a cell-strainer cap. Data was acquired using LSRII C (BD Biosciences) and FACS Diva software. Analysis was performed using FlowJo.

2.3.4 Quantitative PCR (qPCR)

Differentiated cells were washed twice with PBS prior to collection. Cells were lysed directly in the dish using 350 µL of RLT buffer and transferred to eppendorfs before being stored at -80°C. RNA was extracted using RNeasy Mini Kit (Qiagen) according to the manufacturer’s instructions. 1 µg of RNA was used for reverse transcription reaction using SuperScript III (Invitrogen). The resulting 20 µL of cDNA was diluted 1:12 in RNase-free water. qPCR was performed in 384-well plates in a total reaction volume of 10 µL containing 4.5 µL diluted cDNA, 0.5 µL 10 mM primers (forward and reverse) and 5 µL of Platinum SYBR Green qPCR SuperMix-UDG with ROX (Invitrogen). Plates were run using the standard program on the QuantStudio 5 Real-Time PCR system (ThermoFisher Scientific).

2.4 Mouse Procedures

Mouse strains containing the following alleles were used: Pax6(Sey) in strain background C57BL/6Jax and O2e33, O2e33-75 and O2e75 in strain background F1(B6xCBA). The O2e33 alleles were derived using zygote injection of CRISPR gRNA and Cas9 plasmids (see below). Embryos were transferred to psuedo-pregnant females and subsequent pups were genotyped. O2e33 lines were maintained as a heterozygous population. Embryos for analyses were collected from heterozygous crosses at the indicated time points following a mating, with the day of plug detection designated E0.5. Genotyping of mice was initially carried out in-house from DNA extracted from ear-biopsies using described genotyping protocol. A genotyping protocol was subsequently generated using Transnetyx and ear biopsies were sent away for genotyping. All animal procedures were carried out in accordance with the Animal (Scientific Procedures) Act 1986 under the Home Office project licence PPL80/2528 and PD415DD17.
<table>
<thead>
<tr>
<th>Target</th>
<th>Forward and Reverse Primer 5’ - 3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoxc4</td>
<td>AAGCAACCCTAGTCTACCCCTTG</td>
</tr>
<tr>
<td></td>
<td>CTCCTGTTATAATTTGAGGTACCCTG</td>
</tr>
<tr>
<td>Hoxc6</td>
<td>CAGGGAAAGGCAGGGATG</td>
</tr>
<tr>
<td></td>
<td>ATAGGCAGTGGAATTGAGG</td>
</tr>
<tr>
<td>Islet1</td>
<td>TATCAGGTGTTGACGGGATCAA</td>
</tr>
<tr>
<td></td>
<td>CTACACAGGCAGAAACACTCG</td>
</tr>
<tr>
<td>Lhx3</td>
<td>TTAAGCGCTTTGGAGACCAGA AG</td>
</tr>
<tr>
<td></td>
<td>TCCTGGATGCTGTGGGCTTGTGGC</td>
</tr>
<tr>
<td>Nkx2.2</td>
<td>CCGACAATGACAAGGGAGACC</td>
</tr>
<tr>
<td></td>
<td>GCTGGAGAGAGAGCAGCTCG</td>
</tr>
<tr>
<td>Olig1</td>
<td>CCACCACACACATCAACACTTAGG</td>
</tr>
<tr>
<td></td>
<td>ACGGATAAGGAGATAGCGGCG</td>
</tr>
<tr>
<td>Olig2</td>
<td>AGACCCGAGCCAACACCAG</td>
</tr>
<tr>
<td></td>
<td>AAGCTCTCGAATGATCCTTCCTTT</td>
</tr>
<tr>
<td>Phox2b</td>
<td>GATAAAGGACACCTTTGAGG</td>
</tr>
<tr>
<td></td>
<td>GTTTTGTTAGGACTGGGCG</td>
</tr>
<tr>
<td>Sox2</td>
<td>GGCAGCTACAGCATGAGCAGGAGC</td>
</tr>
<tr>
<td></td>
<td>CTGTTGATGAGTGTACTGCAGG</td>
</tr>
<tr>
<td>Tuj1</td>
<td>CAACACGAGGCCATGGCAGG</td>
</tr>
<tr>
<td></td>
<td>CTTGGGCCCCTGGGCGCTCCGA</td>
</tr>
</tbody>
</table>

Table 2.3: Primers for qPCR

2.5 Embryo processing

Embryos were harvested at defined timepoints and were dissected in cold PBS. Surrounding tissue was removed and yolk sacs from individual embryos were collected to use for genotyping. Embryos were transferred to 4% paraformaldehyde (PFA) in PBS and were fixed on ice. The time embryos were fixed for depended on age (Table 2.4). After fixation embryos were washed in cold PBS to remove any remaining fixative and stored at 4°C.

2.5.1 Genotyping embryos

Yolk sacs (e8.5-e11.5) or limb buds (e12.5 and over) collected during initial embryo processing were used to extract DNA. Alkaline lysis buffer (NaOH 10M (25 nM), EDTA 0.5M (0.2 mM)) was added to tissue and incubated at 92°C for 45 minutes. Once tubes had cooled slightly, an equal volume of Neutralising buffer (TRIS-HCl pH 8.0 1M (40 nM)) was added (for E8.5 30µl of each; E9.5 50µl of each and for later stages 100µl of each). 1µl of DNA solution was used for subsequent PCR analysis. As the genotyping primers
were either side of the excised region, in homozygotes and heterozygotes the resulting PCR products when analysed appeared the same as the larger, 3.3kb product would not often amplify. To overcome this, a second pair of genotyping primers, one outside the deleted region and one inside the deleted region, were used to assess whether a wild type allele was present.

2.5.2 Sectioning embryos

Embryos for sectioning were placed in cryopreservation solution of 30% sucrose in 0.12M phosphate buffer pH 7.2 (PB) overnight at 4°C. Embryos were then dissected into fragments for sectioning. Brachial fragments were obtained through cuts above and below the forelimb. Embryo fragments were transferred to 42°C gelatin solution (7.5% gelatin (Sigma), 15% sucrose (Sigma), 0.12M PB). Once equilibrated, embryo fragments were arranged in gelatin solution (7.5% gelatin, 30% sucrose in 0.12M PB) in plastic moulds positioned with anterior side down so sections achieved with be transverse. The arrangement of embryos and subsequent genotyping results were recorded in order to identify mutants and wild type litter mates in future analysis. Moulds were kept at room temperature until embryos were no longer at risk of moving. Moulds were incubated on ice for at least 30 min to solidify. Blocks were removed from moulds and trimmed using a scalpel. Trimmed blocks were placed in -30 to -40°C isopentane to freeze solid. Frozen blocks were stored at -80°C prior to sectioning. Embryo blocks were sectioned into 12µm sections that were collected on Superfrost™ glass slides (Thermo Scientific) using Zeiss Hyrax C 60R cryostat.

2.5.3 Immunohistochemistry on embryo sections

Slides from -80°C storage were placed at room temperature for 30min to dry. Gelatine was removed from the slides by 4 x 5 min washes in PBS at 42°C. Slides were then washed with PBS containing 0.1% Triton X-100 (PBST). To prevent solutions running off slides, a wax pen was used to draw around the slide. Sections were incubated in blocking solution (1% BSA in PBST) for 30min at room temperature inside a dark humidified box. This solution was then replaced with 300µL of blocking solution containing diluted antibodies.
(Table 2.5) before being incubated at 4°C overnight. The following day, slides were rinsed then washed 3 x 5min in PBST before secondary antibodies diluted in PBST were added for 1 hour at room temperature. All secondary antibodies were raised in donkey and conjugated to either Alexa488, Alexa568, Alexa647 (Abcam). Secondary antibodies were removed with 3 x 5 min washes with PBST and one wash containing PBST and DAPI. PBST was removed and 30μL of Prolong Gold (Invitrogen) was added to the slides which were then covered with glass coverslips (thickness 1.5). Clear nail varnish was used to seal slides to prevent coverslips moving.

2.5.4 Immunohistochemistry on whole embryos

Following fixation, embryos were stored in PBS at 4°C for no longer than 1 week before processing allowing time for genotyping results. For antibody staining, embryos were transferred to blocking solution (as above) for an hour at room temperature. Embryos were then incubated over night in blocking solution containing diluted primary antibodies (Table 2.5). The following day, primary antibodies were removed through 3 x 5min washes in PBST followed by a longer incubation of 2 hours. Embryos were then incubated in PBST containing secondary antibodies for 1 hour at room temperature protected from light. Secondary antibodies were removed and embryos were incubated with PBST containing DAPI for 15 minutes at room temperature. All secondary antibodies were raised in donkey and conjugated to either Alexa488, Alexa568, Alexa647 (Abcam). Embryos were then transferred to PBST before mounting.

Embryos were transferred to glass slides and excess PBST was removed. 4 dots of vaseline were placed around the embryo on the slide. A square glass coverslip (1.5 thickness) was then placed on the vaseline before being gently lowered to touch the embryo. 100% glycerol was added to the slide at the edge of the coverslip to fill the cavity.

2.5.5 Imaging embryos and sections

Embryo sections on slides were imaged using a Leica MSP5 confocal microscope with either 20x (Dry) or 40x (Oil) objectives. Z-stacks of sections were taken and compiled in FiJi (FiJi is Just ImageJ) software. Laser settings were
maintained for imaging sections on one slide in order to compare directly between mutant and wild-type embryos. Whole embryos were imaged using the same confocal microscope by acquiring z-stacks and also tile scans.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Fixation Time (Hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E8.5</td>
<td>0.3</td>
</tr>
<tr>
<td>E9.5</td>
<td>0.5</td>
</tr>
<tr>
<td>E10.5</td>
<td>1.0</td>
</tr>
<tr>
<td>E11.5</td>
<td>1.5</td>
</tr>
<tr>
<td>E12.5</td>
<td>3.0</td>
</tr>
<tr>
<td>E13.5</td>
<td>4.0</td>
</tr>
<tr>
<td>over E13.5</td>
<td>O/N</td>
</tr>
</tbody>
</table>

Table 2.4: Embryo fixation time in 4% PFA at 4°C

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Dilution</th>
<th>Host</th>
<th>Catalogue No.</th>
<th>Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chx10</td>
<td>1:100</td>
<td>Mouse</td>
<td>sc-3655195</td>
<td>Santa Cruz</td>
</tr>
<tr>
<td>Flag</td>
<td>1:2000</td>
<td>Rat</td>
<td>200474</td>
<td>Stratagene</td>
</tr>
<tr>
<td>HB9</td>
<td>1:100</td>
<td>Mouse</td>
<td>81.5C10</td>
<td>DSHB</td>
</tr>
<tr>
<td>Lhx3</td>
<td>1:1000</td>
<td>Rabbit</td>
<td>AB14555</td>
<td>Abcam</td>
</tr>
<tr>
<td>Islet1</td>
<td>1:1000</td>
<td>Goat</td>
<td>AF1837</td>
<td>R&amp;D</td>
</tr>
<tr>
<td>Nkx2.2</td>
<td>1:500</td>
<td>Mouse</td>
<td>74.5A5</td>
<td>BD Pharmigen</td>
</tr>
<tr>
<td>Olig2</td>
<td>1:8000</td>
<td>Guinea Pig</td>
<td>(B.Novitch)</td>
<td>(Novitch et al. 2001)</td>
</tr>
<tr>
<td>Olig2</td>
<td>1:1000</td>
<td>Rabbit</td>
<td>AB9610</td>
<td>Millipore</td>
</tr>
<tr>
<td>Pax6</td>
<td>1:1000</td>
<td>Rabbit</td>
<td>AB2237</td>
<td>Millipore</td>
</tr>
<tr>
<td>Sox2</td>
<td>1:500</td>
<td>Goat</td>
<td>AF2018</td>
<td>R&amp;D</td>
</tr>
<tr>
<td>Tuj1</td>
<td>1:1000</td>
<td>Mouse</td>
<td>MMS-435P</td>
<td>Covance</td>
</tr>
</tbody>
</table>

Table 2.5: Antibodies
Chapter 3

Regulation of Olig2 in vitro

3.1 Introduction

Transcription factors Pax6, Olig2, Nkx2.2 and Irx3 are the four nodes of a network used to generate a mathematical model that correctly recapitulates wild type ventral neural tube patterning in silico (Cohen et al. 2014). The weights of edges between nodes of the network were determined using parameter screens but have not been explored experimentally. To investigate how interactions between TFs in the ventral neural tube are coordinated, we need to examine regulatory regions directing expression of these TFs. The first aim of this project was to explore genetic regulation of transcription factor Olig2 expressed in the ventral neural tube. Olig2 stood out as an ideal starting point due to previous analyses on its regulatory input both in vitro and in vivo (Oosterveen et al. 2012, Peterson et al. 2012). In this chapter, I will discuss a previously well described enhancer of Olig2 located 33 kb 5’ to the transcriptional start site. I will explain how, using mouse ES cells, we are able to demonstrate the importance of this enhancer in driving Olig2 expression in a directed differentiation protocol that generates progenitors with ventral neural tube fate. Through this analysis we were able to uncover alternative regulatory elements that direct expression of Olig2 under different conditions.

3.1.1 Olig2 regulation in the ventral spinal cord

One of the key transcription factors within this ventral neural tube network is Olig2 expressed in progenitors in the pMN domain (Lu et al. 2000, Zhou et al. 2000, Takebayashi et al. 2000). These progenitors initially give rise
to motor neurons during neurogenesis before giving rise to oligodendrocytes during gliogenesis (Takebayashi et al. 2002). During initial patterning the pMN domain is located between Nkx2.2-expressing p3 progenitors and Pax6/Irx3-expressing p2 progenitors. Its expression is initiated in mice at E8.5 shortly after exposure to Sonic Hedgehog during ventral neural tube patterning (Jeong & McMahon 2005). We are able to recapitulate the onset of Olig2 expression in ventral neural tube progenitors using an in vitro differentiation protocol using mouse ES cells. This protocol generates a reproducible induction of Olig2 in approximately 70% of cells (Gouti et al. 2014, Sagner et al. 2018). This provides us with a system in which we can explore regulation of Olig2 in spinal cord progenitors in vitro.

In mice, the Olig2 gene is located on chromosome 16 which syntenically maps to chromosome 21 in humans. Its closest neighbour is paralog Olig-gene family member Olig1, approximately 40 kb away. Many regions of regulatory DNA in the vicinity to these two genes have been linked to one of, or both of these genes. Using BAC analysis, a 3.6kb enhancer, termed the K23 enhancer, sits approximately 10kb 3’ to the transcriptional start site of Olig2 was found to drive lacZ reporter expression in the pMN domain during motor neuron differentiation in mice (Sun et al. 2006). In addition, two further enhancer regions were identified using Gli and Sox2 ChIP-seq analyses located at -33 kb and -16 kb from the transcriptional start site (Wang et al. 2011, Peterson et al. 2012, Oosterveen et al. 2012). From these studies, locations of three separate Gli binding sites were identified in the -33 kb enhancer. In chick electroporations, inactivation of all three sites removed GFP reporter expression in the pMN domain compared to the full length enhancer (Oosterveen et al. 2012) and inactivation of two of these binding sites substantially reduced reporter expression in mice (Peterson et al. 2012). Although this approach indicates this enhancer is involved in regulating Olig2 expression in the ventral neural tube, it does not provide information on how these binding site mutations would affect the GRN governing patterned TF expression in the ventral neural tube as expression of the reporter does not in turn influence the expression of
Olig2. In order to understand how this enhancer regulates Olig2 within this network, it needs to be tested in a system where alterations would directly influence Olig2 levels.

### 3.1.2 Bioinformatic analyses of -33 kb enhancer of Olig2

To further assess the requirement of these enhancers in neural progenitor regulation of Olig2 expression we looked at other transcription factors bound to these enhancers as well as chromatin accessibility. Published ChIP-seq analysis has identified transcription factors that bind the -33 kb enhancer under neural progenitor conditions (Fig. 3.1) (Peterson et al. 2012, Kutejova et al. 2016, Nishi et al. 2015). Gli, Nkx2.2, Sox2 and Olig2 all bind to the -33 kb enhancer in in vitro conditions that generate neural progenitors with a ventral spinal cord identity. Equally, ATAC-seq accessibility data generated in the lab had shown this enhancer to be accessible in ventral neural progenitors in vitro (Fig. 3.1) (Metzis et al. 2018). Both the previously predicted -16 kb and K23 enhancers (Peterson et al. 2012, Sun et al. 2006) did not show accessibility in these data sets so we therefore focused our analyses on the -33 kb enhancer.

To predict exact binding locations for these transcription factors I used bioinformatic analysis of the DNA sequence to identify binding motifs. Using the online tool JASPAR, a TF Position Weight Matrix database, I analysed the sequence of the -33 kb enhancer (Sandelin 2004). I screened the -33 kb sequence (1kb) for binding sites for Sox2, Olig2, Nkx2.2 and Gli setting a relative profile score threshold of 80%. The two highest scoring sites for this enhancer were a Sox2 site and a Nkx2-8 site (a paralog of Nkx2.2), the previously located Gli binding sites were not detected in this analysis (Appendix A)(Oosterveen et al. 2012, Peterson et al. 2012). I also analysed the sequence for its conservation by directly comparing to human, this enhancer as well as the binding sites are very highly conserved (Appendix B).

To assess the potential activity of the -33 kb enhancer in our directed differentiation protocol for mouse ES cells to spinal cord neural progenitors I examined data collected from cells differentiated under these conditions (Gouti et al. 2014). In the lab, ATAC-seq analysis had been carried out on cells across dif-
ferent time-points and conditions in the differentiation protocol (Metzis et al. 2018). Genome-wide chromatin accessibility was determined for cells from time-points day 0 to day 5 of the in vitro differentiation protocol. From these data sets, the -33 kb enhancer was accessible from day 4 of the spinal cord differentiation protocol but not before this time point. This is 24 hours after the addition of RA and SAG and is consistent with the onset of Olig2 expression in cells between days 4 and 5 (Sagner et al. 2018). Accessibility was seen at day 5 for a region of 1kb (Fig.3.1), in a location consistent with previous analyses (Oosterveen et al. 2012, Peterson et al. 2012). This 1 kb sequence was contained within a highly conserved sequence 3 kb in length. Indeed, analyses of highly conserved sequences in the genome is how this and many other enhancers were originally detected (Oosterveen et al. 2012, Sandelin et al. 2004) The accessibility at this enhancer region in addition to known binding sites of relevant factors led us to believe that this enhancer played a role in Olig2 expression in our in vitro differentiation protocol.

