DEVELOPMENT OF TOOLS IN MOUSE FOR FUTURE GENE THERAPY:
PROMOTERS FOR THE CNS, AND NOVEL EXPRESSION MODELS OF
NEURAL STEM CELL REGULATOR, \textit{NR2E1}

by

Charles Norval de Leeuw

B.Sc. (Hon.), University of Alberta, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in
THE FACULTY OF GRADUATE STUDIES
(Medical Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

June 2013

© Charles Norval de Leeuw, 2013
Abstract

Three challenges exist for human neurobiology, specifically in the areas of genomics and genetic medicine: understanding of genome regulation, understanding of central nervous system (CNS) development, and the lack of promoters for human gene therapy. To address these issues, we used computational biology strategies, primarily involving phylogenetic footprinting, to identify putative regulatory elements in genes with a regionalized or cell-type specific expression pattern. We generated human MiniPromoter constructs, less than 4 kilobases in size, and made genetically engineered mice by single-copy knock-in at the mouse Hprt locus. Neuroanatomical analyses were performed on brain and eye, primarily. Using this strategy, we generated 50 novel MiniPromoters for use in driving gene therapy constructs. Lastly, we demonstrated retained specificity of three retinal ganglion cell layer MiniPromoters when these were moved from knock-ins to an adeno-associated viral vector, exemplifying the utility of these constructs in other systems. In order to study CNS development, we chose to functionally analyze non-endogenous expression of the neural stem cell regulator NR2E1 in mice. We employed a DCX-based MiniPromoter from the Pleiades Promoter Project, in addition to the ubiquitous CAG promoter, to drive ectopic and ubiquitous expression of human NR2E1. DCX-based expression of human NR2E1 did not result in any overt phenotypes and was unable to rescue the brain and eye defects observed in Nr2e1frc/frc mice. In contrast, the CAG promoter resulted in embryonic death at ~E8, with failure of neural tube closure. We showed that expression of NR2E1 has negative effects on embryonic stem cell growth. Furthermore, we observed altered Pax6 expression in NSCs and embryos. Future work on promoter design and NR2E1 biology will advance our knowledge of genome regulation and CNS development.
Preface

Chapter 1

As they relate to NR2E1, portions of this chapter have been published as part of the human and mouse NR2E1 articles, which I wrote, in:


Note that research article authors retain the copyright of published works in Genome Biology.

Where necessary, minor modifications have been made and the material has been updated as new publications or additional information have come to light. These modifications will also eventually be made to the TFe website: http://cisreg.ca/tfe. Specifically, the mouse Nr2e1 article is located at: http://www.cisreg.ca/cgi-bin/tfe/articles.pl?tfid=181 and the human NR2E1 article at: http://www.cisreg.ca/cgi-bin/tfe/articles.pl?tfid=182.

The remainder of this chapter describing the background data relevant to Chapters 2 and 3 is unpublished.

Chapter 2

This Chapter has been published as:


Minor typographical and grammatical changes were made as part of the incorporation into the thesis.

Note that Proceedings of the National Academy of Sciences USA authors retain the copyright of published articles. Furthermore, PNAS specifically gives authors the right to use a research article as part of a thesis without obtaining explicit permission.

This project was conceived and initiated by Dr. E.M. Simpson at the Centre for Molecular Medicine and Therapeutics (CMMT), University of British Columbia (UBC). This large scale endeavour, known collectively as the Pleiades Promoter Project, was part of an international multi-principal investigator project aimed at developing novel promoter resources for the brain. As such, many people were involved. However, I played a key role in
several aspects of this project including numerous conceptual discussions with team members and Dr. Simpson. I am fourth author on the primary publication of the project, preceded only by post-doctoral researchers on the project. Key involvement in this project included: (1) design and generation of a backbone vector that would allow recombination-mediated cassette exchange (RMCE) of Pleiades reporters with genes of interest by investigators, (2) selection of gene candidates for use in MiniPromoter design – particularly by analysis of appropriate knock-out mouse phenotypes for 173 candidates to corroborate their involvement in brain regions, (3) MiniPromoter design for the selected source gene \textit{SLC6A4}, (4) development of an \textit{in vitro} assay for assessing MiniPromoter-containing ESC gene expression to allow rapid assessment of design success, and (5) analysis by RT-PCR of a selection of candidates \textit{in vitro}. Figure 1 was created by me and is a result of my \textit{in vitro} work. I wrote the associated section of the paper and was further involved in manuscript proof-reading and editing until publication acceptance. This publication has not been previously used in a thesis.

\textbf{Chapter 3}

This project is a continuation of the work started in Chapter 2 and became focused under my direction. I pursued the exploration of neuroanatomical phenotyping of novel Pleiades MiniPromoters to identify promoters of biological interest. Furthermore, I pursued the investigation of MiniPromoter expression patterns in the retina as important key tools for future gene therapy. Imaging of the lacZ 1-mm brain sections was performed by a technician, Stéphanie Laprise, in the Simpson laboratory. The co-staining on pan-neuronal candidates was performed by the Goldowitz laboratory (Dr. D. Goldowitz, CMMT, UBC), specifically Simone C. McInerny and Lisa Borretta. I was in charge of the collaboration with the Hauswirth laboratory (Dr. W. Hauswirth, University of Florida). rAAV constructs were generated by Dr. Frank M. Dyka and tested by the Hauswirth laboratory and images sent to me for further analysis. Under my direction, a transcription factor binding site analysis of \textit{DCX}-based promoter fragments was performed by Michelle Zhou and Alice Y. Chou in the Wasserman laboratory (Dr. Wyeth W. Wasserman). All other experiments were executed by me, including detailed neuroanatomical expression analyses of all strains presented, sectioning and staining of eyes, harvesting and staining of embryos, co-labeling in the retina, and \textit{post-hoc} review of MiniPromoter expression sets. I performed
all data analysis, wrote the paper, and created all of the figures for this manuscript and will see it through to publication:


**Chapter 4**

This project was conceived and initiated by both Dr. E.M. Simpson and myself. I proposed the ectopic expression of *NR2E1* as an *in vivo* model to understand its function, and conceptualized the development of a novel mouse monoclonal due to the lack of availability of a commercial one. I received help with the second set of electroporations from Andrea McLeod and Siu Ling and with expansion of hybridoma clones from Siaw Wong. Ximena Corso-Díaz helped with dot blot screening of hybridomas. Mice were generated by microinjection or *in vitro* fertilization by the Pleiades Promoter Project and Mouse Animal Production Service (MAPS; CMMT, UBC) teams. All other experiments in this chapter were executed solely by me. Furthermore, I performed all data analysis, wrote the paper, and created all of the figures for this manuscript and will see it through to publication:

**de Leeuw CN,** Simpson EM. **Ubiquitous Expression of Nuclear Receptor 2E1 Results in Embryonic Death and Ectopic Expression Fails to Rescue Nr2e1<sup>hefe</sup> Defects. In preparation.**

**Ethics Approval**

The studies described in this thesis have been approved by the Animal Care Committee at the University of British Columbia. All mice were handled according to institutional guidelines. Approval certificates numbers include: A05-1258, A05-1748, A07-0430, A09-0980, A09-0981, and A11-0370.
Table of Contents

Abstract .......................................................................................................................................................... ii
Preface ........................................................................................................................................................... iii
Table of Contents ............................................................................................................................................ vi
List of Tables ................................................................................................................................................... xii
List of Figures ................................................................................................................................................... xiii
List of Abbreviations ..................................................................................................................................... xv
Acknowledgements ......................................................................................................................................... xviii
Dedication ....................................................................................................................................................... xx

Chapter 1: General Introduction .................................................................................................................. 1

1.1 The three challenges for human biology ............................................................................................... 1

1.1.1 Understanding genome regulation ................................................................................................. 1

1.1.1.1 Complexity of CNS gene expression patterns .............................................................................. 1

1.1.1.2 Gene expression analysis projects as resources for understanding the brain ......................... 2

1.1.1.3 A genome-wide enhancer element functional analysis .............................................................. 2

1.1.2 Understanding of CNS development ............................................................................................... 4

1.1.2.1 Brain development – a brief overview ....................................................................................... 4

1.1.2.2 Neural stem cell identity ............................................................................................................. 4

1.1.2.3 NR2E1 is a key regulator of brain and eye development ............................................................... 5

1.1.2.4 Neural stem cells require NR2E1 function ............................................................................... 7

1.1.2.5 NR2E1 interacts with the known stem cell factor Pax6 ............................................................... 8

1.1.2.6 Abnormalities in Nr2e1 mutant and overexpressor mice .......................................................... 9

1.1.2.7 NR2E1 is implicated in psychiatric disorders and cancer ......................................................... 10

1.1.2.8 NR2E1-modulation as a potential therapeutic strategy .............................................................. 10

1.1.3 Promoters for human gene therapy ................................................................................................. 11

1.2 Technologies enabling this work .......................................................................................................... 12
1.2.1 Use of site-specific docking at Hprt on the mouse X-Chromosome ................................................................. 12
1.2.2 AAV as a gene delivery vehicle for the CNS ........................................................................................................ 12
1.3 Thesis objectives ......................................................................................................................................................... 13
  1.3.1 General comments ............................................................................................................................................. 13
  1.3.2 Development of promoters for use in basic and clinical research ................................................................. 13
  1.3.3 Application of Pleiades MiniPromoter to study brain development ................................................................. 14

Chapter 2: A regulatory toolbox of MiniPromoters to drive selective expression in the brain ............... 16

  2.1 Preamble ................................................................................................................................................................. 16
  2.2 Introduction ............................................................................................................................................................ 16
  2.3 Materials and methods ........................................................................................................................................ 17
    2.3.1 Regulatory resolution score characterization ................................................................................................. 18
    2.3.2 MiniPromoter design ...................................................................................................................................... 22
    2.3.3 Hprt targeting vectors and MiniPromoters .................................................................................................. 24
    2.3.4 Knock-in immediately 5' of the Hprt locus .................................................................................................... 25
    2.3.5 PCR analysis of genomic DNA ...................................................................................................................... 25
    2.3.6 In vitro neural differentiation ........................................................................................................................ 25
    2.3.7 Immunohistochemistry and histochemistry ................................................................................................. 26
    2.3.8 Histochemistry of 1 mm brain slices ............................................................................................................ 27
    2.3.9 Regulatory element predictions in Olig1 enhancer sequences ................................................................ 27
  2.4 Results ................................................................................................................................................................. 30
    2.4.1 The Pleiades Promoter Project approach ................................................................................................. 30
    2.4.2 Validation of the approach using a set of previously characterized promoters and ESC neural differentiation ........................................................................................................................ 31
    2.4.3 Brain region-specific gene selection using a regulatory resolution score ................................................. 33
    2.4.4 MiniPs are designed on the basis of genomics analyses ............................................................................... 36
    2.4.5 Validation of the design strategy using the previously characterized promoters ....................................... 36
Chapter 2: Novel MiniPs derived from the bioinformatics analyses

2.4.6 Novel MiniPs derived from the bioinformatics analyses .......................................................... 38

2.4.7 A unique dataset for in silico Studies .................................................................................. 45

2.5 Discussion ................................................................................................................................. 45

2.6 Summary ................................................................................................................................ 46

Chapter 3: Human MiniPromoters for the CNS and demonstration of their use in adeno-associated virus

3.1 Preamble ...................................................................................................................................... 48

3.2 Introduction .................................................................................................................................. 48

3.3 Materials and methods ............................................................................................................... 50

3.3.1 Generation of knock-in mice .................................................................................................. 50

3.3.2 PCR analysis of genomic DNA ............................................................................................... 51

3.3.3 Immunohistochemistry of adult brains and other tissues ....................................................... 52

3.3.4 Embryo harvesting and staining ............................................................................................. 53

3.3.5 Immunohistochemistry on eyes ............................................................................................... 53

3.3.6 Immunofluorescence on eyes .................................................................................................. 53

3.3.7 Adeno-associated virus ............................................................................................................ 54

3.3.8 Image processing ...................................................................................................................... 55

3.3.9 Data and resource distribution ................................................................................................ 55

3.4 Results ........................................................................................................................................ 55

3.4.1 Novel brain MiniPromoter expression ranged from ubiquitous to highly regionalized ............ 58

3.4.2 Two MiniPs identified as candidates for pan-neuronal expression ......................................... 61

3.4.3 MiniPromoter expression was observed in spinal cord and non-CNS tissue ............................. 64

3.4.4 Expression of six individual MiniPromoters in embryos is confined to the heart, hypothalamus, skeletal muscle or is ubiquitous .............................................................. 66

3.4.5 Doublecortin (DCX)-based constructs demonstrated expression driven by a Prom segment and a conserved putative regulatory element in intron four ......................................................... 68
3.4.6 Similarities in MiniPromoter constructs for individual candidate genes indicate basal promoters usually confer specificity................................................................. 72

3.4.7 Seventeen MiniPromoter constructs demonstrated expression in the adult mouse eye................. 73

3.4.8 Knock-in constructs that express in retinal ganglion cells predict expression of promoter-driven GFP using AAV-based intravitreal delivery ........................................................................................................ 75

3.5 Discussion ................................................................................................................................. 77

3.6 Summary ................................................................................................................................ 83

Chapter 4: Ubiquitous expression of nuclear receptor 2E1 results in embryonic death and ectopic expression fails to rescue Nr2e1^frc/frc defects ........................................................................................................ 85

4.1 Preamble.................................................................................................................................. 85

4.2 Introduction ............................................................................................................................... 85

4.3 Materials and methods ............................................................................................................. 88

4.3.1 Generation of the constructs ................................................................................................. 88

4.3.2 Antibody production ............................................................................................................. 88

4.3.3 Transfection and immunofluorescence on HEK293 ............................................................. 89

4.3.4 Generation of knock-in mice ................................................................................................. 89

4.3.5 PCR analysis of genomic DNA ............................................................................................. 90

4.3.6 Embryo harvesting and staining ......................................................................................... 91

4.3.7 RT-PCR of ESCs and embryos ............................................................................................. 91

4.3.8 Generation of ESCs by in vitro fertilization ......................................................................... 92

4.3.9 Cytotoxicity assay on ESCs .................................................................................................. 93

4.3.10 In vitro differentiation of ESCs .......................................................................................... 93

4.3.11 Histological assessments on DCX-FLAG/NR2E1 mice ....................................................... 93

4.3.12 Histology of adult brains ................................................................................................... 94

4.3.13 Histology on eyes ................................................................................................................. 94

4.3.14 Funduscopy ......................................................................................................................... 95

4.3.15 Image processing ............................................................................................................... 95
5.2.2.2 Further characterization of *in vitro* neural differentiation .......................................................... 125
5.2.2.3 Analysis of male CAG-FLAG/NR2E1 embryos at E8.5 & detailed expression characterization differences in embryos .......................................................................................................................... 126
5.2.2.4 Harvesting of embryos at E8.5 for qRT-PCR gene expression comparisons ........................................ 126
5.2.2.5 Extensive histological evaluation of chimeras .................................................................................. 126
5.3 Conclusion ........................................................................................................................................... 127

**References** ........................................................................................................................................... 129

**Appendices** .......................................................................................................................................... 146

Appendix A : Dataset S1 ............................................................................................................................... 146
Appendix B : Dataset S2 ............................................................................................................................... 151
List of Tables

Table 2.1 List of predicted TFBS that were unique to the positive Ple151 construct .............................................. 28
Table 2.2 Predicted TFBS candidates with differential gene expression and literature evidence support ............ 30
Table 3.1 Detailed strain information .................................................................................................................. 51
Table 3.2 Summary of MiniPromoters .............................................................................................................. 57
Table 3.3 MiniPromoters with expression in the mouse eye .................................................................................. 58
Table 3.4 MiniPromoters with expression in tissues other than brain and eye ...................................................... 65
Table 4.1 Genotyping assays and PCR primer sequences .................................................................................... 91
Table 4.2 RT-PCR assays and primer sequences ............................................................................................... 92
List of Figures

Figure 1.1 Temporal and spatial regulation, and independent functioning of two enhancers. ................................. 3

Figure 1.2 Nr2e1frcfrc mice exhibit severe brain and eye abnormalities. ............................................................... 7

Figure 2.1 Genome-wide distribution of regulatory resolution scores. ................................................................. 21

Figure 2.2 Genes with (A) the highest, (B) average and (C) the lowest regulatory resolution scores. ..................... 22

Figure 2.3 In vitro neural differentiation for pre-screening MiniP designs........................................................... 32

Figure 2.4 Targeting of previously characterized MiniPromoters to the Hprt locus validates the Pleiades Promoter Project approach. ......................................................................................................................... 33

Figure 2.5 Regulatory resolution score prioritizes genes for MiniP design............................................................ 35

Figure 2.6 Refined designs of control genes validate the bioinformatics approach............................................. 37

Figure 2.7 Comparison of the Pleiades refined MiniPromoters with the previously characterized constructs .... 38

Figure 2.8 Novel MiniP expression patterns in the adult brain and retina............................................................. 40

Figure 2.9 A view of selected novel MiniP driving lacZ assessed in 1 mm brain slices........................................ 41

Figure 2.10 Selected novel MiniPromoter expression patterns assessed in 1 mm brain slices.......................... 42

Figure 2.11 MiniPromoters as tools to study developmental expression patterns .............................................. 44

Figure 3.1 Brain expression patterns of 18 novel MiniPromoters and two additional MiniPromoter strains that were remade with the lacZ reporter. ............................................................................................................ 60

Figure 3.2 Two MiniPromoters drive near pan-neuronal expression throughout the mouse brain.................... 63

Figure 3.3 Thirteen MiniPromoters expressed in the mouse spinal cord............................................................. 66

Figure 3.4 Expression of six MiniPromoters during development......................................................................... 68

Figure 3.5 Comparison of two Doublecortin-based MiniPromoter constructs with similar but non-identical expression patterns identifies putative functional TFBS........................................................................... 71

Figure 3.6 Basal promoters confer the majority of expression specificity............................................................ 73

Figure 3.7 Fifteen MiniPromoters expressed in the mouse eye........................................................................... 75

Figure 3.8 Three MiniPromoters that expressed in the retinal ganglion cell layer in knock-in mice maintained this restricted expression pattern from AAV viral vectors. ...................................................... 77
Figure 4.1 CAG-FLAG/NR2E1 allele expresses in targeted mouse embryonic stem cells. ................................................................. 97

Figure 4.2 Germline chimeras never have live-born carrier female offspring and female XHprt\textsuperscript{CAG-FLAG/NR2E1}/X mice show failure of cerebral hemisphere fusion at E9.5.................................................................................................................................................. 99

Figure 4.3 ESCs containing CAG-FLAG/NR2E1 grow slower........................................................................................................................................................................ 101

Figure 4.4 \textit{In vitro} differentiation of ESCs show no defect in neurogenesis, but less Pax6+Mcm2+ cells. ......................... 103

Figure 4.5 CAG-FLAG/NR2E1 chimeras made with lacZ-marked targeted ESCs are located non-specifically in eye and brain. ........................................................................................................................................ 105

Figure 4.6 Development is delayed and altered in E9.5 XHprt\textsuperscript{CAG-FLAG/NR2E1}/Y embryos.................................................................................................. 108

Figure 4.7 Ple53-FLAG/NR2E1 mice express in embryos and adult brain with no obvious histologically abnormal tissues......................................................................................................................................................................................... 111

Figure 4.8 Ple53-FLAG/NR2E1 fails to rescue the Nr2e1\textsuperscript{frc/frc} salient defects of reduced brain weight and reduced retinal blood vessels......................................................................................................................................................................................... 113
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>129</td>
<td>129S1/SvImJ</td>
</tr>
<tr>
<td>AAV</td>
<td>Adeno-Associated Virus</td>
</tr>
<tr>
<td>ABA</td>
<td>Allen Mouse Brain Atlas</td>
</tr>
<tr>
<td>aCGH</td>
<td>Array Comparative Genomic Hybridization</td>
</tr>
<tr>
<td>B6</td>
<td>C57BL/6J</td>
</tr>
<tr>
<td>BAC</td>
<td>Bacterial Artificial Chromosome</td>
</tr>
<tr>
<td>BPI</td>
<td>Bipolar Disorder Type I</td>
</tr>
<tr>
<td>BGEM</td>
<td>Brain Gene Expression Map</td>
</tr>
<tr>
<td>BrdU</td>
<td>Bromodeoxyuridine</td>
</tr>
<tr>
<td>CA</td>
<td>Cornus Ammonis</td>
</tr>
<tr>
<td>CldU</td>
<td>Chlorodeoxyuridine</td>
</tr>
<tr>
<td>CNS</td>
<td>Central Nervous System</td>
</tr>
<tr>
<td>DBD</td>
<td>DNA-binding Domain</td>
</tr>
<tr>
<td>DG</td>
<td>Dentate Gyrus</td>
</tr>
<tr>
<td>E(#)</td>
<td>Embryonic Day #</td>
</tr>
<tr>
<td>ENCODE</td>
<td>Encyclopedia of DNA Elements</td>
</tr>
<tr>
<td>ESC</td>
<td>Embryonic Stem Cell</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescence-Activated Cell Sorting</td>
</tr>
<tr>
<td>FRC</td>
<td>Fierce allele</td>
</tr>
<tr>
<td>GCL</td>
<td>Ganglion Cell Layer</td>
</tr>
<tr>
<td>GENSAT</td>
<td>Gene Expression Nervous System Atlas</td>
</tr>
<tr>
<td>HAT</td>
<td>Hypoxanthine Aminopterin Thymidine</td>
</tr>
<tr>
<td>IdU</td>
<td>Iododeoxyuridine</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>ISH</td>
<td><em>In Situ</em> Hybridization</td>
</tr>
<tr>
<td>INL</td>
<td>Inner Nuclear Layer</td>
</tr>
<tr>
<td>IPL</td>
<td>Inner Plexiform Layer</td>
</tr>
<tr>
<td>iNSC</td>
<td>Induced Neural Stem Cell</td>
</tr>
<tr>
<td>iPSC</td>
<td>Induced Pluripotent Stem Cell</td>
</tr>
<tr>
<td>LBD</td>
<td>Ligand-binding Domain</td>
</tr>
<tr>
<td>LOD</td>
<td>Logarithm of Odds</td>
</tr>
<tr>
<td>MiniP</td>
<td>MiniPromoter</td>
</tr>
<tr>
<td>NAbs</td>
<td>Neutralizing Antibodies</td>
</tr>
<tr>
<td>Nr2e1</td>
<td>Nuclear receptor subfamily 2, group E, member 1 – Mouse Gene</td>
</tr>
<tr>
<td>Nr2e1</td>
<td>Nuclear receptor subfamily 2, group E, member 1 – Mouse Protein</td>
</tr>
<tr>
<td>NR2E1</td>
<td>Nuclear receptor subfamily 2, group E, member 1 – Human Gene</td>
</tr>
<tr>
<td>NR2E1</td>
<td>Nuclear receptor subfamily 2, group E, member 1 – Human Protein</td>
</tr>
<tr>
<td>NSC</td>
<td>Neural Stem Cell</td>
</tr>
<tr>
<td>OB</td>
<td>Olfactory Bulb</td>
</tr>
<tr>
<td>ONL</td>
<td>Outer Nuclear Layer</td>
</tr>
<tr>
<td>OPL</td>
<td>Outer Plexiform Layer</td>
</tr>
<tr>
<td>P(#)</td>
<td>Post-natal Day #</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>Quantitative RT-PCR</td>
</tr>
<tr>
<td>RGC</td>
<td>Retinal Ganglion Cell</td>
</tr>
<tr>
<td>RMS</td>
<td>Rostral Migratory Stream</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse Transcriptase PCR</td>
</tr>
<tr>
<td>SGZ</td>
<td>Subgranular Zone</td>
</tr>
<tr>
<td>SNP</td>
<td>Single Nucleotide Polymorphism</td>
</tr>
<tr>
<td>SVZ</td>
<td>Subventricular Zone</td>
</tr>
<tr>
<td>------</td>
<td>---------------------</td>
</tr>
<tr>
<td>TAPs</td>
<td>Transitly Amplifying Progenitors</td>
</tr>
<tr>
<td>TFBS</td>
<td>Transcription Factor Binding Site</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription Start Site</td>
</tr>
<tr>
<td>UCSC</td>
<td>University of California, Santa Cruz</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated Region</td>
</tr>
<tr>
<td>VZ</td>
<td>Ventricular Zone</td>
</tr>
<tr>
<td>Wt</td>
<td>Wild-type</td>
</tr>
</tbody>
</table>
Acknowledgements

Thank you to my PhD supervisory committee for all the years of advice and guidance: Drs. Elizabeth Simpson, Ann Marie Craig, Connie Eaves, William Honer, and Blair Leavitt. I would also like to acknowledge the involvement of Dr. Alaa El-Husseini (posthumous) for being part of my committee initially.

Beth, I deeply appreciate all your patience, guidance and countless hours of support. Thank you for all of the knowledge you have shared with me.

A special thank you also to all of my collaborators, past and present, who have not only helped me pursue rigorous scientific investigations, but who have shaped me as a scientist.

Administrative staff of the Simpson Lab that have been incredibly supportive and helpful throughout my studies include Tracey Weir, Tammy Philippo, Katrina Bepple and Marina Campbell. You were always knowledgeable about grant applications, willing to proofread manuscripts or other documents and helped with organizing meetings.

My Simpson Lab friends will always be dear to my heart. I met some wonderful people who I will continue to cherish as colleagues as each pursues their own dreams and goals. The emotional support I received was second-to-none. Specifically, my fellow grad students past and present – you all know the difficult road of pursuing a PhD and the trials and tribulations inherent in such a pursuit: Dr. Bibiana Wong, Dr. Jean-François (Jeff) Schmouth, Ximena Corso-Díaz and Jack Hickmott. I marvel at the incredible talent you all have and the willingness to share your knowledge or help out in times of need, or just listening when the going gets tough. Thank you. I wish you all the best in your future endeavours – please stay in touch when you become famous, we may need to collaborate 😊. In addition to the students, a special thank you to the two amazing lab managers who made sure that any issues would be resolved right away and put my (ours, as students) needs at the forefront, and who were kind when we did not ‘behave like good lab citizens’ – Kathy Banks and Russell Bonaguro.

To my other friends, thank you so much for all your love, caring and understanding. You provided so much support that I could never repay you.
To my family, thank you for everything. I don’t even know where to start with the gratefulness I feel for letting the scientist in the family do ‘his thing’. To my older brother and sister, thank you for always listening to my ramblings about silly science things, even if they did not always make much sense.

To my mom and dad – I could not have wished for any better parents in this world. You both worked incredibly hard to ensure that I always had the financial support to do whatever I needed and to pursue my dreams, even if I did not always know where I was heading. Regardless of where I end up, I hope you can always be proud of me. I don’t think I could ever thank you for all the emotional support, either. Your selfless love is indescribable, as is the willingness to let me do and be whatever made me happy. Mom, I am finally following in your footsteps. Thank you both for all your love.
Dedication

To my parents
Chapter 1: General Introduction

1.1 The three challenges for human biology

Our knowledge of human biology is increasing at an exponential pace. The overwhelming nature of this continued accumulation of new data will prove challenging for future scientists to enable efficient and effective use of this data. Focused analysis of specific areas of biology will aid in the integration of this knowledge. Specifically, understanding the genetic bases of normal biology and disease will allow the integration of seemingly disparate sets of knowledge. For instance, the identification of a novel interaction between two otherwise unrelated genes allows additional inferences to be made about gene function, pathway involvement, and disease mechanisms. This capability of providing an integrative framework for biological data is the key reason why genetics, and the intrinsically related area of molecular biology, has now become the foundation of many other disciplines including neuroscience, psychiatry and psychology, physiology, and oncology. Also, the ability to manipulate our intrinsic genetic programming provides powerful tools in understanding biological mechanisms. However, should we succeed in the betterment of human disease, particularly in the central nervous system (CNS), three distinct areas of genetic medicine and genome biology will require extensive further investigation: (1) understanding genome regulation, (2) understanding CNS development which is predicated upon complex gene expression patterns and genetic interactions, and (3) the development of promoters for future human gene therapy. This thesis aims at generating new knowledge regarding all three aspects.

1.1.1 Understanding genome regulation

1.1.1.1 Complexity of CNS gene expression patterns

Histologically and physiologically, a large array of neuronal cell types exist with many variations in cell soma size, dendritic branching patterns, connections to and from various brain regions, neurochemical signaling use, and electrical firing rates and patterns (Masland 2004). Evidenced not only by their diversity in

---

physiochemical properties, complex patterns of gene expression also exist throughout development and continuing into the adult. The enormous number of variations in expression further generates discrete cell-types with diverse functions (Lein et al. 2007). Therefore, understanding neuronal function is indisputably linked to understanding neuronal transcriptomes.

1.1.1.2 Gene expression analysis projects as resources for understanding the brain

Large-scale genome-wide analysis of gene expression at high spatio-temporal resolution is now becoming available. GenePaint is a project aimed at characterizing gene expression via in situ hybridization in embryonic day (E)14.5 embryos (Visel et al. 2004). Efforts of St. Jude’s Brain Gene Expression Map focuses on in situ hybridization of transcripts in mouse brain from E11.5 to post-natal day (P)42 (Magdaleno et al. 2006). Other pioneering work in this area is that of the Allen Brain Atlas (ABA) group. Not only have they released an exhaustive investigation of in situ hybridization (ISH) in mouse for whole-embryo from E11.5 – E15.5, and brain from E18.5 – P56 (Lein et al. 2007), but they have now also generated data on gene expression in human brain and produced a similar anatomical atlas of gene expression (Hawrylycz et al. 2012; Zeng et al. 2012). The Gene Expression Nervous System Atlas (GENSAT) project is also characterizing random insertion bacterial artificial chromosome (BAC) transgenic mice to determine the expression pattern of the endogenous promoter and the proximal regulatory elements included in each BAC (Gong et al. 2003). These large scale endeavours result in gene expression maps and tools that researchers can use to study neuronal function.

1.1.1.3 A genome-wide enhancer element functional analysis

The large proportion of non-coding DNA (~98% (Elgar and Vavouri 2008)) and the small binding motifs of regulatory factors render reliable prediction of regulatory modules a challenging feat. On the other hand, promoter identification is less complex due to the ability to identify them using TATA-box and CAAT-box motifs, multiple transcription start sites (TSS), transcriptional evidence based on expressed sequence tags (ESTs) and complementary DNA (cDNA), and 5’ Cap-assisted Analysis of Gene Expression (CAGE) tags (Wasserman and Sandelin 2004). Here I focus on the identification of enhancer elements, which are generally smaller than 1 kb, may be located far from the gene being regulated, and confer gene regulation related to either or both the spatial or temporal pattern of the gene (Fig. 1.1) (Levine and Tjian 2003; Levine 2010; Visel et al. 2009a).
Genome-wide analysis of non-coding DNA to allow identification of putative regulatory enhancer elements has significantly improved in recent years, especially due to the availability of several complete genome sequences. This, in turn, allows multi-species comparisons by which evolutionary conserved, and therefore functionally important, elements can be identified.

An impressive project to identify enhancer elements was recently published, whereby a shotgun approach was taken to identify neural enhancers in *Drosophila*. Pfeiffer and colleagues (2008) demonstrated an initial set of 45 enhancers identified using 3 kb of non-coding fragments spanning the genome, with 1 kb overlap between elements. Since individual genes have core promoters with unique elements, they began by first constructing a synthetic core basal promoter containing a variety of these elements, maximizing the possibility of successfully identifying enhancers. Enhancer-GAL4 activity was demonstrated either by *in situ* hybridization or through expression of the UAS-GFP protein. The entire collection of over 7,000 transgenic fly lines has subsequently been published (Jenett et al. 2012). While a full-genome functional analysis in mice is still financially and practically intractable, similar work of the VISTA Enhancer project using human conserved elements in mice has been highly successful (Visel et al. 2007b; Pennacchio et al. 2006).

**Figure 1.1 Temporal and spatial regulation, and independent functioning of two enhancers.** Two enhancers, located on separate chromosomes, drive expression of lacZ in the developing limb bud of embryos. (a-d) Demonstrates the expression of the E3 enhancer across developmental time and physical space.
(e-h) the E4 enhancer lacks expression at day 10.5, but expresses from E11.5 (f) onwards. When used in combination, both enhancers retain their spatiotemporal regulation. Modified from Visel et al. (2009).

Importantly, for the purpose of this thesis, we have defined the “Prom” or basal promoter element as including both the ~200 bp surrounding the transcription start site (commonly referred to as the core promoter), as well as any immediate upstream conserved region that may be more classically defined as an enhancer region. Thus, we used basal “Prom” or “LongProm” constructs that are greater than 200 bp in size.

1.1.2 Understanding of CNS development

1.1.2.1 Brain development – a brief overview

Once the neural plate is formed (E7.5 in mice (Kaufman 1995)), it folds in forming the neural groove. The resulting folds close forming the neural tube of neuroectodermal origin and the adjacent neural crest (caudally at E8.5 in mice (Kaufman 1995)). The brain is divided into three discrete regions or vesicles in early embryonic development comprising of the prosencephalon, mesencephalon, and rhombencephalon (Bear et al. 2001). The prosencephalon further develops into the telencephalon and diencephalon. The eye and cortex are derived from the telencephalon, and thalamus and hypothalamus from the diencephalon. The mesencephalon generates the midbrain, while the rhombencephalon generates the cerebellum, pons and medulla.

1.1.2.2 Neural stem cell identity

Neural stem cells (NSCs) exist as an ill-defined group of cells with enormous therapeutic promise. They are responsible for generation of the three neural lineages (astrocytes, neurons, and oligodendrocytes) and are present from the earliest identification of neural tissue, the neuroectoderm, through to localized neurogenic niches in the adult brain. There is some lack of agreement on the exact identity of NSCs. In fact, it is plausible that their identity and phenotype changes over developmental time (Nishino et al. 2008), and even differs at a single point in time as has been demonstrated (Merkle et al. 2007), further emphasizing their heterogeneous nature. Even in the prototypical stem cell system of hematopoiesis there is now evidence suggesting heterogeneity of hematopoietic stem cells (Copley et al. 2012; Benz et al. 2012). Despite these issues, some agreement exists regarding certain characteristics (Conti and Cattaneo 2010): expression of neural markers Nestin, Sox1*, Sox2*, Gfap*, RC2*, Pax6*, Nr2e1* (* indicates commonly expressed, but not universal markers), tripotent
differentiation, self-renewal, and clonal expansion. *In vitro* these qualities are demonstrated by growth in neural stem cell proliferation media containing the mitogens bFGF and/or EGF for up to 10 passages (although some groups have demonstrated longer term expansion (Conti et al. 2005)), clonal secondary neurosphere assays indicative of self-renewal, and tripotent differentiation after withdrawal of growth factors. *In vivo* marker expression and incorporation of thymidine-analogs, BrdU, IdU or CldU (Taupin 2007), have been used to identify them in specific sites. Anatomically in the adult brain, NSCs are found in the neurogenic niches of the subgranular zone (SGZ) and subventricular zones (SVZ). In the SVZ, type B neural stem cells (NSC) generate type C transitly amplifying progenitors (TAPs) and subsequently type A neuroblasts (Alvarez-Buylla et al. 2002), whereas in the RMS the transition occurs from type B to type A directly (Doetsch et al. 1997).

For the purpose of this thesis, I will define NSCs as neural cells expressing the marker Pax6, known to play an important role in neurectoderm (Zhang et al. 2010), and the neural stem and progenitor cell marker Nestin (Doetsch et al. 1997; Reynolds and Weiss 1992; Lendahl et al. 1990), recognizing that some cells positive for the markers may not have stem or progenitor properties. As described previously (Doetsch et al. 1997; 1999), I will interchangeably refer to NSCs as type B cells, and will refer to progenitors as type C transitly amplifying progenitors (TAPs).

1.1.2.3 **NR2E1 is a key regulator of brain and eye development**

*Nr2e1* is expressed predominantly in the forebrain and dorsal midbrain regions throughout development (Monaghan et al. 1995) and remains expressed in the adult subventricular and subgranular zones, areas of active neurogenesis (Liu et al. 2008; Zhang et al. 2008; Shi 2007; Shi et al. 2004; Li et al. 2012). Expression begins as early as E8 in the brain and peaks at E12.5 (Monaghan et al. 1995), after which expression drops until birth (Yu et al. 1994). At E12.5, expression is high in the lateral telencephalon (dorsal and lateral pallium, and lateral ganglionic eminence), and low in the dorsomedial pallium and medial ganglionic eminence (Stenman 2003). At E14.5, expression is similar to E12.5 but also in the ventricular zone of the third ventricle (Stenman 2003). In the eye, *Nr2e1* is expressed in the neural retina and optic stalk at E13 (Yu et al. 2000). Specifically, *Nr2e1* is expressed in retinal progenitor cells, Müller glia, and astrocytes (Miyawaki et al. 2004), including proangiogenic retinal astrocytes (Uemura et al. 2006).
Additional studies demonstrated that Nr2e1 is expressed in NSCs and co-stains with Nestin, RC2, and often with indicators of proliferation, BrdU and Ki67 (Li et al. 2012; Shi et al. 2004; Sun et al. 2007; Li et al. 2008; Zhao et al. 2009; Denli et al. 2009). An Nr2e1 BAC driving Cre-recombinase crossed to a ROSA26 reporter mouse indicated that expression is found in the subventricular zone, rostral migratory stream, and dentate gyrus of the hippocampus; all areas of neurogenesis (Liu et al. 2008). This expression is also present in the mouse Allen Brain Atlas by in situ hybridization. Nestin co-expression in the SVZ and SGZ was verified with a β-galactosidase reporter knocked-in at the Nr2e1 locus (Shi et al. 2004; Zhang et al. 2008). Sparse expression of the reporter was also observed in the cortex in inhibitory glutamatergic interneurons, indicative of a unique function in these cells (Zhang et al. 2008). Using this reporter knock-in, it was demonstrated that β-gal-positive presumptive NSCs isolated from brain tissue using FACS proliferate in NSC-proliferation medium and are tripotent (Shi et al. 2004).

The Nr2e1 locus is located on Chromosome 10 in mouse. Multiple null and conditional alleles of mouse Nr2e1 have been discovered and derived (Zhang et al. 2008; Young et al. 2002; Monaghan et al. 1997). The "fierce" allele was identified in the Simpson laboratory as a spontaneous mutation, resulting in a complete deletion of the Nr2e1 coding region (Kumar et al. 2004; Abrahams et al. 2002). In all cases, mice lacking functional Nr2e1 have hypoplastic forebrains, specifically rhinencephalic and limbic structures (cortex, amygdala, hippocampus, olfactory bulbs, striatum), enlarged lateral ventricles, blindness (vision defects - hypocellularity and disorganization of ganglion, inner, and outer nuclear cell layers as well as abnormal ERGs (Yu et al. 2000)), and severe intense aggression towards other mice (Fig. 1.2) (Young et al. 2002; Roy et al. 2002; Monaghan et al. 1997; Stenman et al. 2003; Shi et al. 2004). In particular, the superficial cortical layers (II/III) are decreased in size in null mice, while overall patterning and organization are maintained (Land and Monaghan 2003). Additionally, the amygdalar phenotype is specific to the basolateral and lateral aspects (Stenman 2003). Defects in dendritic structure (abnormal arborization) and in dentate gyrus granule cell long-term potentiation (LTP) has also been demonstrated (Christie et al. 2006). Calretinin GABAergic interneurons and somatostatin-positive neurons were reduced in the cortex of mutant animals (Roy et al. 2002; Monaghan et al. 1997). In adult null animals, both excitatory and inhibitory interneurons are reduced (Roy et al. 2004).

