

**THE USE OF NOVEL HUMANIZED MOUSE MODELS AND TRANSCRIPTOME
CHARACTERIZATION TO STUDY THE NEUROGENESIS FACTOR, *NR2E1*, IN
BRAIN AND EYE DEVELOPMENT**

by

Jean-François Schmouth

M.Sc., Université de Sherbrooke, 2006

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

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2012

© Jean-François Schmouth, 2012

Abstract

For many years to come, a great challenge awaits the field of human genetics, that is, the understanding of the developmental transcriptional network, and its underlying requisite DNA elements. The laboratory mouse has been a model of choice in research for decades, with several examples of mouse models harbouring human genomic elements having been used to study human disease. For my thesis project, I used a strategy aimed at generating humanized mouse models in a high-throughput manner. This approach, namely HuGX (High-throughput Human Genes on the X Chromosome) was used to develop a set of eight humanized mouse strains to document the brain expression pattern resulting from the constructs used. The genes were chosen based on their predicted enriched expression in brain regions of therapeutic interest. The results showed that for each positive gene, expression was found in the predicted brain region, suggesting isolation of the proper underlying regulatory elements. The HuGX strategy was also used to develop a functional allele for *NR2E1*, a gene encoding a transcription factor, known to be a neural stem cell fate determinant of both the developing forebrain and retina. We used this approach to validate the hypothesis that a single copy of human *NR2E1* would be the functional equivalent of a single copy of the mouse gene, whereby the mouse heterozygote has virtually no phenotype. This approach was initially designed to serve as a platform to test the relevance of candidate *NR2E1* mutations found in patient populations. Surprisingly, this mouse model has facilitated discovery of regulatory regions, important in conferring appropriate forebrain expression of *NR2E1*. In my final chapter, I used a large scale mRNA approach to explore the role of the mouse *Nr2e1* gene during forebrain development. Using bioinformatics predictions, I identified a list of 63 genes involved in nervous system development that are predicted to be

direct targets of *Nr2e1*. Binding site prediction revealed three novel candidate co-interactors of *Nr2e1*; *Sox9*, *E2f1*, and *Nr2f1*. Future validation of these predictions will improve our understanding of the biology of *Nr2e1*.

Preface

Chapter 1

The Review highlighting the utility of the High-Throughput Human Gene on the X Chromosome (HuGX) strategy was initiated by Dr. E. M. Simpson and became focused under my direction. The BAC modification strategy was designed and conceived in collaboration with Russell Bonaguro. I wrote the paper, created all of the figures, edited, and saw the manuscript through publication (Schmouh Jean-François, Bonaguro Russell J., Corso-Diaz Ximena, Simpson Elizabeth M. Modelling human regulatory variation in mouse: finding the function in genome-wide association studies and whole-genome sequencing. *PLoS Genet.* 2012 Mar; 8(3):e1002544. [PMID: 22396661]).

Chapter 2

The project described in this chapter was initiated and conceived by Dr. E.M. Simpson and became focused under my direction. The BAC constructs were generated by members of Dr. Robert Holt's laboratory. The mice were generated with the help of, and colonies maintained by, the Pleiades Promoter Project team. All procedures involving animals in this project were in accordance with the Canadian Council on Animal Care (CCAC) and UBC Animal Care Committee (ACC) (Protocol# A09-0980 and A09-0981). Adult tissue collection and *lacZ* staining was performed by me and the Pleiades Promoter Project team. Adult brain cryosectioning, *lacZ* staining, and immunohistochemistry were performed by members of Dr. Daniel Goldowitz's laboratory. Embryo collection was performed by me and Kathleen Banks. Embryo *lacZ* staining and clearing, as well as brain, eye, and embryonic expression pattern characterization, was performed by me. The co-

localization experiments on adult retina, and *Nr2f2* co-localization experiments were performed by me. I also performed all data analysis, wrote the paper, and created all of the figures for this manuscript.

Chapter 3

The project described in this chapter was initiated and conceived by Dr. E.M. Simpson and became focused under my direction. The *Hprt* and *lacZ* cassette retrofitting was performed by members of Dr. Robert Holt's laboratory. The functional and *lacZ*-containing mouse strains generation was facilitated with the help of the Pleiades Promoter Project team. All procedures involving animals in this project were in accordance with the Canadian Council on Animal Care (CCAC) and UBC Animal Care Committee (ACC) (Protocol# A07-0435). I conducted the brain and eye expression pattern characterization and functional strain characterization. The electroretinogram experiments were performed by Drs. Kevin Gregory-Evans and Cheryl Gregory-Evans. I performed the initial bioinformatics investigation and Drs. Wyeth Wasserman and Anthony Mathelier conceived and performed a subsequent statistical analysis, validating my initial findings. I performed all data analysis, wrote the paper, created all of the figures, edited, and saw the manuscript through publication (Schmouh Jean-François, Banks Kathleen G., Mathelier Anthony, Gregory-Evans Cheryl Y., Castellarin Mauro, Holt Robert A., Gregory-Evans Kevin, Wasserman Wyeth W., Simpson Elizabeth M., Retina restored and brain abnormalities ameliorated by single-copy knock-in of human NR2E1 in null mice. *Mol Cell Biol.* 2012 Apr; 32(7):1296-311. [PMID: 22290436]).

Chapter 4

The project described in this chapter was initiated and conceived by Dr. E.M. Simpson and became focused under my direction. All procedures involving animals in this project were in accordance with the Canadian Council on Animal Care (CCAC) and UBC Animal Care Committee (ACC) (Protocol# A11-0412). The SAGE libraries were generated throughout the Mouse Atlas of Gene Expression Project. I oversaw and performed the bioinformatics analyses, and the *Nr2e1* transcription binding site matrix was designed and conceived in collaboration with Dr. Wyeth Wasserman and David Arenillas. Dr. Wyeth Wasserman and David Arenillas also performed the co-factor overrepresentation analysis. The *Lhx2* target validation in ESC and embryonic tissue was performed by me. I performed all data analysis, wrote the paper, and created all of the figures for this manuscript.

Table of Contents

Abstract.....	ii
Preface.....	iv
Table of Contents	vii
List of Tables	xiii
List of Figures.....	xiv
List of Abbreviations	xvii
Acknowledgements	xxi
Chapter 1: General Introduction	1
1.1 Novel Mouse Model Generation to Understand Human Gene Regulation	1
1.1.1 The Problem	1
1.1.2 Mice Serve as an Ideal Model to Understand Human Biology	3
1.2 Humanized Mouse Models; a Powerful Approach to Uncover Regulatory Regions and Study Human Disorders	4
1.2.1 A Large Scale Example	4
1.2.2 Gene Specific Examples.....	5
1.2.3 Traditional Approach to Generate Humanized Mouse Models; the Limitations ..	6
1.3 Excellent Techniques Exist for Single-Copy Non-Random Docking in the Mouse Genome	7
1.3.1 Docking Technologies Independent of the Insertion Site	7
1.3.2 Site Specific Docking Technology	8

1.4	High-Throughput Human Genes on the X Chromosome for High-Throughput Assaying of Human Candidate Regulatory Regions	9
1.4.1	General Description of the Approach.....	9
1.5	High-Throughput Human Genes on the X Chromosome Strategy Specific Examples; Human Gene Expression Characterization	12
1.6	High-Throughput Human Genes on the X Chromosome Strategy Specific Examples; <i>NR2E1</i> Functional Evaluation.....	13
1.6.1	Mouse <i>Nr2e1</i> ; Specific Roles in Development	13
1.6.2	Mouse <i>Nr2e1</i> ; a Role in Behaviour.....	15
1.6.3	<i>NR2E1</i> ; a Conserved Role from Human to Mouse	15
1.6.4	<i>NR2E1</i> ; Implication in Patient Populations.....	16
1.6.5	<i>Nr2e1</i> Target Gene Regulation.....	17
1.7	Large Scale mRNA Profiling Technologies for Target Gene Validation.....	18
1.7.1	Serial Analysis of Gene Expression Technology	19
1.7.2	Target Gene Promoter Analysis	20
1.7.3	Gene Ontology Term Enrichment Classifications.....	21
1.8	Thesis Objectives	23
1.8.1	Human Gene Expression Characterization.....	23
1.8.2	<i>NR2E1</i> Human Gene Functional Complementation Evaluation	23
1.8.3	Mouse <i>Nr2e1</i> Transcriptome Evaluation	24
Chapter 2: BAC Knock-in Mice Generated to Define the Genomic Boundaries of Human Genes Allowing Brain Region-Specific Expression.....		25
2.1	Introduction.....	25

2.2	Methods and Materials.....	29
2.2.1	MaxiPromoter Design	29
2.2.2	BAC Retrofitting.....	30
2.2.3	Mouse Strain Generation, Husbandry and Breeding.....	33
2.2.4	Embryo and Adult Tissue Preparation	36
2.2.5	Histology	37
2.3	Results.....	39
2.3.1	High-Throughput Construction of Humanized Mice to Study Gene Expression	39
2.3.2	Human Genes Deleted Using Flanking Functional <i>loxP</i> Sites.....	41
2.3.3	Human <i>AMOTL1-lacZ</i> Revealed Staining in Mature Thalamic Neurons in the Adult Brain, and Amacrine as well as Ganglion Cells in the Retina.	44
2.3.4	Human <i>MAOA-lacZ</i> Revealed Staining in TH-Positive Neurons in the Locus Coeruleus in Adult Brain as well as Horizontal, and Ganglion Cells in Adult Retina ...	47
2.3.5	Human <i>NOV-lacZ</i> Revealed Staining in Neurons Populating the Hippocampal Formation, Basomedial Amygdaloid Nuclei, and Cortical Layers in the Adult Brain. ..	50
2.3.6	Human <i>NR2F2-lacZ</i> Revealed Staining in Mature Neurons, Immunoreactive for the Nr2f2 Mouse Protein in the Basolateral, and Corticolateral Amygdaloid Nuclei of the Adult Brain.	53
2.3.7	Comparative Genomic Analysis Demonstrated the Boundary for Brain Specific Expression of the Subset of Candidate Human Genes Chosen.	56
2.4	Discussion.....	63
	Chapter 3: Retina Restored and Brain Abnormalities Ameliorated by Single-Copy Knock-in of Human <i>NR2E1</i> in Null Mice.....	66

3.1	Introduction.....	66
3.2	Methods and Materials.....	70
3.2.1	BAC Retrofitting.....	70
3.2.2	Strain Generation, Husbandry and Breeding.....	71
3.2.3	Embryo and Adult Tissue Preparation	75
3.2.4	Histology	76
3.2.5	Funduscopy	78
3.2.6	Electroretinograms	78
3.2.7	Comparative Genomic and Transcription Factor Binding Site (TFBS) Overrepresentation Analysis	78
3.2.8	Nucleotide Shuffling Analysis	79
3.2.9	Mouse Statistical Analysis	80
3.3	Results.....	80
3.3.1	Human <i>NR2E1-lacZ</i> Embryos Displayed an Unexpected Absence of Expression in the Dorsal Pallium in the Brain while Retaining Appropriate Region-Specific Expression in the Eyes.	80
3.3.2	Human <i>NR2E1-lacZ</i> Displayed an Unexpected Absence of Expression in Neurogenic Regions in the Adult Brain while Retaining Appropriate Cell-Type-Specific Expression in the Retina.....	84
3.3.3	<i>NR2E1/fierce</i> Animals Displayed Adult Forebrain Abnormalities and Neurogenesis Defects.....	90
3.3.4	<i>NR2E1/fierce</i> Animals Displayed Appropriate Retinal Architecture.....	94
3.3.5	<i>NR2E1/fierce</i> Animals Displayed Functional Retinas.	98

3.3.6	Comparative Genomic Analysis Revealed Candidate Brain-Specific Stem-Cell-Regulatory Elements.	100
3.4	Discussion.....	106
Chapter 4: Combined Serial Analysis of Gene Expression and Transcription Factor Binding Site Prediction Identifies Novel Targets of <i>Nr2e1</i> in Forebrain Development 110		
4.1	Introduction.....	110
4.2	Methods and Materials.....	113
4.2.1	SAGE Libraries Generation	113
4.2.2	SAGE Data Analysis	113
4.2.3	oPOSSUM Promoter Analysis	114
4.2.4	GO Term Enrichment Analysis.....	115
4.2.5	Clustering	116
4.2.6	Co-Factor Enrichment	116
4.2.7	Embryos Preparation.....	117
4.2.8	Immunofluorescence, and Imaging Analysis	118
4.2.9	Embryonic Stem Cells Culture.....	118
4.2.10	Quantitative RT-PCR	119
4.3	Results.....	120
4.3.1	LongSAGE Libraries Generation by Precise Dissection Using Laser Capture Microdissection.	120
4.3.2	Integration of Three Different Bioinformatics Tools Used for <i>Nr2e1</i> Direct Target Predictions and Functional Gene Classification.	122

4.3.3	LongSAGE Tags Expression Evaluation Results Suggested Distinct Roles for <i>Nr2e1</i> in Different Stages of Neocortex Development.	128
4.3.4	Transcription Factor Binding Sites Overrepresentation Analysis Revealed Novel Candidate NR2E1 Co-Interactors.....	131
4.3.5	Transcription Factor Gene <i>Lhx2</i> Differential Expression Revealed Accurate Prediction Using our Bioinformatics Approach.....	136
4.4	Discussion.....	139
Chapter 5: General Discussion		144
5.1	Challenges for the Future.....	144
5.1.1	Deciphering the Regulatory Sequence Code; a Gene Expression Approach	144
5.1.2	Deciphering the Regulatory Sequence Code; a Functional Approach	148
5.1.3	The Future of Large mRNA Profiling Experiments.....	150
5.2	Conclusions.....	154
References.....		156

List of Tables

Table 1.1 High-Throughput Human Genes on the X Chromosome Strategy to Evaluate the Expression Pattern of Specific Human Genes.	13
Table 2.1 Ten Human Genes Selected for Expression Pattern Characterization in Mice	31
Table 2.2 Primers Used for Reporter Gene Retrofitting	32
Table 2.3 Embryonic Stem Cells and Electroporation Details	34
Table 2.4 Expression Pattern from Reporter Mouse Strains Summary	35
Table 3.1 Human-Specific PCR Assays Designed to Check BAC Integrity.	73
Table 3.2 Relevant GO Terms.	104
Table 3.3 Neurogenesis Transcription Factors Families.....	105
Table 4.1 Gene Ontology Term Analysis Revealed Enrichment in Relevant Biological Processes.	127
Table 4.2 Overrepresentation Analysis Revealed Candidate Proximal Co-Interactors of <i>Nr2e1</i>	132
Table 4.3 Overrepresentation Analysis Gene List Revealed Candidate Direct Target Gene of <i>Nr2e1</i>	135

List of Figures

Figure 1.1 The Literature is Increasing More Slowly for Humanized Mouse Models as Compared to Genome Wide Association Studies and Human Whole Genome Sequencing, or Novel Mouse Models.....	2
Figure 1.2 Strategy for High-throughput Human Genes on the X Chromosome.....	11
Figure 2.1 Whole Bacterial Artificial Chromosomes, Harboring Human DNA, Targeted at the <i>Hprt</i> Locus were Excised Using the <i>cre/loxP</i> Recombination Technology.....	43
Figure 2.2 Human <i>AMOTL1-lacZ</i> Expressed in Mature Neurons in the Thalamus in the Adult Brain, Amacrine and Ganglion Cells in the Adult Retina, and Major Components of the Vasculature System During Development.....	46
Figure 2.3 Human <i>MAOA-lacZ</i> Expressed in TH-Positive Neurons in the Locus Coeruleus in the Adult Brain, Horizontal, and Ganglion Cells in the Retina, and Several Regions of the Developing Hindbrain.....	49
Figure 2.4 Human <i>NOV-lacZ</i> Expressed in the Basomedial Amygdaloid Nuclei, mid Cortical Layers, and Pyramidal Neurons of the Hippocampal Formation.	52
Figure 2.5 Human <i>NR2F2-lacZ</i> Expressed in Mature Neurons Populating the Basolateral and Corticolateral Amygdaloid Nuclei that are Immunoreactive for the Nr2f2 Mouse Protein. ..	55
Figure 2.6 Adult Brain <i>lacZ</i> Staining Results and Comparative Genomic Allowed Prediction of Conserved Regulatory Regions for <i>AMOTL1</i> and <i>MAOA</i>	58
Figure 2.7 Adult Brain <i>lacZ</i> Staining Results and Comparative Genomic Allowed Prediction of Conserved Regulatory Regions for <i>NOV</i> and <i>NR2F2</i>	62

Figure 3.1 <i>NR2E1-lacZ</i> Targeted at the <i>Hprt</i> Locus Demonstrated Unexpected Absence of Expression in the Dorsal Pallium in the Developing Brain while Retaining Appropriate Region-Specific Expression in the Developing Eye.	83
Figure 3.2 <i>NR2E1-lacZ</i> Targeted at the <i>Hprt</i> Locus Demonstrated Unexpected Absence of Expression in Neurogenic Regions of the Adult Forebrain.	85
Figure 3.3 <i>NR2E1-lacZ</i> Targeted at the <i>Hprt</i> Locus Demonstrated Unexpected Absence of Expression in Adult Proliferative Cells of the Dentate-Gyrus-Subgranular Layer.	87
Figure 3.4 <i>NR2E1-lacZ</i> Targeted at the <i>Hprt</i> Locus Demonstrated Unexpected Absence of Expression in Astrocyte-Like Type-B Cells in the Subventricular Zone of the Lateral Ventricle.	88
Figure 3.5 <i>NR2E1-lacZ</i> Targeted at the <i>Hprt</i> Locus Demonstrated Expected Expression in Neurons Populating the Upper Cortical Layers I and II/III.	88
Figure 3.6 <i>NR2E1-lacZ</i> Targeted at the <i>Hprt</i> Locus Demonstrated Appropriate Cell-Type Specific Adult Retina Expression.	90
Figure 3.7 <i>NR2E1/fierce</i> Mice Brains were Partially Corrected for Reduced Olfactory Bulb and Cerebrum Size.	92
Figure 3.8 <i>NR2E1/fierce</i> mice Brains were not Corrected for Adult Neurogenesis Defects. .	93
Figure 3.9 <i>NR2E1/fierce</i> Mice were Corrected for Retinal Blood Vessel Defects.	95
Figure 3.10 <i>NR2E1/fierce</i> Mice were Corrected for Retinal Architecture Defects.	97
Figure 3.11 <i>NR2E1/fierce</i> Mice were Corrected for Retinal Functional Defects.	99
Figure 3.12 Comparative Genomic Analysis Revealed Candidate Regulatory Elements. ...	102
Figure 4.1 LongSAGE Libraries Obtained by Laser Capture Microdissection were Used to Map the Transcriptome of <i>Nr2e1^{frc/frc}</i> , and Wt Embryos.	121

Figure 4.2 An Integrated Approach, Using Three Bioinformatics Tools was Used to Predict Novel Candidate Direct Targets of <i>Nr2e1</i>	124
Figure 4.3 Gene Ontology Term Analysis Revealed Enrichment in “Nervous System Development” Biological Process Category.....	131
Figure 4.4 <i>Lhx2</i> Contains Enriched Overlapping Clusters of NR2E1 Binding Sites in Highly Conserved Regions.	136
Figure 4.5 <i>Lhx2</i> , a Novel Target Gene of <i>Nr2e1</i> is Upregulated in <i>Nr2e1^{frc/frc}</i> when Compared to Wt.....	138

List of Abbreviations

ABA	Allen Mouse Brain Atlas
AHial	Anterolateral Amygdalohippocampal Area
AHiPM	Posteromedial Amygdalohippocampal Area
BAC	Bacterial Artificial Chromosome
BGEM	Brain Gene Expression Map
BLP	Basolateral Amygdaloid Nuclei
BMP	Basomedial Amygdaloid Nuclei
CA1 – CA3	Cornus Ammonis
CGE	Caudal Ganglionic Eminence
ChIP	Chromatin Immunoprecipitation
CMS	Caudal Migratory Stream
CNS	Central Nervous System
CP	Cortical Plate
CR	Conserved Region
DG	Dentate Gyrus
DP	Dorsal Pallium
DS	Down Syndrome
ENCODE	Encyclopedia of DNA Elements
ESC	Embryonic Stem Cells
GCL	Ganglion Cell Layer
GENSAT	Gene Expression Nervous System Atlas
GO	Gene Ontology

GWAS	Genome Wide Association Studies
HAT	Hypoxanthine Aminopterin Thymidine
HD	Huntington Disease
hf	Hippocampal Fissure
HGP	Human Genome Project
HPF	Hippocampal Formation
HuGX	High-Throughput Human Gene on the X Chromosome
HuMM	Humanized Mouse Models
HWGS	Human Whole Genome Sequencing
ILM	Inner Limiting Membrane
INL	Inner Nuclear Layer
IPL	Inner Plexiform Layer
LC	Locus Coeruleus
LCM	Laser Capture Microdissection
LGE	Lateral Ganglionic Eminence
LP	Lateral Pallium
MBP	Medial Parabrachial Nucleus
MePV	Posteroventral part of the Amygdaloid Nuclei
MGE	Medial Ganglionic Eminence
MP	Medial Pallium
NBL	Neuroblastic Layer
NC	Nasal Cavities
OB	Olfactory Bulb

OLM	Outer Limiting Membrane
ONL	Outer Nuclear Layer
OPL	Outer Plexiform Layer
ORE	Optic Recess
OV	Optic Vesicle
PC	Progenitor Cells
PLCo	Posteolateral Cortical Amygdaloid Nuclei
PMCo	Posteromedial Cortical Amygdaloid Nuclei
PSA	Postoptic Area
PSSM	Position Specific Scoring Matrix
QF	Quality Factor
RGC	Retinal Ganglion Cells
RM	Retromammillary Nucleus
RMGR	Recombinase-Mediated Genomic Replacement
RMS	Rostral Migratory Stream
RPC	Retinal Progenitor Cells
SAGE	Serial Analysis of Gene Expression
SNP	Single Nucleotide Polymorphism
SOA	Supra Optic Area
SR	Stratum Radiatum
SVZ	Subventricular Zone
TFBS	Transcription Factor Binding Site
TSS	Transcription Start Site

UCSC	University of California, Santa Cruz
VNO	Vomeronasal Organ
VZ	Ventricular Zone
YAC	Yeast Artificial Chromosome

Acknowledgements

First, I acknowledge my supervisor, Dr. Elizabeth M. Simpson for allowing me to perform my PhD studies in her laboratory. The knowledge and experience gained in her laboratory will be invaluable for my future endeavours, and I am grateful for that. I also thank the current members of my supervisory committee; Drs. Wolfram Tetzlaff, Keith R. Humphries, Wyeth W. Wasserman, and Weihong Song for their judicious suggestions guiding the direction of my research projects.

I acknowledge our many collaborators; Drs. Robert A. Holt, Dan Goldowitz, Wyeth W. Wasserman, Kevin Gregory-Evans, and Cheryl Gregory-Evans, laboratories, for their contributions to the work presented in this thesis. Special thanks go to Mauro Castellarin for the generation of the different BAC constructs, and to Dr. Elodie Portales-Casamar, Dr. Anthony Mathelier, and David Arenillas for their bioinformatics expertise.

I acknowledge present and past members of the Simpson laboratory for their technical support. A special thank-you goes to Kathy Banks and Russell Bonaguro, the past and current lab manager respectively, who made everything run smoothly in the laboratory on a daily basis. I am grateful for the technical help of everyone who worked on the Pleiades Promoter Project, which resulted in the generation of the mouse strains used in this thesis. Special thanks go to Dr. Bibiana K. Wong, Ximena Corso-Diaz, and Charles N. de Leeuw for advices regarding my research, and for all of the good times spent in and out of the laboratory. I thank Katrina Beppe, for all the time spent correcting my typos.

A big thank-you goes to my friends from the west coast; the time spent with you outside of the laboratory allowed me to keep my sanity, as well as explore Vancouver and its

surroundings, making the years of my PhD considerably more special. A special thank you goes to my friends from the east coast; knowing that our friendship can survive five years of time apart demonstrates to me how important you are in my life. I am extremely grateful to my family; my mom, my pop, and my two brothers for always supporting me. Finally, I want to thank Nancy Lévesque for her love and support; your influence in my life helps me to be a better person every day, and I am ever grateful to have you in my life. May our future endeavours be as crazy as this one, just a little bit shorter.

À ma poulette; je te promets que je vais te suivre dans tous tes projets de fou!!!

Chapter 1: General Introduction

1.1 Novel Mouse Model Generation to Understand Human Gene Regulation

1.1.1 The Problem

A decade ago, the Human Genome Project (HGP) published its first DNA sequence draft followed shortly by the full version in 2003 (Lander, Linton et al. 2001; Venter, Adams et al. 2001; Collins, Green et al. 2003). This project, along with the single nucleotide polymorphism (SNP) consortium and the International HapMap Project, have provided geneticists with invaluable tools for their research on human populations (Sachidanandam, Weissman et al. 2001; Frazer, Ballinger et al. 2007). These activities have resulted in an exponential growth of PubMed entries related to Genome Wide Association Studies (GWAS) plus Human Whole Genome Sequencing (HWGS) over the past decade (Figure 1.1, white bars). The increasing numbers of studies cumulated at 2,649 entries in 2010, which mainly focused on understanding the genetic variants affecting the development of diseases and disorders in humans. Protein-coding variants have been the most extensively studied so far. However, an increasing body of literature from GWAS and candidate gene association studies highlights the identification of candidate regulatory variants of potential therapeutic interest in numerous diseases (Bosma, Chowdhury et al. 1995; Nakamura, Kugiyama et al. 2002; Sugatani, Yamakawa et al. 2002; Ono, Ezura et al. 2003; Jinnai, Sakagami et al. 2004; Marzec, Christie et al. 2007; Anttila, Stefansson et al. 2010; Dubois, Trynka et al. 2010; Speliotes, Willer et al. 2010). Furthermore, with the cost of HWGS being driven down by cheaper sequencing technologies, we envision a continued exponential increase in the identification of candidate regulatory variants. In general, the biological role of variants found in putative regulatory regions is harder to predict. This is due in part to our poor

understanding of the functions of non-coding genomic sequence, and to the slow and laborious process of experimental validation of the functional significance of such variants.

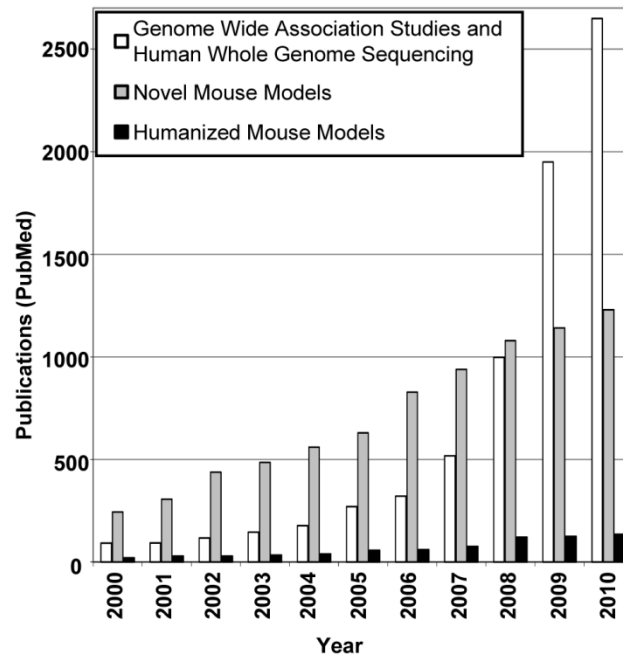


Figure 1.1 The Literature is Increasing More Slowly for Humanized Mouse Models as Compared to Genome Wide Association Studies and Human Whole Genome Sequencing, or Novel Mouse Models.

Interrogation of the PubMed literature database (<http://www.ncbi.nlm.nih.gov/pubmed>) reveals a faster growing body of literature related to Genome Wide Association Studies and Human Whole Genome Sequencing (white bars) and Novel Mouse Models (grey bars) versus Humanized Mouse Models (black bars). Interrogation of the database was done using the online search option from EndNote (<http://www.endnote.com/>). Individual numbers of entries for the search terms “Genome Wide Association Studies” and “Human Whole Genome Sequencing” were added together for the figure. Novel Mouse Models search terms were “Novel knockout mouse”, “Novel knockin mouse”, and “Novel knock-in mouse”. The entries for the search term “Humanized Mouse Models” were not restricted to genetic mouse model but included xenograft mouse models as well. Search terms were interrogated in “all fields” per year.

This highlights one of the challenges awaiting the field of human genetics research, which includes increasing our knowledge of the transcriptional network and its underlying DNA elements through fundamental research in order to create a link between candidate human variants, and their roles in disease development.

1.1.2 Mice Serve as an Ideal Model to Understand Human Biology

The laboratory mouse (*Mus musculus*) has been the human-disease model of choice for geneticists. This is in part due to the short generation time of mice that led to the development of a wide variety of inbred and spontaneous-mutation-harboring strains. Contributing to this was the advancement in technologies allowing the engineering of the mouse genome and resulting in the generation of transgenic random-insertion, knock-out, and knock-in mouse models. Furthermore, the laboratory-mouse-genome sequence was released in 2002 and demonstrated that 99% of mouse genes have clear human homologs, strengthening the importance of mouse models in probing human biology and disease (Paigen 1995; Rossant and McKerlie 2001; Waterston, Lindblad-Toh et al. 2002). This has been reflected by a continually growing literature base describing novel mouse models over the past decade (Figure 1.1, grey bars). However, in contrast to coding regions, human-mouse comparative genomic analysis demonstrated a lower level of conservation in putative regulatory regions of the genome (Waterston, Lindblad-Toh et al. 2002). This strengthened a hypothesis posed more than 25 years ago, suggesting that regulatory regions may play a crucial role in underlying species differences, and human-specific biology and disease (King and Wilson 1975). It also raises a problem for mouse modeling when a strictly mouse DNA based approach is used to validate human candidate regulatory variants, since the equivalent DNA sequence and/or epigenomic environment may not be present.

Humanized Mouse Models (HuMM), in which human genes are introduced into the mouse, suggest an approach to the environment problem. Surprisingly, the number of entries in the literature for HuMM was very modest when compared to the “GWAS-HWGS” and the “Novel Mouse Models” categories (Figure 1.1, black bars). Many of the HuMM entries are

not genetic *per se*, but related to immunity studies using human cells or tissues engrafted in nude mice, and thus unrelated to the data generated by GWAS and HWGS. Nevertheless, there are numerous examples of successful genetic HuMM.

1.2 Humanized Mouse Models; a Powerful Approach to Uncover Regulatory Regions and Study Human Disorders

1.2.1 A Large Scale Example

A HuMM approach was used to study the complex genetic condition arising from Down syndrome (DS), also known as trisomy 21. This syndrome results from an altered dosage of wild-type genes on human Chromosome 21 (*Homo sapiens* 21 (Hsa21)); a phenomenon that can be mimicked by generating trans-species aneuploid mice carrying an additional human chromosome 21 (O'Doherty, Ruf et al. 2005). In this example, the mouse strain generated contained an estimated 92% of all known Hsa21 genes and a large-scale analysis demonstrated that 81% of Hsa21 genes were expressed in mouse tissues (O'Doherty, Ruf et al. 2005; Reynolds, Watson et al. 2010). Additional investigation using a set of conserved and well characterized transcription factors, responsible for hepatocyte development and function, revealed that genetic sequence rather than interspecies differences in epigenetic machinery or cellular environment is largely responsible for directing transcriptional programs (Wilson, Barbosa-Morais et al. 2008). These results demonstrated that human gene regulation is generally conserved in mice, strengthening the argument that HuMM can be a good approach for understanding the role of candidate regulatory regions and subsequent variants in disease development.

1.2.2 Gene Specific Examples

Other examples of successful HuMM to study the role of genetic mutations are found in common monogenic disorders such as Huntington disease (HD) and β -thalassemia, as well as cancer susceptibility genes such as *BRCA1* (Hodgson, Smith et al. 1996; Lane, Lin et al. 2000; Chandler, Hohenstein et al. 2001; Vadolas, Wardan et al. 2005). When placed in HuMM, all three wild-type human genes successfully rescued the embryonic lethal phenotype from the mouse-gene knock-out animals. This provided valuable information regarding the human gene function by demonstrating an interspecies complementation of the human gene in the mouse null background. This was due not only to the similarity of the genes in terms of protein function, but also to the identical tissue expression distribution of the human gene (Hodgson, Smith et al. 1996; Lane, Lin et al. 2000; Chandler, Hohenstein et al. 2001; Vadolas, Wardan et al. 2005). Considering the low percentage of identity between human and mouse for some of the genes in both the regulatory and coding sequences, this was surprising (Chandler, Hohenstein et al. 2001).

These results were invaluable as they demonstrated that HuMM can be used to study the biological role of mutant forms of these human genes. In the case of HD, this has led to the generation of several human yeast artificial chromosome (YAC) harbouring strains to study the biological implication of expanded glutamine repeats in HD development (Hodgson, Agopyan et al. 1999; Slow, van Raamsdonk et al. 2003; Kuhn, Goldstein et al. 2007). Advancements in site-specific bacterial artificial chromosome (BAC) mutagenesis techniques supported the shift to generation of BAC-based mutation-harbouring mouse models (Narayanan, Williamson et al. 1999; Copeland, Jenkins et al. 2001; Swaminathan, Ellis et al. 2001; Yu and Bradley 2001). These included the generation of HuMM harbouring

codon-specific mutations for β -thalassemia and the *BRCA1* cancer susceptibility gene. These HuMM provided information regarding the biological implication of such mutations and their potential underlying role in human health (Yang, Swaminathan et al. 2003; Jamsai, Zaibak et al. 2005). However, the approaches used to generate these HuMM were suitable when protein-coding variants were being tested, but encountered serious limitations in probing the role of human candidate regulatory variants.

1.2.3 Traditional Approach to Generate Humanized Mouse Models; the Limitations

In general, HuMM generation has used microinjection of DNA into zygotic pronuclei (Gordon, Scangos et al. 1980; Brinster, Chen et al. 1981; Gordon and Ruddle 1981). This technique is widely used in the field of mammalian genetics, but not without limitations. For one, it requires extensive characterization of the different founder lines to control for variability in gene expression, a phenomenon due in part to the influence of the genomic environment at the site of insertion (i.e. position effect), and copy numbers, as constructs are usually found tandemly inserted in the genome (Milot, Strouboulis et al. 1996; Pedram, Sprung et al. 2006; Gao, Reynolds et al. 2007; Williams, Harker et al. 2008). This can lead to potential disruption of endogenous gene function and repeat-induced gene silencing; two factors that must be taken into account when generating mice by random insertion pronuclear injection (Garrick, Fiering et al. 1998). Since each strain is unique, reproducibility between the different mouse strains, and consequently assessment of mutant versus wild-type constructs, becomes a major limiting factor when using random insertion as a means to generate HuMM. This is less than ideal for any comparison between transgenes in different mouse strains, but is particularly concerning when probing for candidate regulatory variant

differences. Rather, a method that would control for both the site of insertion and the copy numbers inserted in the genome would be much more appropriate.

1.3 Excellent Techniques Exist for Single-Copy Non-Random Docking in the Mouse Genome

1.3.1 Docking Technologies Independent of the Insertion Site

One type of approach, which allows single-copy insertion in the genome, includes the use of retroviruses and transposon activity (Lois, Hong et al. 2002; Ding, Wu et al. 2005; Mates, Chuah et al. 2009). Although quite successful, these approaches have limitations as they do not provide controls for the site of insertion in the genome, leading to variability in expression due to genomic environment, as well as potential disruption of endogenous genes. Another potentially powerful approach called recombinase-mediated genomic replacement (RMGR) allows the cre-based insertion of a human gene at the site of, and replacing, the endogenous mouse gene (Wallace, Marques-Kranc et al. 2007). This approach provides stringent control over the genomic environment surrounding the insertion site. However, RMGR simultaneously creates two inseparable genetic events in the same gene; heterozygosity at the mouse locus and insertion of the human gene. Thus, the human gene can only be studied on the mutant mouse phenotypic background. Other limitations include the fact that the replacement is a low frequency event, and the “gene by gene” approach will restrict throughput. Another novel approach was described recently using pronuclear injection, coupled to integrase activity, to achieve single-copy site-specific insertion in the mouse genome (Tasic, Hippenmeyer et al. 2011). This approach used ϕ C31 integrase mediated recombination activity between *attB* sites from recombinant DNA with *attP* sites

previously inserted at a specific locus in the mouse genome. Although also quite promising, this approach yielded up to 40% site-specific integration at best, and was only tested on small construct plasmids, another limitation since many genes require large constructs (Tasic, Hippenmeyer et al. 2011).