Results

3.2 Generation of -33 kb enhancer knockout mouse ES cell lines

3.2.1 Targeting approach to knockout -33 kb enhancer

To test the requirement of this enhancer for Olig2 expression in our differentiation system, I decided to try and eliminate all activity at the -33 kb enhancer. To do this I used a paired CRISPR/Cas9 approach to excise the DNA region between these points (Zhang et al. 2012). We hypothesised that if this enhancer region drives Olig2 expression, excision of the DNA would affect levels of Olig2 in differentiated cells. CRISPR/Cas9 targeting has been used extensively to knock out genes by generating small mutations in coding regions with the hope of generating a premature stop codon and therefore removing expression of the gene through nonsense mediated decay of the mRNA (Ran et al. 2013). However, this approach is not appropriate for targeting an enhancer as a much larger deletion is needed to remove all regulatory DNA. Recent publications at the time had shown that using two CRISPR guides
3.2. Generation of -33 kb enhancer knockout mouse ES cell lines

Figure 3.1: The -33 kb enhancer of Olig2 is accessible during in vitro differentiation

Data tracks for ChIP-seq from the lab for Nkx2.2, Olig2 and Sox2 at the Olig2 locus, (Kutejova et al. 2016) as well as ATAC-seq data from cells directly differentiated to neuromesodermal progenitors at day 3 and spinal cord progenitors at days 4 and 5 (Metzis et al. 2018) as well as GERP conservation track. -33 kb enhancer is highlighted in blue. Sequence of visualised data is chr16:91,186,979-91,232,65. Data visualised using UCSC Genome Browser. Scale bar is 10kb.

targeted to either side of a region would occasionally result in the fragment between being excised (Zhang et al. 2015). To ensure that all regulatory DNA would be removed, I designed two pairs of CRISPR guides to target; the first one 1kb apart to target the specific sequence accessible in the ATAC-seq data and the second set was 3.3kb apart to include the surrounding highly conserved sequence. In order to efficiently assess any changes to Olig2 expression using flow cytometry, the CRISPR guides were targeted to a previously generated mouse ES cell line (DOK1-1) containing a reporter for Olig2 expression, Olig2:T2A:mKate2 (Sagner et al. 2018).
3.2. Generation of -33 kb enhancer knockout mouse ES cell lines

Figure 3.2: Locations of CRISPR guides to remove -33 kb enhancer -33 kb enhancer showing locations of 1kb CRISPR guides and 3kb CRISPR guides (orange) as well as predicted binding site locations for Gli (Oosterveen et al. 2012, Peterson et al. 2012), Sox2 and Nkx2.2. GERP mammalian alignment conservation track is shown. Scale bar is 1kb.

<table>
<thead>
<tr>
<th>Name</th>
<th>Deletion (bp)</th>
<th>Location (Chr16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O2e33 (62)</td>
<td>3270</td>
<td>91191287 to 91194556</td>
</tr>
<tr>
<td>O2e33 (52)</td>
<td>3269</td>
<td>91191291 to 91194559</td>
</tr>
<tr>
<td>O2e33 (60)</td>
<td>3272</td>
<td>91191288 to 91194559</td>
</tr>
<tr>
<td>O2e33 (68)</td>
<td>3270</td>
<td>91191287 to 91194556</td>
</tr>
<tr>
<td>O2e33 (23)</td>
<td>999, 1007</td>
<td>91192464 to 91193462</td>
</tr>
<tr>
<td>O2e33 (17)</td>
<td>1007</td>
<td>91192459 to 91193465</td>
</tr>
</tbody>
</table>

Table 3.1: In vitro -33 kb CRISPR deletions

3.2.2 Screening clones for knockout deletions

Following targeting of cells and the growing up of individual clones, cells were genotyped using PCR primers that bound either side of the enhancer region to amplify a 1.6kb fragment or a 3.5kb fragment (Fig. 3.2). Knockout clones generated a fragment of approximately 500bp. Fragments were subsequently sequenced to identify specific deletions. Mouse ES cell clones surviving the targeting and selection process were genotyped and of the 27 clones analysed for the 1kb pair, 4 (14%) had lost both alleles, 13 (48%) were heterozygous and the remaining clones contained point mutations (38%). For the 3.3kb targeting, of the 24 clones genotyped, 7 (29%) were homozygous, 6 (25%) heterozygous and 11 (46%) had point mutations or small deletions at the CRISPR sites.
3.2.3 Analysis of heterozygous and homozygous enhancer knockout cell lines

As the deletions were generated in a cell line containing a reporter for Olig2 expression (Sagner et al. 2018), we used flow cytometry to initially assess any changes to Olig2 expression in different clones. To confirm no clonal differences were present, we analysed two clones for the larger deletions (clones 62, 68) as well as a clone with the shorter deletion (clone 23) in parallel (Table 3.1). All cell lines were differentiated towards a ventral spinal cord fate using a protocol previously established in the lab (Gouti et al. 2014)(Methods 2.2.3). RA and SAG (a Smoothened agonist) are added to day 3 NMPs to differentiate cells to ventral neural fate. The percentage of Olig2 expressing cells in flow cytometry analysis was determined using gates set from mKate fluorescence of DOK1-1 cells only exposed to RA and not SAG from day 4. DOK1-1 cells show a strong induction of Olig2 at day 5 of the spinal cord neural differentiation protocol so we initially looked at this timepoint to compare between clones.

3.2.4 Olig2 expression at day 5

At day 5 of the differentiation, the homozygous knockout line (clone 62) had a near complete loss of cells expressing mKate2 (2%) compared to untargeted (DOK1-1) cells (63%) (Fig. 3.3A). Interestingly, the heterozygous line (clone 52) showed a 15% reduction in the number of mKate expressing cells compared to DOK1-1 (Fig. 3.3A). These flow cytometry data also indicated heterozygous cells do not have as high intensity reporter expression compared to DOK1-1 cells. As well as a reduction in the percentage of expressing cells, the level of expression in individual cells is lower (Fig. 3.3A). Comparison between different clones did not imply any clonal variations were present. The homozygous enhancer knockout cells consistently had very low percentage of positive cells (1-8%) compared to untargeted (63%), the heterozygous lines also had reduced Olig2 expression (40-48%) (Fig. 3.3B). All homozygous and heterozygous cell lines, whether they had a 1kb or 3 kb deletion, generated consistent phenotypes. From these data we decided to use the 1kb deletion cell line (clones 23 (homozygous) and 17 (heterozygous)) for further analysis termed from now on as O2e33.
3.2. Generation of -33 kb enhancer knockout mouse ES cell lines

3.2.4.1 Emergence of Olig2 at later time points

It has previously been demonstrated that day 5 of the mouse ES differentiation protocol is equivalent to day E9.5 spinal cord progenitors in mouse embryos (Gouti et al. 2014). From E9.5 onwards spinal cord neural progenitors begin to differentiate into neurons. We are able recapitulate this in vitro by extending the protocol (Sagner et al. 2018). To understand whether low levels of Olig2 in O2e33 cells at day 5 is a delay or is a loss of expression we differentiated both DOK1-1 and O2e33 cells to day 7 and analysed cells using flow cytometry (Fig. 3.4A). At day 6 and 7 the levels of Olig2 in the homozygous and heterozygous increased compared to day 5 but the percentage of Olig2 expressing cells and the level of Olig2 within individual cells never reached that of DOK1-1 controls (Fig. 3.4A). This behaviour of homozygous cells was observed across multiple repeats (Fig. 3.4B,C).

3.2.5 Expression of Olig1

As Olig2 sits very close to its neighbouring paralog Olig1, we hypothesised that lower levels of Olig2 would result in a change in Olig1 levels. Levels could increase due
3.2. Generation of -33 kb enhancer knockout mouse ES cell lines

Figure 3.4: Olig2 expression is delayed in O2e33 cells (A) Flow cytometry histograms indicate Olig2 expression is lost at day 5 in O2e33 cells but recovers from day 6 onwards. Heterozygous O2e33 cells have lower levels of mKate2 and a lower number of mKate2 positive cells. (B) Flow cytometry histograms at days 5 and 6 comparing O2e33 to DOK1-1 cells. (C) Graph showing percentage of mKate2 positive DOK1-1 and O2e33 cells from days 5-7 of spinal cord differentiation.

to Olig1 acting in a redundant manner with Olig2 as observed in Olig2−/− mice or, because of this enhancers location, levels of Olig1 would also be reduced because of its deletion (Takebayashi et al. 2002). Single cell RNA-seq data in the lab demonstrated that levels of Olig1 are very low in our differentiations (Sagner et al. 2018) but to make sure that removing the -33 kb enhancer did not affect regulation of this gene I analysed expression of Olig1. I used qPCR data to analyse Olig1 expression levels as there was no suitable antibody. qPCR data confirmed that Olig1 expression levels are very low and also that removal of the -33 kb enhancer did not affect levels of Olig1 in O2e33 cells (Fig. 3.5).

3.2.6 Efficiency of cell differentiation

One reason there may be reduced Olig2 positive neural progenitors could be due to a reduction of neural progenitors in the differentiation. It is possible that through the generation of the O2e33 cell line, the resulting cell line does not differentiate as efficiently to neural progenitors. To ensure that the reduced percentage of Olig2
3.2. Generation of -33 kb enhancer knockout mouse ES cell lines

Figure 3.5: *Olig1* expression *in vitro* in DOK1-1 and O2e33 cells differentiated to spinal cord neural progenitors

qPCR data for Olig1 in DOK1-1 and O2e33 cells at days 4-7 of differentiation. CT values normalised to Beta actin.

positive O2e33 cells was not due to reduced differentiation efficiency to neural progenitors. I examined markers expressed in neural progenitors. The identity of cells produced in the *in vitro* differentiation protocol is heterogeneous with both neural and mesodermal derivatives observed. Sox2 is a pan-neural transcription factor expressed throughout the central nervous system in progenitors (Ellis et al. 2004) and can be used to assess the global levels of Sox2 and therefore efficiency of neural differentiation. I used quantitative PCR as well as immunohistochemistry to check Sox2 levels. To compare qPCR data across multiple experiments I normalised the expression levels to 1 for each experiment. Sox2 levels remain consistent between O2e33 and DOK1-1 cells implying the loss of the enhancer does not affect the cells ability to differentiate to neural progenitors (Fig. 3.7A). qPCR for Olig2 confirmed the reduction of Olig2 mRNA levels in O2e33 cells (Fig. 3.7A). Immunofluorescence stainings confirmed consistent protein levels of Sox2 and dramatically reduced protein levels of Flag (a component of the mKate2 reporter construct) in O2e33 cells (Fig. 3.7B).

3.2.7 Alternative cell fates

Although *Olig2* expression is not removed entirely in the O2e33 cells, we hypothesise that the reduction in Olig2 may result in an increase in the number of p2 and p3 identity cells. In Olig2 knockout mice, the domains dorsal and ventral to the pMN domain expand into this area although the p2 domain expands ventrally to a much larger extent than the p3 domain expands dorsally (Zhou & Anderson 2002, Balaskas et al. 2012). The p2 and p3 domains give rise to interneuron populations.
3.3 Motor neuron differentiation in O2e33 cells

In the neural tube, Olig2-expressing cells in the pMN domain differentiate to neurons, the same is observed in this in vitro differentiation protocol (Gouti et al. 2014, Sagner et al. 2018). To examine whether Olig2-expressing O2e33 cells would still

V2 and V3 respectively. p3 progenitors express Nkx2.2 and require a high sustained level of Shh signalling. In the in vitro differentiation protocol the appearance of Nkx2.2 positive cells is seen from day 6 onwards. To examine whether there was an increase in Nkx2.2-expressing cells qPCR and immunofluorescent stainings were used. Analysis of Nkx2.2 levels in O2e33 cells compared to DOK1-1 showed there was no apparent increase of Nkx2.2 expressing cells. However, analysis of qPCR and immunofluorescent stainings do not provide an in depth analysis. To assess changes in Nkx2.2 in a more quantitative manner, flow cytometry of stained cells could be used to analyse Nkx2.2 expression.

3.3 Motor neuron differentiation in O2e33 cells

Figure 3.6: Neural progenitor mRNA and immunofluorescence analysis of O2e33 cells (A) Normalised qPCR levels for Sox2, Olig2 and Nkx2.2 from days 4-7 of spinal cord differentiation for DOK1-1 (black) and O2e33 (red). Points indicate the mean level, error bars are standard deviation for all qPCR data shown. (B) Immunofluorescence images of DOK1-1 and O2e33 cells at day 6. Images show DAPI stain, Sox2, Flag (within the mKate2 reporter) and Nkx2.2. Merge shows Flag in red and Nkx2.2 in cyan.
differentiate to motor neurons, I assayed markers of motor neuron differentiation. Newly born motor neurons express LIM homeodomain proteins Islet1 and Lhx3 (Ericson et al. 1992, Thaler et al. 2002). These two factors have been shown to function within a transcription factor complex that directs motor neuron fate (Thaler et al. 2002, Seo et al. 2015). Knockout mice for Islet1 and double knockout of Lhx3 and Lhx4 disrupts motor neuron generation (Ericson et al. 1992, Sharma et al. 1998). According to population qPCR data, at day 6 the levels of Islet1 in the O2e33 differentiated population compared to DOK1-1 is approximately half. There is even lower levels of Lhx3 implying that although the O2e33 cells are able to differentiate to motor neurons the numbers are severely reduced (Fig.3.7A,B). This is confirmed by a 50% reduction of Tuj1 expression in qPCR analysis, a marker for newly generated neurons. There does not appear to be a difference in the number of Lhx3+/Isl-cells suggesting a similar number of V2 interneurons are generated from both O2e33 and DOK1-1 cells (Fig.3.7B). However, it should be noted that V2 interneurons differentiate later than motor neurons and day 7 of this in vitro differentiation may not correctly capture differentiation of V2 neurons. It is clear from these data that there is a reduced number of motor neuron progenitors in O2e33 cells but the progenitors that exist retain the ability to differentiate to motor neurons. Once differentiated, motor neurons express different markers depending on where they innervate, i.e. which motor column they are present in (Jessell 2000). It would therefore be interesting to evaluate whether motor column identity is affected in in O2e33 motor neurons.
3.4 Bioinformatic analysis of the -33 kb enhancer

The bioinformatic JASPAR analysis of the -33 kb enhancer revealed a clear Sox2 site within the -33 kb enhancer. Removal of the entire enhancer sequence had a dramatic effect on Olig2 expression but did not assess the importance of the different binding sites present. The Gli binding sites had been tested previously and it had been shown that removal of all three Gli binding sites caused a loss of reporter expression in the ventral neural tube but the Sox2 site had not been analysed for its contribution to Olig2 expression (Oosterveen et al. 2013). To test the importance of this Sox2 site, I designed an approach to remove the binding site from the enhancer.

3.4.1 Targeting of the Sox2 binding site

In order to knockout the Sox2 site from the enhancer I designed one CRISPR guide to target the sequence. With this approach I hoped to generate a small deletion or insertion that would remove Sox2 binding at that site. I would need to ensure that any changes in sequence that occurred did not affect neighbouring binding sites.

I targeted the Sox2 CRISPR to DOK1-1 cells to generate clones. Out of the 15 clones I sequenced, 5 clones had identical 13bp deletions across the CRISPR site.
3.4. Bioinformatic analysis of the -33 kb enhancer

This 13bp deletion removed the Sox2 site entirely but did not affect the nearby Gli binding site or create a new binding site in its place (Table 3.2). I therefore selected two of these clones to compare to DOK1-1 and O2e33 cells.

<table>
<thead>
<tr>
<th>Name</th>
<th>Deletion (bp)</th>
<th>Location (Chr16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O2e33 (Sox2)</td>
<td>13</td>
<td>91192839 to 91192848</td>
</tr>
</tbody>
</table>

**Table 3.2:** Sox2 CRISPR deletion

![CRISPR deletion of Sox2 site in -33 kb enhancer](image)

**Figure 3.8:** CRISPR deletion of Sox2 site in -33 kb enhancer

Sequence showing the location of the CRISPR guide target site (green), Sox2 binding site (red) and Gli binding site (blue). Highlighted in the sequence (red box) is the 13bp removed by the CRISPR deletion. Flow cytometry plots show wild type (DOK1-1), O2e33 and two Sox2 clones (1-2) at days 5 and 6 of spinal cord differentiation.

3.4.2 Sox2 binding site is required for correct onset of Olig2 expression in vitro

To analyse any difference in Olig2 expression, Sox2 CRISPR clones were differentiated alongside DOK1-1 and O2e33 cells to day 6. Sox2 CRISPR clones display a reduction in the levels of Olig2 at day 5 of the differentiation compared to DOK1-1 but have higher levels of Olig2 than O2e33 cells at this stage of the differentiation. This reduction recovers by day 6 when Olig2 levels return to that of DOK1-1 cells. Across multiple repeats, it was clear that this delay of Olig2 expression could only be seen for experiments that had a slower onset of Olig2 in DOK1-1 cells (as in Fig. 3.8). In experiments that had a strong induction of Olig2 at day 5 the Sox2 clones would show very similar percentages. In order to quantify this delay, time-points between days 4 and 5 would need to be analysed. Equally, addition of lower levels of SAG from day 4 or delaying SAG addition until day 5 could be used to
3.5 Alternative regulation of Olig2 in vitro

Analyse onset of Olig2. Both of these conditions have been used in the lab to generate neural progenitors with a more dorsal fate at day 5 compared to addition of 500 nm of SAG at day 4 (Lorena Garcia Perez, Joaquina Delas Vives and Andreas Sagner, personal communication).