Expression is also observed in the adult retina and is required for vision (Hollemann et al. 1998; Yu et al. 2000; Kobayashi et al. 1999; Abrahams et al. 2002; Daniel et al. 1999). Adult Nr2e1frc/frc mice are functionally blind.
as assessed by electroretinograms (ERGs) (Schmouth et al. 2012a; Young et al. 2002). By funduscopy there is reduced blood vessels in \( Nr2e1^{frc/frc} \) mice (Abrahams 2005; Schmouth et al. 2012a). Histologically there is a reduction in the size of the INL and ONL layers including displaced ganglion cells in the IPL (Young et al. 2002; Schmouth et al. 2012a; Abrahams 2005).

In addition to hypoplastic retinal layers in \( Nr2e1 \)-null mice, increased apoptosis is also observed in the ganglion cell layer at birth and inner nuclear layer by P7 (Miyawaki et al. 2004). It was demonstrated that Nr2e1 is required for astrocyte maturation, migration, and induction of astrocyte-based vasculogenesis (Miyawaki et al. 2004). Correct lamination of the eye is achieved through the interplay between cyclin D1 and p27(Kip1), and regulation of cyclin D1 by Nr2e1 (Miyawaki et al. 2004). In null mice, the radial extensions of retinal astrocytes are asymmetric, lacking lateral branches, with blood vessels completely absent (Uemura et al. 2006). These retinal astrocytes express Gfap, a target downregulated by Nr2e1, and lack fibronectin (Uemura et al. 2006). The hyaloid vessels in mutant animals fail to regress, become enlarged, and grow toward the avascular retina (Uemura et al. 2006).

This data emphasizes Nr2e1 as a critical regulator of brain and eye development and suggests a functional role in NSCs.

![Figure 1.2](modified from Abrahams et al. (2005) and Schmouth et al. (2012).)

**Figure 1.2** \( Nr2e1^{frc/frc} \) mice exhibit severe brain and eye abnormalities. (A) The forebrain and olfactory bulbs of \( Nr2e1^{frc/frc} \) mice are reduced in size compared to wild-type littermates. (B) Reduced retinal blood vessels, radial asymmetry and mottling is visible by funduscopy in \( Nr2e1^{frc/frc} \) eyes.

### 1.1.2.4 Neural stem cells require \( NR2E1 \) function

During early development (E9-14), neural progenitors lacking \( Nr2e1 \) exhibit a faster cell cycle transit time and premature differentiation (possibly due to premature switching from asymmetric to symmetrical
divisions (Li et al. 2008)), accompanied by a progressive depletion of progenitors after E14.5 (Li et al. 2008). This explains why later cortical structures are severely affected in null animals (Roy et al. 2002). When compared to wild-type at E14.5, reduced numbers of progenitors with prolonged cell cycles are observed in null animals (Li et al. 2008). In addition, a decrease in total dorsal telencephalon cell numbers is observed from E14.5 onwards in 

\( Nr2e1 \)-null animals (Roy et al. 2004).

\( Nr2e1 \) was shown to be required for the transition from radial glial NSCs to astrocyte-like NSCs around P9 (Li et al. 2008). There is also a reduction in Nestin-positive cells, and BrdU incorporation in the dentate gyrus and subventricular zone of adult null mice, but with no evidence of an increase in cell death (Shi et al. 2004). Importantly, FACS-isolated β-gal-positive presumptive NSCs from \( Nr2e1^{lacZ/} \) mice fail to proliferate \textit{in vitro} but this function in driving and maintaining proliferation in NSCs can be rescued by reintroduction of \( Nr2e1 \) (Shi et al. 2004). Furthermore, overexpression of any of the microRNAs which target \( Nr2e1 \) mRNA translation inhibition: \( \text{let-7b, let-7d and miR-9-1} \), results in reduced proliferation of neural stem cells (Zhao et al. 2009; 2013; 2010).

\textbf{1.1.2.5 \( NR2E1 \) interacts with the known stem cell factor \textit{Pax6}}

Both \( Nr2e1 \) and \textit{Pax6} are expressed in NSCs, and both play important roles in the eye. Genetically, the \textit{Pax6 small-eye} (Sey) mutation significantly enhances the \( Nr2e1 \)-null phenotype (described below), resulting in enhanced dorsal shift of subpallial markers, both as \( \text{Sey/+;Nr2e1}^{-/-} \) and even more so with \( \text{Sey/Sey;Nr2e1}^{-/-} \) (Stenman 2003). Other evidence for this genetic interaction comes from work in \textit{Xenopus} where both \textit{Xtll} (\textit{Xenopus} homolog of \( Nr2e1 \)) and \textit{Xpax6} (\textit{Xenopus Pax6}) are expressed in the brain, eye and testes; the fusion of the DBD of \textit{Xtll} to the engrailed repressor domain resulted in specific inhibition of eye vesicle evagination and a reduction in \( Xpax6 \) expression (Hollemann et al. 1998). In \textit{Xenopus} animal cap experiments, \textit{Noggin, Pax6, Six3} and \textit{Lhx2} were each shown to increase \textit{Xtll} expression (Zuber et al. 2003). Unpublished work in our laboratory has demonstrated decreased \textit{Pax6} expression in an overexpressor of mouse \( Nr2e1 \) (Wong 2009). These mice also display a \textit{small-eye}-like phenotype with microphthalmia and corneal opacity.

Combined, this data suggests a complex interplay between these two transcription factors. In the brain, loss of \textit{Pax6} enhances \( Nr2e1 \)-null phenotypes and overexpression of \( Nr2e1 \) functionally inhibits \textit{Pax6} function, at least in eye. To understand this interaction better, reductionist models, such as \textit{in vitro} analysis systems, would be useful.
1.1.2.6 Abnormalities in Nr2e1 mutant and overexpressor mice

Hyperactivity of Nr2e1<sup>frec/frc</sup> animals is observed as early as P18 (Wong et al. 2010). These animals also hypervocalize, jump at and bite handlers, and exhibit a "difficult to handle" phenotype (Roy et al. 2002). Conflicting data exist as to whether null animals may have normal or increased response to pain (Roy et al. 2002; Wong et al. 2010). However, in one study where pain response was not reduced, impaired fear-associative learning was demonstrated (Roy et al. 2002). Mutant animals also showed only residual spatial learning in Morris water maze tests (Monaghan et al. 1997). Additionally, mutants exhibit maternal instinct deficits leading to poor pup survival (Young et al. 2002). Curiously, Nr2e1<sup>frec/frc</sup> mice do not exhibit acoustic startle response, preventing their testing in pre-pulse inhibition (Wong et al. 2010). Despite the inability of lithium chloride, a commonly-used adult-administered small molecule therapeutic for bipolar disorder in humans, to ameliorate either the behavioural or neuroanatomical abnormalities in Nr2e1<sup>frec/frc</sup> mice, some commonalities in phenotypes do exist between these mice and bipolar type I patients (Wong et al. 2010), a disease which has been associated with an non-coding NR2E1 variant (Kumar et al. 2008). Combined, this data suggests that NR2E1 may play a role in human psychiatric disorders.

In mice, it has been shown that overexpression (two additional copies of Nr2e1) results in tumour formation and lack of the septum in the brain (Liu et al. 2010). Two studies have demonstrated that overexpression results in increased proliferation, indicated by Ki67 or BrdU staining, and further validated by increases in the size and number of neurospheres generated in vitro (Liu et al. 2010; Wong et al. 2010). While tumours did develop in overexpressing mice nearing 2 years of age, tumourigenicity was greatly increased on the p53<sup>-/-</sup> background, resulting in hyperproliferative lesions in all mice older than 5 months (Liu et al. 2010). VEGFR2 and CD31 immunostaining demonstrated evidence of a close association of Nr2e1-positive cells with blood vessels. BrdU analysis showed in vivo angiogenesis in these mice (Liu et al. 2010).

A mouse random insertion overexpressor strain was demonstrated to rescue the Nr2e1-null phenotype (Liu et al. 2010), and human NR2E1 has also been shown to rescue the brain and behaviour phenotypes of 'fierce' mice (Abrahams 2005) likely due to high conservation at both the protein and DNA level (Abrahams et al. 2002). This data thus suggests that mouse and human NR2E1 may have similar endogenous functional roles.
1.1.2.7 **NR2E1 is implicated in psychiatric disorders and cancer**

Human NR2E1 is expressed throughout the brain, including forebrain regions (Kumar et al. 2008). The gene is located at Chromosome 6q21 in humans, first identified as a region of minimal deletion in non-Hodgkin lymphoma (NHL), but has not been shown to be involved in NHL disease etiology (Jackson et al. 1998). This location has, however, been identified as a locus within a region of genetic linkage with bipolar disorder and schizophrenia in genome-wide analyses: specifically for Bipolar Disorder type I (BPI), a LOD score of 4.19 was observed in a linkage meta-analysis (McQueen et al. 2005).

Interestingly, candidate regulatory mutations were identified in patients with severe cortical malformations (Kumar et al. 2007b), but not in MMEP (microcephaly, microphthalmia, ectrodactyly, and prognathism) (Kumar et al. 2007a). A significant association was also identified between a non-coding SNP of NR2E1 and bipolar disorder (Kumar et al. 2008). NR2E1 was associated with bipolar disorder I and II ($P = 0.013$), bipolar disorder I ($P = 0.015$), bipolar disorder in females ($P = 0.009$), and bipolar disorder with age of onset < 26 years of age ($P = 0.006$), and 8 novel rare candidate regulatory mutations were identified.

So far, less than a handful of protein coding mutations have been identified in human NR2E1, all of which have unknown functional outcomes. Recently, an R100C mutation in NR2E1 in a glioblastoma multiforme patient with 1332 somatic mutations was identified (Parsons et al. 2008). Increased expression of NR2E1 has also been observed in human brain cancers, including glioblastoma, where elevated expression in the proliferative and mesenchymal subgroups of high-grade glioblastomas was demonstrated (Phillips et al. 2006). In neurocytoma, expression was 5.84-fold that of adult ventricular zone cells ($P = 0.0003$) and 276-fold higher when compared to cultured adult neural stem-progenitor cells ($P = 0.000003$) (Sim et al. 2006). In addition to increased expression within human glioblastomas, one study has also found increased copy number of NR2E1 in two cases of primary human glioblastoma using aCGH (Liu et al. 2010). NR2E1 was also shown to be overexpressed in primary glioblastomas, specifically within a subpopulation of cells that are also Nestin-positive, suggesting a role in brain tumour stem cells (BTSCs) (Liu et al. 2010).

1.1.2.8 **NR2E1-modulation as a potential therapeutic strategy**

The above data indicate that downregulation of NR2E1 in tumours and potential upregulation in brain disorders where neural stem cell dysfunction is involved may be beneficial. However, for either of these
strategies to become therapeutic options, much more research needs to be done on NR2E1 regulation and function. Thus, we propose by using mouse as a model organism, we can perturb NR2E1 levels and evaluate the resulting outcomes.

1.1.3 Promoters for human gene therapy

While many current gene therapies start by employing a ubiquitous promoter to demonstrate a functional outcome both pre-clinically and in trials, it is clear that there is a need to limit the expression of therapeutic molecules. Ubiquitous promoters can increase the immunogenicity of the deliverable (Cordier et al. 2001; Sun et al. 2005). Counter intuitively, viral ubiquitous promoters may even express at lower levels than cell-type specific promoters (Xu et al. 2001; Cordier et al. 2001). Moreover, uncontrolled expression can cause unwanted side-effects. Limited progress has been made on large-scale promoter development. In particular, as described previously, the enormity of gene expression variation in the CNS makes it especially clear that there is a need for a diverse array of promoter constructs for appropriate targeting of therapeutics.

Specific interests for promoters include those that express in neurogenic regions in order to modulate neurogenesis – e.g. KI67 and DCX. Although some promoters already exist for astrocytes and oligodendrocytes, better and smaller novel promoters may be necessary. These genes include GFAP, S100β, UGT8 and OLG1. Lastly, certain regions of the brain hold great therapeutic interest due to disease involvement – e.g. Raphe Nuclei in depression and Alzheimers (Michelsen et al. 2008), and Substantia Nigra in Parkinsons disease (Parent and Parent 2010).

Furthermore, the eye is at the leading edge of gene therapy trials (Cepko 2012; Cideciyan et al. 2008; Hauswirth et al. 2008; Simonelli et al. 2010; Maguire et al. 2008; Ashtari et al. 2011), increasing the need for cell-type specific promoters for this purpose also.

With these issues in mind, the Pleiades Promoter Project was initiated by Dr. Simpson to address these needs.
1.2 Technologies enabling this work

1.2.1 Use of site-specific docking at Hprt on the mouse X-Chromosome

The hypoxanthine phosphoribosyl transferase (Hprt) gene plays an important biochemical role in the salvage pathway of nucleotide synthesis (Caskey and Kruh 1979). It is expressed ubiquitously in tissues (Caskey and Kruh 1979; Stout and Caskey 1985). In humans, mutations in the HPRT gene lead to Lesch-Nyhan Syndrome, a severe biochemical disorder with behavioural abnormalities characterized by mental retardation, aggressive behaviour, and self-injurious behaviour (reviewed by (Torres and Puig 2007)). Two groups developed a mouse model lacking the Hprt gene (Hooper et al. 1987; Kuehn et al. 1987) and to their surprise, no overt phenotypes were observed. Later reports demonstrated mild defects in Hprt-deficient mice (Dunnett et al. 1989), including amphetamine-induced increased locomotor activation and stereotypic behaviour (Jinnah et al. 1991).

Hprt has since been used as location for docking constructs (Bronson et al. 1996). The location of the Hprt gene on the mouse X-Chromosome can be advantageous due to the fact that all female offspring from a male carrier would contain the modified chromosome. On the other hand, to some, the X-inactivation of constructs at this position in female mice may interfere with their research efforts. Regardless, due to a relatively neutral location (Farhadi et al. 2003), and the ability to efficiently select correctly targeted single-copy insertion clones by adding hypoxanthine-aminopterin-thymidine (HAT) supplement (Bronson et al. 1996), we chose this as a technology amenable to high-throughput generation of novel mouse strains. In this strategy the Hprt<sup>b-m3</sup> deletion in mouse ESCs is restored via a chimeric human-mouse homologous fragment reestablishing HPRT function and selecting against aberrant, non-functional clones, with the desired gene knocked-in 5′ of the restored locus (Yang et al. 2009).

1.2.2 AAV as a gene delivery vehicle for the CNS

Adeno-Associated Virus (AAV) is a small 25 nm virus that produces no known disease in humans. As the name applies, AAV requires helper functions of other viruses and was initially identified as a contaminant with adenovirus preparations (Atchison et al. 1965). Although neutralizing antibodies (NAbs) against specific serotypes can be highly prevalent (e.g. prevalence of AAV2 NAbs ranging from 30-60% worldwide) (Calcedo et al. 2009), AAV is considered to be low in immunogenicity (Snyder 1999). A review on strategies for overcoming
pre-existing immunity against AAV can be found in Jeune et al. (2013). Furthermore, long-term gene expression has also been demonstrated using recombinant AAV (Hadaczek et al. 2010).

AAV-mediated delivery of gene products thus provides an option of major interest. However, the small size of the virus and the resulting genome of 4.9 kb, means that optimized promoters and small gene products have to be used in creating recombinant viruses. This poses a problem for future gene therapy, because of the large size of mammalian promoters with additional important regulatory elements required to faithfully recapitulate endogenous expression. Although dual-vector systems have been developed to help manage this problem, there are still inefficiencies compared to that of single vector based delivery methods (Ghosh and Duan 2007). Therefore, it remains critical to have compact promoters for use in AAV.

### 1.3 Thesis objectives

#### 1.3.1 General comments

Two primary thesis objectives were set out: (1) The development of promoters for use in basic research and clinical studies, and (2) the application of said resources to study a key molecule in brain and eye development, namely NR2E1. Throughout the three data chapters several common threads exist. First, the use of Hprt docking technology enabling the genetic engineering of mice in a single-copy and reproducible fashion; second, the development and demonstration of the utility of MiniPromoters; third, the study of the neurogenesis-associated and brain development genes DCX and NR2E1; and fourth, the analysis of mouse brain and eye both for gene expression and histology.

#### 1.3.2 Development of promoters for use in basic and clinical research

Numerous studies have generated data demonstrating regional expression of human genes. But a dearth of expansive resources exist to apply these complex and yet exquisite expression patterns in order to drive the expression of specific genes of interest. By carefully reviewing data indicative of MiniPromoter regional or cell-type expression candidate source genes, we set out in Chapter 2 to determine whether large-scale bioinformatically-driven identification of putative regulatory elements based on phylogenetic footprinting in combination with a regulatory resolution score, and tested in mice, would enable the identification of brain-specific small promoters. Specifically, we anticipated that the use of human core promoters sequences in
addition to selected conserved elements would generate expression related to the endogenous source gene. Furthermore, the use of only human DNA sequence in addition to single-copy knock-in at a neutral genomic locus (Hprt) would result in stable reproducible expression. However, we did recognize that any sub-regional or otherwise specific expression pattern would have utility. We first established a regulatory resolution score to prioritize candidates (previously identified (D’Souza et al. 2008)), which were further analyzed for therapeutically-interesting localized expression, relevant mouse knock-out phenotypes, expression data including other large-scale gene expression databases, RT-PCR, and ChIP-seq data. A pipeline was established for MiniPromoter construction, mouse generation, neuroanatomical histological analysis in adult brain and dissemination of data.

Although we demonstrated retinal gene expression with one MiniPromoter and embryonic with another, it was clear that our rich resource of brain MiniPromoters also expressed in other tissues of promoter-need: the eye, spinal cord and in development. We proposed that since the eye is an outgrowth of the developing brain, and mainly consist of neuronal cells, that the majority of our MiniPromoters would be positive in the eye. Histological analysis was performed on intact eyes and in cryosectioned retinas. Spinal cord tissue, in addition to four other tissues, and a selected subset of MiniPs at E12.5, were also evaluated for lacZ expression. A second hypothesis was that our results from single-copy knock-ins of human DNA would result in reproducible expression across construct delivery mechanisms – that specificity obtained in Hprt knock-ins would translate to other systems like Adeno-Associated Virus. Three retinal ganglion cell promoters were chosen to demonstrate this, and were delivered intravitreally to the mouse retina. Marker analysis was done to demonstrate expression in retinal ganglion cells. Lastly, we evaluated whether the use of an endogenous basal promoter would be the primary driving force in determining MiniPromoter relatedness to that of the source gene by comparing positive MiniPromoters within a set for a given source gene.

### 1.3.3 Application of Pleiades MiniPromoter to study brain development

Due to its role in regulation of the cell cycle and neural stem cell differentiation, we sought to understand the effect of altered expression of human NR2E1 in both the context of the wild-type background and in the Nr2e1<sup>1<sub>low</sub></sup> (null allele) mouse. We hypothesized that ubiquitous expression of NR2E1 would result in
potent inhibition of differentiation in conjunction with severe embryonic disorganization due to increased proliferation. Furthermore, as many transcription factors have been shown to act as cell-fate conferring molecules, we proposed that ectopic expression in neuronal cells using the Ple53 (DCX-based Pleiades MiniP) could act to ameliorate the phenotypes observed in brain and eye of Nr2e1frc/frc mice. Specifically, we chose a promoter to drive expression in a cell-type that is non-overlapping with endogenous Nr2e1, but active in the type A neuroblasts (Doublecortin-positive) (Couillard-Despres et al. 2005; 2006). This choice of promoter was important as newer work has demonstrated Nr2e1 expression in subsets of type C TAPs (Obernier et al. 2011; Li et al. 2012). Furthermore, the transition from type B NSCs to type C TAPs, and then type A neuroblasts is specific to the SVZ neurogenic niche and is not observed in the RMS where the transition is directly to type A neuroblasts (Doetsch et al. 1997). We analyzed ubiquitous expressing embryonic stem cells in an assay that measures cell viability, generated chimeras containing these cells, and studied the embryonic lethal phenotype of germline offspring. In addition, we genetically engineered mice expressing NR2E1 via a Doublecortin-based MiniP from Chapters 2 and 3, and analyzed survival and potential for rescuing Nr2e1frc/frc phenotypes by brain weight and retinal funduscopy.
Chapter 2: A regulatory toolbox of MiniPromoters to drive selective expression in the brain

2.1 Preamble

Disorders of the central nervous system are in dire need of better therapeutics. Unlike current pharmaceuticals, which merely attempt to treat symptoms, gene therapy offers the promise of a potential cure for many debilitating disorders. However, the diverse and highly heterogenous transcriptional landscape of the brain makes it difficult to properly target gene-based therapeutics. Thus, the Pleiades Promoter Project was born to alleviate the problem of targeting via development and testing of novel small promoter constructs of human DNA origin that can be used in future translational research. In this chapter we present the first publication from this project describing our novel promoter tools.

2.2 Introduction

The mammalian central nervous system (CNS) is a complex entity comprising diverse neuronal and non-neuronal cell types. The organization of these cell types, as illustrated in brain atlases (e.g., (Sidman et al. 1971)) and gene expression maps (e.g., (Lein et al. 2007)), is regional within the brain. The Allen Brain Atlas maps the expression of ≈20,000 genes in the adult mouse brain using in situ hybridization (Lein et al. 2007). A similar initiative, the GENSAT project, used large-insert, random-insertion transgenic mouse techniques to profile gene expression patterns in the CNS (Gong et al. 2003). These resources demonstrate the selectivity with which some genes are expressed in the brain. Such specific expression patterns can be driven by modular conserved regulatory regions (RRs) (Carroll 2005). Importantly, patterns of gene expression within the brain tend to be conserved between human and mouse (Strand et al. 2007).

For decades, researchers have struggled to develop tools that direct specific patterns of gene expression in the brain. There is an increasing availability of genetic applications that are predicated upon

---

targeted gene expression in vivo that require cell-specific promoter constructs to restrain the effect to a region of interest. Despite great demand, the discovery of regulatory sequences to drive gene expression in a region-specific manner in the brain has been slow, primarily arising from low-throughput promoter-deletion studies (e.g., L7/Pcp2, (Anderson et al. 1997) and Camk2a, (Mima et al. 2001)).

The Pleiades Promoter Project addressed this challenge using a higher-throughput, bioinformatically directed parallel design process, producing MiniPromoters (MiniPs) vetted in vivo in the adult mouse brain. The Pleiades Promoter Project adopted a three-step strategy wherein we first identified genes specifically expressed in adult brain regions or cell types of therapeutic interest, then we computationally predicted the human RRs responsible for the specific expression and, ultimately, tested compact MiniPs in vivo in knock-in mice. MiniPs contain human DNA sequences to facilitate studies in human cells; they do not use, for example, any prokaryotic sequences thereby reducing the potential for epigenetic inactivation. The Pleiades resource of brain MiniPs has expanded dramatically upon the collection of existing promoters for the study of the brain and is publicly available at http://www.pleiades.org. The resource includes a comprehensive dataset of regulatory information on brain-specific genes, the MiniP designs and sequences, the plasmids for Hprt targeting, the targeted embryonic stem cells, and the mouse strains to allow researchers comprehensive access to the data and tools.

2.3 Materials and methods

The Pleiades Promoter Project is characterizing 128 MiniPs over a 4-year time frame. The pipeline, which involved five specialized laboratories, is summarized below. The general methods used have been described previously (Yang et al. 2009).

MiniP sequences were computationally designed and assembled into DNA molecules at a rate of four per week. The construct DNA was electroporated, typically into B6129F1 ESCs (mEMS1202 or mEMS1204, (Yang et al. 2009)), at a rate of seven per week (including controls). A total of 10–15 clones per construct were picked, expanded, and PCR verified, to obtain approximately four correctly targeted ESC lines per construct. ESCs were typically microinjected into E3.5 blastocysts from BAN2, [N2 backcross of B6(Cg)-Tyr-c2J (JAX stock 000058) onto CD-1 (Charles River strain 022)], a strain combination selected for high blastocyst yield. Germline females were backcrossed to C56BL/6J (JAX stock 00664). The brains of N2–N3 germline males (8–12 wk) were analyzed using
histochemical procedures. A minimum of three brains were processed for each MiniP. Every brain was cryosectioned (20-μm sections at 640-μm intervals) from medial to lateral in the sagittal plane and prepared for brightfield immunodetection of EGFP or lacZ histochemistry. When reporter expression was absent in at least three adult brains, a MiniP strain was classified as negative. Positive MiniP strains underwent further histological analyses to define the cellular pattern of gene expression. Both positive and negative strains were prepared for presentation on the Internet at http://www.pleiades.org.

Appendix A: Dataset S1 provides the list of the mouse strains described in this paper. ESC lines and mouse strains are available at the Mutant Mouse Regional Resource Centers (http://www.mmrrc.org/). All procedures involving animals were in accordance with the Canadian Council on Animal Care (CCAC) and UBC Animal Care Committee (ACC) (protocol no. A05-1258 and A05-1748).

2.3.1 Regulatory resolution score characterization

Identification of genomic sequence boundaries for regulatory resolution scoring – For regulatory resolution scoring of each gene in the Ensembl human genome database (version 46) (Flicek et al. 2007), we defined the transcription start site (TSS) as the start position of the 5′-most exon annotated for the gene. We then determined the boundaries of the region to be analyzed relative to the TSS as follows. In most cases, the upstream boundary was defined as the start/end position of the upstream gene (depending on the upstream orientation of the gene). If the upstream gene was less than 1 kb from the TSS of the gene of interest, we extended the analysis to introns of the upstream gene located within 10 kb of the TSS. In most cases, the downstream boundary was the end of the gene of interest. However, if the gene was longer than 30 kb from the TSS to the end of the last exon, intronic regions within 30 kb downstream of the TSS were used. Conversely, we included 2 kb of sequence downstream of the TSS for genes shorter than 2 kb in length.

Non-exonic conserved regions – PhastCons scores and PhastCons “conserved elements” computed from comparisons of 17-way vertebrate multi-species alignments (Siepel 2005) were downloaded from the UCSC Genome Browser database (Karolchik et al. 2008). Only PhastCons conserved elements that were both 20 bp or longer and non-overlapping with annotated human mRNAs or Ensembl human gene annotations were retained.
for analysis. PhastCons conserved elements separated by less than 100 bp are chained together (excluding the intervening regions) and thereafter considered part of a single longer conserved region.

Score definition – We defined a raw regulatory resolution score as:

\[
raw \ score = \log_{10} \left( \frac{\sum_{i=1}^{n} l(c - b)}{n^2} \right)
\]

where \( l \) is the length of the conserved region, \( c \) is the “conservation level” of the conserved region (i.e., the mean PhastCons score for the conserved region), \( b \) is the baseline conservation level (i.e., the mean PhastCons score for the entire genomic segment analyzed) and \( n \) is the number of conserved regions. Thus, for each conserved region, we considered the amount of conserved sequence, how well distinguished the region was from the background, and we penalized genes with many conserved regions. After computing the raw score, we normalized it to obtain a value between 0 and 1 using the following formula:

\[
normalized \ score = \frac{raw \ score - \min \ raw \ score}{\max \ raw \ score - \min \ raw \ score}
\]

Thus, zero identified a gene with little resolution and 1, one that was highly resolved.

Genome-wide distribution of regulatory resolution scores – Regulatory resolution scores were computed for all human genes as reported in the Ensembl annotations (Fig. 2.1). Of 22,298 genes tested, 2,411 did not contain any conserved PhastCons elements (Fig. 2.1B). Therefore, we could not compute regulatory resolution scores for these genes. The distribution of scores was skewed towards zero, with a median score of 0.34 and a mean score of 0.36 (Fig. 2.1A). Genes with up to 5 conserved regions received higher scores (Fig. 2.1C), with the top 20\textsuperscript{th} percentile having an average of 2.2 conserved regions per gene. The highest scores were assigned to genes with less than 1,000 bp of conserved non-exonic nucleotides (Fig. 2.1D), with an average of 330 bp of conserved sequence per gene for genes scoring within the top 20\textsuperscript{th} percentile.

Features of genes with the highest, average, and lowest regulatory resolution – The ADCK5 locus was assigned the highest regulatory resolution score due to the presence of a single 1,277 bp highly conserved region within the upstream intergenic region (Fig. 2.2A). Two smaller conserved regions directly upstream of
ADCK5 in the “17-Way Most Cons” track were excluded from the analysis. The larger of the two overlaps with human mRNAs, while the smaller conserved element is only 10 bp long (Fig. 2.2A). The low baseline conservation level across the entire region further contributed to the high score. The ELOVL3 locus (Fig. 2.2B) is an example of an average gene, receiving the mean score of 0.35. It contains four small, conserved non-exonic regions containing a total of 186 bp of sequence within the boundaries of the analysis. The lowest scoring gene, NR4A3, features 90 small conserved regions, containing a total of 9,281 bp of non-exonic sequence, that are distributed across the entire locus (Fig. 2.2C). The majority of the NR4A3 locus is conserved and the conservation profile revealed few insights into the location of potential regulatory regions for targeted promoter construct design.

**Manual promoter curation** – Promoters for 100 genes were manually assessed based on a number of gene features, including: (i) the location of the transcription start points, (ii) the boundaries of analysis, i.e., the amount of non-coding sequence to be analyzed upstream and downstream of the gene of interest, and (iii) the number and qualitative conservation level of conserved regions located proximal to the TSS within the defined boundaries. The genes were ranked from 1 to 5 based on the curators’ perception of their suitability for MiniPromoter (MiniP) design, “1” being a gene not suitable and “5” a very good candidate.
Figure 2.1 Genome-wide distribution of regulatory resolution scores. 

(A) Histogram of scores. (B) Summary statistics showing the score by quartiles (Qu.), as well as the median and mean score. NA, not able to score. (C) Boxplot showing the distribution of the number of conserved regions by score intervals. (D) Boxplot showing the distribution of the number of conserved bases by score intervals. The boxes in both boxplots are drawn with widths proportional to the square roots of the number of observations in the groups.
Figure 2.2 Genes with (A) the highest, (B) average and (C) the lowest regulatory resolution scores.
Each screenshot from the UCSC Genome Browser (http://genome.ucsc.edu; NCBI Build 36.1) displays: conserved non-exonic (CNE) PhastCons elements used in the analysis; UCSC gene predictions based on RefSeq, GenBank and UniProt data; transcripts for Ensembl genes based on mRNA and protein evidence; a dense display of human mRNAs from GenBank; CpG islands (≥50% GC content, ≥200 bp in length, and an observed CG to expected CG ratio ≥ 0.6); evolutionary conservation in 17 vertebrates based on Multiz alignments and PhastCons scores; and predictions of conserved elements produced by the PhastCons program (17-way Most Cons).

2.3.2 MiniPromoter design

The MiniP design pipeline is represented below and includes the following resources: PubMed (http://www.ncbi.nlm.nih.gov/pubmed), PAZAR (http://www.pazar.info), the UCSC genome browser (http://genome.ucsc.edu), ORCAtk (http://www.cisreg.ca/cgi-bin/ORCAtk/orca), the VISTA enhancer browser...
(http://enhancer.lbl.gov) and histone modification ChIP (chromatin immunoprecipitation) assays performed on mouse and human cortex (Jones et al. unpublished).

The endogenous promoters of genes were identified using genome annotations for 5′ cap analysis gene expression (CAGE) tags (Shiraki et al. 2003), transcripts (mRNAs, ESTs), and CpG islands (Gardiner-Garden and Frommer 1987). The boundaries for analysis were defined similarly to the regulatory resolution score analysis (see above), except that if one of the neighboring genes has an expression pattern similar to the gene of interest, the boundaries were extended to include the surrounding sequences of this additional gene. In a few cases, the GENSAT project had generated and tested BAC mice for the gene of interest and the expression pattern reported matched the endogenous expression pattern (Gong et al. 2003). In such cases, the BAC sequence defined the boundaries for regulatory sequence analysis.
Phylogenetic footprinting and transcription factor binding site (TFBS) prediction were performed using the ORCA toolkit (Portales-Casamar et al. 2009b) and the following steps:

- Retrieval and alignment of human and mouse orthologous sequences within the defined boundaries;

- Computation of the non-coding conserved regions above a user-defined threshold (ranging from 50 to 85% identity in our analyses);

- Prediction of TFBS in those conserved regions for the transcription factors that have been described to be relevant for the expression of this specific gene or for expression in the brain region of interest in general. The TF binding models were extracted from the JASPAR database (Sandelin 2004) or custom-generated from the PAZAR database based on the manually curated “Pleiades Genes” project (Portales-Casamar et al. 2007).

2.3.3 Hprt targeting vectors and MiniPromoters

The Hprt targeting vectors used in this study were pEMS1306 (EGFP reporter (Yang et al. 2009)), pEMS1313 (lacZ reporter), and pEMS1302 (EGFP/cre reporter). The pEMS1313 and pEMS1302 fragments from the multiple cloning site (MCS) to the end of the reporter gene were synthesized by GeneArt (Regensburg, Germany) and cloned into the Hprt targeting plasmid pJDH8A/246b (Heaney et al. 2004) using EcoRI restriction sites.

MiniPs typically comprised up to 4 distinct genomic segments joined by fusion PCR. Each genomic segment was first PCR-amplified independently using AccuPrime Pfx DNA Polymerase (Invitrogen, Carlsbad, U.S.A.), PCR primers (Integrated DNA Technologies, San Diego, CA), and BAC DNA template (10 pg to 200 ng). PCR primers for the outermost 5’ and 3’ segments were tailed with the appropriate restriction sites to allow for cloning. For MiniPs with two segments or more, PCR products of upstream segments were 3’ tailed with 18 bp linkers homologous to the first 18 bp of the adjacent downstream genomic segment. Reaction conditions were 0.25 Unit enzyme, 1x AccuPrime Pfx reaction mix, 1.0 μM each primer mix in a 20 μl volume. A 2-minute denaturation at 95 ºC was followed by 30 cycles of 95 ºC for 15 seconds, 30 seconds (at Tm corresponding to primer pair) and 68 ºC for 90 seconds, plus a final extension at 68 ºC for 10 minutes. The PCR reaction was run on a 1% low melting point agarose gel, visualized using SYBR Green (Invitrogen), excised and recovered from the gel using QIAquick gel extraction kit (Qiagen, Germantown, MD). Reaction products were eluted using 30 μl of Ultrapure water (Gibco Invitrogen) then quantified using the NanoDrop (Thermo Fisher Scientific, Waltham, MA).
For MiniPs with multiple elements, fusion PCR was performed as above, but using 2.0 μl of gel purified first round reaction products (10 pg to 200 ng). Additional binary fusions were executed as above until the full-length was obtained. A subset of 9 MiniPs was generated by direct synthesis at GeneArt.

The final MiniPs were cloned into one of our Hprt targeting vectors and sequence verified with primers located every 300 bp along the construct on both strands. All discrepancies between the designed and constructed sequences were inspected using the UCSC Genome Browser annotations (hg18) (Karolchik et al. 2008). We tolerated discrepancies if they were known polymorphisms, located in a non-conserved region (PhastCons Vertebrate Multiz Alignment & Conservation (17 Species) score below 0.7), or if analysis did not show any further regulatory implication. We rejected any sequence with insertion or deletion bigger than 10 bp.

2.3.4 Knock-in immediately 5’ of the Hprt locus

The mEMS1204 (B6129F1-Gt(ROSA26)Sor^m1Sor/+, Hprt^-m3/Y), mEMS1202 (B6129F1-Gt(ROSA26)Sor^/+), and E14TG2A (Hooper et al. 1987) ESC lines were electroporated with constructs built in pEMS1302, pEMS1306, or pEMS1313, respectively. Clones were maintained under HAT selection for 3-4 days of expansion in 96 well plates and then transferred to 2 x 24 wells and cultured in HT media. Once cells reached confluence, both wells were frozen in HT-freeze media and stored in liquid nitrogen (LN2).

2.3.5 PCR analysis of genomic DNA

Vector NTI (http://www.invitrogen.com) software was used to design PCR assays for the different constructs. MiniP-specific PCR genotyping assays are available on the http://www.pleiades.org website.

2.3.6 In vitro neural differentiation

Neural differentiation of ESCs was conducted as previously described (Barberi et al. 2003), with the following modifications. Once confluent, ESCs were trypsinized and seeded in duplicate wells onto confluent MS-5 feeder layers at 500 cells/cm² for seven time-points. Total cell RNA was extracted with the RNeasy Plus Mini Kit (Qiagen) and used in RT-PCR analysis in both +RT and –RT conditions, using the OneStep RT-PCR Kit (Qiagen) according to manufacturer’s instructions. Ple53-EGFP immunohistochemistry was performed on day 11 of differentiation. Cells were washed once with 1x PBS and fixed using 4% paraformaldehyde in PBS for 15 minutes at room temperature. Cells were then blocked using Image-iT FX signal enhancer (Invitrogen) reagent and
subsequently incubated with 1:1000 rabbit polyclonal anti-GFP antibody (Abcam, Cambridge, MA) followed by 1:1000 Alexa-488 secondary anti-rabbit antibody (Invitrogen). Cells were imaged on a Zeiss Axiovert 200M microscope at 20x with the FITC filter set. Ple88-lacZ staining was performed as outlined at http://openwetware.org/wiki/LacZ_staining_of_cells, on day 14 of differentiation. Brightfield images were taken with the 10x objective on an Olympus Bx61 microscope.

2.3.7 Immunohistochemistry and histochemistry

The mice were anesthetized with avertin and perfused transcardially with 4% paraformaldehyde (PFA) in 0.1 M phosphate buffer (pH7.4) for 15 min, brains were dissected and post fixed in the same fixation solution for 2 hours and transferred into 25% sucrose-PBS overnight. Each brain was sectioned in a cryostat and 20 µm sagittal sections were collected. Enhanced Green Fluorescent Protein (EGFP) expression was detected with anti-GFP using the Vectastain Elite ABC kit (Vector Labs, Burlingame, CA) and DAB, as a brown chromogen, following the manufacturer’s directions. Expression of the beta-galatosidase (lacZ) or the EGFP/cre fusion protein (following recombination of the Gt(ROSA)26Sor<sup>tm1Sor</sup> locus (Soriano 1999)) was detected with X-gal (5-Bromo-4-chloro-3-indolyl-ß-D-galactopyranoside) staining as previously described (Reiner et al. 2007). High resolution serial images of brightfield material were acquired using a Nikon Optiphot-2 microscope with a LEP motorized stage connected to a Dell Precision 390 computer equipped with hardware and software from MicroBrightField, Inc. Images were captured and tilled using MBF Neurolucida Virtual Slice v8.2.3.0.

Double-label immunofluorescence for colocalization of EGFP and endogenous proteins was performed as previously described (Liu et al. 2007). Either native EGFP fluorescence (nGFP) or anti-GFP detection with an Alexa-488 secondary antibody was combined with a second primary antiserum and detection with a Cy3 or Alexa-555 secondary antibody. Co-staining of lacZ activity and tyrosine hydroxylase or NeuN was performed sequentially as previously described (Liu et al. 2007). Primary antibodies used for these studies included: rabbit anti-DCX (1:500, Cell Signaling), rabbit anti-orexin (1:500, Millipore, Billerica, MA), mouse anti-GFAP (1:1000, Millipore), mouse anti-S100β (1:1000, Abcam), mouse anti-NeuN (1:500, Chemicon (Millipore), Billerica, MA), mouse anti-TH (1:3000, Chemicon), mouse anti-RIP (1:500, Chemicon). Secondary antibodies included: goat anti-rabbit Alexa-488 (1:500, Molecular probes (Invitrogen)), goat anti-rabbit-Cy3 (1:500, Jackson ImmunoResearch
Laboratories, West Grove, PA), goat anti-mouse Alexa-555 (1:500, Molecular probes), goat anti-mouse Alexa-488 (1:500, Molecular probes), donkey anti-goat-Cy3 (1:500, Jackson ImmunoResearch Laboratories). Sections were counterstained with TOTO3 (2 µM, Molecular Probes) and mounted with anti-fade reagent FluroSave Reagent (Calbiochem, San Diego, CA). Detection of double immunofluorescence was performed using a BioRad confocal laser-scanning microscope (CLSM, BioRad, Hercules, CA). Whole-mount X-gal histochemistry was performed on 4% PFA-fixed embryos (E10.5, E11.5) or dissected brains (E15.5, P0.5) following a similar protocol described above after preincubation of the tissue in 0.1 M PBS containing 0.3% Triton X-100. Stained embryos and brains were photographed, cryosectioned, and counterstained with neutral red for localization of lacZ-expressing cells.