1.3.2 Site Specific Docking Technology

Traditionally, two mouse genes have been used as genomic docking sites; the autosomal *Rosa26* (the reverse orientation splice acceptor 26) and X-Chromosome *Hprt* (hypoxanthine guanine phosphoribosyl transferase) (Doetschman, Gregg et al. 1987; Friedrich and Soriano 1991). The *Rosa26* locus has most often been used to dock constructs when strong-ubiquitous expression is required (Zambrowicz, Imamoto et al. 1997; Mao, Fujiwara et al. 1999; Soriano 1999; Madisen, Zwingman et al. 2010; Abe, Kiyonari et al. 2011). Plasmid-size docking is readily achieved; however large-BAC insertions have not been reported. Also, insertion at the *Rosa26* locus typically results in disruption of the gene, which in turn may lead to mild phenotypic consequences (Kohlhepp, Hegge et al. 2001). The *Hprt* docking site has also been widely used in the literature, and despite the wide expression of *Hprt* itself, this locus is more often chosen for tissue or cell-type specific expression of the targeted construct (Heaney, Rettew et al. 2004; Yurchenko, Friedman et al. 2007; Portales-Casamar, Swanson et al. 2010). This locus readily accepts plasmid-size constructs but also large >200 kb BAC constructs (Doetschman, Gregg et al. 1987; Heaney, Rettew et al. 2004). In the past, docking has been done in such a way as to disrupt the gene, resulting in mice with a mild phenotype (Finger, Heavens et al. 1988; Dunnett, Sirinathsinghji et al. 1989; Jinnah, Gage et al. 1991). However, this is typically avoided by the use of embryonic stem

cells (ESC) carrying a spontaneous deletion that removed the 5' end of the *Hprt* gene (Hooper, Hardy et al. 1987). In this strategy, docking involves construct insertion 5' of *Hprt* and repairing the expression of the *Hprt* gene itself (Bronson, Plaehn et al. 1996; Heaney, Rettew et al. 2004; Yurchenko, Friedman et al. 2007). This repair of *Hprt* enables direct selection of high-frequency correctly-targeted ESC clones (Bronson, Plaehn et al. 1996).

1.4 High-Throughput Human Genes on the X Chromosome for High-Throughput Assaying of Human Candidate Regulatory Regions

1.4.1 General Description of the Approach

As part of my thesis, I have used a strategy for high-throughput generation of humanized mouse models, namely High-throughput Human Genes on the X Chromosome (HuGX), aimed at understanding the role of regulatory regions on human gene expression. The strategy comprises; 1) the use of the BAC-adapted-recombineering technology to create a human modified BAC, 2) knock-in of this BAC into the mouse genome using the *Hprt*-docking technology, and 3) allele comparison by either interspecies complementation or reporter gene expression evaluation.

Our approach highlights the use of the RPCI-11 human male BAC library (<http://bacpac.chori.org/hmale11.htm>), which was built in the pBACe3.6 vector (Figure 1.2A). The backbone of this BAC vector contains a *SacB* gene that can be used as a targeting site for the first retrofitting step, adding the *HPRT* homology regions from plasmids pJDH8A/246b or the pEMS1306 series (Heaney, Rettew et al. 2004; Yang, Banks et al. 2009; Portales-Casamar, Swanson et al. 2010). This BAC construct can then be used as the substrate for subsequent retrofitting steps, to add a reporter gene (*lacZ* or *EGFP*) within the

start codon of the human gene (example, *NR2E1*) (Figure 1.2A). Each modified BAC construct contains the homology regions that allow proper targeting at the *Hprt* locus, and can be electroporated into ESC, selected in hypoxanthine aminopterin thymidine (HAT) media, identified using molecular biology techniques, and microinjected into mouse embryos (Figure 1.2A, Figure 1.2B) (Bronson, Plaehn et al. 1996). Male chimeras are bred to generate germline females that carry a site-specific single-copy BAC on their X Chromosome. The approach can be applied to either delineate the minimal human regulatory regions for tissue specific expression characterization or functional complementation evaluation. The latter is achieved by generating mice carrying a functional allele of the human BAC, targeted at the *Hprt* locus, and bred through two-generations of mating. A simplified version of the breeding scheme used to evaluate the complementation capacity of the human *NR2E1* gene is highlighted in Figure 1.2C. In this case, the resulting animals will carry a single copy of the human BAC on the mouse null background. Animals studied on the null background will be males, thus avoiding X-inactivation (Liskay and Evans 1980; Heaney, Rettew et al. 2004).

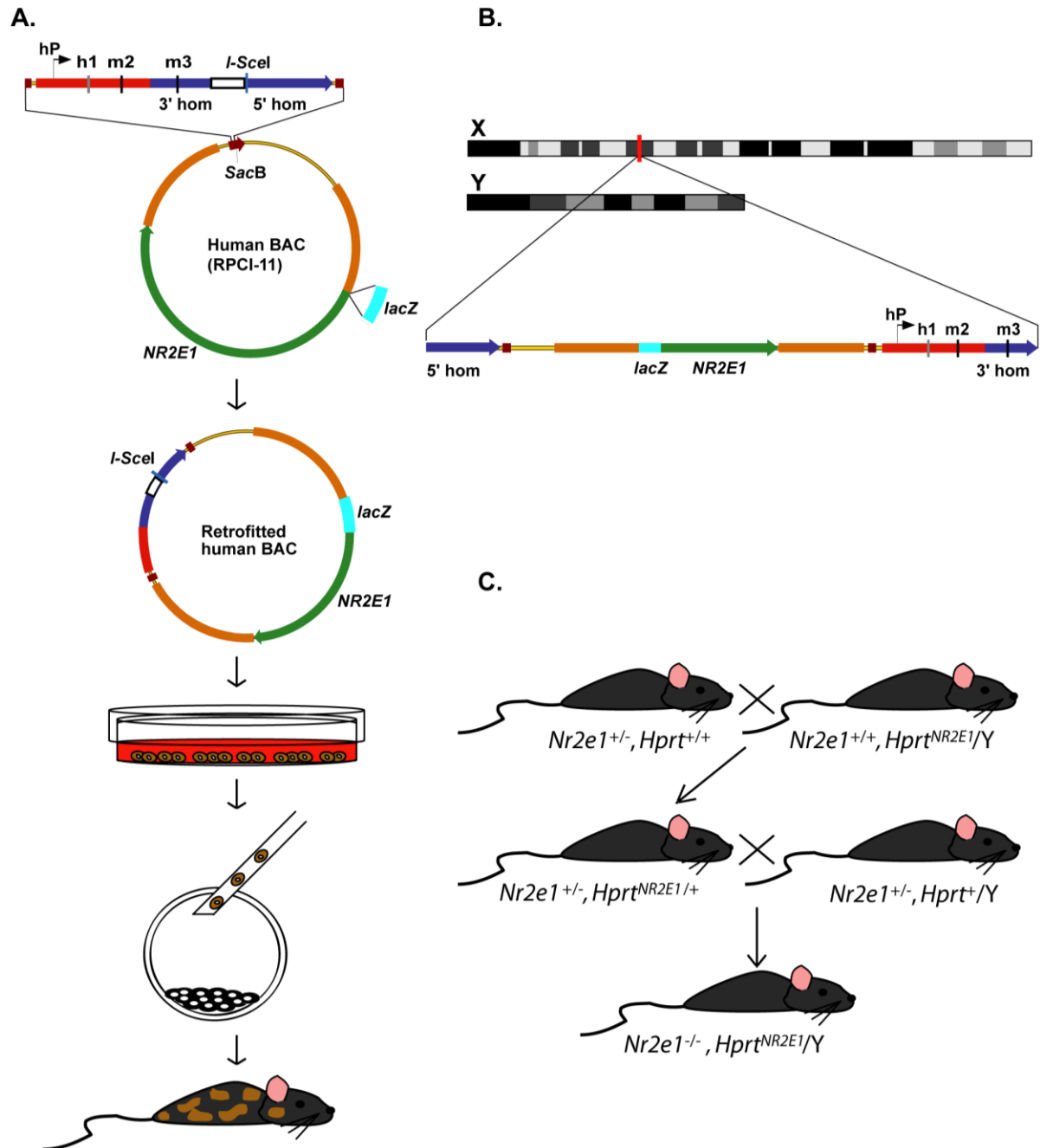


Figure 1.2 Strategy for High-throughput Human Genes on the X Chromosome

A. Flow diagram representing the major steps of the HuGX strategy, which builds on previous methods (Heaney, Rettew et al. 2004; Yang, Banks et al. 2009). Starting with a human bacterial artificial chromosome (BAC) from the RPCI-11 library, for ease of demonstration, we used *NR2E1* as an example. Two retrofitting steps are employed; 1) addition of the *HPRT* homology regions for recombination, and 2) introduction of a reporter gene (*lacZ* or *EGFP*) at the ATG of the human gene. The resulting BAC is linearized, typically using *I-SceI*, and electroporated into embryonic stem cells (ESC). 129P2/OlaHsd, B6129F1 hybrid, and C57BL/6NTac ESC are all available carrying the 36 kb (*Hprt^{b-m3}*)

deletion used for docking. Selection of homologous recombinant clones was performed using hypoxanthine-aminopterin-thymidine and clones carrying correctly-targeted-complete-BAC inserts are injected into blastocysts to generate chimeras. Schematic, not to scale. **B.** Details of knock-in 5' of the *Hprt* locus on X Chromosome. The linearized BAC construct is introduced into the *Hprt*^{b-m3} deletion by electroporation. The *Hprt* gene expression is restored by the presence of the human *HPRT* promoter (hP), first exon (h1), and second mouse exon (m2). Mouse homology arms (blue); *Hprt* coding regions (red); vector backbone (yellow with black edges); *SacB* gene from BAC vector backbone (brown); 5' and 3' untranslated regions (orange); reporter gene (turquoise); coding region of *NR2E1* (green); hP (black arrow); h1 (grey); m2, m3 (black). Schematic, not to scale. **C.** Breeding strategy to achieve complementation. Males wild-type for the mouse *Nr2e1* (*Nr2e1*^{+/+}), and harbouring the human BAC construct targeted at the *Hprt* locus (*Hprt*^{NR2E1}/Y) are bred to females heterozygous for a null allele of the mouse *Nr2e1* (*Nr2e1*^{+/-}) to generate females heterozygous for *Nr2e1* (*Nr2e1*^{+/-}) and *Hprt*^{NR2E1/+}. These females are then mated with an *Nr2e1*^{+/-} male resulting in males for study carrying a single copy of the human retrofitted *Hprt*^{NR2E1} and the mouse null (*Nr2e1*^{-/-}) gene.

1.5 High-Throughput Human Genes on the X Chromosome Strategy Specific

Examples; Human Gene Expression Characterization

The work presented in this thesis builds on efforts generated through the Pleiades Promoter project (<http://pleiades.org/>) that identifies brain-specific regulatory elements using HuMM (Portales-Casamar, Swanson et al. 2010). For this project, an initial set of 237 human genes were selected based on their mouse homologs brain-region-enriched expression patterns using a genome-wide approach (D'Souza, Chopra et al. 2008). A curation process that included looking at the potential relevance to human disease, comparative genomics, and transcription factor binding site predictions, allowed to narrow this list to 57 human genes containing a small number of well-defined conserved non-coding regions that were close to the transcription start site (TSS) (Portales-Casamar, Swanson et al. 2010). An additional set of 10 genes that were interesting based on their expression pattern and/or relevance to disease but with a structure too complex for MiniPromoter design, were set aside (details, Table 1.1).

These genes were challenging because they were particularly large or had multiple conserved regulatory regions or multiple TSS.

Gene Name	Brain Regions
<i>AMOTL1</i>	Thalamus
<i>GLRA1</i>	Brainstem, Pons, Medulla
<i>LCT</i>	Hippocampus, Dentate Gyrus
<i>MAOA</i>	Locus Coeruleus
<i>NEUROD6</i>	Hippocampus, Ammon's Horn
<i>NGFR</i>	Basal Nucleus of Meynert
<i>NOV</i>	Amygdala, Basolateral Complex
<i>NR2E1</i>	Neurogenic Regions
<i>NR2F2</i>	Amygdala, Basolateral Complex
<i>PITX2</i>	Subthalamic Nucleus

Table 1.1 High-Throughput Human Genes on the X Chromosome Strategy to Evaluate the Expression Pattern of Specific Human Genes.

The genes chosen for brain expression evaluation using the HuGX strategy are presented. Column one highlights the gene name, and column two highlights the brain region for which the gene was chosen.

1.6 High-Throughput Human Genes on the X Chromosome Strategy Specific

Examples; *NR2E1* Functional Evaluation

Amongst these genes, Nuclear Receptor 2E1 (*NR2E1*) has been the main focus for an additional project in which the HuGX strategy was used to develop a new HuMM carrying a functional allele of human *NR2E1*. This approach was used to evaluate the complementation capacity of the *NR2E1* gene when inserted as a single copy in the genome and crossed to the mouse null background.

1.6.1 Mouse *Nr2e1*; Specific Roles in Development

Nr2e1 encodes an orphan nuclear receptor (also known as *Tlx*), which is expressed in the progenitor cells populating both the forebrain lateral telencephalon, and developing

neural retina (Stenman, Yu et al. 2003; Miyawaki, Uemura et al. 2004). Additional expression of *Nr2e1* has also been reported in migrating astrocytes along the developing optic nerve (Miyawaki, Uemura et al. 2004; Uemura, Kusuhara et al. 2006).

During forebrain development, *Nr2e1* has been shown to regulate the timing of neurogenesis in the cortex by maintaining the neural stem cell population in an undifferentiated and proliferative state (Monaghan, Grau et al. 1995; Stenman, Wang et al. 2003; Li, Sun et al. 2008). Absence of expression of *Nr2e1* in forebrain development results in a progressive depletion of the neural stem cell population, a phenomenon partly attributable to precocious neuronal differentiation (Roy, Kuznicki et al. 2004; Li, Sun et al. 2008). This in turn affects subsequently generated structures such as the upper cortical layers (layers II/III), the dentate gyrus, and the olfactory bulb in the adult brain (Land and Monaghan 2003; Roy, Kuznicki et al. 2004).

In eye development, *Nr2e1* has been shown to prevent retinal dystrophy, a phenomenon partly attributable to the role of *Nr2e1* in maintaining the number of proliferating cells in the neuroblastic layer (NBL) (Miyawaki, Uemura et al. 2004; Zhang, Zou et al. 2006). Absence of expression of *Nr2e1* results in a dynamic phenotype that affects all of the layers of the adult retina. *Nr2e1*-null retina display a progressive depletion of proliferating cells found in the NBL and an increase in apoptotic cells in the ganglion cell layer (GCL) during development (Miyawaki, Uemura et al. 2004). This in turn results in a reduction of all the retinal layers in the eyes of adult null animals (Young, Berry et al. 2002; Miyawaki, Uemura et al. 2004; Zhang, Zou et al. 2006). Furthermore, *Nr2e1* has been shown to be involved in the proper vasculature of the retina through deposition of fibronectin matrices in pro-angiogenic astrocytes (Uemura, Kusuhara et al. 2006). Defects in the

deposition of the fibronectin matrices results in inappropriate distribution of the major retinal veins, and arteries in *Nr2e1*-null animals (Young, Berry et al. 2002).

1.6.2 Mouse *Nr2e1*; a Role in Behaviour

Studies using *Nr2e1*-null mice have demonstrated that the adult animals suffer from brain defects, including reduced cortical and olfactory bulb size, enlarged ventricles, and reduced dentate gyrus (Monaghan, Bock et al. 1997; Young, Berry et al. 2002). *Nr2e1*-null animals also suffer from eye abnormalities such as reduced retinal vasculature, as well as retinal, and optic nerve dystrophy resulting in blindness (Yu, Chiang et al. 2000; Young, Berry et al. 2002). The resulting brain defects were shown to affect the behaviour of the *Nr2e1*-null animals which have been shown to suffer from extreme aggressive behaviour, deficits in spatial learning and memory, hyperactivity, and reduced anxiety (Monaghan, Bock et al. 1997; Roy, Thiels et al. 2002; Young, Berry et al. 2002; Wong, Hossain et al. 2010).

1.6.3 *NR2E1*; a Conserved Role from Human to Mouse

Sequence comparison studies performed both on coding, and regulatory sequences of *NR2E1* have revealed a strong level of conservation throughout evolution. The predicted protein sequence of *NR2E1*, based on the complementary DNA (cDNA) sequence has revealed 99%, and 97% conservation to the mouse or chick proteins respectively, suggesting a conserved role from one species to another (Jackson, Panayiotidis et al. 1998). Additional sequence comparisons of *NR2E1* performed on genomic DNA from human, mouse, and puffer fish (*Fugu rubripes*) have revealed a high degree of conservation across coding and non-coding parts of the gene. A low level of interspersed repeats was also reported,

suggesting that the *NR2E1* gene expression is regulated by an array of regulatory elements embedded within the gene (Abrahams, Mak et al. 2002). Based on the sequence comparison results, a novel mouse model containing multiple copies of the human *NR2E1* gene under its endogenous promoter was generated. The results obtained from this HuMM revealed that the brain and behaviour phenotype were rescued, while the retinal defects were only ameliorated (Abrahams, Kwok et al. 2005). That the retinal defects were not completely rescued was attributed to gene dosage sensitivity during eye development, while the rescue of the brain and behaviour defects demonstrated the functional equivalence of the human gene in mouse (Abrahams, Kwok et al. 2005).

1.6.4 *NR2E1*; Implication in Patient Populations

The functional complementation of the human *NR2E1* gene in adult mouse brain and behaviour defects prompted investigators to perform population screening analysis for *NR2E1*. An evolutionary study performed on unaffected humans representing different parts of the world, as well as great apes, and monkeys, has revealed strong purifying selection for this gene, with low genetic diversity (Kumar, Leach et al. 2007). Recently, positive association results between *NR2E1* and bipolar disorder, along with an increase in detection of rare variants have been reported in patients suffering from microcephaly, bipolar disorder I, schizophrenia, psychopathy, mental retardation, and psychosis as well as paraphilia (Kumar, Leach et al. 2007; Kumar, McGhee et al. 2008).

1.6.5 *Nr2e1* Target Gene Regulation

Nr2e1 encodes a transcription factor known to recognize the canonical DNA sequence AAGTCA (Yu, McKeown et al. 1994). To date, only a handful of target genes and co-regulators of *Nr2e1* have been identified using various approaches. For instance, *Nr2e1* has been shown to regulate cell proliferation through its repressive action on the phosphatase encoding gene *Pten*, which is a known repressor of the cell cycle gene *CyclinD1* (Li, Sun et al. 2008; Liu, Belz et al. 2008). DNA sequence analysis has revealed canonical binding sites recognized by the *Nr2e1* protein product within the promoter regions of the *Pten* gene, and luciferase reporter assays have further demonstrated the repressive mechanism of *Nr2e1* on this gene (Zhang, Zou et al. 2006). The repressive action of *Nr2e1* on *Pten* involves the action of a histone demethylase encoding gene *Lsd1* (*Kdm1*) in the Y79 human retinoblastoma cell line (Yokoyama, Takezawa et al. 2008). Additionally, in the developing retina, *Nr2e1* interacts with the co-repressor atrophin1 (*Atn1*), suggesting another possible mechanism by which *Nr2e1* can regulate downstream target genes expression (Wang, Rajan et al. 2006).

A recent publication suggested that *Nr2e1* acts as a self-regulating gene via a competitive regulation loop involving a member of the SRY-box family gene, *Sox2* (Shimozaki, Zhang et al. 2012). Specific binding sites recognized by both *Sox2* and *Nr2e1* protein products are found within the promoter regions of the *Nr2e1* gene. *Sox2* binds these sites to positively regulate *Nr2e1* gene expression, while *Nr2e1* acts as a repressor on itself (Shimozaki, Zhang et al. 2012). Another example involving a regulatory loop includes the gene encoding for a member of the microRNA family, *miR-9* (Zhao, Sun et al. 2009). In this case, *miR-9* suppresses *Nr2e1* expression through its ability to bind to sequences found in the

3' untranslated region (3'UTR) of the gene. Nr2e1 in turn represses *miR-9* expression via its interaction in 3' genomic DNA sequence, closing the regulatory loop (Zhao, Sun et al. 2009). These results highlight the fact that *Nr2e1* gene expression regulation appears to rely on a dynamic balance between activators and repressors, as well as on *Nr2e1* itself.

Furthermore, *Nr2e1* regulates neurogenesis through its repressive action on promoter sequences found within astrocyte marker genes, *Gfap*, *S100 β* , and *Aqp4* (Shi, Chichung Lie et al. 2004). Adult *Nr2e1*-null neural stem cells in culture proliferate poorly, and tend to spontaneously form astrocytes, suggesting a role for *Nr2e1* in suppressing gliogenesis. This role is strengthened by the presence of canonical consensus binding sites within the promoter regions of these genes, and the ability for Nr2e1 to bind within these sites (Shi, Chichung Lie et al. 2004). These results highlight one of the many possible ways used by *Nr2e1* to regulate expression of downstream target genes. However, our ability to clearly understand the biological role of *Nr2e1* is restricted by both the simplistic models, and the tight focus of the approaches used in these examples, resulting in poor knowledge of the pathways wherein *Nr2e1* acts. Only two approaches in the literature have used large scale mRNA profiling methods in an attempt to understand *Nr2e1* importance in biological processes, one was using a model for retinal development, and the other one was looking at *Nr2e1* role in adult neural stem cells (Zhang, Zou et al. 2006; Zhang, Zou et al. 2008).

1.7 Large Scale mRNA Profiling Technologies for Target Gene Validation

Large scale mRNA profiling approaches have been used in various processes, ranging from understanding the specific role of a gene in a particular biological process, to understanding cancer biology, and to exposing the differences in expression profiles found in

different regions of the brain (Seth, Krop et al. 2002; Siddiqui, Khattra et al. 2005; Sansom, Griffiths et al. 2009). These approaches traditionally define the activity of a gene by its transcription into messenger RNA (mRNA) molecules. In most cases, with the exception of microRNA genes, the expression of an mRNA molecule is considered as the first step for the synthesis of a specific protein product. Under this definition, inactive genes do not produce any corresponding mRNA. Two predilection approaches exist to quantify mRNA expression levels in large scale experiments; microarray technology, and serial analysis of gene expression (SAGE) technology. These two methods became amenable to high-throughput levels with the evolution of automated technologies and commonly rely on a comparative approach, where samples from a control mRNA pool are compared to an experimental mRNA pool. Both technologies have their benefices and disadvantages, which are highlighted in the discussion of this thesis. For the purpose of this introduction, we will focus our attention on the high-throughput approach used in chapter 4, namely the SAGE technology.

1.7.1 Serial Analysis of Gene Expression Technology

The SAGE procedure relies on the advancements in sequencing and mapping technologies to quantify the mRNA content within a sample. The sequencing results are filed in a digital format called libraries. SAGE libraries, generally relying on the use of the mRNA or cDNA content as a starting material. The mRNA or cDNA samples are purified and cut into smaller fragments, ranging in size from 10 to 21 base pairs (bps) depending on the protocol used. These fragments are called tags, and can be cloned and sequenced in a high-throughput manner. The sequencing technology usually allows the tags to be sequenced to a

depth greater than 100,000, meaning that more than 100,000 cloned fragments per samples are being sequenced. This in turn allows for quantification of each tag with respect to the size of the library generated, and computational analysis will determine if the differential tag counts between the control and experimental samples are due to chance or statistically significant (Audic and Claverie 1997; Chen, Centola et al. 1998; Vencio, Brentani et al. 2004). The differential tag ratio between the control and experimental samples are then mapped to the corresponding genome allowing the identification of the potential target genes. A novel approach, called SAGE-lite, uses an initial PCR amplification step allowing SAGE libraries to be generated from a small amount of RNA starting material (as low as 50 ng) (Peters, Kassam et al. 1999). Modifications in the tag generation procedures, allowing generation of LongSAGE tags, have also resolved issues regarding the mapping resolution of the SAGE libraries (Wahl, Heinzmann et al. 2005). These advances made the SAGE technique a good high-throughput approach to attempt to understand the role of *Nr2e1* during forebrain development.

1.7.2 Target Gene Promoter Analysis

One of the challenges when collecting large data sets from large scale mRNA expression analysis is to identify relevant pathways explaining the differences in the gene regulation observed. New bioinformatics algorithms, designed for predicting regulatory regions found within the promoter regions of candidate genes are proving to be useful. Additionally, knowing that *Nr2e1* acts as a transcription factor recognizing the canonical DNA sequence AAGTCA, becomes an advantage in designing a high-throughput approach to identify novel direct targets (Yu, McKeown et al. 1994). Generally speaking, gene

expression regulation requires the interaction of numerous proteins that can be found either within the proximity of the transcription start site (TSS), or in distal enhancer regions. Transcription factors regulate gene expression by binding specific DNA sequences and interacting with other regulatory proteins (e. g.: chromatin modifiers, RNA polymerases). The binding specificity of these proteins can be modeled using a position specific scoring matrix (PSSM) (Stormo 2000). These matrices generally rely on published observations that have been validated experimentally, or that have been generated in high-throughput approaches (Pollock and Treisman 1990; Bulyk, Gentalen et al. 1999). Phylogenetic footprinting, an approach to identify regulatory regions by comparing conservation across genomic sequence between related species, has demonstrated to be a powerful technique when coupled to PSSM in transcription factor binding site predictions (Dermitzakis and Clark 2002; Lenhard, Sandelin et al. 2003; Ho Sui, Mortimer et al. 2005). In this project, we generated a PSSM designed to represent the binding properties of the Nr2e-subfamily of transcription factors. This PSSM was coupled to a phylogenetic footprinting analysis performed on the set of up or down regulated genes identified by SAGE analyses and included automation, which define the throughput of the method. This approach was used to predict novel direct target genes of *Nr2e1* in forebrain development.

1.7.3 Gene Ontology Term Enrichment Classifications

Another approach for the assessment of gene expression data is to evaluate enrichment of functional annotations of genes. One of the most commonly used annotation systems is the Gene Ontology (GO) resource, which represents biological information in a way that is independent of species (Ashburner, Ball et al. 2000; Consortium 2001; Harris,

Clark et al. 2004). The system addresses gene properties in three different categories, 1) molecular function, 2) biological process, and 3) cellular compartment. Bioinformatics software such as GOstats or web based services such as DAVID allows for automatic retrieval of the GO terms related to a gene list of interest and performs statistical analysis looking at enrichment of terms from the submitted list in comparison to a defined background (Dennis, Sherman et al. 2003; Falcon and Gentleman 2007; Huang da, Sherman et al. 2009). In focused studies, the list of genes obtained from large scale mRNA expression analyses will contain an enriched population of genes involved in a common biological process. By performing a GO term enrichment analysis on such a list, it is possible to identify interesting candidate genes to be evaluated in subsequent experimental studies. We used this approach to assess the biological roles of the genes predicted to be direct targets of *Nr2e1* from our list obtained from SAGE analyses.

The work presented in this thesis develops and applies genetic and genomic approaches to expand the characterization of the neurogenesis factor *NR2E1*. The two first projects demonstrate the efficiency of the HuGX strategy in generating and characterizing novel mouse models. There is an emphasis on method development in the first project (chapter 2) and application to the characterization of novel *NR2E1* harbouring mice in the second (chapter 3). Results from this second project led to the generation of hypotheses in relation to *NR2E1* expression regulation during forebrain and retina development. Thus, the main topic of the third and final project (chapter 4) is to map regulatory pathway(s) in which *Nr2e1* may act during forebrain development. The specific objectives of each project are highlighted in the following paragraphs.

1.8 Thesis Objectives

1.8.1 Human Gene Expression Characterization

The future of gene therapy lies in part on the availability of promoter regulatory regions capable of directing targeted expression of a therapeutic construct to a tissue- or cell type of therapeutic interest. Ideally, these promoters will be small in size (less than 4 kb) and consist of entirely human DNA sequence to minimize the risk of gene silencing and immunogenicity. The brain, with its diverse regions and cell types, presents a great challenge for the creation of such tools. We used a modified version of the HuGX strategy to generate HuMM for specific human genes predicted to be enriched in brain regions of therapeutic interest. These animals harboured a *lacZ* reporter cassette, which facilitated the documentation of the expression pattern for each human gene. In this objective, we sought to validate our regulatory region prediction approach. This objective, forming the first chapter of my thesis, laid the foundation for additional prediction investigations to be undertaken in an attempt to generate small promoters (MiniPromoters) to drive the expression of a human gene specifically in a brain region of therapeutic interest.

1.8.2 *NR2E1* Human Gene Functional Complementation Evaluation

Knowing that enrichment in rare variants was found in patients suffering from various brain disorders in regulatory regions of the *NR2E1* human gene, we have developed the HuGX strategy to test the relevance of these variants using mouse models. The objective was to look at whether or not the human *NR2E1* gene, inserted as a single copy in the mouse genome had the same complementation capacity of the mouse gene. This hypothesis was based on the high level of sequence similarity between the human and mouse gene, and the

fact that animals heterozygous for the mouse *Nr2e1* gene have virtually no phenotype. The ability of the human gene to complement the mouse null phenotype was evaluated both in the eyes and brain of the animals. The results coming from this approach formed the second chapter of my thesis, which aimed to establish a platform for future testing of candidate human mutations in *NR2E1*.

1.8.3 Mouse *Nr2e1* Transcriptome Evaluation

Nr2e1 has been demonstrated to be a key transcription factor regulating the neural stem cell populations of the developing forebrain. To date, only a handful of target genes have been published for this gene, and the identification of specific molecular pathway(s) on which *Nr2e1* acts remain to be identified. In an attempt to reveal such pathway(s), we used a large scale mRNA profiling approach coupled to a binding site prediction approach, and a GO term enrichment evaluation system. This objective formed the third chapter of my thesis, and the results obtained will be useful for future studies looking at deciphering the molecular pathway(s) on which *Nr2e1* acts during forebrain development.

Chapter 2: BAC Knock-in Mice Generated to Define the Genomic Boundaries of Human Genes Allowing Brain Region-Specific Expression

2.1 Introduction

The protein-coding regions of the human genome have been thoroughly characterized, but our understanding of the transcriptional network, and its underlying DNA elements remains elusive. Pioneer examples in this field include efforts from groups such as the Encyclopedia of DNA Elements (ENCODE) consortium (<http://genome.ucsc.edu/ENCODE/index.html>), seeking to catalog regulatory elements in the human genome or the Pleiades Promoter Project (<http://pleiades.org/>), identifying brain specific regulatory elements using humanized mouse models (Portales-Casamar, Swanson et al. 2010; Myers, Stamatoyannopoulos et al. 2011). Building on the expertise developed in the Pleiades Promoter Project, and to refine our knowledge behind the complexity of gene regulation, we have characterized the expression pattern of bacterial artificial chromosome (BAC) harbouring humanized mouse models for four human genes. The strategy used to generate the animals was built on a novel method for high-throughput single-copy site-specific generation of humanized mouse models, entitled HuGX (High-throughput Human Genes on the X Chromosome) (Schmuth, Bonaguro et al. 2012). The human genes (*AMOTL1*, *MAOA*, *NOV*, and *NR2F2*) were chosen based on sequence similarities between the orthologous human, and mouse genes as well as enriched expression of mouse orthologs in brain regions of therapeutic interest (D'Souza, Chopra et al. 2008). Characterization of the expression was done in development at embryonic day 12.5 (E12.5), in postnatal (P7), and adult brain as well as in adult eyes.

AMOTL1 (angiomin-like 1), initially known as junction-enriched and -associated protein (JEAP), encodes a member of the motin protein family (Bratt, Wilson et al. 2002; Nishimura, Kakizaki et al. 2002). *In vitro* and *in vivo* studies have demonstrated that the protein encoded by the *Amotl1* gene localizes at “tight” junctions in cells (Nishimura, Kakizaki et al. 2002). *Amotl1* regulates sprouting angiogenesis by affecting tip cell migration, and cell-cell adhesion *in vivo* (Zheng, Vertuani et al. 2009). Northern blot analysis demonstrated high levels of expression of the *Amotl1* gene in the brain, heart, lung, skeletal muscle, kidney, and uterus (Bratt, Wilson et al. 2002). These results differed from previously reported immunohistochemical analysis demonstrating absence of expression in the brain, heart, and kidney (Nishimura, Kakizaki et al. 2002). Discrepancy between the studies can be partly explained by the existence of different isoforms of the Amotl1 protein, highlighting the need for further characterization (Zheng, Vertuani et al. 2009).

MAOA (monoamine oxidase A), is a gene encoding a membrane-bound mitochondrial flavoprotein that deaminates monoaminergic neurotransmitters (Weyler, Hsu et al. 1990; Vitalis, Alvarez et al. 2003). In mice, expression pattern characterization of *Maoa* by combining *in situ* hybridization, and immunohistochemistry during central nervous system (CNS) development demonstrated expression in a variety of neurons, including noradrenergic and adrenergic neurons as well as dopaminergic cells in the substantia nigra (Vitalis, Fouquet et al. 2002). *Maoa* is expressed in neurons populating the developing brainstem, amygdala, cranial sensory ganglia, and the raphe (Vitalis, Fouquet et al. 2002). Transient expression in cholinergic motor nuclei in the hindbrain, and in non-aminergic neurons populating the thalamus, hippocampus, and claustrum has been detected during development (Vitalis, Fouquet et al. 2002). In rodent adult brain, *Maoa* is expressed in

neurons populating the cerebral cortices, the hippocampal formation, and the cerebellar granule cell layer (Luque, Kwan et al. 1995). *Maoa* knockout models implicate the gene as a regulator of neurochemical pathways, leading to increased level of serotonin (5-HT), norepinephrine, dopamine, and noradrenaline neurotransmitters in adult brain (Cases, Seif et al. 1995; Kim, Shih et al. 1997). This increased level of neurotransmitters lead to behavioural abnormalities including aggression, that can be rescued by generation of a *Maoa* forebrain-specific transgenic mouse model (Cases, Seif et al. 1995; Kim, Shih et al. 1997; Popova, Skrinkskaya et al. 2001; Chen, Cases et al. 2007). The role of *Maoa* in regulating neurotransmitters applies to the telencephalic neural progenitors, and retinal projections from retinal ganglion cells (RGC) during development (Upton, Salichon et al. 1999; Cheng, Scott et al. 2010).

NOV (nephroblastoma overexpressed gene), belongs to the CCN secreted proteins family (Natarajan, Andermarcher et al. 2000). *NOV* protein stimulates fibroblast proliferation via a tyrosine phosphorylation dependent pathway (Liu, Liu et al. 1999). In mice, *Nov* expression characterization by *in situ* hybridization is first observed at E10 in a subset of dermomyotomal cells of muscular origin along the entire embryonic rostro-caudal axis (Natarajan, Andermarcher et al. 2000). In the CNS, *Nov* expression was first detected at E11.5 in scattered cells of the olfactory epithelium, the developing cochlea and the trigeminal ganglia (Natarajan, Andermarcher et al. 2000). Expression observed in the olfactory epithelium extends later to cells populating the olfactory lobe (E13.5) (Natarajan, Andermarcher et al. 2000). At E12.5, and onward, *Nov* is expressed in muscle cell types of diverse developing tissues including; vertebral muscles, limb muscles (femur, and hindfoot), aorta, and other major vessels, maxillary muscles, and extra-ocular muscles (Natarajan,

Andermarcher et al. 2000). Additionally, at E12.5, *Nov* expression is observed in developing motorneurons in the ventral horns of the spinal cords (Natarajan, Andermarcher et al. 2000).