A subtle delay in Olig2 induction in cells lacking this binding site suggests that this Sox2 binding site is involved in the early initiation of Olig2 expression. Certainly Sox2 is believed to be a “pioneer factor” (Iwafuchi-Doi & Zaret 2016), a transcription factor that is able to bind to closed chromatin and recruit further transcription factors to initiate activity. It could be that without this optimum Sox2 binding site present, a slight delay occurs in Olig2 expression due to a delay in the accessibility of this enhancer. In addition, from the mathematical model of the ventral neural tube, changes in basal input alone to any of the nodes has effects on overall patterning. When basal input to Olig2 is reduced by half there is a dramatic reduction in the pMN domain size (Cohen et al. 2014). It is possible that due to the structure of the GRN, even a small reduction in basal input to a target gene would have a noticeable effect. It would be interesting to examine accessibility at this enhancer on day 4 with the Sox2 site removed (see Discussion).

3.5 Alternative regulation of Olig2 in vitro

Although removal of the -33 kb enhancer nearly completely removes Olig2 expression in differentiated O2e33 cells at day 5, the Olig2 expression from day 6 onwards suggested that an alternative enhancer was driving this expression. At day 6, neural progenitors are differentiating to neurons which could coincide with a change in accessibility around progenitor transcription factors. A second Olig2 enhancer becoming accessible at day 6 could therefore be the reason for Olig2 expression in a subset of O2e33 cells. One alternative idea is that a sequence of non-coding DNA that does not normally drive expression is becoming accessible at day 6 because of the removal of the -33 kb enhancer thus being a “shadow enhancer” which would not be accessible in wild-type cells (Barolo 2012).

3.5.1 Analysis of Olig2 locus for alternative enhancers

First we considered previously predicted enhancer regions within the Olig2 locus to check if they were accessible in our data sets. A 3.6 kb enhancer, termed the K23
enhancer, sits approximately 10kb 3’ to the transcriptional start site of Olig2 was found to drive lacZ reporter expression in the pMN domain during motor neuron differentiation in mice (Sun et al. 2006). However, both this enhancer and another at -17kb from the transcriptional start site (Peterson et al. 2012), were not accessible in our conditions at day 5 suggesting that these enhancers are not crucial for Olig2 expression at this time under these in vitro conditions. To analyse any changes in accessibility in the Olig2 locus at day 6, we carried out ATAC-seq on O2e33 and DOK1-1 day 6 differentiated cells. I used flow cytometry assisted sorting (FACS) to collect Olig2 (mKate2) positive cells from differentiations of both O2e33 and DOK1-1 cell lines at day 6.

3.5.2 Alternative Olig2 enhancer at +75 kb

From the ATAC-seq data, accessibility in the Olig2 locus appeared fairly similar between day 5 and day 6 for DOK1-1 and O2e33 cells. It was also comforting to note that no accessibility was observed for the -33 kb region, confirming that there was not a population of cells where a deletion had not occurred (Fig. 3.9). There was one change in accessibility when comparing the two cell lines, a clear increase in accessibility was observed at a peak 75kb downstream of the Olig2 transcriptional start site in the O2e33 cells compared to DOK1-1 cells. As no other changes were observed within reasonable distance of the Olig2 gene, this peak appeared to be a good candidate enhancer for driving Olig2 expression at day 6 in O2e33 cells.

The +75 kb enhancer is accessible at day 6 in DOK1-1 cells but the peak is much smaller compared to O2e33 (Fig. 3.9). The reduced accessibility in the DOK1-1 cells at the +75 kb enhancer is likely to be because when sorting Olig2 positive cells, those sorted will likely have Olig2 expression due to the -33 kb enhancer as well as the +75 kb either acting individually or together. Increased accessibility at the +75 kb enhancer in O2e33 cells does not inform us whether this enhancer is driving expression in more cells than it would in the DOK1-1 cell line as in both case the same number of Olig2 expressing cells were analysed.

I wanted to explore whether this enhancer region had previously been investigated or linked to Olig2 expression. The enhancer sequence aligned with a previously published report investigating transcription factors involved in myelination in the spinal cord (Fulton et al. 2011). They had shown this +75 kb region was able to
3.5. Alternative regulation of Olig2 in vitro

drive LacZ reporter expression in the spinal cord in post-natal days 5 and 10 and mature mice and had linked it with Olig1 through proximity. Due to the low levels of Olig1 expression in our in vitro differentiation and no change in Olig1 levels was observed between DOK1-1 and O2e33 cells we concluded that +75 kb enhancer contributes mainly to Olig2 expression under these conditions. Analysis of previously published ChIP-seq data displayed peaks for Olig2 and Sox2 and Gli at the +75 kb enhancer (Peterson et al. 2012, Kutejova et al. 2016). The +75 kb enhancer is also at a location with high GERP conservation values. The combination of these data encouraged me to explore its importance within our system further.

Figure 3.9: ATAC-seq data for Olig2 locus in showing predicted enhancer at +75 kb

ATAC-seq data tracks of cells from different time points in the differentiation protocol. Day 3 NMP (green), Day 4 (dark blue), Day 5 (light blue), Day 6 mKate2 positive DOK1-1 cells (red) and Day 6 mKate2 positive O2e33 cells (pink). Highlighted are the -33 kb enhancer and +75 kb enhancer neither of which are accessible at day 3. The -33 kb enhancer is accessible in all populations apart from in O2e33 cells where this region has been removed. The +75 kb enhancer is increasing accessible over time and is most accessible in O2e33 mKate2 positive cells. Tracks are of the Olig2 locus chr16:91190706-91305745, mm9 assembly. Scale bar is 50kb. All ATAC-seq data in this chapter was collected and assembled by Vicki Metzis (Metzis et al. 2018)
3.5.3 Accessibility of +75 kb enhancer across differentiation protocol

I first examined other ATAC-seq data tracks previously generated in the lab for different time-points and conditions in the differentiation protocol to assess the +75 kb behaviour throughout the differentiation. When comparing to spinal cord day 4 and 5 a peak at +75 kb could be seen but not to the same extent as in the day 6 cells. It was also clear that neither the +75 or the -33 enhancers are accessible at day 3 suggesting that these are neural enhancers responding to RA and SAG addition at day 4.

I wanted to assess whether the accessibility of this enhancer changed depending on the anterior-posterior identity of the cells. The neural progenitors generated throughout experiments so far have all had the same A-P identity of brachial spinal cord, according to Hox gene expression from single cell seq analysis (Sagner et al. 2018). However, in the embryo at E9.5-E10.5 Olig2 is expressed in the neural tube pMN domain in the hindbrain and spinal cord as well as areas of expression in the forebrain. ATAC-seq data generated in the lab had shown protocols generating progenitors from different A-P positions in the nervous system have different accessibility profiles (Metzis et al. 2018) so I used these data to examine accessibility at the +75 kb enhancer across different anterior-posterior protocols.

As Olig2 is not expressed when cells are differentiated using a protocol that mimics forebrain, where neural progenitors are exposed to retinoic acid and not SAG (Gouti et al. 2014, Metzis et al. 2018), I focused on the hindbrain and spinal cord conditions. Remarkably, the +75 kb enhancer shows increased accessibility at day 4 and 5 in hindbrain conditions compared to days 4-6 in spinal cord conditions. The -33 kb enhancer appears equally accessible in both hindbrain and spinal cord conditions. These data imply that expression of Olig2 has different regulation depending on anterior-posterior position within the embryo, even if it is expressed in the same tissue (the neural tube) in the same pattern. I wanted explore Olig2 expression in O2e33 cells differentiated under hindbrain conditions, as the +75 kb enhancer shows more accessibility in these conditions from day 4 and therefore may play distinct roles in the regulation of Olig2 in the hindbrain.
3.5. Alternative regulation of Olig2 in vitro

Figure 3.10: ATAC-seq data for Olig2 locus in hindbrain and spinal cord differentiation conditions

-33 kb enhancer is accessible under both hindbrain and spinal cord conditions at day 4-6 in wild-type cells. Increased accessibility is seen for the +75 kb enhancer under hindbrain conditions compared to spinal cord conditions. In Olig2 positive O2e33 cells there is increased accessibility at the +75 kb enhancer. Tracks are of the Olig2 locus chr16:91190706-91305745, mm9 assembly. Scale bar is 50kb.

3.5.4 Differential usage of Olig2 enhancers along the A-P axis

To assess whether Olig2 expression was affected by hindbrain and spinal cord conditions in O2e33 and DOK1-1 cells I differentiated both cell lines under spinal cord and hindbrain conditions in parallel. To confirm cells were of the correct AP identity I used qPCR to analyse mRNA levels for Phox2b, Hoxc6 and Hoxc4 (Fig. 3.11). At day 5, DOK1-1 hindbrain cells had lower levels of mKate2 compared to spinal cord. However, there was a slight increase in mKate2 positive cells in O2e33 hindbrain cells compared to spinal cord. A much clearer phenotype was observed at days 6 and 7 where O2e33 hindbrain cells had a higher percentage of mKate2 cells (33.7% and 62.7%) compared to spinal cord (6.45% and 18.0% respectively). The O2e33
3.5. Alternative regulation of Olig2 in vitro

Figure 3.11:
Quantitative PCR data to check anterior posterior markers of hindbrain and spinal cord differentiation. Values are calculated using Beta actin qPCR data. NB these data are from a single experiment.

hindbrain cells show a strong recovery of Olig2 expression compared to spinal cord conditions but it should be noted that these cells never reach the same fluorescence intensity of DOK1-1. Equally the overall percentage of cells expressing Olig2 in O2e33 cells reach that of DOK1-1 cells. It is clear that under hindbrain conditions the -33 kb enhancer is still required for correct Olig2 expression in cells.

Figure 3.12: Flow cytometry comparing O2e33 and DOK1-1 in hindbrain and spinal cord conditions
3.5. Alternative regulation of Olig2 in vitro

Table 3.3: In vitro +75 kb CRISPR deletions

<table>
<thead>
<tr>
<th>Name</th>
<th>Deletion (bp)</th>
<th>Location (Chr16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O2e33-75 (+75) (1-14)</td>
<td>1352</td>
<td>91300820 to 91302185</td>
</tr>
<tr>
<td>O2e75 (2-10)</td>
<td>1250</td>
<td>91300841 to 91302104</td>
</tr>
</tbody>
</table>

3.5.5 Generation of O2e75 and O2e33-75 cells

As the +75 kb enhancer was the only observed difference in the accessibility at the Olig2 locus when comparing O2e33 and DOK1-1 cells I wanted to assess whether removal of the +75 kb enhancer would completely remove Olig2 expression in our system. Due to the successful approach in excising the -33 kb enhancer from DOK1-1 cells I decided to use the same approach to remove the +75 kb enhancer (Appendix D). During this targeting process, I also took the opportunity to remove the +75 kb enhancer on its own in DOK1-1 cells to see if it would have any affect compared to removal of the -33 kb enhancer.

I designed a pair of CRISPR guides flanking the +75 kb enhancer approximately 1.2kb apart. I targeted both DOK1-1 cells and O2e33 cells and genotyped with the same approach as before by using primers flanking the enhancer region. For each targeting I screened 15 clones and from the PCR had 5 clones that appeared positive for the deletion that I went on to sequence. Of these only two had clear homozygous deletions of the enhancer region. Within the DOK1-1 background, clone 2-10 had a 1250bp deletion and within the O2e33 background, clone 1-14 had a 1352bp deletion (Table 3.3). It was these two clones that I used for the following analyses referred to as O2e75 and O2e33-75 respectively.

3.5.6 Analysis of O2e75 and O2e33-75 cells

To compare Olig2 expression between cells lines with different deletions, DOK1-1, O2e33, O2e75 and O2e33-75 lines were differentiated in parallel under spinal cord conditions and analysed using flow cytometry to detect mKate2 fluorescence. Dramatically, O2e33-75 cells had no expression of mKate2 across all 3 days analysed (Fig. 3.13). Although ATAC-seq analysis points to other areas of accessibility in the Olig2 locus at these time points, these data imply that deletion of both -33 kb and +75 kb enhancers from cells is able to completely remove Olig2 expression in this system under spinal cord conditions.

The O2e75 cells also had a reduced level of mKate2 fluorescence but not to the
3.5. Alternative regulation of Olig2 in vitro

same extent as O2e33 cells (Fig. 3.13). From these data, cells in spinal cord conditions require the +75 kb enhancer for correct expression but not to the same extent as the -33 kb enhancer. As there are different inductions of Olig2 depending on whether the cells are put through a spinal cord or hindbrain differentiation conditions (Fig. 3.13, I wanted to assess whether this would also apply for the O2e33-75 and O2e75 cells. To test this, I differentiated the four lines in parallel under both spinal cord and hindbrain differentiation protocols.

![Flow cytometry analysis of mKate2 fluorescence levels comparing DOK1-1 (black), O2e33 (red), O2e75 (blue) and O2e33-75 (orange) differentiated under spinal cord conditions. Dashed line at highlights extend of fluorescence intensity in DOK1-1 cells compared to other lines.](image)

**Figure 3.13:** Flow cytometry analysis of mKate2 fluorescence levels comparing DOK1-1 (black), O2e33 (red), O2e75 (blue) and O2e33-75 (orange) differentiated under spinal cord conditions. Dashed line at highlights extend of fluorescence intensity in DOK1-1 cells compared to other lines.

When differentiated under hindbrain conditions, O2e75 cells display a more severe phenotype compared to O2e75 cells differentiated under spinal cord conditions (Fig. 3.14). As the +75 kb enhancer displays more accessibility under hindbrain conditions (Fig. 3.10), these data support the hypothesis that this enhancer is active and required for neural differentiation under hindbrain conditions.

In spinal cord conditions, removal of the +75 kb enhancer does not display the same reduction observed in hindbrain conditions suggesting that in spinal cord conditions, the -33 kb enhancer contributes more to the expression of Olig2. The O2e33 cells under hindbrain conditions have more Olig2 induction over the 3 days compared to the spinal cord conditions confirming that the +75 kb enhancer contributes more to expression in hindbrain conditions.

O2e33-75 cells, that have deletion of both enhancers, do not have any expression when differentiated under hindbrain conditions (Fig. 3.14). It appears that removing both of these enhancers has removed possibility of generating Olig2 expressing motor
neuron progenitors within these in vitro protocols. It would be interesting to test whether these cells could be differentiated into cells expressing Olig2 with a forebrain identity (Discussion) and to assess the identity of these cells. Analysis of expression of ventral neural tube genes would provide insight into the identity of these cells.

From these data it should be noted that none of the knockouts achieve similar mKate2 intensity or percentage of expression cells to DOK1-1 cells indicating that both enhancers contribute to Olig2 expression in this system (Fig. 3.14).

![Flow cytometry analysis comparing mKate2 fluorescence in DOK1-1, O2e33, O2e75 and O2e33-75 in spinal cord and hindbrain conditions from days 5-7 of differentiation.](image)

**Figure 3.14:** Flow cytometry analysis comparing mKate2 fluorescence in DOK1-1, O2e33, O2e75 and O2e33-75 in spinal cord and hindbrain conditions from days 5-7 of differentiation.
Chapter 4

-33 kb enhancer deletion *in vivo*

The *in vitro* knockout of the Olig2 -33 kb enhancer demonstrated the importance of this regulatory region for correct expression of Olig2 in mouse ES cells differentiated to ventral neural progenitors. This prompted me to test whether removal of this enhancer would have an effect on Olig2 expression *in vivo* in the ventral neural tube. We were unsure to what extent a phenotype would be evident in mice, if at all, as many previous studies had shown removal of enhancer regions *in vivo* tended to not yield a phenotype, even when multiple regions were deleted.

4.1 O2e33 Mouse line generation

To test whether this enhancer was required for correct Olig2 expression in the ventral neural tube we decided to use the same approach as in cells. The same 3kb CRISPR guides within the pX459 V2.0 plasmid were injected into to fertilised ova before being transferred to pseudo-pregnant females (Chapter 3). The Genetic Manipulation Service (GeMS) at the Francis Crick Institute carried out this protocol for the project. Pups were ear clipped following weaning and genotyped as in cells by using primer pairs outside the deleted region as previously described (Chapter 3). In the initial round of targeting 2/42 pups (5%) contained a deletion at the -33 kb enhancer. Both of these mice were crossed to wild-type mice but in only one of the mice was the deletion transmitted to the germline. This F0 mouse was mosaic for two different deletions, 3279 bp deletion, with and without a 3 bp insertion at genome position chr16:91191294-91194553 (mm9). F1s from this female contained one of the two deletions generating two separate mouse lines with these alleles. These mouse lines were maintained as a heterozygous lines and staged embryos were collected following timed matings of heterozygous adult mice.
4.2 O2e33 patterning phenotype

To initially assess O2e33 embryos for any differences in *Olig2* expression, I collected brachial spinal cord sections of E9.5 embryos, a stage equivalent to when we see a reduction of *Olig2* *in vitro*. I also stained for ventral tube GRN transcription factors, Nkx2.2 and Pax6, to assess if there were any differences in patterning linked with the removal of the -33 kb enhancer. Within litters of embryos from heterozygous crosses, homozygous, heterozygous and wild-type embryos were observed at Mendelian ratios implying that this enhancer deletion did not affect embryonic survival at this stage. I did not observe any differences between heterozygous mice and wild type littermates and therefore analysed only knockout and wild type embryos.

At E9.5 O2e33 embryos had a smaller pMN domain and a ventral expansion of Pax6 expression compared to stage-matched wild-type littermates (Fig. 4.1). As well as a reduction in size due to the dorsal boundary shift, the ventral boundary also
was altered in O2e33 embryos. The ventral boundary appeared shifted due to an expansion of Nkx2.2 however this shift was not clear as this boundary was no longer clearly defined. This loss of sharpness between the domains was not due to cells expressing both Nkx2.2 and Olig2, as seen in human ventral neural tube patterning (Marklund et al. 2014), but rather Nkx2.2 positive cells intermingled with Olig2 positive cells. It also appeared from the images that the overall levels of Olig2 in the O2e33 embryos was lower than in the wild type consistent with in vitro results.