2.3.8 Histochemistry of 1 mm brain slices

Mice were perfused with 4% PFA, and post-fixed 2-4 hours as described above. The brains were then removed from 4% PFA and immediately sectioned. The dissected brains were placed ventral side up into the adult mouse coronal or sagittal, Rodent Brain Matrix (ASI Instruments, Warren, MI). Slices were sectioned lateral through medial to lateral (for sagittal) or rostral to caudal (for coronal) using single-edge razor blades (Electron Microscopy Sciences, Hatfield, PA). All slices from one brain were placed into one well of a 12-well plate containing 1X PBS (Invitrogen, Carlsbad, CA), or 0.1% sodium azide (in 1X PBS) until staining was performed. The staining was performed with 3-5 ml of X-gal staining solution (25 mg/ml X-gal, 1 M MgCl2, 50 mM potassium ferri-cyanide, 50 mM potassium ferro-cyanide, and 1X PBS to volume) per well in a 24-well plate. The plate was wrapped in aluminum foil and incubated at 37°C for 10-16 hours. Subsequently, the sections were transferred into PBS, examined under a dissecting microscope, and photographed using a CoolSNAP-Procf color camera (Media Cybernetics, Bethesda, MD) mounted on a Leica MZ12.5 Stereomicroscope (Leica Microsystems, Wetzlar, Germany) and Image-Pro Express v.4.5.1.3 software (Media Cybernetics, Inc., Bethesda, MD).

2.3.9 Regulatory element predictions in OLIG1 enhancer sequences

Identification of “most conserved” aligned sequences in OLIG1 construct sequences – The genomic coordinates for each of the conserved regions constituting the tested OLIG1 MiniPs were retrieved using the BLAT sequence search tool at the UCSC browser against the Human March 2006 assembly (Karolchik et al. 2008). The genomic coordinates were used as input to the UCSC Table Retrieval function to extract the human
sequence alignment in the 17-way multiple mammalian-species and 28-way placental mammals “most conserved” alignments (“17-way” and “28-way” respectively in Table 2.1) and each aligned human sequence (with gaps) was stored in FASTA format.

Table 2.1 List of predicted TFBS that were unique to the positive Ple151 construct.

<table>
<thead>
<tr>
<th>TF Predictions</th>
<th>TFs Differentially Expressed (U34A-C Chips)</th>
<th>17-way 80%</th>
<th>17-way 75%</th>
<th>28-way 80%</th>
<th>28-way 75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>No</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CREB1</td>
<td>No</td>
<td></td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DDIT3 (JP: Ddit3-Cebpa)</td>
<td>No</td>
<td></td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2F1</td>
<td>Probe mapping not available</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>EGR1</td>
<td>Yes - Down-regulated from OPCs over time and up-regulated at D7-D9</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGR4</td>
<td>No</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELK1</td>
<td>No</td>
<td></td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELK4</td>
<td>Probe mapping not available</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EVI1</td>
<td>No</td>
<td></td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOXD1</td>
<td>No</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOXF2</td>
<td>Probe mapping not available</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GABPA</td>
<td>Probe mapping not available</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLI1</td>
<td>Probe mapping not available</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HINF (JP: MIZF)</td>
<td>Probe mapping not available</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLF</td>
<td>No</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRF2</td>
<td>No</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>MAX</td>
<td>No</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFAT1</td>
<td>Yes - Down-regulated between OPC and D2</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFYA</td>
<td>No</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NKX2-2</td>
<td>Probe mapping not available</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NKX3-1</td>
<td>Probe mapping not available</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NKX3-2 (JP: Bapx1)</td>
<td>Probe mapping not available</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NKX3C1</td>
<td>Yes - Down-regulated between D9 and acute OLs</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>PAX4</td>
<td>No</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAX5</td>
<td>Probe mapping not available</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>POU2F1</td>
<td>No</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>RORA (JP: RORA_2)</td>
<td>Probe mapping not available</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>RREB1</td>
<td>No</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RXRA (JP: RXRA-VDR)</td>
<td>No</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>STAT1</td>
<td>Yes - Up regulated between OPC and D2</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T brachyury</td>
<td>Probe mapping not available</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEAD1</td>
<td>Probe mapping not available</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZNF143 (JP: Staf)</td>
<td>Probe mapping not available</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>ZNF423 (JP: Roax)</td>
<td>No</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

JP: Jaspar profile name - included if different than TF HUGO gene name; OPC: oligodendrocyte progenitor cells; D2: day 2 time point in the Dugas et al. dataset; D9: day 9 time point in the Dugas et al. dataset; OLs: oligodendrocytes.
TFBS predictions in “most conserved” sequences – Each of the conserved regions making up a MiniP was subjected to a TFBS prediction analysis. A PERL script was developed using the TFBS PERL module (Lenhard and Wasserman 2002) and the JASPAR CORE database (Sandelin 2004) (supplemented with additional model annotations for Glia-related TFBS: POU2F1; EGR1; EGR2; EGR3; EGR4; POU3F1; NKX2-2; NKX2-5) to evaluate vertebrate TFBS models across the most conserved sequence elements of each region using profile score threshold levels of 75% and 80% (“75%” and “80%” respectively in Table 2.1). TFBS predictions were written to BED formatted files for each analyzed alignment.

Analysis of in vitro oligodendrocyte gene expression profile data – To identify TF candidates that could be directing OLIG1 regulation in oligodendrocyte cells, we analyzed an in vitro 8-day time point oligodendrocyte differentiation dataset produced by Dugas et al. (2006). This dataset is comprised of recorded gene expression profiles across a timescale of differentiated, purified, rat cortical oligodendrocyte progenitor cells (OPC) using the Affymetrix RG_U34-A, RG_U34-B, and RG_U34-C chips. 96 Affymetrix CEL files (8 time points x 4 biological replicates x 3 chips) were obtained from J.C. Dugas. We developed R code (http://www.R-project.org) and used the Bioconductor packages (Gentleman et al. 2004) to perform a robust multichip analysis (RMA) (Irizarry et al. 2003) on each chip dataset to obtain a probe-level summarization. All pairwise experiments were subjected to a two-sample T-test with a random variance model (Wright and Simon 2003) implemented in the BRB-array software (http://linus.nci.nih.gov/~brb/). The Rat Affymetrix chip probes were mapped to Entrez rat genes using Bioconductor packages. The rat Entrez Genes were mapped to mouse Entrez Genes (where possible) using Homologene (Wheeler et al. 2008). A set of mouse TF genes (Fulton et al. 2009) was mapped to the rat Affymetrix probes. PERL software was written to convert all HTML-formatted expression analysis results to text files and extract and report all significantly (P-value ≤ 0.001), differentially expressed genes across the pairwise expression profiles and mapped TF genes in this set were identified.

Evaluation of TFBS predictions – TFBS predictions in the positively expressed MiniP construct (Ple151) and the MiniP constructs that had no reporter gene expression (Ple148 and Ple150) were compared using a PERL script to identify the TFBS predictions that were unique to the expressed MiniP. These unique TFBS predictions were then compared against the expression profile analysis results (Table 2.1).
Prioritization of candidate TFBS – The compiled TFBS predictions and expression data analyses were reviewed to rank the TFBS candidates. TFBS predictions that were unique to the positive Ple151 construct with differential gene expression and correlated literature evidence support were reported (Table 2.2).

Table 2.2 Predicted TFBS candidates with differential gene expression and literature evidence support.

<table>
<thead>
<tr>
<th>TF Predictions</th>
<th>TFs Differentially Expressed (U34A-C Chips)</th>
<th>Literature Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGR1</td>
<td>Yes - EGR1 (KROX-24) expression is down-regulated after OPC time point</td>
<td>EGR1 (KROX-24) may be involved in the initial oligodendrocyte differentiation primary response [20]</td>
</tr>
<tr>
<td>FOS</td>
<td>Yes - FOSL1 (FRA-1) expression is down-regulated between OPC and D2</td>
<td>The AP-1 family of TFs may play a role in oligodendrocyte differentiation [21]</td>
</tr>
</tbody>
</table>

OPC: oligodendrocyte progenitor cells; D2: day 2 time point in the Dugas et al. dataset.

2.4 Results

2.4.1 The Pleiades Promoter Project approach

The goal of the Pleiades Promoter Project strategy is to computationally identify RRs responsible for the selective expression observed in manually selected brain-region or cell-type enriched genes (D’Souza et al. 2008), specifically focused on assembling RRs that drive selective expression patterns. Once candidate RRs are identified, the strategy was to design up to seven MiniPs per gene, each including a different combination of one to six regions. To produce physiological levels of expression and define a MiniP entirely derived from human DNA sequences, we used the endogenous promoter of the gene instead of an exogenous promoter. MiniPs of interest need to be small (≤4 kb) so that they can be easily manipulated and suitable for most space-restricted molecular constructs. Identified MiniP sequences are then cloned 5′ of a reporter gene, such as EGFP, lacZ, or an EGFP/cre fusion protein, and then introduced by homologous recombination immediately 5′ of the Hprt locus on the X chromosome of mouse ESCs, as previously described (Bronson et al. 1996; Yang et al. 2009), to provide a single-copy knock-in insertion to enable reproducible, predictable expression for subsequent analyses (Farhadi et al. 2003).
2.4.2 Validation of the approach using a set of previously characterized promoters and ESC neural differentiation.

We selected the CAG ubiquitous promoter as well as five previously characterized multicopy random-insertion promoters that drive expression in a subset of neurons or glial cells, generated vector constructs, and targeted insertion to the \( Hprt \) locus (Appendix A: Dataset S1 “Control”). The CAG promoter drives ubiquitous reporter gene activity, demonstrating that widespread, adult expression can be achieved from the locus, consistent with previous studies (Yang et al. 2009).

Initially, we adopted an ESC-differentiation strategy that generates NSCs within 6 days, as confirmed by gene expression patterns, as a test of cell-type specificity from the control promoters inserted 5’ of \( Hprt \) (Barberi et al. 2003). Fig. 2.3 shows the characterization of Ple88 (\( GFAP \) RRs) for glial expression and Ple53 (\( DCX \) RRs) for neuronal expression (Fig. 2.3A and E), selected because both GFAP and DCX are selectively expressed during CNS development and therefore the promoters are likely to drive expression in the differentiation assay. Expression of the MiniPs follows a pattern similar to the corresponding endogenous genes, as shown by RT-PCR (Fig. 2.3B and F) and staining (Fig. 2.3C and G). The positive ESC pre-screening results were then confirmed in the corresponding knock-in adult mouse brains (Fig. 2.3D and H). This validation test, using these two previously characterized promoters, demonstrates the suitability of the Pleiades Promoter Project approach, using single-copy knock-in 5’ of the \( Hprt \) locus, for the characterization of brain MiniPs.

The five selected control promoters were further characterized in transgenic animals by neurohistological analysis. As detailed in Appendix A: Dataset S1 “Control” and Fig. 2.4, four of the five controls reproduced the expected expression patterns. The last control MiniP, Ple59 (\( DDC \) RRs), did not produce detectable reporter gene expression. As the original study characterizing this construct reports only one of six founder lines expressing the transgene, we concluded that favorable conditions at that site in the genome may permit expression and such insertion-site properties are not present at the other five locations nor at \( Hprt \) (Chatelin et al. 2001). In conclusion, the analysis of these previously characterized brain-specific promoters validates our knock-in strategy 5’ of the \( Hprt \) locus as a tool to monitor brain-specific expression.
Figure 2.3 In vitro neural differentiation for pre-screening MiniP designs. (A, E) Ple53 and Ple88 are cloned upstream of EGFP or lacZ, respectively. (B, F) RT-PCR assays across 7 time points of ESC neural differentiation for endogenous and reporter genes demonstrate appropriate temporal expression. (C, G) Immunohistochemistry or X-gal staining demonstrates appropriate spatial expression. Scale bars, 100 μm (C), 200 μm (G). (D, H) Germline knock-in adult mouse brain sagittal sections analyzed by immunohistochemistry or X-gal staining confirms expression in the appropriate regions (i.e., olfactory bulb and rostral migratory stream for Ple53, and glia throughout the brain for Ple88).
Figure 2.4 Targeting of previously characterized MiniPromoters to the Hprt locus validates the Pleiades Promoter Project approach.

Neuro-histological analysis of four control strains carrying previously characterized random-insertion constructs knocked-in at the Hprt locus. EGFP is detected using anti-GFP immunocytochemistry (brown signal in brightfield images) and lacZ is detected using X-gal histochemistry (blue signal counter-stained with neutral red in brightfield images). Cb, cerebellum; Ctx, cortex; Hip, hippocampus; Hyp, hypothalamus; LC, locus coeruleus; RMS, rostral migratory stream. A. Ple48-lacZ expression (previously characterized construct based on DBH regulatory regions) is enriched in noradrenergic cells in the locus coeruleus and the adrenal gland but also present in other regions such as the cortex. The last image shows no co-staining of beta-galactosidase activity (blue) with tyrosine hydroxylase (brown) in the locus coeruleus. B. Ple53-EGFP (previously characterized construct based on DCX regulatory regions) expression is observed in multiple regions of the brain as seen on the whole brain image with enrichment in the olfactory bulb and the rostral migratory stream. The last image shows co-staining (yellow) of EGFP (green) with the endogenous Dcx protein (red). C. Ple88-EGFP (previously characterized construct based on GFAP regulatory regions) is expressed in astrocytes throughout the brain. The last image shows co-staining (yellow) of EGFP (green) with the endogenous Gfap protein (red); the nuclear counterstain is TOTO3 (blue). D. Ple112-EGFP (previously characterized construct based on HCRT regulatory regions) is specifically expressed in a cluster of hypothalamic cells. The last image shows co-staining of EGFP (green) with the endogenous Hcrt protein (red); the nuclear counterstain is TOTO3 (blue). Scale bars, 100 µm.

2.4.3 Brain region-specific gene selection using a regulatory resolution score

To generate new MiniPs expressing with a cell type or regional specificity in the brain, we first identified, using a genomewide approach, 237 candidate genes with region-enriched expression patterns that included 30
target adult brain regions of therapeutic interest (D’Souza et al. 2008). To narrow the list of genes, we took into account the potential relevance of each gene to human disease and, importantly, the predicted suitability of the gene for MiniP design. For disease relevance, we reviewed the literature and, when available, examined the phenotypic consequences of mouse gene knockouts. The suitability for MiniP design highlights those genes in which RRs are more easily distinguished. Comparative sequence analysis, or phylogenetic footprinting, has proven useful in delineating RRs with the expectation that sequences under selective pressure are more conserved than those that are not. We thus based our gene prioritization on the following criteria: (i) the existence of known RRs within a gene responsible for a expression pattern in the brain, (ii) transcript evidence supporting the presence of a single transcription initiation site, (iii) the length of sequence to be analyzed, (iv) the number of conserved regions between the analysis boundaries, and (v) how well distinguished conserved regions are relative to the overall conservation level for a gene. Thus, we sought genes containing a small number of well-defined conserved noncoding regions close to the transcription start site. To this end, we developed a “regulatory resolution score” intended to reflect human perception of what constitutes a good candidate gene for MiniP design (Materials and Methods and Figs. 2.1 and 2.2). The scoring procedure captures aspects of the manual curation process as demonstrated by comparison with scores manually assigned for 100 curated genes (Fig. 2.5A and Materials and Methods). The 57 genes selected for MiniP design are heavily skewed toward higher resolution scores in our set of 237 brain region-selective genes (Fig. 2.5B and C).
Figure 2.5 Regulatory resolution score prioritizes genes for MiniP design.

(A) Score distribution for 100 manually curated genes. The width of the boxes are proportional to the number of observations in the groups. The increases in scores from “1” to “4” and “5” are significant (p = 1.4e-03 and p = 7e-04, respectively; Wilcoxon test), as well as from “2” to “5” (p = 4.5e-02; Wilcoxon test). (B) Score frequency of the selected 57 genes (black) compared to all other brain region selective genes (white). The dotted grey line shows that the proportion of MiniP genes relative to total increases with the score (linear regression; individual values are marked as grey boxes). (C) Regulatory resolution scores for the genes selected for MiniP design.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Score</th>
<th>Gene</th>
<th>Score</th>
<th>Gene</th>
<th>Score</th>
<th>Gene</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADORA2A</td>
<td>0.52</td>
<td>FIBCD1</td>
<td>0.23</td>
<td>NTSR1</td>
<td>0.36</td>
<td>SLC7A5</td>
<td>0.47</td>
</tr>
<tr>
<td>ATP6V1C2</td>
<td>0.41</td>
<td>GABRA6</td>
<td>0.21</td>
<td>OLIG1</td>
<td>0.49</td>
<td>SLITRK6</td>
<td>0.39</td>
</tr>
<tr>
<td>AVP</td>
<td>0.65</td>
<td>GAL</td>
<td>0.61</td>
<td>OXT</td>
<td>0.51</td>
<td>STX1A</td>
<td>0.25</td>
</tr>
<tr>
<td>C8orf46</td>
<td>0.31</td>
<td>GCHFR</td>
<td>0.48</td>
<td>PCP2</td>
<td>0.48</td>
<td>TAC1</td>
<td>0.33</td>
</tr>
<tr>
<td>CARTPT</td>
<td>0.29</td>
<td>GPR88</td>
<td>0.52</td>
<td>PITX3</td>
<td>0.37</td>
<td>TAC3</td>
<td>0.44</td>
</tr>
<tr>
<td>CCKBR</td>
<td>0.33</td>
<td>GPX3</td>
<td>0.42</td>
<td>PKP2</td>
<td>0.16</td>
<td>TBR1</td>
<td>0.37</td>
</tr>
<tr>
<td>CCL27</td>
<td>0.32</td>
<td>GRP</td>
<td>0.32</td>
<td>POGZ</td>
<td>0.15</td>
<td>THY1</td>
<td>0.36</td>
</tr>
<tr>
<td>CD68</td>
<td>0.40</td>
<td>HAP1</td>
<td>0.41</td>
<td>RAMP3</td>
<td>0.59</td>
<td>TNNT1</td>
<td>0.31</td>
</tr>
<tr>
<td>CLDN5</td>
<td>0.47</td>
<td>HBEGF</td>
<td>0.19</td>
<td>RGS16</td>
<td>0.54</td>
<td>TRH</td>
<td>0.44</td>
</tr>
<tr>
<td>CRH</td>
<td>0.32</td>
<td>HSPA12B</td>
<td>0.35</td>
<td>RLBP1L2</td>
<td>0.20</td>
<td>UGT8</td>
<td>0.26</td>
</tr>
<tr>
<td>CRLF1</td>
<td>0.30</td>
<td>HTR1A</td>
<td>0.26</td>
<td>S100B</td>
<td>0.37</td>
<td>VIM</td>
<td>0.30</td>
</tr>
<tr>
<td>CX3CR1</td>
<td>0.30</td>
<td>ICMT</td>
<td>0.46</td>
<td>SLC6A2</td>
<td>0.28</td>
<td>VIP</td>
<td>0.35</td>
</tr>
<tr>
<td>DRD1</td>
<td>0.50</td>
<td>LCT</td>
<td>0.40</td>
<td>SLC6A3</td>
<td>0.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEV</td>
<td>0.40</td>
<td>MKI67</td>
<td>0.33</td>
<td>SLC6A4</td>
<td>0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEZF2</td>
<td>0.27</td>
<td>NR2E1</td>
<td>0.25</td>
<td>SLC6A5</td>
<td>0.27</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.4.4 **MiniPs are designed on the basis of genomics analyses**

For some selected genes, published information about their regulation influenced designs. The regulatory literature pertaining to the 57 selected genes was curated and stored in the PAZAR database (Portales-Casamar et al. 2007) within the “Pleiades Genes” project. This data collection represents a unique public dataset dedicated to transcriptional regulation in the brain, uniting information about both cis-regulatory sequences and mediating transcription factors (TFs), accompanied by details of the experimental evidence.

A main MiniP design intention is the identification of sequences with the capacity to drive expression in both human and mouse cells. Although gene selection was primarily informed by mouse expression data, MiniPs were designed with human DNA sequences and tested in mice. This dichotomy motivated the use of phylogenetic footprinting between the human and mouse genomes to delineate candidate RRs. Within these conserved regions we generated TF binding site predictions when TFs were known to be important mediators of expression in the specific cells or regions of interest. The TF REST (also known as NRSF) was included in all analyses as it has been described to bind to many neuronal genes to prevent their expression in nonneuronal tissues (Schoenherr and Anderson 1995). The TF binding models used in this study were compiled from the JASPAR database (Sandelin 2004) supplemented with brain-specific TFs annotated in the PAZAR “Pleiades Genes” project introduced above. As detailed in Materials and Methods, the MiniP design pipeline takes into account information from the literature, genome annotations, and computational analyses.

2.4.5 **Validation of the design strategy using the previously characterized promoters**

As a proof of principle validation of our bioinformatics approach, we applied our MiniP design pipeline to the four genes for which a promoter had been previously characterized (Appendix A: Dataset S1 “Refined Control,” Fig. 2.6, and Fig. 2.7).

For both the DBH and GFAP genes, we identified a minimal promoter and a well-conserved upstream sequence, both included in the previously characterized promoters, and we then fused them to generate Ple49 and Ple90 for DBH and GFAP, respectively. In the case of DCX, two additional MiniPs were designed, Ple54 and Ple55, which include additional putative RRs located within DCX introns. Finally, for the HCRT gene, the previously characterized promoter contains an upstream RR followed by a minimal promoter but our
bioinformatics analysis suggested that the minimal promoter alone could be sufficient to induce expression and was designed as Ple111. All our refined control MiniPs were expressed similarly to the original control MiniPs, demonstrating the suitability of the bioinformatics design approach (Fig. 2.4 and Fig. 2.6).

Figure 2.6 Refined designs of control genes validate the bioinformatics approach.
We analyzed adult brain sagittal sections of 4 positive strains carrying MiniPs designed using our bioinformatics pipeline on control genes. For Ple49, sections of the adrenal gland were also analyzed. EGFP is detected using anti-GFP immunochemistry (brown) and lacZ is detected using X-gal histochemistry (counter-stained with neutral red). AG, adrenal gland; Bs, brainstem; Cb, cerebellum; Ctx, cortex; Hyp, hypothalamus; LC, locus coeruleus; OB, olfactory bulb; RMS, rostral migratory stream. (A) Ple49-lacZ (DBH RRs) expression is enriched in the LC (but in cells that do not do-stain with anti-TH) and the AG. The last image shows co-staining of X-gal (blue) with tyrosine hydroxylase (brown) in the AG. (B) Ple54-EGFP (DCX RRs) expression is observed in different regions of the brain with enrichment in the OB as seen on the whole brain image. The last image shows co-staining of EGFP (green) with the endogenous Dcx protein (red) in the RMS. (C) Ple90-EGFP (GFAP RRs) is expressed in astrocytes throughout the brain. The last image shows co-staining of EGFP (green) with the endogenous GFAP protein (red). (D) Ple111-EGFP (HCRT RRs) is specifically expressed in a few cells of the lateral hypothalamus. The last image shows co-staining of EGFP (green) with the endogenous HCRT protein (red). Scale bars, 100 µm.
Figure 2.7 Comparison of the Pleiades refined MiniPromoters with the previously characterized constructs.

The diagrams represent the human genomic sequences upstream and within (A) *DBH*, ~30 kb displayed, (B) *DCX*, ~130 kb displayed, (C) *GFAP*, ~15 kb displayed and (D) *HCRT*, ~5 kb displayed. Black boxes indicate the exons, black arrows for the transcription start sites, red features for the regulatory regions contained in the control promoters, and blue features for the regulatory regions contained in the refined MiniPromoters. The yellow boxes outline the MiniPromoter designs analyzed.

2.4.6 Novel MiniPs derived from the bioinformatics analyses

Applying the validated bioinformatics approach to 57 brain region-specific genes, a panel of new MiniPs was designed. This resulted in the contribution of 27 novel-expressing MiniPs for which detailed sequence and design information can now be found at the project website (http://www.pleiades.org) and in Appendix A: Dataset S1 “novel MiniP.” MiniPs were tested with different reporter genes as the 4-year project progressed: EGFP, EGFP/cre, and LacZ. Each MiniP was characterized in adult brain of germline mouse strains,
with selected strains studied for expression in spinal cords and eyes. MiniPs driving expression of the EGFP/cre reporter fusion protein were analyzed for lacZ staining after recombination of the Gt(ROSA)26Sor<sup>tm1Sor</sup> allele (Soriano 1999). Here, lacZ is acting as a historical marker and was visualized in cells descended from progenitors expressing Cre during development. The most recent subset of Pleiades MiniPs was constructed with the lacZ reporter gene to increase detection sensitivity and enable an early-rapid–analysis method in which brains were grossly sectioned in 1-mm slices, stained, and observed using a dissecting microscope. These MiniP-driven lacZ samples were taken largely from chimeras. Where both chimeras and germline animals were studied, the chimera proved to be a good predictor of a germline expression pattern.

As summarized in Figs. 2.8 and 2.9 and Fig. 2.10 and Appendix A: Dataset S1, the novel MiniPs presented unique expression patterns, as exemplified by Ple178, specific to neurons (Fig. 2.8F), Ple151, specific to glia (Fig. 2.8C), Ple162, restricted to a discreet region of the brain (Fig. 2.8D), and Ple34, broadly distributed to cells adjacent to blood vessels (Fig. 2.9 and Fig. 2.10D). Some MiniPs drove expression similarly to the endogenous source gene (e.g., Ple24; Fig. 2.9 and Fig. 2.10I) and others gave an unrelated expression pattern (e.g., Ple140; Fig. 2.9 and Fig. 2.10L). However, in both cases, the MiniPs provided immediate value in that they directed expression to a subset of brain cells.

To demonstrate the value of the EGFP/cre strains carrying lacZ as a historical marker, we performed a developmental analysis to assess the history of the lacZ-positive neurons located in the mid brain of the Ple162-EGFP/cre (<i>PITX3</i> RR s) mice (Fig. 2.11). The expression of the lacZ reporter was observed at E11.5 in the ventricular region at the mesencephalic flexure, but not in neural tissue at E10.5, delineating the onset of Ple162 MiniP expression. In a recent study, several mouse Pitx3 promoter constructs, one overlapping in sequence with Ple162, were analyzed in E12.5 mouse transgenic embryos but the authors did not report a similar brain expression pattern (Coulon et al. 2007).
Figure 2.8 Novel MiniP expression patterns in the adult brain and retina.

EGFP is detected using anti-GFP immunochemistry (brown) or EGFP/cre, detected using X-gal histochemistry (counter-stained with neutral red). All sections are sagittal unless otherwise stated. Hip, hippocampus; MB, midbrain; Ret, retina; VTA, ventral tegmental area. (A) Ple67-EGFP (FEV RRs) expression is enriched in all raphe nuclei: from left to right, dorsal (coronal section), magnus (sagittal section), pallidus/obscurus (coronal section). (B) Ple103-EGFP/cre (HAP1 RRs) shows sporadic staining in various regions of the brain. (B) Ple151-EGFP (OLIG1 RRs) expresses throughout the brain in a puffy-like manner. The second detail image shows no co-staining of EGFP (green) with neuronal marker NeuN (red). The last image shows co-staining of EGFP (green) with the myelinating oligodendrocyte marker RIP (red). (D) Ple162-EGFP/cre (PITX3 RRs) is very specifically expressed in cells just dorsal to the VTA as well as the retina. The last image shows no co-staining of X-gal (blue) with tyrosine hydroxylase.
(brown). (E) Ple167-EGFP/cre (POGZ RRs) is expressed in patches of cells across the brain. The last image shows co-staining of X-gal (blue) with NeuN (brown). (F) Ple178-EGFP/cre (RGS16 RRs) is expressed in various regions of the brain. The last image shows a co-staining of X-gal (blue) with NeuN (brown). (G) Ple185-EGFP (S100B RRs) expresses in Bergmann glia of the cerebellum and myelinated fibers in the cortex. The last image is a co-staining of EGFP (green) with the endogenous S100B protein (red). Scale bars, 100µm, except C right-most image, 50 µm.

Figure 2.9 A view of selected novel MiniP driving lacZ assessed in 1 mm brain slices.
At least two chimeric and/or germline brains were analyzed for each strain using X-gal staining of 1 mm brain slices, and presented similar phenotypes. More detailed images can be found in Fig. 2.10 and at http://www.pleiades.org.
Figure 2.10 Selected novel MiniPromoter expression patterns assessed in 1 mm brain slices.
Figure 2.10. Selected novel MiniPromoter expression patterns assessed in 1 mm brain slices (continued).

(A) Ple3-lacZ (ADORA2A RRs) expression is observed in the dorsal lateral geniculate and brainstem region. (B) Ple17-lacZ (C8ORF46 RRs) expression is strong in the deep cortex, hippocampus, and posterior lobe of cerebellum. (C) Ple26-lacZ (CCL27 RRs) expression is strong throughout the brain except for the olfactory region and thalamus. (D) Ple34-lacZ (CLDNS RRs) expression is present in, or around, blood vessels throughout the brain. (E) Ple119-lacZ (HTR1A RRs) staining is relatively sparse but nicely localized in the ventral thalamic/posterior hypothalamic territories, cortex layer IV, hippocampus area CA1c, and retrosplenial cortex. (F) Ple131-lacZ (MKI67 RRs) expression is strong surrounding the ventricles, in the RMS, and the dentate gyrus. (G) Ple153-lacZ (OXT RRs) expression is limited to the anterior thalamic territory. (H) Ple160-lacZ (PITX3 RRs) expression is strong in the anterior thalamus, ventral-lateral hippocampus, and present in the VTA region. (I) Ple24-lacZ (CCKBR RRs) expression is enriched in the cortex, basal lateral amygdala, hippocampal pyramidal cells and in the red nucleus. (J) Ple123-lacZ (ICMT RRs) expression is strong throughout the brain. (K) Ple139-lacZ (NR2E1 RRs) presents a very regional staining, heavy up through dorsal midbrain, then virtually absent going more ventral and posterior. (L) Ple140-lacZ (NR2E1 RRs) expression is strong in the hypothalamus and present the amygdala.
Figure 2.11 MiniPromoters as tools to study developmental expression patterns

X-gal staining was performed in whole-mount (w) and sagittal sections (s) in Ple162-EGFP/cre mice (based on PITX3 regulatory regions) across development from embryonic day (E) (A) 10.5, (B) 11.5, (C) 15.5, and (D) postnatal day (P) 0.5. NE, neuro-epithelium; VTA, ventral tegmental area. Scale bars, 50 µm (Cw, As, Bs, Ds), 100 µm (Cs), 500 µm (Aw), 750 µm (Dw), 1,000 µm (Bw, Cw inset, Dw inset).
2.4.7 A unique dataset for in silico Studies

The expression profiles of active MiniPs can provide insights into the mechanisms governing transcriptional regulation in specific cells and regions of the brain. For the MiniPs associated with OLIG1, one MiniP was found to drive expression in oligodendrocytes in the adult brain (Ple151), recapitulating the endogenous OLIG1 expression pattern (Arnett et al. 2004), whereas two other MiniPs did not direct detectable expression in adult brain (Ple148 and Ple150). For oligodendrocyte-enriched TFs, we compared predicted TFBS between the positive and negative MiniP sequences (Materials and Methods and Tables 2.S1 and 2.S2). Our results highlight the potential involvement of EGR1 (KROX-24) and FOS (AP-1) in the Ple151 expression pattern (Fig. 2.6 and Table 2.2). These findings are in agreement with published data suggesting a role for OLIG1, KROX-24, and the AP-1 family of TFs in oligodendrocyte differentiation (Zhou and Anderson 2002; Sock et al. 1997; Barnett et al. 1995) and demonstrate the utility of validated MiniP sequences in predicting gene regulatory mechanisms.

2.5 Discussion

The Pleiades Promoter Project uses a high-throughput, bioinformatically-driven approach to produce brain-specific MiniPs. To date, 27 novel MiniPs have demonstrated positive brain expression (32% of constructs tested), greatly increasing the availability of brain promoters for new research initiatives. New promoters for the blood-brain barrier (e.g., Ple34), proliferating cells (e.g., Ple131), and glia (e.g., Ple185) may hold particular impact.

The bioinformatically driven design of MiniPs for adult-brain gene expression is an important development compared with traditional tedious promoter-deletion studies. It also allows the generation of small constructs by removing nonessential sequences, making these tools potentially more widely applicable. The GENSAT project has generated reporter gene expression using BACs (100–200 kb) driving EGFP regionally in the brain (Gong et al. 2003), but the results do not specify define RRs. The achieved success of the Pleiades design process was facilitated by the availability of large-scale gene expression studies (Lein et al. 2007; Magdaleno et al. 2006; Visel et al. 2004) comparative genomics tools (Siepel 2005), and bioinformatics software for regulatory sequence prediction (Wasserman and Sandelin 2004). By introducing the regulatory resolution scoring
procedure to target the design efforts on the most tractable genes, the probability of design success was increased, a necessity given the expense of transgenic studies of the developing brain. However, the bioinformatics approach does not need to be brain-specific and therefore could be applied for compact promoter design in other tissues. Moreover, we report on our website both positive and negative constructs to facilitate future designs.

The MiniPs defined in this study are anticipated to be of wide utility. They can be used for brain-, spinal cord-, and eye-directed delivery of molecules such as siRNA, cre recombinase, fluorescent reporters, and research proteins. Driving specific reporters, they can be used in flow-sorting experiment to enrich or exclude cells of specific neural types. We have already demonstrated their function in mouse ESCs and a future critical step will be assaying their function in human stem cells. This will serve as a unique cross-species test of the Pleiades human MiniPs, and, if they function similarly, will deepen our understanding of regulatory program conservation.

The Pleiades mouse strains can also be used as marker strains that can be crossed with mutant strains to reveal specific cellular effect of the mutation. Further characterization of the germline mouse strains will delineate specific cellular subtypes targeted by the MiniPs, as well as expression outside the CNS and over development. As exemplified by the analysis of the OLIG1 data, the MiniPs hold basic utility in ongoing efforts to understand the transcriptional regulatory networks governing cellular phenotypes. Ultimately, the greatest impact of the Pleiades MiniPs is anticipated to be the added specificity for therapeutic gene delivery into the human brain. Although this may be accomplished using viruses, site-specific delivery to the human genome directly or in cell therapy is an area of active research (Thyagarajan et al. 2008; Kuduvalli et al. 2005) and the suitability of the Pleiades promoters for such therapeutic delivery will require further study. The availability of a large collection of new MiniPs will play an important role in research and treatment for incurable brain diseases.

2.6 Summary

We demonstrated throughout this study the applicability of our human MiniPromoters for use in in vitro studies and in adult mouse brain. Consistent expression patterns were observed between mice. Our bioinformatics approach generated 27 novel MiniPromoters and 5 refined controls. Nearly ~75% showed related
or overlapping expression compared to the endogenous mouse gene. Many of these novel MiniPromoters express in regions or cell-types of therapeutic interest. Pleiades currently remains the only large-scale promoter development effort.

The Pleiades Promoter Project was highly successful in several aspects: (1) large-scale promoter production for a defined tissue type and the generalizable pipeline for application in other tissues, (2) the deliverables of tools for improving both basic science and clinical research in brain, and (3) improving our knowledge of gene regulation of key brain genes, including developmental genes such as DCX and NR2E1 that are further studied in the following chapters.
Chapter 3: Human MiniPromoters for the CNS and demonstration of their use in adeno-associated virus

3.1 Preamble

Although Pleiades had developed many new MiniPromoters for the adult brain and hinted at the utility of their use in development and other nervous system tissue, this aspect of MiniPromoter expression had not been fully evaluated. Since many brain genes are reused in other CNS tissues, for example the Brn3 transcription factor family in the eye, brain and dorsal root ganglion (Ninkina et al. 1993; Badea et al. 2012; Nadal-Nicolas et al. 2009), we hypothesized that many of our MiniPromoters would also express in the eye – an outgrowth of the diencephalon. Furthermore, delivery of a gene therapy via commonly used viral vectors may target non-CNS tissues also emphasizing the importance of characterizing expression in off-target tissues. Just as important is to also demonstrate that these single-copy knock-ins of MiniPromoters in mouse would recapitulate the same expression pattern when delivered via viral vectors. This validation step would provide further evidence of our design strategy presented in Chapter 2, as a viable methodology for translational studies. Thus in the current chapter; we set out to characterize additional MiniPs, expression outside the brain, and demonstrate their use in adeno-associated virus. Lastly, we wanted to further pursue the expression pattern of our MiniPromoters designed based on the brain development genes DCX and NR2E1, which would set up further gene-specific studies in Chapter 4.

3.2 Introduction

Several groups have focused on genome-wide expression analyses in mouse brain (Ng et al. 2009; Gong et al. 2003; Visel et al. 2004; Magdaleno et al. 2006). However, these projects are limited in their ability to provide information on the location and function of specific regulatory elements that drive the expression pattern. Recently the VISTA enhancer project has generated data regarding specific putative regulatory elements (Visel et al. 2006).

---

Gene therapies for severe brain and eye disorders hold great therapeutic promise. Adeno-associated virus (AAV) is likely to be a key delivery mechanism due to its non-pathogenic, non-insertional and low immunogenicity characteristics (Snyder 1999). However, because of its small size, the gene payload is severely limited. To accommodate such space restrictions, compact promoters will need to be developed. Many studies have employed ubiquitous promoters to drive expression of genes of interest; however, this strategy is similarly susceptible to side effects associated with standard pharmaceuticals. In order to limit off target effects, region or cell-type specific promoters will be crucial. In addition, physiological levels of gene expression is likely to be more appropriate for gene-based therapeutics and in fact, using functional endogenous promoters may result in higher expression levels than ubiquitous, viral-based ones (Xu et al. 2001). Viral promoters can also enhance transgene inactivation, resulting in failure of long-term therapeutic results, increased immunogenicity (Cordier et al. 2001; Sun et al. 2005), and also preventing vector readministration.

The Pleiades Promoter Project aims to overcome these challenges by generating small promoters of purely human DNA-content, which exhibit highly-specific expression patterns and that could be used in space-constrained vectors. We previously published the first effort to generate novel human-only DNA MiniPromoters, less than 4 kilobases long, to drive regional or cell-type specific expression in mouse brain (Portales-Casamar et al. 2010). With this work as a stepping-stone, we set out to continue promoter development and to characterize more deeply some of our previous designs. One such aspect was to evaluate MiniPromoter expression in the spinal cord. Furthermore, we analyzed a set of MiniPromoters in the embryo for use in developmental studies. While our primary design goal was for adult brain expression, other CNS tissues use similar transcriptional programs and are also likely to be positive for a subset of constructs. Thus, we also sought to characterize the function of MiniPromoters in the mouse retina, a key tissue of need when it comes to promoters.

Eighteen novel brain MiniPromoters are described herein, with two identified as near pan-neuronal expression in the adult mouse brain and six validated for use in developmental studies. We further demonstrate 17 with expression in the eye, including a few with expression outside the neural retina in the cornea or lens.
Importantly, we validated three of these MiniPromoters, which were positive in the ganglion cell layer, for use in the adeno-associated viral vector.