NR2F2 (nuclear receptor 2f2), also known as *COUP-TFII/ARPI*, is a gene encoding for a transcription factor belonging to the orphan receptors group. In mice, expression of *Nr2f2* by *in situ* hybridization during CNS development is first observed at embryonic day 8.5 (E8.5), peaks at E14-15 and declines after birth (Qiu, Cooney et al. 1994). *Nr2f2* expression in the developing telencephalon is restricted to the caudal lateral domains, with positive staining in the medial ganglionic eminence (MGE), and caudal ganglionic eminence (CGE) (Qiu, Cooney et al. 1994; Kanatani, Yozu et al. 2008). *Nr2f2* function in the CGE is essential for the migration of inhibitory interneurons during forebrain development (Kanatani, Yozu et al. 2008). These interneurons migrate from the CGE via the caudal migratory stream (CMS) to populate the neocortex, hippocampus, and amygdala (Nery, Fishell et al. 2002; Yozu, Tabata et al. 2005). Recent studies demonstrated a role for *Nr2f2*, and its closest relative (*Nr2f1*) in regulating the temporal specification of neural stem/progenitor cells in the ventricular zone of the developing CNS (Naka, Nakamura et al. 2008). In the adult CNS, *Nr2f2* is expressed in a subpopulation of calretinin-positive interneurons in the postnatal cortex, and a population of amacrine cells in the mouse adult retina (Kanatani, Yozu et al. 2008; Inoue, Iida et al. 2010). Finally, *Nr2f2* participates in the development and proper function of multiple organs, including the inner ear, the limbs, and skeletal muscle, heart, and pancreas (Pereira, Qiu et al. 1999; Lee, Li et al. 2004; Tang, Alger et al. 2005; Qin, Chen et al. 2010). The expression of *Nr2f2* in heart, and pancreatic development highlights the role for this gene in regulating angiogenesis, a property that in turn affects tumor growth, and metastasis in cancer (Qin, Chen et al. 2010).

In this study, we characterized for the first time, the expression of human *MAOA*, and *NR2F2*, two genes from which the mouse homologs have been extensively studied in CNS development (Qiu, Cooney et al. 1994; Upton, Salichon et al. 1999; Vitalis, Alvarez et al. 2003; Kanatani, Yozu et al. 2008; Naka, Nakamura et al. 2008; Cheng, Scott et al. 2010; Inoue, Iida et al. 2010), and *AMOTL1*, and *NOV* for which roles in CNS development are unclear.

2.2 Methods and Materials

2.2.1 MaxiPromoter Design

The BAC constructs came from the RPCI-11 human male BAC library (<http://bacpac.chori.org/hmale11.htm>, accessed January 4, 2012, BAC numbers, see Table 2.1). Suitable BACs were selected based on coverage of the gene of interest and its upstream sequence. Our approach looked at including as much proximal and distal regulatory regions as possible. Under the criteria applied, the ideal BAC would cover the entire gene up to, but not including, the neighbouring genes. If multiple BACs were available, priority was given to the one including the most upstream sequence. Two 50bp oligonucleotide recombination arms were designed for the insertion of the reporter gene in the BAC. The left arm targets immediately upstream of the gene's endogenous ATG. Ideally, the right arm targets just after the end of the same exon. Because of sequence composition, initial oligonucleotides designs in some cases were altered for the right arm to target further downstream.

2.2.2 BAC Retrofitting

BACs were modified by two subsequent steps of retrofitting using the lambda recombination system (Yu, Ellis et al. 2000). The first retrofitting step allowed the insertion of *Hprt* homologous recombination targeting arms as described in the literature (Schmouth, Banks et al. 2012). The second step allowed the insertion of a reporter cassette (*lacZ* or *EGFP*) at the ATG of the specified gene. Only the *lacZ* or *EGFP* coding sequence, including a STOP codon, was added at the ATG. No intron or polyA signals were added, as the endogenous splicing and polyA were to be used to best reflect the natural gene expression. The specific primers used for the retrofitting of each of the constructs are listed in Table 2.2. The retrofitting of the reporter cassette was done as described in the literature (Schmouth, Banks et al. 2012). Briefly, the reporter cassette was designed to contain a kanamycin gene, allowing the selection of correctly retrofitted clones. This resistance gene was designed with flanking full *frt* sites (Lyznik, Mitchell et al. 1993), which were used to excise the kanamycin gene via induction of flpe recombinase (Schlake and Bode 1994; Buchholz, Angrand et al. 1998). The resulting constructs, contained a reporter gene under the influence of the human regulatory regions of the specified gene (complete list of genes, see Table 2.1).

Brain Region(s)	Gene	Parental BAC Name	Construct Size (bp)	Reporter	BAC Name	Status
Thalamus	<i>AMOTL1</i>	RP11-936P10	211,681	<i>lacZ</i>	bEMS90	Successful
Brainstem, Pons, Medulla	<i>GLRA1</i>	RP11-602K10	188,541	N/A	bEMS89	† Failed
Hippocampus, Dentate Gyrus	<i>LCT</i>	RP11-406M16	156,570	EGFP	bEMS75	Successful
Locus Coeruleus	<i>MAOA</i>	RP11-475M12	197,191	<i>lacZ</i>	bEMS84	Successful
Hippocampus, Ammon's Horn	<i>NEUROD6</i>	RP11-463M14	N/A	N/A	N/A	¶ Failed
Basal Nucleus of Meynert	<i>NGFR</i>	RP11-158L10	175,851	<i>lacZ</i>	bEMS88	Successful
Amygdala, Basolateral Complex	<i>NOV</i>	RP11-840I14	202,342	<i>lacZ</i>	bEMS91	Successful
Neurogenic Regions	<i>NR2E1</i>	RP11-144P8	138,165	<i>lacZ</i>	bEMS86	Successful
Amygdala, Basolateral Complex	<i>NR2F2</i>	RP11-134D15	213,112	<i>lacZ</i>	bEMS85	Successful
Subthalamic Nucleus	<i>PITX2</i>	RP11-268I1	200,209	<i>lacZ</i>	bEMS87	Successful

N/A, not applicable; †*GLRA1* failed retrofitting with *Hprt* homology arms; ¶*NEUROD6* failed BAC isolation from the library.

Table 2.1 Ten Human Genes Selected for Expression Pattern Characterization in Mice

Gene Name	Forward Primers (5'-3')	Reverse Primers (5'-3')
<i>AMOTL1</i>	CCGGCAGCCGTCTTCCCCAGCCGAGGGACTGAACTAGCCATGATCGCCTC atggcggatcccgtcgtttt	GTGGGAGACTCGGAGACGCCCTCCCGGCACCTCGAGTGGGGGCTGGTTAC gaagttcctatactttctag
<i>LCT</i>	TTGCAGTTATAAAGTAAGGGTTCCACATACCTCCTAACAGTTCCCTAGAAA atggtgagcaagggcgagga	TGTGTGATGAAGGTTGCCGAGGGGTCACCATCAGGTCAATGTGTACTCAC gaagttcctatactttctag
<i>MAOA</i>	TTGCCGTCCCCACTCCTGTGCCTACGACCCAGGAGCGTGTGAGCCAAAGC atggcggatcccgtcgtttt	ACCCCTCACTGGCCAGGGTCCCCAGGCCACCGCTACGGTCCACACTGAC gaagttcctatactttctag
<i>NGFR</i>	CCGCAAAGCGGACCGAGCTGGAAGTCGAGCGCTGCCGCGGGAGGCGGGCG atggcggatcccgtcgtttt	GGAGTTCTGATCCCGGAAAGGGAGCGGGCCCCCTCCGGCTAACACTCAC gaagttcctatactttctag
<i>NOV</i>	TACAGCGAAGAAAGTCTCGTTTGGTAAAAGCGAGAGGGGAAAGCCTGAGC atggcggatcccgtcgtttt	ACCAAGGCGGGCAAAGTAACTTGGGGCATCTTAAGGTTGTGCCACTTAC gaagttcctatactttctag
<i>NR2E1</i>	GCCGGGACTCGGGCAGCGCCACCAACCGCTCCGCCCCGGGACAGCCAGC atggcggatcccgtcgtttt	TCGCCCCAGGCTGCGCGCCTAGGCCCCACGGCGGCCCGAGAGGTACCCAC gaagttcctatactttctag
<i>NR2F2</i>	CGCCGCCCCGAGCCAGGGGAGCAGGAAGTCCGGACGCAGCCCCATAGAT atggcggatcccgtcgtttt	CCAGGACCCCGGGACCCAGGACGAGGGAAGGAGAAATGAGAGGCCGATAC gaagttcctatactttctag
<i>PITX2</i>	CCGCCGCTTCTTACAGCCTTCCTTCTTCTGTTTTGCAGATAACGGGGA atggcggatcccgtcgtttt	GTGGCGCGGCCTCCCGTCCGATGACCCGGGCAGGAGAAGGGGTTCTTAC gaagttcctatactttctag

*Capital letters, sequence matching the mouse genome; lowercase letters, sequence matching the reporter gene (forward primer) or the kanamycin cassette (reverse primer).

Table 2.2 Primers Used for Reporter Gene Retrofitting

2.2.3 Mouse Strain Generation, Husbandry and Breeding

The strains were generated using a variation of the previously described strategy to insert constructs 5' of *Hprt* on the mouse X Chromosome (HuGX) (Bronson, Plaehn et al. 1996; Heaney, Rettew et al. 2004; Yang, Banks et al. 2009; Schmouth, Bonaguro et al. 2012). Briefly, BAC DNA was purified using the Nucleobond BAC 100 kit (Clontech laboratories, Mountain View, California) and linearized with *I-SceI*. The BAC constructs were electroporated in ESC using the following conditions: voltage, 190 V; capacitance, 500 μ F; resistance, none; using a BTX ECM 630 Electro cell manipulator (BTX, San Diego, California) (Schmouth, Banks et al. 2012). ESC clones were selected in hypoxanthine aminopterin thymidine (HAT), isolated, and DNA purified. Specific human PCR assays were designed to characterize the integrity of the entire BAC construct inserted in the mouse genome as previously described (Yang, Banks et al. 2009; Schmouth, Banks et al. 2012). Table 2.3 list all the different ESC lines used and their associated genotype. ESC derivation and culture was conducted as previously described (Yang, Banks et al. 2009). Correctly targeted ESC clones were microinjected into B6(Cg)-*Tyr^{c-2}*/J (JAX Stock#000058) blastocysts to generate chimeras that were bred to C57BL/6J (B6) (JAX Stock#000664) to obtain offspring carrying the BAC insert. Backcrossing to B6 continued such that mice used in this study were N3 or higher. Table 2.4 list the details about the strains used in this study, these are available at The Mutant Mouse Regional Resource Center (MMRRC) (<http://www.mmrrc.org/>). Male animals were used in all studies to avoid any variability due to random X-inactivation of the knock-in alleles at *Hprt*.

Gene	Parental ESC Line Name	Genotype	ESC Clones		
			No. Isolated	% Correctly Targeted	Final ESC Line mEMS
<i>AMOTL1</i>	mEMS1202.04	B6129F1-Gt(<i>ROSA</i>)26Sor ^{+/+} , <i>Hprt1</i> ^{b-m3} /Y	8	50	4645
<i>LCT</i>	mEMS1204.31	B6129F1-Gt(<i>ROSA</i>)26Sor ^{tm1Sor/+} , <i>Hprt1</i> ^{b-m3} /Y	31	48	3305
<i>MAOA</i>	mEMS1202.04	B6129F1-Gt(<i>ROSA</i>)26Sor ^{+/+} , <i>Hprt1</i> ^{b-m3} /Y	15	53	4442
<i>NGFR</i>	mEMS1202.04	B6129F1-Gt(<i>ROSA</i>)26Sor ^{+/+} , <i>Hprt1</i> ^{b-m3} /Y	20	20	4583
<i>NOV</i>	mEMS1202.04	B6129F1-Gt(<i>ROSA</i>)26Sor ^{+/+} , <i>Hprt1</i> ^{b-m3} /Y	24	42	4521
<i>NR2E1</i>	mEMS1204	B6129F1-Gt(<i>ROSA</i>)26Sor ^{tm1Sor/+} , <i>Hprt1</i> ^{b-m3} /Y	29*	21	4751
<i>NR2F2</i>	mEMS1292.02	B6129F1-A ^{w-J} /A ^w , Gt(<i>ROSA</i>)26Sor ^{tm1Sor/+} , <i>Hprt1</i> ^{b-m3} /Y	25*	28	4990
<i>PITX2</i>	mEMS1202.04	B6129F1-Gt(<i>ROSA</i>)26Sor ^{+/+} , <i>Hprt1</i> ^{b-m3} /Y	15	20	4496

* Obtained from multiple electroporations

Table 2.3 Embryonic Stem Cells and Electroporation Details

Gene	Embryo E12.5	P7 Brain	Adult			MMRRC Strain Name	MMRRC. Stock No
			Brain	Eye	Other		
<i>AMOTL1</i>	Positive	Positive	Positive	Positive	Positive	B6.129P2(Cg)- <i>Hprt</i> ^{tm66(Ple5-lacZ)Ems} /Mmjax	012354
<i>LCT</i>	Negative	Negative	Negative	N/D	N/D	N/A	N/A
<i>MAOA</i>	Positive	Positive	Positive	Positive	Positive	B6.129P2(Cg)- <i>Hprt</i> ^{tm68(Ple127-lacZ)Ems} /Mmjax	012583
<i>NGFR</i>	N/D	Negative	Negative	Negative	Negative	N/A	N/A
<i>NOV</i>	Positive	Positive	Positive	Negative	Positive	B6.129P2(Cg)- <i>Hprt</i> ^{tm69(Ple134-lacZ)Ems} /Mmjax	012584
<i>NR2E1</i>	Positive	Positive	Positive	Positive	N/D	B6.129P2(Cg)- <i>Hprt</i> ^{tm73(Ple142-lacZ)Ems} /Mmjax	032962
<i>NR2F2</i>	Positive	Positive	Positive	Negative	Positive	B6.129P2(Cg)- <i>Hprt</i> ^{tm75(Ple143-lacZ)Ems} /Mmjax	014536
<i>PITX2</i>	N/D	Negative	Negative	Negative	Negative	N/A	N/A

N/D, not determined; N/A, not applicable

Table 2.4 Expression Pattern from Reporter Mouse Strains Summary

The strains used for the *cre/loxP* experiment were bred as follow: females, heterozygotes for the human BAC reporter genes (*Hprt*^{tm66(Ple5-lacZ)Ems/+}, *Hprt*^{tm69(Ple134-lacZ)Ems/+}, *Hprt*^{tm75(Ple143-lacZ)Ems/+}) were crossed to males, heterozygotes for the ACTB-cre gene (ACTB-cre/+). The resulting offspring contained experimental males, hemizygotes for the reporter genes, and heterozygotes for the ACTB-cre gene (*Hprt*^{tm66(Ple5-lacZ)Ems/Y}, ACTB-cre/+; *Hprt*^{tm69(Ple134-lacZ)Ems/Y}, ACTB-cre/+; *Hprt*^{tm75(Ple143-lacZ)Ems/Y}, ACTB-cre/+), and control males, hemizygotes for the reporter genes only (*Hprt*^{tm66(Ple5-lacZ)Ems/Y}, +/+; *Hprt*^{tm69(Ple134-lacZ)Ems/Y}, +/+; *Hprt*^{tm75(Ple143-lacZ)Ems/Y}, +/+). Animals of the appropriate genotype were kept and processed for *lacZ* staining as described below.

All mice were maintained in the pathogen-free Centre for Molecular Medicine and Therapeutics animal facility on a 7 am–8 pm light cycle, 20 ± 2°C with 50 ± 5% relative humidity, and had food and water *ad libitum*. 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).

2.2.4 Embryo and Adult Tissue Preparation

Time-pregnant mice were euthanized by cervical dislocation and embryos at E12.5 were dissected, and then fixed in 4% paraformaldehyde (PFA) with 0.1 M PO buffer (pH 7.2-7.4) for 4 hours at 4°C. Whole embryos were incubated in *lacZ* staining solution (X-gal (1 mg/ml), MgCl₂ (2 mM), K₃Fe(CN)₆ (4 mM), K₄Fe(CN)₆ (4 mM) in 1x phosphate buffered saline (PBS)) overnight at 37°C and were subsequently washed in three volumes of 1x PBS before being photographed. Embryos having the proper genotype were cleared as described

in the literature (Schatz, Golenser et al. 2005) and pictures were taken in 100% glycerol solution.

Intracardial perfusions were performed on avertin-anesthetized mice with 4% PFA with 0.1 M PO buffer (pH 7.2-7.4). Brain tissues destined to be 1 mm sectioned were collected and post fixed in 4% PFA for an additional 2 hours at 4°C before being sectioned in a rodent brain matrix (RBM-2000S/RBM-2000C, ASI instruments, Michigan, USA). The sectioned brains were incubated in *lacZ* staining solution for a duration varying between 2 hours to overnight at 37°C and were washed in 1x PBS before being photographed. Brain tissues destined to be cryosectioned were directly transferred to 20% sucrose with 0.05 M PO buffer overnight at 4°C and embedded the next day in OCT on dry ice. Eye tissues destined to be cryosectioned were incubated in *lacZ* staining solution overnight at 37°C and were washed in 1x PBS before being post-fixed for two hours in 4% PFA at 4°C. The eyes were washed in 1x PBS, then transferred to 20% sucrose with 0.05 M PO buffer overnight at 4°C and embedded the next day in OCT on dry ice.

2.2.5 Histology

For immunofluorescence, 25 µm cryosections from adult brains (floating sections) were rehydrated in sequential washes of PBS, permeabilized in PBS with 0.1% triton, and quenched in 0.1 M glycine-PBS solution. The cryosections were blocked with 1% bovine serum albumin (BSA) in PBS with 0.1% triton for 1 hour at room temperature before applying the primary antibodies. Co-localization experiments were performed using chicken anti-β-gal antibody (Abcam, San Francisco, California, ab9361) (1:5,000), mouse anti-NR2F2 antibody (R&D systems, PP-H7147-00) (1:100), and incubated overnight at 4°C.

Corresponding secondary antibodies coupled to Alexa 488 or Alexa 594 (Invitrogen, Burlington, Ontario) were incubated at room temperature for 2 hours in the dark (1:1,000). Hoechst 33342 was used for nuclear staining on all sections.

For immunohistochemistry staining on adult brain, 25 μm cryosections (floating sections) were rehydrated in sequential washes of PBS, permeabilized in PBS with 0.1% triton before being incubated in *lacZ* solution overnight at 37°C. The following day, the sections were rinsed, post-fixed in 2% PFA for 10 minutes, and blocked with 0.3% BSA, 10% normal goat serum (NGS) solution for 20 minutes. Primary antibodies were incubated overnight at 4°C using the following dilutions: rabbit anti-TH antibody (Pel-freez, P40101-0) (1:500), mouse anti-NeuN (Millipore, MAB377) (1:500), mouse anti-GFAP (NEB, cell signaling technology, mAB3670) (1:200), rabbit anti-Brn3 (Santa Cruz, sc-28595) (1:1,000), rabbit anti-calbindin (Abcam, ab49899). The third day, sections were rinsed and corresponding secondary antibody coupled to biotin (Vector laboratories, Burlingame, CA) were incubated at room temperature for 1 hour (1:200). The sections were finally processed for standard avidin–biotin immunocytochemical reactions using the ABC kit from Vector Laboratories (Burlingame, CA). Immunolabeling was visualized using 3,3-Diaminobenzidine tetrahydrochloride (DAB) (Roche, St. Louis, MO). Sections were dehydrated in subsequent washes in 50-100% ethanol, and xylene before being mounted for microscopy.

For adult eyes, stained with *lacZ*, 20 μm cryosections were mounted directly on slides and pressed for 30 min before being processed for counter staining or antibody labeling. Antibody staining was performed as previously described for brain floating sections. Counter staining was performed using neutral red as follow: the slides were washed once in PBS for 2 min, followed by a 2 min wash in water. They were then incubated 45 sec in neutral red

solution before being subsequently dehydrated in 70-100% ethanol and xylene and then mounted for microscopy.

2.3 Results

2.3.1 High-Throughput Construction of Humanized Mice to Study Gene Expression

Previously, the Pleiades Promoter Project studied gene expression in the brain using human MiniPromoters of less than 4 kb (Portales-Casamar, Swanson et al. 2010). For this project, 57 human genes were selected that contained well-defined and conserved non-coding regions that were close to the transcription start site (TSS) (Portales-Casamar, Swanson et al. 2010). An additional set of ten genes that were interesting based on their expression pattern and/or relevance to disease were omitted from the MiniPromoter project because they either had regulatory regions that were too large, had multiple conserved regulatory regions or they had multiple TSS. For these genes, we designed MaxiPromoters as an alternative (D'Souza, Chopra et al. 2008). A MaxiPromoter consists of a Bacterial Artificial Chromosome (BAC) that has a reporter sequence (*lacZ* or *EGFP*) inserted at the start codon. For a BAC to be suitable, it had to cover the whole gene sequence plus the upstream sequence up to the neighbouring gene. It also could not include the promoter sequence of any additional genes to avoid misexpression.

The ten human genes used in this project had enriched expression of their mouse homologs in brain regions of therapeutic interest (Table 2.1) (brain regions, column one; gene name, column two). The expression pattern of these human genes was analysed using our novel mouse models, generated through the HuGX method, i.e., harbouring a human BAC reporter construct docked at the *Hprt* locus (Schmuth, Bonaguro et al. 2012). This

approach was used to validate the predicted minimal boundaries for adult brain specific expression of the selected human genes, and to allow documentation of the expression pattern resulting from the selected BAC constructs at different developmental stages.

Of the list of ten selected genes, eight custom BACs were successfully constructed that contained an *Hprt* complementation cassette, and a reporter gene (*lacZ* or *EGFP*) (Table 2.1, column five). One BAC construct, RP11-463M14 (*NEUROD6*), was discarded in the process due to rearrangement in the parental clone, resulting in a smaller DNA construct than expected. This construct was not advanced in the pipeline to be retrofitted. Using our retrofitting approach, a success rate of ~89% was obtained for the nine constructs remaining, with the largest construct generated spanning ~213 kb (Table 2.1, column four). *GLRA1* failed to be properly retrofitted partly due to an irresolvable primer design issue (Table 2.1, column seven). While many technical challenges can occur when using the recombineering technology in a high-throughput manner, the 89% success rate demonstrates the efficiency of the technology.

The eight retrofitted constructs were electroporated in embryonic stem cells (ESC), and positive recombinant clones were selected, and screened as described in materials and methods. A success rate of ~35% was obtained for this step (Table 2.3, column five). The corresponding ESC were microinjected to generate chimeric animals, and expression analysis was evaluated in P7, adult brains, and adult eyes of N3 to N9 animals as well as E12.5 embryos. Gross dissection and investigation of staining in the whole body of the animals was also performed. Based on this approach, a rate of ~56% was observed for positive expression of the human gene in the mouse tissues (Table 2.4, columns two to six). As the first processed MaxiPromoter construct, the *LCT* gene contained an *EGFP* reporter, which was

subsequently excluded due to sensitivity concerns. The failure to detect expression from the *LCT* MaxiPromoter could be attributable to sensitivity issues regarding direct read-out of fluorescence from the *EGFP* reporter gene. All the other constructs were retrofitted with *lacZ* which offers the advantage of greater sensitivity than *EGFP* when a direct read-out of the enzymatic product is used as the principal means of detection. The two other negative results, obtained with *NGFR* and *PITX2* could be attributable to the complexity of the human gene structures by themselves or the fact that the conserved regions, included in the BAC construct used in this study were non-functional in the mouse genome. For example, the Gene Expression Nervous System Atlas (GENSAT) (<http://www.gensat.org/index.html>) has reported positive expression in brain regions for their *Pitx2* BAC harbouring mouse model (RP24-215O15) (Gong, Zheng et al. 2003). Sequence comparison analysis, using relative coordinates between human and mouse, revealed that our human RP11-268I1 BAC construct was shorter at the 5' end by approximately 30 kb (data not shown). This suggests an important role for this 30 kb in regulating *PITX2* expression, and will require additional investigation to assess the functionality of the missing 30 kb. In this study, we highlight the expression patterns observed from four of the five positive humanized mouse strains generated in this project: *AMOTL1-lacZ*, *MAOA-lacZ*, *NOV-lacZ* and *NR2F2-lacZ*. The detailed analysis of the *NR2E1-lacZ* strain was demonstrated in a previous publication and will be discussed in chapter 3 (Schmouth, Banks et al. 2012).

2.3.2 Human Genes Deleted Using Flanking Functional *loxP* Sites

The HuGX method used in this study consisted in modifying existing human BAC constructs to harbour an *Hprt* complementation cassette that allowed site specific targeting in

the genome of *Hprt*^{b-m3} ESC by homologous recombination (Schmuth, Bonaguro et al. 2012). The resulting complementation targeting event was detectable using described PCR assays (Yang, Banks et al. 2009) The insertion of the *Hprt* complementation cassette involved recombineering, and resulted in the disruption of a *SacB* gene present on the original BAC construct as well as the addition of *loxP* sites, flanking the entire human gene. To verify that these *loxP* sites were functional, we designed an experiment involving crossing of *AMOTL1-lacZ*, *NOV-lacZ* and *NR2F2-lacZ* females to ACTB-cre males (Figure 2.1A, Figure 2.1B). Brain expression pattern resulting from male offspring, positives for the complementation at the *Hprt* locus by genotyping, were evaluated based on the presence or absence of the ACTB-cre transgene. The results showed that in all strains, *lacZ* staining (blue) was only found in males, positive for the complementation at the *Hprt* locus and negative for the presence of the ACTB-cre transgene (*AMOTL1-lacZ*, *NOV-lacZ*, and *NR2F2-lacZ*) (Figure 2.1C-E). Absence of *lacZ* staining was only found in brains of the animals that were positive for both the complementation at the *Hprt* locus and the ACTB-cre transgene (*AMOTL1-lacZ*, ACTB-cre; *NOV-lacZ*, ACTB-cre; *NR2F2-lacZ*, ACTB-cre), suggesting complete excision of the BAC construct from the genome and proper function of the *loxP* sites (Figure 2.1C-E).

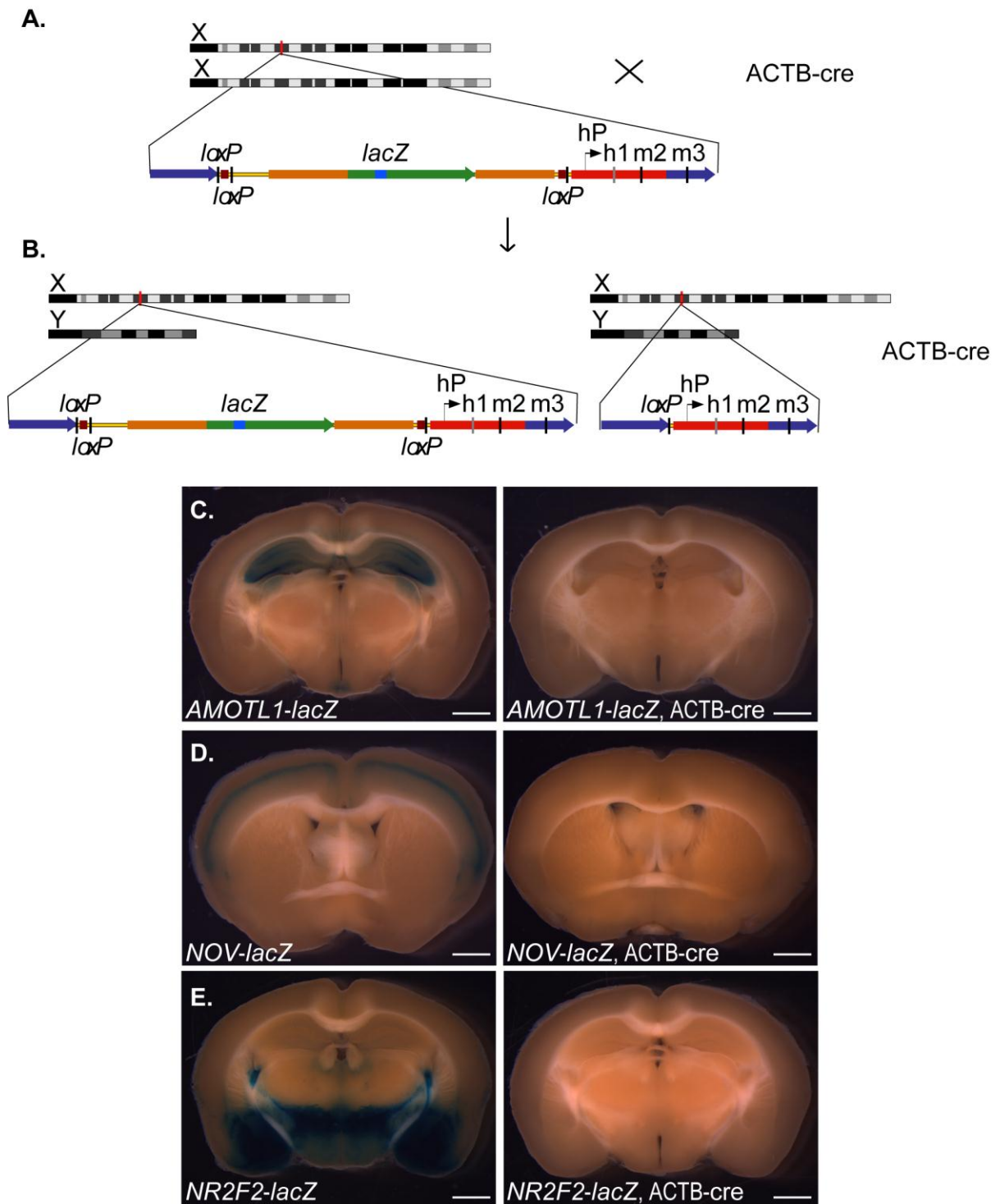


Figure 2.1 Whole Bacterial Artificial Chromosomes, Harboursing Human DNA, Targeted at the *Hprt* Locus were Excised Using the *cre/loxP* Recombination Technology.

Human bacterial artificial chromosome (BAC) constructs positive for *lacZ* (BAC-*lacZ*), targeted at the *Hprt* locus were excised using the *cre/loxP* recombination technology. **A.** The integration in the genome of the BAC-*lacZ* reporter constructs generated in this project resulted in the presence of three *loxP* sites in the genome (two in the 5' end and one in

the 3' end) **B.** Crossing the BAC-*lacZ* reporter females to ACTB-cre males should result in the generation of two different males offspring; BAC-*lacZ* reporter animals, wild type for the ACTB-cre transgene; and BAC-*lacZ* reporter animals harbouring the ACTB-cre transgene. Only the reporter animals, positive for the ACTB-cre gene should recombine the *loxP* sites, resulting in an excision of the BAC construct from the genome, and leaving only one *loxP* site, resulting in an absence of *lacZ* positive signal. hP, human *HPRT* promoter; h1, human first exon; m2 and m3, mouse second and third exons; mouse homology arms (dark blue); *Hprt* coding regions (red); vector backbone (yellow with black edges); *SacB* gene from BAC vector backbone (brown); 5' and 3' untranslated regions of the human gene (orange); coding region of the human gene (green); *lacZ* reporter gene (light blue). Schematic, not to scale. **C, D, E.** *lacZ* expression results comparison from *AMOTL1-lacZ*, *NOV-lacZ*, and *NR2F2-lacZ* females bred to the ACTB-cre males are presented. *lacZ* positive staining (blue) was found in *AMOTL1-lacZ*, *NOV-lacZ*, and *NR2F2-lacZ* males not harbouring the ACTB-cre allele whereas absence of staining was found in males positive for ACTB-cre by genotyping (*AMOTL1-lacZ*, ACTB-cre; *NOV-lacZ*, ACTB-cre; *NR2F2-lacZ*, ACTB-cre), suggesting whole BAC excision from the genome. N=3 for all genotype (Scale bar, 1 mm [**C, D, E**]).

2.3.3 Human *AMOTL1-lacZ* Revealed Staining in Mature Thalamic Neurons in the Adult Brain, and Amacrine as well as Ganglion Cells in the Retina.

Gene choosing in this work was based in part on results from SAGE expression profile of brain regions of therapeutic interest, as well as data mining of both the Allen Mouse Brain Atlas (ABA) (<http://www.brain-map.org/>), and the Brain Gene Expression Map (BGEM) (<http://www.stjudebgem.org/web/mainPage/mainPage.php>) (Magdaleno, Jensen et al. 2006; Lein, Hawrylycz et al. 2007; D'Souza, Chopra et al. 2008) The expression patterns for all humanized mouse strains generated were subsequently evaluated in E12.5 embryos, P7 developing brain as well as adult tissues.

AMOTL1 was proposed as a candidate gene for the thalamus in the adult brain. Whole mount *lacZ* E12.5 stained embryos revealed expression of the *AMOTL1* reporter construct in components of the vasculature system such as the jugular vein, the posterior cerebral artery, and the vertebrate arteries (Figure 2.2A). Additional analysis using embryos subjected to a clearing protocol confirmed staining in the previously described regions, and revealed

expression in the basilar artery, the dorsal aorta, and vertebrate arteries (Figure 2.2B). Staining was observed in the omphalomesenteric vascular system, and the lateral nasal prominence (Figure 2.2B). P7 developing mouse brain sections revealed expression in the anterior thalamic nuclei, and hippocampal formation (HPF) (Figure 2.2C). Adult mouse brain sections revealed staining in the anteromedial, and anteroventral thalamic nuclei, the pontine nuclei, and the HPF, with sharp staining in the cornu ammonis fields (CA1 and CA3) (Figure 2.2D, Figure 2.2E). Additional staining was detected in the subiculum that extended to the lower layers of the cortex (layer VI) (Figure 2.2E). Positive co-localization between the β -gal staining product (blue), and a NeuN specific antibody (brown) indicated that the *AMOTL1* reporter construct was expressed in mature neurons in the anteroventral thalamic nuclei (Figure 2.2F). In the same brain region, the β -gal positive cells (blue) did not co-localize with GFAP (brown), a marker of astrocytes (data not shown). Further *AMOTL1-lacZ* expression pattern characterization revealed staining in the adult retina, extending from the inner limiting membrane (ILM) to the junction between the inner plexiform layer (IPL) and the inner nuclear layer (INL) (Figure 2.2G). The β -gal staining product (blue) was detected at the junction of the IPL, and INL, co-localized with anti-Tyrosine Hydroxylase (anti-TH) (brown), demonstrating expression of the *AMOTL1* reporter construct in a subpopulation of amacrine cells found in the INL (Figure 2.2H) (Haverkamp and Wassle 2000). Co-localization using an anti-Brn3 antibody revealed expression in the ILM corresponding to ganglion cells in the retina (Figure 2.2I). Additional staining was found in the nose structures, extending from the naris, to the nasopharynx, and at the sternum junction, in the extremities of the ribs in the *AMOTL1-lacZ* animals in gross dissections (data not shown).