![Figure 4.2: O2e33 embryos have reduced Olig2 expression](image)

**Figure 4.2: O2e33 embryos have reduced Olig2 expression**

**A** Brachial sections of E9.5 wild-type and O2e33 embryos stained for Olig2. Scale bar is 50 μm.

**B** Normalised Olig2 expression extracted from images using method in (Chapter 5). Data analysed and assembled by Edgar Herrera Delgado. Figure adapted from (Exelby et al. 2019).

To explore why there was a difference in boundary sharpness I analysed any changes in expression of other transcription factors in the GRN. In the neural tube, Pax6 is initially expressed throughout the dorsal ventral axis before Shh signalling induces Olig2 then Nkx2.2 (Dessaud et al. 2008). Expression of these transcription factors represses Pax6 expression in the p3 and pMN domains. However, it is known that although Olig2 represses Pax6, Pax6 does not in turn repress Olig2 (Briscoe et al. 2000, Novitch et al. 2001). This results in the gradient of Pax6 expression becoming a step change of expression with Pax6 still expressed in the pMN domain but at a lower level that in the p2 domain (Ericson et al. 1997). When we compare the Pax6 staining between wild type and O2e33 embryos (Fig. 4.1), we can see a clear boundary between high and low Pax6 expression at the dorsal boundary of the pMN domain. In comparison, in O2e33 embryos, Pax6 expression remains as gradient until reaching the p3/pMN boundary. This expansion of Pax6, even though
4.3 Onset of Olig2 in O2e33 embryos

Olig2 cells remain, suggests that the levels of Olig2 are not strong enough to repress Pax6 in the pMN domain to the same extent observed in wild type. Indeed, in the mathematical modelling of the ventral neural tube, reduction of the basal input to Olig2 significantly decreased the pMN domain with a ventral expansion of Pax6 but minimal change in pMN/p3 boundary (Cohen et al. 2014). However, this expansion of Pax6 expression into the pMN domain does not explain why the pMN/p3 boundary is no longer sharp and this cannot be explored in the mathematical model as it is deterministic.

In order to quantify levels of Olig2 from images I used a Fiji quantification method to extract intensity data from images (Chapter 5). Quantification from the images showed that Olig2 protein level in O2e33 embryo sections is on average lower than in wild type due to an increased number of cells having a lower level of Olig2 (Fig. 4.2B). It appears that in the embryo, although the average expression levels are lower, certain cells are able to express Olig2 to the same level observed in the wild type (Fig. 4.2B). This is not what we observe in vitro where average expression level is shifted but O2e33 cells are not able to reach same mKate2 intensity as untargeted (Chapter 3). The distribution of Olig2 expression levels in individual cells has a smaller spread in wild type compared to mutants which can also be appreciated visually from the images of the neural tube sections (Fig. 4.2A).

4.3 Onset of Olig2 in O2e33 embryos

The smaller pMN domain and lower levels of Olig2 raised the question of whether the onset of Olig2 expression was delayed in vivo in line with the delayed onset in O2e33 cells. To assay the onset of Olig2 expression in vivo I collected both wild-type and O2e33 embryos at E8.5. In order to pinpoint the time of onset I counted the number of somites of each embryo and to define exactly where in the embryos the Olig2 expression was, collected embryos were stained and imaged whole. Previously published data showed onset of Olig2 expression in the ventral neural tube at 6 somites so I initially started analysis of embryos around this stage (Jeong & McMahon 2005).

At 7 somites, wild-type embryos showed clear expression of Olig2 in the neural tube consistent with previous reports (Fig. 4.3). However, in O2e33 embryos no Olig2 was observed at 7 somites. The exact time point of onset in the wild type appears
4.3. Onset of Olig2 in O2e33 embryos

Figure 4.3: *Olig2* expression occurs later in O2e33 mutants
Whole mount images of wild type and O2e33 embryos each with 7 somites stained for DAPI and Olig2. Zoomed area shows no *Olig2* expression in O2e33 embryo at this stage.

Figure 4.4: Onset of *Olig2* expression is delayed in O2e33 mutants
Whole mount images of wild type and O2e33 embryos stained for DAPI and Olig2. Both images show the earliest observed *Olig2* expression, in wild type this is at 5 somites and in O2e33 embryo this is 8 somites. Zoomed area shows the initial *Olig2* expression in O2e33 embryo at this stage. Scale bar is 150µm.

to be at 5 somites and for the O2e33 embryos at 8 somites (Fig. 4.4). The location of the Olig2 expressing cells appears consistent between embryos, the anterior limit is just anterior to the first pair of somites. The 3 pairs of somites difference between the onset links to approximately a 6 hour time delay in onset of Olig2 in the neural tube (Hubaud & Pourquié 2014). Although we see a difference in Olig2 onset in ESC derived neural progenitors *in vitro*, we have not quantified this difference. To do this cells would need to be collected between 4 and 5 days. These embryos were also stained for Nkx2.2 but expression could not be seen for the wild type or O2e33 embryos at this stage consistent with previous reports of expression onset around 8
4.4 Motor neuron differentiation in O2e33 embryos

To explore how a reduction in pMN domain size and the subtle phenotype of loss of sharpness would manifest as development proceeded I analysed embryos at later stages. In the in vitro analysis, although O2e33 cells they have very little Olig2 expression at day 5 they do have expression from day 6 onwards. As E9.5 is considered an equivalent stage to day 5 in the in vitro differentiation, it is possible that after E9.5 Olig2 expression could recover to wild type levels in embryos. At E10.5, the pMN domain is decreasing in size due to Olig2 positive cells differentiating to neurons and migrating out of the pMN domain to the ventral horns (Kicheva et al. 2014). At this developmental time, instead of recovering levels of Olig2 in the pMN domain, O2e33 embryos continue to display a smaller domain than wild-type embryos (Fig. 4.5). The Pax6 and Nkx2.2 domains remain expanded in O2e33 embryos with a consistent boundary between Pax6 (high) and Nkx2.2 expressing cells. At this stage, there continued to be intermingling between Olig2 and Nkx2.2 positive cells although in some cases this could not be seen as a boundary phenotype because the pMN domain consisted of so few cells.

![Image of Olig2 expression in wildtype and O2e33 embryos](image)

**Figure 4.5: Olig2 expression does not recover in vivo** Brachial sections of E10.5 embryos stained for Pax6, Nkx2.2 and Olig2. Scale bar is 50µm.

The pMN domain undergoes a large expansion between E9.0 and E9.5. This expansion generates a large progenitor pool that will then differentiate into motor somites (Jeong & McMahon 2005).
neurons of the spinal cord (Kicheva et al. 2014). As the O2e33 pMN domain is not able to reach wild type size it is possible that this will effect the number of motor neurons generated. To assess motor neuron differentiation in O2e33 embryos I examined spinal cord sections for early markers of motor neuron differentiation. Isl1 and HB9 two of the initial markers expressed in post-mitotic motor neurons in the spinal cord. At E9.5 there are a few Isl1 positive cells in O2e33 embryos and these do not co-express HB9 unlike in the wild type. At E10.5 it can also be observed that the ventral horns are reduced in size, although to count the number of motor neurons, HB9/Isl1 staining would need to be done at this stage as well. From these data, motor neuron differentiation is severely reduced in O2e33 embryos, consistent with \textit{in vitro} observations.

To assess how this early reduction of MNs would manifest at later stages, I analysed E12.5 embryo sections for Isl1 expression. In O2e33 embryos at this stage there are reduced Isl1+ cells in the ventral half of the neural tube implying a reduction in motor neurons. However, to confirm this, different factors involved in motor column divergence would need to be analysed as Isl1 is not expressed in all MNs (Jessell 2000).
4.4. Motor neuron differentiation in O2e33 embryos

Figure 4.6: Reduced number of motor neurons in O2e33 embryos at E9.5

Brachial sections of E9.5 wild type and O2e33 embryos. Sections stained for Sox2 (green) labelling neural progenitors, Isl1 (red) and HB9 (blue) labelling motor neurons. O2e33 embryos have fewer Isl1 expressing cells and no HB9 expressing cells. Scale bar is 50µm.
4.4. Motor neuron differentiation in O2e33 embryos

Figure 4.7: Reduced number of motor neurons in O2e33 embryos at E13.5
Brachial sections of E13.5 wild type and O2e33 embryos. Sections stained for Islet1 and Olig2. O2e33 embryos have reduced Isl1 staining in the ventral half of the neural tube and have no Olig2 staining at this stage. Scale bar is 100µm.
4.5 Oligodendrocytes in O2e33 embryos

After the patterning of neural progenitors at stages E8.5 to E10.5 in the neural tube, a period of neurogenesis occurs. Neural progenitors from each of the different domains begin to differentiate and migrate outwards to the periphery of the neural tube. The spinal cord neurogenesis period lasts until approximately E13.5, from which point on, gliogenesis occurs (Deneen et al. 2006). At this point, remaining Olig2 positive cells in the pMN domain divide and re-expand the Olig2 positive progenitor domain although this is a matter of debate as some lines of evidence suggest cells switch fate to express Olig2 in the ventral neural tube (Richardson et al. 2006). Equally, there is evidence of dorsally derived oligodendrocytes (Cai et al. 2005, Vallstedt et al. 2005). These Olig2 positive progenitors from this point are oligodendrocyte precursor cells (OPCs) that will divide and differentiate into oligodendrocytes. When OPCs differentiate into oligodendrocytes, they migrate out of the progenitor domain and begin to populate the developing spinal cord, initially ventrally before populating more dorsal regions (Zhou & Anderson 2002, Lu et al. 2002, Takebayashi et al. 2002).

In the O2e33 embryos at E10.5 there are very few Olig2 positive cells in the pMN domain. This prompted me to examine whether the progenitor pool would repopulated to give rise to oligodendrocytes in these embryos. I collected embryos at E12.5, a point at which the pMN domain begins to give rise to oligodendrocytes (Takebayashi et al. 2002), to compare the wild type and O2e33 embryos. At E12.5 there are very few cells expressing Olig2 in O2e33 embryos compared to wild type where a clear domain of Olig2 expression can be seen as well as a few Olig2 positive cells beginning to migrate throughout the neural tube. In order to discern whether oligodendrocytes develop I needed to look later in development.

At E14.5, there was a mixed phenotype for O2e33 embryos. Of three homozygous embryos, two had Olig2 expression at E14.5 compared to one that had none (Fig. 4.8). The O2e33 embryos that had Olig2 expression appeared the same as wild type. One possibility for this variability in phenotype may be due to the penetrance of mutation (Discussion). There is an Olig2 positive population of cells centrally in progenitor cells as well as migrating Olig2 positive cells in the ventral half with some in the dorsal half. In order to investigate further, more embryos would need
4.5. Oligodendrocytes in O2e33 embryos

to be collected and analysed at this stage. At E16.5, there appears to be reduced number of Olig2-positive cells within the spinal cord of O2e33 embryos compared to wild type (Fig. 4.9). Again to confirm that this reduction is seen for all embryos, more would need to be collected and analysed. An alternative method to explore oligodendrocyte differentiation would be to stain for OPC markers which in addition to Olig2 include Sox10. Although this antibody failed to work on embryo sections, others have had success using it on in vitro cells (Andreas Sagner, personal communication). It would be interesting to assess oligodendrocyte differentiation efficiency in O2e33 cells in vitro and these protocols are now beginning to be established in the lab.

<table>
<thead>
<tr>
<th>Wild type</th>
<th>O2e33</th>
<th>O2e33</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Wild type" /></td>
<td><img src="image2.png" alt="O2e33" /></td>
<td><img src="image3.png" alt="O2e33" /></td>
</tr>
</tbody>
</table>

Figure 4.8: Olig2 expression at E14.5

Brachial sections of wild type and O2e33 embryos stained for DAPI and Olig2. Scale bar is 100µm.
4.6 Anterior-Posterior Olig2 patterning

I had previously shown that the -33 enhancer responded differently in vitro depending on whether cells were differentiated towards a caudal hindbrain identity or a spinal cord identity. We hypothesised that at hindbrain levels the phenotype would be less severe as we are able to generate more Olig2 expressing cells when we differentiate cells under these conditions. I analysed hindbrain sections from E9.5 embryos to explore whether there was a difference in Olig2 expression. To ensure I was collecting sections from the correct location, I stained embryos for hindbrain markers Hoxb4 and Phox2b, which is expressed in visceral motor neuron progenitors (Pattyn et al. 1997).

Figure 4.9: Olig2 expression at E16.5
Brachial sections of wild type and O2e33 embryos stained for DAPI and Olig2. Scale bar is 200µm.
4.6. Anterior-Posterior Olig2 patterning

In O2e33 embryos at hindbrain positions, there appears to be a similar phenotype to that observed at more brachial sections (Fig. 4.10). Olig2 levels appear lower and the pMN domain smaller. These data suggest that, in the hindbrain, the -33 kb enhancer is required for correct expression even though the phenotype in vitro is less severe. It should be acknowledged however, that although the percentage of Olig2 expressing cells is higher under hindbrain conditions compared to under spinal cord conditions, the average and peak expression levels never reach the same intensity of wild type cells implying that even under hindbrain conditions, Olig2 expression is still affected by removal of the -33 kb enhancer.

When examining whole mount embryos for Olig2, expression in the midbrain of O2e33 embryos was not affected compared to wild type (data not shown) implying that deletion of the -33 kb enhancer does not affect expression in the anterior part of the embryo. This is consistent with data showing the -33 kb enhancer is not accessible at day 5 in cells differentiated under anterior conditions (Metzis et al. 2018).

Figure 4.10: Hindbrain expression of Olig2 in wild type and O2e33 embryos
Hindbrain sections of wild type and O2e33 embryos stained for Sox2, Nkx2.2 and Olig2. Scale bar is 50um.
4.7 Viability of O2e33 mice

This mouse line was maintained as a heterozygous line as when heterozygous mice were crossed of 42 pups weaned only 2 were homozygous. We were struck by so few homozygous so decided to cross these two homozygous mice. There were no pups from this breeding even though the female became pregnant three times. These mice did not display any abnormal behaviours.

Olig2<sup>−/−</sup> mice are not homozygous viable and pups die shortly after birth. This is due to failure to correctly form the hypoglossal nerve resulting in newly born pups being unable to feed (Zhou & Anderson 2002). When heterozygous crosses were set up for live births, no dead pups were found which could either due to them being removed by the mother or the embryos could be dying in utero. Interestingly, Olig1<sup>−/−</sup>Olig2<sup>−/−</sup> mice fail to be born and at e18.5 can be identified by size compared to heterozygous and wild type litter mates (Zhou & Anderson 2002). It may be that by removal of the -33 kb enhancer we have affected Olig2 as well as Olig1 and the combination in which both genes are affected results in failure of embryos to survive.

In addition, the observed differences in Olig2 in O2e33 embryos at e14.5 could be due to a variability in Olig2 expression in O2e33 embryos. Possibly, through noise the number of Olig2 positive cells may be at a critical point at E12.5 in order to repopulate the Olig2 progenitor pool in order to give rise to oligodendrocytes which may be the reason for survival of the two homozygous embryos.

To understand why homozygous mice are not being born, later time points could be collected and embryo size could be assessed. Equally, whole mount staining for Tuj1 would enable us to observe any differences in hypoglossal nerve generation.
Chapter 5

Boundary sharpness in O2e33 and \textit{Pax6}^{-/-} embryos

The defined domains of transcription factors in the neural tube are required for the generation of different neurons across the dorsal-ventral axis. Loss of certain genes can disrupt this patterning and cause a change in the neuronal populations generated. In addition to the previously observed loss of sharpness in \textit{Pax6}^{-/-} embryos (Balaskas et al. 2012), we now had a second mutant that also displayed a loss of sharpness at the same boundary although in this case without any nodes of the network being removed. I began to work with Edgar Herrera Delgado (EHD), a mathematician in the lab, on a collaborative project to explore whether the loss of sharpness phenotype in O2e33 mice could be modelled \textit{in silico} and if, in addition to \textit{Pax6}^{-/-} embryos, would enable us to understand how GRNs encode sharpness between domains of expression. The data in this chapter has been submitted for publication (Exelby et al. 2019).

5.1 Measuring sharpness

We were very interested in the loss of sharpness phenotype at the p3/pMN domain boundary (Fig. 4.1). A loss of sharpness at this boundary has been previously observed in \textit{Pax6} knockout embryos but had not been extensively analysed (Balaskas et al. 2012). We wanted to understand how disruption to the GRN by way of modifying interactions between nodes could result in a loss of sharpness instead of just a change of position of a boundary.

We needed a method to quantify the loss of sharpness and levels of expression and therefore needed to extract data from individual nuclei in the images of the neu-
5.1. Measuring sharpness

Measuring sharpness of neural tube sections. Automatic segmentation of nuclei proved extremely challenging and none of the available software at the time performed successfully. Therefore, I decided the best approach would be to manually select individual nuclei for extraction of data. For selecting and extracting fluorescence intensity data, I used the point tool in Fiji image analysis software. An ROI (Region of Interest) was placed at the centre of each nuclei, using all 4 channels of staining (DAPI, Pax6, Nkx2.2 and Olig2) to determine exactly where different nuclei were (Fig. 5.1A). I then used a macro that enlarged each point by a radius of 2 µm and extracted intensity data (Fig. 5.1B, AppendixE). These data were: mean, standard deviation, minimum and maximum for the intensity levels across all channels as well as XY co-ordinate of the point in the image. A fiduciary mark was placed at the most dorsal and ventral parts of the neural tube in order to determine size and align the section.

![Figure 5.1: Method to quantify single nuclei in neural tube sections](image)

**Figure 5.1: Method to quantify single nuclei in neural tube sections**

A Individual points (yellow) placed at the centre of each nuclei using Fiji point tool. B Points expanded by a radius of 2 µm.