### 3.3 Materials and methods

#### 3.3.1 Generation of knock-in mice

Candidate genes for MiniPromoter design were identified in previous datasets (Siddiqui et al. 2005; D’Souza et al. 2008) as previously described (Portales-Casamar et al. 2010). MiniPromoters were designed and single-copy knock-in mice at the *Hprt*<sup>Δm3</sup> locus on the X-chromosome were generated as previously described (Yang et al. 2009; Portales-Casamar et al. 2010). Animals were housed at the Centre for Molecular Medicine and Therapeutics, in a barrier pathogen-free facility. Animals were kept on a 14-hour ON light cycle, and provided with food and water *ad libitum*. MiniPromoter strains were backcrossed to either C57BL/6J (JAX Stock#000664) or 129S1/SvImJ (JAX Stock#0022448). Animals used for analysis were at least two male germline mice unless otherwise noted with the chimera indicator (CH) - chimeras were analyzed for constructs that never entered the germline despite extensive breeding of chimeras and often multiple rounds of microinjection of different clones. Each strain or associated embryonic stem cell line in the case of chimeras, were deposited at The Jackson Laboratory as indicated in Table 3.1.

All procedures involving animals were in accordance with the Canadian Council on Animal Care (CCAC) and UBC Animal Care Committee (ACC) (Protocol #A09-0980 and A09-0981).
### Table 3.1 Detailed strain information

<table>
<thead>
<tr>
<th>Ple number</th>
<th>Gene RRs</th>
<th>Plasmid pEMS number</th>
<th>Animal category</th>
<th>Strain name or ESC mEMS number</th>
<th>MMBRC Stock</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>AVP</td>
<td>1482</td>
<td>Germline</td>
<td>B6.129P2(Cg)-Hprt&lt;sup&gt;tm12293lacZ/Ems&lt;/sup&gt;/Mmjax</td>
<td>032943-JAX</td>
</tr>
<tr>
<td>15</td>
<td>CBORF46</td>
<td>1485</td>
<td>Chimera</td>
<td>mEMS4328</td>
<td>035891-JAX</td>
</tr>
<tr>
<td>16</td>
<td>CBORF46</td>
<td>1486</td>
<td>Chimera</td>
<td>mEMS4698</td>
<td>035984-JAX</td>
</tr>
<tr>
<td>17</td>
<td>CBORF46</td>
<td>1487</td>
<td>Germline</td>
<td>B6.129P2(Cg)-Hprt&lt;sup&gt;tm12948lacZ/Ems&lt;/sup&gt;/Mmjax</td>
<td>032948-JAX</td>
</tr>
<tr>
<td>22</td>
<td>CCKBR</td>
<td>1493</td>
<td>Chimera</td>
<td>mEMS4421</td>
<td>035915-JAX</td>
</tr>
<tr>
<td>23</td>
<td>CCKBR</td>
<td>1494</td>
<td>Chimera</td>
<td>mEMS4267</td>
<td>035870-JAX</td>
</tr>
<tr>
<td>24</td>
<td>CCKBR</td>
<td>1495</td>
<td>Germline</td>
<td>B6.129P2(Cg)-Hprt&lt;sup&gt;tm12951lacZ/Ems&lt;/sup&gt;/Mmjax</td>
<td>032949-JAX</td>
</tr>
<tr>
<td>25</td>
<td>CCKBR</td>
<td>1496</td>
<td>Germline</td>
<td>B6.129P2(Cg)-Hprt&lt;sup&gt;tm12957lacZ/Ems&lt;/sup&gt;/Mmjax</td>
<td>032965-JAX</td>
</tr>
<tr>
<td>26</td>
<td>CCL27</td>
<td>1497</td>
<td>Germline</td>
<td>B6.129P2(Cg)-Hprt&lt;sup&gt;tm12962lacZ/Ems&lt;/sup&gt;/Mmjax</td>
<td>032952-JAX</td>
</tr>
<tr>
<td>27</td>
<td>CCL27</td>
<td>1498</td>
<td>Chimera</td>
<td>mEMS4354</td>
<td>035898-JAX</td>
</tr>
<tr>
<td>28</td>
<td>CCL27</td>
<td>1499</td>
<td>Chimera</td>
<td>mEMS4428</td>
<td>035916-JAX</td>
</tr>
<tr>
<td>29</td>
<td>CCL27</td>
<td>1500</td>
<td>Chimera</td>
<td>mEMS4153</td>
<td>035835-JAX</td>
</tr>
<tr>
<td>32</td>
<td>CLDN5</td>
<td>1503</td>
<td>Chimera</td>
<td>mEMS4730</td>
<td>035997-JAX</td>
</tr>
<tr>
<td>34</td>
<td>CLDN5</td>
<td>1505</td>
<td>Germline</td>
<td>B6.129P2(Cg)-Hprt&lt;sup&gt;tm12950lacZ/Ems&lt;/sup&gt;/Mmjax</td>
<td>032950-JAX</td>
</tr>
<tr>
<td>53</td>
<td>DCX</td>
<td>1524</td>
<td>STOCK</td>
<td>Hprt&lt;sup&gt;tm521lacZ/Ems&lt;/sup&gt;/Mmjax</td>
<td>032945-JAX</td>
</tr>
<tr>
<td>55</td>
<td>DCX</td>
<td>1526</td>
<td>Germline</td>
<td>Hprt&lt;sup&gt;tm525lacZ/Ems&lt;/sup&gt;/Mmjax</td>
<td>032942-JAX</td>
</tr>
<tr>
<td>67</td>
<td>FEV</td>
<td>1538</td>
<td>Germline</td>
<td>B6.129P2(Cg)-Hprt&lt;sup&gt;tm12941lacZ/Ems&lt;/sup&gt;/Mmjax</td>
<td>032941-JAX</td>
</tr>
<tr>
<td>88</td>
<td>GFAP</td>
<td>1559</td>
<td>Germline</td>
<td>B6.129P2(Cg)-Hprt&lt;sup&gt;tm12947lacZ/Ems&lt;/sup&gt;/Mmjax</td>
<td>032947-JAX</td>
</tr>
<tr>
<td>97</td>
<td>GPX3</td>
<td>1568</td>
<td>Chimera</td>
<td>mEMS4232</td>
<td>035857-JAX</td>
</tr>
<tr>
<td>112</td>
<td>HCRT</td>
<td>1583</td>
<td>Chimera</td>
<td>mEMS3867</td>
<td>035774-JAX</td>
</tr>
<tr>
<td>122</td>
<td>ICMT</td>
<td>1593</td>
<td>Chimera</td>
<td>mEMS4123</td>
<td>035825-JAX</td>
</tr>
<tr>
<td>123</td>
<td>ICMT</td>
<td>1594</td>
<td>Germline</td>
<td>B6.129P2(Cg)-Hprt&lt;sup&gt;tm12954lacZ/Ems&lt;/sup&gt;/Mmjax</td>
<td>032954-JAX</td>
</tr>
<tr>
<td>131</td>
<td>MKI67</td>
<td>1602</td>
<td>Germline</td>
<td>B6.129P2(Cg)-Hprt&lt;sup&gt;tm12940lacZ/Ems&lt;/sup&gt;/Mmjax</td>
<td>032940-JAX</td>
</tr>
<tr>
<td>140</td>
<td>NR2E1</td>
<td>1612</td>
<td>Germline</td>
<td>B6.129P2(Cg)-Hprt&lt;sup&gt;tm12951lacZ/Ems&lt;/sup&gt;/Mmjax</td>
<td>032951-JAX</td>
</tr>
<tr>
<td>146</td>
<td>NTSR1</td>
<td>1617</td>
<td>Chimera</td>
<td>mEMS3776</td>
<td>035770-JAX</td>
</tr>
<tr>
<td>155</td>
<td>PCP2</td>
<td>1626</td>
<td>Germline</td>
<td>B6.129P2(Cg)-Hprt&lt;sup&gt;tm12960lacZ/Ems&lt;/sup&gt;/Mmjax</td>
<td>032960-JAX</td>
</tr>
<tr>
<td>167</td>
<td>POGZ</td>
<td>1691&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Germline</td>
<td>B6.129P2(129S4)-Hprt&lt;sup&gt;tm16/129S4-Egfp/cre/Ems&lt;/sup&gt;/Mmjax</td>
<td>032928-JAX</td>
</tr>
<tr>
<td>170</td>
<td>POGZ</td>
<td>1641</td>
<td>Germline</td>
<td>B6.129P2(Cg)-Hprt&lt;sup&gt;tm170lacZ/Ems&lt;/sup&gt;/Mmjax</td>
<td>032944-JAX</td>
</tr>
<tr>
<td>201</td>
<td>SLC6A5</td>
<td>1673</td>
<td>Chimera</td>
<td>mEMS4184</td>
<td>035842-JAX</td>
</tr>
<tr>
<td>232</td>
<td>TNN1T</td>
<td>1704</td>
<td>Germline</td>
<td>B6.129P2(Cg)-Hprt&lt;sup&gt;tm12964lacZ/Ems&lt;/sup&gt;/Mmjax</td>
<td>032964-JAX</td>
</tr>
<tr>
<td>235</td>
<td>TRH</td>
<td>1707</td>
<td>Chimera</td>
<td>mEMS3902</td>
<td>035779-JAX</td>
</tr>
<tr>
<td>238</td>
<td>TRH</td>
<td>1710</td>
<td>Germline</td>
<td>B6.129P2(Cg)-Hprt&lt;sup&gt;tm12956lacZ/Ems&lt;/sup&gt;/Mmjax</td>
<td>032956-JAX</td>
</tr>
<tr>
<td>240</td>
<td>UGTB</td>
<td>1712</td>
<td>Germline</td>
<td>B6.129P2(Cg)-Hprt&lt;sup&gt;tm12959lacZ/Ems&lt;/sup&gt;/Mmjax</td>
<td>032959-JAX</td>
</tr>
</tbody>
</table>

<sup>a</sup>Plasmid constructs available from AddGene (http://www.addgene.org/).

<sup>b</sup>All constructs drive a lacZ reporter, except Ple167 which drives EGFP/cre.

RRs, regulatory regions.

### 3.3.2 PCR analysis of genomic DNA

Vector NTI software (Invitrogen) was used to design PCR assays for the different constructs. MiniP-specific PCR genotyping assays are available on the http://www.pleiades.org website or by contacting the corresponding author.
3.3.3 Immunohistochemistry of adult brains and other tissues

1 mm sectioned tissues. Mice were perfused with 4% PFA, and post-fixed 2–4 hours as described above. The brains were then removed from 4% PFA and immediately sectioned. The dissected brains were placed ventral side up into the adult mouse coronal or sagittal, Rodent Brain Matrix (ASI Instruments). Slices were sectioned lateral through medial to lateral (for sagittal) or rostral to caudal (for coronal) using single-edge razor blades (Electron Microscopy Sciences). All slices from one brain were placed into one well of a 12-well plate containing 1× PBS (Invitrogen), or 0.1% sodium azide (in 1× PBS) until staining. Expression of β-galactosidase (lacZ) was detected with X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) staining as previously described (Portales-Casamar et al. 2010). Briefly, the staining was performed with 3–5 mL of X-gal staining solution (25 mg/mL X-gal, 1 M MgCl₂, 50 mM potassium ferri-cyanide, 50 mM potassium ferro-cyanide, and 1× PBS to volume) per well in a 24-well plate. The plate was wrapped in aluminum foil and incubated at 37 °C for 10–16 hours. Subsequently, the sections were transferred into PBS, examined under a dissecting microscope, and photographed using a CoolSNAP-Procf color camera (Media Cybernetics) mounted on a Leica MZ12.5 stereomicroscope (Leica Microsystems) and Image-Pro Express v.4.5.1.3 software (Media Cybernetics).

Other tissues. Spinal cord, heart, liver, lung and earnotch were collected along with perfused brains and stained as described above.

Cryosectioned tissues. The mice were anesthetized with avertin and perfused transcardially with 4% PFA (paraformaldehyde) in 0.1 M phosphate buffer (pH 7.4) for 15 minutes. Brains were dissected and post-fixed in the same solution for 2 hours and transferred into 25% sucrose-PBS overnight. Each brain was sectioned using a cryostat and 20-μm sagittal sections were collected. X-gal staining and imaging was performed as described for 1-mm sections. Co-staining of X-gal and NeuN or Gfap was performed sequentially. Primary antibodies used were mouse anti-GFAP (1:1000, Millipore) and mouse anti-NeuN (1:500, Chemicon). Bound primary antibodies were detected using the VectaStain Elite ABC Kit (Vector Labs, PK-6100), as per manufacturer’s instructions. Slides were gradually dehydrated in ethanol and xylene series and mounted with coverslips. Brightfield images were taken using Olympus BX61 motorized microscope using the DP Controller software.
3.3.4 Embryo harvesting and staining

Female heterozygous MiniPromoter mice were crowded for 7-10 days and set-up in the afternoon for timed-pregnancies using C57BL/6J or 129S1/SvImJ studs. Plugs were checked for 5 days in the morning. The day of plug was considered embryonic day 0.5 (E0.5). At E12.5, embryos were harvested, yolk sac tissue taken for PCR analysis, and embryos fixed in 4% PFA/PBS for 2-6 hours. Embryos were stained in X-gal solution for 18 hours, post-fixed overnight at 4°C and stored in 70% ethanol. Pictures were taken of uncleared embryos. The embryos were then partially cleared according to (Schatz et al. 2005) by incubating in Step 1 solution (20% glycerol with 0.8% KOH) for three days at 37°C with swirling once daily. Further clearing was done in Step 2 solution (50% glycerol with 0.5% KOH) for 3 days at 37°C. Post-fixing of fragile cleared embryos was done in 4% PFA for 2 hours at room temperature. Embryos were stored until imaging in 70% ethanol, after which they were equilibrated to 100% glycerol for imaging. Imaging was performed as per 1-mm brain sections. All embryo expression patterns were seen in at least two animals.

3.3.5 Immunohistochemistry on eyes

LacZ staining was detected either directly from MiniPromoters driving β-galactosidase, or by driving expression of EGFP/cre resulting in recombination of the Gt(ROSA)26Sor^ms1Sor locus (Soriano 1999). After whole body perfusion of mice, eyes were post-fixed in 4% PFA along with brains. LacZ staining was performed using the X-gal substrate for 18 hours at 37°C. Eyes were then post-fixed in 4% PFA for 2 hours prior to cryoprotection in 25% sucrose-PBS overnight. Eyes were embedded in OCT, along with positive and negative controls, and cryosectioned at 12 μm and mounted on SuperFrost Plus slides. Eye sections were counterstained with neutral red for 45 seconds to 1 minute. Eye expression was confirmed in at least two animals unless otherwise indicated. Brightfield images were taken using Olympus BX61 motorized microscope using the DP Controller software.

3.3.6 Immunofluorescence on eyes

Eyes were harvested and fixed in 4% PFA for 2 hours prior to cryoprotection in 25% sucrose-PBS overnight. Eyes were embedded in OCT, along with positive and negative controls, and cryosectioned at 12 μm and mounted on SuperFrost Plus slides prior to immunofluorescence. Standard immunofluorescence procedures were followed. Primary antibodies include: chicken anti-β-gal (Abcam, ab9361), rabbit anti-β-tubulin.
III (Covance, PRB-435P), rabbit anti-Brn3 (Santa Cruz, sc28595), and rabbit anti-Sox2 (Abcam, ab97959).

Secondary antibodies used were (Invitrogen): chicken anti-rabbit Alexa-594, goat anti-chicken Alexa-488, goat anti-chicken Alexa-594, goat anti-mouse Alexa-488, goat anti-mouse Alexa-594, goat anti-rabbit Alexa-488, goat anti-rabbit Alexa-594. Nuclei were stained using 1 μg/mL of Hoechst 33342 (Sigma, 14533). Slides were mounted with Prolong Gold anti-fade reagent and imaged using an Olympus BX61 microscope and the InVivo software, or alternatively using confocal laser scanning microscopy with the Leica SPS II microscope and LAF software.

### 3.3.7 Adeno-associated virus

Recombinant AAV (rAAV) vector plasmids containing either a 2202 bp Ple67 (FEV RRs), a 3310 bp Ple53 (DCX RRs) or 3312 bp Ple25 (CCKBR RRs) promoter construct, an SV40 splice donor acceptor site followed by the cDNA for green fluorescent protein (GFP) were constructed. To accommodate the packaging limit of rAAV, Ple53 and Ple25 were shortened relative to the knock-in promoter allele as follows: promoters were amplified by polymerase chain reaction (PCR) and the respective restriction sites were added via the oligonucleotides. The following oligonucleotides were used: Ple25 FWD KpnI : GTCGGGTACCTTATCCCATCCATTGTTGTATC, Ple25 REV XhoI : GTCGCTCGAGTGCCTTACTCAGCTCGACC, Ple53 FWD KpnI : GTCGGGTACCTACCCATCCATTGTGTATC, Ple53 REV XhoI : GTCGCTCGAGGTTGAACCTCAGAGACCTG, Ple67 FWD EcoRI : GCTGGAATTCGGAGAAGATCAAGACTCAG, and Ple67-REV-XhoI : GCTGCTCGAGGGAGGCTTTTCGCTC, (restriction sites underlined).

Identity for all promoters was confirmed by sequencing. rAAV vector plasmids were packaged into AAV2-(quad Y272, 444, 500, 730 F). AAV2-(quad Y-F) capsid contains four surface exposed tyrosine to phenylalanine mutations (Petr-Silva et al. 2011). Vectors were purified and titrated according to previously published methods (Zolotukhin 2005; Jacobson et al. 2006). At least 5 x 10⁹ vector genomes (vg) per eye were injected intravitreally in four-week old C57BL/6 mice.

GFP expression was analyzed four weeks post injection. Mice were euthanized and the eyes were fixed in 4% PFA for 2 hours at RT. Eyes were then treated with a stepwise sucrose gradient (from 4 to 20% sucrose in PBS) and embedded and frozen in Tissue-Tek OCT compound (Sakura Finetek, Torrance, CA). Ten micron thick sections of each eye were then washed in PBS for 15 minutes, mounted in VECTASHIELD® mounting medium.
(Vector Laboratories, Burlingame, CA) with 4',6-diamidino-2-phenylindole (DAPI) and analyzed for GFP expression by confocal microscopy.

All experiments were approved by the University of Florida’s Institutional Animal Care and Use Committee and were conducted in accordance with the ARVO Statement for the Use of Animals in Ophthalmic and Vision Research and with National Institutes of Health regulations.

3.3.8  Image processing

Images were processed using ImageJ, Adobe Photoshop and Adobe Illustrator. Brightness, contrast and scaling adjustments were performed as necessary. For pictures where brains were not properly aligned, they were traced with magnetic lasso and repositioned on a black background.

3.3.9  Data and resource distribution


3.4  Results

As part of ongoing efforts in promoter development we have continued to characterize previous published designs as well as developing novel MiniPromoters (MiniPs; Table 3.2). We present additional expression data for 12 previously published MiniP designs from our group (Portales-Casamar et al. 2010). We also tested three previously positive MiniPs with the more sensitive lacZ reporter construct. More importantly, we have developed 18 novel MiniPromoter constructs. The 33 construct designs herein cover 75,777 bp of the human genome comprising of sequence from genes located on Chromosomes 1-6, 8-11, 17, 19-20, 22, and X.

While the primary aim of our promoter design had been to develop tools for adult brain gene expression, we also recognize the need for promoter constructs to study developmental processes as well. For this purpose we analyzed the expression of six developmentally-relevant MiniPs at E12.5 (embryonic day 12.5). Lastly, due to the relatively non-invasive therapeutic model and immunoprivileged status (Streilein 1995) of the
eye imparted by the blood-retinal barrier (Runkle and Antonetti 2010), we pursued the development of small promoter constructs for future eye gene therapy applications. As such, we generated a unique resource of 17 MiniPromoters for expression in the mouse eye, including the neural retina, lens, cornea and optic nerve (Table 3.3).

All of our MiniPromoters are based on constructs which contain either a Prom or a longer Prom (LongProm) segment with both usually spanning the TSS of the parental gene and includes any contiguous upstream region that is conserved and likely functions as an enhancer. Therefore, our basal promoter constructs are larger than the core promoter element. In general, we designed and constructed four MiniPs for each source gene. While some constructs only contained the minimal promoter element, most contained additional conserved putative regulatory elements in a configuration 5' of the minimal promoter. All of the sequence used in generating the constructs is entirely the human sequence generated by PCR from the RP11 BAC library or synthesized using the human genome reference sequence (hg18, Mar. 2006).
<table>
<thead>
<tr>
<th>Ple number</th>
<th>Gene RRs</th>
<th>Plasmid pEMS number</th>
<th>Strain category</th>
<th>Brain expression</th>
<th>Eye expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>AVP</td>
<td>1482</td>
<td>Novel MiniP</td>
<td>This study</td>
<td>Negative</td>
</tr>
</tbody>
</table>
| 15         | C8orf46  | 1485                | Novel MiniP     | This study      | Negative⁺ Capt 
| 16         | C8orf46  | 1486                | Novel MiniP     | This study      | Negative⁺ Capt 
| 17         | C8orf46  | 1487                | Expanded characterization | Portales-Casamar et al. 2010 | Negative |
| 22         | Cckbr    | 1493                | Novel MiniP     | This study      | Negative⁺ Capt 
| 23         | Cckbr    | 1494                | Novel MiniP     | This study      | Negative⁺ Capt 
| 24         | Cckbr    | 1495                | Expanded characterization | Portales-Casamar et al. 2010 | Positive |
| 25         | Cckbr    | 1496                | Expanded characterization | Portales-Casamar et al. 2010 | Positive |
| 26         | Ccl27    | 1497                | Expanded characterization | Portales-Casamar et al. 2010 | Positive |
| 27         | Ccl27    | 1498                | Novel MiniP     | This study      | Negative⁺ Capt 
| 28         | Ccl27    | 1499                | Novel MiniP     | This study      | Positive 
| 29         | Ccl27    | 1500                | Novel MiniP     | This study      | Negative⁺ Capt 
| 32         | Cldns    | 1503                | Expanded characterization | Portales-Casamar et al. 2010 | Positive |
| 34         | Cldns    | 1505                | Expanded characterization | Portales-Casamar et al. 2010 | Positive |
| 53         | Dcx      | 1524                | Different reporter | This study | Positive |
| 55         | Dcx      | 1526                | Expanded characterization | Portales-Casamar et al. 2010 | Positive |
| 67         | Fev      | 1538                | Different reporter | This study | Positive |
| 88         | Gfap     | 1559                | Expanded characterization | Portales-Casamar et al. 2010 | Positive |
| 97         | Gpx3     | 1568                | Novel MiniP     | This study      | Negative⁺ Capt 
| 112        | Hcrt     | 1583                | Different reporter | This study | Negative⁺ Capt 
| 122        | Icmt     | 1593                | Novel MiniP     | This study      | Negative⁺ Capt 
| 123        | Icmt     | 1594                | Expanded characterization | Portales-Casamar et al. 2010 | Positive |
| 131        | Mk67     | 1602                | Expanded characterization | Portales-Casamar et al. 2010 | Positive |
| 140        | Nrze1    | 1612                | Expanded characterization | Portales-Casamar et al. 2010 | Negative |
| 146        | Ntsr1    | 1617                | Novel MiniP     | This study      | Positive |
| 155        | Pcp2     | 1626                | Novel MiniP     | This study      | Positive |
| 167        | Pog2     | 1091⁺ Capt | Expanded characterization | Portales-Casamar et al. 2010 | Positive |
| 170        | Pog2     | 1641                | Novel MiniP     | This study      | Positive |
| 201        | Slcd6as  | 1673                | Novel MiniP     | This study      | Negative⁺ Capt 
| 232        | Tnnt1    | 1704                | Novel MiniP     | This study      | Negative⁺ Capt 
| 235        | Trh      | 1707                | Novel MiniP     | This study      | Negative⁺ Capt 
| 238        | Trh      | 1710                | Novel MiniP     | This study      | Negative⁺ Capt 
| 240        | Ugt8     | 1712                | Novel MiniP     | This study      | Positive |

**MiniPromoter constructs presented in this article are shown. Data for 18 new MiniPromoters are presented in addition to further characterization of 12 constructs, and in the case of Ple53, Ple67 and Ple112 as a novel strain using the lacZ reporter.**

⁺Plasmid constructs available from AddGene (http://www.addgene.org/).

⁺⁺All constructs were positive in the brain.

⁺⁺⁺All constructs drive a lacZ reporter, except Ple167 which drives EGFP/cre.

⁺⁺⁺⁺These constructs were analyzed in chimeras and might be false negatives in the eye.

RRs, regulatory regions.
Table 3.3 MiniPromoters with expression in the mouse eye

<table>
<thead>
<tr>
<th>Ple number</th>
<th>Gene RRs</th>
<th>Plasmid number</th>
<th>Retinal expression regions</th>
<th>Extra-retinal expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>CCKBR</td>
<td>1495</td>
<td>GCL, rare INL</td>
<td>-</td>
</tr>
<tr>
<td>25</td>
<td>CCKBR</td>
<td>1496</td>
<td>GCL, rare INL</td>
<td>-</td>
</tr>
<tr>
<td>26</td>
<td>CCL27</td>
<td>1497</td>
<td>All layers</td>
<td>cornea, nerve</td>
</tr>
<tr>
<td>28</td>
<td>CCL27</td>
<td>1499</td>
<td>All layers</td>
<td>lens</td>
</tr>
<tr>
<td>32</td>
<td>CLDN5</td>
<td>1503</td>
<td>GCL, INL</td>
<td>-</td>
</tr>
<tr>
<td>34</td>
<td>CLDN5</td>
<td>1505</td>
<td>GCL, INL, IPL, PR</td>
<td>-</td>
</tr>
<tr>
<td>53</td>
<td>DCX</td>
<td>1524</td>
<td>GCL, rare INL</td>
<td>nerve</td>
</tr>
<tr>
<td>55</td>
<td>DCX</td>
<td>1526</td>
<td>GCL, rare INL</td>
<td>nerve</td>
</tr>
<tr>
<td>67</td>
<td>FEV</td>
<td>1538</td>
<td>GCL, rare INL</td>
<td>-</td>
</tr>
<tr>
<td>88</td>
<td>GFAP</td>
<td>1559</td>
<td>Astro</td>
<td>nerve</td>
</tr>
<tr>
<td>123</td>
<td>MKI67</td>
<td>1594</td>
<td>All layers</td>
<td>cornea, nerve</td>
</tr>
<tr>
<td>131</td>
<td>MKI67</td>
<td>1602</td>
<td>-</td>
<td>cornea</td>
</tr>
<tr>
<td>146</td>
<td>NTSR1</td>
<td>1617</td>
<td>GCL, rare INL</td>
<td>-</td>
</tr>
<tr>
<td>155</td>
<td>PCP2</td>
<td>1626</td>
<td>INL, IPL</td>
<td>-</td>
</tr>
<tr>
<td>167</td>
<td>POGZ</td>
<td>1091b</td>
<td>All layers</td>
<td>cornea</td>
</tr>
<tr>
<td>170</td>
<td>POGZ</td>
<td>1641</td>
<td>All layers</td>
<td>cornea</td>
</tr>
<tr>
<td>240</td>
<td>UGT8</td>
<td>1712</td>
<td>GCL, INL, PR</td>
<td>cornea, nerve</td>
</tr>
</tbody>
</table>

Astro, astrocyte layer; GCL, ganglion cell layer; INL, inner nuclear layer; IPL, inner plexiform layer; PR, photoreceptor; RRs, regulatory regions.

aPlasmid constructs available from AddGene (http://www.addgene.org/).
bAll constructs drive a lacZ reporter, except Ple167 which drives EGFP/cre.

3.4.1 Novel brain MiniPromoter expression ranged from ubiquitous to highly regionalized.

We first pursued analysis of constructs in the context of the adult mouse brain. Previously, we encountered examples of constructs initially negative with EGFP, subsequently tested with lacZ and found to be positive (unpublished observation and (Portales-Casamar et al. 2010)). This data implied that the lacZ marker is more robust than EGFP for evaluating MiniPromoter expression. Thus, we set out to determine whether the use of lacZ on previous designs would result in a change in expression pattern due to the increased sensitivity. We evaluated the expression of three of our previously published promoters driving the EGFP reporter (Portales-Casamar et al. 2010): Ple53 (expression in neurons, also see Fig. 3.5), Ple67 (expression in serotonergic brain regions) and Ple112 (expression in hypothalamus), but instead driving the more sensitive lacZ reporter (Fig. 3.1). All three demonstrated an overlapping but expanded expression pattern when compared to their published
EGFP counterparts, suggesting that our previous EGFP reporter was expressed below antibody-detection limits in some cells.

Novel MiniPromoters were analyzed for brain expression (Fig. 3.1) and detailed expression analysis can be found in Appendix B: Dataset S2. Positive MiniPromoters can be grouped by their primary site of expression in the mouse brain. Four constructs (Ple27, Ple28, Ple122 and Ple170) expressed in nearly all brain regions. Three of our constructs, Ple15, Ple16 and Ple23, demonstrated expression primarily limited to the cortex. Ple240 expression was restricted to the white matter. Several MiniPs expressed strongest in the diencephalon (Ple12, Ple29, Ple97, Ple112, Ple146, Ple155 and Ple232). Ple67 was expressed in regions outside of the telencephalon. Four MiniPromoters expressed primarily in the hindbrain – Ple22, Ple201, Ple235 and Ple238. These MiniPs represent a novel diverse collection of brain promoters.

The majority of our constructs showed unique expression patterns. However, we found a large number (33.3%; Ple12, Ple22, Ple23, Ple29, Ple155 and Ple232) of the 18 novel MiniPromoters with expression in the zona incerta of the brain, indicative of a common function in this brain region despite otherwise diverse expression. Variability in expression was observed for 3 of the 9 strains (Ple12, Ple232 and Ple240). However, only one strain (Ple232) demonstrated extreme variability from nearly negative in the brain to strongly positive. Thus we conclude that variability of MiniP constructs docked at Hprt is generally low.
Figure 3.1 Brain expression patterns of 18 novel MiniPromoters and two additional MiniPromoter strains that were remade with the lacZ reporter.
Figure 3.1 Brain expression patterns of 18 novel MiniPromoter and two additional MiniPromoter strains that were remade with the lacZ reporter. (continued from previous page)

Adult mice were perfused, brains post-fixed, sectioned sagittally or coronally at 1 mm, and stained overnight with X-gal. At least two germline animals were analyzed for each established strain. Ple12 (AVP RRs) expressed in amygdala and thalamus (first panel), and in the ventral hippocampus and cortex layer IV (second panel). Ple15 (C8ORF46 RRs) expressed throughout the cortex and throughout the hippocampal granule layers with very strong signal in the dentate gyrus. Ple16 (C8ORF46 RRs) showed a similar but weaker expression pattern to Ple15. Ple22 (CCKBR RRs) had lacZ-positive cells located in the periaqueductal grey (first panel, white arrow) and in the hypothalamus (second panel, white arrow). Ple23 displayed cortical staining throughout the brain, colliculi staining, a strong localized area at the anterior thalamus (first panel), and in the midbrain colliculi, cerebellar granular and Purkinje cells (second panel, white arrows). Ple27 and 28 (CCL27 RRs) showed expression throughout most brain regions, with Ple28 being much stronger. Ple29 (CCL27 RRs) expressed weakly in small localized areas of the thalamus and hypothalamus (second panel, white arrows). Ple67 (FEV RRs) expressed in cortical layer II/III, cerebellar Purkinje cells, and strongly throughout most mid- and hindbrain areas, with a definitive exclusion of striatal tissue (first panel). Raphe nuclei are clearly positive (second panel). Ple97 (GPX3 RRs) stained cells localized at the anterior thalamus (first panel) and ventrolateral hippocampus (second panel), and also in large blood vessels (first panel). Ple112 (HCRT RRs) expressed in the anteroventral thalamus and lateral hypothalamic area (first panel). The lateral geniculate complex was positive (second panel). Ple122 (ICMT RRs) expressed strongly throughout all major brain regions, with comparatively weaker expression found in the thalamus, and hippocampal CA2, 3, 4 and dentate gyrus (first panel). Ple146 (NTSR1 RRs) had scattered staining in the anterior thalamus (first panel) and in the periaqueductal grey and Purkinje cells (second panel). Ple155 (PCP2 RRs) expressed strongly in the anterodorsal, anteroventral and anteromedial nuclei (first panel). Additional staining was seen in the retrosplenial area and in the periaqueductal grey (second panel, white arrows). Ple170 (POGZ RRs) expressed strongly throughout all brain regions, with weaker staining in the striatal nuclei. For example, the olfactory bulb mitral, glomerular, and granular layers are all stained (second panel). Ple201 (SLC6A5 RRs) expressed weakly in the cortex and brain stem. Particularly, there was staining in the anterior and retrosplenial areas of the cortex and in the medulla. Ple232 (TNNT1 RRs) expressed in the anterior cingulate and zona incerta (first panel, white arrows) and in the lateral hippocampus (second panel, white arrow). Ple235 (TRH RRs) expressed in the periaqueductal grey. Ple238 (TRH RRs) had scattered expression in many brain regions, including the cerebellar interposed and fastigial nuclei and in parts of the medulla (second panel). Ple240 (UGT8 RRs) showed strong fiber tract associated staining in all parts of the brain, with an emphasis at the boundary between layers II/II and V in the cortex and in thalamic intralaminar and anterior nuclei groups. In addition, the olfactory bulb glomeruli and corpus callosum were clearly labeled. Sagittal and coronal negative control annotated for reference. 3V, third ventricle; Bs, brainstem; Cb, cerebellum; CH, chimera; Ctx, cortex; Hipp, hippocampus; Hth, hypothalamus; Mb, midbrain; Ob, olfactory bulb; RRs, regulatory regions; Str, striatum; Th, thalamus.

3.4.2 Two MiniPs identified as candidates for pan-neuronal expression.

Establishing a pan-neuronal promoter for the adult brain has proven to be challenging due to the highly diverse collection of neuronal subtypes represented with numerous non-overlapping transcriptional programs (Molyneaux et al. 2007; DeFelipe 1993; Nelson et al. 2006). Recently Heimer-McGinn and Young (2011) published a random-insertion transgenic mouse driving Cre recombinase in an apparently pan-neuronal fashion. However, although a great resource for lineage tracing studies, it is not amenable as a general tool for driving pan-neuronal expression (i.e., there is no promoter fragment that can be used to create other mouse strains, or can be used in viral or plasmid delivery systems). In our set of MiniPromoters, we identified three constructs,
Ple26 (CCL27 RRs) and Ple123 (ICMT RRs), originally described in (Portales-Casamar et al. 2010), and Ple170 (POGZ RRs; a novel MiniPromoter), as pan-neuronal candidates. In this study, we expanded our characterization and assessed the extent to which these constructs label neurons. We observed highly variable staining for Ple123 in adult brain, similar to the extent seen in embryos (described later), and excluded it from this detailed analysis. These remaining two constructs, Ple26 and Ple170, were assessed for neuronal co-staining using NeuN, and off-target glial expression using Gfap (Fig. 3.2).

We then looked at three regions of the brain: the cortex, the hippocampus and the cerebellum (Fig. 3.2A, B). Both constructs were positive throughout these regions. The olfactory bulb was weaker stained in Ple26 as compared to the rest of the brain and to Ple170. The general overlap with NeuN is more for Ple26 than Ple170. This is particularly true for the cerebellum where comparatively weak label is observed for Ple170. Using hippocampal neurons and estimating the average coverage based on co-labeling in CA2-3 regions, we found a significant difference of 91.34% co-labeling with NeuN for Ple26 and 82.89% for Ple170 (n=3 animals for each, and 3 sections averaged for each animal; *t*-test, *P* < 0.05). Both Ple26 and Ple170 showed numerous small X-gal puncta in the cerebellar molecular layer. This data suggests that Ple26 would be superior to Ple170 for use in driving near pan-neuronal expression. Although the Ple26 construct stained greater proportions of NeuN-positive cells, it also stained extensively in fiber tracts and white matter regions. In contrast, Ple170 had less NeuN-overlap, but the non-neuronal staining was weaker.
Figure 3.2 Two MiniPromoters drive near pan-neuronal expression throughout the mouse brain.
Ple26 (CCL27 RRs) and Ple170 (POGZ RRs). For each, a representative 1-mm sagittal and coronal whole brain section is shown, as well as cryosections of NeuN (neuronal marker) and Gfap (glial marker) antibody co-labeling with X-gal in the cortex, hippocampus and cerebellum. (A) Ple26 (CCL27 RRs) showed marked expression throughout the adult brain with slightly less labeling in the olfactory bulb. The whole brain images show that there was significant X-gal staining in white-matter tracts. In the cortex, NeuN closely co-labeled many of the X-gal positive cells (black arrows). However, there were some NeuN cells that do not co-label with X-gal (red arrows), as well as many smaller punctate X-gal nuclei that do not co-label with NeuN (white arrows). Nearly all hippocampal neurons co-label with X-gal, and extensive co-labeling was also observed in the cerebellum, however X-gal only staining was present in the molecular layer. Also, rare X-gal/Gfap-positive cells were observed (black arrow). (B) Ple170 (POGZ RRs) demonstrated a similar expression pattern to that of Ple26, but with increased olfactory bulb staining and reduced white-matter track labeling. The CA2 and CA3 regions of the hippocampus showed less staining than the dentate gyrus and CA1 regions. Many co-labeled cells were visible (black arrows). However, some NeuN-positive cells in the cortex were not labelled with X-gal (red arrows), as well as some punctate X-gal nuclei were NeuN-negative (white arrows). Reduced co-labeling was seen in the hippocampus and cerebellum with NeuN. Also, occasionally a cell was found to be double-labeled with Gfap and X-gal (black arrow). Cb, cerebellum; Ctx, cortex; Hipp, hippocampus; RRs, regulatory regions. (Scale bars = 100 μm)
3.4.3 MiniPromoter expression was observed in spinal cord and non-CNS tissue.

For a subset of MiniPromoters (21 of the 33 presented) we also collected adult spinal cord, heart, liver, lung and ear-notch tissue samples to be stained for β-gal (Table 3.4). Within this analysis, we looked at all germline strains, except for Ple25 which was analyzed prior to this protocol. Of the 21 analyzed, 13 MiniPs stained positive in spinal cord (Fig. 3.3), 12 in heart, 6 in liver, 8 in lung and 14 in ear-notch samples. Nearly half (46%) of the 13 CNS-positive constructs were also expressed in all four other tissues. Interestingly, only four (31%) were only positive in spinal cord, and hence CNS-specific.

Constructs that were positive in the spinal cord were often expressed in the nerve fibers exiting the spinal column and/or dorsal root ganglia, with cell bodies located in the spinal cord. Overview images for spinal cord expression are shown in Fig. 3.3A and summarized in Fig. 3.3B. We observed expression patterns that consisted of only the central spinal cord (n=1, Ple88), spinal cord with nerve fibers (n=4; Ple17, Ple28, Ple32 and Ple170), spinal cord with dorsal root ganglia (n=4; Ple24, Ple26, Ple53 and Ple55), and spinal cord with nerve fibers and dorsal root ganglia positive (n=4; Ple34, Ple123, Ple167 and Ple240).
<table>
<thead>
<tr>
<th>Ple number</th>
<th>Gene</th>
<th>RRs</th>
<th>Plasmid pEMS number</th>
<th>Spinal cord</th>
<th>Heart</th>
<th>Liver</th>
<th>Lung</th>
<th>Ear notch</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>AVP</td>
<td>1482 Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>C8orf46</td>
<td>1485 ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>C8orf46</td>
<td>1486 ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>C8orf46</td>
<td>1487 Positive</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>CCKBR</td>
<td>1493 ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>CCKBR</td>
<td>1494 ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>CCKBR</td>
<td>1495 Positive</td>
<td>Positive</td>
<td>Negative</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>CCKBR</td>
<td>1496 ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>CCL27</td>
<td>1497 Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>CCL27</td>
<td>1498 ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>CCL27</td>
<td>1499 Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>CCL27</td>
<td>1500 ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>CLDN5</td>
<td>1503 Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>CLDN5</td>
<td>1505 Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>DCX</td>
<td>1524 Positive</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>DCX</td>
<td>1526 Positive</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>FEV</td>
<td>1538 Negative</td>
<td>Positive</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>GFAP</td>
<td>1559 Positive</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>GPIX3</td>
<td>1568 ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>HCRT</td>
<td>1583 ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>122</td>
<td>ICMT</td>
<td>1593 ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>123</td>
<td>ICMT</td>
<td>1594 Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>131</td>
<td>MKI67</td>
<td>1602 Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>140</td>
<td>NR2E1</td>
<td>1612 Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>146</td>
<td>NTSR1</td>
<td>1617 Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>155</td>
<td>PCP2</td>
<td>1626 Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>POGZ</td>
<td>1641 Positive</td>
<td>Positive</td>
<td>Negative</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>201</td>
<td>SLC6A5</td>
<td>1673 ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>232</td>
<td>TNNT1</td>
<td>1704 Negative</td>
<td>Positive</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>235</td>
<td>TRH</td>
<td>1707 ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>238</td>
<td>TRH</td>
<td>1710 Negative</td>
<td>Positive</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>240</td>
<td>LITB</td>
<td>1712 Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td></td>
</tr>
</tbody>
</table>

ND not done; RRs, regulatory regions; number of positives for each tissues indicated.