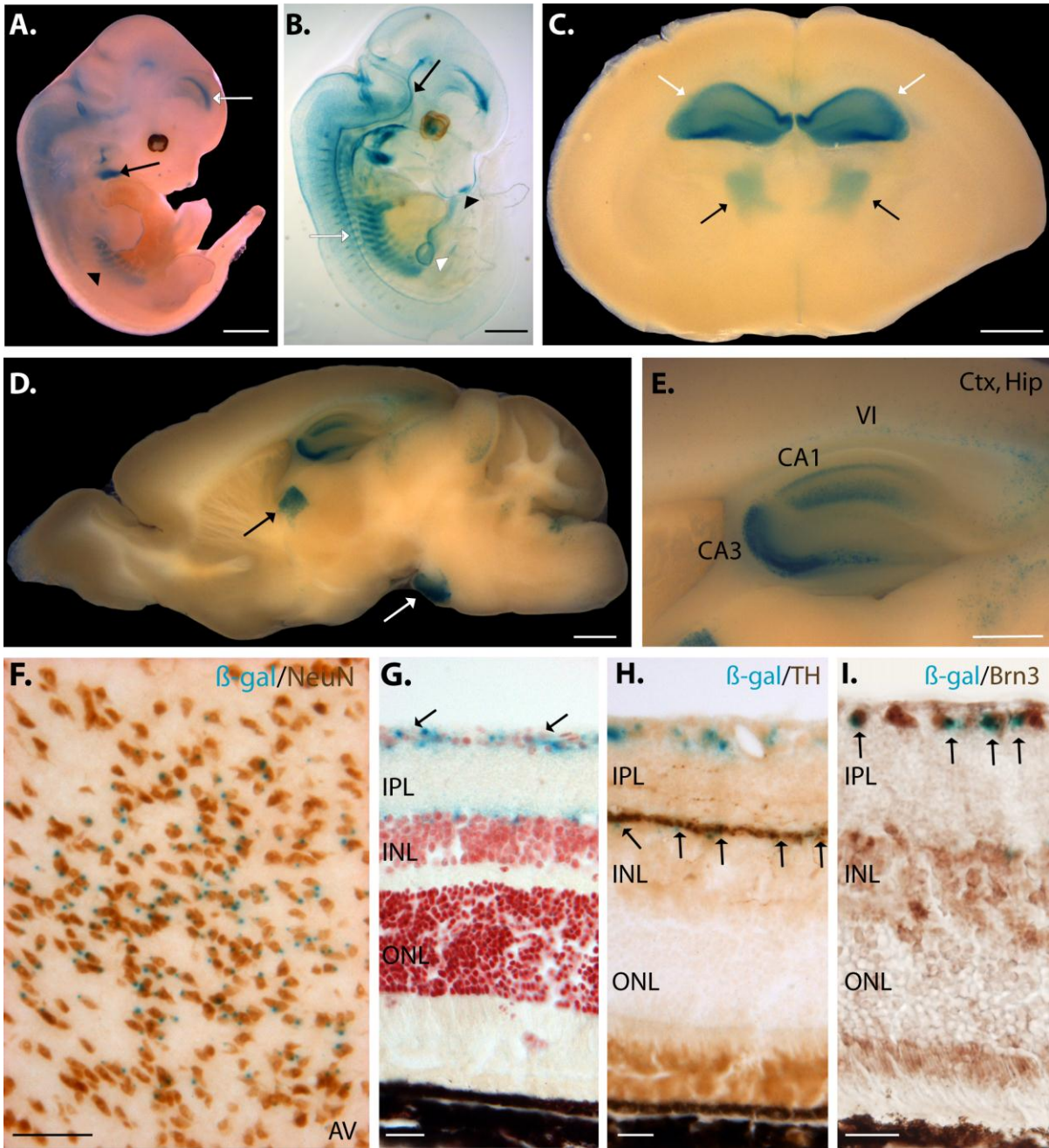


Figure 2.2 Human *AMOTL1-lacZ* Expressed in Mature Neurons in the Thalamus in the Adult Brain, Amacrine and Ganglion Cells in the Adult Retina, and Major Components of the Vasculature System During Development.

Expression analysis of the human *AMOTL1-lacZ* strain was examined using β -galactosidase (β -gal) histochemistry (blue). **A.** E12.5 whole embryos stained in the jugular vein (black arrow), posterior cerebral artery (white arrow), and vertebral arteries (black arrowhead). **B.** E12.5 cleared embryos additionally demonstrated staining in the basilar artery (black arrow), the dorsal aorta (white arrow), the omphalomesenteric vascular system (white arrowhead), and the lateral nasal prominence (black arrowhead). **C.** P7 brains stained in the developing anterior thalamic nuclei (black arrows), and hippocampal formation (HPF) (white arrows). **D.** Adult brains stained in the anteromedial and anteroventral thalamic nuclei

(black arrow). Staining was present in the pontine nuclei (white arrow). **E.** The HPF showed sharp staining in the cornu ammonis fields (CA1, and CA3). Staining was also present in the dorsal subiculum, which extended to the lower cortical layers of the cortex (layer VI). **F.** Co-localization using β -gal staining and anti-NeuN immunocytochemistry (brown) on adult brain cryosections revealed co-localization of *AMOTL1-lacZ* in thalamic neurons. **G.** Staining was found in two layers of the adult retina, the inner limiting membrane (ILM) (black arrows), and the junction between the inner plexiform layer (IPL) and the inner limiting membrane (ILM). **H.** Co-localization experiment using β -gal staining and an anti-Tyrosine Hydroxylase (anti-TH) antibody (brown) performed on adult eye cryosections indicated co-localization of *AMOTL1-lacZ* in amacrine cells populating the INL (black arrows). **I.** Co-localization experiment using β -gal staining and an anti-Brn3 antibody (brown) performed on adult eye cryosections indicated co-localization of *AMOTL1-lacZ* in ganglion cells populating the ganglion cell layer (GCL) (black arrows). Ctx: cortex, Hip: Hippocampus, AV: anteroventral thalamic nucleus, ONL: outer nuclear layer (Scale bar, 1 mm [A, B, C, D], 500 μ m E, 100 μ m F, 20 μ m [G, H, I]).

2.3.4 Human *MAOA-lacZ* Revealed Staining in TH-Positive Neurons in the Locus Coeruleus in Adult Brain as well as Horizontal, and Ganglion Cells in Adult Retina

MAOA expression was predicted to be observed in the locus coeruleus (LC) in the adult brain. Whole mount *lacZ* E12.5 stained embryos revealed expression in the prepontine hindbrain that extended to the basal midbrain and the prosomeres 1 and 2 of the diencephalon (Figure 2.3A). Staining extended from the pontine hindbrain (pons proper) to the medullary hindbrain (medulla) (Figure 2.3A). E12.5 cleared embryos confirmed staining in the previously described regions, and suggested expression of the *MAOA* reporter construct in major neuron fibers extending from the prepontine hindbrain, and the pontine hindbrain (Figure 2.3B). Staining was present in fibers of the medullary hindbrain that extended into the thoracic cavity, and the ventral region of the somites, starting in the upper thoracic cavity that extended towards the posterior limbs (Figure 2.3B). P7 developing mouse brains sections revealed expression at the lateral extremities of the fourth ventricle, demonstrating expression of the *MAOA* reporter construct in the region of the LC at this developmental stage (Figure 2.3C). Adult mouse brain sections staining revealed high level of expression in

the LC, and lateral cerebellar nuclei (Figure 2.3D). Staining was present in the medial and lateral vestibular nuclei (Figure 2.3D). Further co-localization experiments on cryosections revealed that the *MAOA-lacZ* reporter gene (blue) was expressed in mature neurons, positive for NeuN (brown), in the LC and medial parabrachial nucleus (MBP) (Figure 2.3E, higher magnification on the LC, Figure 2.3F). The β -gal staining product (blue) found in the LC co-localized with TH positive cells (brown), demonstrating expression of the *MAOA* reporter construct in TH positive neurons populating the LC (Figure 2.3G, Figure 2.3H).

Interestingly, the β -gal positive cells (blue) in the LC did not co-localize with GFAP (brown), a marker of astrocytes (data not shown). Furthermore, *MAOA-lacZ* expression pattern characterization revealed staining in several layers in the adult retina such as the ILM, the outer plexiform layer (OPL), and the outer limiting membrane (OLM) (Figure 2.3I).

Calbindin is a marker for three different cell types in the retina; the ganglion cells that populate the ganglion cell layer (GCL), the amacrine cells found in the INL, and the horizontal cells populating the OPL (Haverkamp and Wassle 2000). Co-localization using an anti-Calbindin antibody revealed that our *MAOA-lacZ* construct was expressed in both the horizontal cells populating the OPL and the ganglion cells found in the GCL (Figure 2.3J, Figure 2.3K). Additional staining was found along the spinal cord in the *MAOA-lacZ* animals in gross dissections (data not shown).

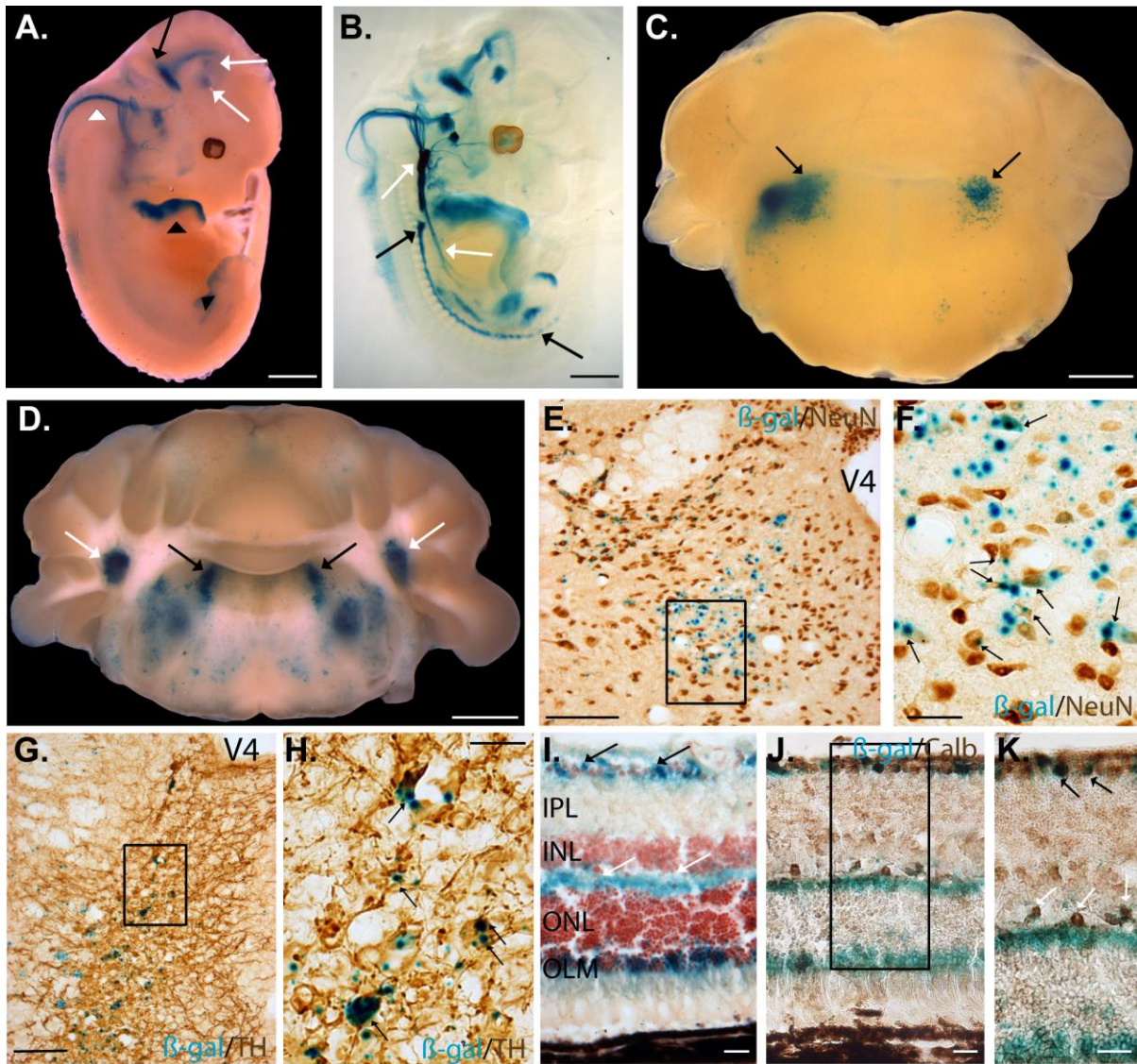


Figure 2.3 Human *MAOA-lacZ* Expressed in TH-Positive Neurons in the Locus Coeruleus in the Adult Brain, Horizontal, and Ganglion Cells in the Retina, and Several Regions of the Developing Hindbrain.

Expression analysis of the human *MAOA-lacZ* strain was examined using β -galactosidase (β -gal) histochemistry (blue). **A.** E12.5 whole embryos stained in the preoptine hindbrain (black arrow) that extended to the basal midbrain and the prosomeres 1 and 2 of the diencephalon (white arrows). Staining extended from the pontine hindbrain (pons proper) to the medullary hindbrain (medulla) (white arrowhead). Staining was noticeable in the anterior part of the developing limbs (black arrowhead). **B.** E12.5 cleared embryos additionally demonstrated staining in fibers of the preoptine hindbrain, and the pontine hindbrain. Staining was present in fibers of the medullary hindbrain that extended in the thoracic cavity (white arrows), and in the ventral region of the somites, starting in the upper thoracic cavity and extending towards the posterior limbs (black arrows). **C.** P7 brains showed staining in nuclei surrounding the fourth ventricle, a region where the locus coeruleus is located (LC) (black arrows). **D.** Adult brains stained in the LC (black arrows), and lateral cerebellar nuclei (white arrows). Staining was present in the medial vestibular

nuclei, and lateral vestibular nuclei. **E.** Co-localization experiment using β -gal staining and anti-NeuN immunocytochemistry (brown) performed on adult brain cryosections suggested expression of *MAOA-lacZ* in mature neurons in the LC. Sparse staining was found in neurons populating the medial parabrachial nucleus. Boxed region in **E** is shown in **F**. **F.** Higher magnification revealed expression of B-gal in neurons in the LC (black arrows). **G.** Co-localization experiment using β -gal staining and an anti-Tyrosine Hydroxylase (anti-TH) antibody (brown) performed on adult brain cryosections suggested expression of *MAOA-lacZ* in TH-positive neurons populating the LC. Boxed region in **G** is shown in **H**. **H.** Higher magnification revealed expression of B-gal in TH-positive neurons in the LC (black arrows). **I.** Staining was found in the retina, extending from the inner limiting membrane (ILM) (black arrows) to the outer plexiform layer (OPL) (white arrows), and the outer limiting membrane (OLM). **J, K.** Co-localization experiment using β -gal staining and an anti-Calbindin (anti-Calb) antibody (brown) performed on adult eye cryosections suggested expression of *AMOTL1-lacZ* in horizontal cells populating the OPL (white arrows), and ganglion cells populating the ganglion cell layer (GCL) (black arrows). Boxed region in **J** is shown in **K**. V4: fourth ventricle, IPL: inner plexiform layer, INL: inner nuclear layer, ONL: outer nuclear layer. (Scale bar 1 mm [**A, B, C, D**], 100 μ m [**E, G**], 25 μ m [**F, H**], 20 μ m [**I, J, K**]).

2.3.5 Human *NOV-lacZ* Revealed Staining in Neurons Populating the Hippocampal Formation, Basomedial Amygdaloid Nuclei, and Cortical Layers in the Adult Brain.

NOV expression was predicted to be observed in the amygdala and basolateral complex in the adult brain, using our approach. Whole mount *lacZ* E12.5 stained and cleared embryos, revealed expression throughout the olfactory epithelium and the primitive nasopharynx (Figure 2.4A, Figure 2.4B). P7 developing mouse brains sections revealed strong expression of the *NOV* reporter construct in the retromammillary nucleus (RM) and light expression in the hippocampal formation (HPF) (Figure 2.4C). Adult mouse brain sections revealed high expression in the mid cortical layers, and the HPF (Figure 2.4D). Pronounced staining was seen in the posterior basomedial amygdaloid nuclei (BMP), and the adult RM (Figure 2.4D, Figure 2.4E). We investigated the nature of the positive cells expressed in the upper cortical layers, HPF, and BMP using co-localization experiments on cryosections. Positive co-localization between the β -gal staining product (blue) and NeuN (brown) revealed that the *NOV* reporter construct was expressed in mature neurons in the

cortical layers II, IV and VI, as well as the HPF, and BMP (Figure 2.4F, Figure 2.4G, Figure 2.4H). The staining pattern in the HPF suggested expression in pyramidal neurons, with staining of the cell body in the CA1 that extended in the apical dendrites found in the stratum radiatum (SR) (Figure 2.4G). A similar experiment, using GFAP as a marker revealed absence of co-localization of the *NOV-lacZ* reporter construct in astrocytes populating the cortex, HPF, and BMP (data not shown).

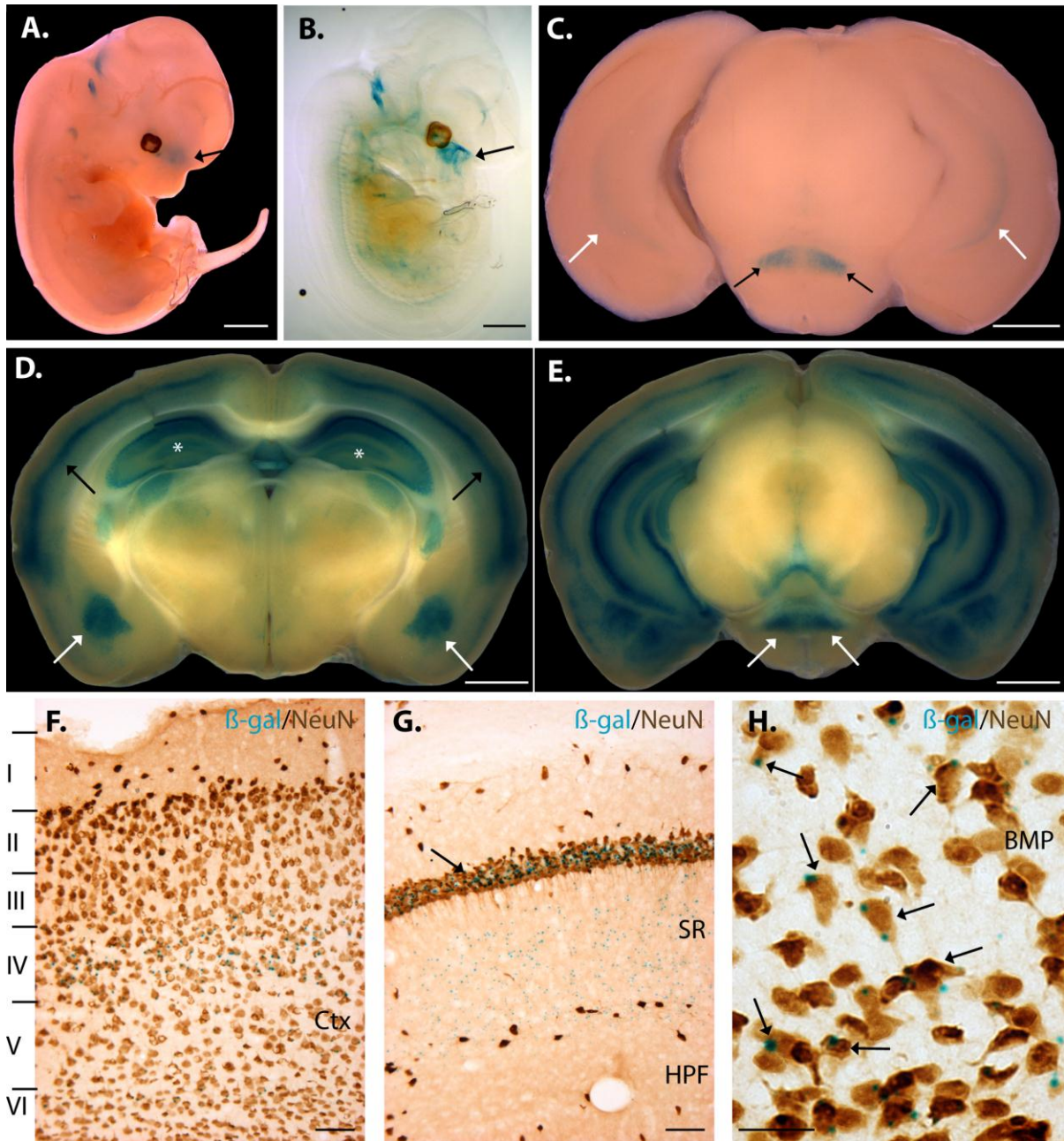


Figure 2.4 Human *NOV-lacZ* Expressed in the Basomedial Amygdaloid Nuclei, mid Cortical Layers, and Pyramidal Neurons of the Hippocampal Formation.

Expression analysis of the human *NOV-lacZ* strain was undertaken by examination of β -galactosidase (β -gal) histochemistry (blue). **A, B.** E12.5 whole, and cleared embryos showed staining in the olfactory epithelium, and the primitive nasopharynx (black arrows). **C.** P7 brains stained in the developing retromammillary nucleus (RM) (black arrows), and hippocampal formation (HPF) (white arrows). **D.** Adult brains stained in the mid cortical layers (black arrows), and the HPF (white asterisks). Staining was found in the basomedial amygdaloid nuclei (BMP) (white arrows). **E.** *NOV-lacZ* staining was present in the adult RM (white arrows). **F.** Co-localization experiment using β -gal staining and a NeuN antibody (brown) performed on adult brain cryosections revealed expression of *NOV-lacZ* in mature

neurons in the cortical layers II, IV and VI. **G.** Co-localization experiment revealed expression of *NOV-lacZ* in pyramidal neurons populating the hippocampal formation (HPF) (NeuN, brown). Staining was found in the cell body in the cornu ammonis 1 region (CA1) (black arrow) and extended in the projections found in the stratum radiatum (SR). **H.** Co-localization experiment revealed expression of *NOV-lacZ* in mature neurons (NeuN, brown) found in the BMP. Ctx: cortex (Scale bar, 1 mm [A, B, C, D, E], 50 μ m [F, G], 25 μ m H).

2.3.6 Human *NR2F2-lacZ* Revealed Staining in Mature Neurons, Immunoreactive for the Nr2f2 Mouse Protein in the Basolateral, and Corticolateral Amygdaloid Nuclei of the Adult Brain.

NR2F2 expression was predicted to be in the amygdala and basolateral complex in the adult brain. Whole mount *lacZ* E12.5 stained embryos revealed strong expression in the rostral secondary prosencephalon that extended throughout all three prosomeric regions of the diencephalon (Figure 2.5A). Staining was present in the nasal cavity, the vestibulochochlear ganglion, and mesenchyme of the posterior limbs (Figure 2.5A). E12.5 cleared embryos confirmed staining in the previously described regions, and suggested expression of the *NR2F2* reporter construct in the developing bladder (Figure 2.5B). P7 developing mouse brains sections revealed expression of the *NR2F2* reporter construct in the amygdala, and subthalamic nuclei (Figure 2.5C). Adult mouse brain sections revealed expression in brain regions including the posterior basolateral amygdaloid nuclei (BLP), the basomedial amygdaloid nuclei (BMP), the posterolateral cortical amygdaloid nuclei (PLCo), the posteromedial cortical amygdaloid nuclei (PMCo), and the posteroventral part of the amygdaloid nuclei (MePV) (Figure 2.5D). Strong staining was found in various thalamic nuclei (Figure 2.5D). Co-localization experiments on cryosections revealed that the *NR2F2-lacZ* reporter gene (blue) was expressed in mature neurons, positive for NeuN (brown), in the BLP, and, BMP (Figure 2.5E). Co-localization was found in the PLCo, PMCo, and MePV

(Figure 2.5E). Lower level of *NR2F2-lacZ* expression was found in mature neurons in the anterolateral amygdalohippocampal area (AHial), and posteromedial amygdalohippocampal area (AHiPM) (Figure 2.5E, higher magnification, Figure 2.5F). The β -gal positive cells (blue) did not co-localize with GFAP (brown), a marker of astrocytes (data not shown). Additional co-localization experiment, performed on adult brain cryosections revealed strong β -gal labelling in cells expressing the *Nr2f2* mouse gene in brain regions extending from the PMCo to the MePV (Figure 2.5G, Figure 2.5H). Lower levels of β -gal were found in the AHiAL (Figure 2.5G, Figure 2.5H).

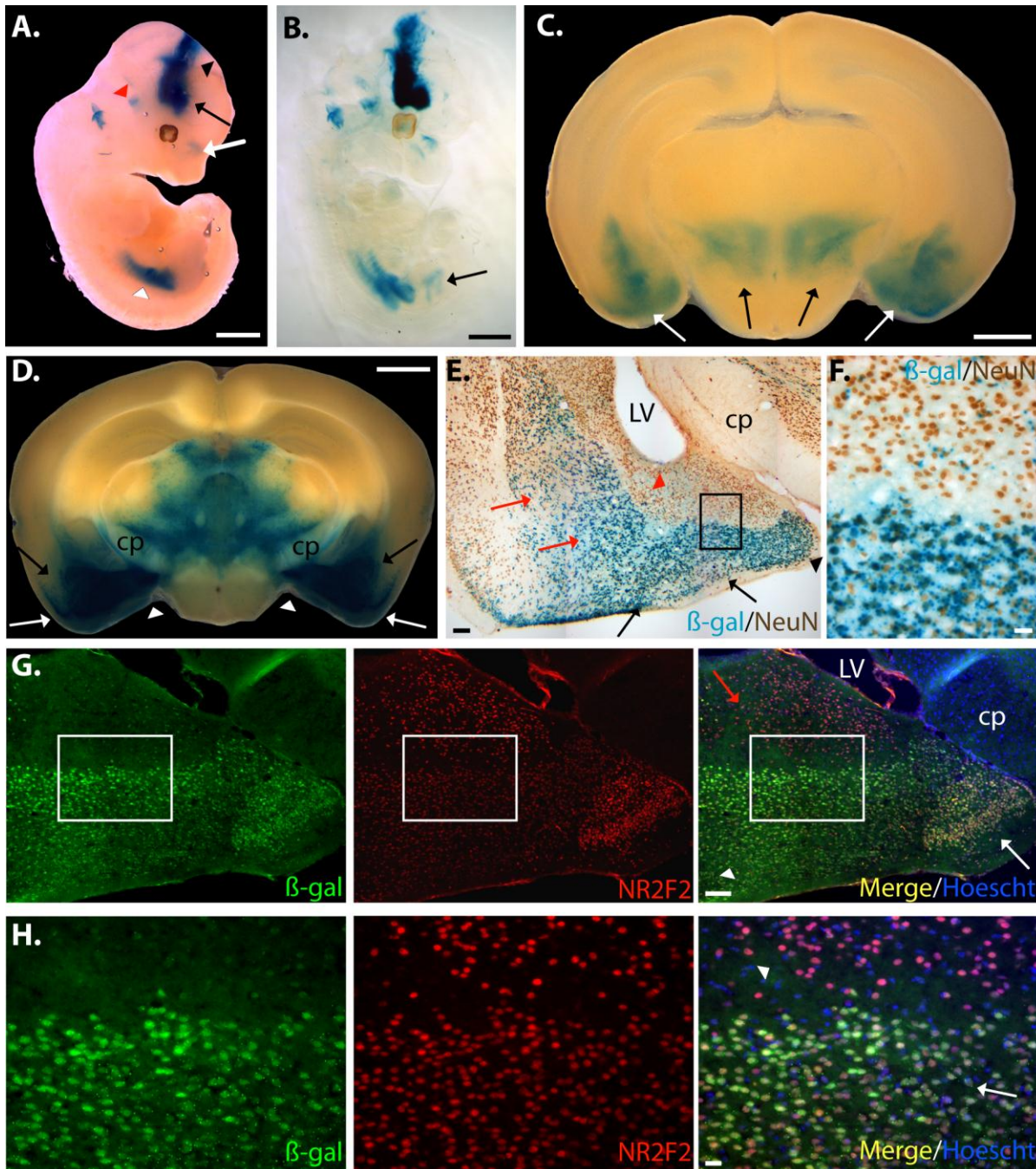


Figure 2.5 Human *NR2F2-lacZ* Expressed in Mature Neurons Populating the Basolateral and Corticolateral Amygdaloid Nuclei that are Immunoreactive for the Nr2f2 Mouse Protein.

Expression analysis of the human *NR2F2-lacZ* strain was undertaken by examination of β -galactosidase (β -gal) histochemistry (blue). **A.** E12.5 whole embryos revealed staining in the rostral secondary prosencephalon (black arrow) that extended throughout all three prosomeric regions of the diencephalon (black arrowhead). Staining was present in the nasal cavity (white arrow), the vestibulochochlear ganglion (red arrowhead) and mesenchyme of the posterior limbs (white arrowhead). **B.** E12.5 cleared embryos additionally demonstrated staining in the developing bladder (black arrow). **C.** P7 brains stained in the amygdala nuclei

(white arrows), and the subthalamic nuclei (black arrows). **D.** Adult brains revealed strong staining extending from the posterior basolateral amygdaloid nuclei (BLP) (black arrows) to the posterolateral cortical amygdaloid nuclei (PLCo) (white arrows), and the posteroventral part of the amygdaloid nuclei (MePV) (white arrowheads). Broad staining was found in the ventral thalamic area, excluding the cerebral peduncle (cp). **E.** Co-localization experiment using β -gal staining and a NeuN antibody (brown) performed on adult brain cryosections revealed strong expression of *NR2F2-lacZ* in mature neurons populating the BLP, and the basomedial amygdaloid nuclei (BMP) (red arrows). Co-localization was found in the PLCo, and the posteromedial cortical amygdaloid nuclei (PMCo) (black arrows), and the MePV (black arrowhead). Lower level of β -gal staining was found in mature neurons in the anterolateral amygdalohippocampal area (AHial) (red arrowhead). Boxed region in **E** is shown in **F**. **F.** Higher magnification revealed strong expression of β -gal in mature neurons in the PMCo and sparse expression in mature neurons in the AHial. **G.** Co-localization experiment, using an anti β -gal antibody (green), and a NR2F2 antibody (red), performed on adult brain cryosections revealed strong β -gal labelling in cells expressing the *Nr2f2* mouse gene in brain regions extending from the PMCo (white arrowhead) to the MePV (white arrow). Lower levels of β -gal were found in the (AHiAL) (red arrow). Boxed region in **G** is shown in **H**. **H.** Higher magnification revealed strong expression of β -gal (green) in *Nr2f2* positive cells (red) in the (PMCo) (white arrow) and lower expression in the AHiAL (white arrowhead). LV: lateral ventricle, cp: cerebral peduncle. (Scale bar, 1 mm [A, B, C, D], 100 μ m [E, G], 20 μ m [F, H]).

2.3.7 Comparative Genomic Analysis Demonstrated the Boundary for Brain Specific Expression of the Subset of Candidate Human Genes Chosen.

To understand the relevance of the expression pattern results obtained from our four different humanized mouse models, we performed comparative sequence alignment of the BAC constructs used in this study against multiple genomes using the University of California, Santa Cruz (UCSC) genome browser (<http://genome.ucsc.edu/>). We compared our expression results against available data from the literature, and public databases such as the ABA, BGEM, and *EGFP* reporter mouse models generated throughout GENSAT (Gong, Zheng et al. 2003; Siegert, Scherf et al. 2009).

The *AMOTL1-lacZ* animals showed staining in the thalamic nuclei at P7 which correlated with available expression data from the ABA at similar developing time-points (P4 and P14). Absence of expression in the developing HPF was noticeable for both time-points

and suggested a discrepancy between the expression results coming from our human construct and the ones from the ABA. The expression results obtained from the ABA in the adult brain for the *Amotl1* gene correlated with the results obtained with our human construct, with expression in the thalamic nuclei, and CA1 and CA3 regions. This suggests that the discrepancy between the expression results from our human construct and the ones from the ABA could be due to missing key regulatory regions, important for silencing expression of the *AMOTL1* human gene in the developing HPF (Figure 2.6A). This suggested that the regulatory regions allowing expression of the *AMOTL1* gene in the anteromedial, and anteroventral thalamic nuclei as well as the CA1 and CA3 regions in the adult brain are conserved from human to mouse and were included in our construct

The *MAOA-lacZ* animals showed a staining pattern suggesting expression in the developing LC at P7 which correlated with available expression data from the ABA (P4 and P14) and BGEM (P7) at similar developing time-points. Staining of adult brain sections in our mouse model revealed expression of *MAOA* in the LC, results which were again correlating with the ABA. Co-localization experiments revealed that the *MAOA-lacZ* reporter gene was expressed in TH positive neurons in the LC which correlated with previously observed results for the mouse gene (Vitalis, Fouquet et al. 2002). Additional staining was found in the medial, and lateral vestibular nuclei. This could partly be due to the fact that important 3' regulatory regions, acting as silencers to narrow down the expression pattern of the gene, were missing from the construct used in our current mouse model (Figure 2.6B, red rectangle box). These regions are conserved from human to mouse, and are rich in DNases clusters and transcription factor binding sites according to ENCODE, suggesting that they have regulatory properties. Nevertheless, the expression results in TH-positive neurons in the

LC in the adult brain suggested that the regulatory regions allowing expression of the *MAOA* gene in these cells are conserved from human to mouse, and were included in our construct (Figure 2.6B).

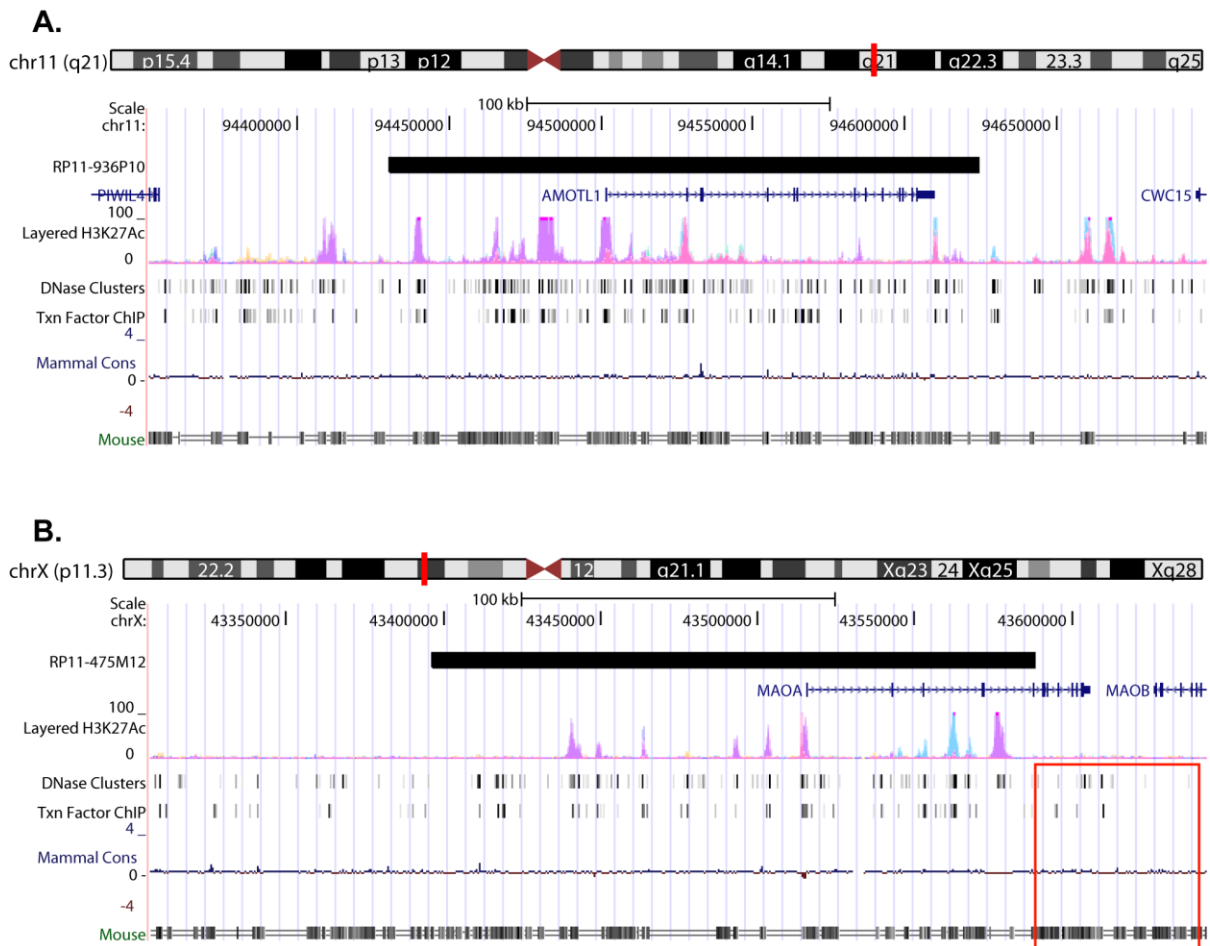


Figure 2.6 Adult Brain *lacZ* Staining Results and Comparative Genomic Allowed Prediction of Conserved Regulatory Regions for *AMOTL1* and *MAOA*.

Coordinates corresponding to the human bacterial artificial chromosome (BAC) constructs used in this manuscript were retrieved and visualized using the UCSC genome browser. **A.** The alignment results from the *AMOTL1* BAC (RP11-936P10) suggested that the construct contained all the elements allowing proper expression of this human gene in the anterior thalamic nuclei in both developing (P7) and adult mouse brain. The results suggested that additional conserved regulatory elements, outside of the current genomic DNA fragment used are necessary for proper silencing in the developing hippocampal formation (HPF) at P7. **B.** The alignment results from the *MAOA* BAC (RP11-475M12) suggested that the construct contained all the elements allowing proper expression of this human gene in the locus coeruleus (LC) in both developing (P7) and adult mouse brain. The results suggested

that additional conserved regulatory elements in the 3' regions (red rectangle) could be essential in narrowing the brain specific expression pattern of this human gene.