We collected and analysed data from embryo sections between stages E8.5-10.5. Initially cells were classified into different identities, Nkx2.2 (green), Pax6 (blue), Olig2 (red), floor plate (black) or were excluded from analysis (cyan) (Fig. 5.2A). Intensity data from the points was normalised for each section between the highest expressing point and lowest. Positional information was used to restrict identities to certain areas of the neural tube, this was in order to remove anomalies from the stainings for example, background staining in the Nkx2.2 channel frequently labelled blood cells (see Fig. 4.2) so Nkx2.2 identity cells are only able to exist in...
the bottom 1/3 of the section. As Olig2 positive cells transition to Nkx2.2 positive cells depending on the time exposed to and distance from the Shh source there were occasionally transitioning cells within the sections that would express both Nkx2.2 and Olig2. These few cells were removed from the analysis.

Figure 5.2: Measuring sharpness of domain boundaries
A Map of nuclei from neural tube section image classified into identities.
B, C Quantification of domain size and boundary width in WT (grey) and O2e33 mutants (red). The pMN domain in WT is larger than in O2e33 embryos at E9.5 (250-350µm) (Mann-Whitney test p = 0.004). The p3-pMN boundary is wider, i.e. less sharp, in O2e33 mutants compared to WT (n = 6 (WT), n = 12 (O2e33), Mann-Whitney test p = 0.009).

Once each nucleus had been given an identity we then had to decide the best way to determine “sharpness” of the boundary. The conclusion we came to was to calculate the width of the area where both Nkx2.2 and Olig2 cells were present. In a perfectly straight boundary this area would be the width of two cells, whereas in an area that had intermingling of two fates the width would be much larger. We calculated the width of the area containing 87.5% of the intermingled cells and used these limits, the midpoint of which was determined as the boundary position. These quantitative measurements allowed us to compare the sharpness and position of the p3/pMN boundary for embryos across different stages from both wild type and mutants. We used neural tube DV length as a measure to determine stage of different embryos. For E8.5, embryos were between 150µm-250µm, for E9.5 250µm-350µm and for E10.5 larger than 350µm. At E9.5, O2e33 embryos had a smaller pMN
5.1. Measuring sharpness

domain and also a larger boundary width (Fig. 4.2B,C). It was more complicated to calculate boundary sharpness at E10.5 as at this stage there were very few Olig2 cells in the O2e33 embryos (Fig. 3.7).
5.2 Boundary sharpness in *Pax6*\(^{-/-}\) embryos

![Figure 5.3: Boundary sharpness in *Pax6*\(^{-/-}\) embryos](image)

Brachial neural tube sections of both wild type and *Pax6*\(^{-/-}\) embryos stained for Pax6 (blue), Olig2 (red), Nkx2.2 (green). Over time *Pax6*\(^{-/-}\) embryos display an increasing loss of sharpness at the pMN-p3 boundary. Scale bar is 50\(\mu\)m.

To analyse loss of sharpness in ventral neural tube patterning, we also wanted to quantitatively measure the loss of sharpness at the p3/pMN domain in *Pax6*\(^{-/-}\) embryos. We wanted to determine whether the observed loss of sharpness in *Pax6*\(^{-/-}\) embryos was more or less severe than that calculated for O2e33 embryos. I collected embryos between E9.0-E10.5 and we used the same approach to collect and analyse data from images of sections. In *Pax6*\(^{-/-}\) embryos, the p3/pMN boundary became progressively less sharp from E9.0 and by E10.5, Nkx2.2 positive cells were also seen dorsal to the pMN domain (Fig. 5.3). The boundary was also shifted dorsally in agreement with previous reports (Fig. 5.3, 5.4A)(Balaskas et al. 2012). Quantitative
analyses revealed that the loss of sharpness at the p3/pMN boundary in $\text{Pax6}^{-/-}$ embryos was more severe than in O2e33 embryos (Fig. 5.4B).

**Figure 5.4: Boundary sharpness and position $\text{Pax6}^{-/-}$ embryos**

(A) Position of the pMN-p3 boundary in WT and $\text{Pax6}^{-/-}$. Box plots show upper and lower quartile and mean; $n = 7$ (WT), $n = 8$ ($\text{Pax6}^{-/-}$), Mann-Whitney test $p = 0.005$.

(B) Quantification of the pMN-p3 boundary width in WT and $\text{Pax6}^{-/-}$, Quantification indicates that the p3-pMN boundary is narrower and sharper in WT than $\text{Pax6}^{-/-}$ mutants analogous to boundary position in (A) (Mann-Whitney test $p = 0.0006$).

5.3 Modelling sharpness

To model boundary sharpness in ventral neural tube patterning, EHD built upon a previous mathematical model of the GRN generated from the lab (Fig. 5.5A)(Cohen et al. 2014). This model was able to accurately recapitulate $\text{Pax6}^{-/-}$ in silico in terms of the p3/pMN boundary shift but not the loss of sharpness at the boundary. To model the loss of sharpness, we needed to transform the model from a deterministic to stochastic. At p3/pMN and pMN/p2 expression boundaries, bistable switches result in an all-or-nothing transition between one fate and the other meaning sharpness cannot be observed. In the deterministic model, the transition between the two steady states is determined solely by the level of Shh signalling, and the system remains in the pMN state until the level of signalling increases above the bistable region. However, in stochastic modelling, the presence of intrinsic noise causes fluctuations in gene expression that can result in spontaneous transitions between pMN and p3 identity within the bistable region (Perez-Carrasco et al. 2016). To transition to a stochastic model, a noise term was added to each ODE which would account for fluctuations in gene expression. Addition of a noise term as sufficient for the model to correctly recapitulated loss of sharpness at the p3/pMN boundary in
5.3. Modelling sharpness

Pax6\(^{-/-}\) embryos as well retaining sharp boundaries in wild type and other mutant patterning phenotypes (Fig. 5.5B).

We wanted to explore parameter adjustments that would generate an \textit{in silico} phenotype comparable to the observed O2e33 patterning phenotype. As the O2e33 mutant was not the knockout of a node but a removal of an enhancer, an element where transcription factor binding drives or represses expression, the parameters influencing Olig2 expression were screened to find alterations that would result in patterning observed \textit{in vivo}. From the network it is known that Irx3, Gli, Nkx2.2 and Sox2 all input to Olig2 expression. As these factors are all predicted to bind to the -33 kb enhancer which has been removed, we explored adjustments to all of these parameters. Strengths of these inputs were screened to find a range of parameters that would result in an \textit{in silico} O2e33 phenotype. From this screen, a range of parameters was found that generated an O2e33 patterning phenotype (Fig. 5.5B). This parameter range fit with our \textit{in vivo} and \textit{in vitro} observations (Exelby et al. 2019). The O2e33 \textit{in silico} patterning showed a delay in Olig2 onset consistent with the observed delay in Olig2 expression in embryos. Additionally, a large reduction in Sox2 (basal) binding probability was a key feature of parameters that generated O2e33 patterning which is consistent with deletion of only the Sox2 binding site \textit{in vitro} demonstrating an effect on Olig2 onset.

As we had two loss of sharpness p3/pMN boundary phenotypes which we could correctly recapitulate \textit{in silico}, we wanted to understand the basis for these alterations to the model, knockout of Pax6 or reduction of Olig2, would result in a loss of sharpness in this GRN. At the p3/pMN boundary cells are transitioning from a Pax6\(^{LOW}\)-Olig2 (pMN) to a Nkx2.2 (p3) state due to the concentration and the time exposed to Shh. The structure of the network results in a bistable regime between these two states (Fig. 5.5C). This bistability means that because of the addition of noise to the system, noise driven transitions between these two states occur in this region, i.e. at the p3/pMN boundary. Within Pax6\(^{-/-}\) and O2e33 patterning, the region where noise driven transitions occur is larger, visually appearing as a less sharp boundary. We calculated the “jump time”, i.e. the probability of transitioning from Pax6\(^{LOW}\)-Olig2 state to Nkx2.2, for each position along the neural tube (Fig. 5.5C). In wild type patterning it very quickly becomes less likely for jumps to
occur as position along the neural tube moves away from the ventral pole meaning there is a smaller region where these noise driven transitions can occur, visually displayed as a sharper boundary. In the \( \text{Pax6}^{\neq/}\) the change in jump time is slower meaning a larger region at the p3/pMN boundary where noise driven transitions can occur and in O2e33 the size of this region is somewhere between wild type and \( \text{Pax6}^{\neq/}\).

Figure 5.5: Modelling boundary sharpness

(A) A GRN comprising the transcription factors Pax6, Olig2, Nkx2.2 and Irx3 is responsible for positioning the p3 and pMN domains in the ventral neural tube.

(B) Simulations of the stochastic dynamics of the GRN account for the experimentally observed dorsal expansion of the p3 domain and loss of boundary precision in \( \text{Pax6}^{\neq/}\) and O2e33 compared to WT. Diagram illustrates the pattern of gene expression in WT neural tube with progenitor domains (p2, pMN, p3) and differentiated neuron populations (motor neurons, MNs).

(C) A 3D bifurcation diagram of the mathematical model illustrates a region of bistability for pMN (red; expressing Olig2 and Pax6) and p3 (green; expressing Nkx2.2) separated by a transition point (unstable fixed point of dynamics, purple). The noise driven transition pathway from pMN to p3 is indicated by black arrows (solid: pMN to transition point, dashed: transition point to p3).

(D) Fate jump times calculated from simulations: average time for noise driven transitions from pMN to p3 in WT (black), \( \text{Pax6}^{\neq/}\) (blue) and O2e33 (red). Relative position refers to distance from the bifurcation point. Grey shading indicates the time regime where transitions can occur on relevant developmental timescales. In WT, jump times change more rapidly as a function of position, leading to a narrower boundary than in \( \text{Pax6}^{\neq/}\) and O2e33.

Caption continued on next page
5.3. Modelling sharpness

Figure 5.5: Caption continued from previous page

(E) Gene expression space view of the transition path from pMN (red point) to p3 (green point) steady state via the transition point (purple point). Simulated trajectory (dots) shows stochastic fluctuations from the pMN steady state remain close to the most likely transition path (dotted line) and are not oriented directly towards the transition point in WT. Axes show relative expression levels. In \(Pax6^{-/-}\) the loss of the Pax6 dimension means that fluctuations are oriented towards the transition point. Hence, similar noise levels will result in more fluctuation-induced jumps. Data are shown for neural tube position at 0.1 fraction of total neural tube length dorsal to the bifurcation point and are representative of the behaviour within the bistable region. Figure adapted from (Exelby et al. 2019)

To understand why the jump times were different between the phenotypes, the model was used to calculate the minimum action path (MAP) of the transition between \(Pax6^{LOW}\)-Olig2 state to Nkx2.2 (Fig. 5.5E) (Perez-Carrasco et al. 2016). This predicts the most likely gene expression trajectory that a stochastic transition resulting from a small fluctuation in gene expression will take, thereby providing a portrait of the dynamical landscape that leads to a noise-induced transition. For a cell to transition from a pMN state to a p3 state it must reach a point between the two states called the transition point in the system. Once cells reach this point, cells will transition to the alternative state (Fig. 5.5E). In this system, it is incredibly unlikely for cells to transition back, Nkx2.2 expressing p3 progenitors will not transition back to Olig2 expressing pMN progenitors. This has been previously been demonstrated using cre-reporter lines (Dessaud et al. 2007). It is more difficult to reach the transition point in the wild type system as the path pitches away from the transition point before reaching it. This means that noise will only account for a few transitions at this boundary. For \(Pax6^{-/-}\) patterning, the transition is only from Olig2 to Nkx2.2 as there is no \(Pax6\) expression in the pMN state, the removal of this dimension means that noise driven fluctuations in gene expression are only in the Olig2 plane and therefore are directly aligned with the transition point (Fig. 5.5E). In \(Pax6^{-/-}\), even as the system moves further away from the boundary, noise-induced transitions are still likely to occur. The outcome is a much larger region where Nkx2.2 and Olig2 cells intermingle which manifests as a fuzzy boundary.

In the O2e33 system, because the levels of Pax6 and Olig2 are different at the p3/pMN boundary (Pax6 is higher and Olig2 is lower) the orientation of the
MAP and therefore the fluctuations in gene expression are aligned more towards the
transition point. Fluctuations in gene expression around the steady state bring cells
closer to the transition point than in wild type. As in $Pax6^{-/-}$, this means that
even as the system moves away from the boundary, noise driven transitions are still
likely to occur resulting in a larger boundary region with both Nkx2.2 and Olig2
expressing cells.
Chapter 6

**Regulation of Olig2 by -33 kb and +75 kb enhancer *in vivo***

After discovering that removal of both the -33 and +75 kb enhancers *in vitro* resulted in a complete loss of Olig2 expression, even though the gene itself had not been mutated, I was eager to generate a mouse line with removal of both of these enhancers to test whether removal of these enhancers *in vivo* would result in loss of Olig2 expression as well. During this mouse line generation, I took this opportunity to develop a +75 kb enhancer mutant on its own to further explore different anterior-posterior regulation of Olig2.

### 6.1 O2e33-75 mouse line generation

In order to generate a mouse knockout for both enhancers, I needed to target the locus already missing the -33 kb enhancer. Due to the distance of 100 kb between the two sites it would be near impossible (0.05% according to centimorgan calculations) to generate the double enhancer mutant through crosses of individual enhancer knockouts. Therefore heterozygous O2e33 males were used to fertilise super-ovulated wild type females. This would result in 50% of the fertilised ova being heterozygous for O2e33. These ova were then injected with the +75 kb enhancer CRISPR guides that were used to generate the O2e33-75 cell line (Chapter 3, Appendix D). Following an initial unsuccessful round of targeting, the second attempt yielded three mice that had the +75 kb enhancer excised. In one mouse, this deletion was on the O2e33 chromosome and two in wild type background (Table 6.1). The knockout in the O2e33 background was homozygous implying that the +75 kb enhancer was deleted on the same chromosome as the -33 kb enhancer. This was also thankfully
confirmed through genotyping of F1 mice. Both of these lines were maintained as heterozygotes. This also demonstrated that mice were viable with a single copy of the -33 kb enhancer of \( \text{Olig2} \) even when both copies of the +75 kb enhancer were missing.

\[
\begin{array}{|c|c|c|}
\hline
\text{Name} & \text{Deletion (bp)} & \text{Location (Chr16)} \\
\hline
\text{O2e33} & 3259 & 91191295 to 91194570 \\
\text{O2e33-75} & 1250 & 91300841 to 91302104 \\
\text{O2e75} & 1252 & 91300842 to 91302107 \\
\hline
\end{array}
\]

\textbf{Table 6.1: In vivo CRISPR deletions}

6.2 \textbf{O2e33-75 phenotype}

![Image of Figure 6.1]

\textbf{Figure 6.1: Reduction of Olig2 in O2e33-75 embryos at brachial levels}

Brachial sections of wild type and O2e33-75 embryos at E9.0. Sections are stained for Pax6 (blue), Olig2 (red) and Nkx2.2 (green). O2e33-75 show complete loss of Olig2 in the hindbrain. Scale bar is 50\,\mu m.

To initially assess any changes in phenotype, I analysed embryos at early neural tube patterning stages. At E9.0 there was very little Olig2 in O2e33-75 embryos (Fig. 6.1). There were so few positive cells, it was difficult to distinguish a clear domain. Consistent with the O2e33 and Olig2\(^{-/-}\) phenotype, Pax6 and Nkx2.2 expression had expanded into the pMN domain (Fig. 6.1)(Takebayashi et al. 2002). These data imply that alternative enhancers are still able to drive Olig2 expression in the spinal cord at this timepoint. This expression could be driven from the K23
enhancer located between *Olig2* and *Olig1* that is not accessible in our *in vitro* system.

<table>
<thead>
<tr>
<th>Merge</th>
<th>Pax6</th>
<th>Nkx2.2</th>
<th>Olig2</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Wild type" /></td>
<td><img src="image" alt="Pax6" /></td>
<td><img src="image" alt="Nkx2.2" /></td>
<td><img src="image" alt="Olig2" /></td>
</tr>
<tr>
<td><img src="image" alt="O2e33-75" /></td>
<td><img src="image" alt="Pax6" /></td>
<td><img src="image" alt="Nkx2.2" /></td>
<td><img src="image" alt="Olig2" /></td>
</tr>
</tbody>
</table>

**Figure 6.2: Loss of Olig2 in O2e33-75 embryos at hindbrain levels**

Hindbrain sections of wild type and O2e33-75 embryos at e9.0. Sections are stained for Pax6 (blue), Olig2 (red) and Nkx2.2 (green). O2e33-75 show complete loss of Olig2 in the hindbrain. Scale bar is 50um.

From *in vitro* experiments, the data suggested that the +75 kb enhancer had a stronger influence on *Olig2* expression under hindbrain conditions than under spinal cord conditions. I therefore wanted to assess *Olig2* expression in the hindbrain of O2e33-75 embryos. Hindbrain sections at E9.5 showed the complete loss of *Olig2* expression in O2e33-75 embryos (Fig. 6.2). Deletion of only two enhancer regions we were able to abolish *Olig2* expression from the hindbrain at this stage. To assess exactly where Olig2 was present along the AP axis I used whole-mount imaging so expression of *Olig2* could be linked directly to AP position through anatomical landmarks.

Whole-mount imaging of a 22-somite embryo enabled us to examine where Olig2 is expressed in O2e33-75 embryos. The brachial expression we observed in sections is consistent with expression observed brachially in the whole mount embryo between somites 7 and 15 (Fig. 6.3). As well as this area of expression in the spinal cord, Olig2 remains expressed in the forebrain (Fig. 6.3). As with O2e33 embryos, forebrain
expression of Olig2 remained unaltered suggesting that both the +75 and -33 kb enhancers do not contribute to Olig2 expression in this area.

![Figure 6.3: Reduced expression of Olig2 in O2e33-75 embryos](image)

Whole mount image of O2e33-75 embryo at E9.5 (22 somites) stained for Olig2 (red) and Nkx2.2 (cyan). The spinal cord Olig2-expressing pMN domain is reduced along both dorsal-ventral and anterior-posterior axis. Olig2 expression appears unaltered in the forebrain.