*aPlasmid constructs available from AddGene (http://www.addgene.org/).
*bPle146 is the only case of chimeras assessed in other tissues and found to be negative.
*cAll constructs drive a lacZ reporter, except Ple167 which drives EGFP/cre.
Figure 3.3 Thirteen MiniPromoters expressed in the mouse spinal cord. LacZ expression was detected by X-gal immunohistochemistry in whole mount thoracic spinal cord. (A) Ple17 (C8orf46 RRs), Ple28 (CCL27 RRs), Ple32 (CLDN5 RRs), and Ple170 (POGZ RRs) showed scattered expression in the central spinal cord and in nerve bundles. Ple24 (CCKBR RRs) expression was seen primarily in the spinal cord with rare puncta in the dorsal root ganglia (data not shown). Ple26 (CCL27 RRs), Ple53 (DCX RRs), and Ple55 (DCX RRs) expressed throughout the gray matter regions and also in dorsal root ganglia, the latter particularly strong for Ple53 and Ple55. Ple34 (CLDN5 RRs) and Ple123 (ICMT RRs) expressed throughout all of the spinal cord, while Ple123 demonstrated weaker expression in nerve bundles but strong expression in dorsal root ganglia. Ple88 (GFAP RRs) was expressed only in patches along white matter regions of the central spinal cord. Ple167 (POGZ RRs) expressed in patches in the spinal cord with some nerve fibers and occasional dorsal root ganglia puncta (data not shown) also positive. Ple240 (UGT8 RRs) is expressed strongly and predominantly along spinal nerve cord bundles with weaker expression in the central spinal cord tissue and in dorsal root ganglia. (B) Summary of expression pattern for each construct in the three key regions of the spinal cord. CH, chimera; DRG, dorsal root ganglion; RRs, regulatory regions; ✓, expression present; X, expression absent.

3.4.4 Expression of six individual MiniPromoters in embryos is confined to the heart, hypothalamus, skeletal muscle or is ubiquitous.

Six MiniPromoters were analyzed at E12.5 and assessed by X-gal staining to gain insight into their developmental expression (Fig. 3.4). For each construct, at least two hemizygous males and one litter-matched wild-type were stained for X-gal. Three adult brain pan-neuronal candidates, previously discussed, were selected.
based on preliminary brain expression data: Ple26 (CCL27 RRs), Ple123 (ICMT RRs) and Ple170 (POGZ RRs). We also selected three constructs of specific interest: Ple131 (MKI67 RRs) for the role of MKI-67 in proliferation (Gerdes et al. 1983), Ple140 (NR2E1 RRs) because of its highly regionalized hypothalamic expression in adults (Portales-Casamar et al. 2010), and Ple232 (TNNT1 RRs) for the role of TNNT1 in skeletal muscle (Novelli et al. 1992; Johnston et al. 2000).

Ple26 (CCL27 RRs) (Fig. 3.4A), a pan-neuronal candidate, had very strong expression throughout the embryo, indicative of ubiquitous expression at this developmental stage. Ple123 (ICMT RRs) and Ple131 (MKI67 RRs) were both predominantly expressed in the heart (Fig. 3.4B,C), with Ple131 also marking scattered cells throughout all other regions of the embryo. Our previously published highly-specific adult hypothalamus promoter, Ple140 (NR2E1 RRs) (Fig. 3.4D), demonstrated exclusive expression to the developing hypothalamus, consistent with our observation in the adult (Portales-Casamar et al. 2010). This early developmental specificity suggests expression in a cell type that maintains its identity throughout development. Interestingly, Ple170 (POGZ RRs) (Fig. 3.4E), the second pan-neuronal candidate, expressed more strongly throughout the limbs and tail, with a blue-green haze of expression throughout the rest of the embryo, including the nervous system. The Ple232 (TNNT1 RRs) (Fig. 3.4F) construct showed strong expression in the developing skeletal muscles of the embryo, consistent with the known function of this gene in slow skeletal muscle (Kee and Hardeman 2008). Mutations in TNNT1 result in nemaline myopathy, emphasizing its role in skeletal muscle (Johnston et al. 2000).

At E12.5, a subset of our MiniPromoters: Ple123, Ple131 and Ple170, showed variability in expression. The patterns shown here represented the most frequently observed staining. For Ple123 (N=7 embryos) expression levels ranged from primarily heart-only (n=5) to ubiquitous (n=2). Ple131 (N=5 embryos) displayed heart-only (n=1) to weak ubiquitous (n=3) to strong ubiquitous (n=1) expression. Ple170 (N=3 embryos) ranged from negative (n=2) to ubiquitous (n=1) expression, in contrast with lack of variability in the adult. Interestingly, in contrast to the adult brain, we did not observe variability in the embryo for Ple232 expression. Although three of six constructs showed variability at the E12.5 developmental timepoint, this variability was not observed generally for other MiniPromoter constructs docked at the Hprt locus, at least in adult animals.
Figure 3.4 Expression of six MiniPromoters during development. E12.5 embryos were harvested and stained for lacZ expression using X-gal. Whole-mount (first panels) and partially-cleared whole-mount embryos (second panels) were photographed. (A) Ple26 (CCL27 RRs) stained throughout the entire embryo. (B) Ple123 (ICMT RRs) displayed variability, with the most common expression pattern being limited to the developing heart. (C) Ple131 (MKI67 RRs) was expressed throughout the embryo at varying levels, most prominently in the heart. (D) Ple140 (NR2E1 RRs) showed a very strongly defined pattern in the hypothalamus of the embryo. (E) Ple170 (POGZ RRs) demonstrated staining in all parts of the embryo, but generally stronger outside of the nervous system. (F) Ple232 (TNNT1 RRs) stained strongly in skeletal muscle and the tongue of the embryo, but was not visible in the brain or spinal cord. RRs, regulatory regions.

3.4.5  *Doublecortin (DCX)*-based constructs demonstrated expression driven by a Prom segment and a conserved putative regulatory element in intron four.

We then focused on a comparison of two constructs designed using the *DCX* gene to understand differences in the putative regulatory regions used (Fig. 3.5). The first design, Ple53, was based on a known promoter used in the literature (Couillard-Despres et al. 2005; 2006), but with minor sequence changes during PCR synthesis, and found to be expressed as anticipated in neurogenic regions of adult brain. The second *DCX*-
based construct (Ple55) was designed by analysing the promoter region (LongProm element; Ple53 construct) to isolate a putative minimal promoter sequence. A Prom segment was identified and used in combination with an additional highly-conserved putative regulatory element from within intron 4.

We first looked at the cortical layer specificity of Ple53 and Ple55 (Fig. 3.5C,D). Ple53 expression, as detected by X-gal and counter-stained with neutral red, was limited to the deepest cortical layer VI (Fig. 3.5C left panel). Rare positive cells in layers II, IV and V were observed. In contrast, Ple55 was expressed in all cortical layers, including layer I (Fig. 3.5D left panel). Nearly all layer II, IV and VI neurons were labeled. Both constructs were expressed in the known adult neurogenic regions of the hippocampal subgranular zone, and the subventricular zone and rostral migratory stream axis.

To assess whether the positive cells in the cortex were primarily neurons, we compared the extent of NeuN and Gfap co-labeling with X-gal in the cortex of these two strains (Fig. 3.5C,D center and right panels). Cells were found in both strains that co-labeled (black arrows) and did not co-label (white arrows) with NeuN (Fig. 3.5C,D). We did not observe significant labeling with Gfap. However, we saw extensive non-neuronal labeling of the hippocampal formation in Ple55 (Fig. 3.5D right panel), suggesting oligodendrocyte labeling. The overlapping cortical and neurogenic expression between the two designs suggests an expansion of the expression pattern with Ple55.

At E12.5 (Fig. 3.5E,F) we saw an overall conserved staining pattern for Ple53 and Ple55, consistent with the developing nervous system. Specifically, expression was observed throughout the developing brain, eye, spinal cord and dorsal root ganglia. The labeling extends along both the dorsal and ventral ramus. Expression in the developing kidney was found in partially-cleared embryos (Fig. 3.5E,F white arrows). Ple55 was also expressed in the face and limbs (black arrows), indicative of either pre-cartilage or pre-muscle tissue. Both the nervous system expression and kidney expression has been previously documented for the mouse Dcx gene (Reiner et al. 2006).

In the retina (Fig. 3.5G,H), Ple53 and Ple55 were expressed nearly exclusively in the GCL, with rare INL cells positive. Both constructs were also expressed in the optic nerve. Studies in rat retina suggested horizontal cell labeling for endogenous Dcx (Lee et al. 2003; Wakabayashi et al. 2008). However, only the rare INL cells found to be expressing (indicated with white arrows in Fig. 3.5G,H top-middle panel) our MiniPs may be consistent with
their localized expression in horizontal cells. Thus, we conclude that the retinal expression for Ple53 and Ple55 is ectopic and unexpected. In order to further narrow the characterization of the retinal expression, we performed co-labeling studies. Strong overlap between NeuN-positive nuclei and X-gal (Fig. 3.5G,H top-middle panel), and a lack of overlap with the astroglial marker Gfap (Fig. 3.5G,H top-right panel) was observed. Further analysis by immunofluorescence of β-galactosidase using co-labeling with β-Tubulin III (Fig. 3.5G,H bottom-left panel) and Brn3 (Fig. 3.5G,H bottom-middle panel), both markers of retinal ganglion cells, demonstrated significant overlap, while only rare cells were positive for the cholinergic amacrine marker, Sox2 (Fig. 3.5G,H bottom-right panel, white arrows).

To account for the differences observed between Ple53 and Ple55, we undertook a bioinformatics analysis to identify their putative differential regulation (Fig. 3.5B). The 5′ extension of the Prom segment found in the LongProm used in Ple53, and the additional element from intron 4 added together with the Prom segment in Ple55, were used to search for predicted TFBS. Ple53 was predicted to contain a MafK binding site (UCSC/ChIP-seq ENCODE data) and a BRACH (TFBIND database (Tsunoda and Takagi 1999)) TFBS that is located in the 5′ non-overlapping region of the LongProm and thus missing in Ple55. Furthermore, Ple55 contained two predicted SRF (CONSITE (Sandelin et al. 2004) and JASPAR (Sandelin 2004; Portales-Casamar et al. 2009a) databases) and two RXR/RAR dimer (JASPAR database) TFBSs within the conserved element 3, derived from a region within intron four. These unique TFBSs may account for the expression differences observed.

Interestingly, we were unable to maintain the DCX-based construct strains on the C57BL/6J background after N3 due to lack of allele transmission. We were able to restore normal allele transmission after crossing the allele to the 129S1/SvImJ strain, and hypothesized an unknown genotoxic effect for this allele in C57BL/6J mice.
Figure 3.5 Comparison of two Doublecortin-based MiniPromoter constructs with similar but non-identical expression patterns identifies putative functional TFBSs.

(A) Diagram showing the differences between Ple53 and Ple55. Both constructs contained a minimal ‘Prom’ sequence of 2,353 bp, however Ple53 contained an additional but contiguous sequence and thus was named ‘LongProm’, whereas Ple55 contains the non-contiguous segment 3 in addition to the Prom. (B) In the UCSC genome browser, predicted binding sites for MafK and BRACH were observed in the extended LongProm sequence specific to Ple53. Predicted RXR/RAR heterodimer and SRF sites were found in the segment 3 specific to Ple55. Comparison of the expression of lacZ using X-gal substrate (blue) demonstrated that (C, E, G) Ple53-lacZ
and \((D, F, H)\) Ple55-lacZ were similar. \((C, D)\) However, the comparison also demonstrated that with the Ple53 construct, the majority of cells labeled in the cortex were located in the deeper cortical layers, whereas with Ple55, they were distributed evenly throughout. There were also more cells in Ple55 that did not co-label with NeuN. \((E, F)\) Expression analysis in E12.5 X-gal stained whole-mount and partially cleared embryos from each construct demonstrated that the Ple55 construct contained all of the neuronal-based expression observed in Ple53 plus staining in the pre-cartilage or pre-muscle of the face and limb buds. Both constructs exhibited expression around the periphery of the limb buds, and both included staining of the developing kidneys (white arrows). A close-up of the face and front limb showed the additional staining in Ple55 (black arrows). \((G, H,\) upper panels\) Construct lacZ expression as detected by X-gal reaction in the retina (left panel). Expression for the DCX-based designs was limited to the ganglion cell layer (GCL) for both Ple53 and Ple55, with rare cells observed in the inner nuclear layer (data not shown). NeuN (marking all retinal ganglion and amacrine cells) co-labels with X-gal (center panel), and Gfap (retinal astrocytes) does not (right panel). Rare cells do not co-label with NeuN (white arrows). \((G, H,\) lower panels\) Immunofluorescence co-labeling with Brn3, βtubIII, and Sox2. Co-labeling demonstrated that most lacZ-positive cells stain with βtubIII and Brn3 markers (ganglion cells) for both constructs (left and center panels), and only rarely in Sox2-positive cholinergic amacrine cells (right panel, white arrows). Ctx, cortex; Hipp, hippocampus; NR, neutral red; RRs, regulatory regions; TFBS, transcription factor binding sites. (Scale bars = 100 μm, except for \((G, bottom panel)\) fluorescence images 20 μm.)

### 3.4.6 Similarities in MiniPromoter constructs for individual candidate genes indicate basal promoters usually confer specificity.

It has been proposed that the minimal promoter for a given gene confers the general specificity of expression, but that additional regulatory elements alter either the timing and or fine-tune the regional specificity (Boulaire et al. 2009; Haeussler and Joly 2011). To test the assertion that basal promoters generally confer specificity, we analyzed sets of constructs designed from the same source gene in our combined dataset. For the purpose of this analysis we classified constructs as expressing (with respect to brain only) as: (1) related, due to the use of a known promoter from the literature or having at least some expression similar to endogenous, or; (2) unrelated, having no pattern similar to the endogenous gene. We had 16 such sets where multiple MiniPromoters were generated and tested based on the same gene (Fig. 3.6). Nine of these sets contain constructs which all have an identical basal Prom (or minimal promoter) element within the set. Of these nine, six sets of MiniPs \((\text{C8ORF46, CLDN5, HCRT, ICMT, MKI67, and S100β RRs})\) displayed related expression while the remaining two \((\text{POGZ, RGS16 RRs})\) displayed unrelated expression, and one set had one related and one unrelated MiniPromoter \((\text{NR2E1 RRs})\).

An additional six sets included a combination of constructs that either contained the Prom element or a LongProm element (usually a known promoter in the literature and contiguous but larger than the Prom
element) within the given set (i.e. overlapping basal promoters). All six sets displayed related expression. Five of these six sets (CCKBR, DBH, DCX, GFAP, and PITX3 RRs) had 5′ extensions compared to the Prom, and one set (TRH RRs) contained a 3′ extension compared to the Prom.

Lastly, the CCL27-based set contained constructs with two non-overlapping Prom1 and Prom2 designs. Prom1 was located further upstream and annotated by our team as a putative cortical expression promoter based on CAGE data. A Prom2-based construct, located near the known TSS, was still related, but distinctly different from those based on Prom1. Due to this difference, we classified this as a mixed set.

75% (12/16) of the MiniP sets displayed related expression in all positive constructs. Thus, we conclude that the common Prom element is generally sufficient to confer specificity, as compared to the endogenous mouse gene, and added putative regulatory elements either enhance or reduce expression levels, or alternatively expand or narrow the expression pattern.

**Figure 3.6 Basal promoters confer the majority of expression specificity.**

16 MiniPromoter sets each designed for a different gene and containing more than one expressing MiniPromoter, were analyzed for their relationship to the endogenous gene expression. Sets were grouped into (1) identical basal promoter, (2) overlapping basal promoters, or (3) distinct basal promoters. Sets were scored: related, MiniPromoter and endogenous gene have substantial overlap in expression; unrelated, MiniPromoter and endogenous gene have substantially distinct expression; mixed relatedness, at least one member of the set is related, and at least one member unrelated.

- **Identical Basal Promoter**
  - 6 sets: related expression (C8ORF46, CLDNS5, HCRT, ICMT, MKI67, S100β)
  - 2 sets: unrelated expression (POGZ, RGS16)
  - 1 set: mixed relatedness (NR2E1)

- **Overlapping Basal Promoters**
  - 6 sets: related expression (CCKBR, DBH, DCX, GFAP, PITX3, TRH)
  - 0 sets: unrelated expression

- **Distinct Basal Promoters**
  - 1 set: mixed relatedness (CCL27)

**3.4.7 Seventeen MiniPromoter constructs demonstrated expression in the adult mouse eye.**

The gene set for MiniPromoter design was originally chosen for region- and cell-type specific adult brain expression, leading to the hypothesis that this selection-bias may result in a unique set of promoters which also express in the eye, and is unique compared to currently available eye promoters (see (Hashimoto 2008) for a brief list). We analyzed the expression pattern of 17 MiniPromoters found to be positive in the adult mouse eye (Tables 3.2, 3.3). Within this set, we observed expression in nearly all major retinal cell types. In extra-retinal
tissue, we identified 6 MiniPs positive in the cornea, 6 in the optic nerve and 1 positive in the lens. Fifteen eye positive MiniPs are presented in Fig. 3.7, and two additional ones are compared directly in Fig. 3.5, as part of our DCX MiniPromoter comparison.

We found Ple24 and 25 (CCKBR RRs) were expressed in the GCL and occasionally in the inner part of the INL. Ple26 (CCL27 RRs) was expressed in the cornea, optic nerve, and all retinal layers. Another design for this gene, Ple28, showed expression in all layers and also in the lens, but no expression in the cornea or optic nerve. Ple32 (CLDN5 RRs) was expressed in the GCL and INL. Ple34 (CLDN5 RRs) was positive in the INL and GCL, and was expressed rarely in cells of the ONL (photoreceptors). The morphology of cell nuclei located in the INL with processes extending to the GCL was suggestive of bipolar ON cells. Ple67 (FEV RRs) stained cells in the GCL, and rare INL cells. Ple88, based on a known promoter for GFAP (Brenner et al. 1994), was expressed in the astrocytic layer as expected. Ple123 (ICMT RRs) stained all retinal layers including the cornea and optic nerve. Ple131 (MKI67 RRs), based on proliferative maker Ki-67, stained in patches of the cornea only, consistent with other studies on Ki67+ cells in the cornea (Joyce et al. 1996). X-linked mosaic studies have also demonstrated a similar striping pattern for corneal epithelial wound repair (Mort et al. 2009; 2012). Ple146 (NTSR1 RRs) stained rare GCL and inner INL cells (INL-positive cell not pictured). Ple155 (PCP2 RRs) stained predominantly in the INL and IPL, and was strongly suggestive of bipolar ON cells. Ple167 and 170 (POGZ RRs) stained cells in all layers of the neural retina including the cornea. Ple240 (UGT8 RRs) stained positive in the retinal inner segment (photoreceptors) and occasionally in the INL with fibers extending into the astrocytic layer, suggesting a Müller glial or bipolar cell-type. The cornea and optic nerve was also positive for Ple240.
Figure 3.7 Fifteen MiniPromoters expressed in the mouse eye.
Mice were perfused, eyes harvested, and stained with X-gal. Eyes were cryoprotected, embedded and sectioned. Ple24 (CCKBR RRs), Ple25 (CCKBR RRs), Ple32 (CLDN5 RRs), Ple67 (FEV RRs), and Ple146 (NTSR1 RRs) expressed in a subset of GCL cells. Rarely, an amacrine cell in the innermost part of the INL was positive in Ple24, or a rare positive horizontal cell in Ple32. Ple26 and Ple28 (both based on CCL27 RRs) expressed in the INL, and in photoreceptors as evidenced by the columnar staining of the ONL and photoreceptor segments. Ple26 showed extensive IPL and OPL fiber staining and Ple28 was also positive in the GCL. Ple34 (CLDN5 RRs) and Ple155 (PCP2 RRs) have staining characteristic of bipolar cells, with cell bodies in the INL and processes extending to the GCL (Ple34) or inner half of the IPL (Ple155). Ple88 (GFAP RRs) stained in the retinal astrocytic layer. Ple123 (ICMT RRs) expressed throughout the retina. Ple131 (MKI67 RRs) expressed in cells of the corneal epithelium. Ple167 and Ple170 (both based on POGZ RRs) demonstrated similar staining patterns in the GCL, INL, OPL and photoreceptor segments. Ple240 (UGT8 RRs) had fiber-like staining spanning the IPL and clear cell-body labeling in a subset of INL cells. An annotated negative control retina is shown for reference. CH, chimera; RRs, regulatory regions. (Scale bars = 100 μm)

3.4.8 Knock-in constructs that express in retinal ganglion cells predict expression of promoter-driven GFP using AAV-based intravitreal delivery.

In order to assess the feasibility of using the MiniPromoters for AAV-mediated delivery, we cloned three (Ple25, Ple53 and Ple67) constructs that were expressed primarily in the retinal ganglion cell layer of the retina, into the AAV2-(quad Y-F) vector. This vector, containing four tyrosine to phenylalanine mutations, has previously
been shown to infect cells of the retina uniformly (Petr-Silva et al. 2011). Two (Ple25 and Ple53) of the three constructs were 5' trimmed by ~200-400 bp to fit into the viral vector due to limits on the carrying capacity of AAV constructs. Ple25 was trimmed from 3,740 bp (KI allele) to 3,312 bp (AAV) and Ple53 trimmed from 3,520 bp (KI allele) to 3,310 bp (AAV). Ple67 (2,202 bp) did not require trimming. Mice were injected intravitreally at P28, harvested four weeks later and analyzed for GFP expression. Fig. 3.8A,C,E shows a side-by-side comparison of the knock-in mouse strain and the virus-injected mice. All three viruses induced expression of the reporter gene in the retinal ganglion cell layer. For Ple53 and Ple67 we also observed rare positive cells in the innermost aspect of the INL. We conclude that for at least these three constructs, the expression pattern in the knock-in strain adequately predicted that obtained using a virus-based delivery to the mouse retina. These data suggest that promoter specificity may be maintained in the AAV vectors.

The retinal ganglion cell layer in mice contains approximately 40% ganglion cells and 60% amacrine cells (Jeon et al. 1998). Nadal-Nicolas et al. (2009) estimated that Brn3a+ cells label ~92.2% of RGCs, as co-labelled with the Fluorogold (FG) tracer in rat retina. Similarly, FG was estimated to label 97.8-98.4% of RGCs in rats (Salinas-Navarro et al. 2009). Thus, we considered that ~90% of RGCs in the adult mouse retina should be labeled with Brn3a, and could be used as a reliable marker of retinal ganglion cells. We undertook Brn3a-co-labeling on virus-injected mice for Ple25, Ple53 and Ple67 to assess the extent of expression in ganglion cells, with the remainder of GFP-positive cells in the GCL assumed to be amacrine. As shown in Fig. 3.8B,D,E, the constructs ranked qualitatively from most to least overlap with Brn3a: Ple53 > Ple25 > Ple67. We were also interested in how much expression was observed in amacrine (Brn3a-negative) cells in the GCL. Similarly qualitative ranking of the same constructs from most to least expression in Brn3a-negative but GFP-positive cells ordered the constructs: Ple67 > Ple25 > Ple53. Thus, we conclude that Ple53 stains the greatest proportion of retinal ganglion cells and is most specific to this cell-type, while Ple67 stains a subset of retinal ganglion cells in addition to numerous amacrine cells.
Figure 3.8 Three MiniPromoters that expressed in the retinal ganglion cell layer in knock-in mice maintained this restricted expression pattern from AAV viral vectors.

Three MiniPromoter constructs (A, B) Ple25, (C, D) Ple53, and (E, F) Ple67 that expressed in the GCL when docked 5’ of Hprt driving β-galactosidase were recloned into AAV2(quad Y-F) driving hGFP. (A, C, E) Knock-in mouse retinas stained with X-gal substrate for lacZ (blue) and counterstained with neutral red (first panel) showed similar ganglion expression to intravitreal injected AAV2(quad) hGFP (second panel). (B, D, F) Virus-injected retinas were immunostained for Brn3a (a marker of ganglion cells). First panel shows Brn3a (red) with DAPI (blue), second panel hGFP (green) with DAPI (blue), and third panel the merge (yellow) of Brn3a (red), hGFP (green), with DAPI (grey). GCL, ganglion cell layer; INL, inner nuclear layer; ONL, outer nuclear layer; RRs, regulatory regions. (Scale bars = 100 µm, except for fluorescence images 20 µm.)

3.5 Discussion

Interest in the identification of small promoters for capturing the expression pattern of biologically interesting genes is longstanding. However, historical approaches using the ‘promoter bashing’ technique is laborious and can result in failure with complex gene structures where important regulatory elements may not be located close to the TSS (Visel et al. 2009a; 2009c). The Pleiades approach, with methodology described in detail previously (Portales-Casamar et al. 2010), has the ability to generate novel promoters in a higher-throughput manner. Here we present 18 additional promoters identified by this technique, and further the characterization of 15 previously identified designs.

For clinical purposes, numerous groups are directing their efforts at enhancing the expression levels of human promoters in viral vectors; however, our approach focuses on the importance of physiologically-relevant
expression. The reasoning for this is three-fold: (1) overexpression of therapeutic molecules can be deleterious (Schedl et al. 1996); (2) overexpression of purely research-related molecules such as Cre can cause cytotoxicity (Buerger et al. 2006; Boulaire et al. 2009; Silver and Livingston 2001; Forni 2006); and (3) introduction of artificial elements to enhance expression can pose regulatory challenges due to their unknown specificity, and increase the size of promoter constructs. Thus, our strategy of mimicking endogenous expression levels is likely to overcome these challenges.

However, there are some limitations to the Pleiades design pipeline. First, a small number of constructs displayed highly variable expression. We were unable to find any previous reports in the literature of variable expression for constructs knocked-in at Hprt. Since we observed variability in both embryos and adult brains for most of these constructs, we hypothesize that the variability observed is construct- and sequence-specific. For the most part, the extent of variability in embryos correlated with that observed for adults. It remains to be seen whether such construct-dependent variability is genome-context specific, or would be similarly reproduced in other systems (e.g. virus-based delivery or knock-in at the ROSA26 locus). However, variability between independently-derived mouse lines for other Hprt-docked constructs has been demonstrated to be low and not significant (Farhadi et al. 2003), suggesting that expression variability is not a general characteristic of the Hprt locus. Expression variability was also not a general feature of our previous MiniPromoter work (Portales-Casamar et al. 2010).

Additionally, our data was limited by the inability to attain germline animals for a small subset of constructs. In these cases, chimeras were studied. Although each chimera may not display the entire cellular expression profile, the observed chimeric expression has been a reliable indicator of germline expression (Portales-Casamar et al. 2010).

Lastly, we were limited in characterization of the specific cell-types of expression, particularly for the brain where 100s or 1,000s of different neuronal cell types are predicted to exist (Masland 2004; Stevens 1998). However, this data on regionalized expression of our promoter constructs remain a valuable scientific resource. Importantly, this data represents numerous novel and biologically-significant findings:

First, we identified the zona incerta as a common region of expression for our designs. We hypothesize that overrepresentation of the zona incerta brain region in our set is due to the diverse range of neurochemical
cells located in this region, which includes neurons varying in soma size and shape (see (Mitrofanis 2005) for a comprehensive review on the zona incerta). In addition, extensive networking exists between cells of the zona incerta and numerous other brain regions (e.g., the cerebral cortex, diencephalon, basal ganglia, brainstem and spinal cord) (Mitrofanis 2005), further suggesting a heterogeneous population in this region. Our persistent finding of staining in this region suggests caution for therapeutic delivery where expression in the zona incerta could cause harm.

The brain-positive MiniPromoters provide novel tools for researchers and clinicians alike. Within our toolbox of MiniPs we have now added a collection of constructs that would be applicable to high levels of expression in the cortex and hippocampus (Ple15, Ple16), in oligodendroglia (Ple240), and for use in driving near pan-neuronal expression (Ple26, Ple170), or in the developing nervous system (Ple53, Ple55). Regarding pan-neuronal expression, our MiniPs present a superior solution to the SLICK-H pan-neuronal transgenic mice due to the ability of using our constructs in viral vectors, and better targeting of nearly all brain regions in the adult, as well as extensively in the embryonic brain.

For this purpose, we analyzed two pan-neuronal candidates in detail, Ple26 and Ple170. Both constructs extensively co-stain with NeuN, and both have some non-neuronal staining. We assessed the extent of overlap with NeuN in the CA2-3 region of the hippocampus and found that Ple26 is likely a better pan-neuronal candidate, but also exhibits stronger expression in non-neuronal cells. Thus, the choice between these two constructs would depend on (1) how critical near-complete overlap with neuronal cells is to the application and (2) how important it is to limit off-target non-neuronal expression.

Selected designs were also validated for their use in developmental studies. We tested six adult brain-positive MiniPs and showed expression in the brain for five of these at E12.5. The remaining MiniP, Ple232, was expressed only outside of the embryonic brain, in contrast to its expression in the adult. This finding is not entirely unexpected, since not all genes will be expressed in the brain at the time point used in this study (E12.5, mid-point of neurogenesis). Thus some constructs would not be expected to be positive (e.g. Ple88 based on GFAP RRs, a key gene expressed in astrocytes (Bignami et al. 1972)). A separate study showed robust expression of the human GFAP promoter in mouse from E13.5 (Brenner et al. 1994), further suggesting onset at a later developmental stage. Surprisingly, we showed highly specific expression of Ple140 (NR2E1 RRs) in the
hypothalamus at E12.5, with no evidence of its expression in other embryonic tissue. This data suggests that the Ple140 construct drives gene expression in a cell-type that appears early in neural development and this continues in the adult. Our small subset of embryonic data suggests that many other Pleiades MiniPs are likely to be positive during development. This embryonic validation has also led us to discover the unique functioning of MiniPs outside of the CNS, like Ple232 (TNN1 T Rs) which could prove a valuable tool for researchers studying, and developing treatments for, skeletal muscle disorders.

To better understand how different putative regulatory elements drive MiniPromoter expression, we analyzed both the expression pattern and regulatory element differences for two constructs based on the DCX gene. Both Ple53 and Ple55 expressed in a highly similar fashion with extensive staining in the developing nervous system at E12.5, throughout deep cortical layers and regions of active neurogenesis in the adult brain, and also in the ganglion cell layer of the retina. However, distinct expression was observed at E12.5 in Ple55 that was indicative of pre-cartilage or pre-muscle in the limb buds, as well as in the tongue muscle. SRF is a transcription factor critical for muscle development (Zhao et al. 2005; Li et al. 2005), and a TFBS for it is located in element 3 of Ple55. We propose that this SRF site in Ple55 is likely responsible for the additional expression pattern in embryos and that the tissue identity is likely pre-muscle.

We were further surprised by finding many of our MiniPromoters were expressed in the mouse eye. This represents a novel toolset for driving retinal gene expression and these constructs pose great therapeutic interest. Bipolar cells are the next step in retinal processing after photoreceptors are activated by light (Dowling 1987). In diseases where photoreceptors are non-functional or have undergone cell death, one strategy is to target the bipolar cells. Similarly, since retinal ganglion cells form the last step in the visual processing pathway (Dowling 1987), here also expression of therapeutic molecules could be a viable option, with perhaps more generic retinal disease application. Our efforts have developed at least two MiniPromoters with bipolar cell expression (Ple34 and Ple155) and six with ganglion cell layer expression (Ple24, Ple25, Ple53, Ple55, Ple67, and Ple146). The finding of such a large proportion (35.3%) of eye positives specific to the GCL, implies that selection of candidates for MiniPromoter design using brain expression data may be biased towards generating RGC-positive designs. We hypothesize this may be due to the similarities between RGCs and other CNS neurons with regards to (1) the firing of action potentials, which seemingly no other retinal neurons perform apart from All
amacrines (Wu et al. 2011); and (2) similarities in gene expression between RGC-specific transcripts (Trimarchi et al. 2007) and other CNS neurons which fire action potentials, such as NF-L (Kong and Cho 1999; Julien et al. 1987) and Gap43 (Reh et al. 1993; Capone et al. 1991) genes. Interestingly, the ganglion cell-marking Brn3 transcription factors are also highly expressed in the sensory neurons of the dorsal root ganglia (Badea et al. 2012; Ninkina et al. 1993). Of the six GCL-positive MiniPs, we were able to assess DRG expression in three and found all three positive (Ple24, Ple53, Ple55). Thus our selection strategy has not only preferentially generated MiniPs positive in the GCL of the retina, but also in the DRG (8 out of 10 MiniPs tested).

Furthermore, other MiniPs that expressed in the eye also demonstrate great utility: Ple88 (GFAP RRs) could be used to aid in the treatment of retinal disorders with increased astrogliosis (defined as increased GFAP expression), such as reactive gliosis (Bringmann et al. 2006; Bringmann and Wiedemann 2012) or massive retinal gliosis (MRG) (Inayama et al. 2005; Deshmukh et al. 2011), and Ple131 (MKI67 RRs) may be useful for corneal repair. Altogether this represents a powerful set of tools for driving gene expression in the eye.

Curiously we found occasional discordance between cell-type expression in the brain and eye. Our Ple155 (PCP2 RRs) MiniPromoter was expressed as expected in the mouse retina (bipolar ON cells) (Berrebi et al. 1991; Oberdick et al. 1990), but with seemingly unrelated expression in the brain (our expression was present in cortex, thalamus and hypothalamus but absent in Purkinje cells for which PCP/L7 is famous for (Oberdick et al. 1990; 1988; Berrebi and Mugnaini 1992)). This data further supports the notion that cell-type specificity is regulatory element-specific in specific organs or tissues (Visel et al. 2007a; 2009a). Alternatively, the combination of these regulatory elements at the Hprt locus may negatively regulate their Purkinje expression.

One concern was whether the observed expression pattern of MiniPs would be reliably reproduced when tested using viral vectors. Viral vector gene delivery is the preferred therapeutic method for the brain and eye, primarily since the realm of treatment options is greater for gene delivery than cell-based therapeutics. Thus for this application it is critical to confirm that our strategy of MiniPromoter design results in clinically-useful tools. We hypothesized that our use of single-copy knock-ins and entirely human DNA sequence would limit the failure of these constructs with the methodological transition of transgenic mice to use in viruses, since our constructs were designed using high levels of phylogenetic conservation and should be less prone to transgene inactivation and expression variability compared to more distantly-related sequence or viral promoters. We set
out to preliminarily demonstrate this application of our constructs with three MiniPromoters. We showed that a subset of three MiniPromoters (Ple25, Ple53 and Ple67) that express in the retinal ganglion cell layer in knock-in mice exhibited a similar expression pattern in the mouse retina when delivered intravitreally and expressing from rAAV2. The remarkable paralleling of expression pattern and easy detection with hGFP epifluorescence demonstrate functional equivalency of our promoters when moved from knock-ins at Hprt to AAV2.

An interesting pattern emerged using these three constructs: we observed an inverse qualitative relationship between the overlap of stained retinal ganglion cells (defined here as Brn3a-positive in the adult retina) and amacrine cells (defined as Brn3a-negative) located in the GCL. This result was surprising, as we were anticipating that the MiniPromoter expression may constitute different subsets of ganglion or amacrine cells, with no expression in the other cell type. However, we observed expression overlapping both cell types and our data implies that transcriptional regulation between amacrine and RGC cells may not be distinctly separated. Indeed, in support of this idea are the results of Kunzevitzky et al. (2010) who evaluated the transcriptional profiles using microarrays of amacrine and retinal ganglion cells, isolated using immunopanning, at E20, P5 and P11 in the rat. Surprisingly, they found ~70% overlap in expression between the two cell types at all time points (see Fig. 5 in Kunzevitsky et al. 2010). Both amacrine and ganglion cells are born between E14-E18 (Young 1985a; 1985b). Evidence for potential overlapping transcriptional regulation between the two cell types also comes from work that showed VC1.1+ progenitors can generate both VC1.1+ and VC1.1- progenitors (Alexiades and Cepko 1997; Cepko et al. 1996). The former produces amacrine and horizontal progeny, while the later produces retinal ganglion cells and cones (Cepko et al. 1996).

While our success with ganglion-cell promoters is greatly encouraging and provides preliminary validation of our tools, it remains to be seen whether cell-type specificity is similarly conserved for constructs with different retinal cell-type specificity. Furthermore, consistency between expression pattern in the eye may be a unique characteristic and found to be different for brain. With these concerns in mind, we are currently further testing some of our existing promoters in an intravenous AAV-based delivery system (Foust et al. 2009) as part of the CanEuCre (http://caneucre.org/) effort (Smedley et al. 2011; Bucan et al. 2012; Murray et al. 2012).

A myriad of new data is emerging from large genomic analyses on transcription factor binding sites, such as the efforts of the ENCODE project (Consortium et al. 2013) and VISTA Enhancer Browser project (Visel et
al. 2007b). Using this data, we aim to develop new promoters and make refinements to existing designs in order to maximize the remaining capacity for genes of interest in virus delivery systems.

In conclusion, we presented several new MiniPromoters, especially with application to the mouse retina, and demonstrated that our constructs express similarly when used in viral vectors. This work confirms that MiniPromoter design is a tractable high-throughput option for tissues and shows that it can identify novel promoters that translate across transgene delivery systems. We attribute the success of our design strategy to: (1) the use of evolutionary conserved sequence and human-only DNA limiting transgene inactivation; (2) paring of putative regulatory elements along with their respective core basal promoters instead of unrelated or viral core promoters, the importance of which is summarized in Haeussler et al. (2011); and (3) characterization of reproducible, single-copy knock-ins in a consistent neutral genomic environment. These tools and further expansion of small promoter resources will greatly benefit both basic research and future human gene therapy.