The *NOV-lacZ* animals showed staining in the developing RM, and light staining in the HPF at P7 which partly correlated with available expression data from the ABA (P4 and P14), BGEM (P7), and GENSAT (P7). The *NOV-lacZ* animals at P7 displayed absence of expression in developing cortical layers, demonstrating a stark difference with results obtained at similar time-points for the ABA, BGEM, and GENSAT. Expression results obtained in the adult brain using our *NOV-lacZ* animals demonstrated strong expression in mid cortical layers, the HPF, and BMP, results that correlated with the ones obtained for the ABA, BGEM, and GENSAT datasets. Additional analysis revealed staining of the *NOV-lacZ* reporter gene in layers II, IV and VI of the cortex, results correlated with expression data coming from both the ABA, and GENSAT. Expression in the HPF was found in CA1 pyramidal neurons in our *NOV* mouse model. These results correlated with the mouse model generated by GENSAT, suggesting that the regulatory regions allowing expression of the *NOV* gene in the adult cortical layers, HPF and BMP are conserved from human to mouse, and were included in our construct (Figure 2.7A). Additionally, sequence alignment comparison between the BAC construct used in our study (RP11-840I14) and the one used by GENSAT (RP23-235B13) suggested that additional regulatory regions, either in the 5' and/or the 3' end were missing resulting in a delay in temporal expression of our reporter gene in the cortical layers during development. Presumably, these regions would be acting from a distant genomic location.

The *NR2F2-lacZ* animals showed staining in the developing amygdala, and subthalamic nuclei at P7, which partly correlated with available expression data from BGEM

and GENSAT at the same time-point. The *NR2F2-lacZ* animals at P7 displayed absence of expression in the developing hypothalamus, demonstrating a difference with results obtained at similar time-points in BGEM, and GENSAT. Expression results obtained in the adult brain using our *NR2F2-lacZ* animals demonstrated expression in the anterior hypothalamic nuclei and the basomedial amygdalar nuclei, results that correlated with the ones obtained for the ABA, and GENSAT. Co-localization experiments performed using a β -gal, and NR2F2 antibody revealed cell specific expression of the *NR2F2-lacZ* reporter construct in cells expressing the *Nr2f2* endogenous mouse gene of various amygdala regions. Our results demonstrated that similar expression levels were obtained when comparing the β -gal staining to the endogenous *Nr2f2* mouse gene in cells expressed in regions extending from the PMCo to the MePV in the amygdala. Interestingly, our *NR2F2-lacZ* reporter strain revealed lower level of expression in the AHiAL, an amygdala region that has shown robust level of expression of the endogenous *Nr2f2* mouse gene. The discrepancy between the levels of expression resulting from our reporter gene when compared to the endogenous mouse gene could result from the absence of distal enhancer regulatory regions important in regulating the expression level in the AHiAL. Additionally, expression pattern characterization using the same NR2F2 antibody revealed that the *Nr2f2* mouse gene is expressed in amacrine cells in the retina (Inoue, Iida et al. 2010). Expression results using our *NR2F2-lacZ* reporter strain demonstrated absence of expression in cells populating the retina (data not shown). The GENSAT project has reported expression of their *Nr2f2-EGFP* reporter strain in amacrine cells of the retina (RP23-109L9) (alignment results Figure 2.7B) (Siegert, Scherf et al. 2009). This suggests a discrepancy between the expression results obtained using our reporter mouse model and the expression from both the endogenous *Nr2f2* mouse gene, and the

mouse model generated by GENSAT. An investigation looking at the sequence alignment on the *NR2F2* human gene revealed an overlap between our *lacZ* reporter cassette retrofitted in the BAC construct and the epitopes coordinates recognized by the NR2F2 antibody used in previous studies (Figure 2.7C) (Kanatani, Yozu et al. 2008; Inoue, Iida et al. 2010). This suggests that the absence of expression of our construct in the amacrine cells of the retina is not attributable to the recognition of a different isoform of the gene by the antibody, but rather to an absence of conserved regulatory regions. Presumably, these conserved regulatory regions would be present in the additional 5'distal DNA portion of the BAC construct used in the GENSAT project (Figure 2.7B).

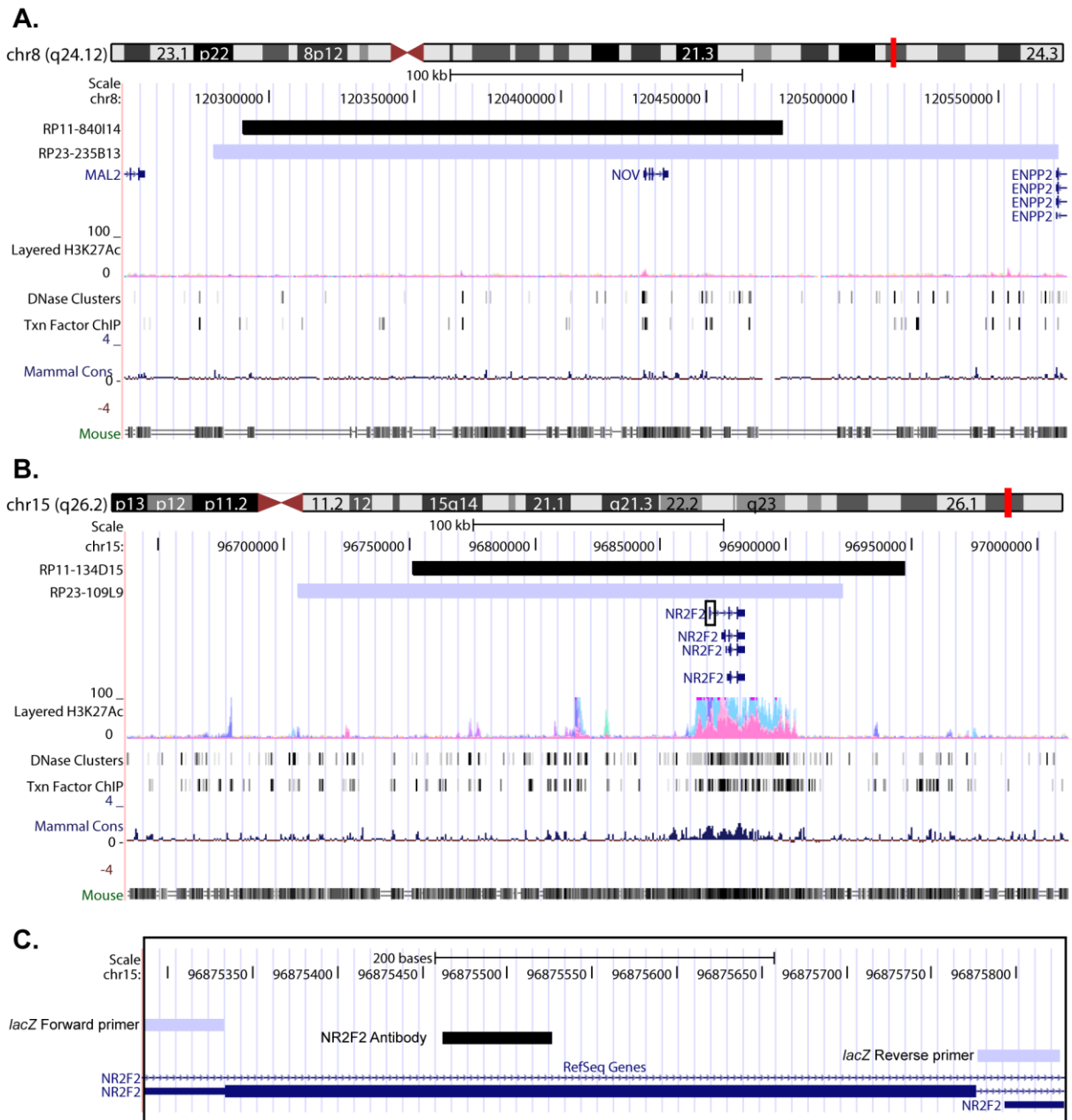


Figure 2.7 Adult Brain *lacZ* Staining Results and Comparative Genomic Allowed Prediction of Conserved Regulatory Regions for *NOV* and *NR2F2*.

Coordinates corresponding to the human bacterial artificial chromosome (BAC) constructs used in this manuscript were retrieved and visualized using the UCSC genome browser. **A.** The alignment results from the *NOV* BAC (RP11-840I14) suggested that the construct contained all the elements allowing proper expression of this human gene in the basomedial amygdaloid nuclei (BMP), cortical layers and pyramidal neurons in the cornu amonis 1 (CA1) regions in the adult brain. The results suggested that additional conserved regulatory elements in the 3' regions, present in the RP23-235B13 BAC construct used in the GENSAT mouse model, are necessary for proper expression in the developing cortical layers at P7. **B.** The alignment results from the *NR2F2* BAC (RP11-134D15) suggested that the BAC construct used in this study contained all the proper elements allowing region specific

expression of this human gene in the basolateral, and corticolateral amygdaloid nuclei in the adult brain. Sequence alignment comparison demonstrated that the RP11-134D15 BAC construct was shorter at the 5' end and longer at the 3' end than the BAC construct RP23-109L9 used in the GENSAT mouse model. This slight difference appears to have only minor consequences when comparing expression pattern from both constructs. Black rectangle box in **B** is shown in **C**. **C**. Sequence alignment comparison analysis using the coordinates corresponding to the primers used in the BAC *lacZ* retrofitting process (grey bars) and the cDNA sequence (black bar) used to generate an anti-NR2F2 antibody suggested that the absence of expression of the *NR2F2-lacZ* constructs in amacrine cells in the retina was not attributable to the recognition of different isoform of *NR2F2*.

2.4 Discussion

Here, we describe an approach looking at refining our understanding in gene expression regulation. We first generated a list of human genes predicted to be enriched in brain regions of therapeutic interest and tested the veracity of these prediction using novel knock-in reporter mouse models. This approach using the pipeline suitable HuGX method was scalable to higher-throughput (Schmuth, Bonaguro et al. 2012). Out of the nine recovered BAC DNA constructs, eight were fully modified to contain the *Hprt* homology arms and a reporter gene (*lacZ* or *EGFP*). The success rate of ~89% in the modification step of the BAC DNA construct, demonstrates the efficiency of the approach.

Furthermore, the success rate of correctly targeted ESC clones varied between 20 and 50%, with an average of 35%. The relatively small amount of clones to be picked and the high rate of correctly targeted clones made it a very efficient method, easily suitable for the current project. Germline offspring were evaluated for positive expression in E12.5 embryos, P7 developing brain, and adult brain, and eyes. Based on these criteria, five out of eight constructs were considered positive for expression, and four of the five positive mouse strains, including *AMOTL1-lacZ*, *MAOA-lacZ*, *NOV-lacZ*, and *NR2F2-lacZ* were characterized in this manuscript. Careful analysis of the expression pattern from each of these

reporter genes demonstrated slight variations, either during development or in specific adult tissue, with publicly available mouse data. These variations could be either attributable to species specific differences or missing regulatory regions in the constructs used to generate the humanized reporter mouse models in this study. Another possibility lies on the effects resulting from the insertion site chosen on the X chromosome. The *Hprt* locus has been considered a neutral insertion site based on the expected restricted expression results coming from tissue-specific promoter insertion at the locus (Cvetkovic, Yang et al. 2000; Evans, Hatzopoulos et al. 2000; Guillot, Liu et al. 2000; Minami, Donovan et al. 2002). It was suggested that the introduction of larger DNA construct (i.e. BAC DNA) could minimise the risk of influences from the *Hprt* insertion site by providing the essential chromatin environment, allowing proper gene regulation (Heaney, Rettew et al. 2004). In our case, one possibility remains; that the slight variation observed in our results could be attributable to differential epigenetic markers, specific to the *Hprt* locus, and resulting in different chromatin conformation from that of the endogenous gene location. Additional analysis, looking at different epigenetic markers along the BAC sequences used and comparing them to both the endogenous mouse gene and human gene could allow us to investigate this possibility. This would be of particular interest in understanding the negative results obtained for both *PITX2* and *NGFR* constructs. Nevertheless, the results obtained using our approach demonstrated that the expression results for every human gene matched the predicted specific adult brain region for which they were chosen. This shows that a careful investigation, using both elements from publicly available resources and bioinformatics can lead to accurate prediction of gene expression.

This is important as the future of gene therapy may rely upon the development of small human promoters to finely regulate the expression of therapeutic genes in a cell specific manner. In that sense, the work done in this project was looking at identifying the minimum boundaries for expression in brain regions of therapeutic interest, a major goal achieved for all four genes described in this manuscript. In the near future, refined mouse models using subsets of the regulatory regions defined within these boundaries could lead to the generation of MiniPromoters, driving the expression of a gene of therapeutic interest specifically either in the thalamus, locus coeruleus or various amygdala nuclei in the brain.

Chapter 3: Retina Restored and Brain Abnormalities Ameliorated by Single-Copy Knock-in of Human *NR2E1* in Null Mice

3.1 Introduction

During embryogenesis, the proper development of an organism relies on orchestration between proliferation, differentiation, and death of different cell populations. The resulting dynamic balance depends on both cell intrinsic regulators and environmental factors. One such intrinsic regulator is *Nr2e1* (also known as *Tlx*, *Tll*, and *Tailless*), which encodes a highly conserved transcription factor known to be a key stem-cell-fate determinant in both the developing mouse forebrain and retina (Pignoni, Baldarelli et al. 1990; Yu, McKeown et al. 1994; Jackson, Panayiotidis et al. 1998; Miyawaki, Uemura et al. 2004; Li, Sun et al. 2008). In mouse, abnormal regulation of this gene results in blindness, behaviour abnormalities, and brain tumour initiation and progression (Yu, Chiang et al. 2000; Roy, Thiels et al. 2002; Young, Berry et al. 2002; Liu, Wang et al. 2010; Park, Kim et al. 2010). In humans, candidate regulatory mutations have been found in patients suffering from microcephaly, bipolar disorder, schizophrenia and aggression (Kumar, Leach et al. 2007; Kumar, McGhee et al. 2008). Upregulation of *NR2E1* expression was also found in cancer and a somatic-protein-coding mutation was found in glioblastoma (Phillips, Kharbanda et al. 2006; Sim, Keyoung et al. 2006; Parsons, Jones et al. 2008; Liu, Wang et al. 2010; Park, Kim et al. 2010). These data argue in favour of strengthening our knowledge regarding the regulation and function of human *NR2E1* and thereby the potential for this gene to have a role in disease. To date, the main information regarding human *NR2E1* expression comes from assembled analyses done on homogenized tissue samples and information from a humanized mouse model (Jackson, Panayiotidis et al. 1998; Abrahams, Kwok et al. 2005;

Kumar, McGhee et al. 2008). This model demonstrated complete correction of the null-brain phenotype while only ameliorating the eye phenotype. The uncorrected eye phenotype was proposed to be due to gene dosage sensitivity during eye development; a hypothesis that has not been verified to date.

During development, the mammalian forebrain can be divided into two regions, the telencephalon and diencephalon. The telencephalon will become the cerebrum, comprising the cerebral cortex and basal ganglia, and the diencephalon will give rise to the brain thalamus regions and optic vesicles (OV); the latter will mature into the optic nerve and eye. The mouse *Nr2e1* gene is first detected at embryonic day 8 (E8) in the ventricular zone (VZ) of the neuroepithelium layer and later spreads posteriorly into the diencephalon (E10.5) (Yu, McKeown et al. 1994; Monaghan, Grau et al. 1995). At E12.5, the expression is detected in the VZ region of the lateral telencephalon; comprising the dorsal and lateral pallium (DP, LP), the medial pallium (MP), the lateral ganglionic eminence (LGE), and the medial ganglionic eminence (MGE) (Stenman, Yu et al. 2003). Mice lacking the *Nr2e1* gene (*Tlx*^{-/-}, *Nr2e1*^{frc/frc} a.k.a. fierce, referred here as *Nr2e1*-null) show forebrain defects, leading to cerebrum and olfactory bulb (OB) hypoplasia (Monaghan, Bock et al. 1997; Young, Berry et al. 2002). The cerebrum and OB defects are mainly attributed to the role of *Nr2e1* in controlling the proliferation rate of neural stem cells in the VZ of the telencephalon during development (Stenman, Wang et al. 2003; Roy, Kuznicki et al. 2004; Li, Sun et al. 2008). The role of *Nr2e1* in maintaining the neural stem cell population also appears to be important in the adult brain. Expression studies using *lacZ* reporter mice demonstrate a positive signal for the *Nr2e1* gene in proliferative cells lining the subventricular zone (SVZ), rostral migratory stream (RMS), and the subgranular layer of the dentate gyrus (DG) (Shi, Chichung

Lie et al. 2004; Li, Sun et al. 2008; Liu, Belz et al. 2008). Functional studies in the adult brain, using conditional knock-out and transgenic mouse models, revealed that *Nr2e1* is involved in controlling the number of proliferative cells in both the SVZ and DG (Liu, Belz et al. 2008; Zhang, Zou et al. 2008). This control in the SVZ appears to be highly dependent on the copy number found in the genome (Liu, Wang et al. 2010). Surprisingly, non-proliferative cells populating the Cornu Ammonis regions (CA1, CA3) and DG regions of the hippocampus, as well as the striatum and cortex, have also been shown to express sparse to strong levels of *Nr2e1* in the adult brain (Zhang, Zou et al. 2008).

In the developing mouse eye, *Nr2e1* is detected in the optic processes of the developing embryo as early as E9 (Monaghan, Grau et al. 1995). At E11.5, the expression becomes restricted to the innermost surface of the retina, corresponding to the end feet of retinal progenitor cells (RPC) found in the neuroblastic layer (NBL) of the developing embryo (Miyawaki, Uemura et al. 2004). *Nr2e1* expression in the NBL peaks at E15.5, suggesting a role for this gene in early phase of retinogenesis during embryonic development (Miyawaki, Uemura et al. 2004; Zhang, Zou et al. 2006) *Nr2e1*-null mice have retinal and optic nerve dystrophy leading to blindness (Yu, Chiang et al. 2000; Young, Berry et al. 2002; Zhang, Zou et al. 2006). During development, the null animals display a deregulation in the proliferation rate of RPC and an increase in apoptotic levels in the ganglion cell layer (GCL), which results in a marked reduction in thickness of the distinct layers in the adult retina (Young, Berry et al. 2002; Miyawaki, Uemura et al. 2004). *Nr2e1*-null animals also suffer from retinal vasculature defects which are explained by the role of this gene in the proper assembly of fibronectin matrices secreted by proangiogenic astrocytes (Uemura, Kusuhara et al. 2006). Absence of *Nr2e1* expression in these cells during development has been attributed

to defects in normal vasculature formation in the null mice. In the adult retina, *Nr2e1* expression is restricted to the Müller cells, a retinal glia subtype population that is immunoreactive for Cellular Retinaldehyde-Binding Protein (CRALBP) (Miyawaki, Uemura et al. 2004; Zhang, Zou et al. 2006).

With the advancement of technologies allowing easy modification of large DNA constructs, and the use of novel docking sites to transfer these large constructs to the mouse genome, it is now possible to generate multiple humanized mouse strains in a relatively short time (Copeland, Jenkins et al. 2001; Heaney, Rettew et al. 2004). In the present study, we sought to investigate the functionality of the human gene in both brain and eye, using a novel humanized mouse strain carrying a single-copy BAC insert of *NR2E1* knocked in at the *Hprt* locus. The high level of sequence similarity at the *NR2E1* locus between the mouse and human, and the fact that mice carrying a single copy of the mouse *Nr2e1* gene (*Nr2e1^{frc/+}*) have virtually no phenotype, were elements justifying the choice of our approach (Monaghan, Bock et al. 1997; Jackson, Panayiotidis et al. 1998; Abrahams, Mak et al. 2002; Young, Berry et al. 2002; Abrahams, Kwok et al. 2005). Thus, we tested the hypothesis that a single-copy human gene is functionally equivalent to a single copy mouse gene and will correct both the brain and eye phenotypes of the *Nr2e1*-null mice. We also documented, for the first time, the expression pattern of the human gene using an *NR2E1-lacZ* reporter strain generated with the same technology. Finally, employing the site-specific, single-copy, docking technology allowed us to develop a platform for future testing of candidate human mutations in *NR2E1* (Schmouth, Bonaguro et al. 2012).

3.2 Methods and Materials

3.2.1 BAC Retrofitting

Two BAC constructs were generated: first bEMS223, which is a modified version of BAC RP11-144P8 (<http://bacpac.chori.org/hmale11.htm>, accessed June 6, 2011) containing the human *NR2E1* gene and retrofitted to contain *Hprt* homologous recombination targeting arms; and second bEMS86, which was a further retrofitting of bEMS223 to carry a *lacZ* reporter cassette inserted at the ATG of *NR2E1*. Both BAC constructs were generated using the lambda recombination system (Yu, Ellis et al. 2000).

To generate the bEMS223 construct, the *HPRT* region in pJDH8A/246b (Heaney, Rettew et al. 2004) was modified by *NotI* digestion resulting in a 1.3 kb fragment that contained the Ori site. This fragment was extracted and amplified using the following primers:

5'-**AATTGCGGCCGCTGTTCACTGATTCACGCGGTTCAAAAATGACGATCG**
ATGGTATTA^{*ACTCAAACGATA*}*TTTAAATCGCTCTTCCGCTTCCTCG*-3'

5'-**AATTGCGGCCGCTCAGCGTTTTGCAGCGGCCAGCTGTCCCACACATCAA**
GTCTTTTGCAGACTCAAT^{*TTTAAATTGGATGGAGGCGGATAAA*}-3'

These primers added 60 bp targeting arms for integration into the *SacB* gene and *NotI* (bold) and *SwaI* (italics) restriction sites to the Ori linear DNA cassette. The modified Ori cassette replaced the original Ori site in pJDH8A/246b by *NotI* digestion and ligation. This produced the modified pJDH8A/246b construct pEMS1907, which contained a *HPRT* sequence flanked by *SacB* gene and *SwaI* sequence. The *HPRT* cassette was removed from pEMS1907 using *SwaI* and recombineered into the *SacB* gene of the *NR2E1* parental BAC (RP11-144P8), using the lambda recombination system in SW102 cells.

The bEMS86 construct was generated by adding a *lacZ* reporter cassette in the first exon starting at the first codon of *NR2E1* and creating a downstream 25 bp deletion of genomic DNA. A custom synthesized *lacZ*/kanamycin cassette pEMS1908 (GeneArt, Regensburg, Germany) was PCR amplified to contain BAC homology arms using the following primers:

5'-GCCGGGACTCGGGCAGCGCCACCAACCGCTCCGCCCCGGGACAGCCA
GCATGGCGGATCCCGTCGTTTT-3'

5'-TCGCCCCAGGCTGCGCGCCTAGGCCCCACGGCGGCCCGAGAGGTACCC
ACGAAGTTCCTATACTTTCTAG-3'

The resulting fragment was retrofitted into bEMS223 as described above and transformed into SW105 cells. The kanamycin gene adjacent to the *lacZ* gene was designed with flanking full *frt* sites (Lyznik, Mitchell et al. 1993), which were used to excise the kanamycin gene via induction of flpe recombinase (Schlake and Bode 1994; Buchholz, Angrand et al. 1998). The resulting construct bEMS86, contained a *lacZ* reporter gene under the influence of the *NR2E1* human regulatory regions.

3.2.2 Strain Generation, Husbandry and Breeding

Three different strains were generated; B6.129P2(Cg)-*Hprt*^{tm86(NR2E1,bEMS223)Ems} from embryonic stem cell (ESC) clone mEMS2044 harboured the BAC bEMS223, B6.129P2(Cg)-*Hprt*^{tm73(Ple142-lacZ)Ems} from mEMS4751 and B6.129P2(Cg)-*Hprt*^{tm87(Ple142-lacZ)Ems} from mEMS4749 both harboured the BAC bEMS86. These strains were generated using a variation of the previously described strategy to insert constructs 5' of *Hprt* on the mouse X Chromosome (Bronson, Plaehn et al. 1996; Heaney, Rettew et al. 2004; Yang, Banks et al.

2009). Briefly, BAC DNA was purified using the Nucleobond BAC 100 kit (Clontech laboratories, Mountain View, California) and linearized with *I-SceI* as described in the literature (Heaney, Rettew et al. 2004). bEMS223 was electroporated into 4×10^7 mEMS1204 ESCs (Yang, Banks et al. 2009) and bEMS86 similarly into mEMS1202 (Yang, Banks et al. 2009) using the following conditions: voltage; 190 V, capacitance; 500 μ F, resistance; none, using a BTX ECM 630 Electro cell manipulator (BTX, San Diego, California) (Yang, Banks et al. 2009). ESC clones were selected in hypoxanthine aminopterin thymidine (HAT), isolated, and DNA was purified as described previously (Yang, Banks et al. 2009). Note, B6.129P2(Cg)-*Hprt*^{tm73(Ple142-lacZ)Ems} and B6.129P2(Cg)-*Hprt*^{tm87(Ple142-lacZ)Ems} were obtained from two independent ESC clones from the same electroporation but different plates. Table 3.1 lists the PCR assays used for BAC DNA integrity characterization. These human-specific PCR assays were designed to span the entire BAC construct inserted in the mouse genome; they had a maximum of 11 kb between them, with an average of 6 kb apart. The PCR assays were performed as previously described (Yang, Banks et al. 2009). PCR positive ESC clones for bEMS223 were subsequently tested in Southern blot assays to verify the integrity of the insertion site using a 5' *Hprt* probe and RSA probe (Heaney, Rettew et al. 2004). Copy number quantification for bEMS223 ESC clones, which had passed both the PCR and Southern blot screening, was performed on genomic DNA extracted with the Qiagen DNeasy blood and tissue kit (Qiagen Inc., Toronto, Ontario). A custom TaqMan® copy-number assay was designed in an intronic region with identical sequence in mouse and human *NR2E1* genes, and used in accordance to the manufacturer (Life technologies, Carlsbad, California).

Assay Name	Assay Description	Forward Primer Name	Forward Primer Sequence (5' - 3')	Reverse Primer Name	Reverse Primer Sequence (5' - 3')	Tm (°C)
HPRT correction*	HPRT1-CS (CS) allele	oEMS2267	TCAGGCCGAACCTCTCGGCTT	oEMS2269	TGCTGGACATCCCTACTAACCCA	61
5'SACB	In vector, 5'SACB fragment	oEMS2686	GCAAGGACAGCTGACAGTCA	oEMS2689	GATGCAAGTGTGTCGCTGTC	61
Assay 1*	NR2E1 5' distal region	oEMS3446	TGGAAAAGCATTTCCTCCTATTGT	oEMS3447	TGGCCAAGATCACAATAGGTGGTTA	61
Assay 2*	NR2E1 5' distal region	oEMS3448	CTTTAAAGTCCATATTTTCGGCCAGC	oEMS3449	CCACCCGCCAGCTATATTTTGT	61
Assay 3*	NR2E1 5' distal region	oEMS3452	CCAAAGGCGTTTTTGTAAATGGTG	oEMS3453	CCTACCCCAAACAGTTGTCAACTCA	61
Assay 4*	NR2E1 5' distal region	oEMS3454	CAGATAGAGTGCCAAGGCAAAACC	oEMS3455	ATGGATCATTGGCTGGGCC	61
Assay 5*	NR2E1 5' distal region	oEMS3456	GCCACTGTACCAGGCTGAAAAGACA	oEMS3457	GGGCAAGTCCCTTACCTTATTCTT	61
Assay 6*	NR2E1 5' distal region	oEMS3460	TCAGGCAGGTTCACTCATACTTCC	oEMS3461	ATTCACACCCTTGGTTGGACAAAA	61
Exon 1**	NR2E1 first exon	oEMS800	CCCAGCAGCTGCGTTTTTGC	oEMS801	GCAGCGCTCCAGGCAGGAC	58
Exon 2**	NR2E1 second exon	oEMS829	GACTAGGAGGCAGGCCAAC	oEMS810	GGTGACATCGCTCTGCTCTC	58
Exon 3	NR2E1 third exon	oEMS736	GGACTGGCCCTCTGAAGTA	oEMS737	TCCCAGCATCTGAAAGAAG	58
Exon 4	NR2E1 fourth exon	oEMS738.1	CTCCCTCAGATTCCCTCTCC	oEMS804	GGGTGCGTCCCTCTCCATTGC	61
Exon 5	NR2E1 fifth exon	oEMS1972	TGTA AACGACGCGCCAGTTACCCACCAATGTCAACTGC	oEMS1973	CAGAAACAGCTATGACAACCCACAGGAAGAAGCAAG	58
Exon 6	NR2E1 sixth exon	oEMS742	TGGGAAAATAAGGGAAAGCTAGA	oEMS743	ATTTAAATAACAATGCAAGCAGTCA	55
Exon 7	NR2E1 seventh exon	oEMS744	CTTTCATACAATATAGCCGGTTTACA	oEMS745.1	AACATGCAGGTTCCCATAGC	55
Exon 8	NR2E1 eighth exon	oEMS746.1	GATTACAGACACATGCCACCAT	oEMS747	CACCCACCCTGAGAGATAGG	55
Exon 9	NR2E1 ninth exon	oEMS748	GACAACAGTGCCGTGCCAGA	oEMS749	TTCCTGAAGGCTACACATTCC	55
Assay 7*	NR2E1 3' distal region	oEMS3473	CCGGCCACAGCAAGATCGAT	oEMS3474	GGGCCTTCTCAAATGTTGAAAGCC	61
Assay 8*	NR2E1 3' distal region	oEMS3477	CAGGGTAGAATCCCAGACTGGTCTC	oEMS3478	AGCAGGGGAAATTAGCCGGG	61
Assay 9*	NR2E1 3' distal region	oEMS3481	TGTGCTTGGTGGACATCCTAGTTTG	oEMS3482	AACTTGCGGTGGTTTGGGGA	61
Assay 10*	NR2E1 3' distal region	oEMS3485	TGGTAGAACAAATAACCTGCTGCCC	oEMS3486	GCACAAGGAAGGCCTCACTCTA	61
Assay 11*	NR2E1 3' distal region	oEMS3489	GGCAAGAACCAAAAAGTAAGCCACA	oEMS3490	TGGGCCCTGACATAAGCTTTAAGTG	61
Assay 12*	NR2E1 3' distal region	oEMS3493	CCCATCGCTGCACAAAATAATTAA	oEMS3494	AACATTGGCAGCAAACAGTGGG	61
3'SACB	In vector, 3'SACB fragment	oEMS2692	AAGCCTTCGCGAAAGAAAAT	oEMS2695	GACGCTGAGGGGTATGTGAT	61

*5% DMSO added to the PCR reaction

**10% DMSO added to the PCR reaction

Table 3.1 Human-Specific PCR Assays Designed to Check BAC Integrity.

Primers used to characterize the integrity of the bacterial artificial chromosome (BAC) construct inserted at the *Hprt* locus are presented. All assays were human specific.

The custom assay employed the following primers:

5'-AAGCTCTGGAAAGTAGTGTTATGAA-3'

5'-TAATAGGCATCCCAAACACAAA-3'

5'-TGGGAATGCTCTGTGAATGA-3' [Probe]

Copy number evaluation was done using raw Ct value comparisons between the custom assay and two different TaqMan® copy number mouse reference assays; Tfr^c and Tert (Life Technologies, Carlsbad, California). The Q-PCR results from three technical replicates were pooled together and compared to wild-type ESC Ct values. Correctly targeted ESC clones were microinjected into B6(Cg)-Tyr^{c-2}/J (JAX Stock#000058) blastocysts to generate chimeras that were bred to C57BL/6J (B6) (JAX Stock#000664) to obtain offspring carrying the BAC insert. Backcrossing to B6 continued such that mice used in this study were N5 or higher. The Mutant Mouse Regional Resource Center (MMRRC) at The Jackson Laboratory is distributing strain B6.129P2(Cg)-Hprt^{tm73(Ple142-lacZ)Ems}/Mmjax (MMRRC Stock #032962).

Male animals were used in all studies to avoid any variability due to random X-inactivation of the knock-in alleles at *Hprt*. Experimental animals to study the functional *NR2E1* BAC (bEMS223) were generated through a breeding scheme similar to that used previously (Abrahams, Kwok et al. 2005). Briefly, B6 females heterozygous for the BAC insert and for the fierce deletion (B6.Cg-Hprt^{tm86(NR2E1,bEMS223)Ems}/X, *Nr2e1*^{frc/+}), were crossed to 129 males heterozygous for the fierce mutation (129S1/SvImJ.Cg-*Nr2e1*^{frc/+}). This produced first-generation hybrid offspring (B6129F1) abbreviated here as: Wt (X/Y, *Nr2e1*^{+/+}), *NR2E1*/fierce (*Hprt*^{tm86(NR2E1,bEMS223)Ems}/Y, *Nr2e1*^{frc/frc}), fierce (X/Y, *Nr2e1*^{frc/frc}), and *NR2E1* (*Hprt*^{tm86(NR2E1,bEMS223)Ems}/Y, *Nr2e1*^{+/+}). Males harbouring the *NR2E1 lacZ* reporter BAC (bEMS86) were used to evaluate the expression pattern of the human *NR2E1*

gene (B6.129P2(Cg)-*Hprt*^{tm73(Ple142-lacZ)Ems}/Y and B6.129P2(Cg)-*Hprt*^{tm87(Ple142-lacZ)Ems}/Y) abbreviated here *NR2E1-lacZ*. The age of adult animals used in the different experiments ranged from 8 to 20 weeks.

All mice were maintained in the pathogen-free Centre for Molecular Medicine and Therapeutics animal facility on a 6 am–6 pm light cycle, 20 ± 2°C with 50 ± 5% relative humidity, and had food and water *ad libitum*. All procedures involving animals were in accordance with the Canadian Council on Animal Care (CCAC) and UBC Animal Care Committee (ACC) (Protocol# A07-0435).

3.2.3 Embryo and Adult Tissue Preparation

Time-pregnant mice were euthanized by cervical dislocation and embryos at E12.5 or E15.5 were dissected, and then fixed in 4% paraformaldehyde (PFA) with 0.1 M Na₂HPO₄ (PO) buffer (pH 8.0) for 1 hour at 4°C. Whole embryos were incubated in *lacZ* staining solution overnight at 37°C (X-gal (1 mg/ml), MgCl₂ (2 mM), K₃Fe(CN)₆ (4 mM), K₄Fe(CN)₆ (4 mM) in 1x phosphate buffered saline (PBS)), before being post fixed in 4% PFA for an additional 4 hours. The embryos were then cryoprotected as described in the literature and embedded in optimal cutting temperature (OCT) compound (Tissue-tek, Torrance, California) on dry ice (Li, Sun et al. 2008). Embryos were then sectioned at 16 µm using a Cryo Star HM550 cryostat (MICROM International, Kalamazoo, Michigan) and mounted for imaging or cleared as described in the literature (Schatz, Golenser et al. 2005) and pictures were taken in 100% glycerol solution.

Intracardial perfusions were performed on avertin-anesthetized mice with 4% PFA with 0.1 M PO buffer (pH 8.0). For *lacZ* expression analysis, brain and eye tissues were

collected and post fixed in 4% PFA for an additional 30 min at 4°C, then transferred to 20% sucrose with 0.05 M PO buffer overnight at 4°C and embedded the next day in OCT on dry ice. For luxol fast blue/cresyl violet staining and immunofluorescence, brains were collected and post fixed in PFA overnight at 4°C, then transferred to 20% sucrose with 0.05 M PO buffer overnight at 4°C and embedded the next day in OCT on dry ice. For haematoxylin and eosin staining, eyes were collected and transferred to Davison fixative for post fixing overnight at 4°C, then subsequently washed in 1x PBS and 50-70% ethanol solution before being processed for paraffin embedment.

3.2.4 Histology

For *lacZ* staining, 20-25 µm cryosections from adult brains (floating sections) or adult eyes (sections on slides) were rehydrated in sequential washes of PBS, permeabilized in PBS with 0.1% triton before being incubated in *lacZ* solution overnight at 37°C (X-gal (1mg/ml), MgCl₂ (2 mM), K₃Fe(CN)₆ (4 mM), K₄Fe(CN)₆ (4mM) in 1x PBS-0.1% triton, deoxycholate 0.01%, NP40 0.02%).