To examine how the levels of Olig2 changed over time, I analysed later stage embryos. At E10.5 there were very few Olig2 positive cells, even less compared to O2e33 embryos. By E11.5 there were no Olig2 positive cells in O2e33-75 (Fig. 6.4). Equally, at E12.5 no Olig2 expression was observed in O2e33-75 embryos (data not shown). At these stages, there is a boundary directly between Pax6 and Nkx2.2 compared to wild type that has reduced expression of Pax6 in the pMN domain. In addition to the loss of Olig2 positive cells, the size of the ventral horns in O2e33-75 embryos was much smaller compared to wild type implying this near loss of Olig2 resulted in severely reduced motor neuron differentiation.
6.2. O2e33-75 phenotype

Figure 6.4: Loss of Olig2 in O2e33-75 embryos after E9.5

Brachial sections of wild type and O2e33-75 embryos at E10.5 and E11.5. Sections are stained for Pax6 (green), Olig2 (blue) and Nkx2.2 (red). There are only a few cells expressing Olig2 in O2e33-75 embryos at E10.5 and no cells at E11.5. Scale bar is 100um
6.2.1 Motor neuron generation in O2e33-75 embryos

In Olig−/− embryos, cells that would normally differentiate to motor neurons instead acquire an interneuron fate. One marker of V2 neurons is Chx10 and when analysed an increased number of Chx10+ cells could be seen in O2e33-75 embryos compared to wild type (Fig. 6.5). Consistent with this, fewer Isl1 positive cells populated the ventral horns suggesting differentiation of fewer motor neurons. To confirm the reduced Isl1 cells was due to reduce motor neuron differentiation, I examined embryos for co-expressing Isl1/HB9 cells. At E11.5, there are very few Isl1/HB9 cells in O2e33-75 compared to wild type (Fig. 6.6).

<table>
<thead>
<tr>
<th>Islet1</th>
<th>Chx10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>O2e33-75</td>
</tr>
</tbody>
</table>

Figure 6.5: O2e33-75 embryos have increased V2 interneuron differentiation
Brachial sections of wild type and O2e33-75 embryos at E10.5. Sections are stained for Islet1 (red) and Chx10 (cyan). Scale bar is 50µm

To assess the whether the reduced motor neuron differentiation affects a par-
6.2. O2e33-75 phenotype

Figure 6.6: O2e33-75 embryos have reduced number of motor neurons
Brachial sections of wild type and O2e-33-75 embryos at E11.5. Sections are stained for Islet1 (green) and HB9 (red) and Sox2 (cyan). There are very few Isl1/HB9 cells in O2e33-75 (white arrows) compared to wild type. Scale bar is 100 um

particularly motor neuron column I examined embryos at E12.5. Embryos were stained for Chx10, Isl1 and Lhx3 to assess motor neuron number (Fig. 6.7). In wild type embryos, motor neurons in the medial motor column co-express Isl1 and Lhx3 consistent with previous reports. Neighbouring the MMC, neurons in the lateral motor column (LMC) express Isl1 and HB9. A pool of neurons next to the Lhx3/Isl1 expressing cells can be seen with higher levels of Isl1 most likely identifying the LMC although HB9 staining would confirm this population identity. In O2e33-75 embryos co-expressing Lhx3/Isl1 cells are severely depleted from MMC area and a smaller number of Isl1 positive cells are observed in the location of the LMC (Fig. 6.7). O2e33-75 embryos display an increased number of Chx10/Lhx3 positive cells at the edge of the progenitor zone centrally implying a larger number of V2 interneuron generation consistent with the earlier staged O2e33-75 embryos (Fig. 6.7).
6.2. O2e33-75 phenotype

<table>
<thead>
<tr>
<th>Wild type</th>
<th>O2e33-75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chx10</td>
<td>Red</td>
</tr>
<tr>
<td>Isl1</td>
<td>Cyan</td>
</tr>
<tr>
<td>Lhx3</td>
<td>Green</td>
</tr>
</tbody>
</table>

Figure 6.7: O2e33-75 embryos have reduced number of motor neurons

Brachial sections of wild type and O2e-33-75 embryos at E12.5. Sections are stained for Chx10 (red), Isl1 (cyan) and Lhx3 (green). Scale bar is 100um
6.2.2 Oligodendrocyte generation in O2e33-75 embryos

In O2e33 embryos, expression of Olig2 can recover at later stages to generate oligodendrocytes. To assess whether oligodendrocytes are generated in O2e33-75 embryos, I analysed embryos at E15.5. Following neurogenesis, oligodendrocyte precursor cells (OPCs) are born from the neuroepithelium at the location of the pMN domain (Richardson et al. 1997, 2000). They migrate rapidly throughout the surrounding grey and white matter where they undergo further proliferation and differentiation to mature oligodendrocytes. At E15.5, oligodendrocytes can be identified throughout the neural tube by expression of Olig2 (Zhou & Anderson 2002, Lu et al. 2002, Takebayashi et al. 2002). In O2e33-75 embryos there was a marked reduction in Olig2 positive cells at this stage compared to wild type implying a severe reduction in oligodendrocytes (Fig. 6.8).

Figure 6.8: O2e33-75 embryos display reduced numbers of oligodendrocytes
Brachial sections of wild type and O2e33-75 embryos at E15.5. Sections are stained for DAPI and Olig2 Scale bar is 100um
6.3 O2e75 phenotype

I wanted to generate the +75 enhancer deletion in a wild type background because of the interesting observation in vitro that the two enhancers have different impacts on Olig2 expression depending on the anterior-posterior identity of the cells. I wanted to see if I could visualise these impacts in vivo.

Both hindbrain and brachial sections of O2e75 embryos did not reveal any phenotype compared to wild type at E9.5 (Fig. 6.9). At both E9.5 and E14.5 Olig2 expression appears unaltered between wild type and O2e75 embryos (Fig. 6.9, data not shown). Removal of the +75 kb enhancer in directly differentiated cells in vitro had a less severe phenotype than removal of the -33 kb. More analyses would need to be undertaken to understand if removal of this enhancer impacts hindbrain Olig2 expression or oligodendrocyte development in vivo. As removal of both enhancers severely impacts Olig2 expression, it is likely that in O2e33 mice, the +75 kb enhancer contributes to Olig2 expression. This could make the +75 kb enhancer a shadow enhancer as it may only direct expression in vivo when the -33 kb enhancer is lost.
Figure 6.9: O2e75 embryos do not display a neural tube patterning phenotype
Brachial and hindbrain sections of wild type and O2e-75 embryos at E9.5. Sections are stained for Pax6 (green), Olig2 (blue) and Nkx2.2 (red). Scale bar is 50μm
Chapter 7

Discussion

During development, graded signals provide positional information to cells within developing tissues. In the developing spinal cord, a ventral to dorsal gradient of Shh protein directs the patterning of different progenitor domains. Neural progenitors within these domains express a distinct code of transcription factors that direct their subsequent differentiation to molecularly distinct neurons in the spinal cord. Shh signalling feeds into a gene regulatory network (GRN) that includes activatory and cross-repressive inputs which restricts expression of different transcription factors to progenitor domains delimited by sharp boundaries. This GRN has been modelled in a deterministic mathematical model that recapitulates wild type and mutant patterning (Balaskas et al. 2012, Cohen et al. 2014). To explore how this GRN has developed to ensure robustness in patterning, the interactions between nodes of the network required further investigation.

In this study, we focused on interactions at a particular node, Olig2. This is a key transcription factor within the ventral neural tube GRN. Olig2 is expressed ventrally in the pMN domain that gives rise to motor neurons and later oligodendrocytes (Choi et al. 1983, Hirano & Goldman 1988, Pringle & Richardson 1993, Richardson et al. 2000). We focused our attention on the -33 kb enhancer of Olig2 that has binding sites for all major inputs in the GRN (Oosterveen et al. 2012, Peterson et al. 2012, Nishi et al. 2015, Kutejova et al. 2016). We demonstrate, through analyses of Olig2 enhancers both in vitro and in vivo, how different enhancers influence Olig2 expression along the anterior-posterior axis and how enhancers contribute to robustness in gene regulatory networks driving patterning.

As well as having the predicted binding locations of TFs for the -33 kb enhancer,
our own ATAC-seq data on the *Olig2* locus suggested the -33 kb enhancer was a key regulatory region of *Olig2* due to its accessibility arising at the same time as *Olig2* expression. Excision of this enhancer *in vitro* severely disrupted *Olig2* expression in mouse ES cells directly differentiated towards cells with a spinal cord progenitor identity. Remaining *Olig2* expression highlighted a second regulatory region that becomes accessible at later stages located +75 kb downstream of *Olig2*. Differentiation of cells with either the -33 kb or +75 kb enhancer removed under different conditions suggested that the +75 kb enhancer plays a more important role in directing expression of *Olig2* in the hindbrain compared to spinal cord. When both -33 and +75 kb enhancers were excised *in vitro*, *Olig2* expression was completely absent.

Excision of the -33 kb enhancer *in vivo* resulted in embryos having a reduced *Olig2* expression in a smaller pMN domain as well as a loss of precision at the p3/pMN boundary. Interestingly, although deletion of the +75 kb enhancer in combination with deletion of the -33 kb enhancer resulted in a more severe reduction in *Olig2*, deletion of the +75 kb alone did not reveal a phenotype implying its function as a so-called shadow enhancer.

We have modelled the loss of precision observed in O2e33 embryos *in silico* and have been able to build our understanding of the dynamical landscape of the ventral neural tube GRN. From this we have begun to explore mechanisms by which sharpness at boundaries of gene expression domains can be encoded in the structure of a gene regulatory network.

### 7.1 Gene regulatory networks

Throughout development, complex developmental processes are coordinated by gene regulatory networks that direct expression of genes in the right place at the right time (Davidson 2010). GRNs that drive patterns of gene expression have been described in tissues such as the Drosophila blastoderm and early the early sea urchin embryo (Jaeger 2011, Oliveri et al. 2008). Through assembly of these networks we can explore how interactions between different genes determines the positioning of gene expression and the precision of developmental patterning. GRNs for both the Drosophila blastoderm, early sea urchin embryo and the ventral neural tube in mice have been modelled mathematically enabling further exploration of how the

The gene expression profile of a developing cell determines its behaviour. Removal of key genes directing developmental processes often leads to disruption or termination of events, for example, removal of Nodal in the early mouse embryo, Bicoid in the Drosophila blastoderm or Sox9 in sex determination (Conlon et al. 1994, Frohnhöfer & Nüsslein-Volhard 1986, Chaboissier et al. 2004). Therefore to explore how these graded signals dictate important patterning events, we need to explore how their expression is controlled and how they in turn control expression of downstream targets. We are beginning to understand more about these processes through analysis of regulatory regions directing expression of these genes and how their expression feeds into downstream GRNs (Jaeger 2011, Simon et al. 2017, Gonen et al. 2018).

In line with this logic, removal of Shh prevents correct patterning of the ventral neural tube but through analysis of the downstream GRN we are able to explore how graded Shh signalling directs precise patterning of neural progenitors. Previous work had determined the requirement and interactions of different transcription factors in ventral neural tube patterning and this has led to the description and mathematical modelling of a GRN determining patterning (Briscoe et al. 2000, Dessaud et al. 2010, Balaskas et al. 2012, Cohen et al. 2014). By excising an enhancer of a single gene in the ventral neural tube GRN, we modified the inputs on this node without altering levels of individual transcription factors. By deleting the enhancer within its genomic locus any subsequent changes in gene expression are able to contribute to the GRN. Therefore, we are able to assess how the structure of the GRN incorporates these changes and leads to alterations in downstream patterning. This is in contrast to many previous studies that explore enhancer regulation using enhancer driven reporters (Oosterveen et al. 2012, Peterson et al. 2012, Perry et al. 2011). Through this modification, we can begin to explore the logic used by enhancers to integrate multiple input signals to direct a single output expression level. Equally, the effect removal of this enhancer has on downstream patterning can be used to investigate mechanisms employed by regulatory systems to ensure precision and sharpness in patterning.
7.1.1 Enhancer logic in gene regulatory networks

For correct expression of a gene, enhancers need to integrate signals from multiple different sources. Although there are many ways to identify different proteins bound at enhancers, how enhancers integrate multiple inputs to generate a single gene expression output is still not well understood. There is still much discussion on how enhancer syntax, the spacing and orientation of binding sites, influences activity (Arnosti & Kulkarni 2005, Swanson et al. 2010, Farley et al. 2016). Expression of a gene at a certain time and location within development often relies on integration of both positive and negative inputs from the gene regulatory network within which it resides.

For a gene to be expressed, permissive activatory transcription factors are often required to bind to regulatory elements. In Drosophila, the transcription factor Zelda is uniformly expressed in the early embryo (Liang et al. 2008) where its binding at enhancers of early patterning genes is required to enable correct response to graded signals (Xu et al. 2014, Crocker et al. 2017). Sox2 has been suggested to play a similar role in the ventral neural tube by binding to neural tube enhancers (Bergsland et al. 2011, Peterson et al. 2012, Oosterveen et al. 2012). Deletion of either Sox2 or Zelda binding sites from enhancers responding to patterning signals prevents gene expression (Foo et al. 2014, Crocker et al. 2017, Peterson et al. 2012). For example removal of Sox2 binding sites in enhancers for Nkx6.1 and Nkx2.2 resulted in a loss of enhancer reporter expression (Peterson et al. 2012). We extend this finding in this work by showing the importance of the Sox2 binding site in the -33 kb enhancer for correct onset of Olig2 expression *in vitro*. This is mirrored by the substantial reduction in the basal input parameter in the mathematical modelling recapitulating the O2e33 ventral neural tube patterning (Exelby et al. 2019).

As well as enabling activation of specific transcription factors within a GRN, both Sox2 and Zelda exhibit “pioneering activity”, meaning that these TFs prime cis-regulatory elements prior to gene activation (Liang et al. 2008, Lee et al. 2013, Zaret & Carroll 2011). These factors also direct the opening of chromatin which enables subsequent binding by transcription factors (Iwafuchi-Doi & Zaret 2016). Previous work has shown that both these factors bind to regulatory elements prior to their activation with Sox2 binding to neural specific genes and Zelda to early
activation genes (Peterson et al. 2012, Nien et al. 2011). The delay observed in vitro when the -33 kb Sox2 site is deleted could therefore be because this enhancer is not correctly primed for Shh signalling.

In the ventral neural tube, Sox2 binds at enhancers to ensure that in response to Shh signalling, correct genes will be expressed in comparison to Shh signalling in different areas of development. Interestingly, ATAC-seq data from day 6 and 8 in vitro differentiated cells not exposed to SAG (Shh signalling) show accessibility at the -33 kb enhancer of Olig2 although Olig2 is not expressed and removal of 3 Gli binding sites from -33 kb enhancer removes reporter expression (Oosterveen et al. 2012). These data suggest that Sox2 binding without Gli is not sufficient to drive Olig2 expression in the ventral neural tube.

In addition to binding of basal permissive factors such as Sox2 in the ventral neural tube and Zelda to early patterning gene enhancers, these enhancers also need to integrate morphogen signalling and different transcription factors within the network. To explore this within our system, testable predictions could be generated using the mathematical model that was previously generated (Cohen et al. 2014, Exelby et al. 2019). For example, completely removing Nkx2.2 or Gli input on Olig2. One approach to testing these hypothesis could be to utilise the system we have generated using enhancer knock outs. As O2e33-75 cells demonstrate a complete loss of Olig2 expression in spinal cord progenitors in vitro, we know that when one of these two enhancers is deleted, all Olig2 expression is being driven from the second enhancer. We could therefore use O2e75 cells to target Nkx2.2 or Gli binding sites within the -33 kb enhancer to test their requirement. Equally, these cells could be used to understand how different factors bound at the -33 kb enhancer initiate gene expression. One way in which this could be done is using techniques such as CAPTURE (Liu et al. 2017). This technique is able to identify protein complexes bound at chromatin in specific loci and long-range DNA interactions. Using this technique, factors binding to the -33 and +75 enhancers under different conditions could be identified. The differences in factors binding at these enhancers may explain why deletion of either enhancer in vitro does not reach untargeted expression levels of Olig2.

Understanding enhancer logic continues to rely on synthetic systems (Farley
et al. 2016, Crocker et al. 2017). By developing a system where a single enhancer controlling gene expression within a GRN can be manipulated, understanding of how multiple GRN inputs are computed by regulatory regions can begin to be assessed. In addition, similarities between binding logic such as Sox2 in vertebrates and Zelda in Drosophila, general rules for transcription factor binding logic could be uncovered. Finally, new techniques that identify TF binding could be used to explore bound TFs at enhancers when expression is active vs. when it is suppressed.

7.1.2 Multiple enhancers combine to drive gene expression

One aspect that makes understanding regulation of a gene complicated, is that for any one gene, multiple enhancer regions may exist. Different enhancers can direct expression of a gene under different spatial and temporal conditions. During embryonic development, Sox2, the pan-neural transcription factor expressed in all neural progenitors, is expressed under the control of 5 different conserved enhancers depending on location in the nervous system (Uchikawa et al. 2003). Alternatively, instead of different enhancers directing expression at defined locations and time points, multiple similar enhancers can play a redundant role in gene expression for a single location. These co-called shadow enhancers have similar structure to the “primary” enhancer of a gene but are located further away and have been shown to contribute to robustness in gene expression (Perry et al. 2011, Barolo 2012). The presence of multiple redundant enhancers has been suggested for many developmental genes, highlighted by only mild phenotypes being exhibited when a single enhancer is deleted (Dickel et al. 2018, Osterwalder et al. 2018, Cunningham et al. 2018). In comparison, some enhancer deletions cause a reduction in gene expression and lead to severe developmental phenotypes. In these cases, gene expression is reduced but not completely removed implying that although alternative enhancers contribute to expression they are unable to act in a redundant manner (Gonen et al. 2018, Simon et al. 2017). It is not clear why deletion of some enhancers results in reduction of gene expression whereas other deletions appear to have a minimal effect. In this work we show that deletion of one or two enhancers of Olig2 leads to a reduction but not complete removal of its expression in the ventral neural tube implying that multiple enhancers combine to drive gene expression. In comparison to reduced Olig2 in the developing spinal cord, expression of Olig2 in the develop-
7.1. Gene regulatory networks

ing telencephalon does not appear affected by the removal of the -33 and +75 kb enhancers implying that for Olig2 distinct enhancers direct expression in different tissues, as observed for Sox2 (Uchikawa et al. 2003).