3.6 Summary

In this chapter we expanded the utility of the tools generated through the Pleiades Promoter Project. We characterized the expression of 18 additional novel MiniPromoters for the brain, explored the value of these tools in spinal cord and non-CNS tissues, and importantly, showed that 17 of our MiniPromoters express in the mouse eye. The three clinical trials for Leber’s Congenital Amaurosis published in 2008 (Hauswirth et al. 2008; Cideciyan et al. 2008; Maguire et al. 2008) where AAV2 was used to deliver the RPE65 gene to the eye demonstrates both the cutting edge studies in the field and the importance of gene therapy tools for eye. But it is also important to remain aware of the impact that our brain MiniPs will have in using optogenetic stimulation of defined populations of neurons to uncover their functional role in the live behaving animal. Our MiniPs will allow delivery of ChannelRhodopsin molecules to discrete and defined populations of neurons in the brain, aiding in improving the ‘resolution’ of these functional studies. Furthermore, we showed that three of our retina-positive MiniPromoters retained their restricted expression pattern when used in AAV2(quad). Thus, we have demonstrated the applicability of using these tools in a vector that is highly desirable for its use in gene therapy. We have now established several collaborations with investigators who intend on performing pre-clinical and clinical studies with our MiniPromoters.
In addition to the importance discussed above, the completion of the characterizing of Pleiades MiniPromoters has allowed us to further reevaluate positive designs to generate much smaller constructs, for example those based on the CCKBR, DCX and FEV genes. These more compact designs, tested as part of the CanEuCre endeavour, will enable the delivery of larger therapeutic gene products ensuring that there is greater utility of our tools. Further work will demonstrate the biological use of these tools to address fundamental issues in basic and clinical sciences. A study demonstrating the use of Pleiades tools to evaluate the functional role of a key brain development gene, NR2E1, follows in the next chapter.
Chapter 4: Ubiquitous expression of nuclear receptor 2E1 results in embryonic death and ectopic expression fails to rescue \( \text{Nr2e1}^{frc/frc} \) defects

4.1 Preamble

Appropriate brain development is predicated on the fine balance between proliferation and differentiation of neural stem cells. A key molecule, \( NR2E1 \), is an orphan nuclear receptor transcription factor that plays a significant role in this balanced developmental potential of neural stem cells. Since discovery of the null allele for \( \text{Nr2e1} \) in mice (e.g. \( \text{Nr2e1}^{frc/frc} \)), we have been incredibly interested in both identifying methods for performing rescue of the \textit{fierce} phenotype (as a pre-clinical step for neural stem cell modulation), but also in better understanding of the role of this transcription factor in development. Although \( NR2E1 \) was a gene chosen for MiniPromoter design as part of the Pleiades Promoter Project, we were unsuccessful in developing MiniPromoters that recapitulate a subset of, or an accurate representation of, the endogenous expression pattern. Instead, we have pursued another commonly-used genetic paradigm for understanding gene function: overexpression and ectopic expression. In this chapter, I explored the phenotypic outcomes of CAG promoter based ubiquitous/overexpression and Ple53 (DCX RRs) based ectopic expression. This is the first study employing Pleiades MiniPromoters and tools to address a biological question.

4.2 Introduction

Neural stem cells (NSCs) control the fine balance of proliferation and differentiation during embryonic brain and eye development as well as in the adult brain. NSC dysfunction has now been implicated in various psychiatric disorders, including bipolar disorder, depression and schizophrenia (Thomas and Peterson 2008; Young 2009; Mao et al. 2009; Higashi et al. 2008; Hitoshi et al. 2007; Laeng et al. 2004). In addition, NSC function also extends to a role in tumour initiation and development (Singh et al. 2004; 2003; Zaidi et al. 2009;)

---

4 This chapter is in preparation. de Leeuw CN, Simpson EM. (201_). \textit{Ubiquitous Expression of Nuclear Receptor 2E1 Results in Embryonic Death and Ectopic Expression Fails to Rescue \( \text{Nr2e1}^{frc/frc} \) Defects}. \textit{In preparation.} See Preface for details of my contribution.
Emmenegger and Wechsler-Reya 2008). Thus, future work on the modulation of NSCs may result in viable therapeutic strategies for a variety of disorders with neural stem cell dysfunction.

The nuclear receptor 2E1 (NR2E1, also known as Tailless or Tlx) is an important regulator of neural stem cells, particularly in the forebrain and dorsal midbrain (Monaghan et al. 1995; Roy et al. 2002). NR2E1 has itself been implicated in psychiatric disorders (Kumar et al. 2008), behavioural abnormalities in mice (Abrahams 2005; Wong et al. 2010), cortical malformation (Kumar et al. 2007b), found expressed in human brain tumours (Park et al. 2010; Sim et al. 2006; Phillips et al. 2006) and is involved in tumour formation in mice (Zou et al. 2012; Liu et al. 2010). However, little is known about the cellular function of Nr2e1 and how it regulates stem cell dynamics. It has been demonstrated that Nr2e1 represses inhibitors of cell cycle progression and genes expressed in differentiated cell types (Shi et al. 2004; Sun et al. 2007; Zhang et al. 2006), but little is known about how it performs those functions. Furthermore, it is unknown whether these functions are conserved in other cell types, or would only be observed when it is expressed according to its endogenous pattern.

One method for deciphering the cellular role of a transcription factor is to pursue overexpression or ectopic expression studies. Ectopic expression of the Drosophila homolog of Nr2e1, tailless, has previously been shown to induce supernumerary neuroblasts (Kurusu et al. 2009), suggesting a positive effect on proliferation. However, in mice only endogenous-based overexpression has been studied (Liu et al. 2010; Wong 2009). Although these mice do display some phenotypes, it is only in aged mice or on the p53 mutant background that severe histological and phenotypic abnormalities were identified (Liu et al. 2010).

When it comes to Nr2e1 interactors, little is known. It has been demonstrated that Nr2e1 genetically interacts with another neural stem cell gene, Pax6, through specification of the pallial-subpallial boundary in the telencephalon (Stenman 2003). Furthermore, it has been shown in Xenopus that exogenous Xtl1 (Xenopus homolog) fused with the engrailed repressor interferes with Xpax6 (Pax6 homolog) expression and eye formation (Hollemann et al. 1998). Similarly, Pax6 expression was shown to be reduced in the eyes of mice overexpressing Nr2e1 (Wong 2009). However, in the mature retina Pax6 is a marker of amacrine, ganglion and horizontal cells (de Melo et al. 2003; Stewart et al. 2005), and thus the relationship between Nr2e1 levels and Pax6 in neural stem cells or brain remains unclear. Interestingly, Pax6- ESCs demonstrate reduced capability of generating neurons in vitro, despite normal growth parameters in ESC cultures and in their ability to generate
embryoid bodies (Quinn et al. 2010). Work in ESCs has shown that while PAX6 plays an important function in neurectoderm specification in humans ESCs, it is not required in mouse (Zhang et al. 2010). This data necessitates the need to better understand the interaction with Nr2e1 and Pax6.

We wanted to address several hypotheses: first, we hypothesized that expression of NR2E1 in ESCs would result in a growth advantage. Second, we hypothesized that forced non-specific ubiquitous expression of NR2E1 would result in complete failure of embryogenesis. Third, we hypothesized that ubiquitous expression of NR2E1 would generate NSCs in vitro, and the resultant constitutive overexpression in NSCs would block their ability to differentiate into neurons. Lastly, we hypothesized that neuronal-based expression in mice would cause overt brain abnormalities.

To address these hypotheses, and to establish the safety of NR2E1 modulation, we chose to develop novel ubiquitous- and ectopic-expressing mouse strains of human NR2E1 using the Hprt knock-in strategy (Bronson et al. 1996; Yang et al. 2009; Schmouth et al. 2012b; 2012a; Portales-Casamar et al. 2010). Previously we demonstrated near complete functional equivalence of the human and mouse genes using large genomic clones containing human NR2E1 (Abrahams 2005; Schmouth et al. 2012a), in part due to the high conservation both at the DNA and protein levels (Abrahams et al. 2002). This data along with human disease-relevance drove the decision to use human NR2E1 for these studies.

In this report we demonstrate that ubiquitous expression, driven by the CAG promoter, of NR2E1 leads to embryonic developmental failure at ~E8. We also show that ectopic neuronal (DCX promoter driven) expression in migrating neurons does not produce gross phenotypic abnormalities. We then investigated whether the fierce (an Nr2e1-null allele) phenotype could be rescued by NR2E1 expression via a DCX-based promoter. Although our DCX mice failed to rescue Nr2e1frc/frc phenotypes, we observed a genetic interaction with Nr2e1frc/+ mice, which was manifested as sudden death in the adult. Importantly, we demonstrated a reduced number of Pax6-positive putative NSCs, when these were generated from ubiquitously expressing NR2E1 ESCs in vitro. These findings suggest that NR2E1 can function outside of its normal domain in mammals. Furthermore, the data suggests that as with Pax6 (Sansom et al. 2009), there may be a fine balance of NR2E1 expression required to maintain the appropriate levels of self-renewal and progenitor generation from NSCs. Lastly, this
work indicates that brain-delivered NR2E1 or siRNA may remain viable therapeutic options, but that caution should be taken as systemic distribution of NR2E1 could have unwanted effects.

4.3 Materials and methods

4.3.1 Generation of the constructs

A novel construct was designed containing a multiple cloning site (MCS), intron, Full Flip-Recombinase Target (FullFRT) site mutant F5 site, N-terminal FLAG-tagged human NR2E1 cDNA, FullFRT wild-type site, SV40 polyadenylation signal and an insulator sequence. This EcoRI-flanked fragment containing the pEMS1348 backbone elements was synthesized by GeneArt (Germany). The pEMS1307 backbone (Yang et al. 2009) was digested with EcoRI and the Hprt complementary fragment containing the AmpR gene and 5’ and 3’ Hprt homology arms retained. The synthetic fragment from GeneArt and the Hprt homology fragments were ligated together to yield the pEMS1348, according to standard cloning procedures. The CAG promoter fragment was cloned into this backbone using SacI and MluI enzymes to yield the pEMS1349 CAG-FLAG/NR2E1 vector. The Ple53 promoter fragment was excised from pEMS1199 (Portales-Casamar et al. 2010) and cloned into this backbone using FseI and AscI enzymes to yield pEMS1452 Ple53-FLAG/NR2E1 (also known as DCX-FLAG/NR2E1).

4.3.2 Antibody production

Due to the lack of successful commercial Nr2e1 antibodies, we generated a novel mouse monoclonal antibody against human NR2E1. Human NR2E1 cDNA was fused in frame with the N-terminal Lumio and 6xHis tags using the pET160GW-D-TOPO plasmid. E. coli BL21* was transformed using this plasmid and recombinant human NR2E1 protein produced using MagicMedia™ E. coli Expression Medium (containing IPTG; Invitrogen) and isolated using the ProBond Purification system (Invitrogen). 50 µg of purified protein was used to immunize a female BALB/cJ mouse, boosted twice with two weeks between injections, and then given a final IV and IP boost of protein 5 days prior to harvesting splenocytes. Hybridoma cell lines were created according to protocol (ClonaCell™–HY Hybridoma Kit, StemCell Technologies) by fusion with P3X63-Ag8.653 myeloma cells. We picked 612 hybridoma clones. Of these, 571 hybridomas were screened using ELISA with BL21* lysate for negative selection and human NR2E1 BL21* lysate for positive selection. Fifty-six clones that showed specific
signal above background were selected, expanded and 27 successfully frozen. Subsequently clones were screened with dot blots, Western blots, and were tested on HEK293 mock and CAG-FLAG/NR2E1 transfected cells. One particularly strong clone, 1A7 was identified and used further.

4.3.3 Transfection and immunofluorescence on HEK293

HEK293 cells were grown in 24-wells on coverslips and transfected with pEMS1349 (CAG-FLAG/NR2E1) using Lipofectamine-2000 (Invitrogen) according to manufacturer’s instructions. Cells were harvested 48 hours after transfection, fixed with 4%PFA/PBS for 10 minutes at room temperature and stored in 0.02% Azide/PBS at 4°C. Immunofluorescence was performed according to standard procedures. For FLAG-detection, we used the mouse monoclonal anti-FLAG antibody (Sigma-Aldrich, F1804) or rabbit polyclonal anti-FLAG antibody (Sigma-Aldrich, F7425). Secondary antibodies used included goat anti-mouse Alexa-488 and goat anti-rabbit Alexa-594. Coverslips were mounted with Prolong Gold anti-fade reagent (Invitrogen).

4.3.4 Generation of knock-in mice

Single-copy knock-in mice at the Hprt\textsuperscript{b-m3} locus on the X-chromosome were generated as previously described (Yang et al. 2009; Portales-Casamar et al. 2010). Briefly, 30 μg of the pEMS1348, pEMS1349 and pEMS1452 vectors were linearized using I-SceI restriction enzyme and electroporated in B6129F1-\textit{Hprt}\textsuperscript{b-m3}/\textit{Y} (mEMS1202.02 passage 10) mouse ESCs (Yang et al. 2009) or newly generated 129B6-\textit{Hprt}\textsuperscript{b-m3}/\textit{Y}, \textit{Gt}(ROSA)26Sor ESCs (made using B6;129S-\textit{Gt}(ROSA)26Sor/J mice (JAX Stock#002073; expresses β-galactosidase ubiquitously (Zambrowicz et al. 1997)). Electroporated ESCs were plated on wild-type mitomycin C-inactivated mouse embryonic fibroblasts (MEFs) and selected over a period of 10-12 days using hypoxanthine-aminopterin-thymidine (HAT) containing media. Independent clones were identified, isolated and grown. ESCs were harvested for DNA and assayed by PCR for \textit{Hprt} complementation sequence (\textit{HprtCS}), \textit{Hprt}\textsuperscript{b-m3} deletion (\textit{Hprt-Null}), FLAG/NR2E1 (FLAG), DCX/Ple53 and CAG promoter assays to ensure a pure population of ESCs were selected. Only ESCs that successfully passed all PCR assays were further employed.

Chimeras were generated by blast microinjection of correctly targeted ESC clones into B6(Cg)-\textit{Tyr}\textsuperscript{22}/\textit{J} (JAX Stock#000058; B6-Albino) host blastocysts. Chimeras, as identified by coat colour (see (Yang et al. 2009) for details), were bred to B6-Albino females and germline offspring identified as agouti in colour.
Three novel untargeted ESC lines (mEMS55529.02, mEMS55537.02 and mEMS55544.02), which contain the ubiquitous β-gal expressing Gt(ROSA)26Sor (Zambrowicz et al. 1997) allele in combination with the Hprt\textsuperscript{-m3} deletion, were generated. The CAG-FLAG/NR2E1 plasmid (pEMS1349) was retargeted into a germline capable ESC line (mEMS55529.02) and three correctly targeted CAG-FLAG/NR2E1-containing clones (mEMS55651, mEMS55652 and mEMS55658) were selected and expanded as per our standard procedure.

Chimeras generated with lacZ-marked cells were initially crossed to B6-Albino, and subsequently crossed to 129-XXist\textsuperscript{10lox}/X females as described previously (Yang et al. 2012), resulting in 129B6-XXist\textsuperscript{10lox}/XHprt\textsuperscript{CAG-FLAG/NR2E1} germline female animals.

For all work with the 129-mEMS3791 (Ple53-FLAG/NR2E1) strain, only male animals were used. For the aging study, mice were N4 on 129S1/SvlMJ background. For the brain weight and retinal funduscopy experiments mice were N5-N6 for 129-mEMS3791, subsequently crossed to 129-Nr2e\textsuperscript{1frc/+} (N28) to generate double heterozygous females and lastly bred to B6-Nr2e\textsuperscript{1frc/+} (N28) to yield 129B6F1-mEMS3791, frc animals.

Animals were housed at the Centre for Molecular Medicine and Therapeutics, in a barrier pathogen-free facility. Animals were kept on a 14-hour ON light cycle, and provided with food and water \textit{ad libitum}. All procedures involving animals were in accordance with the Canadian Council on Animal Care (CCAC) and UBC Animal Care Committee (ACC) (Protocol #A07-0430, A09-0980, A09-0981, A11-0370).

### 4.3.5 PCR analysis of genomic DNA

Vector NTI software (Invitrogen) was used to design constructs and devise cloning strategies and PCR protocols. Genotyping primers are shown in Table 4.1. Tissue for genotyping was digested with 200 μl of mouse homogenization buffer [50 mM KCl, 10 mM Tris-HCl, 2 mM MgCl\textsubscript{2}, 0.1 mg/ml gelatin, 0.45% Igepal CA-630, 0.45% Tween-20, pH 8.3 with 24 μg Protease K] overnight at 55°C. Digested samples were heat-inactivated at 95°C for 10-20 minutes.
### Table 4.1 Genotyping assays and PCR primer sequences

<table>
<thead>
<tr>
<th>Assay</th>
<th>Primer</th>
<th>Direction</th>
<th>Sequence (5'→3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>HprtCS</td>
<td>oEMS2267</td>
<td>Forward</td>
<td>TCAGCCGAGACCTCTGGCCTT</td>
</tr>
<tr>
<td></td>
<td>oEMS2267</td>
<td>Reverse</td>
<td>TCAGCCGAGACCTCTGGCCTT</td>
</tr>
<tr>
<td>Hprt-Null</td>
<td>oEMS2236</td>
<td>Reverse</td>
<td>TCCAGGGAATCCATCAACTAC</td>
</tr>
<tr>
<td></td>
<td>oEMS2240</td>
<td>Forward</td>
<td>GCATCCAGTGCTCTCCCT</td>
</tr>
<tr>
<td>CAG</td>
<td>oEMS2364</td>
<td>Forward</td>
<td>GCGTATCACGGGCCTCC</td>
</tr>
<tr>
<td></td>
<td>oEMS2670</td>
<td>Reverse</td>
<td>GCGTATCACGGGCCTCC</td>
</tr>
<tr>
<td>DCX</td>
<td>oEMS4210</td>
<td>Forward</td>
<td>GCACGCTAAAGAACCATTATTAT</td>
</tr>
<tr>
<td></td>
<td>oEMS4211</td>
<td>Reverse</td>
<td>CAGGGGATAATGACTCTTTTC</td>
</tr>
<tr>
<td>FLAG/NR2E1</td>
<td>oEMS2641</td>
<td>Reverse</td>
<td>AGCATCACTGGTCTTGGCA</td>
</tr>
<tr>
<td></td>
<td>oEMS2667</td>
<td>Forward</td>
<td>GCGGCCACCATGGATTACA</td>
</tr>
<tr>
<td>X/Y</td>
<td>oEMS2255</td>
<td>Forward</td>
<td>CTGAGCTTTTGCCCTT</td>
</tr>
<tr>
<td></td>
<td>oEMS2256</td>
<td>Reverse</td>
<td>CCACACCAATTTGTTC</td>
</tr>
<tr>
<td>IL-2</td>
<td>oEMS2200</td>
<td>Forward</td>
<td>CTGAGCAGCAGATGAAAGATCT</td>
</tr>
<tr>
<td></td>
<td>oEMS2201</td>
<td>Reverse</td>
<td>GAGGTGGAATTCTAGCATCC</td>
</tr>
</tbody>
</table>

### 4.3.6 Embryo harvesting and staining

Female B6-Albino and 129-XXist<sup>llox</sup>/X females mice were crowded for 7-10 days and set-up in the afternoon for timed-pregnancies using chimeric mEMS5658 or C57BL/6J studs, respectively. Plugs were checked for 5 days in the morning. The day of plug was considered embryonic day 0.5 (E0.5). Embryos were harvested at various time-points (E9.5, E12.5, E14.5 and E18.5) and were fixed in 4% PFA/PBS for 2-6 hours for histology, or alternatively harvested and flash-frozen for RNA or protein studies. Yolk sac DNA was extracted to confirm genotypes. Embryos were sectioned at 12 μm on Superfrost Plus slides and stained by standard protocol for hematoxylin and eosin.

For immunofluorescence embryos were washed with 1X PBS, blocked in 3% BSA and 5% NGS in PBS for 30 minutes, and incubated in primary antibodies at room temperature in blocking solution overnight. Slides were washed several times with 1X PBS and incubated in secondary antibodies in blocking solution for 1 hr. Slides were washed again and mounted in Prolong Gold antifade reagent.

### 4.3.7 RT-PCR of ESCs and embryos

Embryos from timed-pregnancies and mouse tissues were flash-frozen in N2(l) and stored at -80°C until extracted for RNA. RNA extracts were prepared using the Qiagen RNeasy Mini Kit according to manufacturer's
instructions. 1 µg of RNA was treated with DNase I to destroy any genomic DNA contamination and was subsequently used for RT-PCR. Mouse Gapdh was used as an internal control.

*Embryonic Stem Cells.* 23.8 ng of DNase-treated RNA was used directly in OneStep RT-PCR (Qiagen) and amplified for 34 cycles in each assay.

*Embryos.* 200 ng DNase-treated RNA was used for cDNA synthesis using the qScript cDNA SuperMix kit (Quanta Biosciences). For Gapdh, 2 µl of a 1:10 dilution of cDNA was used in the PCR reaction and 32 cycles of amplification was performed. For Nr2e1 and FLAG/NR2E1, 2 µl of undiluted cDNA was used and 40 cycles of amplification was performed.

*Mice.* 23.8 ng DNase-treated RNA was used directly in OneStep RT-PCR (Qiagen) and amplified for 32 cycles for Gapdh and 34 cycles for FLAG/hNR2E1.

Primers for RT-PCR are in Table 4.2.

### Table 4.2 RT-PCR assays and primer sequences

<table>
<thead>
<tr>
<th>Assay</th>
<th>Primer</th>
<th>Direction</th>
<th>Sequence (5’-3’)</th>
<th>Amplicon size</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gapdh</td>
<td>oEMS2574</td>
<td>Sense</td>
<td>CTCGTCCTAGACAAAGATGGTGGAAG</td>
<td>305 bp</td>
<td>Barberi T. (2003)</td>
</tr>
<tr>
<td></td>
<td>oEMS2575</td>
<td>Antisense</td>
<td>AGACTCCACGACATACTCGACACC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nr2e1</td>
<td>oEMS1633</td>
<td>Sense</td>
<td>GGCATGAATCTGACAACACACGAC</td>
<td>324 bp</td>
<td>Abrahams B. (2005)</td>
</tr>
<tr>
<td></td>
<td>oEMS1637</td>
<td>Antisense</td>
<td>GCTAATTGACCGTAAAGCTGGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLAG/hNR2E1</td>
<td>oEMS3532</td>
<td>Sense</td>
<td>TGCTTTTGGCAGCTTTACG</td>
<td>300 bp</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>oEMS3533</td>
<td>Antisense</td>
<td>TTAACACAAACATATCGACATT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### 4.3.8 Generation of ESCs by in vitro fertilization

Sperm from a germline CAG-FLAG/NR2E1 chimera (mEMS3836 ESC line) was frozen and tested successfully via in vitro fertilization. ESC lines were derived as per standard protocol for embryos generated by breeding as previously described (Yang et al. 2009). Briefly, IVF-generated embryos were cultured to 3.5 days post fertilization (dpf) in KSOM+AA media under oil at 37°C. Each blastocyst was transferred to a single 96-well containing previously grown wild-type mitomycin C-inactivated MEFs and cultured in KSR media. Once large enough clumps were visible, colonies were trypsinized and transferred to 1x24-well. Once confluent, cells were split to 3x24-well into 100% ESC media. Subsequently all three wells were combined at confluency and frozen as 3 vials for each individual clone.
4.3.9 Cytotoxicity assay on ESCs

Three untargeted 129B6-Hprt<sup>b<sup>m</sup>Y</sup>, Gt(ROSA)26Sor ESC lines (mEMS55229.02, mEMS5537.02 and mEMS5544.02) and three CAG-FLAG/NR2E1 targeted ESC clones in the parental mEMS5529.02 cell line (mEMS5565, mEMS55652 and mEMS55658) were seeded at 500 cells per 96-well on wild-type mitomycin C-inactivated MEFs. Cells were grown for 4 days. alamarBlue® (Invitrogen) was added to the growth media as per manufacturer’s instructions and cells incubated at 37°C (bacterial incubator) for 4 hrs. Fluorescence (544 nm excitation, 590 nm emission, 200 flashes, 900 gain) was measured on a POLARstar Omega plate reader (BMG Labtech).

4.3.10 In vitro differentiation of ESCs

Two untargeted 129B6-Hprt<sup>b<sup>m</sup>Y</sup>, Gt(ROSA)26Sor ESC lines (mEMS55229.02 and mEMS5544.02) and two CAG-FLAG/NR2E1 targeted clones in the parental mEMS5529.02 cell line (mEMS55651 and mEMS55652) were in vitro differentiated according to the protocol by (Barberi et al. 2003). Briefly, MS-5 stromal feeder cells were grown in 24-well plates on glass coverslips. After the MS-5 cells reached confluency, two sets of plates were seeded with ESCs at a density of 500 cells/cm<sup>2</sup> and the media changed to serum replacement media (SRM; K-DMEM media, 15% KSR, 2 mM L-glutamine, 10 μM β-mercaptoethanol). Cells were fed with new media every second day. On day 6, one set of plates were harvested by trypsinizing the cells for 5 minutes (0.25% Trypsin-EDTA), inactivating with soybean trypsin inhibitor (STI) and replating the cells in the same wells. Cells were grown for another 24 hrs. This procedure was performed to spread out the highly three-dimensional colonies that had formed. The other set of day 6 plates were changed to N2+AA media (D-MEM high glucose media, 1X N2-supplement, 2 mM L-glutamine, 100 μM ascorbic acid) and fed on days 9 and 12. On day 14, cells were harvested as described for day 6 and allowed to recover for 24 hours. After the recovery period, both groups were fixed with 4% PFA/PBS solution for 10 minutes at room temperature. Plates were stored in 0.02% Azide/PBS solution.

4.3.11 Histological assessments on DCX-FLAG/NR2E1 mice

Two Ple53-FLAG/NR2E1 males and two Wild-type littermates were couriered to the University of Toronto’s Center for Modeling Human Disease (CMHD) to undergo whole body phenotype investigation and
characterization. This process involved the full necropsy and analysis of 40 tissues, followed by histopathology analysis by a veterinary pathologist. The animals were provided to CMHD staff blind to genotype and the specific genetic manipulation to allow unbiased assessment. After the report was completed, we provided CMHD with this information and requested a reassessment of the nervous system and vital organs.

4.3.12 Histology of adult brains

Adult mice were anesthetized with avertin and perfused transcardially with 4% PFA (paraformaldehyde) in 0.1 M phosphate buffer (pH 7.4) for 15 minutes. Brains were dissected and post-fixed in the same solution for 2 hours and transferred into 25% sucrose-PBS overnight. Each brain was sectioned using a Microm HM 550 Cryostat and 16-μm sagittal sections were collected. Expression of the β-galactosidase (lacZ) was detected with X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) staining as previously described (Portales-Casamar et al. 2010; Schmouth et al. 2012a), slides counterstained with neutral red, gradually dehydrated in ethanol and xylene and mounted with Permount and coverslips prior to imaging. Brightfield images were taken using Olympus BX61 motorized microscope using the DP Controller software. For immunofluorescence, standard procedures were followed. Briefly, brains were permeabilized and blocked non-specifically with 3% BSA, 5% NGS and 0.5% Triton-X 100 in PBS for 1 hour. Slides were incubated in primary antibody in 3% BSA/PBS solution overnight. Primary antibodies used were: mouse anti-Gfap (NEB, mAB3670), rabbit anti-Olig2 (Santa Cruz, sc-48817), mouse anti-NeuN (Chemicon, MAB377) and chicken anti-β-gal (Abcam, ab9361). After several washes of 1xPBS, slides were incubated with secondary antibodies goat anti-mouse Alexa-488, goat anti-rabbit Alexa-594 and goat anti-chicken Alexa-647 (Invitrogen) in 3% BSA/PBS for 1-2 h at room temperature. Nuclei were stained using 1 μg/mL of Hoechst 33342 (Sigma, #14533). Finally, slides were washed with 1xPBS, mounted with Prolong Gold anti-fade reagent (Invitrogen) and imaged using confocal laser scanning microscopy on a Leica SP5 II microscope and LAF software.

4.3.13 Histology on eyes

After whole body perfusion of mice, eyes were post-fixed in 4% PFA along with brains. Eyes were then post-fixed in 4% PFA for 2 hours prior to cryoprotection in 25% sucrose-PBS overnight. Eyes were embedded in OCT, along with positive and negative controls, and cryosectioned at 16 μm and mounted on SuperFrost Plus
slides. For lacZ immunohistochemistry, staining was performed using the X-gal substrate for 18 h at 37°C. Eye sections were counterstained with neutral red for 45 seconds to 1 minute. Slides were mounted dehydrated step-wise in ethanol and xylene, mounted with Permount and coverslipped. Brightfield images were taken using Olympus BX61 motorized microscope using the DP Controller software.

4.3.14 Funduscopy

Funduscopy was performed as previously described (Schmouth et al. 2012a; Abrahams 2005; Hawes et al. 1999). Eyes were dilated with 1% atropine/PBS and photographed after 30 minutes. Animals were manually restrained without sedation.

4.3.15 Image processing

Images were processed using ImageJ, Adobe Photoshop and Adobe Illustrator. Brightness, contrast and scaling adjustments were performed as necessary.

4.3.16 Statistical analyses

Statistical analyses were performed using Microsoft Excel and STATISTICA. Results were considered significance with a P-value < 0.05. For samples representative of a population, standard error of the mean (SEM) was indicated with error bars. Otherwise standard deviations were used for error bars.

4.4 Results

4.4.1 Generation of mice ectopically expressing NR2E1

A FLAG/NR2E1-containing backbone was synthesized, which contains the cDNA for human NR2E1 protein with an N-terminal FLAG tag. We then cloned the human cytomegalovirus immediate-early enhancer/modified chicken beta-actin (CAG) and Ple53 (based on DCX regulatory regions (RRs), from (Portales-Casamar et al. 2010) and de Leeuw CN and Simpson EM, in preparation) promoters into the FLAG/NR2E1 backbone (Fig. 4.1A). We first demonstrated appropriate expression of FLAG/NR2E1 in transfected HEK293 cells using antibodies against human NR2E1 and the FLAG tag by Western blot (data not shown) and immunofluorescence (Fig. 4.1B). As expected, the protein was located in the nucleus (Fig. 4.1B) with both antibodies co-localized. In many cells, we noted the FLAG/NR2E1 protein in discrete foci of the nucleus with notable exclusion of nucleoli (Fig. 4.1C). This punctate pattern of localization in the nucleus is a hallmark of
activated nuclear receptors (Stenoien et al. 2000; Fejes-Tóth et al. 1998). Our data confirms that the N-terminal FLAG tag does not interfere with NR2E1 localization or activation.

Subsequently, we targeted CAG-FLAG/NR2E1 and Ple53-FLAG/NR2E1 into Hprt<sup>b-m3</sup>/Y mouse embryonic stem cells (ESCs) (Yang et al. 2009). No difference was observed in successful targeting events between the three constructs (Fig. 4.1D). We analyzed expression of the CAG-FLAG/NR2E1 allele by RT-PCR in ESCs and detected positive signal for three clones (mEMS3829, mEMS3831 and mEMS3836; Fig 4.1E), demonstrating appropriate expression in ESCs for this promoter.

Chimeras were generated and bred to B6-Albino to obtain germline animals. For one CAG-containing clone, three germline chimeras were generated (chimerism ranged from 6-30.5%), while another clone yielded one germline chimera with only one male pup, despite extensive breeding efforts. For the DCX-containing cell line, we obtained 2 germline chimeras (12-22% chimerism) with a total of 26 females and 26 males (P = 1.00). Germline female animals for the DCX chimera were obtained (containing the restored chimeric HPRT locus and the expression construct 5' of HPRT (Bronson et al. 1996)), as assessed by coat colour (Yang et al. 2009) and verified by genotyping. However, only germline males (which do not contain the modified X-Chromosome) for CAG-FLAG/NR2E1 were born. Exhaustive breeding was performed and yet no female germline animals were obtained (0 females, 44 males; P = 3.28 X 10<sup>-11</sup> χ-test). This data indicated lethality associated with the targeted X-Chromosome containing CAG-FLAG/NR2E1. Rarely, a reduced proliferation phenotype has been noted in Nr2e1<sup>-/-</sup> mice (Liu et al. 2008), thus we attempted breeding of the CAG allele to both the 129S1/SvImJ-Nr2e1<sup>frc/+</sup> and C57BL/6J-Nr2e1<sup>frc/+</sup> strains in an attempt to rescue the lethality. Again, we were unable to get live-born female pups (0 females, 31 males; P = 2.58 X 10<sup>-8</sup> χ-test). We conclude that the CAG-FLAG/NR2E1 allele represents a dominant embryonic lethal phenotype, which cannot be rescued by random X-inactivation in females nor by a heterozygous deletion of Nr2e1.
Figure 4.1 CAG-FLAG/NR2E1 allele expresses in targeted mouse embryonic stem cells.

(A) FLAG/NR2E1 allele diagram. An EcoRI-flanked fragment was synthesized containing an MCS, intron, Full-FRT F5 mutant, FLAG/NR2E1 cDNA, Stop codon, Full-FRT Wt, SV40 early polyadenylation signal, and an insulator fragment. The MCS was used to clone in the CAG and Ple53 (DCX regulatory regions) promoter constructs. Once inserted into the genome, the promoter-cDNA segment is located 5’ of the restored chimeric human-mouse HPRT locus. (B) HEK293 cells were mock transfected and transfected with the vector containing the CAG promoter driving FLAG/NR2E1 expression (pEMS1349). An antibody against human NR2E1 (in-house generated monoclonal) and an anti-FLAG antibody were used to show localization of transfected cells (co-localized in yellow). (C) A close-up of a transfected clone shows nuclear localization of the FLAG/NR2E1 protein, including exclusion from nucleoli and nuclear puncta resembling that of activated nuclear receptors. (D) Many clones were observed for promoter-less and promoter-containing constructs. No difference in the percentage of clones passing molecular biology verification was observed. (E) RT-PCR demonstrating expression of the CAG promoter-driven FLAG/NR2E1 in undifferentiated mouse embryonic stem cells. MCS, multiple cloning site; N/A, not applicable; RRs, regulatory regions. Scale bar for B is 100 µm; C is 20 µm..

4.4.2 Female CAG-FLAG/NR2E1 embryos show developmental delay and failure of cerebral lobe fusion at E9.5, and start to die by E12.5

To establish the estimated embryonic age of death of CAG-FLAG/NR2E1 embryos, we set-up timed pregnancies using the germline chimera as a stud male and harvested embryos at four time-points: E9.5, E12.5,
E14.5 and E18.5 (Fig. 4.2A). Embryos were classified as histologically ‘analyzable’ if they were not resorbed and structures were still visible. At E9.5 all male (wild-type) and female (CAG-FLAG/NR2E1) embryos were analyzable. Eighty % of male embryos were normal according to staging criteria (EMAP eMouse Atlas Project (http://www.emouseatlas.org)), while only 20% of female embryos carrying the CAG-FLAG/NR2E1 were normal ($P = 0.0322$, Fisher’s exact test one-tailed). At E12.5, 91% of male embryos were analyzable, and 90% of female embryos analyzable. Of the analyzable embryos, 100% of male embryos were normal and 0% of female embryos were normal ($P = 0.0003$, Fisher’s exact test one-tailed). At E14.5, 83% of male embryos were analyzable but only 17% of female embryos could be analyzed ($P = 0.04$, Fisher’s exact test one-tailed). Furthermore, of the analyzable embryos, 100% of male embryos were normal and 0% of female embryos were normal ($P = 0.1667$, Fisher’s exact test one-tailed). Note that for E14.5 limited numbers of analyzable female embryos resulted in lack of statistical power. At our last time-point of E18.5, 95% of male embryos were analyzable and 33% of female embryos analyzable ($P = 0.0014$, Fisher’s exact test one-tailed). Of the analyzable embryos, 100% of male embryos were normal and 0% of female embryos were normal ($P = 0.0008$, Fisher’s exact test one-tailed). Based on this data we conclude that by E9.5 developmental abnormalities are visible between male Wt and female CAG-FLAG/NR2E1 embryos, and that by E14.5 death and embryo resorption has started to occur resulting in few analyzable embryos.

By visual examination, E9.5 embryos demonstrated similar numbers of somites but a failure of cerebral lobe fusion (Fig. 4.2B). RT-PCR analysis on a separate set of embryos demonstrated expression of the FLAG/NR2E1 construct at E9.5 (Fig. 4.2C). Thus ubiquitous expression of FLAG/NR2E1 in female embryos interferes with normal embryonic brain development, resulting in delay of cerebral lobe fusion.
Germline chimeras never have live-born carrier female offspring and female \( \text{XHprt}^{\text{CAG-FLAG/NR2E1}}/\text{X} \) mice show failure of cerebral hemisphere fusion at E9.5.

Embryos were collected by timed-pregnancy. For each age in (A), \( n \) numbers in the graphs indicated total number of embryos collected for each group in the first graph, and total number of analyzable embryos for the second graph. (A) Female CAG-FLAG/NR2E1 embryos collected at E9.5, E12.5, E14.5 and E18.5 demonstrate developmental abnormalities as early as E9.5 and reduced number of analyzable (i.e. not completely resorbed; structure still visible) embryos from E14.5 onwards, compared to wild-type male embryos. (B) Whole-mount comparison of an E9.5 wild-type male embryo and a female CAG-FLAG/NR2E1 embryo, showing failure of cerebral lobe fusion (arrow). (C) RT-PCR on E9.5 embryos demonstrated \( \text{Nr2e1} \) expression in all embryos, as well as FLAG/NR2E1 expression in the CAG female embryos.

4.4.3 CAG-FLAG/NR2E1 ESCs exhibit growth defects

As developmental defects may be explained by a general growth delay, we were interested in testing whether a growth defect was present in ESCs containing the CAG-FLAG/NR2E1 allele. \( \text{Nr2e1} \) is not normally expressed in ESCs (unpublished observations), and this aspect of the CAG promoter would be considered ectopic overexpression. In order to generate a population of ESCs containing this allele and to avoid any differences in pre- and post-selection of targeted ESCs, we first generated novel ESC lines using sperm from a
germline chimera fertilized in vitro (IVF) with B6-Albino oocytes. Using a standard protocol we had developed for generating ESCs by natural mating (Yang et al. 2009), we generated 10 ESC lines of wild-type X/Y and 14 ESC lines of CAG-FLAG/NR2E1 X/X by IVF. During the procedure of seeding blastocysts on MEFs for generating the ESC lines, we took note of how long it took for each individual clone to be established as an ESC line. After clones were successfully frozen and DNA genotyped, we retrospectively analyzed the amount of time required for each clone to go through the last two steps of the procedure (Fig 4.3A). For both expansion of ESC clones from 1x24-well to 3x24-wells and the amount of days to reach confluency for freezing we saw a significantly slower growth time for CAG-FLAG/NR2E1 containing ESC clones ($P < 0.000005$, $t$-test, for expansion step; and $P < 0.001$, $t$-test, for freezing step), suggesting that ubiquitous expression of FLAG/NR2E1 has negative growth effects on ESCs, contrary to the known role of NR2E1 in repressing cell-cycle inhibitors $Pten$ and $p21$ (Sun et al. 2007; Zhang et al. 2006).

We also observed ESC defects via a mitochondrial activity assay, alamarBlue® (Invitrogen), with the original targeted CAG-FLAG/NR2E1 cell lines (data not shown). To strengthen our quantitative analysis, we generated new cell lines. These ESC lines would provide additional untargeted controls and confirm the phenotype in a separate retargeting event. Three novel ubiquitous β-gal expressing Gt(ROSA)26Sor (Zambrowicz et al. 1997) $Hprt^{b-m3}$ deletion cell lines, and three targeted CAG-FLAG/NR2E1 cell lines were generated. Herein, we refer to the untargeted cell lines as "Wt" due to their parental nature to that of targeted clones. These cell lines were also tested in the alamarBlue® assay and CAG-FLAG/NR2E1 targeted cell lines showed less mitochondrial activity ($P < 0.05$, ANOVA with Bonferroni post-hoc) than parental "Wt" cell lines as indicated by fluorescence (Fig. 4.3B). This data strengthens our observation that ubiquitous expression of $NR2E1$ leads to slower growth, and/or reduced viability of ESCs.
101

Figure 4.3 ESCs containing CAG-FLAG/NR2E1 grow slower.
(A) New mESC lines were derived via in vitro fertilization of sperm from a germline mEMS3836 CAG-FLAG/NR2E1 chimera and C57BL/6J oocytes. The average number of days it took individual clones to reach confluency for passaging from 1x24-well to 3x24-wells is indicated, as well as the average number of days for the 3x24-wells to be ready for freezing. Both steps were significantly slower for CAG-FLAG/NR2E1 X/X ESCs (n=14) compared to wild-type X/Y ESCs (n=10) (t-test: Step 1x24 -> 3x24, P < 0.000005; Step 3x24 -> freezing, P < 0.001). (B) 129B6- Hprt<sup>b-m<sup>i</sup>/Y, Gt(ROSA)26Sor ESCs (n=3) and 129B6- Hprt<sup>CAG-FLAG/NR2E1</sup>/Y, Gt(ROSA)26Sor ESCs (n=3) were compared in an alamarBlue® assay for mitochondrial activity after 4 days of growth on MEFs. Cells were incubated in alamarBlue for 4 hrs prior to assessing fluorescence (excitation 544 nm, emission 590 nm) using a plate reader. Experiment was performed twice and aggregate data is shown. CAG-FLAG/NR2E1 male ESCs showed less reduction of the growth indicator resazurin (non-fluorescent, blue) to resorufin (fluorescent, red), as demonstrated by reduced fluorescence (ANOVA: P < 0.05). Wild-type male ESCs showed significantly higher fluorescence than wells containing only MEFs (ANOVA: P < 0.01).