For antibody staining, 20-25 µm cryosections from adult brains (floating sections) or E15.5 embryos and adult eyes (sections on slides) were rehydrated in sequential washes of PBS, permeabilized in PBS with 0.1% triton (adults) or 0.3% triton (embryos), and quenched in 0.1 M glycine-PBS solution. The cryosections were blocked with 1% BSA in PBS with 0.1% triton (adults) or 0.3% triton (embryos) for 1 hour at room temperature before applying the primary antibodies. Co-localization experiments were performed using chicken anti-β-galactosidase (β-gal) antibody (Abcam, San Francisco, California, ab9361) (1:5,000), rabbit anti-Sox2 antibody (Abcam, ab97959) (1:1,000), mouse anti-CRALBP antibody (Abcam,

ab15051) (1:1,200), rabbit anti-Ki67 antibody (Abcam, ab15580) (1:1,000), mouse anti-Glial Fibrillary Acidic Protein (GFAP) antibody (Vector, Burlington, Ontario, VP-G805) (1:1,000), or mouse anti-NeuN antibody (Chemicon, Billerica, Massachusetts, MAB377) (1:1,000), and incubated overnight at 4°C. Corresponding secondary antibodies coupled to Alexa 488 or Alexa 594 (Invitrogen, Burlington, Ontario) were incubated at room temperature for 2 h in the dark (1:1,000). Hoechst 33342 was used for nuclear staining on all sections. Ki67 positive cell counting was performed as described in the literature (Wong, Hossain et al. 2010).

For brain histology, coronal floating sections of 25 µm were mounted on slides, and dehydrated with 95% ethyl alcohol before being left overnight at 56°C in a luxol fast blue solution (0.1% luxol fast blue in 95% ethyl alcohol). Excess stain was then rinsed with 95% ethyl alcohol and the slides were rinsed in distilled water before being differentiated in a lithium carbonate solution for 30 sec (0.05% lithium carbonate in distilled water). The differentiation was then continued using 70% ethyl alcohol for 15 sec and counterstained in cresyl echt violet solution (0.1% cresyl violet in distilled water) for 6 min before being dehydrated and mounted for microscopy.

For retinal histology, 5 µm paraffin sections were rehydrated and incubated in haematoxylin solution for 5 min, then washed in tap water and incubated in 1% lithium carbonate solution for 30 sec before being washed in tap water again and incubated in acid alcohol 1% for 5 sec. The sections were then washed again in tap water before being incubated in eosin Y solution for 5 min and finally washed in tap water and dehydrated in 95-100% ethanol and xylene before mounting for microscopy.

3.2.5 Funduscopy

Direct funduscopy was performed as described in the literature (Hawes, Smith et al. 1999). Eyes were dilated with 1% atropine, 30 min before examination. The animals were restrained from moving without sedation.

3.2.6 Electroretinograms

Electroretinogram responses were recorded as previously described (Guerin, Gregory-Evans et al. 2008). Briefly, animals were dark-adapted, anesthetized with xylazine (13 mg/ml) and ketamine (87 mg/ml) and maintained on a heating pad. The corneas were locally anesthetized with 0.5% proparacaine hydrochloride and the pupils were dilated with 2.5% phenylephrine and 1% atropine. Experiments were conducted using an Espion E2 System with a Colordome mini-Ganzfeld stimulator (Diagnosys LLC, Lowell, Massachusetts). DTL Plus corneal electrodes were used (Diagnosys LLC, Lowell, Massachusetts). Dark-adapted responses were recorded by averaging fifteen responses for each stimulus intensity (0.01 and 3.16 candela-sec (cd-s)/m²).

3.2.7 Comparative Genomic and Transcription Factor Binding Site (TFBS)

Overrepresentation Analysis

For the comparative genomic analysis, relative sequence coordinates were retrieved for the BAC construct used in this and three other studies (Gong, Zheng et al. 2003; Abrahams, Kwok et al. 2005; Liu, Belz et al. 2008) and then visualized using the University of California, Santa Cruz (UCSC) genome browser (<http://genome.ucsc.edu/index.html>, assembly Feb. 2009).

For the TFBS overrepresentation analysis on the highly conserved regions, phylogenetic foot printing was performed by extracting a phastCons profile to identify regions with significant identity (default parameters for multi-species analysis were used). TFBS predictions were performed using the ORCA toolkit system (<http://burgundy.cmmt.ubc.ca/cgi-bin/ORCAtk/orca>, version 1.0.0). TFBSs were predicted using the “all JASPAR CORE vertebrate” profiles (default parameters). A gene ontology (GO) term over-representation analysis on the enriched transcription factor population was performed using “GOHyperGAll function” of “GOstats library” from “R package” (Falcon and Gentleman 2007; Horan, Jang et al. 2008). Terms with a Bonferroni-corrected P -value (< 0.05) were considered significantly enriched in the analysis.

3.2.8 Nucleotide Shuffling Analysis

Sequences from the highly conserved regions were used in a random mono-nucleotide shuffling analysis. The nucleotides were shuffled to obtain random sequences, conserving the original mono-nucleotide distribution amongst the conserved regions. TFBS prediction analysis and GO term over-representation analysis were performed as described in the previous paragraph, on the obtained random sequences. The following terms were retained for the shuffling analysis: “nervous system development”, “central nervous system development”, “neurogenesis”, “eye development”, and “camera-type eye development” for both the 5' and 3' conserved regions. The analysis compiled for each GO term, the number of times (n) a Bonferroni-corrected P -value, lower than or equal to the original Bonferroni-corrected P -value, was found on the shuffled sequences (10,000 times). The associated P -value obtained for each GO term (randomised P -value) correspond to n divided by 10,000.

3.2.9 Mouse Statistical Analysis

All analyses were performed using STATISTICA 6.0 (Statsoft, Inc., Tulsa, Oklahoma). Non parametric analysis using Kruskal–Wallis was performed as described in the literature (Abrahams, Kwok et al. 2005).

3.3 Results

3.3.1 Human *NR2E1-lacZ* Embryos Displayed an Unexpected Absence of Expression in the Dorsal Pallium in the Brain while Retaining Appropriate Region-Specific Expression in the Eyes.

To understand the functionality, and document the expression pattern, of the human *NR2E1* gene in brain and eye development, we generated novel humanized mouse strains for *NR2E1* as described in material and methods. Frequency of properly targeted ESC clones varied between 24 % for bEMS223 (*NR2E1* functional construct) and 21 % for bEMS86 (*NR2E1-lacZ* construct).

Two independent human *NR2E1-lacZ* strains, B6.129P2(Cg)-*Hprt*^{tm73(Ple142-lacZ)Ems} and B6.129P2(Cg)-*Hprt*^{tm87(Ple142-lacZ)Ems} were generated and yielded similar embryonic expression results, all the data presented here was obtained using B6.129P2(Cg)-*Hprt*^{tm73(Ple142-lacZ)Ems}. Mouse *Nr2e1* gene expression at E12.5 is detected at high levels in the VZ of the dorsal and lateral pallium (DP, LP) as well as the lateral ganglionic eminence (LGE). Lower expression levels are also detected in the medial pallium (MP), as well as the medial ganglionic eminence (MGE) at this time point (Stenman, Yu et al. 2003). As expected, whole mounts of E12.5 *NR2E1-lacZ* embryos showed expression in the nasal cavities (NC), eyes and the diencephalon. Expression in the telencephalon was distributed in

a rostral caudal manner where low expression was apparent at the rostral/basal level and extended strongly to the caudal/lateral regions. Unexpectedly, no expression in the dorsal telencephalon was apparent in these animals (black arrow, Figure 3.1A). Cryosections of E12.5 *NR2E1-lacZ* embryos revealed a staining pattern that extended from the LP to the optic recess (ORE) in the caudal regions of the developing telencephalon (Figure 3.1B). On this section, it was also possible to see staining in the innermost surface of the NC (black arrows, Figure 3.1B). Higher magnification focusing on the lateral telencephalon demonstrated that the staining extended from the LP to the ventral pallidum (VP), the LGE, the MGE and the supra optic area (SOA) as well as the MP (Figure 3.1C). Unexpectedly, no *lacZ* staining was found in the DP, a region that has been reported to strongly express mouse *Nr2e1* at this time point (Stenman, Yu et al. 2003). Further caudally, the staining became more restricted to the postoptic area (PSA) in the diencephalon and the vomeronasal organ (VNO) in the NC (Figure 3.1D and Figure 3.1E, black arrows). Staining in E12.5 cleared embryos showed strong expression in the NC (black arrows) and the innermost surface of the neural retina (white arrows), two regions known to express mouse *Nr2e1* at this time point (Figure 3.1F) (Monaghan, Grau et al. 1995; Yu, Chiang et al. 2000; Miyawaki, Uemura et al. 2004). The eye expression at E12.5 was restricted to the innermost surface of the retina and extended toward the periphery (Figure 3.1F, Figure 3.1G). Lower levels of expression were present in the developing optic nerve (black arrow) (Figure 3.1G). This staining pattern in eye development from *NR2E1-lacZ* mimicked that obtained from a mouse *Nr2e1-lacZ* reporter strain previously reported (Miyawaki, Uemura et al. 2004; Zhang, Zou et al. 2006). From E15.5 onward, the mouse *Nr2e1* gene expression becomes uniformly localized in RPC populating the NBL (Miyawaki, Uemura et al. 2004). Sox2, a known marker of RPC in the

NBL of the developing retina was used to perform co-localization experiments on retinal cryosections from E15.5 *NR2E1-lacZ* embryos with a β -gal specific antibody (Taranova, Magness et al. 2006). The results revealed that the human *NR2E1-lacZ* reporter gene was also expressed uniformly along the NBL with strong staining within the outer NBL region (Figure 3.1H). Higher magnification revealed that the staining was mainly found in the cytoplasm of Sox2-positive cells in the NBL of the developing retina (Figure 3.1I). Overall, these results suggested proper expression of the human *NR2E1* gene in the developing eye and NC but also demonstrated absence of expression in the VZ of the DP in the developing telencephalon, a region critical for the development of the neocortex (Butler 1994; Butler 1994).

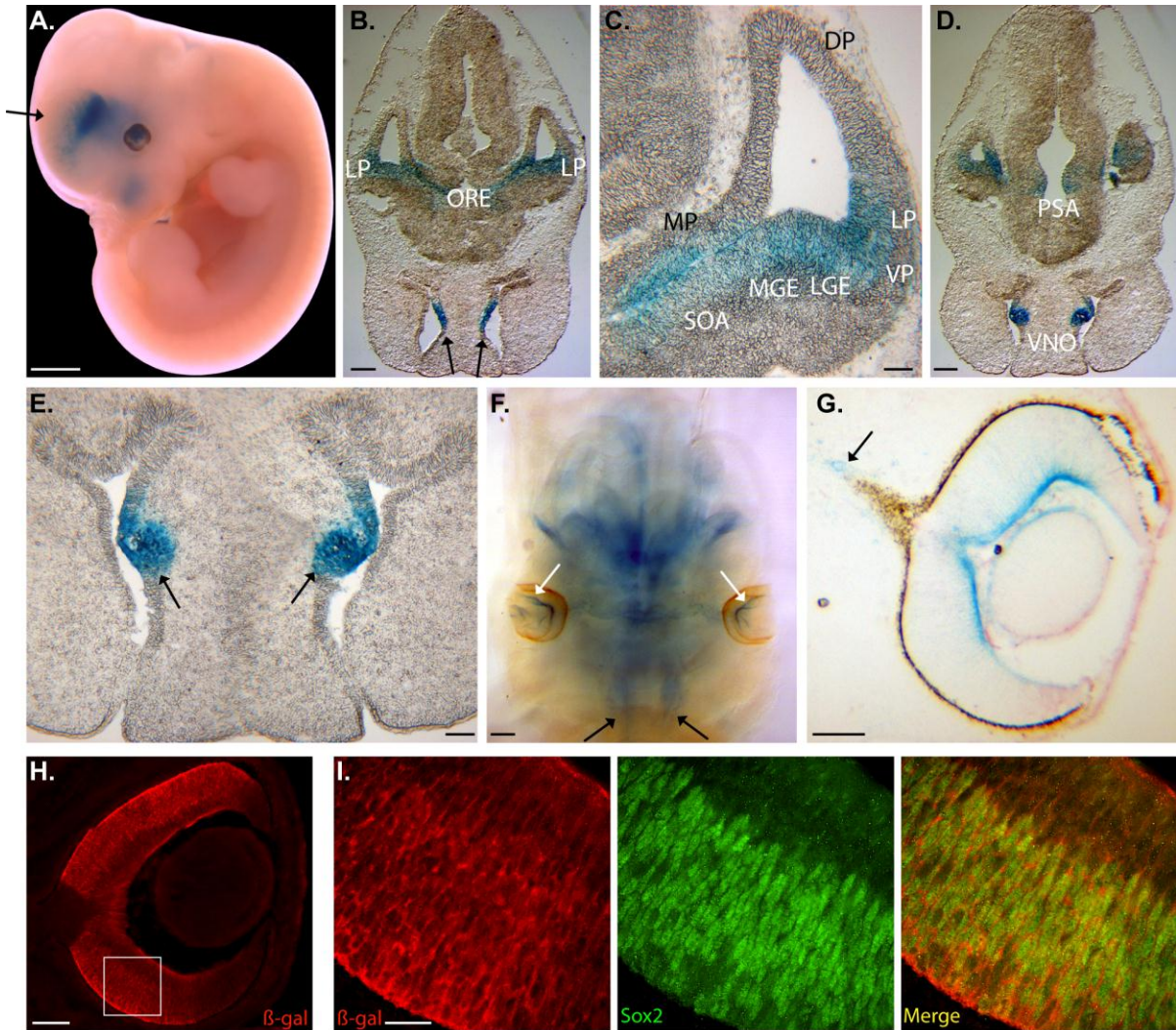


Figure 3.1 *NR2E1-lacZ* Targeted at the *Hprt* Locus Demonstrated Unexpected Absence of Expression in the Dorsal Pallium in the Developing Brain while Retaining Appropriate Region-Specific Expression in the Developing Eye.

A. BAC *NR2E1-lacZ* was expressed in the lateral and caudal regions of the telencephalon, the diencephalon, nasal cavities (NC) and developing eyes as shown in embryonic day 12.5 (E12.5) embryos. The arrow highlights the absence of expression in the dorsal telencephalon. **B.** Expression was restricted to ventral structures of the caudal telencephalon and extended from the lateral pallium (LP) to the optic recess (ORE). Strong expression in the NC was also observed (arrows). **C.** Higher magnification demonstrated absence of *lacZ* expression in the ventricular/subventricular zone (VZ/SVZ) of the dorsal pallium (DP). The medial pallium (MP), LP, ventral pallium (VP), lateral ganglionic eminence (LGE), medial ganglionic eminence (MGE), and supra optic area (SOA) showed staining for *lacZ*. **D.** Staining in the diencephalon was restricted to the postoptic area (PSA). **E.** The vomeronasal organ (VNO) also displayed strong *lacZ* staining (arrows). **F.** Eye expression was restricted to the innermost surface of the developing retina (white arrows). Expression was also in the NC (black arrows). **G.** Strong expression was observed in the innermost surface of the developing retina, which extended to the periphery of the developing eye. Staining was also in the developing optic nerve (arrow). **H.**

Immunofluorescence performed on E15.5 embryo retina using an anti β -gal antibody (red) revealed staining along the innermost surface of the retina as well as staining in cells found in the neuroblastic layer. Boxed region in **H** is shown in **I**. **I**. Co-localization performed between β -gal (red) and Sox2 (green), focusing on the NBL revealed β -gal positive staining in the cytoplasm of Sox2 positive cells, suggesting expression in the retinal progenitor cells. **A, F**, whole mounts; **B, C, D, E, G, H, I** cryosections. (Scale bars: **A, B, D, F**, 2 mm; **C, E, G, H**, 100 μ m, **I**, 20 μ m.)

3.3.2 Human *NR2E1-lacZ* Displayed an Unexpected Absence of Expression in Neurogenic Regions in the Adult Brain while Retaining Appropriate Cell-Type-Specific Expression in the Retina.

Mouse *Nr2e1* gene expression in the adult brain is found in proliferative cells lining the SVZ, RMS and the subgranular layer of the dentate gyrus (DG) as well as neurons populating the HPF, striatum and cortex (Shi, Chichung Lie et al. 2004; Li, Sun et al. 2008; Liu, Belz et al. 2008; Zhang, Zou et al. 2008). Thus, as expected, adult brain sections stained from the BAC *NR2E1-lacZ* reporter animals demonstrated sparse expression in the hippocampal formation (HPF). The staining was restricted to the region bordered by the CA1, CA2 and CA3 dorsally and the hippocampal fissure (hf) ventrally (Figure 3.2A, black arrows Figure 3.2B). Very few lightly stained cells were found at the border of the subgranular layer in the DG (Figure 3.2C, black arrows Figure 3.2D) but there was an unexpected absence of staining in the SVZ of the lateral ventricles (Figure 3.2E-F). As expected, the cortex contained *lacZ* positive cells in the upper layers (layers I-II/III) (Figure 3.2G). Overall, these results showed unexpectedly, an absence of expression of the human BAC *NR2E1-lacZ* reporter in the SVZ of the lateral ventricles but an appropriate expression in the subgranular layer of the DG, the HPF, and superficial layer of the cortex.

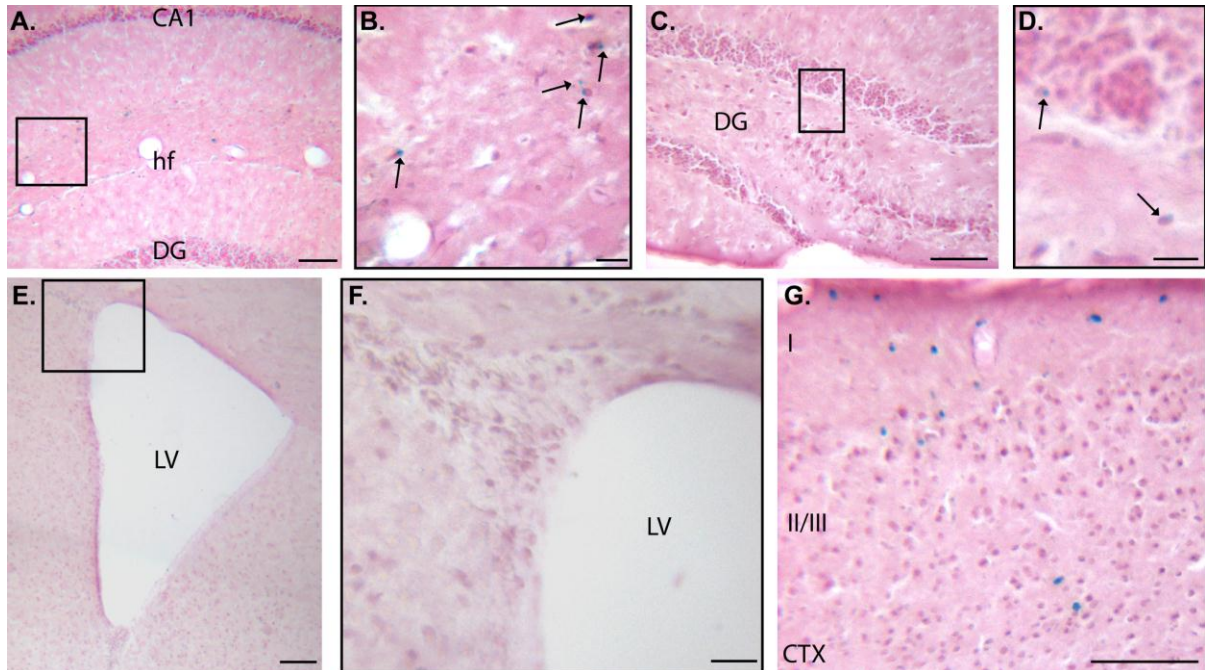


Figure 3.2 NR2E1-lacZ Targeted at the *Hprt* Locus Demonstrated Unexpected Absence of Expression in Neurogenic Regions of the Adult Forebrain.

A, B. BAC *NR2E1-lacZ* was sparsely expressed in the hippocampal formation (arrows). Boxed region in **A** is shown in **B**. CA1, Cornu Ammonis 1; hf, hippocampal fissure; DG, dentate gyrus. **C, D.** Very few *lacZ* positive cells were in the subgranular layer of the DG (arrows). Boxed region in **C** is shown in **D**. **E, F.** *lacZ* staining demonstrated absence of expression in the subventricular zone of the lateral ventricles (LV). Boxed region in **E** is shown in **F**. **G.** *lacZ* staining was in layers I, II and III of the cortex (CTX). (Cryosections. Scale bars: **A, C, E, G**, 100 μm ; **B, D, F**, 20 μm .)

We sought to investigate the nature of these *lacZ*-positive cells using co-localization analyses with a β -gal antibody. Co-localization with NeuN and β -gal revealed that the reporter gene was expressed in neurons populating the HPF; including the subgranular layer of the DG (Figure 3.3A, Figure 3.3B). These results were confirmed by an absence of co-localization with either Ki67 or GFAP antibodies in these regions (HPF; Figure 3.3C, Figure 3.3D, Figure 3.3E, DG; Figure 3.3F, Figure 3.3G).

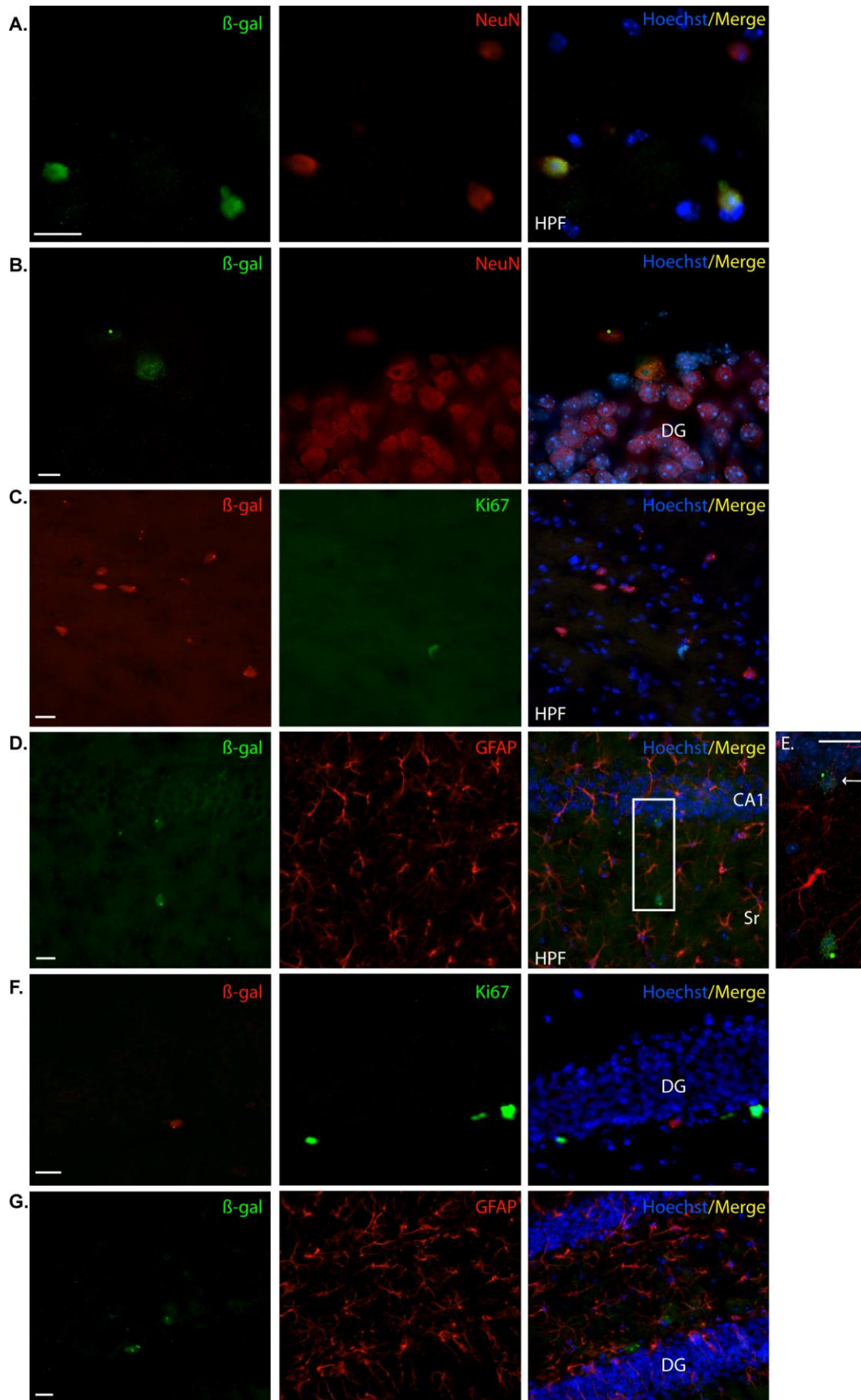


Figure 3.3 NR2E1-lacZ Targeted at the Hprt Locus Demonstrated Unexpected Absence of Expression in Adult Proliferative Cells of the Dentate-Gyrus-Subgranular Layer.

A. In BAC *NR2E1-lacZ* mice, immunofluorescence using an anti β -gal antibody (green) revealed positive cells in the hippocampal formation (HPF) that co-localized with NeuN (red), suggesting expression in mature neurons. **B.** Co-localization revealed few β -gal immunoreactive cells (green) in the subgranular layer of the dentate gyrus (DG) that co-localized with NeuN (red), suggesting expression in mature neurons. **C.** β -gal immunopositive cells (red) in the HPF did not co-localize with Ki67 (green). **D.** β -gal immunopositive cells (green) in the HPF did not co-localize with GFAP (red). Boxed region in **D** is shown in **E**. **E.** Higher magnification of the co-localization between β -gal (green) and GFAP (red) revealed β -gal positive cells in the Cornu Ammonis1 (CA1) of the hippocampus (arrow). These cells did not co-localize with GFAP. **F.** β -gal immunopositive cells (red) in the DG did not co-localize with Ki67 (green). **G.** β -gal immunopositive cells (green) in the DG did not co-localize with GFAP (red). Sr; stratum radiatum. (Cryosections. Scale bars: **A-G**, 20 μ m.)

In contrast to direct staining, immunolocalization revealed very few β -gal positive cells in the SVZ. However, these cells were not positive for Ki67, indicating that they were non-proliferative (Figure 3.4A). The mouse *Nr2e1* gene has been shown to be expressed in astrocyte-like type-B cells, a population immunoreactive for GFAP in the mouse SVZ (Liu, Belz et al. 2008). The results obtained in our study demonstrated that the few β -gal positive cells surrounding the SVZ were not immunoreactive for GFAP, confirming absence of staining of our reporter gene in astrocyte-like type B cells (Figure 3.4B).

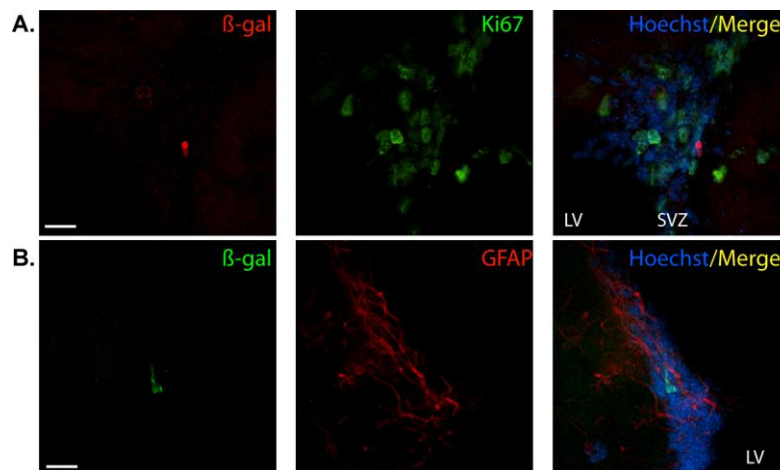


Figure 3.4 NR2E1-lacZ Targeted at the *Hprt* Locus Demonstrated Unexpected Absence of Expression in Astrocyte-Like Type-B Cells in the Subventricular Zone of the Lateral Ventricle.

A. In BAC *NR2E1-lacZ* mice, co-localization revealed the presence of few β -gal immunoreactive cells (red) in the subventricular zone (SVZ) of the lateral ventricle (LV). These cells were not positive for Ki67 (green). **B.** β -gal immunopositive cells (green) in the SVZ of the LV did not co-localize with GFAP (red). (Cryosections. Scale bars: **A, B**, 20 μ m.)

The β -gal positive cells in the cortex were immunoreactive for NeuN and not immunoreactive for Ki67 or GFAP, together implying that the BAC *NR2E1-lacZ* reporter was only expressed in mature neurons in the cortex (Figure 3.5A, Figure 3.5B, Figure 3.5C).

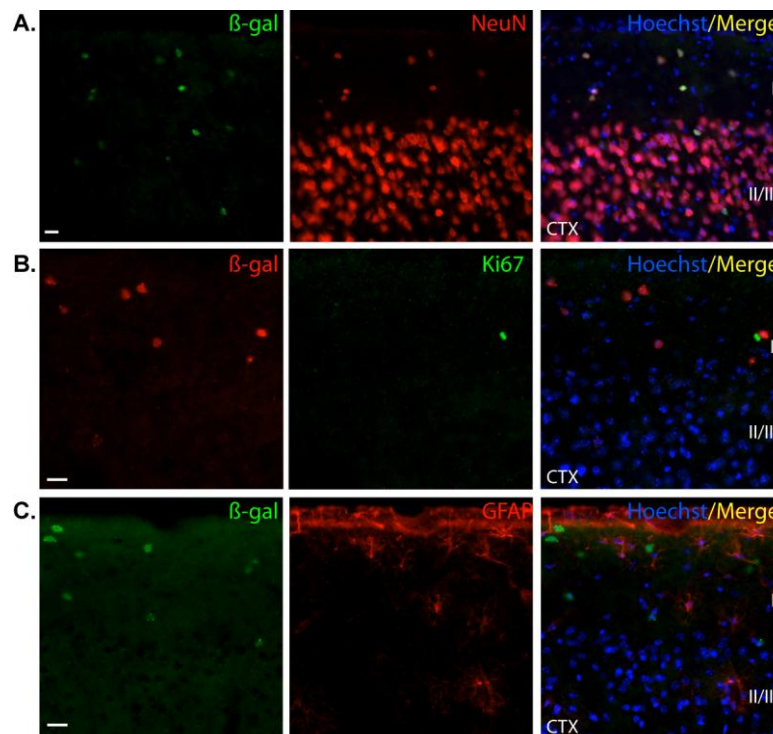


Figure 3.5 NR2E1-lacZ Targeted at the *Hprt* Locus Demonstrated Expected Expression in Neurons Populating the Upper Cortical Layers I and II/III.

A. In BAC *NR2E1-lacZ* mice, co-localization revealed the presence of β -gal positive cells (green) in the cortex (CTX) that were positive for NeuN (red), suggesting expression in mature neurons. **B.** β -gal immunopositive cells (red) in the CTX did not co-localize with Ki67 (green). **C.** β -gal immunopositive cells (green) in the CTX did not co-localize with GFAP (red). I, II/III, cortex layers. (Cryosections. Scale bars: **A-C**, 20 μ m.)

Finally, both the *lacZ* staining and the β -gal antibody revealed absence of expression in the striatum and RMS in the BAC *NR2E1-lacZ* animals (data not shown). Overall, our

results showed that like the mouse gene, human *NR2E1* was sparsely expressed in neurons in the cortex and hippocampus (Zhang, Zou et al. 2008), but the results also demonstrated that the human BAC *NR2E1-lacZ* reporter was not expressed in proliferative cells in regions where neurogenesis normally occurs in adult mice.

In the adult eye, the mouse *Nr2e1* gene is known to express in Müller cells populating the retina (Miyawaki, Uemura et al. 2004). We thus examined where *NR2E1-lacZ* expressed in adult retina sections using both the staining method and colocalization experiments. The stained sections revealed characteristic expression in Müller cells, with the body of cells located in the inner nuclear layer (INL) and fibres extending radially throughout the inner plexiform layer (IPL) and outer nuclear layer (ONL) (Figure 3.6A). The end feet of these stained cells were located in the inner limiting membrane (ILM) and the outer limiting membrane (OLM) (Figure 3.6B). Colocalization was shown with the β -gal antibody and CRALBP; the latter a known marker of Müller cells (Figure 3.6C). Thus, as expected, these results demonstrated Müller cell-type specific expression of the human BAC *NR2E1-lacZ* reporter in the adult mouse retina.

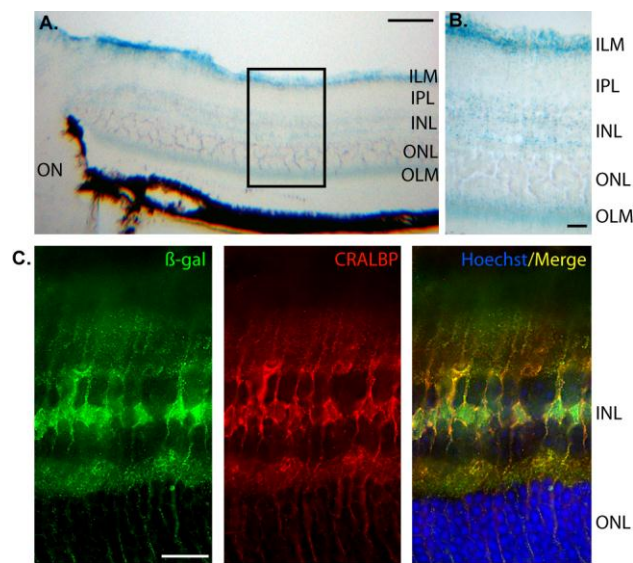


Figure 3.6 *NR2E1-lacZ* Targeted at the *Hprt* Locus Demonstrated Appropriate Cell-Type Specific Adult Retina Expression.

A, B. In BAC *NR2E1-lacZ* mice, *lacZ* expression in the adult eye was strong in the inner nuclear layer (INL) and extended radially to the inner limiting membrane (ILM) and outer limiting membrane (OLM). This staining pattern was consistent with Müller cells in the adult retina. IPL, inner plexiform layer; ON, optic nerve; ONL, outer nuclear layer. **C.** Co-localization, using an anti β -gal antibody (green) and anti CRALBP antibody (red), revealed cell appropriate expression in Müller cells in the adult retina. The β -gal expression pattern was localized to the large nucleus and perinucleus structure in the INL. CRALBP staining was mainly in the cytoplasm of these cells as expected. (Cryosections. Scale bars: **A.** 100 μ m, **B, C,** 20 μ m.)

3.3.3 *NR2E1*/fierce Animals Displayed Adult Forebrain Abnormalities and

Neurogenesis Defects.

The results obtained with the BAC *NR2E1-lacZ* reporter mouse strain showed absence of expression in key neurogenic regions and cells, in both the developing and adult brain. To understand the importance of this absence of expression, we used the same human BAC, but without the *lacZ* reporter and thus containing a functional *NR2E1* gene, knock-in to the same chromosomal *Hprt* locus. This single-copy-human knock-in allele (abbreviated here as *NR2E1*) was bred onto the fierce (*Nr2e1^{frc/frc}*) background, which is null for mouse *Nr2e1*, to generate experimental animals referred here as *NR2E1*/fierce. In these resulting animals, brain and eye development relied solely on the single-copy functional human *NR2E1*. Comparison analyses were performed with three different controls; wild-type (Wt) animals, fierce animals, and *NR2E1* animals with the human BAC on the Wt background.

Fierce mice are named for their aggressive behaviour and display gross brain abnormalities such as hypoplasia of the OB and cerebrum (Young, Berry et al. 2002). Our results demonstrated that a single copy of the human *NR2E1* gene can only partially correct the fierce brain phenotype. *NR2E1*/fierce animals were difficult to handle and aggressive towards other mice. They also displayed OB and cerebrum hypoplasia with exposed colliculi,

phenotypes that were absent in Wt and *NR2E1* animals (Figure 3.7A). Quantitative analyses of the surface area from the OB (Figure 3.7B) and cerebrum (Figure 3.7C) demonstrated significant reduction of the forebrain structures in *NR2E1/fierce* animals when compared to Wt and *NR2E1* (P -values < 0.001). However, *NR2E1/fierce* animals displayed larger OB and cerebrum size when compared to the *fierce* animals, showing that the *NR2E1* BAC can partially correct the *fierce* phenotype (P -values, OB size < 0.01 , cerebrum size < 0.001). Nevertheless, no significant difference was observed when comparing the OB and cerebrum size of Wt and *NR2E1* animals, indicating no effects of an additional human copy on the developing brain of Wt animals.