We demonstrate that removal of a key Olig2 enhancer causes a reduction of Olig2 expression in vitro and in vivo which leads to a loss of precision in patterning in the ventral neural tube. Equally, we highlight the requirement of a second Olig2 enhancer at +75 kb that has previously been linked to Olig1 expression in mature oligodendrocytes (Fulton et al. 2011). Deletion of both enhancers generates a more severe reduction in Olig2 expression in vivo and causes a complete loss of expression in vitro. Although the additional deletion of the +75 kb enhancer in vivo generates a more severe phenotype, deletion of the +75 kb on its own did not reveal a phenotype. This suggests that regulation of Olig2 using these two enhancers is not directly additive which poses the question of how these two enhancers interact together to drive Olig2 expression.

How two or more simultaneously active enhancers integrate their activities to determine the levels and timing of gene expression is poorly understood. The in vitro model, where we have removed expression of Olig2 through deletion of two regulatory regions, provides a system to investigate this question. Recent studies in the Drosophila embryo have explored interactions between a pair of shadow enhancers driving expression of the gap gene Krüppel (Scholes et al. 2019). They were able to demonstrate that swapping enhancer positions or placing identical enhancers in both positions resulted in changes to the gene expression output confirming that the both enhancers combine to drive expression output (Scholes et al. 2019). Following this logic, it is possible that swapping the positions of the +75 and -33 kb enhancers in the Olig2 locus would alter Olig2 expression. A system where the Olig2 enhancers can be moved to different positions could also be used to explore how two enhancers generate 3D DNA interactions that result in recruitment of transcriptional machinery (Hahn 2018). Finally, as previously mentioned, deletion of the Sox2 site in the -33 kb enhancer of Olig2 led to a delay in expression onset. Bioinformatic analyses of the two enhancers indicated that they do not have the same binding sites, the +75 kb enhancer has one predicted Gli binding site compared to 3 in the -33 kb, although both have a Sox2 site. If Sox2 is required for correct response to Shh signalling, it is
possible that removal of Sox2 sites from both enhancers \textit{in vitro} may delay further the expression of Olig2 or prevent expression entirely.

It is clear that multiple enhancers exist for the majority of genes and this appears to contribute to correct expression in development. What remains an outstanding question is how these enhancers contribute to patterning. Removal of the -33kb enhancer for Olig2 disrupts ventral neural tube patterning. Our analysis suggests that this enhancer is required to correctly integrate Olig2 within the ventral neural tube GRN and also drive robust Olig2 expression.

### 7.1.3 Enhancers ensure robustness in patterning

Although enhancers have been studied for nearly 40 years (Banerji et al. 1981), we are still not able to fully explain how non-coding DNA sequences direct expression of genes. It is also not clear how highly conserved enhancer regions have remained within the genome when removal of them does not appear to result in a phenotype (Osterwalder et al. 2018, Cunningham et al. 2018, Ahituv et al. 2007). It is not apparent why deletion of regulatory elements for certain genes results in a phenotype when for others removal of enhancers has no affect. Genomic deletion of two highly conserved enhancers that drive reporter expression in the developing forelimb resulted in no phenotype (Cunningham et al. 2018). Removal of “ultraconserved” enhancers for the neural transcription factor Arx led to viable offspring although subtle brain defects were observed (Dickel et al. 2018). As these deletions led to viable offspring, one hypothesis for the conservation of these sequences within the species is that viability in these mice being is due to the favourability of standard laboratory conditions and possibly, under non laboratory conditions, these mice would not survive. Evidence in support of this idea has come from studies in Drosophila that only exhibit a phenotype linked to removal of an enhancer under stress linked to higher temperatures (Frankel et al. 2010, Perry et al. 2010).

In contrast to these studies, removal of a single distal enhancer for Sox9 leads to sex reversal in mice and removal of an enhancer for Eomes, a transcription factor expressed in the anterior visceral endoderm (AVE) required for A-P determination in mice, results in a variably penetrant defects in axis formation (Gonen et al. 2018, Simon et al. 2017). Studies such as these are beginning to shed light on GRNs directing vital developmental events. We show in this work how genomic deletion of
a single enhancer for Olig2 results in disruption to ventral neural tube patterning and a sub viable phenotype. The reason these enhancers are vital for correct gene expression may be due to their importance in integrating signals within a GRN to ensure precise expression. Both Sox9 and Eomes expression rely on tight time scales. Sox9 expression is required to reach a critical threshold within a few hours in the gonads to ensure testis formation (Gonen et al. 2018, Chaboissier et al. 2004) and Eomes expression in the AVE is required for its migration at the onset of gastrulation (Arnold et al. 2008, Ciruna & Rossant 1999). It is possible that single enhancers directing specific expression are more likely to be vital for development than those that combine with multiple factors to drive gene expression over a larger area over a longer period of time such as Arx in the developing brain (Visel et al. 2013, Dickel et al. 2018).

Deletion of the -33 kb enhancer of Olig2 results in reduced precision at the p3/pMN boundary in the ventral neural tube. Precision generated through used of multiple enhancers has previously been explored in Drosophila (Perry et al. 2011). Analysis of patterning genes in Drosophila has shown that these genes often have primary and shadow enhancers that produce overlapping activities (Hong et al. 2008, Cannavò et al. 2016, Perry et al. 2010, Frankel et al. 2010). It has been demonstrated that the combined activities of these enhancers leads to robustness and precision in patterning (Perry et al. 2011). However, the combination of these enhancer activities was assessed using transgenic reporter embryos so the GRN dictating expression remained complete. Therefore the precision in reporter activity could not be explored in relation to the structure of the GRN. If removal of a key enhancer of a gap gene did alter precision in boundary formation, it is possible that generation of sharpness within the GRN is due to mechanisms described using mathematical modelling of precision in the ventral neural tube (Exelby et al. 2019).

7.1.4 Structure of GRNs encodes precision in patterning in the presence of noise

It has previously been suggested that GRN topology is able to define precision in patterning (Li et al. 2018, Cotterell & Sharpe 2010). In addition, theoretical studies have shown that network structure influences boundary sharpness between domains of gene expression (Perez-Carrasco et al. 2016). To determine the structure of differ-
Gene regulatory networks not only do the different genes need to be identified but also how different nodes interact and regulate one another via cis-regulatory elements. It is becoming increasingly easier to explore enhancer interactions and describe the edges in GRNs due to the development of multiple techniques to explore chromatin accessibility, 3D genome interactions and specific transcription factor binding (Cusanovich et al. 2015, Liu et al. 2017, Rowley & Corces 2018) but testing predicted network structure through manipulation of edges within a GRN has been largely unexplored. We demonstrate in this work how manipulation of an enhancer, edges of the network onto the node Olig2, results in a change in precision of boundary formation. To visualise this \textit{in silico}, noise within the system needed to be modelled (Perez-Carrasco et al. 2016).

Using the previously generated mathematical model (Cohen et al. 2014), the position of the p3/pMN boundary can be modelled. However, due to the deterministic nature of the model, stochastic fluctuations in gene expression were not considered. As the p3/pMN boundary is generated due to a bistable switch, fluctuations in gene expression would be expected to generate variations in the position and precision at which cells switch from pMN to p3 state (Perez-Carrasco et al. 2016). Therefore to incorporate the known fluctuations that arise due to the inherently noisy process of gene expression (Raser 2005), the previous model was elaborated to a stochastic form by the addition of noise terms (in collaboration with Edgar Herrera-Delgado (Exelby et al. 2019)). This addition successfully recapitulated the loss of precision observed in Pax6^{−/−} embryos (Balaskas et al. 2012). Using this model we could assess how the structure of the GRN in the ventral neural tube is able to suppress noise induced transitions at the p3/pMN boundary in order to generate a sharp transition between these two domains.

How the structure of a GRN in response to graded signals determines patterning has previously been explored in the formation of stripes of expression in Drosophila (Cotterell & Sharpe 2010). It was shown that 6 different dynamical mechanisms produced by 6 distinct families of network topology could influence stripe formation including bistability but how and whether bistability buffered noise to generate precise expression domains was not explored (Cotterell & Sharpe 2010). Within the neural tube, bistable switches between Nkx2.2, Olig2 and Olig2 and Irx3 determine
the stripe of Olig2-expressing pMN progenitors but these switches alone do generate sharpness in patterning, as demonstrated by Pax6−/− embryos (Balaskas et al. 2012). What we are able to demonstrate in this work is how the third node in the network (Pax6) influences the shape of the dynamical landscape of the system which forces fluctuations in gene expression to occur away from the most direct transition path between the two states within the bistable switch (pMN to p3 identity). Thus by controlling the configuration of stochastic fluctuations the structure of the GRN contributes to precision of gene expression.

The importance of precision in patterning is evident across development, not only in spatial domains of expression but precision in timing and expression levels. In the examples discussed, the presence of multiple enhancer regions are required to correctly incorporate a gene within its GRN and to drive correct expression levels (Gonen et al. 2018, Simon et al. 2017). In addition, work that has explored precision using enhancer reporters suggests that deletion of these enhancers may also affect precision in patterning (Perry et al. 2011). These reports support the idea that enhancers within these systems are required to buffer noisy gene expression. For example, the variability in penetrance observed when the AVE enhancer of Eomes is deleted in mouse embryos suggests the presence of noise in the system (Simon et al. 2017). We suggest in this work that precision in the ventral neural tube occurs due to the initial design in GRN structure. This is in comparison to alternative mechanisms that have been shown to correct anomalies in patterning to produce precision (Dahmann et al. 2011, Batlle & Wilkinson 2012). For example, in the zebrafish hindbrain, cells in a particular location with the wrong fate are targeted to switch fate to sharpen boundaries of expression (Addison et al. 2018). This is in contrast to the GRN in the ventral neural tube that drives precision by its design. The presence of intermingling between Olig2 and Nkx2.2 expression cells at the p3/pMN boundary also implies that alternative secondary mechanisms to drive sharpening such at that in the zebrafish hindbrain are not present in the case of the p3/pMN boundary.
7.2 Robust Olig2 expression is required during development

In this work, we focus on how two enhancers drive correct expression of Olig2 in spinal cord progenitors both in vitro and in vivo. In addition to the loss of precision in patterning at the p3/pMN boundary in the ventral neural tube, the reduced Olig2 expression has downstream affects on the correct generation of motor neurons and oligodendrocytes.

7.2.1 Expression of Olig2

Correct expression of Olig2 is vital for development of motor neurons and also oligodendrocytes during embryogenesis. Olig2 is expressed in the pMN domain in the neural tube where it demarcates progenitors that will give rise to motor neurons and oligodendrocytes (Zhou et al. 2000, Lu et al. 2000). We demonstrate in this work that the -33 kb enhancer of Olig2 is required for correct expression of Olig2 in spinal cord progenitors. In vitro, O2e33 cells differentiated to spinal cord progenitors have fewer Olig2 positive cells present at day 5 compared to DOK1-1 and these express lower levels of Olig2. Analysis of the population of differentiated cells at days 5 and 6 was not able to assess the change in ratio of alternative ventral neural tube cell types. In order to assess the number of p3 and p2 progenitors present at days 5 and 6, flow cytometry analysis using antibodies against Nkx2.2 and Pax6 would provide quantitative data on these cell types. These data would also confirm whether we are able to recapitulate the increase in Pax6 expression observed in Olig2 positive cells in vivo.

In the developing spinal cords of O2e33 and O2e33-75 embryos, there are lower levels of Olig2 and a reduced number of Olig2 positive cells. This reduction during neurogenesis results in reduced numbers motor neurons and oligodendrocyte precursor cells at gliogenesis. In contrast to spinal cord expression, both O2e33 and O2e33-75 embryos it appears there is no change in expression of Olig2 in the forebrain at E9.5. In the developing brain, Olig2 is expressed in diencephalon and telencephalon in cells that will differentiate to cholinergic neurons that integrate sensory and motor signals (Ono et al. 2009). The persistent expression of Olig2 in the brain suggests that regulatory control of Olig2 in the forebrain is distinct to the hindbrain and spinal cord. This is unsurprising given the lineage of cells
that will give rise to the brain differs from those of the spinal cord (Henrique et al.
2015, Metzis et al. 2018). It would be interesting to assess whether O2e33-75 cells
could differentiate in vitro under conditions that would give rise to Olig2 positive
cells with a forebrain identity, especially as these protocols are currently being es-
tablished in the lab to investigate comparisons between developmental and disease
directed differentiation. As Olig2 is a key transcription factor in the development of
glioblastoma (Trépant et al. 2015), O2e33-75 cells may be a useful tool to ensure all
Olig2 positive cells generated under forebrain in vitro conditions are homogeneous.

Although forebrain expression appears unaltered in the currently analysed en-
hancer knockouts, between O2e33 and O2e33-75 embryos, differences were observed
at hindbrain locations. In O2e33 embryos, there was no apparent difference in hind-
brain Olig2 expression but with the additional deletion of the +75 kb enhancer,
Olig2 expression was lost from hindbrain regions of E9.5 embryos. It has recently
been shown that spinal cord and hindbrain neural progenitors differentiated in vitro
have different chromatin landscapes due to their prior exposure to Wnt signalling
and expression of transcription factor Cdx (Metzis et al. 2018). At E9.5 in O2e33-75
embryos, Olig2 is expressed in the pMN domain between somites 7 and 15. The
loss of Olig2 expression in the anterior spinal cord is interesting as this location
may be where NMP contribution to the spinal cord begins (Vicki Metzis, personal
communication). Previous studies have shown that Cdx2 expression is required
for correct axial elongation with only somites up until number 5 forming normally
(van den Akker et al. 2002, Chawengsaksophak et al. 2004). In addition, Phox2b,
expressed in the hindbrain in visceral motor neuron progenitors is expressed up un-
til the boundary between somite 6 and 7 (data not shown)(Ikeda et al. 2015). It
is plausible that Olig2 expression in the spinal cord of O2e33-75 embryos occurs
only in more posterior cells that have previously been exposed to Wnt signalling
and expressed Cdx2. Alternatively, transcription factors expressed in the hindbrain
for example, Phox2b, could be contributing to Olig2 expression and through the
deletion of the +75 and -33 kb enhancers, these transcription factors can no longer
regulate Olig2.
7.2. Robust Olig2 expression is required during development

7.2.2 Differentiation of motor neurons is affected by Olig2 expression levels

In this study we show that reduction of Olig2 expression during development leads to a reduced number of motor neurons and oligodendrocytes being generated. We show this initially by using an in vitro approach that recapitulates the generation and differentiation Olig2 positive neural progenitors with a spinal cord identity to motor neurons (Sagner et al. 2018). By excising the -33 kb enhancer of Olig2 in mouse ES cells, we show that this enhancer is required to drive robust Olig2 expression in vitro. Although these cells are still able to differentiate to motor neurons, as demonstrated by the expression of Islet1 and HB9, fewer motor neurons are generated.

When O2e33 cells are differentiated in vitro, expression levels of Olig2 never reach that of untargeted (DOK1-1) cells. Previous work in the lab demonstrated that Olig2 expression increases prior to a cell differentiating to a motor neuron (Sagner et al. 2018). This work also demonstrated that Olig2^{HIGH} cells were also positive for pro-neural transcription factor Neurogenin2 (Ngn2) (Sagner et al. 2018). In addition, using mKate2 as a short-lived lineage tracer, mKate2^{+ve}Olig2^{-ve} cells were positive for Islet1 (Sagner et al. 2018). Although O2e33 cells do not reach the same intensity as Olig2^{HIGH} DOK1-1 cells, O2e33 cells do differentiate to motor neurons. O2e33 cells could be stained for Olig2 and Ngn2 in addition to mKate2 fluorescence to confirm whether the higher intensity Olig2 positive cells are indeed the cells transitioning to motor neurons even if they do not reach the same level as DOK1-1. It is plausible that Olig2^{HIGH} driven motor neuron differentiation does not rely on absolute levels of Olig2 but rather just a comparative increase in Olig2 expression. It could also be explored whether motor neurons generated through differentiation via an O2e33 Olig2^{HIGH} state are different to those generated via a DOK1-1 Olig2^{HIGH} cells.

Consistent with the in vitro data presented, differentiation of Olig2 positive cells to motor neurons occurs in O2e33 and O2e33-75 embryos although at reduced levels. Equally, the presence of Olig2 positive cells at later embryonic stages imply the ability of cells in these mice to differentiate to oligodendrocytes (Rowitch 2004). In O2e33 embryos, it is not clear how the reduction in motor neurons affects the assignment of motor neurons across different motor columns. In O2e33-75 embryos
staining with Lhx3 and Isl1 demonstrates a reduction in the number of lateral motor 
column neurons and a near loss of medial motor column neurons however it is not 
clear why differences in motor column assignment occur.

The identity of differentiating neurons in the spinal cord changes over 
time (Delile et al. 2019). This work from our lab developed a spinal cord gene 
expression atlas using single cell RNA-seq of E10.5 and E13.5 embryos. From these 
data they were able to identify two phases of neurogenesis, identified by expression 
of different pro-neural genes (Delile et al. 2019). If the time of differentiation deter-
mines the identity of different motor neurons, then due to the shortened window of 
motor neuron differentiation in O2e33 and O2e33-75 embryos, specific sub-types of 
motor neurons could be missing. To answer this question, further analysis of motor 
neuron markers at later stages would be required. Single cell RNA-seq of O2e33 and 
O2e33-75 embryos at E13.5 would also be able to identify changes in motor neuron 
identity and could be directly compared to data sets previously generated in the lab 
(Delile et al. 2019).