4.4.4 In vitro differentiated CAG-FLAG/NR2E1 ESCs show a reduction in Pax6+ cells and are capable of generating neurons

Due to the surprising negative effect observed in ESCs for CAG-FLAG/NR2E1, we hypothesized that ubiquitous expression would also alter the generation of neural stem and progenitor cells. Furthermore, as Nr2e1 plays a strong role in suppressing glial differentiation (e.g. repressing the Gfap, Aqp4 and S100β promoters) – and
enhancing proliferation of neural stem/progenitor cells (Shi et al. 2004), we proposed that a proportional increase in NSCs would be found when ESCs are differentiated in vitro. Two parental and two CAG-FLAG/NR2E1 ESC clones were seeded as previously described (Portales-Casamar et al. 2010; Barberi et al. 2003) and stimulated to differentiate for 6 days to NSC-containing colonies. The Pax6 marker was used to identify putative NSCs (Conti and Cattaneo 2010; Zhang et al. 2010; Gómez-López et al. 2011), and Mcm2 was used as an indicator of quiescence and stem cell activation (Niu et al. 2011), since Nr2e1 may also play a role in NSC quiescence (Li et al. 2012).

Nestin, which is present in neural stem and progenitor cells only, was visible throughout and surrounding each colony (Fig. 4.4A, first column). Fewer Pax6 positive cells were observed for CAG-FLAG/NR2E1 (Fig. 4.4A, second column), and approximately half of the colony expressed Mash1 (Fig. 4.4A, third column), a marker for the transiently amplifying type C cells (Uda et al. 2007). Quantification of Pax6 and Mcm2 double positive cells showed a significant 35.9% reduction (Fig. 4.4B; \( P < 0.03 \), t-test) in CAG-FLAG/NR2E1 colonies (identified using β-galactosidase staining, blue) but with no reduction in Mcm2-positive cells (not shown), suggesting a shift away from stem cells with no change in quiescence or stem cell activation. Furthermore, average colony size was similar (16.7 ± 10.8 cells for Wt and 17.8 ± 13.9 cells for CAG), indicating CAG-FLAG/NR2E1 expression did not increase proliferation rate. Lastly, at day 14 of differentiation we assessed the expression of the immature migrating neuronal marker, βtubIII, and found βtubIII+β-gal+ cells in both groups (Fig. 4.4C).

Combined, this data suggests that ubiquitous overexpression of human NR2E1 shifts the balance away from self-renewal of NSCs without any change in their quiescence. In contrast to our initial hypothesis, the cells retained the ability to differentiate into neurons.
Figure 4.4 *In vitro* differentiation of ESCs show no defect in neurogenesis, but less Pax6+Mcm2+ cells. (A) Differentiated ESCs colonies at day 6 were completely positive for Nestin (panel one, green). Nearly complete labeling of nuclei in β-gal-positive colonies was observed for Mcm2 (panel two, red), with ~40-60% of cells also positive for Pax6 (panel two, green). Mash1 labeled some cells of each colony (panel three, green). (B) A significant reduction of Pax6+MCM2+ cells was seen in CAG-FLAG/NR2E1 ESC-derived colonies at day 6 compared to Wt (60.8 ± 2.4% [Wt] vs 39.0 ± 4.5% [CAG], *P* < 0.05, *t*-test; n=2 for each genotype with > 170 cells counted per n). (C) At day 14 of differentiation, neurons have started to form throughout the culture. Scale bars 50 μm.

4.4.5 LacZ-marked chimeras show contribution of CAG-FLAG/NR2E1 cells to all three neural lineages without any specific regionalization

In order to study the phenomenon of toxic expression of CAG-FLAG/NR2E1 *in vivo*, we generated additional chimeric animals. Chimeras are an excellent tool to assess the ability of a ubiquitous NR2E1 cell to become any cell type in a permissive and physiological wild-type environment, and to study otherwise lethal phenotypes. To label CAG-FLAG/NR2E1 cells we made novel Hprt<sup>b<sub>m3</sub>/Y</sup> mouse ESCs which also contain the ubiquitously expressed β-galactosidase marker (Zambrowicz et al. 1997) and retargeted the construct. We microinjected three untargeted clones to generate control chimeras and microinjected three CAG-FLAG/NR2E1 clones. Overall, chimerism for untargeted controls (n=7) varied from 19.5-57.5% and for CAG-FLAG/NR2E1 chimeras (n=12) from 1-17.5% (*P* < 0.0004, *t*-test). Two chimeras (4-11% chimerism) proved germline competent (as assessed by offspring coat colour, details in Yang *et al.* (2009)), but also failed to yield any female CAG-FLAG/NR2E1 germline offspring (*P* = 1.21 X 10<sup>-13</sup> *χ*-test), as was previously observed. This result demonstrates the
embryonic lethality phenotype in a second, independently targeted cell line and thus excludes the possibility of an unrelated defect in the original CAG-targeted ESCs as responsible for the phenotype.

Next we analyzed the brain and eye, regions of normal \(Nr2e1\) expression, to determine if CAG-FLAG/NR2E1 lacZ-marked cells can contribute to all parts of the eye and retina. Generally, weaker lacZ signal was observed in CAG-FLAG/NR2E1 chimeras, likely due to the lower percent chimerism attained. LacZ-positive cells were observed throughout the retina and \(\beta\)-gal protein was localized primarily to 2-3 perinuclear puncta (Fig. 4.5A), similar to the puncta in \textit{in vivo} data from a previous study using this allele (Liu et al. 2012). Furthermore, lacZ-positive cells were observed in all aspects of the adult brain throughout forebrain, midbrain and hindbrain regions (cortex, hippocampus and cerebellum shown in Fig. 4.5B). To confirm that CAG-FLAG/NR2E1 cells can contribute to all three neural lineages (astrocytes, oligodendrocytes and neurons) \textit{in vivo}, we performed co-staining of cell-type markers. \(\beta\)-galactosidase-marked cells were identified co-labeling with Gfap (astrocyte marker; Fig. 4.5C), Olig2 (oligodendrocyte marker; Fig. 4.5D), and both Dcx (newborn migrating neurons; Fig. 4.5E) and Pcp2-positive cerebellar Purkinje neurons (Fig. 4.5F). This data demonstrates that in the milieu of a mixed population of Wt (blastocyst host) and CAG-FLAG/NR2E1 cells, \(Nr2e1\) expression does not prevent generation of mature differentiated cell types.
Figure 4.5 CAG-FLAG/NR2E1 chimeras made with lacZ-marked targeted ESCs are located non-specifically in eye and brain.

(A) Adult chimeric eyes were stained for β-galactosidase activity using the X-gal substrate (blue product) and counter-stained with neutral red. Although fainter than the untargeted control chimera, staining was observed in all retinal layers for CAG-FLAG/NR2E1 chimeras. (B) Three representative areas (cortex, dentate gyrus, and cerebellum) of the adult brain is shown, stained with X-gal and counterstained with neutral red. There were no specific areas of exclusion or overrepresentation. (C) Co-labeling of β-gal (green) and Gfap (red) in the adult brain. (D) Co-labeling of β-gal (green) and Olig2 (red) in the adult brain. (E) Some newborn neurons (Dcx-positive, red) are positive for β-gal (green) in the hippocampus. (F) Some Purkinje cells (Pcp2-positive, red) are positive for β-gal (green). Scale bars 200 µm.
4.4.6 The 129-XistKO mouse strain rescues the X-linked CAG-FLAG/NR2E1 embryonic lethality

Due to the inherent variability that X-inactivation presents in this scenario of an X-linked lethal phenotype, we devised a strategy for generating male CAG-FLAG/NR2E1 animals. An intermediate female rescue of the X-linked lethality can be performed by employing the use of an Xist knock-out. The Xist RNA acts in cis and coats the inactive X-Chromosome to prevent gene expression. We undertook this experiment by breeding a germline CAG chimera to the 129-XistKO strain, which is deleted for the Xist transcript on one X-Chromosome and results in complete X-inactivation of the opposite chromosome (Brockdorff et al. 1991; Borsani et al. 1991; Csankovszki et al. 1999). This breeding strategy generated 7 XHprtCAG-FLAG/NR2E1Xist+ / XHprtXistKO viable female animals with no overt phenotype, further demonstrating that the embryonic lethality associated with the CAG-FLAG/NR2E1 allele is localized to the X-Chromosome. Females were also fertile and generated XHprtXistKO offspring (males and females), but no live-born offspring carrying the XHprtCAG-FLAG/NR2E1Xist+ chromosome. Since genetic recombination could lead to XHprtCAG-FLAG/NR2E1XistKO and XHprtXist+ chromosomes, we analyzed for both alleles in the offspring of XHprtCAG-FLAG/NR2E1Xist+ / XHprtXistKO in a test-cross and observed one recombination event (a male XHprtXist+-carrier) in 19 live-born mice. This data reconfirms the localization of the CAG-FLAG/NR2E1 embryonic phenotype to the modified X-Chromosome and the use of the XistKO allele in rescuing X-linked lethal phenotypes in mice.

4.4.7 Male CAG-FLAG/NR2E1 embryos harvested at E9.5 are severely delayed and staged at ~E8

To evaluate the effects of ubiquitous expression of human NR2E1 on embryonic development we harvested germline male embryos from female timed-pregnant XHprtCAG-FLAG/NR2E1Xist+ / XHprtXistKO mice for histological examination. As our previous data suggested that embryonic abnormalities are clearly visible by age E9.5 in female animals, we chose this age as a starting point for our study of male animals.

All CAG-FLAG/NR2E1 X/Y animals were found to be developmentally delayed and staged at ~E8 (Fig. 4.6A). Embryos had not completed turning. Some disorganization of both brain (Fig. 4.6B) and heart tissue (Fig. 4.6C) was visible. The neural tube was spayed open and not closed as found in the E9.5 Wt embryo (Fig. 4.6C, white arrow), resembling similarities to Wnt7a-overexpressing embryos (Horn et al. 2007), a Wnt/β-catenin
signalling molecule through which Nr2e1 has been shown to act. Nr2e1 and Wnt7a has been shown to be co-expressed in NSCs (Qu et al. 2009).

We chose a set of markers to analyze the development program of the heart and brain in Wt male and CAG-FLAG/NR2E1 male embryos. For the brain, we assessed expression of Pax6, Dcx, β-catenin, Wnt7a, and Sox2. For the heart we assessed Isl1/2, Sema5, mGluR1, Pax6, Sox9, B-catenin, PDGFRα and Notch1. No differences were observed for Isl1/2, Sema5, mGluR1, PDGFRα, or Notch1 staining (data not shown).

Surprisingly, in brain sections Pax6 and Dcx localization was similar in the neural epithelium (Fig. 4.6D). β-catenin-positive (antibody against phosphorylated β-catenin, the active form) cells were observed in the neural epithelium of CAG-FLAG/NR2E1 embryos. Wnt7a localization also looked similar (Fig. 4.6E). Furthermore, Sox2 and Sox9 were expressed in both genotypes (Fig. 4.6F,G). Ectopic Pax6 expression was observed in CAG-FLAG/NR2E1 embryos in the rostral neural epithelium, but more caudally demonstrated expression patterns with similar boundaries in the neural tube (Fig. 4.6G, white arrows). Both genotypes showed autofluorescence of primitive gut tissue (Fig. 4.6G, yellow arrows), and some non-specific amnion fluorescence (green; red arrows) was present for the CAG-FLAG/NR2E1 (verified to be non-specific across multiple different IF experiments). To our surprise, this data showed that despite the developmental delay, neural markers are still present in the CAG-FLAG/NR2E1 embryos, and that the embryos are capable of producing neurons in vivo also, correlating with our previous observations on in vitro differentiated ESCs.
Figure 4.6 Development is delayed and altered in E9.5 XHprt<sup>CAG-FLAG/NR2E1</sup>/Y embryos.
(A) Whole mount images of wild-type male and CAG-FLAG/NR2E1 wild-type male embryos collected at E9.5. The CAG-FLAG/NR2E1 embryos show significant developmental delay and were staged to ~E8. (B) Hematoxylin & eosin staining of brain of E9.5 wild-type and CAG-FLAG/NR2E1 embryos. (C) Hematoxylin & eosin staining of heart tissue of E9.5 wild-type and CAG-FLAG/NR2E1 embryos. (D) Ectopic Pax6 (green) and Doublecortin (red) expression was observed near the neural epithelium in CAG-FLAG/NR2E1 embryos. (E) β-catenin (green) labeling was observed throughout the embryos with stronger localization along the basal edge of the neural epithelium. Wnt7a (red) expression is seen throughout, with stronger localization along the apical edge of the neural epithelium. (F) Sox2 (red) expression is strong in the neural epithelium of both genotypes. (G) Comparison of
more caudal sections containing the neural tube and developing heart. The neural epithelium has Pax6 (green) and low Sox9 (red) expression in embryos (white arrows). Non-specific signal in the green channel was observed in the amnion (red arrows). Autofluorescence is observed in the gut for both embryos (yellow arrows). Insets for D-F show a magnification of the neural tube. Scale bars for B, C are 100 µm and for D-G 200 µm.

### 4.4.8 Ple53-FLAG/NR2E1 is expressed throughout the brain and does not exert any obvious histological abnormalities in any of the tissues assessed

To evaluate the effects of neuronal expression of NR2E1 in vivo, we generated the Ple53-FLAG/NR2E1 strain which is expressed in Doublecortin-positive neurons (type A neuroblasts). This strategy was based on a published human DCX promoter (Karl et al. 2005; Couillard-Despres et al. 2005; 2006), which we have validated as a knock-in at Hprt (Portales-Casamar et al. 2010). As the CAG-FLAG/NR2E1 embryonic lethality prevented us from assessing amelioration of the phenotypes observed in Nr2e1\(^{frc/frc}\) mice, we were also interested in the potential of rescuing Nr2e1\(^{frc/frc}\) phenotypes using this new strain, should there be no overt abnormalities in the DCX mice. Previously, we have shown amelioration of two characteristic phenotypes in Nr2e1\(^{frc/frc}\) mice, namely severely reduced brain weight and a reduction in retinal blood vessels, with a human PAC random insertion transgenic mouse (Abrahams 2005) and with a single-copy targeted BAC knock-in at Hprt (Schmouth et al. 2012a). Thus, in order to study the effects of NR2E1 in a specific cell population and its potential to rescue the Nr2e1\(^{frc/frc}\) phenotype, we generated the Ple53-FLAG/NR2E1 mice.

Ple53-FLAG/NR2E1 mice were both viable and fertile. We first examined the expression pattern of the Ple53-FLAG/NR2E1 allele by RT-PCR in both adult mouse tissue and in the embryo at E12.5 (Fig. 4.7A). Strong expression for this allele was observed in all brain regions tested (olfactory bulb, cortex and striatum, hippocampus, midbrain, cerebellum and hindbrain), as predicted by the previous reporter mice (Portales-Casamar et al. 2010). Furthermore, expression was also detected in the eye. Lastly, as predicted from our Ple53-lacZ strain (de Leeuw CN and Simpson EM, in preparation), both the head and body of an E12.5 embryo were positive.

Next we assessed key tissues for any obvious histological abnormalities by hematoxylin and eosin staining (Fig. 4.7B-I). Brain, heart, lung, liver, kidney, pancreas, spleen and thymus tissues were harvested and
examined and no abnormalities or differences between Wt and Ple53-FLAG/NR2E1 mice were found (n=2 for each genotype).

Surprisingly and in stark contrast to CAG, this data suggests there is no overt phenotype associated with neuronal-based ectopic expression of NR2E1.
Figure 4.7 Ple53-FLAG/NR2E1 mice express in embryos and adult brain with no obvious histologically abnormal tissues.

(A) RT-PCR was performed on adult and E12.5 Ple53-FLAG/NR2E1 and wild-type mice. For adults, expression was tested in ear, eye, olfactory bulb, cortex and striatum, hippocampus, midbrain, cerebellum, hindbrain, lung, heart, liver, and spinal cord. E12.5 embryos were divided into head and body regions. Brain regions showed the strongest signal. Both the head and body were positive for the embryo. (B-I) Several tissues were evaluated for histological abnormalities by H&E staining. (B) brain, (C) heart, (D) lung, (E) liver, (F) kidney, (G) pancreas, (H) spleen, and (I) thymus. Scale bars 400 μm for B, F and 100 μm for C, D, E, G, H, I.
4.4.9 Ple53-FLAG/NR2E1 mice fail to rescue the salient Nr2e1\textsuperscript{frc/frc} phenotypes

In order to test whether ectopic neural expression of human NR2E1 could suffice in rescuing the Nr2e1\textsuperscript{frc/frc} phenotypes, we first crossed the 129-Ple53-FLAG/NR2E1 to 129-Nr2e1\textsuperscript{frc/+} animals to yield 129-Ple53-FLAG/NR2E1, Nr2e1\textsuperscript{frc/+} females which were then crossed to B6-Nr2e1\textsuperscript{frc/+} males to yield 129B6F1-Ple53-FLAG/NR2E1, Nr2e1\textsuperscript{frc/frc} offspring and control littermates.

Our set of 129B6F1-Ple53-FLAG/NR2E1, Nr2e1\textsuperscript{frc/frc} mice were generated and either harvested at or monitored till ~7-10 months. Curiously, we observed sudden unexplained adult deaths which were unexpectedly prominent on the Nr2e1\textsuperscript{frc/+} background (53% of genotype, $P < 0.05$ for a genotype effect in survival analysis; Fig. 4.8A). This data suggests a genetic interaction between the Nr2e1 locus and the restored HPRT locus carrying the Ple53-FLAG/NR2E1 allele.

The two most salient phenotypes observed in Nr2e1\textsuperscript{frc/frc} mice are reduced brain weight and severe aggression (Young et al. 2002; Abrahams 2005), and blindness as assessed by electroretinogram, which is further correlated with reduced and disorganized retinal blood vessels (Schmouth et al. 2012a).

Brain weight of adult animals was measured and no difference between Hprt\textsuperscript{+} and Hprt\textsuperscript{Ple53-FLAG/NR2E1} mice was observed (Fig. 4.8B), regardless of whether the genotype was Nr2e1\textsuperscript{+/+} or Nr2e1\textsuperscript{frc/frc}. As expected, Hprt\textsuperscript{+}, Nr2e1\textsuperscript{+/+} brains were significantly larger than their Hprt\textsuperscript{+}, Nr2e1\textsuperscript{frc/frc} counterparts (Wt; Wt vs Wt; Frc, $P < 0.001$), as were Hprt\textsuperscript{Ple53-FLAG/NR2E1}, Nr2e1\textsuperscript{+/+} compared to Hprt\textsuperscript{Ple53-FLAG/NR2E1}, Nr2e1\textsuperscript{frc/frc} brain weight (Hemi; Wt vs Hemi; Frc, $P < 0.001$). This data demonstrates that the Ple53-FLAG/NR2E1 allele is unable to rescue the hypoplastic forebrain phenotype observed in Nr2e1\textsuperscript{frc/frc} mice.

Visual examination of Nr2e1\textsuperscript{frc/frc} retinas confirmed the previously described fierce defects regardless of the Hprt allele (Fig. 4.8C). There was a mottling of the retina, disorganization of the radial pattern of blood vessels, and a clear reduction in the number of retinal blood vessels. Quantitatively, there was no difference between the Hprt\textsuperscript{+} and Hprt\textsuperscript{Ple53-FLAG/NR2E1} phenotypes for any of the Nr2e1-genotypes (Wild-type, heterozygous, or homozygous fierce; Fig. 4.8D). Both Nr2e1\textsuperscript{+/-} and Nr2e1\textsuperscript{+/-} animals contained significantly more retinal blood vessels than Nr2e1\textsuperscript{frc/frc} animals ($P < 0.001$). Thus we conclude that the Ple53-FLAG/NR2E1 allele is unable to rescue the retinal funduscopic phenotypes present in the Nr2e1\textsuperscript{frc/frc} background.
Figure 4.8 Ple53-FLAG/NR2E1 fails to rescue the Nr2e1\textsuperscript{frc/frc} salient defects of reduced brain weight and reduced retinal blood vessels.

The Ple53-FLAG/NR2E1 allele was bred onto the Nr2e1\textsuperscript{frc/frc} background, and brain weight measurements and retinal funduscopy performed. (A) Survival data of a cohort of 129B6F1-Ple53-FLAG/NR2E1, Nr2e1 mice demonstrate increased mortality rates by ~9 months in mice containing the Ple53-FLAG/NR2E1 allele on the Nr2e1\textsuperscript{frc/+} background. (B) Brain weights of both Nr2e1\textsuperscript{frc/frc} and Ple53-FLAG/NR2E1; Nr2e1\textsuperscript{frc/frc} animals was reduced compared to Nr2e1\textsuperscript{frc/+} animals (\(P < 0.001\)). (C) Pictures of retinal funduscopy of Nr2e1\textsuperscript{frc/frc} and Ple53-FLAG/NR2E1; Nr2e1\textsuperscript{frc/frc} mice clearly show a less retinal blood vessels, lack of radial symmetry, and mottling of the retinas (white arrow). (D) Severe reduction in number of blood vessels in the retinas of Nr2e1\textsuperscript{frc/frc} mice, with no rescue observed in Ple53-FLAG/NR2E1; Nr2e1\textsuperscript{frc/frc} (\(P < 0.001\) between all Nr2e1\textsuperscript{frc/+} and Nr2e1\textsuperscript{frc/frc} containing genotypes, and both Nr2e1\textsuperscript{frc/frc} groups).
4.5 Discussion

Here we investigated the role of human NR2E1 in the context of ubiquitous CAG promoter-driven and neuronal ectopic DCX promoter-driven expression in the mouse. We showed that ubiquitous expression of NR2E1 leads to embryonic death, which cannot be rescued by random X-inactivation but can instead be rescued with the Xist 

non-random X-inactivation allele, confirming our hypothesis. Surprisingly, we showed that there is a growth deficiency phenotype in ESCs which contain the ubiquitous allele. Differentiation of ESCs to NSCs demonstrated a reduction in Pax6-positive putative NSCs. Analysis of chimeric mice showed that CAG-FLAG/NR2E1 cells can generate all three neural lineages. Next we evaluated a narrower ectopic expressor of human NR2E1, driven by a Doublecortin promoter fragment (PleS3, de Leeuw CN and Simpson EM, in preparation). No overt histological phenotype was observed despite widespread expression in both the brain, eye and non-CNS tissues. A breeding cross to the Nr2e1 

frc/frc background demonstrated failure to rescue the salient brain weight and forebrain hypoplasia phenotype of the fierce mouse, as well as lack of rescue in the reduction of retinal blood vessels, mottling, and radial asymmetry.

Although the CAG promoter would also result in overexpression in endogenous Nr2e1-expressing cells, at least three viable models of NR2E1/Nr2e1 overexpression suggests this is not the primary driver of embryonic death (Wong 2009; Liu et al. 2010); rather, we propose that the ubiquitous expression of NR2E1 leading to embryonic death suggests that the NR2E1 protein can function outside of the brain. The Lim1- mutant which lacks a brain, although rarely, can still survive to term in utero (Shawlot and Behringer 1995), suggesting that a primary brain-only defect should be insufficient in of itself for the CAG-FLAG/NR2E1 embryonic phenotype. We hypothesize that either (1) the necessary co-factors for NR2E1-function are also present in other tissues, or (2) that there is an altered/reduced requirement of co-factors in regions outside of endogenous expression that results in partial or novel functions in other tissues.

The poor growth and viability of mouse ESCs expressing human NR2E1 suggests a negative effect on ESC global functioning. Recently Heng et al. (2010) sought a replacement for Oct4 in the reprogramming of fibroblasts to induced pluripotent stem cells (iPSCs). They identified Nr5a2 as a viable factor, but also showed a severe reduction in iPSC formation when they tried to use Nr2e1 in combination with the four known
reprogramming factors (Oct4, Klf4, Sox2 and c-Myc). Although this analysis was not specifically done with respect to Nr2e1, their data suggests that Nr2e1 expression may be refractory, and in fact antagonistic, to ESC generation and/or maintenance. Other work by Thier et al. (2012) demonstrated that only a limited time-course of Oct-3/4 expression is required initially in order to reprogram fibroblasts to induced neural stem cells (iNSCs). However, it has also been demonstrated that Nr2e1 induces Oct-3/4 expression under certain conditions (Chavali et al. 2011). Lastly, it was reported that precise levels of Oct-3/4 expression may be required in ESC maintenance, since overexpression leads to endo- and mesoderm differentiation of ESCs (Niwa et al. 2000). Together this literature along with our data suggests that Nr2e1 exerts negative consequences on ESC viability. This may be due to reduced pluripotency as the Oct4 literature suggests.

Contrary to our expectations, CAG-FLAG/NR2E1 cells could differentiate to all three neural lineages in chimeras. Although not assessed in vitro, we did not anticipate co-labeling of B-gal in astrocytes due to the known repression of the Gfap and S100β promoters (Shi et al. 2004). This data, in combination with the in vitro generation of neurons at day 14, demonstrates that CAG-FLAG/NR2E1 cells possess the necessary developmental programming to generate all three neural lineages, but may require exogenous signalling to do so as implicated by the chimera data. Although the MS-5 cells in our co-culture based neural induction assay are of stromal origin, signalling by these cells may also have provided the necessary exogenous signalling to allow neuronal differentiation of CAG-FLAG/NR2E1 cells. Thus, an alternative system of neural induction in ESCs without co-culture (Gaspard et al. 2008) could provide novel clues to NR2E1 function. Interesting, in vitro work demonstrated negative effects on proliferation of Wt NSCs by co-culture with Nr2e1 siRNA-transduced NSCs (Qu and Shi 2009). This data further supports the hypothesis that autocrine or paracrine signalling may regulate Nr2e1 functions.

Interestingly, we showed abnormal expression of Pax6 for the CAG-FLAG/NR2E1 allele. We observed a reduced number of Pax6-positive cells in neural stem cell colonies generated in vitro, with minor disturbances in vivo. This data further supports the genetic interaction between Nr2e1 and Pax6 (Stenman 2003; Hollemann et al. 1998). Furthermore, it emphasizes that a delicate balance of Nr2e1 expression levels is critically important in normal development of the brain and eye, key tissues of Pax6 function also. This conclusion is further supported
by our data showing reduced Pax6 expression, and eye abnormalities, in an overexpressor of mouse Nr2e1 (Wong 2009).

Even though we found developmental delay of CAG-FLAG/NR2E1 embryos and did not observe any pups surviving to term, we were still surprised to see that many neural markers were present in ~E8 staged embryos. This suggests that despite ESC growth defects and reduced Pax6-positive cells in vitro, the neural developmental defects may be secondary and that abnormal development of another organ is the primary cause of embryonic death.

We were surprised by the lack of obvious phenotypes in the Ple53-FLAG/NR2E1 mice especially given the CAG-FLAG/NR2E1 results. The Ple53 promoter has been demonstrated to be expressed early in the developing nervous system (Portales-Casamar et al. 2010), thus we anticipated a greater effect. This data suggests that, our genetic manipulation resulting in ectopic neuronal NR2E1 expression is still strictly regulated by other cellular processes, a conclusion further supported by the chimera data. Although one could hypothesize the interference of the FLAG tag with NR2E1 function, this seems unlikely as many other nuclear receptors have previously been N-terminally FLAG-tagged without loss of function (Ding 2006; Hosseinpour 2006; Wang and Fondell 2001; Schodin et al. 1995). Shimozaki and colleagues (2012) also demonstrated protein complex formation of FLAG-tagged Nr2e1 with Sox2 in an EMSA assay. Similarly, HA-tagged Nr2e1 has also been used in other studies (Sun et al. 2007). Additionally, localization of this construct in HEK293 is restricted to the nucleus, as predicted, and generates distinct subnuclear foci which are associated with nuclear receptor activation. In addition, we have clearly elicited a phenotype with the CAG-FLAG/NR2E1 construct and the only difference between that and the DCX strain is the promoter fragment. Lastly, the cDNA portion of the Ple53-FLAG/NR2E1 genomic integration was sequenced and was found to be intact with no mutations.

The failure to rescue any of the salient Nr2e1<sup>frc/frc</sup> phenotypes lead us to hypothesize that type A neuroblasts are refractory to reprogramming to NSCs using NR2E1 alone and that the timing of expression of FLAG/NR2E1 in Ple53-positive cells is too late for such a reversion of cell fate. Recently, Tian and colleagues demonstrated that as part of a 5-factor system (Nr2e1, Sox2, Bmi1, c-Myc and Brn2), Nr2e1 can be used to reprogram fibroblasts to iNSCs (Tian et al. 2012; 2013). However, a combination of only 3-factors (Nr2e1, Sox2 and Bm2) was unable to do so. Thus, perhaps our failure to rescue the Nr2e1<sup>frc/frc</sup> phenotype could be due to requiring
additional genetic manipulations encoding these factors. Based on in situ hybridization data from the Allen Brain Atlas and the experiments of Tian et al. we hypothesize that candidate factor(s) may be either or both c-Myc and Bmi1.

Data by Liu et al. (2010) demonstrated that the p53-/- background greatly enhances the phenotype of a two-copy overexpressor of mouse Nr2e1. This leads us to hypothesize that the p53-/- background may be necessary to elicit a strong phenotype in the Ple53-FLAG/NR2E1 mice. Furthermore, reduced penetrance of histological abnormalities could result in the lack of an overt phenotype in our observations. Lastly, the animals may need to be aged beyond 10 months to uncover an overt phenotype, similar to the presentation of tumours in the overexpressor study by Liu et al. (2010).

In conclusion, we demonstrated that ubiquitous NR2E1 expression results in growth defects in ESC, contrary to our initial hypothesis. We have also demonstrated a Pax6 interaction with NR2E1 at the cellular level, in neural stem/progenitor cell colonies derived from ubiquitously expressed NR2E1. We also show that type A neuroblast-based expression, as per a DCX promoter fragment, of NR2E1 is insufficient to rescue Nr2e1frc/frc phenotypes. Importantly, for future genetic manipulation to modulate the function of NSCs, our work demonstrates that expression of NR2E1 in neuronal cells seems unlikely to rescue a primary stem cell defect, but that care should be taken as systemic delivery of NR2E1 may result in detrimental effects, which will also depend on tissue distribution and transduction efficiency.

4.6 Summary

In this chapter we showed that ubiquitous and overexpression of NR2E1 results in detrimental phenotypes, including embryonic lethality, whereas ectopic expression within the nervous system failed to rescue any of the Nr2e1frc/frc phenotypes. This is the first time, to our knowledge, that either of these genetic perturbations have been attempted and studied in mammals.

This work demonstrates that although it would be critical to restore gene function in the type B NSCs of the brain to rescue fierce phenotypes, there is still a possibility of off-target effects. Importantly, because NSCs are part of an actively dividing cell-type, even if an appropriate MiniPromoter had been developed, there may be
problems in delivery of this therapeutic via AAV, as used in Chapter 3. This would necessitate a viral vector that can be used effectively in the long-term in dividing cells.

Lastly, although interesting preliminary changes in Pax6 were observed, further work is necessary to verify these findings using other techniques and models to further explore the interaction between these two important transcription factors.

The study presented here demonstrate that our Pleiades tools can be applied to study gene function in an in vitro and in vivo model.
Chapter 5: Discussion

The work presented in this thesis was a major part of the Pleiades Promoter Project, which has demonstrated the utility of a large-scale genome-wide strategy to identify novel small promoter constructs of therapeutic potential. Their strength and specificity was tested in knock-in mice in CNS tissues at both embryonic and adult time-points. Furthermore, I analyzed the functional outcome of NR2E1, a key brain development gene included in the Pleiades project, when it is expressed ubiquitously using the CAG promoter, and ectopically in neurons using a Pleiades MiniPromoter to drive expression. In this discussion, I will summarize and discuss the major findings from each chapter, and suggest future directions. Both aspects of this thesis – development of promoter resources for the CNS, and functional analyses of the neural stem cell regulator NR2E1 – represent important findings for the future development of gene therapies for the brain and eye.

5.1 Overview of major findings

In Chapter 2, we described a pipeline used to identify regulatory elements from brain genes with regional or cell-type specificity using a computational biology approach, including phylogenetic analysis, and validated by single-copy knock-ins at Hprt, a neutral genomic locus in mice. Most regulatory elements tested recapitulated aspects of their source gene expression pattern. Through successful use of this strategy, several novel MiniPromoters with brain-specific expression patterns were generated. Of particular interest are MiniPromoters which are expressed in glia (e.g., Ple90, Ple185, and Ple240), proliferating cells of the adult brain (Ple131), and others in discrete cell populations (e.g. Ple67 for serotonergic neurons in Raphe Nuclei) or defined brain regions (e.g. Ple140 for the hypothalamus). Our hope is that the Pleiades tools will contribute to future human gene therapies.

Although many gene therapy strategies involve viral transduction in the adult tissue, we recognize that some diseases may require early genetic interventions and consequently pursued developmental analysis of MiniPromoters as well. In particular, the specificity of the elements included in any one MiniP design may greatly vary depending on tissue and development age, since different regulatory elements for the same source gene act individually for appropriate time course or spatial specificity of gene expression (Visel et al. 2007a; 2009a).
To address the above concern, in Chapter 3, we not only generated additional MiniPromoters expressing in brain, but also studied their expression pattern in other tissues, with a selected subset also analyzed during development. Of particular interest were elements that drove expression in the retina. Gene therapy for the eye is translating to the clinic at a much faster pace as less restrictive regulations surrounding approval for human trials are in place. This is due largely to a much greater benefit to harm ratio and the relative simplicity of the retinal system compared to the brain. Not only did we demonstrate that 17 MiniPromoters drive expression in the eye, we also showed that three ganglion cell layer specific promoters drive similar expression patterns when packaged in the common gene therapy vector, AAV2, and delivered to the mouse retina via intravitreal injections. Combined, this data demonstrates the potential of moving our tools from knock-in mice to model systems that are amenable to therapeutic uses.

In summary, the Pleiades Promoter Project, led by Dr. Simpson, set out the lofty goal of developing a large-scale resource for driving gene expression in a cell- or region-specific manner in the brain (primary outcome). To our knowledge, Pleiades remains the only large-scale resource for human-based brain promoters. Although greatly successful in generating novel MiniPromoters with high specificity, further improvements (discussed in 5.2.1) will enable even better promoter designs.

Throughout the thesis, I have also studied a key regulator of brain development and NSCs, the NR2E1 gene. In Chapters 2 and 3, two MiniPromoters for human NR2E1 were developed based on conserved elements, identified as possibly having regulatory function (Ple139 & Ple140). The Ple139 construct drove some expression mirroring endogenous NR2E1 expression patterns, namely throughout the walls of the lateral ventricles (Li et al. 2012; Zhang et al. 2008). In contrast, the Ple140 construct behaved unexpectedly – expression was localized to the hypothalamus in adult brain and in E12.5 embryos, with no expression elsewhere. The hypothalamus is not a known region of Nr2e1 expression or function. However, curiously a study has shown that tanycytes in the rat ependymal layer of the 3rd ventricle may represent hypothalamic neural progenitor cells (Xu et al. 2005). Perhaps then, the regulatory elements in Ple140 may still have a stem/progenitor cell-specific function. Further work would need to verify the identity of the positive cells in Ple140 in both embryos and adults, and the finding of Xu et al. (2005). Regardless, the embryonic work suggested that the elements used in Ple140 only drive expression in the hypothalamus and nowhere else in the brain or the rest of the animal, and thus this MiniP remains a very
valuable tool. In addition to the poor regulatory resolution score (Chapter 2 and (Abrahams et al. 2002)) for NR2E1, this data suggests that design of MiniPromoters for this gene will continue to be challenging.

While it would be of great scientific interest to pursue sub-regional rescue of the Nr2e1<sup>frc/frc</sup> (an Nr2e1-null allele) phenotype in order to localize the exact nature of Nr2e1<sup>frc/frc</sup> sub-phenotypes, efforts will need to continue at improving MiniP design for difficult cases such as this. Therefore, instead of a sub-regional rescue experiment in Chapter 4, I explored the functional outcome of ubiquitous and ectopic NR2E1 expression.

Neither ubiquitous nor ectopic expression has previously been studied in mice, or other higher organisms. Since regulatory mutations are the most commonly found for human NR2E1 (Corso-Diaz et al. 2012; Schmou et al. 2012b; Kumar et al. 2008), we deemed the evaluation of the effects of dysregulated NR2E1 expression as critical in understanding its function and potential for disease. For ubiquitous expression I used the CAG promoter and for neuronal-based ectopic expression, the Ple53 (DCX based) Pleiades MiniPromoter.

Using the CAG promoter I conducted an analysis of the effect of NR2E1 expression in ESCs, and observed reduced growth and viability of ESCs. In addition, via two independent retargeting events in ESCs, I showed that ubiquitous expression of NR2E1 leads to embryonic death at the start of neurogenesis (~E8). I further demonstrated that overexpression in NSCs <em>in vitro</em> leads to a reduction in Pax6-positive cells, potentially altering the balance between stem cell self-renewal and progenitor generation. This is, to my knowledge, the first cellular level analysis of the effects on NR2E1 on Pax6 expression in a simplified culture system.

I also pursued a neuronal-based ectopic expression model of NR2E1. Using the Ple53 promoter, we expressed NR2E1 in a distinct but closely-related cell type (Dcx-positive type A neuroblasts). Based on our current knowledge of the DCX-based promoter we used, the expression is non-overlapping with the endogenous Nr2e1-positive cell population. Current data suggests Nr2e1 expression is limited to type B NSCs, with only limited evidence for expression of Nr2e1 in type C transitly amplifying progenitors also (Obernier et al. 2011; Li et al. 2012). Using this strain, I demonstrated that there was no functional rescue of the severe Nr2e1<sup>frc/frc</sup> phenotype (Abrahams 2005; Young et al. 2002). Brain weight remained reduced to levels that reflect those of Nr2e1<sup>frc/frc</sup> mice, and retinal funduscopy did not show any amelioration of the reduced blood vessels or mottling of Nr2e1<sup>frc/frc</sup> retinas, suggesting no rescue of vision either.
The data from this chapter shows that embryos are more permissive to high levels of ectopic expression of NR2E1 than we previously thought, surviving until E8 and having formed embryonic structures of the neural tube and heart. So too does our data of no overt phenotype in Ple53-FLAG/NR2E1 mice demonstrate tolerance of NR2E1 expression in immature neurons.

Interestingly, Nr2e1 has been shown to dynamically interact in a negative feedback loop with miRNA-9-1 (Denli et al. 2009; Zhao et al. 2009), whereby sequence complementarity sites of miR-9-1 are located in the 3’ UTR of Nr2e1 mRNA and similarly binding sites for Nr2e1 are located in the downstream region of the miR-9-1 locus. Since our FLAG/NR2E1 coding-only cDNA lacks the endogenous 3’ UTR it is not subject to such regulation but could downregulate miR-9. Furthermore, miR-9 has been demonstrated to be required for neuronal differentiation, and have mixed effects on proliferation depending on the location of stem/progenitor cells in the anterior-posterior neuraxis (Bonev et al. 2011). Therefore, one explanation for the failure of cerebral lobe fusion observed in CAG-FLAG/NR2E1 female embryos and the open neural tube in male embryos may be due to the inability of miR-9 to downregulate expression of CAG-FLAG/NR2E1, and the constitutive repression of miR-9 by FLAG/NR2E1. More recently, the let-7b and let-7d miRNAs have also been shown to regulate Nr2e1 mRNA through the 3’ UTR (Zhao et al. 2010; 2013), further complicating this matter.