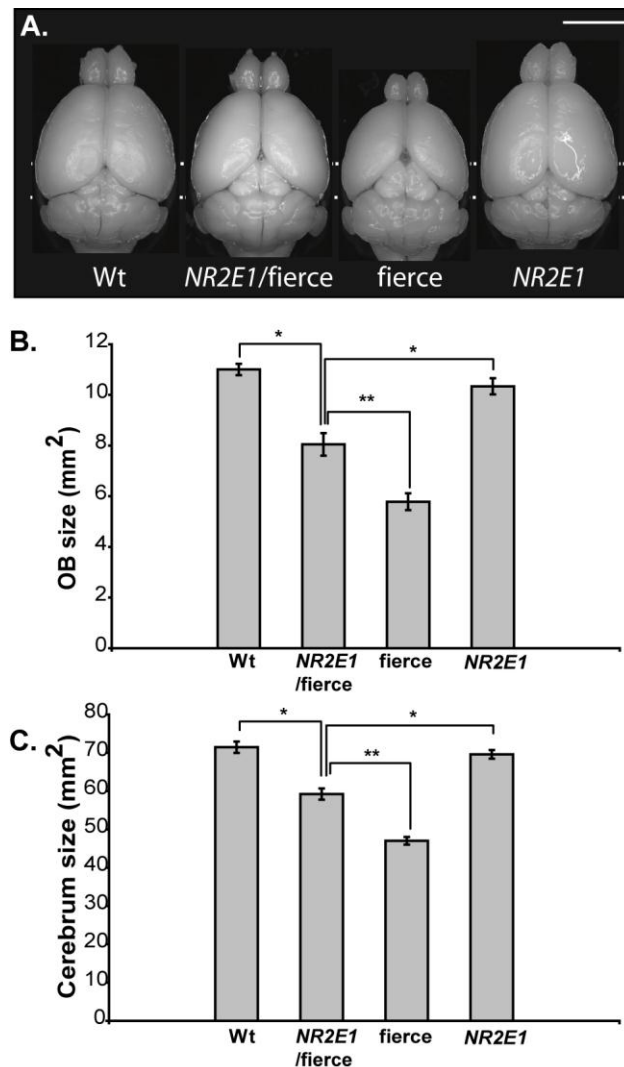


Figure 3.7 *NR2E1*/fierce Mice Brains were Partially Corrected for Reduced Olfactory Bulb and Cerebrum Size.

A. Gross brain dissection revealed reduced olfactory bulb (OB) and cerebrum size in fierce; a known characteristics of the null phenotype. To a lesser extent, *NR2E1*/fierce mice also displayed these abnormalities, suggesting incomplete correction of the fierce brain phenotype by the human gene. (Scale bar, 5 mm.) **B.** Surface quantification revealed significant reduction of the OB in *NR2E1*/fierce mice compared to Wt and *NR2E1* mice ($*P < 0.001$). The *NR2E1*/fierce OB was also significantly bigger than the fierce OB ($**P < 0.01$). No significance was found when comparing Wt and *NR2E1* mice. **C.** Similarly, cerebrum area was significantly reduced in *NR2E1*/fierce mice compared to Wt and *NR2E1* mice ($*P < 0.001$). The *NR2E1*/fierce cerebrum was also significantly bigger than the fierce cerebrum ($**P < 0.001$). No significance was found when comparing Wt and *NR2E1* mice. **B, C.** Kruskal–Wallis H-test was performed on N = 6-10 mice for all genotypes. Error bars represent standard errors of the mean.

To understand the importance of absence of *NR2E1* expression in the adult brain, we looked at the morphology of the SVZ of the lateral ventricles and the DG of the hippocampus. Fierce mice display enlarged ventricles as well as reduced and poorly defined DG due in part to neurogenesis defects in these areas (Monaghan, Bock et al. 1997; Young, Berry et al. 2002). Histological analyses demonstrated no difference in morphology between *NR2E1*/fierce and fierce mice, suggesting similar neurogenesis defects in both brain regions (Figure 3.8A). No morphological differences were observed between Wt and *NR2E1* animals in either brain region, again showing no effect of an additional human copy (Figure 3.8A). *Nr2e1* has been shown to be a critical regulator of adult neural stem cell population in both the SVZ and DG in adult mice (Liu, Belz et al. 2008; Zhang, Zou et al. 2008). We used Ki67 immunolabeling to understand the cortical hypoplasia defects found in *NR2E1*/fierce adult mice. Initial results obtained from both brain regions suggested a similar number of proliferative cells in *NR2E1*/fierce animals when compared to fierce (Figure 3.8B). No apparent differences in proliferative cell numbers were found when comparing Wt and *NR2E1* animal sections (Figure 3.8B). These results were confirmed by Ki67-positive cell

counting in both the SVZ and DG regions of these animals (Figure 3.8C). No statistical differences were found in the number of Ki67-positive cells in both the SVZ and DG between *NR2E1*/fierce and fierce animals, whereas significance was obtained between *NR2E1*/fierce and Wt or *NR2E1* animals (P -values < 0.05). Also, no statistical differences were obtained when comparing the number of Ki67-positive cells between Wt and *NR2E1* animals in both neurogenic regions, again showing no effect of an additional human copy. Overall, the results correlated with the expression data obtained with the BAC *NR2E1-lacZ* animals, and strongly suggested that an absence of *NR2E1* expression in proliferative cells of both the SVZ and DG led to an inability to correct proliferative and structural defects in these brain regions. This also suggests that the partial correction of both the OB and cerebrum size occurred during embryonic development rather than in the adult brain.

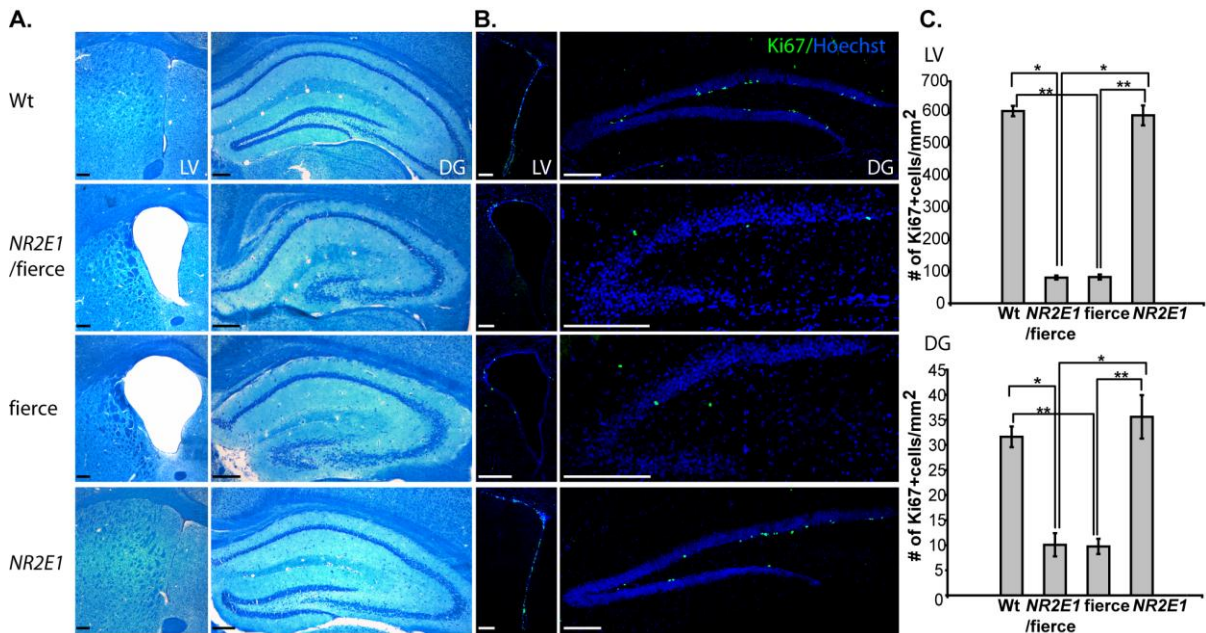


Figure 3.8 *NR2E1*/fierce mice Brains were not Corrected for Adult Neurogenesis Defects.

A. Luxol fast blue-cresyl violet stained cryosections analysis revealed enlarged lateral ventricles (LV), and reduced and poorly defined dentate gyrus (DG) in *NR2E1*/fierce mice, a phenotype indistinguishable from that found in fierce mice. No differences were found when comparing Wt and *NR2E1* mice. **B.** Ki67 immunopositive cell numbers were reduced in *NR2E1*/fierce mice in the LV and DG compared to Wt and *NR2E1* mice. **C.** Quantitative

analysis showed a significant reduction of Ki67 immunopositive cell numbers in *NR2E1/fierce* mice compared to Wt and *NR2E1* mice in both the LV ($*P < 0.05$) and DG ($*P < 0.05$). No significance was found when comparing Ki67 immunopositive cell numbers from *NR2E1/fierce* and *fierce* mice in either the LV or DG. Also, no significance was found when comparing Ki67 immunopositive cell numbers from Wt and *NR2E1* mice in either the LV or DG. As expected, there was a significant reduction in Ki67 positive immunopositive cell numbers in *fierce* mice compared to Wt and *NR2E1* mice in both the LV ($**P < 0.05$) and DG ($**P < 0.05$). (Scale bars: **A**, **B**, 200 μm .) **C**. Kruskal–Wallis H-test was performed on $N = 3$ mice for all genotypes for both the LV and DG. Error bars represent standard errors of the mean.

3.3.4 *NR2E1/fierce* Animals Displayed Appropriate Retinal Architecture.

We looked at the eyes of the *NR2E1/fierce* mice, which exhibited an appropriate expression pattern in the BAC *NR2E1-lacZ* mice. We began our investigation by evaluating the retinal architecture of the *NR2E1/fierce* animals using fundus microscopy. *Fierce* mice characteristically display a reduced and abnormal vascular patterning of the major retinal arteries and veins (Young, Berry et al. 2002; Abrahams, Kwok et al. 2005). Our investigation revealed that the blood vessels populating the retina of the *NR2E1/fierce* animals showed comparable structural characteristics to Wt animals, suggesting a correction of this aspect of the *fierce* phenotype (Figure 3.9A). Quantitative analysis of the vessel numbers in *NR2E1/fierce* showed no differences with Wt or *NR2E1* animals, whereas a significant reduction in blood vessel count was observed when comparing *fierce* to the other genotypes ($P\text{-value} = 0.001$) (Figure 3.9B). As described in the literature, radial asymmetry was only observed in *fierce* ($P\text{-value} < 0.001$) (Figure 3.9C) (Young, Berry et al. 2002; Abrahams, Kwok et al. 2005). These results render the eyes from the *NR2E1/fierce* animals indistinguishable from Wt, demonstrating complete correction of the vasculature defects of the *fierce* phenotype.

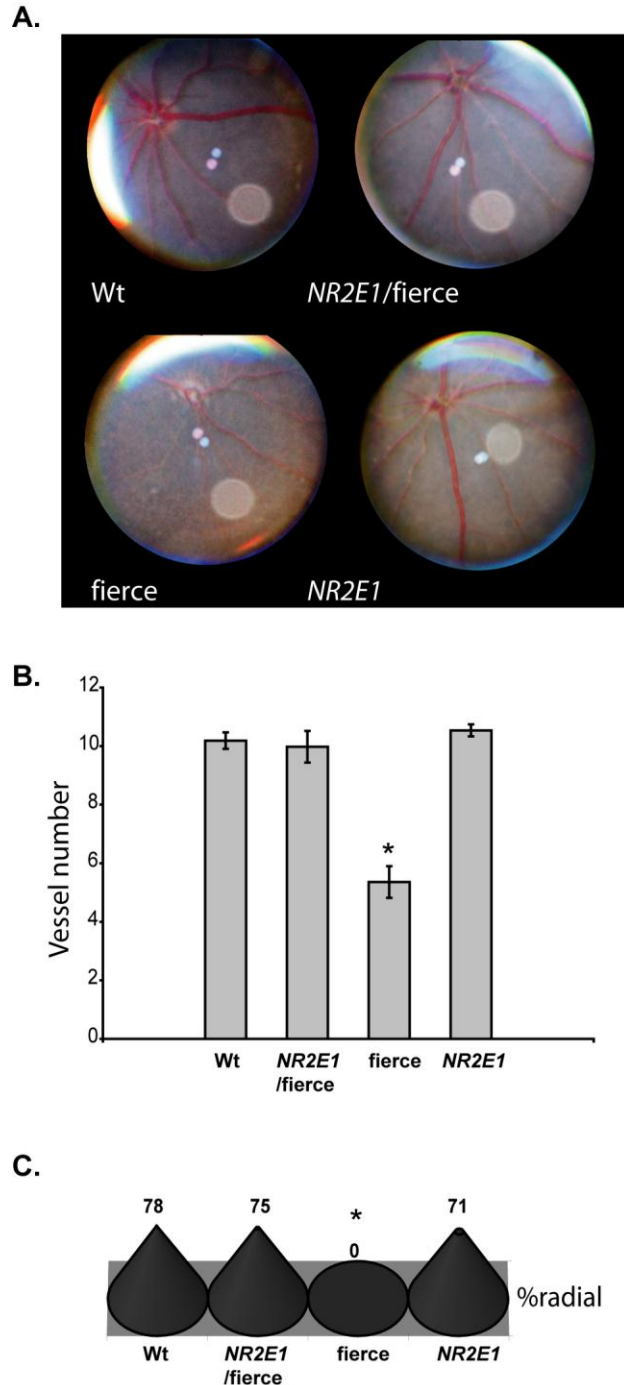


Figure 3.9 NR2E1/fierce Mice were Corrected for Retinal Blood Vessel Defects.

A. Eye fundus photos showed normal blood vessel organization in Wt, NR2E1/fierce, and NR2E1 retinal surface. The expected blood vessel abnormalities were seen in fierce. **B.** No significant difference was found in the blood vessel numbers of Wt, NR2E1/fierce, and NR2E1 mice. The blood vessel number was significantly reduced in fierce eyes compared to the other mice (* $P = 0.001$). Error bars represent standard errors of the mean. **C.** No significant differences were found between Wt, NR2E1/fierce, and NR2E1 mice for

asymmetry, only fierce showed asymmetry of the blood vessels ($*P < 0.001$). **B, C.** Kruskal–Wallis H-test was performed on $N = 6-9$ mice for all genotypes.

The eye defects in the fierce animals also include thinning of the INL and ONL in comparison to Wt (Young, Berry et al. 2002). Adult retinas from *NR2E1*/fierce animals revealed normal thickness of the INL and ONL when compared to Wt and *NR2E1* retinas (Figure 3.10A). The fierce animals INL measured $23.3 \pm 2.40 \mu\text{m}$, demonstrating a significant reduction when compared to $36.4 \pm 0.844 \mu\text{m}$ (Wt), $35.6 \pm 4.48 \mu\text{m}$ (*NR2E1*/fierce) and $33.6 \pm 3.84 \mu\text{m}$ (*NR2E1*) ($P\text{-value} < 0.01$) (Figure 3.10B). The fierce animals ONL also revealed a marked reduction of thickness measuring $32.3 \pm 2.87 \mu\text{m}$ compared to $56.0 \pm 2.72 \mu\text{m}$ (Wt), $55.6 \pm 4.88 \mu\text{m}$ (*NR2E1*/fierce) and $50.4 \pm 6.23 \mu\text{m}$ (*NR2E1*) ($P\text{-value} < 0.05$) (Figure 3.10C). The fierce animals GCL also demonstrated incomplete differentiation, a phenomenon characterized by the presence of remnant retinal ganglion cells (RGC) in the IPL (Young, Berry et al. 2002). We observed a significant increase in displaced RGC only in fierce retinas with an average of 20.4 ± 2.44 displaced ganglion cells in comparison to 8.20 ± 0.970 cells (Wt), 8.50 ± 0.957 cells (*NR2E1*/fierce) and 7.75 ± 1.18 cells (*NR2E1*) ($P\text{-value} < 0.01$) (Figure 3.10D). These results, combined with the expression pattern analysis of the BAC *NR2E1-lacZ*, revealed that the genomic fragment used to generate these animals contained all of the elements for proper temporal and spatial expression of the human gene in the eye. An additional human copy of *NR2E1* on the Wt background had no effect, and corrected the retinal architecture deleterious phenotype of the fierce background. Overall, these results suggested an appropriate functional role of the *NR2E1* human protein in the developing eyes, resulting in a correction of the blood vessels and retinal defects in adult animals.

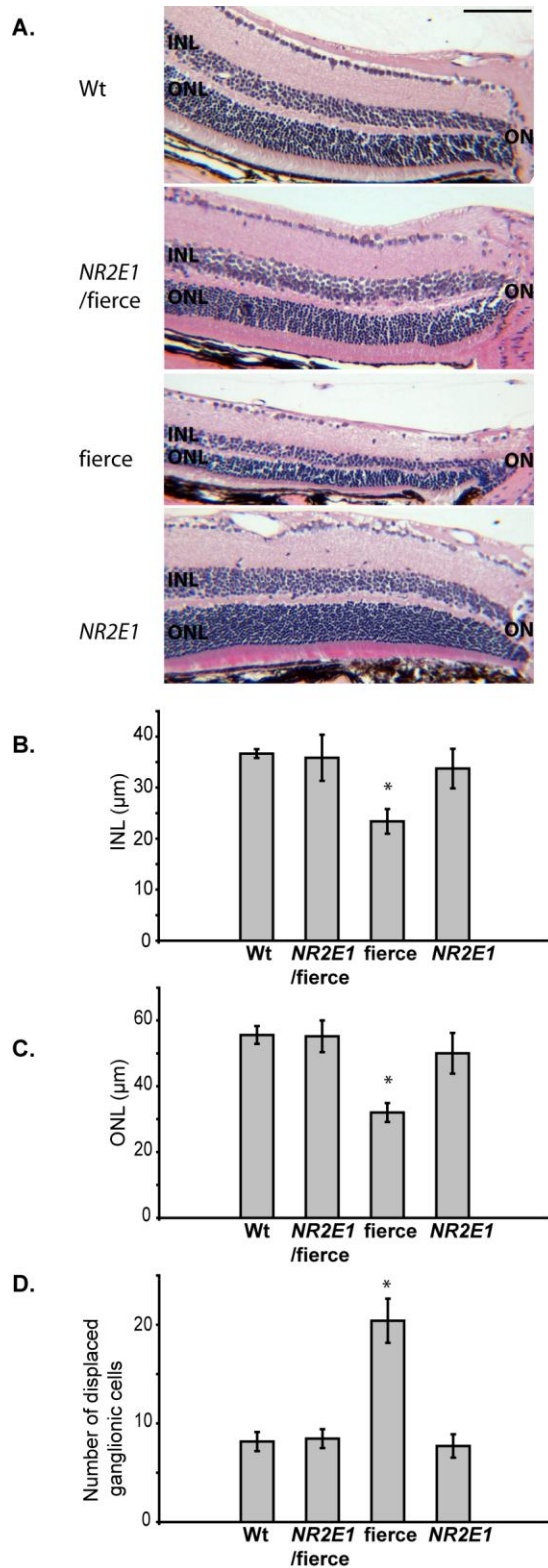


Figure 3.10 *NR2E1/fierce* Mice were Corrected for Retinal Architecture Defects.

A. Haematoxylin and eosin stained paraffin sections analysis revealed that the retina of Wt, *NR2E1/fierce*, and *NR2E1* mice appeared similar when the inner nuclear layer (INL) and outer nuclear layer (ONL) thickness was examined. Only the *fierce* retina demonstrated a

reduction in the INL and ONL compared to the three other genotypes. ON, optic nerve. (Scale bar, 100 μm). **B.** Quantitative analysis showed no significant difference when comparing the INL thickness of Wt, *NR2E1/fierce*, and *NR2E1* mice. Only the fierce retina demonstrated a reduction in INL thickness compared to the three other genotypes ($*P < 0.01$). **C.** No significant difference was found when comparing the ONL thickness of Wt, *NR2E1/fierce*, and *NR2E1* mice. Only the fierce retina demonstrated a reduction in ONL thickness compared to the three other genotypes ($*P < 0.05$). **D.** No significant difference was found when comparing the number of displaced ganglionic cells in the inner plexiform layer (IPL) of Wt, *NR2E1/fierce*, and *NR2E1* mice. Only the fierce retina demonstrated increased displaced retinal ganglion cells in the IPL compared to the other three genotypes ($*P < 0.01$). **B, C, D.** Kruskal–Wallis H-test was performed on $N = 4\text{-}6$ mice for all genotypes. Error bars represent standard errors of the mean.

3.3.5 *NR2E1/fierce* Animals Displayed Functional Retinas.

To confirm the functionality of the *NR2E1/fierce* retinas, we performed electroretinogram (ERG) experiments. In the literature, fierce mice subjected to an ERG demonstrate reduced to absent a-wave and b-wave signals in adult animals (Yu, Chiang et al. 2000; Young, Berry et al. 2002). We obtained a similar pattern for fierce animals with an average a-wave of $11.3 \pm 3.91 \mu\text{V}$ and b-wave of $29.3 \pm 9.00 \mu\text{V}$ at 3 cd.s/m^2 (Figure 3.11). These results demonstrated a significant reduction of amplitude for both a-wave and b-wave when compared to animals of the three other genotypes at the same intensity (P -values < 0.01) (Figure 3.11). No significant difference was found between Wt, *NR2E1/fierce* and *NR2E1* amplitude values, suggesting proper functionality of their retinas as well as no effect of the additional copy of *NR2E1*, and complete correction of the retina-null phenotype.

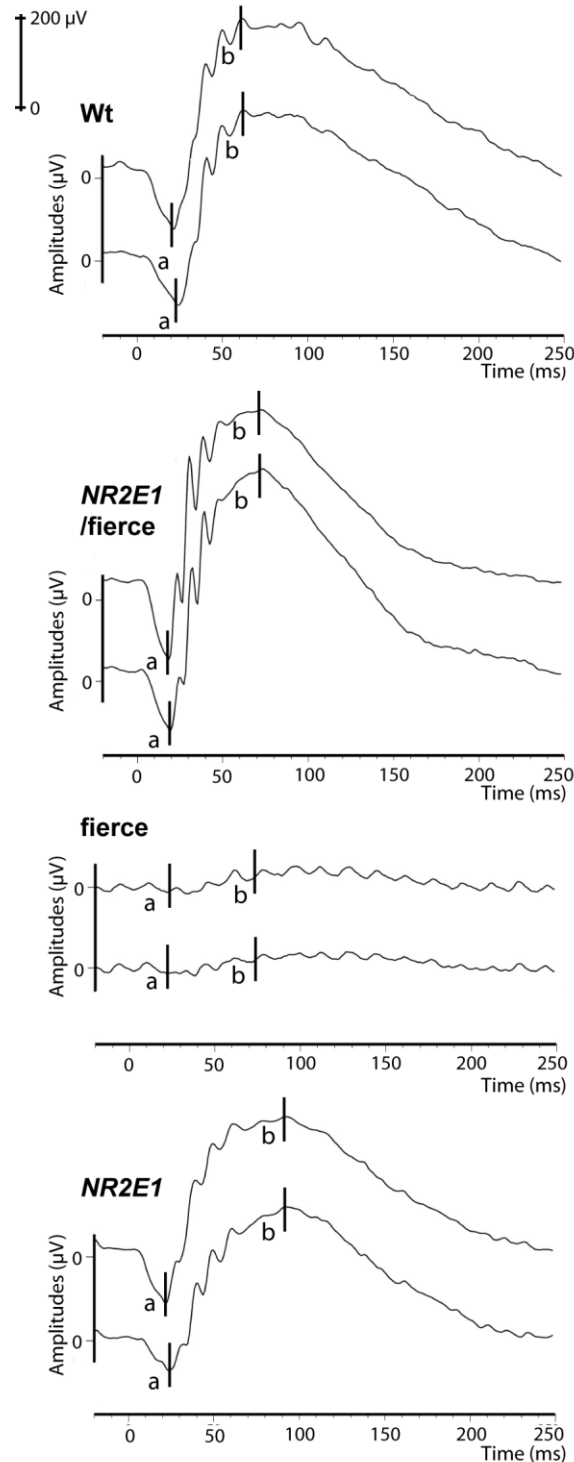


Figure 3.11 *NR2E1/fierce* Mice were Corrected for Retinal Functional Defects.

Electretinogram (ERG) experiments demonstrated normal a-wave and b-wave amplitudes at 3 cd.s/m^2 in Wt, *NR2E1/fierce*, and *NR2E1* mice. Only the *fierce* ERG values differed significantly compared to the other three genotypes (a-wave, $P < 0.01$, b-wave, $P < 0.01$). Each line presents measurements from one eye; each graph presents two eyes from one mouse. Kruskal–Wallis H-test was performed on $N = 3$ mice for all genotypes.

3.3.6 Comparative Genomic Analysis Revealed Candidate Brain-Specific Stem-Cell-Regulatory Elements.

To understand the discrepancy between the results obtained from the human BAC derived mice used in this manuscript to those obtained with a previous humanized mouse model for *NR2E1*, which demonstrated complete correction of the brain phenotype while only ameliorating the eye phenotype, we decided to undertake a comparative genomic analysis (Figure 3.12A) (Abrahams, Kwok et al. 2005). Sequence alignment comparison between our current mouse model and three others, which were proven successful in either functionally correcting the brain phenotype using the human gene or conferring proper expression in the brain using the mouse gene, revealed that our sequence was the shortest at the 5' end (Figure 3.12A) (Gong, Zheng et al. 2003; Abrahams, Kwok et al. 2005; Liu, Belz et al. 2008). Our construct was missing 25 kb of sequence which contained four highly conserved regions, located in the intragenic region, and spanning a distance of approximately 6 kb. These regions correspond to the following coordinates according to the UCSC genome browser (assembly Feb 2009): conserved region (CR) 1, chr6:108,435,521 to 108,436,101; CR 2, chr6:108,437,653 to 108,438,062; CR 3, chr6:108,439,859 to 108,440,601; CR 4, chr6:108,441,507 to 108,442,084 (grey rectangle box, Figure 3.12A). The construct used in the current study also contained additional 3' sequence that was not found in the previously humanized mouse model, which demonstrated partial correction of the eye phenotype (Figure 3.12A) (Abrahams, Kwok et al. 2005). Within this additional sequence of 24 kb, a series of five conserved regions were found, located within intron two of the neighbouring *SNX3* gene, and spanning a distance of approximately 11 kb. These regions correspond to the following coordinates according to the UCSC genome browser (assembly Feb 2009): CR 5,

chr6:108,544,148 to 108,544,255; CR 6, chr6:108,547,452 to 108,547,478; CR 7, chr6:108,547,888 to 108,548,739; CR 8, chr6:108,550,671 to 108,550,708; CR 9, chr6:108,553,391 to 108,554,246 (black rectangle box, Figure 3.12A).

We used TFBS prediction analysis to identify, and evaluate the over-representation of, transcription factors that could bind the four 5' intragenic conserved regions. The results revealed an enriched presence for TFBSs predicted for transcription factors with gene ontology (GO) terms involved in biological processes such as “nervous system development” (P -value = 9.23×10^{-12}), “central nervous system development” (P -value = 8.45×10^{-10}), “neurogenesis” (P -value = 2.63×10^{-07}), “eye development” (P -value = 9.45×10^{-06}), and “camera-type eye development” (P -value = 1.18×10^{-04}) (Table 3.2). A randomised statistical analysis performed on these four conserved regions also demonstrated that these GO terms were significant and not due to chance (Table 3.2, randomised P -value). The transcription factors corresponding to the most significantly enriched GO term “neurogenesis” (Table 3.2, randomised P -value = 0.0122) were particularly relevant to this study. This category contains 25 transcription factors including the “high mobility group” family (HMG) (Sox2, Sox5, Sox9, Sox10), the “homeobox” family (including Pax2 and Pax6), and the “hormone-nuclear receptor” family (Nr2f1, Nr2e3, Nr4a2), all of which are known to have important roles in brain development (Table 3.3, family binding site distribution; Figure 3.12B) (Stoykova, Treichel et al. 2000; Graham, Khudyakov et al. 2003; Kitambi and Hauptmann 2007; Naka, Nakamura et al. 2008).

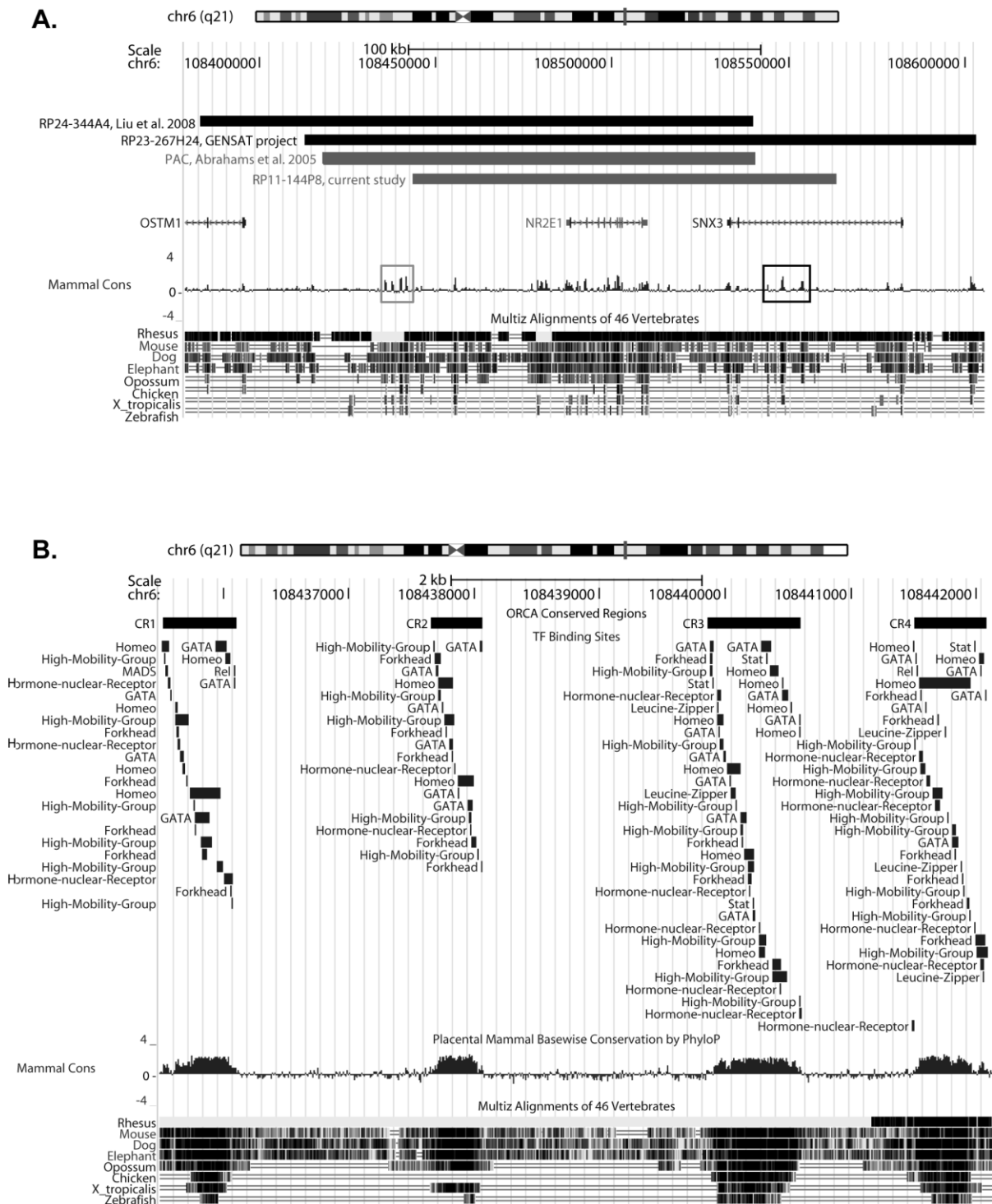


Figure 3.12 Comparative Genomic Analysis Revealed Candidate Regulatory Elements.

A. Relative sequence coordinates from constructs used in four different mouse models were retrieved and visualized using the UCSC genome browser (<http://genome.ucsc.edu/>). The results showed that the Bacterial Artificial Chromosome (BAC) used in the current study model terminated prior to a 5' conserved element, composed of four conserved regions (grey rectangle box). This element was present in all three other mouse models generated. An additional 3' conserved element, composed of five conserved

regions (black rectangle box) was also present in the BAC construct used in this study. **B.** The four conserved regions, found in the 5' conserved element, and described in **A** contained an enriched population of binding sites for transcription factors involved in neurogenesis. The transcription factor binding sites are grouped by family for easier representation on this figure.

GO Term	GO ID	Transcription Factors	5' conserved regions analysis		3' conserved regions analysis	
			Bonferroni-Corrected <i>P</i> -value	Randomised <i>P</i> -value	Bonferroni-Corrected <i>P</i> -value	Randomised <i>P</i> -value
Nervous system development	GO:0007399	STAT3 NR2F1 ESR2 SOX2 MEF2A PAX5 FEV PAX2 CREB1 PAX6 TLX1 ZEB1 NHLH1 FOXC1 FOXL1 SOX10 FOXD1 FOXD3 RELA TFAP2A FOXA2 MAFB SOX5 GATA2 NR3C1 NFATC2 E2F1 EN1 HNF1B RXRA PBX1 NKX2-5 FOXA1 NR2E3 SRF SOX9 PPARG NR4A2	9.23094119467758e-12	0.0327	1.23502929422786e-11	0.0220
Central nervous system development	GO:0007417	PAX6 TLX1 ZEB1 NHLH1 FOXC1 FOXL1 ESR2 FOXA2 MAFB SOX2 SOX5 CREB1 GATA2 NR3C1 NFATC2 NR2F1 E2F1 EN1 HNF1B SOX9 SOX10 PPARG	8.4507010705991e-10	0.0147	2.99022615274573e-09	0.0356
Neurogenesis	GO:0022008	STAT3 NR2F1 ESR2 SOX2 PAX2 CREB1 RELA FOXA2 SOX5 GATA2 TLX1 RXRA PBX1 NKX2-5 FOXA1 NR2E3 PAX6 SRF EN1 SOX9 SOX10 PPARG FOXD1 FOXD3 NR4A2	2.63390159936542e-07	0.0122	5.7376060069322e-07	0.0344
Eye development	GO:0001654	PAX6 SOX2 FOXC1 STAT3 FOXL1 YY1 SP1 ZEB1 PAX4 MAX NR2E3	9.45330733200263e-06	0.0300	6.69632064726976e-05	0.2595
Camera-type eye development	GO:0043010	FOXC1 FOXL1 YY1 SP1 ZEB1 PAX4 MAX NR2E3 SOX2	1.18317513845212e-04	0.0473	9.3334459117566e-04	0.4437

Table 3.2 Relevant GO Terms.

The most relevant Gene Ontology (GO) terms that have been found in the over-representation analysis are presented. GO terms are given in the first column with their GO ID in the second column. Gene symbols for the transcription factors that belong to those GO terms and have at least one binding site predicted within the conserved regions are given in the third column. A Bonferroni corrected *P*-value, associated with the over-representation of the corresponding GO term is given in the fourth and sixth columns. The fifth and seventh columns contain the randomized *P*-values obtained after the nucleotide shuffling analysis (Material and Methods).

Family	Transcription Factors
Stat	Stat3
Hormone-nuclear Receptor	NR2F1 Esr2 RXRA Nr2e3 PPARG Nr4a2
High mobility group	Sox2 SOX5 SOX9 Sox10
Homeo	PAX2 Tlx1 Pbx1 NKX2-5 PAX6 EN1
Leucine zipper	CREB1
Rel	RELA
Forkhead	FOXA2 Foxa1 FOXD1 FOXD3
GATA	GATA2
MADS	SRF

Table 3.3 Neurogenesis Transcription Factors Families.

The family classification used in Figure 3.12B is presented. The first column gives the family classification of the different transcription factors (second column) that are associated with the Neurogenesis Gene Ontology (GO) term. Structural classification came from JASPAR (http://jaspar.genereg.net/cgi-bin/jaspar_db.pl?rm=struct_browse).