7.2.3 Differentiation of oligodendrocytes is affected in O2e33 and 
O2e33-75 embryos

Following neurogenesis, Olig2 is expressed in oligodendrocytes and their precursor 
cells (Zhou & Anderson 2002). At E14.5, Olig2 positive cells were not always present 
O2e33 embryos, however, collection of more embryos at this stage would confirm this. 
Collecting embryos at later stages could confirm the reduction in oligodendrocytes 
at E16.5. At E14.5 in O2e33 embryos, it appears that Olig2 expression re-appears 
in a small population of cells in the position of the pMN domain that are then able 
to give rise to migratory oligodendrocyte precursor cells (OPCs). It is not clear 
whether this population arises from the few remaining Olig2 positive cells present 
in O2e33 embryos or whether neural progenitors in the location of the pMN switch 
on Olig2 expression. This indeed has been debated within the field (Richardson 
et al. 2006). If existing cells expand to generate OPCs, then this is not affected 
by removal of the -33 kb enhancer and the reduced number instead may be due 
to the reduced number of Olig2 positive cells prior to this expansion. One method 
to test this would be to differentiate O2e33 cells towards oligodendrocyte fate in 
vitro to quantify oligodendrocyte differentiation efficiency in cells without the -33
7.2. Robust Olig2 expression is required during development

O2e33-75 embryos display a more severe phenotype than O2e33 embryos with Olig2 expression in the spinal cord nearly completely gone by E10.5. The presence of Olig2 positive OPCs at E15.5 implies that a population of oligodendrocytes can be generated without input from the -33 and +75 kb enhancers. From the stainings at E15.5, there appears to be more Olig2 positive cells in the dorsal half of the neural tube in O2e33-75 embryos although confirmation of this would require further quantitative analyses. Increased numbers of oligodendrocytes in the dorsal half of the neural tube is consistent with previous work demonstrating that Olig2 positive OPCs can be generated from the dorsal neuroepithelium in Nkx6.1/6.2 double knockouts and Shh-null embryos (Cai et al. 2005, Vallstedt et al. 2005). In these embryos, although Olig2 expression is lost from the ventral neuroepithelium during neurogenesis and gliogenesis stages, OPCs are still generated from the dorsal neuroepithelium. These OPCs retain the ability to develop into mature oligodendrocytes but their development is delayed compared to earlier born ventral OPCs (Cai et al. 2005, Zhu et al. 2011). It is therefore likely that the Olig2 positive cells present in O2e33-75 embryos at E15.5 arise from the dorsal neuroepithelium.

7.2.4 Viability of O2e33 mice

The O2e33 mouse line is sub-viable, only two homozygous O2e33 mice survived to adulthood out of 42 tracked births from heterozygous crosses. It should be noted that when paired, these two homozygous mice failed to generate litters even though the female became pregnant three times. This implies that embryos are either dying shortly after birth or dying in utero. To discern which of these is the case, later time points would need to be collected. Although survival of O2e33-75 mice has not been tested, it is most likely that these mice do not survive as they have lower levels of Olig2 in a smaller number of cells compared to O2e33. As the founder of the O2e33-75 line was homozygous for the +75 kb deletion it is likely that O2e75 mice are viable, this is in line with no phenotype observed in E9.5 or E14.5 embryos.

Olig2−/− embryos are born but die very quickly. This is because the hypoglossal cranial nerve does not form correctly therefore pups cannot feed (Takebayashi et al. 2002). In contrast, Olig2−/−Olig1−/− die in utero and can be identified at E18.5 by their smaller size compared to wild type (Takebayashi et al. 2002). It is possible
that through deletion of the -33 kb enhancer we affect regulation of both \textit{Olig2} and \textit{Olig1} which manifests as a phenotype similar to the double knockout. This could be assessed by analysing the hypoglossal nerve in O2e33 embryos at E10.5 and also compare embryo size at E18.5. As two mice were born and survived with no obvious phenotype, it is possible that there is variability in \textit{Olig2} expression across embryos, for example, there could be a critical number of Olig2 positive pMN progenitors that allows generation of enough motor neurons or correct formation of the hypoglossal nerve, and enough progenitors to then populate the spinal cord with oligodendrocytes. This line of argument would also be consistent with the idea of noise within the system. It is possible that through noise, variability in the phenotype could result in a small number of embryos reaching a critical number of Olig2 positive cells.

\textbf{7.3 Concluding Remarks}

In this study, we demonstrate the importance of two regulatory regions, -33 kb and +75 kb, for the correct expression of \textit{Olig2} in spinal cord progenitors both \textit{in vitro} and \textit{in vivo}. The data presented contributes to previous observations that multiple enhancer regions for different genes increase the robustness of gene expression. Removal of a key enhancer of \textit{Olig2} \textit{in vivo} results in a reduced pMN domain and loss of precision at the p3/pMN boundary. In collaborative work with EHD, we developed a previous mathematical model of ventral neural tube patterning to include noise induced fluctuations in gene expression. This transformation of the model was able to successfully recapitulate loss of sharpness in O2e33 embryos and previously described Pax6\(^{-/-}\) embryos. Using this model we were able to explore how structure of a GRN is able to restrict noise-induced transitions in order to generate sharp boundaries of gene expression during development.
## Appendix A

### JASPAR analysis of -33 kb enhancer

### Table: Predicted sites

<table>
<thead>
<tr>
<th>Model ID</th>
<th>Model name</th>
<th>Score</th>
<th>Relative score</th>
<th>Start</th>
<th>End</th>
<th>Strand</th>
<th>Predicted site sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA0673.1</td>
<td>MA0673.1</td>
<td>3.083</td>
<td>0.825637793820569</td>
<td>42</td>
<td>50</td>
<td>1</td>
<td>ggCTTGGCT</td>
</tr>
<tr>
<td>MA0673.1</td>
<td>MA0673.1</td>
<td>6.733</td>
<td>0.894633003787526</td>
<td>56</td>
<td>64</td>
<td>1</td>
<td>CCCATCGGG</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>2.511</td>
<td>0.818288725070966</td>
<td>117</td>
<td>125</td>
<td>1</td>
<td>CCTCCTCTC</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>Sox2</td>
<td>5.849</td>
<td>0.8009096275687171</td>
<td>123</td>
<td>132</td>
<td>1</td>
<td>CCCCTGTC</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>Sox2</td>
<td>6.693</td>
<td>0.8705617930287171</td>
<td>134</td>
<td>141</td>
<td>1</td>
<td>CCTTGGCT</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>4.046</td>
<td>0.846945605444679</td>
<td>140</td>
<td>148</td>
<td>1</td>
<td>CCTATTTTCC</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>OLIG2</td>
<td>3.897</td>
<td>0.8037591913052588</td>
<td>188</td>
<td>196</td>
<td>1</td>
<td>GCCGATTGTT</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>Sox2 Sox2</td>
<td>5.501</td>
<td>0.801250687253447</td>
<td>201</td>
<td>209</td>
<td>1</td>
<td>TCATGGTGC</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>2.544</td>
<td>0.818550438806668</td>
<td>316</td>
<td>324</td>
<td>1</td>
<td>CCCCCTGGCG</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>OLIG2</td>
<td>5.549</td>
<td>0.802412138049483</td>
<td>327</td>
<td>335</td>
<td>1</td>
<td>AGGATTGGTC</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>OLIG2</td>
<td>5.093</td>
<td>0.820364378101213</td>
<td>327</td>
<td>335</td>
<td>1</td>
<td>AGGAAATGGTC</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>Sox2</td>
<td>5.265</td>
<td>0.847492682949088</td>
<td>353</td>
<td>361</td>
<td>1</td>
<td>CCTTGGTG</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>Sox2</td>
<td>13.311</td>
<td>0.9717054029322066</td>
<td>406</td>
<td>413</td>
<td>1</td>
<td>Cwagg</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>2.544</td>
<td>0.818066442886966</td>
<td>450</td>
<td>458</td>
<td>1</td>
<td>CCCCCTGCC</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>1.653</td>
<td>0.802736095207969</td>
<td>463</td>
<td>471</td>
<td>1</td>
<td>CCAATGGGG</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>4.283</td>
<td>0.820272486232689</td>
<td>505</td>
<td>513</td>
<td>1</td>
<td>GCAATTGTA</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>1.501</td>
<td>0.857230977458598</td>
<td>519</td>
<td>527</td>
<td>1</td>
<td>CTATTCCAG</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>Sox2 Sox2</td>
<td>5.305</td>
<td>0.84742462634906</td>
<td>539</td>
<td>546</td>
<td>1</td>
<td>CCTTGGTC</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>Sox2</td>
<td>5.849</td>
<td>0.800909627308171</td>
<td>555</td>
<td>562</td>
<td>1</td>
<td>TCTTGGTC</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>2.348</td>
<td>0.815342521311986</td>
<td>571</td>
<td>579</td>
<td>1</td>
<td>CCTTCCTCCC</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>OLIG2</td>
<td>5.170</td>
<td>0.824971486222447</td>
<td>584</td>
<td>593</td>
<td>1</td>
<td>AGCCAATTTT</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>OLIG2</td>
<td>4.390</td>
<td>0.811692383605895</td>
<td>584</td>
<td>593</td>
<td>1</td>
<td>ACAATGGTT</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>5.076</td>
<td>0.84667721234216</td>
<td>634</td>
<td>642</td>
<td>1</td>
<td>GCTTCCCTG</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>2.322</td>
<td>0.814979328632324</td>
<td>647</td>
<td>655</td>
<td>1</td>
<td>CCTTGGCAT</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>1.817</td>
<td>0.820794622697958</td>
<td>680</td>
<td>688</td>
<td>1</td>
<td>ACTTCCTGAG</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>3.054</td>
<td>0.8201863968483232</td>
<td>678</td>
<td>686</td>
<td>1</td>
<td>CACATCGAG</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>8.190</td>
<td>0.820364378101213</td>
<td>681</td>
<td>689</td>
<td>1</td>
<td>GCCCTTGAG</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>Sox2</td>
<td>5.849</td>
<td>0.800909627308171</td>
<td>688</td>
<td>695</td>
<td>1</td>
<td>GCTTGGTC</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>2.528</td>
<td>0.818597127279746</td>
<td>719</td>
<td>727</td>
<td>1</td>
<td>CCCTTCCAG</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>Sox2</td>
<td>6.693</td>
<td>0.8705617930287171</td>
<td>734</td>
<td>741</td>
<td>1</td>
<td>CCTTGGTC</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>Sox2</td>
<td>6.752</td>
<td>0.8705617930287171</td>
<td>752</td>
<td>759</td>
<td>1</td>
<td>TCTTGGTC</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>4.461</td>
<td>0.82059132502689</td>
<td>786</td>
<td>794</td>
<td>1</td>
<td>GCTATTTTA</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>6.184</td>
<td>0.847065630709599</td>
<td>792</td>
<td>799</td>
<td>1</td>
<td>GCCCTTGGT</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>Sox2</td>
<td>6.229</td>
<td>0.820750401096966</td>
<td>782</td>
<td>789</td>
<td>1</td>
<td>CCCCTTGG</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>1.694</td>
<td>0.832307019639443</td>
<td>807</td>
<td>815</td>
<td>1</td>
<td>GCTTCTGG</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>3.574</td>
<td>0.837511686250513</td>
<td>833</td>
<td>841</td>
<td>1</td>
<td>CCTCCTCTC</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>3.053</td>
<td>0.820364378101213</td>
<td>844</td>
<td>852</td>
<td>1</td>
<td>GTATTG</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>OLIG2</td>
<td>6.700</td>
<td>0.848913696852386</td>
<td>854</td>
<td>863</td>
<td>1</td>
<td>AGGATCTG</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>OLIG2</td>
<td>5.093</td>
<td>0.820364378101213</td>
<td>854</td>
<td>863</td>
<td>1</td>
<td>AGGAAATGGT</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>6.722</td>
<td>0.849431175264183</td>
<td>895</td>
<td>903</td>
<td>1</td>
<td>GCCTTCCTG</td>
</tr>
<tr>
<td>MA0143.3</td>
<td>MA0673.1</td>
<td>12.352</td>
<td>0.9840398274483866</td>
<td>910</td>
<td>927</td>
<td>1</td>
<td>CCAATTCCAG</td>
</tr>
</tbody>
</table>
### Figure A.1: Results from analysing the -33kb enhancer sequence using JASPAR database (jaspar2016.genereg.net).

PWM percentage score for TF binding sites for Sox2, Nkx2.8 and Gli. Minimum percentage threshold set as 80%. The two highest scoring sites are ones for Nkx2-8 and Sox2. The Gli sites were not uncovered in these analyses.

<table>
<thead>
<tr>
<th>Model ID</th>
<th>Name</th>
<th>Affine Score</th>
<th>Relative Score</th>
<th>Start</th>
<th>End</th>
<th>Strand</th>
<th>Predicted Site Sequence</th>
</tr>
</thead>
</table>
| MA0143.3 | Nkx2-8 | 5.201 | 0.9669 | 1092 | 1 | 1 | ACCCTTGA |}

<table>
<thead>
<tr>
<th>Model ID</th>
<th>Name</th>
<th>Affine Score</th>
<th>Relative Score</th>
<th>Start</th>
<th>End</th>
<th>Strand</th>
<th>Predicted Site Sequence</th>
</tr>
</thead>
</table>
| MA0673.1 | Nkx2-8 | 6.982 | 0.9669 | 1092 | 1 | 1 | ACCCTTGA |}

<table>
<thead>
<tr>
<th>Model ID</th>
<th>Name</th>
<th>Affine Score</th>
<th>Relative Score</th>
<th>Start</th>
<th>End</th>
<th>Strand</th>
<th>Predicted Site Sequence</th>
</tr>
</thead>
</table>
| MA0673.1 | Nkx2-8 | 7.435 | 0.9669 | 1092 | 1 | 1 | ACCCTTGA |}

<table>
<thead>
<tr>
<th>Model ID</th>
<th>Name</th>
<th>Affine Score</th>
<th>Relative Score</th>
<th>Start</th>
<th>End</th>
<th>Strand</th>
<th>Predicted Site Sequence</th>
</tr>
</thead>
</table>
| MA0673.1 | Nkx2-8 | 4.620 | 0.9669 | 1092 | 1 | 1 | ACCCTTGA |}

<table>
<thead>
<tr>
<th>Model ID</th>
<th>Name</th>
<th>Affine Score</th>
<th>Relative Score</th>
<th>Start</th>
<th>End</th>
<th>Strand</th>
<th>Predicted Site Sequence</th>
</tr>
</thead>
</table>
| MA0673.1 | Sox2 | 5.581 | 0.9669 | 1092 | 1 | 1 | ACCCTTGA |}

<table>
<thead>
<tr>
<th>Model ID</th>
<th>Name</th>
<th>Affine Score</th>
<th>Relative Score</th>
<th>Start</th>
<th>End</th>
<th>Strand</th>
<th>Predicted Site Sequence</th>
</tr>
</thead>
</table>
| MA0673.1 | Sox2 | 6.713 | 0.9669 | 1092 | 1 | 1 | ACCCTTGA |}

The highest scoring sites are ones for Nkx2-8 and Sox2. The Gli sites were not uncovered in these analyses.
Appendix B

-33 kb enhancer conservation in humans

Figure B.1: Mouse -33kb enhancer sequence BLAST against human sequence. Mouse 999bp sequence is 84% conserved compared to human. Binding sites for Gli (green), Sox2 (red) and Nkx2.2 (blue) are highlighted.
Appendix C

Olig2 -33 kb locus
Figure C.1: Olig2 -33kb enhancer locus showing locations of primers and CRISPR guides as well as predicted binding sites. Sequence is chr16: 91190530 to 91194675
Appendix D

Olig2 +75 kb locus
Figure D.1: Olig2 +75 kb enhancer locus showing locations of primers and CRISPR guides. Sequence is chr16: 91300721 to 91302436
Appendix E

FiJi macro to obtain immunofluorescence intensity data from selected nuclei

Figure E.1: FiJi macro to collect data from selected nuclei. Text in blue details what each line of code represents

```javascript
// directory = getDirectory("Choose a directory"); // Opens window to select where you will save data
1 index = indexOf(name, "."); // Collects name of image being analysed
2 name = substring(name, 0, index); // Selects all ROIs to analyse
3 count = roManager("count"); // Each ROI is selected in turn, enlarged by 2 microns
4 for(i=0; i<counts; i++) { // then re-added to the ROI manager
6 roManager("Select", i); // An array is built to contain all the ROI points
8  run("Enlarge..", "enlarge=2"); // All ROIs are selected by selecting the array
9  roManager("Update"); // Multi-measure collects mean, max and min intensity values for each ROI
10 array = newArray(counts); // as well as XY position and channel number
12  for(i=0; i<counts; i++) { // Data values are saved as a .txt file
14    array[i] = i; // ROI points are saved as a .zip file
16  }
18  roManager("Select", array); // saveAs("Results", directory + "Results_" + name + ".txt");
20  roManager("multi-measure measure_all"); // saveAs("Results", directory + "Results_" + name + ".zip");
22  saveAs("Results", directory + "Results_" + name + ".txt");
24  roManager("Save", directory + "ROISet" + name + ".zip");
26
```
Bibliography


BIBLIOGRAPHY


marker starting from comparative analysis of data from similar DNA chip microarray platforms’, *Tumor Biology* **36**(3), 1943–1953.


van den Akker, E., Forlani, S., Chawengsaksophak, K., de Graaff, W., Beck, F., Meyer, B. I., Deschamps, J., de Graaff, W., Beck, F., Meyer, B. I. & Deschamps, J. (2002), ‘&lt;em&gt;Cdx1&lt;/em&gt; and &lt;em&gt;Cdx2&lt;/em&gt; have overlapping functions in anteroposterior patterning and posterior axis elongation’, *Development* **129**(9), 2181 LP – 2193.