Overall, this data paints a complex picture of NSC regulation involving multiple factors and whereby appropriate levels are critical in normal brain development. Additional studies on these details will be required in order to gain a better understanding of these complexities.

5.2 Future directions

5.2.1 Promoter design & resources

5.2.1.1 Application of the Pleiades pipeline to other tissues – e.g. retina

The success of the Pleiades Pipeline suggests that a similar strategy could be employed to generate promoters for other complex tissue types. Despite having produced several MiniPromoters for use in the eye, a more systematic approach aimed at regionalized or cell-type specific gene transcripts in the retina specifically, would generate a rich resource with many immediate gene therapy applications.
For the brain we had the extensive resources of the Allen Brain Atlas, BGEM, GenePaint and GENSAT upon which to draw and to use as selection methods for candidate genes. A detailed discussion on candidate gene selection is described in D’Souza et al. (2008) and in Chapter 2, methods. Similarly, there are many extensive gene expression studies in the retina that could be used. I propose the use of publications aimed at cell-type specific gene expression in the retina (Trimarchi et al. 2007; Blackshaw et al. 2004), including developing a genetic signature for retinal cell types (that is, photoreceptors, horizontal cells, bipolar cells, GABAergic and glycinergic amacrine cells, ganglion cells, astrocytes and Müller glia), both in normal and diseased retinas (Siegert et al. 2009; 2012). Combined, these resources would allow a Pleiades-like project to develop MiniPromoters specifically for use in cell-types of the mouse retina.

5.2.1.2 The refinement of previously positive MiniPromoters

The Pleiades promoter design criteria placed an upper limit of 4 kilobases (kb) on MiniPromoter design. This size criterion would be sufficient given most viral vectors; however, since there are numerous benefits to the use of AAV for gene therapy, we propose the need to minimize the size of previously generated positive MiniPs. The standard AAV genome has an optimal capacity of approximately 4.9 kb, with a current upper limit of 5.2 kb (Dong et al. 1996). Efforts toward larger genomes via the use of cross-viral splicing and recombination in two-vector systems have shown some promise (Lai et al. 2005; Xu et al. 2004; Samulski 2000), but can generate complications both in efficiency and quality control for therapeutics. Therefore, considering the average human cDNA is 1,340 bp in size (Lander et al. 2001), this would leave approximately 2,720 bp for the size of promoter constructs (assuming ~840 bp needed for an intron, poly-adenylation signal and LTRs). We have thus begun focusing efforts on reducing the sizes of existing promoter constructs via bioinformatics reanalysis. Three sources of additional information have allowed us to proceed with redesigns: (1) the evaluation of multiple constructs for a given source gene from our data, and comparison of their expression patterns has enabled the identification of overlapping sequence which could explain similarities in expression; (2) the continuing increase of large-scale datasets or genome analyses that allow better predictive value of putative regulatory elements; and (3) new publications that have further analyzed the promoter region, and the regulatory elements of a given source gene (e.g. Dcx and (Piens et al. 2010)).
An example of novel data enabling better prediction of enhancers is the discovery of homotypic clusters of transcription factor binding sites (HCTs) in the human genome (Gotea et al. 2010). In contrast to the co-occurrence of different transcription factors in the same region, HCTs are defined by several binding sites for the same transcription factor. Importantly, HCTs can be used for identification of enhancers both in a tissue and developmental time-point specific fashion. Furthermore, there is significant overlap between HCTs and the Ep300 protein, known to be associated with enhancers (Visel et al. 2009b).

In our dataset we generated 16 sets of MiniPromoters where more than one construct was positive in knock-in mice (see Chapter 3, Fig. 3.8). Using this data we found that for 6 such sets with identical basal promoters present in all positives, all constructs demonstrated a related expression pattern, suggesting the basal promoter element confers the primary specificity of gene expression in both the endogenous locus but also in our set. Furthermore, an additional 6 sets wherein overlapping basal promoter fragments were used, also showed related expression within each set for all the positive designs. Though not as strongly supportive, this second piece of data also suggests that the common basal promoter fragment confers the majority of specificity in adults. Using this information, we have now pursued the development of more compact designs of MiniPromoters for three of these genes: CCKBR, CLDNS, and DCX. We were able to construct versions of the promoters that range from 1,000 – 1,650 bp in size, which is below the 2,720 bp allotted in our calculation. We are currently pursuing functional testing of these designs to verify that we have retained the desired expression pattern of the previous designs (Chapters 2 and 3).

5.2.1.3 Verification of MiniPs in viral constructs

The discovery of MiniPromoter specificity retained for three retinal ganglion cell promoters when moved to the recombinant AAV2 genome was hugely exciting. This data supported our hypothesis that large scale testing of MiniPromoters in knock-in mice has predictive value of MiniPromoter function in other systems. Despite this positive finding, it is important to recognize that many other verification steps are necessary.

First, for the eye, the specificity of other eye MiniPromoter positives (e.g. those in bipolar cells, namely Ple34 and Ple155), remains to be evaluated. Although we expect that these constructs would similarly retain their specificity, it remains an important caveat of the work. Furthermore, testing needs to be performed on brain expression of MiniPromoters as well using viral vectors. Since regulation of expression in the brain is more
complex due to the much greater diversity of subtypes represented compared to the eye, it may prove more
difficult to retain expression specificity in this complex transcriptional environment. As part of the CanEuCre
project (http://www.caneucre.org) (Smedley et al. 2011; Bucan et al. 2012; Murray et al. 2012; Bradley et al. 2012),
we are now pursuing both these questions using intravenous delivery of rAAV9 viral vectors into ROSA26
reporter mice at P0 (Post-natal day 0) (Foust et al. 2009) where MiniPromoters drive a Cre recombinase reporter.
The Cre recombinase reporter in turn recombines the Gt(ROSA)26Sor tm1Sor locus (Soriano 1999), resulting in β-
galactosidase expression in cells that have undergone recombination.

In addition to the above experiments, it is essential to conduct a comparison of expression of MiniPromoters
delivered via AAV versus other commonly used viral vectors, such as Lentivirus. This data would be important to
ensure the universal applicability of our resources.

5.2.2 Studies on NR2E1 function

5.2.2.1 Quantitative assessment of embryonic stem cell pluripotency genes in CAG-FLAG/NR2E1 ESCs

Overexpression of FLAG/NR2E1 in mouse ESCs proved detrimental to their growth. A potential
mechanism for this is dysregulated Oct4 levels, as studies suggest that Nr2e1 can upregulate Oct4 under hypoxic
conditions in NSCs (Chavali et al. 2011). Oct4 levels are carefully regulated in ESCs (Muñoz Descalzo et al. 2012;
Shi and Jin 2010), and as little as a 2-fold increase may push the pluripotent mESCs towards endo- and
mesoderm fate (Niwa et al. 2000). Thus, assessment of endo- and mesoderm fate markers would also be
informative. Future work on identifying the exact molecular underpinnings of the ESC defects would involve
quantitative assessment of gene expression of known ESC pluripotency genes such as Oct4, Sox2 and Nanog.

5.2.2.2 Further characterization of in vitro neural differentiation

In this thesis I presented a quantification of Pax6-positive cells in in vitro derived neural stem and
progenitor cell colonies, showing a reduction of Pax6-positive cells upon CAG promoter expression of human
NR2E1 (Fig. 4.4). In order to determine the fate of the Pax6-negative cells, quantitative analysis of Mash1-positive
cell numbers and early neural differentiation markers would be valuable. Wnt7a overexpressing embryos show a
delay in β-tubulin III expression (Horn et al. 2007), and Nr2e1 increases Wnt7a levels in SVZ stem cells (Qu et al.
2009). Thus, we hypothesize that Wnt7a expression may be increased in CAG-FLAG/NR2E1 NSC colonies and that
there may be precocious neuronal differentiation, should elevated Wnt7a levels be found. Further work on
testing these predictions would help in understanding the functional outcome of reduced Pax6-positive cells in
CAG-FLAG/NR2E1 colonies.

5.2.2.3 Analysis of male CAG-FLAG/NR2E1 embryos at E8.5 & detailed expression characterization
differences in embryos

Morphology suggests that CAG-FLAG/NR2E1 embryos harvested at E9.5 better resemble ~E8 wild-type
embryos (Fig. 4.6 and (Kaufman 1995)). Secondary effects due to the large morphological disparities may
influence gene expression. This necessitates an analysis of Wt and CAG-FLAG/NR2E1 differences at an earlier age.
Further work using cryosectioning and histological stains will be required to confirm this finding.

5.2.2.4 Harvesting of embryos at E8.5 for qRT-PCR gene expression comparisons

Similarly, structural abnormalities in CAG-FLAG/NR2E1 embryos make quantitative marker assessment
by immunofluorescence challenging. In order to assess the transcript levels of neurogenesis related factors –
Nr2e1, Pax6 and Wnt7a for stem cell regulation, Mash1 for type C cells, and β-tubulin III and Dcx for early
neuronal cells – generating E8.5 embryos for harvesting and analyzing RNA should allow better detection of
differences between genotypes while maintaining the complexity of in vivo interactions between these factors
and embryonic development.

5.2.2.5 Extensive histological evaluation of chimeras

While no apparent mislocalization of markers of the three neural cell types were identified in CAG-
FLAG/NR2E1 chimeras, it remains unlikely even in low percent chimeras which have wide-spread distribution of
CAG-FLAG/NR2E1, that no phenotype is present. Thus, an extensive marker based analysis on these chimeras
may reveal an otherwise hidden phenotype, which could elucidate important functions of NR2E1. For instance,
expression of Nr2e1 is observed in cortical glutamatergic inhibitory interneurons (Zhang et al. 2008) and the role
that Nr2e1 plays in these cells is currently unknown. Therefore, staining of chimera brains for glutamatergic
markers mGluR1-4 and other neuronal subtypes may prove fruitful. Furthermore, as calretinin-positive GABAergic
and somatostatin-positive populations have previously been shown to be altered in Nr2e1-null mice (Roy et al.
2002; Monaghan et al. 1997), these two markers may also be useful.
5.3 Conclusion

This thesis consists of a large-scale development of promoters for use in the central nervous system and the first ectopic expression analysis of a key regulator of CNS development, the NR2E1 gene. The work presented in this thesis used the Hprt-docking technology to study genes involved in the central nervous system both in adult mice and during development.

The results of Chapter 2 represent novel small promoter resources for driving brain gene expression. It also demonstrated that a combined computational biology and phylogenetic footprinting approach to promoter development resulted in a high success rate of positive constructs. Suggested future work would consist of developing additional promoter constructs for brain regions that has remained more challenging, application of this pipeline to other key tissues, and the reconfiguration of the pipeline to take advantage of new studies on transcriptional regulation.

Chapter 3 highlighted that many of our MiniPromoters express in non-brain CNS tissue, such as the eye and spinal cord. This data demonstrates the unique resource that the Pleiades Promoter Project represents outside of its main founding goals of developing promoters for the adult brain. With great excitement we showed preliminary data suggesting our overall strategy is predictive of the expression pattern of MiniPromoters when delivered via viruses. Our resources uniquely positions Pleiades MiniPromoters as a key player in future gene therapy designs. Further studies will focus not only on the testing of other eye positives in viral vectors but also on the refinement of existing MiniPromoters to reduce their size.

Lastly, Chapter 4 not only studies one of our Pleiades source genes, nuclear receptor 2E1 (NR2E1), in greater detail, but also demonstrates the application of Pleiades resources. We first showed the effects of a ubiquitous promoter driving NR2E1 in the milieu of mouse development. We then proceeded with the DCX-based Ple53 construct as a more refined and biologically-relevant ectopic expression model of human NR2E1. This data showed that NR2E1 has effects outside of its known physiological range of functioning. Future work would elucidate other phenotypes in the DCX-based ectopic expressor and further evaluate NR2E1 effects driven by promoters of other cell types.
Combined, the results in this thesis set the foundation for future studies on promoter development and on NR2E1 function.
References


Farhadi HFH, Lepage PP, Forghani RR, Friedman HCHH, Orfali WW, Jasmin LL, Miller WW, Hudson TJT, Peterson ACA. 2003. A combinatorial network of evolutionarily conserved myelin basic protein regulatory sequences


Kumar RA, Chan KL, Wong AHW, Little KQ, Rajcan-Separovic E, Abrahams BS, Simpson EM. 2004. Unexpected embryonic stem (ES) cell mutations represent a concern in gene targeting: Lessons from *fierce* mice. *Genesis*


Wong B. 2009. Evaluating the effects of variable NR2E1 levels on gene expression, behaviour, and neural and ocular development. *University of British Columbia, PhD Dissertation*.


## Appendices

### Appendix A : Dataset S1

<table>
<thead>
<tr>
<th>MinIP</th>
<th>Gene</th>
<th>Existing in vivo promoter studies</th>
<th>Reporter</th>
<th>Expression relative to source gene</th>
<th>Expression pattern description</th>
<th>Pleiades URL (images)</th>
<th>Mouse strain name</th>
<th>JAX catalog nb</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control</strong></td>
<td>Ple48</td>
<td>DBH</td>
<td>PHID 8182421</td>
<td>lacZ</td>
<td>as published</td>
<td>Ple48 expression is enriched in noradrenergic nuclei as initially reported (construct DBH1.5 in PHID 8182421), however, additional expression is observed in exogenous regions including the cortex. Interestingly, lacZ staining overlaps with the sites of tyrosine hydroxylase expression but no clear co-localization in individual cells was detected in the brain (Fig. S1). The co-localization analysis with TH in the original papers was performed with different constructs (DBH5.8 and DBH8.1) and reported as sparse (PHID 8182421; 1742021).</td>
<td><a href="http://pleiades.org/promoter/4191/index.php">http://pleiades.org/promoter/4191/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm17(Ple48-lacZ)Emsy/J</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>Ple53</td>
<td>DCX</td>
<td>PHID 17004917</td>
<td>EGFP</td>
<td>as published</td>
<td>Ple53 directs expression primarily in young neurons of the olfactory bulb and rostral migratory stream where it co-localizes with endogenous Dcx (Fig. S1). We also observed weak expression in the subventricular zone and the subgranular zone of the dentate gyrus. A similar expression pattern was described for this construct in the adult brain (PHID 17004917).</td>
<td><a href="http://pleiades.org/promoter/4199/index.php">http://pleiades.org/promoter/4199/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm20(Ple53-EGFP-NLS)Emsy/J</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>Ple59</td>
<td>DDC</td>
<td>PHID 11750071</td>
<td>EGFP, lacZ</td>
<td>not detected</td>
<td>No expression could be detected.</td>
<td>-</td>
<td>B6.129P2-Hprt&lt;tm4(Ple59-EGFP-NLS)Emsy/J</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>Ple88</td>
<td>GFAP</td>
<td>PHID 8120611</td>
<td>EGFP, lacZ</td>
<td>as published</td>
<td>Ple88 expresses in astroglial-like cells throughout the brain, spinal cord, and the eye with a subset of cells co-expressing the Gfap endogenous gene (Fig. S1) similarly to published results with this construct (PMID 8120611).</td>
<td><a href="http://pleiades.org/promoter/4173/index.php">http://pleiades.org/promoter/4173/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm44(Ple88-EGFP-NLS)Emsy/J</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>Ple112</td>
<td>NCR7</td>
<td>PHID 11854267</td>
<td>EGFP</td>
<td>as published</td>
<td>Ple112 directs EGFP expression in a discrete population of Hcrt-positive neurons in the lateral hypothalamic area (Fig. S1) as initially described (construct delt-Alu in PMID 11854267).</td>
<td><a href="http://pleiades.org/promoter/4148/index.php">http://pleiades.org/promoter/4148/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm32(Ple112-EGFP-NLS)Emsy/J</td>
</tr>
<tr>
<td><strong>Refined Control</strong></td>
<td>Ple49</td>
<td>DBH</td>
<td>yes</td>
<td>lacZ</td>
<td>related</td>
<td>As for Ple48, Ple49 shows an enriched expression in noradrenergic nuclei with a strong expression in the adrenal gland where it co-expressed with TH (Fig. 3).</td>
<td><a href="http://pleiades.org/promoter/4144/index.php">http://pleiades.org/promoter/4144/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm44(Ple49-lacZ)Emsy/J</td>
</tr>
<tr>
<td><strong>Refined Control</strong></td>
<td>Ple54</td>
<td>DCX</td>
<td>yes</td>
<td>EGFP</td>
<td>related</td>
<td>Both MiniPs drive expression similarly to the control MiniP Ple53 (Fig. 3). Ple54 is the strongest expression compared to the other two but with a broader expression pattern (same expressing cells are found in the cortex in addition to the main OB and RMS expression; Fig. 3). As for Ple53, Ple54 does express the EGFP reporter in neuronal populations that are immunopositive for endogenous Dcx (Fig. 3).</td>
<td><a href="http://pleiades.org/promoter/4136/index.php">http://pleiades.org/promoter/4136/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm13(Ple54-EGFP-NLS)Emsy/J</td>
</tr>
<tr>
<td><strong>Refined Control</strong></td>
<td>Ple55</td>
<td>DLX</td>
<td>yes</td>
<td>EGFP, lacZ</td>
<td>related</td>
<td>The 1.4 kb Ple55 drives the expression of EGFP in a pattern similar to the previously characterized 2.2 kb Ple55 and co-localizes with the endogenous gene (Fig. 3). This result was supported by a recent independent study analyzing similar GFAP promoter constructs (PMID 18240313).</td>
<td><a href="http://pleiades.org/promoter/4126/index.php">http://pleiades.org/promoter/4126/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm18(Ple55-EGFP-NLS)Emsy/J</td>
</tr>
<tr>
<td><strong>Refined Control</strong></td>
<td>Ple60</td>
<td>GFAP</td>
<td>yes</td>
<td>EGFP</td>
<td>related</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MiniP</td>
<td>Gene</td>
<td>Existing in vivo promoter studies</td>
<td>Reporter</td>
<td>Expression relative to source gene</td>
<td>Expression pattern description</td>
<td>Pleiades URL (images)</td>
<td>Mouse strain name</td>
<td>JAX catalog nb</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td>----------------------------------</td>
<td>----------</td>
<td>-----------------------------------</td>
<td>--------------------------------</td>
<td>----------------------</td>
<td>------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Ple111</td>
<td>MC4R</td>
<td>yes</td>
<td>EGFP</td>
<td>related</td>
<td>As for Ple112, EGFP expression in the Ple111 brain is co-localized with Hirt in a discrete population of neurons in the hypothalamus (Fig. 3). A similar construct, slightly shorter, was previously tested but didn't show any expression (construct 0.4 in PMID 11854267). This contrasting observation could be due to the random insertion strategy the authors employed.</td>
<td><a href="http://pleiades.org/promoters/1417/index.php">http://pleiades.org/promoters/1417/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm15(Ple111-EGFP/NLS)Ems&gt;/J</td>
<td>009115</td>
</tr>
<tr>
<td>Ple117</td>
<td>C8ORF46</td>
<td>no</td>
<td>lacZ</td>
<td>related</td>
<td>The following observations are based on three chimeras. A consistent regional staining is observed in the dorsal lateral geniculate and brainstem region, ventral nucleus of the lateral lemniscus (Fig. SSA). One chimera shows expression in the posterior cortex, a region of the posterior brainstem, and the mesencephalic V. This is entirely distinct from the endogenous gene expression pattern (<a href="http://mouse.brain-map.org">http://mouse.brain-map.org</a>).</td>
<td><a href="http://pleiades.org/promoters/1473/index.php">http://pleiades.org/promoters/1473/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm55(Ple117-lacZ)Ems&gt;/J</td>
<td></td>
</tr>
<tr>
<td>Ple24</td>
<td>CCKBR</td>
<td>no</td>
<td>lacZ</td>
<td>related</td>
<td>Staining is enriched in the cingulate gyrus, throughout the deep layers of cortex (strong in layers III, IV and more sporadically in V, VI), in a region of the anterior thalamus, in a discrete area of the amygdala spanning the entire basal lateral amygdala, in hippocampal pyramidal cells and in the red nucleus (Fig. S5I). Some lighter labelling is observed in the subthalamic area zona incerta, and in medial septal moving out lateral to the diagonal band. There is also some cerebellar granular cell label and some other diffuse mids and brainstem labelling. This expression is nicely related to the endogenous gene (<a href="http://mouse.brain-map.org">http://mouse.brain-map.org</a>). Ple24 and 25 present a similar staining with the exception of the additional staining in the dorsal lateral geniculate in Ple25.</td>
<td><a href="http://pleiades.org/promoters/1495/index.php">http://pleiades.org/promoters/1495/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm39(Ple24-lacZ)Ems&gt;/J</td>
<td></td>
</tr>
<tr>
<td>Ple26</td>
<td>CCL27</td>
<td>no</td>
<td>lacZ</td>
<td>overlapping</td>
<td>Cell bodies are strongly labelled throughout the brain except for the olfactory region and thalamus which are only lightly stained (Fig. S5C). In the cerebellum, the granular cells are strongly labelled and the Purkinje and molecular layers are only lightly labelled. This expression pattern is much broader but overlapping that of the endogenous gene, which is highly expressed in the cortex (<a href="http://mouse.brain-map.org">http://mouse.brain-map.org</a>).</td>
<td><a href="http://pleiades.org/promoters/1497/index.php">http://pleiades.org/promoters/1497/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm57(Ple26-lacZ)Ems&gt;/J</td>
<td></td>
</tr>
<tr>
<td>Ple32</td>
<td>CLDN5</td>
<td>no</td>
<td>lacZ</td>
<td>related</td>
<td>For all three MiniPs, a strong staining is observed in or around small capillaries and blood vessels throughout the brain (Fig. S5D). This expression is directly related to that of the endogenous gene CLDN5 (aka BEC1, TMVC), an integral membrane protein and component of tight junctions. In situ hybridization and immunocytochemical analyses revealed Cldn5 in brain capillary endothelial</td>
<td><a href="http://pleiades.org/promoters/1503/index.php">http://pleiades.org/promoters/1503/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm51(Ple32-lacZ)Ems&gt;/J</td>
<td></td>
</tr>
<tr>
<td>MiniP</td>
<td>Gene</td>
<td>Existing in vivo promoter studies</td>
<td>Reporter</td>
<td>Expression relative to source gene</td>
<td>Expression pattern description</td>
<td>Pleiades URL (images)</td>
<td>Mouse strain name</td>
<td>JAX catalog nb</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>----------------------------------</td>
<td>----------</td>
<td>-----------------------------------</td>
<td>--------------------------------</td>
<td>----------------------</td>
<td>------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Novel MiniP</td>
<td>Ple34</td>
<td>CLDN5</td>
<td>no</td>
<td>lacZ related</td>
<td>cells, as well as in a subset of other endothelial and epithelial cells (PMID 9520948). Mice deficient in Cldn5 show size-selective loosening of the blood-brain barrier (PMID 12743111). Thus, we anticipate these MiniPs are driving expression in the cells of the blood brain barrier.</td>
<td><a href="http://pleiades.org/promoter/1505/index.php">http://pleiades.org/promoter/1505/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm40(Ple34-lacZ)Ems&gt;/J</td>
<td></td>
</tr>
<tr>
<td>Novel MiniP</td>
<td>Ple35</td>
<td>CLDN5</td>
<td>no</td>
<td>lacZ related</td>
<td>Mice deficient in Cldn5 show size-selective loosening of the blood-brain barrier (PMID 12743111). Thus, we anticipate these MiniPs are driving expression in the cells of the blood brain barrier.</td>
<td><a href="http://pleiades.org/promoter/1506/index.php">http://pleiades.org/promoter/1506/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm52(Ple35-lacZ)Ems&gt;/J</td>
<td></td>
</tr>
<tr>
<td>Novel MiniP</td>
<td>Ple67</td>
<td>FEV</td>
<td>yes</td>
<td>EGFP related</td>
<td>Expression is specific to the raphe nuclei (dorsal, obscurus/magnus and pallidus) which all show EGFP positive cells (Fig. 4A). This pattern is related to that of the endogenous gene as described in the literature (PMID 15003288; 10575032; 11875656). At the time of the designs, a published study reported expression from three mouse Fev constructs (40 kb, 12 kb and 1.8 kb of upstream sequence respectively; PMID 15758173). The authors showed targeting of the 5-HT neurons but if the 40 kb sequence is specific to these neurons, the shorter constructs showed ectopic expression and more variability depending on the integration site. Ple67 is similar to the 1.8 kb construct although slightly longer and from human rather than mouse. The better result we obtain might be due to our controlled integration site. The Ple67 expression pattern was confirmed recently in a new study using a 2.2 kb human FEV promoter sequence (PMID 19036967).</td>
<td><a href="http://pleiades.org/promoter/1401/index.php">http://pleiades.org/promoter/1401/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm31(Ple67-EGFP/NLS)Ems&gt;/J</td>
<td></td>
</tr>
<tr>
<td>Novel MiniP</td>
<td>Ple103</td>
<td>HAP1</td>
<td>no</td>
<td>EGFP/cre unrelated</td>
<td>There is weak beta-galactosidase activity in a small but consistent population of cells throughout the brain; typically these cells are small and found in juxtaposition to blood vessels (Fig. 4B). Rare large lacZ-positive cells are detected in neuropil of the forebrain and mid brain. EGFP expression, however, was not detected in these adult brains by either native fluorescence or immunocytochemistry. The absence of EGFP expression and the small number of total cells labeled suggests the historical nature of the functional activation of the promoter at an earlier time in development. This expression is unrelated to that anticipated for the endogenous gene (<a href="http://mouse.brain-map.org">http://mouse.brain-map.org</a>).</td>
<td><a href="http://pleiades.org/promoter/1083/index.php">http://pleiades.org/promoter/1083/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm14(Ple103-EGFP/cre/NLS)Ems&gt;/J</td>
<td></td>
</tr>
<tr>
<td>Novel MiniP</td>
<td>Ple119</td>
<td>HTR1A</td>
<td>yes</td>
<td>lacZ related</td>
<td>The staining is relatively sparse but nicely localized in a few regions: ventral thalamic/posterior hypothalamic territories, cortex layer IV, hippocampus area CA1c, and retrosplenial cortex (Fig. 5E). Some brainstem nuclei are also labelled. This pattern of expression matches well with the endogenous gene expression (<a href="http://mouse.brain-map.org">http://mouse.brain-map.org</a>) but perhaps is more discrete.</td>
<td><a href="http://pleiades.org/promoter/1590/index.php">http://pleiades.org/promoter/1590/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm58(Ple119-lacZ)Ems&gt;/J</td>
<td></td>
</tr>
<tr>
<td>Novel MiniP</td>
<td>Ple123</td>
<td>ICMY</td>
<td>no</td>
<td>lacZ overlapping</td>
<td>We observe a strong cell body staining throughout the brain, the most heavily labelled areas being: cortex layers IV and VI, pyriform cortex, islands of Calleja, anterior hypothalamus, parts of the amygdala, hippocampus regio-superior, and nucleus of darkschewitzi (Fig. 5B). Weak or non-stained areas include peri-aqueductal grey, caudate, dentate gyrus CA3a-c, and ventral posterior thalamus. There is fairly light staining in the cerebellum, but discrete staining in the Purkinje cell layer. Overall, this staining does not reflect the specificity of the endogenous staining, but does overlap in the Purkinje cell layer (<a href="http://mouse.brain-map.org">http://mouse.brain-map.org</a>).</td>
<td><a href="http://pleiades.org/promoter/1594/index.php">http://pleiades.org/promoter/1594/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm59(Ple123-lacZ)Ems&gt;/J</td>
<td></td>
</tr>
<tr>
<td>MiniP</td>
<td>Gene</td>
<td>Existing in vivo promoter studies</td>
<td>Reporter</td>
<td>Expression relative to source gene</td>
<td>Expression pattern description</td>
<td>Pleiades URL (images)</td>
<td>Mouse strain name</td>
<td>JAX catalog nb</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td>----------------------------------</td>
<td>----------</td>
<td>----------------------------------</td>
<td>--------------------------------</td>
<td>----------------------</td>
<td>------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Novel MiniP</td>
<td>Ple129</td>
<td>MKI67</td>
<td>no</td>
<td>lacZ</td>
<td>related</td>
<td>Staining is strongest surrounding the ventricles, in the RMS, and dentate gyrus (Fig. S5F). Lightly stained throughout the internal granular cell layer of the cerebellum, but particularly strong in the posterior lobe. This expression is highly related to the endogenous Mki67 gene with the exception of the cerebellar label (<a href="http://mouse.brain-map.org">http://mouse.brain-map.org</a>; PMID 10653597; 16262632; 15564594). Ple129 and Ple131 present a similar staining. Though Ple129 shows additional eXogenous areas of expression including cingulate cortex and hippocampal pyramidal cells.</td>
<td><a href="http://pleiades.org/promoters/1600/index.php">http://pleiades.org/promoters/1600/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm60(Ple129-lacZ)Ems&gt;/J</td>
</tr>
<tr>
<td>Novel MiniP</td>
<td>Ple131</td>
<td>MKI67</td>
<td>no</td>
<td>lacZ</td>
<td>related</td>
<td>The staining is strongly regional. The walls of the lateral ventricles have positive cells (Fig. S5K).Anterior dorsal cortex light, ventral cortex heavy. Up through dorsal midbrain heavy staining is present, then going more ventral and posterior the staining is virtually absent. However, the hippocampus is only sporadically labelled. This staining is overlapping the endogenous, which is present in the ventricle, and matches the developmental range of expression.</td>
<td><a href="http://pleiades.org/promoters/1612/index.php">http://pleiades.org/promoters/1612/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm61(Ple139-lacZ)Ems&gt;/J</td>
</tr>
<tr>
<td>Novel MiniP</td>
<td>Ple139</td>
<td>NR2F2</td>
<td>no</td>
<td>lacZ</td>
<td>unrelated</td>
<td>Strong lacZ-positive staining is found in the periventricular hypothalamus, with weaker staining in the apparent amygdala region (Fig. S5L). Although highly regional, this pattern of expression is unrelated to the endogenous gene, which is expressed in neurogenic regions of the mouse brain (PMID 7720587).</td>
<td><a href="http://pleiades.org/promoters/1612/index.php">http://pleiades.org/promoters/1612/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm62(Ple153-lacZ)Ems&gt;/J</td>
</tr>
<tr>
<td>Novel MiniP</td>
<td>Ple140</td>
<td>NR2F2</td>
<td>no</td>
<td>lacZ</td>
<td>unrelated</td>
<td>The observed staining is very discrete, limited to the anterior thalamic territory (Fig. S5G). In addition, a linear array of lacZ-positive cells is located in the super-mammillary region, best seen in the coronal section. It is not overlapping the endogenous expression pattern (<a href="http://mouse.brain-map.org">http://mouse.brain-map.org</a>). lacZ-positive cells are enriched in the anterior thalamus (Fig. S5M), with further expression observed in the ventral-lateral hippocampus, and the VTA region. The latter (VTA region) location reflects the expression of the endogenous Pitx3 gene (PMID 9371841; 16190884).</td>
<td><a href="http://pleiades.org/promoters/1612/index.php">http://pleiades.org/promoters/1612/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm61(Ple153-lacZ)Ems&gt;/J</td>
</tr>
<tr>
<td>Novel MiniP</td>
<td>Ple151</td>
<td>Dlg1</td>
<td>no</td>
<td>EGFP</td>
<td>related</td>
<td>Positive EGFP immunoreactivity was detected across numerous brain regions including the brainstem, cortex, and thalamus, as well as, to a lesser extent, in the olfactory bulb, striatum and cerebellum (Fig. 4C). EGFP-positive cells show darker staining in the nucleus with a lighter staining halo of processes, consistent with published studies of oligodendrocyte morphology (PMID 18824157). Endogenous Olig2 expression is detected in myelinating and non-myelinating oligodendrocytes (PMID 11526205). Accordingly, EGFP-positive cells overlap with the oligodendrocyte myelin-related protein marker RIP and do not co-stain with the neuronal marker NeuN and the astrocyte marker GFAP (Fig. 4C). These results suggest that Ple151 reporter expression overlaps with Dlg1 expression in myelin-forming oligodendrocytes.</td>
<td><a href="http://pleiades.org/promoters/1612/index.php">http://pleiades.org/promoters/1612/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm61(Ple151-EGFP-NLS)Ems&gt;/J</td>
</tr>
<tr>
<td>Novel MiniP</td>
<td>Ple153</td>
<td>Grik4</td>
<td>yes</td>
<td>lacZ</td>
<td>unrelated</td>
<td>The observed staining is very discrete, limited to the anterior thalamic territory (Fig. S5G). In addition, a linear array of lacZ-positive cells is located in the super-mammillary region, best seen in the coronal section. It is not overlapping the endogenous expression pattern (<a href="http://mouse.brain-map.org">http://mouse.brain-map.org</a>). lacZ-positive cells are enriched in the anterior thalamus (Fig. S5M), with further expression observed in the ventral-lateral hippocampus, and the VTA region. The latter (VTA region) location reflects the expression of the endogenous Pitx3 gene (PMID 9371841; 16190884).</td>
<td><a href="http://pleiades.org/promoters/1612/index.php">http://pleiades.org/promoters/1612/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm61(Ple153-lacZ)Ems&gt;/J</td>
</tr>
<tr>
<td>Novel MiniP</td>
<td>Ple160</td>
<td>Pitx3</td>
<td>yes*</td>
<td>lacZ</td>
<td>related</td>
<td>Staining is strong surrounding the ventricles, in the RMS, and dentate gyrus (Fig. S5F). Lightly stained throughout the internal granular cell layer of the cerebellum, but particularly strong in the posterior lobe. This expression is highly related to the endogenous Mki67 gene with the exception of the cerebellar label (<a href="http://mouse.brain-map.org">http://mouse.brain-map.org</a>; PMID 10653597; 16262632; 15564594). Ple129 and Ple131 present a similar staining. Though Ple129 shows additional eXogenous areas of expression including cingulate cortex and hippocampal pyramidal cells.</td>
<td><a href="http://pleiades.org/promoters/1612/index.php">http://pleiades.org/promoters/1612/index.php</a></td>
<td>B6.129P2-Hprt&lt;tm41(Ple160-lacZ)Ems&gt;/J</td>
</tr>
</tbody>
</table>
**MiniP** | **Gene** | **Expression in vivo promoter studies** | **Reporter** | **Expression relative to source gene** | **Expression pattern description** | **Pleiades URL (images)** | **Mouse strain name** | **JAX catalog nb**  
--- | --- | --- | --- | --- | --- | --- | --- | ---  

* Promoter analyses of the PITX3 gene were published after production of transgenic MiniP animals and did not influence the work.
<table>
<thead>
<tr>
<th>Ple number</th>
<th>Gene RRs</th>
<th>Strain Category</th>
<th>Existing promoter studies</th>
<th>Reporter</th>
<th>Brain expression relative to source gene</th>
<th>Brain expression pattern description</th>
<th>Pleiades URL (images)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ple number</td>
<td>Gene RRs</td>
<td>Strain Category</td>
<td>Existing promoter studies</td>
<td>Reporter</td>
<td>Brain expression relative to source gene</td>
<td>Brain expression pattern description</td>
<td>Pleiades URL (images)</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>-----------------</td>
<td>---------------------------</td>
<td>----------</td>
<td>------------------------------------------</td>
<td>-----------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Ple number</td>
<td>Gene RRs</td>
<td>Strain Category</td>
<td>Existing promoter studies</td>
<td>Reporter</td>
<td>Brain expression relative to source gene</td>
<td>Brain expression pattern description</td>
<td>Pleiades URL (images)</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>-----------------</td>
<td>---------------------------</td>
<td>----------</td>
<td>------------------------------------------</td>
<td>-------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Ple number</td>
<td>Gene RR</td>
<td>Strain Category</td>
<td>Existing promoter studies</td>
<td>Reporter</td>
<td>Brain expression relative to source gene</td>
<td>Brain expression pattern description</td>
<td>Pleiades URL (images)</td>
</tr>
<tr>
<td>------------</td>
<td>---------</td>
<td>----------------------</td>
<td>--------------------------</td>
<td>----------</td>
<td>------------------------------------------</td>
<td>--------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Ple number</td>
<td>Gene RRs</td>
<td>Strain Category</td>
<td>Existing promoter studies</td>
<td>Reporter</td>
<td>Brain expression relative to source gene</td>
<td>Brain expression pattern description</td>
<td>Pleiades URL (images)</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>-----------------</td>
<td>---------------------------</td>
<td>----------</td>
<td>------------------------------------------</td>
<td>-----------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Ple number</td>
<td>Gene RRs</td>
<td>Strain Category</td>
<td>Existing promoter studies</td>
<td>Reporter</td>
<td>Brain expression relative to source gene</td>
<td>Brain expression pattern description</td>
<td>Pleiades URL (images)</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>----------------</td>
<td>--------------------------</td>
<td>----------</td>
<td>----------------------------------------</td>
<td>-------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Ple number</td>
<td>Gene RRs</td>
<td>Strain Category</td>
<td>Existing promoter studies</td>
<td>Reporter</td>
<td>Brain expression relative to source gene</td>
<td>Brain expression pattern description</td>
<td>Pleiades URL (images)</td>
</tr>
<tr>
<td>------------</td>
<td>---------</td>
<td>----------------</td>
<td>---------------------------</td>
<td>---------</td>
<td>-----------------------------------------</td>
<td>-------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Ple number</td>
<td>Gene RRs</td>
<td>Strain Category</td>
<td>Existing promoter studies</td>
<td>Reporter</td>
<td>Brain expression relative to source gene</td>
<td>Brain expression pattern description</td>
<td>Pleiades URL (images)</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>----------------</td>
<td>--------------------------</td>
<td>---------</td>
<td>------------------------------------------</td>
<td>------------------------------------</td>
<td>---------------------</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Ple number</th>
<th>Gene RR</th>
<th>Strain Category</th>
<th>Existing promoter studies</th>
<th>Reporter</th>
<th>Brain expression relative to source gene</th>
<th>Brain expression pattern description</th>
<th>Pleiades URL (images)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ple number</td>
<td>Gene RRs</td>
<td>Strain Category</td>
<td>Existing promoter studies</td>
<td>Reporter</td>
<td>Brain expression relative to source gene</td>
<td>Brain expression pattern description</td>
<td>Pleiades URL (images)</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>----------------</td>
<td>---------------------------</td>
<td>----------</td>
<td>----------------------------------------</td>
<td>-------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Ple number</td>
<td>Gene RRs</td>
<td>Strain Category</td>
<td>Existing promoter studies</td>
<td>Reporter</td>
<td>Brain expression relative to source gene</td>
<td>Brain expression pattern description</td>
<td>Pleiades URL (images)</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>-----------------</td>
<td>---------------------------</td>
<td>----------</td>
<td>--------------------------------------</td>
<td>-------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Ple number</td>
<td>Gene RR</td>
<td>Strain Category</td>
<td>Existing promoter studies</td>
<td>Reporter</td>
<td>Brain expression relative to source gene</td>
<td>Brain expression pattern description</td>
<td>Pleiades URL (images)</td>
</tr>
<tr>
<td>------------</td>
<td>---------</td>
<td>-----------------</td>
<td>---------------------------</td>
<td>----------</td>
<td>------------------------------------------</td>
<td>-----------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Ple number</td>
<td>Gene RRs</td>
<td>Strain Category</td>
<td>Existing promoter studies</td>
<td>Reporter</td>
<td>Brain expression relative to source gene</td>
<td>Brain expression pattern description</td>
<td>Pleiades URL (images)</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>------------------</td>
<td>---------------------------</td>
<td>----------</td>
<td>------------------------------------------</td>
<td>-------------------------------------</td>
<td>----------------------</td>
</tr>
</tbody>
</table>

RRs, regulatory regions.