Similarly, we found enrichment for TFBSs prediction in the five 3' *SNX3*-intronic-conserved regions for all the previously reported GO terms; “nervous system development” (P -value = 1.24×10^{-11}), “central nervous system development” (P -value = 2.99×10^{-09}), “neurogenesis” (P -value = 5.74×10^{-07}), “eye development” (P -value = 6.70×10^{-05}), and “camera-type eye development” (P -value = 9.33×10^{-04}) (Table 3.2). However, only the terms “nervous system development”, “central nervous system development” and “neurogenesis” remained significant after performing the randomised analysis (Table 3.2). In addition, since the *SNX3* gene is also expressed in the central nervous system (Mizutani, Nakamura et al. 2011), these regions may be present to regulate the gene in which they reside, and there is no requirement to hypothesize they regulate *NR2E1*. In contrast, the location proximal to the *NR2E1* promoter supports the involvement of the four 5' conserved regions in *NR2E1* regulation. This allows us to predict that the incomplete *NR2E1* expression and function observed in the brains of our animals was most likely due to the absence of these four 5' conserved regions.

3.4 Discussion

In the current study, we have demonstrated for the first time the expression resulting from the insertion of a BAC carrying the human *NR2E1* gene in both developing and adult mice. We show that human *NR2E1-lacZ* expression mimics the mouse homolog in both developing and adult eyes, resulting in normal retina in the animals harbouring only one functional allele of the human gene, thereby correcting the retina-null phenotype. These results show that the BAC construct used in this study contains all of the elements for appropriate spatial and temporal eye expression of the human gene. Further, the results support the hypothesis that the human gene is functionally equivalent to the mouse gene in the mouse eye, and that the previous incomplete correction of the fierce eye phenotype may indeed be attributed to excess copy number (Abrahams, Kwok et al. 2005). Finally, the functional conservation of *NR2E1* highlights the potential role of this gene in human eye diseases of unknown aetiology.

In stark contrast, our results also show that the BAC used in this study demonstrates an unexpected absence of expression of the human *NR2E1* gene in key neurogenic regions of both the developing and adult brain. The *Nr2e1* mouse gene has been previously shown to regulate the proliferation rate of neural stem cell populations in the developing telencephalon, and the absence of expression of the *Nr2e1* gene results in premature neuronal differentiation, leading to depletion of the neural stem cell populations (Roy, Kuznicki et al. 2004). This affects the development of subsequent forebrain structures such as the upper layer of the cortex, the DG and OB. In this study, animals harbouring the functional *NR2E1* allele generated from the human BAC display a fierce-like phenotype in adult neurogenic brain regions. This phenotype includes morphological defects in the DG of the hippocampus

and SVZ of the lateral ventricles, which correlate with an absence of expression of the BAC *NR2E1-lacZ* reporter gene in proliferating cells in these regions. These results provide a distinct separation of the regulatory mechanisms governing the *NR2E1* gene; leading to an abnormal brain expression pattern while retaining normal cell-specific retinal expression.

In addition to the phenotype found in adult neurogenic brain regions, the *NR2E1/fierce* brain displays an attenuated version of cortex and OB hypoplasia; a hallmark of the *fierce* phenotype. We conclude that this is due to a partial correction of the *fierce* phenotype during embryonic development as the formation of these structures involves the radial migration of neurons from the VZ of the DP, but also the tangential migration of neurons from the VZ of the subpallium (Angevine and Sidman 1961; de Carlos, Lopez-Mascaraque et al. 1996; Anderson, Eisenstat et al. 1997; Tamamaki, Fujimori et al. 1997; Nadarajah, Brunstrom et al. 2001). In our current study, the expression results obtained in the BAC *NR2E1-lacZ* embryos demonstrate absence of expression in the DP of the telencephalon, while preserving expression in the MP, LP/VP and subpallium regions. This suggests that the partial correction found in the adult animals harbouring the current human BAC construct is due to expression of the gene in the LP/VP and subpallium regions during development.

Our hypothesis that the human gene is capable of completely correcting the *fierce* brain phenotype was based primarily on previous published results using a random-insertion multiple-copy mouse model (Abrahams, Kwok et al. 2005). Thus, the current incomplete correction could be due to the single-copy insert or location on the X Chromosome. X-inactivation makes it impossible for us to test two expressed copies of the human BAC in female mice (Liskay and Evans 1980; Heaney, Rettew et al. 2004). However, taking into

consideration the complete absence of expression of the human gene in the mouse DP, we favour an explanation involving missing key regulatory regions. Following this path, lead us to the identification of 5' conserved regulatory regions that are statistically enriched for binding sites of transcription factors involved in neurogenesis. Amongst these, *Pax6* has been shown to genetically interact with *Nr2e1* in the establishment of the pallio-subpallial boundary (Stenman, Yu et al. 2003), and *Sox2* has been shown to play an important role in controlling *Nr2e1* transcription in cultured neural stem cells (Shimozaki, Zhang et al. 2012). Furthermore, *Pax6* and *Sox2* can form a co-DNA binding partner that regulates initiation of lens development (Kamachi, Uchikawa et al. 2001). Thus, we hypothesize that we have identified important regulatory regions that may interact with *Pax6* and *Sox2* to control specific expression of *NR2E1* in both the developing and adult brain.

The findings from the current study are also of critical importance for future human genetic disease studies. The candidate novel regulatory regions found in the current study were not included in previous association studies and deep sequencing analysis of *NR2E1* (Kumar, Everman et al. 2007; Kumar, Leach et al. 2007; Kumar, McGhee et al. 2008). Two of these studies have reported the finding of candidate regulatory mutations in patients with various brain disorders, and some of these mutations were predicted to affect the binding sites of specific regulators involved in brain development (Kumar, Leach et al. 2007; Kumar, McGhee et al. 2008). Hence, we argue that future patient studies, screening for mutations at the *NR2E1* locus, should include these new candidate regulatory regions in their analysis. In addition, *NR2E1* is now a very strong candidate to be involved in human eye disease, a role never examined in patients to date. Importantly, the docking technology used to generate the animals in this study can be reapplied using BAC recombineered to carry candidate human

mutations, including regulatory mutations, to test their ability to correct the fierce phenotype. Finally, this work serves as a paradigm, and can be generalized to the study of other human genes for which a BAC construct can be derived and a mutant mouse phenotype exists.

Chapter 4: Combined Serial Analysis of Gene Expression and Transcription Factor Binding Site Prediction Identifies Novel Targets of *Nr2e1* in Forebrain Development

4.1 Introduction

The proper development of the mammalian neocortex involves a fine tuning between cells intrinsic developmental programs and environmental factors. In this process, neurons acting as the backbone of the neuronal circuitry are generated first. These cells arise from two different brain regions; the dorsal telencephalon, generating cortical excitatory neurons by radial migration, and the ventral telencephalon, giving rise to cortical inhibitory interneurons by tangential migration (Angevine and Sidman 1961; de Carlos, Lopez-Mascaraque et al. 1996; Anderson, Eisenstat et al. 1997; Tamamaki, Fujimori et al. 1997; Nadarajah, Brunstrom et al. 2001). This developmental step, called the neurogenic stage, is followed by the integration of glial cells in the circuitry during the gliogenic stage. In rodents, these phenomena are temporally segregated, with neurons being generated from embryonic days 12 (E12) to E18 and astrocytes appearing at around E18 (Bayer and Altman 1991; Miller and Gauthier 2007). Ultimately, the neocortex will comprise six different radial layers with cell populations having distinct molecular identities (Job and Tan 2003).

The complexity behind this process involves a careful balance between proliferation of neural stem cells (NSC), and the proper differentiation of progenitor cells (PC) depending on the developmental time-point (i.e. neurons, or glia cells). One particular NSC fate determinant, called *Nr2e1* (also known as *Tlx*, *Tll*, and *Tailless*) encodes a transcription factor expressed along the ventricular zone (VZ) of the dorsal telencephalon during development (Li, Sun et al. 2008). Absence of *Nr2e1* expression in null embryos negatively affects the

numbers of progenitor cells (PC) populating the VZ and subventricular zone (SVZ) during development, resulting in reduced thickness of the cortical plate (CP) (Roy, Kuznicki et al. 2004). The reduction in PC populating the VZ is more prominent in the caudal telencephalon whereas the reduction in the SVZ is seen at all rostrocaudal levels during development. This phenomenon ultimately results in defects in later generated structures such as the upper cortical layers (layers II and III), the dentate gyrus and the olfactory bulb in adult brain (Land and Monaghan 2003; Roy, Kuznicki et al. 2004). Premature neurogenesis, affecting the development of the upper cortical layers, is also observed in *Nr2e1*-null embryos from E9.5 to E14.5 (Roy, Kuznicki et al. 2004).

To date, *Nr2e1* gene regulation has been well documented in adult neural stem cells but only a few genes regulated by *Nr2e1* during forebrain development have been identified. In forebrain development, *Nr2e1* has been shown to regulate cell cycle progression via its interaction with the tumor suppressor encoding gene *Pten*, and the cyclin-dependent kinase inhibitor *p21* (Li, Sun et al. 2008). This phenomenon involves a repressive mechanism mediated via the interaction of *Nr2e1* with chromatin modifier proteins such as members of the histone deacetylase family (HDACs), and the demethylase protein *Lsd1* (*Kdm1a*) (Sun, Yu et al. 2007; Yokoyama, Takezawa et al. 2008). Furthermore, the balance between neural stem cell proliferation and differentiation has been demonstrated to be under the control of a regulation loop involving both *Nr2e1*, and a microRNA encoding gene, *mir-9* (Zhao, Sun et al. 2009). In this case, *mir-9* was demonstrated to be able to promote neural stem cell differentiation by directly repressing *Nr2e1* gene expression, whereas *Nr2e1* was demonstrated to promote neural stem cell proliferation by repressing *mir-9*. Moreover, *Nr2e1* has been shown recently to be able to act as a transcriptional activator of the deacetylase gene

Sirt1, which has been demonstrated to have a role in promoting neuronal differentiation (Hisahara, Chiba et al. 2008; Iwahara, Hisahara et al. 2009). These contradictory results suggest a dynamic role of *Nr2e1* in controlling both the fate and proliferation rate of neural stem cells in the developing forebrain. This also highlights the fact that the reductionist approaches used in these studies limit our clear understanding of the role of *Nr2e1* in forebrain development.

To identify biological pathways affected by *Nr2e1* during forebrain development, and identify novel regulatory partners acting with the protein product of this gene, we used an unbiased approach employing both the precision of laser capture microdissection (LCM), and the power of serial analysis of gene expression (SAGE). Our approach involved dissection of the VZ/SVZ of the dorsal-lateral telencephalon from both Wild type (Wt), and *Nr2e1*-null embryos harbouring the fierce mutation (referred here as *Nr2e1^{frc/frc}*) using LCM at embryonic stages that spanned both the neurogenic and early gliogenic stages in neocortex development (E13.5, E15.5, and E17.5) (Young, Berry et al. 2002; Miller and Gauthier 2007). Based on the literature, we generated a binding profile corresponding to the NR2E1 binding site, and used it in combination with our list of genes differentially expressed according to our SAGE results to reveal potential primary targets. A group of genes involved in biological functions related to the development of the nervous system emerged as direct targets of *Nr2e1*. For a subset of genes from the nervous system development list, SOX9, NR2F1, and E2F1 may act as co-interactors of NR2E1. *Nr2e1* may directly regulates *Lhx2*, a gene encoding for a LIM-homeodomain (LIM-HD) transcription factor involved in the patterning of the dorsal telencephalon early in development, and also in the neurogenic to gliogenic switch during hippocampal development (Monuki, Porter et al. 2001; Subramanian,

Sarkar et al. 2011). Our data reinforce that *Nr2e1* is a key intrinsic regulator of neurogenesis during neocortex development, and this mechanism might involve a dynamic balance with specific co-interactors that in turn affect the regulation of *Lhx2*.

4.2 Methods and Materials

4.2.1 SAGE Libraries Generation

Libraries were generated from tissue samples obtained by laser capture microdissections (LCM) of dorsal ventricular/subventricular zones from the telencephalon of *Nr2e1^{frc/frc}*, and Wild-type (Wt) embryos at E13.5, E15.5, and E17.5 as described in the literature (D'Souza, Chopra et al. 2008). Briefly, one embryo per genotype at each time-point was sectioned at 20 µm thickness to generate the LCM tissue samples which were pooled and prepared for RNA extraction using an RNeasy Micro Kit (Qiagen). The LongSAGE-lite method was used to construct the libraries using 15 to 86 ng of high quality RNA from each embryo (Peters, Kassam et al. 1999; Khattri, Delaney et al. 2007; D'Souza, Chopra et al. 2008). Each library was sequenced to a depth of >100,000 raw tags and the processed data is accessible on the Mouse Atlas of Gene Expression project website (<http://www.mouseatlas.org/>) and the NIH SAGEmap data repository (<http://www.ncbi.nlm.nih.gov/projects/SAGE/>) (Lash, Tolstoshev et al. 2000).

4.2.2 SAGE Data Analysis

LongSAGE libraries were analyzed using the DiscoverySpace 4.0 application (<http://www.bcgsc.ca/platform/bioinfo/software/ds>) (Robertson, Oveisi-Fordorei et al. 2007). The library data was electronically filtered based on previously published procedures

(Siddiqui, Khattra et al. 2005; Romanuik, Wang et al. 2009). Briefly, duplicated ditags (identical copies of a ditag) and singletons (tag counted only once) were retained for analysis. Sequence data were filtered for bad tags (tags with one N-base call), and linker-derived tags (artifact tags). Only tags with a sequence quality factor (QF) greater than 99% were included in the analysis, and constituted the useful tags population. Sequence tag comparisons between *Nr2e1^{frc/frc}*, and Wt libraries were performed and a *P*-value cutoff ≤ 0.05 using the Audic-Claverie statistical method was used (Audic and Claverie 1997). LongSAGE tags exhibiting differential expression levels were mapped to transcripts in the NCBI Reference Sequence (Refseq) collection (version 52, released March 8th 2012) and Ensembl gene collection (version 66, released February 2012).

4.2.3 oPOSSUM Promoter Analysis

A pooled list of the RefSeq accession numbers exhibiting differential expression levels from the three different time-points was used to perform an oPOSSUM promoter analysis, and a gene ontology (GO) term enrichment analysis. A modified version of an oPOSSUM promoter analysis was used in our procedure (<http://www.cisreg.ca/oPOSSUM/>) (Ho Sui, Mortimer et al. 2005; Ho Sui, Fulton et al. 2007). Briefly, for each Refseq accession numbers, oPOSSUM automatically retrieved the genomic DNA sequences around any annotated transcription start sites (TSS) in Ensembl (plus 5,000 bps of upstream and downstream non-coding sequence), performed an alignment of the orthologous sequences (human to mouse), and extracted non-coding DNA sequences that are conserved above a predefined threshold (default value, 70%). It then searched the subsequences for matches to a transcription factor binding site (TFBS) profile corresponding to a position frequency matrix

of the NR2E1 binding sites (AAGTCA) developed to represent the binding properties of Nr2e-subfamily transcription factors. This matrix was generated using the MEME motif discovery program applied to a collection of 23 selex-derived binding sites for Nr2e3 and 4 published Nr2e1 binding sites reported in the PAZAR database (Chen, Rattner et al. 2005; Portales-Casamar, Kirov et al. 2007; Bailey, Boden et al. 2009). MEME was applied using the following parameters: “-dna -mod anr -revcomp -minsites 46 -w 7”, which restricted the profile to a width of 7 bps, required at least 46 sites within the sequence collection to generate the final NR2E1 half-site matrix. The list of corresponding genes obtained after this analysis were thought to be enriched in NR2E1 binding sites within their promoter regions. This list was then submitted to a GO term enrichment analysis to look if certain biological process were overrepresented in the data.

4.2.4 GO Term Enrichment Analysis

Refseq accession numbers corresponding to the list of genes predicted to be containing NR2E1 binding sites were submitted to a GO term enrichment analysis using the DAVID service (<http://david.abcc.ncifcrf.gov/summary.jsp>) (Dennis, Sherman et al. 2003; Huang da, Sherman et al. 2009). The Refseq list was first converted to DAVID IDs, using the DAVID knowledgebase, a tool that collects and integrates identifiers from various sources and compare them to more than 40 well-known publicly available categories (Sherman, Huang da et al. 2007). Gene enrichment was evaluated using the mouse genome as a background list with a threshold count of 2 to eliminate orphan genes. List of genes from GO term “Biological process all” with an EASE score value of ≤ 0.1 were retained as an initial screen (Hosack, Dennis et al. 2003). Terms having a significant *P*-value after correction

(Bonferroni, P -value ≤ 0.05) were evaluated based on their relevance to brain development or cell functions.

4.2.5 Clustering

Clustering was performed on the differential tag ratios corresponding to each genes found in the GO terms category “nervous system development” using the Gene Cluster software (de Hoon, Imoto et al. 2004). Hierarchical clustering was used on both the gene lists, and the embryonic stages using a centered correlation (Pearson correlation) with the average linkage clustering option.

4.2.6 Co-Factor Enrichment

Over-representation analysis was performed on TFBSs proximal to NR2E1 dimer sites (AAGTCA, plus 0-8 bps spacers) for genes found in the GO terms “nervous system development”. Sites within 100 bps of NR2E1 dimers were retrieved from the oPOSSUM database. Sites overlapping an NR2E1 dimer were excluded unless they were completely contained within the gap between the two half-sites. Both the NR2E1 sites and proximal sites were retrieved using default oPOSSUM parameters of conservation level (top 30% of conserved regions with a minimum percentage identity of 60%), threshold level (default matrix score threshold of 80%) and, search region level (5,000 bps upstream and downstream of TSS). The analysis was performed against a background of 500 genes selected randomly from the oPOSSUM database. Over-representation results were considered significant based on the results from the Z-score (> 10) and Fisher score (< 0.01) according to the literature (Ho Sui, Mortimer et al. 2005). Candidate transcription factors were scored based on the

evaluation of their expression pattern at E13.5-15.5-18.5 using images from the Allen Mouse Brain Atlas (<http://www.brain-map.org/>) (Lein, Hawrylycz et al. 2007). The expression pattern was scored according to the specificity, and the strength of the expression level along the VZ/SVZ. Transcription factors expression patterns were scored from high (+++), moderate (++) to low (+), where a strong and, restricted expression along the VZ/SVZ was score as +++, and a weak and ubiquitous expression in the entire embryos was scored as +. The transcription factors having a high score (+++) were retain as the most interesting candidates. Further investigation, looking at the number of tags found in the SAGE libraries generated in this project for each candidate transcription factors as well as data mining from the literature was also performed.

4.2.7 Embryos Preparation

All procedures involving animals were in accordance with the Canadian Council on Animal Care (CCAC) and UBC Animal Care Committee (ACC) (Protocol# A11-0412). Time-pregnant mice were euthanized by cervical dislocation, and embryos at E15.5 were dissected and fixed in 4% paraformaldehyde (PFA) with 0.1 M PO buffer (pH 8.0) for 6 hours at 4°C. The embryos were then cryoprotected as described in the literature, and embedded in optimal cutting temperature (OCT) compound (Tissue-tek, Torrance, California) on dry ice (Li, Sun et al. 2008). Embryos were sectioned at 20 µm using a Cryo Star HM550 cryostat (MICROM International, Kalamazoo, Michigan), and mounted for immunofluorescence.

4.2.8 Immunofluorescence, and Imaging Analysis

For antibody staining, 20 μm sagittal cryosections from embryos were rehydrated in subsequent washes of PBS, permeabilized in PBS with 0.3% triton, and blocked with 1% BSA in PBS triton 0.3% for 1 hour at room temperature. Goat anti-Lhx2 primary antibody (Santa Cruz, sc-19344), (1:1,000) was incubated overnight at 4°C. Rabbit anti-goat Alexa 488 (Invitrogen, A21222), (1:1,000) was incubated for 2 hours at room temperature in the dark. Tiled images were retrieved with an Olympus BX61 motorized fluorescence microscope at 20x magnification (Olympus America Inc., Center Valley, PA, USA). Intensity quantification was performed using Image-Pro (Media Cybernetics Inc., Bethesda, MD, USA). The relative intensity level of Lhx2 was calculated as described in the literature (Singaraja, Huang et al. 2011). Briefly, the sum of intensity was divided by the area selected and multiplied by the thickness, and number of sections. A background correction was applied by subtracting the sum of intensity from a section stained using the secondary antibody only. All values represent the mean \pm standard errors of the mean (SEM). Statistical analysis was performed using Student's *t*-test.

4.2.9 Embryonic Stem Cells Culture

Embryonic stem cells (ESC) from *Nr2e1^{frc/frc}*, and Wt blastocysts were derived, and maintained in culture as described in the literature (Yang, Banks et al. 2009). The two cell lines used in this study were mEMS1239 (B6129F1-*Nr2e1^{frc/frc}*, *Hprt1^{b-m3}/Y*), and mEMS1271 (B6129F1-*Nr2e1^{+/+}*, *Hprt1^{b-m3}/Y*). Expansion, and handling of these cell lines was performed as described in the literature (Yang, Banks et al. 2009).

The ESC differentiation procedure involved the use of an adapted method of neurogenesis from adherent monoculture (Gaspard, Bouschet et al. 2008; Gaspard, Bouschet et al. 2009). Briefly, the cells were seeded at low density ($\sim 10,000$ cells/mm²) on gelatin coated dishes, in a chemically defined medium (DDM) exempt of cycloamine, and maintained in culture for 12 days (fresh media changed every two days). RNA aliquots at 12 days of differentiation were retrieved for quantitative RT-PCR (qRT-PCR).

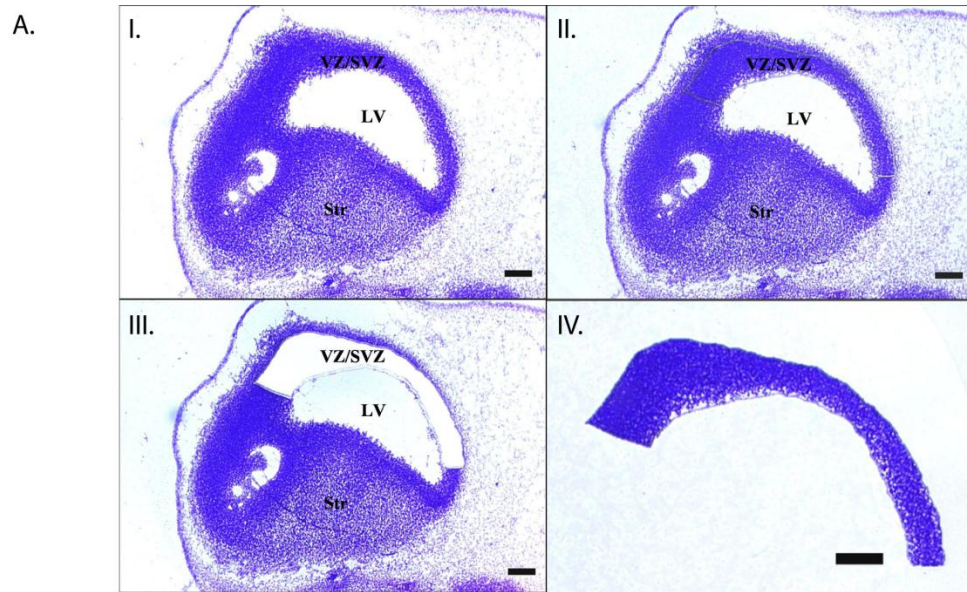
4.2.10 Quantitative RT-PCR

RNA from ESC grown in an adapted method of neurogenesis from adherent monoculture, collected at day 12 of differentiation was extracted using Qiagen RNA Mini Plus kit (Qiagen Inc., Mississauga, Canada). RNA was cleaned with Qiagen DNase kit (Qiagen Inc., Mississauga, Canada), and cDNA generated using Superscript III Master Mix kit (Invitrogen, Carlsbad, USA). cDNA quantification was performed using ABI Taqman® assays specifically designed for *Nr2e1* (Mm00455855_m1), and *Lhx2* (Mm00839783_m1). The 7500 fast real-time PCR system, and Taqman® fast universal PCR Master Mix (Applied Biosystems Inc., Foster city, USA) was used for all the qRT-PCR runs. The cycle threshold (Ct) value was defined as the number of cycles required for the fluorescent signal to cross a threshold above the background signals, and is inversely proportional to the amount of target cDNA. All values represent the mean \pm SEM. Statistical analysis was performed using Student's *t*-test.

4.3 Results

4.3.1 LongSAGE Libraries Generation by Precise Dissection Using Laser Capture Microdissection.

The SAGE libraries used in this manuscript, were obtained by LCM of the VZ/SVZ of the dorsal-lateral telencephalon from *Nr2e1^{frc/frc}*, and Wt embryos at three different developmental time-points (E13.5, E15.5, and E17.5) (Figure 4.1A). The RNA isolated from the corresponding tissue was used to generate SAGE libraries that were sequenced to a depth $\geq 100,000$ tags (total number of tags per libraries, see Figure 4.1B, column five). Useful tags were retained using a filtering procedure involving the DiscoverySpace 4.0 application (<http://www.bcgsc.ca/platform/bioinfo/software/ds>) (filtering details, see Methods and Materials) (Robertson, Oveisi-Fordorei et al. 2007). Details of the number of tags constituting the SAGE libraries used in this manuscript are summarized in Figure 4.1B. On average, $\sim 24\%$ of the total tags per library were discarded in this procedure resulting in a useful tag population averaging $\sim 83,000$ tags per library, and corresponding to $\sim 25,000$ tag types per library (Figure 4.1B, column six, and nine). Singleton tags (tags counted only once) constituted $\sim 18\%$ of the useful tags population per library and $\sim 68\%$ of the tag type population per library (Figure 4.1B, column seven). These numbers were consistent with previously published results, obtained using a similar filtering procedure (Romanuik, Wang et al. 2009). The useful tags population (including singletons) was used in the subsequent analyses.



B.

Library Name	Genotype	Theiler stage (Embryonic stage)	Starting RNA (ng)	Total tags	Total useful tags	Singletons tags	Total tags excluding singletons	Tag types
SM197	<i>Nr2e1^{frclfrc}</i>	TS21 (E13.5)	20	110,020	90,737	15,559	75,178	23,274
SM301	<i>Nr2e1^{frclfrc}</i>	TS23 (E15.5)	86	108,230	80,939	14,784	66,155	21,171
SM168	<i>Nr2e1^{frclfrc}</i>	TS25 (E17.5)	19	108,157	82,799	16,126	66,673	23,497
SM167	Wild type	TS21 (E13.5)	18	105,208	80,618	13,863	66,755	20,625
SM178	Wild type	TS23 (E15.5)	16	117,920	85,077	15,241	69,836	22,591
SM175	Wild type	TS25 (E17.5)	15	105,486	80,227	14,875	65,352	21,413

Figure 4.1 LongSAGE Libraries Obtained by Laser Capture Microdissection were Used to Map the Transcriptome of *Nr2e1^{frclfrc}*, and Wt Embryos

A. Laser capture microdissection (LCM) procedure details. **A-I.** Embryonic day 13.5 sagittal sections were stained with cresyl violet. **A-II.** The ventricular/subventricular zone (VZ/SVZ) of the lateral telencephalon was cut with a laser. **A-III.** The VS/SVZ was removed by LCM. **A-IV.** The VS/SVZ of dorsal telencephalon was captured by LCM for RNA extraction. LV, Lateral Ventricle, Str, striatum. Scale bars, **I** to **IV** 100 μ m. **B.** The LongSAGE libraries composition is presented. The first column gives the name identifying the library in DiscoverySpace. The second and third columns give information regarding the genotype and developmental stage for each library. Column four gives the amount of RNA used as starting material. Column five to nine give information regarding the tag numbers for each library depending on the filtering criteria used.

4.3.2 Integration of Three Different Bioinformatics Tools Used for *Nr2e1* Direct Target Predictions and Functional Gene Classification.

The DiscoverySpace 4.0 application was used to perform statistical analyses on the useful tag populations retained after the initial filtering steps. Tags differentially expressed between *Nr2e1^{frc/frc}*, and Wt libraries at each time-points (E13.5, E15.5, and E17.5), and falling within the confidence interval of 95% ($P \leq 0.05$), according to the Audic-Claverie significance test, were retained for further analyses (Audic and Claverie 1997). The number of tags up or down regulated at each time-point are highlighted in Figure 4.2A, column three. The proportion of up or down regulated tags varied between 15 to 25% when compared to the combined numbers of useful tags found in the *Nr2e1^{frc/frc}*, and Wt library at each time-point. On average, ~ 52% of the up or down regulated tags ($P \leq 0.05$) mapped to genes in both the Ensembl (v66) and Refseq (v52) collection databases (Figure 4.2A, column four). The numbers of Refseq accession IDs corresponding to each up or down regulated tags at the three different time-points are highlighted in Figure 4.2A, column five. These accession numbers, corresponding to Refseq genes, were retrieved and used in the subsequent analyses.

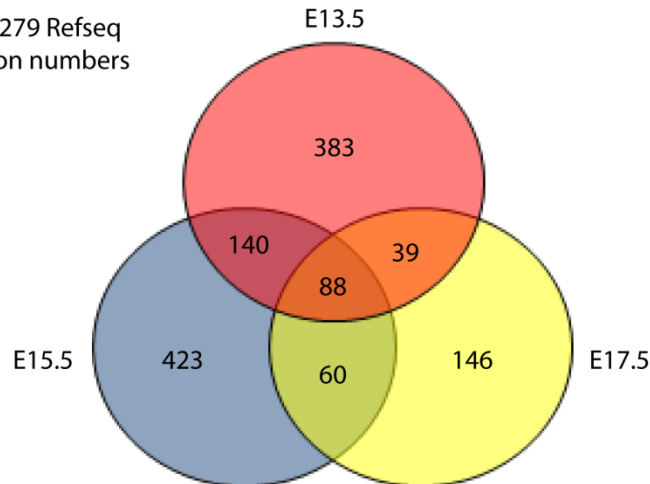
We next looked at the distribution of genes considered to be up or down regulated between the *Nr2e1^{frc/frc}*, and Wt libraries at E13.5, E15.5, and E17.5 time-points. This resulted in a total of 1,279 Refseq accession numbers distributed according to the Venn Diagram found in Figure 4.2B. Interestingly, this distribution revealed that, on average, a number of 6 genes per time-point corresponded to tags that were found in both the up and down regulated populations. This suggested that the tags mapped to these genes were corresponding to alternative transcripts that were expressed in opposing directions when comparing *Nr2e1^{frc/frc}*, and Wt libraries. The distribution results also demonstrated that on

average ~ 53% of the up or down regulated genes are specific for each time-point, and that ~ 31% of the genes overlapped between at least two time-points and only ~ 7% of the genes overlapped between the three time-points. A difference also existed when comparing the total number of up or down regulated genes at each time-point. The combined list of genes up or down regulated at E13.5, and E15.5 was greater in proportion by ~ 49% and ~53% when compared to the one obtained at E17.5. The same type of analysis between the E13.5 and E15.5 time-points revealed only a difference of ~9% between both libraries, suggesting that during development, *Nr2e1* expression has an effect on a greater number of genes in early and mid-stages of neurogenesis (E13.5 and E15.5), than during the switch from neurogenesis to gliogenesis occurring around E17.5. These results correlated with the previously published observations, demonstrating an evolution of the *Nr2e1*-null phenotype during neocortex development, with a greater effect between E13 and E15 (Roy, Kuznicki et al. 2004).

A.

Direction of change	Embryonic days	Tags ($P \leq 0.05$)	Tags mapped to gene	Corresponding number of genes
Up regulated	13.5	35,057	17,691	245
Down regulated	13.5	36,495	16,492	413
Up regulated	15.5	42,166	21,576	240
Down regulated	15.5	31,742	15,506	477
Up regulated	17.5	25,078	15,081	218
Down regulated	17.5	27,631	16,285	119

B. Total: 1,279 Refseq accession numbers



C.

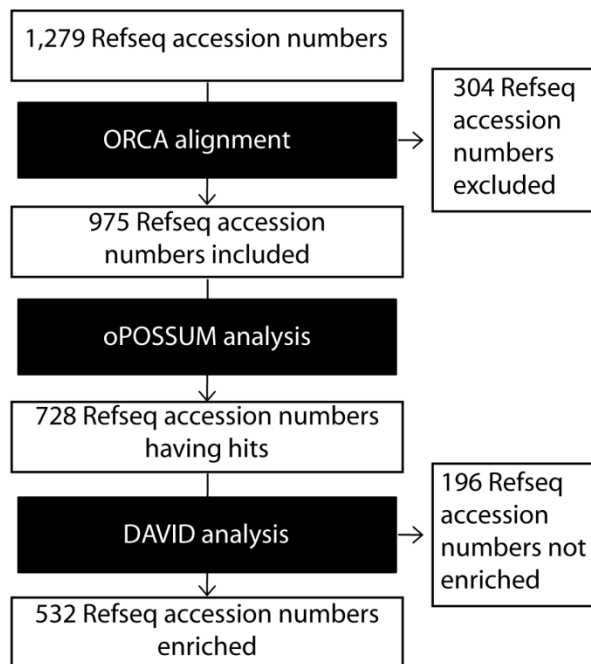


Figure 4.2 An Integrated Approach, Using Three Bioinformatics Tools was Used to Predict Novel Candidate Direct Targets of *Nr2e1*.

A. The number of tags up or down regulated between *Nr2e1^{frc/frc}*, and Wt embryos are presented. Column one presents the direction of change; column two, the corresponding

embryonic days; column three, the number of tags that had significant different counts between *Nr2e1^{frc/frc}*, and Wt embryos; column four the number of tags, having significant different counts that mapped to genes found in the Ensembl (v66) and Refseq (v52) gene collections, and column four highlights the number of Refseq genes mapped by the corresponding tags. **B.** The compiled distribution of up and down regulated Refseq genes at each embryonic day is presented. The Venn diagram presents the number of up and down regulated genes exclusive and shared at each embryonic day (E13.5, E15.5, and E17.5). **C.** Flow chart describing the number of Refseq accession numbers retained for both the oPOSSUM and DAVID analyses. A compiled list of 1,279 Refseq accession numbers, corresponding to differentially regulated genes between *Nr2e1^{frc/frc}*, and Wt embryos from the three different time-points was used in a modified oPOSSUM analysis (details see methods and material). Of this list, 975 Refseq accession numbers were included in the analysis, and 728 Refseq accession numbers were found to have hits for NR2E1 binding sites. The 728 Refseq accession numbers list was submitted to a gene ontology term analysis using DAVID. Of this analysis, 532 Refseq accession numbers were found to be enriched in a category related to biological processes.

The pooled list of 1,279 Refseq accession numbers was used to perform two additional analyses: 1) an *Nr2e1* binding sites prediction analysis, using a modified version of oPOSSUM (<http://burgundy.cmmt.ubc.ca/oPOSSUM/>) (Ho Sui, Mortimer et al. 2005; Ho Sui, Fulton et al. 2007) to unravel genes predicted to have *Nr2e1* binding sites within their promoter regions, and 2) a GO term analysis, using the DAVID service (<http://david.abcc.ncifcrf.gov/summary.jsp>) (Dennis, Sherman et al. 2003; Huang da, Sherman et al. 2009) to evaluate if the resulting gene list can be found in any relevant biological processes.

The results coming from these two additional analyses are summarized in the flowchart of Figure 4.2C. In the modified oPOSSUM analysis, initial orthologous sequence alignments between human and mouse for each corresponding genes were performed using ORCA (Portales-Casamar, Arenillas et al. 2009). In this process, 304 Refseq accession numbers were discarded due to poor conservation between human and mouse, within the promoter sequences of the genes (default value, 70%). This resulted in 975 Refseq accession

numbers that were used in the modified oPOSSUM analysis (Figure 4.2C). Of these 975 accession numbers, 728 (~ 75%) were found to have predicted binding sites for NR2E1 within their promoter regions. These 728 accession numbers were used in a GO term enrichment analysis using the DAVID service. The 728 Refseq accession numbers were first converted to DAVID ID using the DAVID knowledgebase before being compared to the list of genes coming from the mouse background of the DAVID service (Sherman, Huang da et al. 2007; Huang da, Sherman et al. 2009). The enrichment results were visualized using the functional annotation module based on the relevance for each enriched gene to “biological process” with an initial P -value ≤ 0.1 , using the modified fisher exact test (EASE score) (Dennis, Sherman et al. 2003; Hosack, Dennis et al. 2003; Huang da, Sherman et al. 2007). In this process, 196 Refseq accession numbers were discarded as they were deemed not to be enriched in our list in comparison to the mouse background (Figure 4.2C). The remaining 532 Refseq accession numbers were interrogated based on their “biological process” terms. Only terms with a P -value ≤ 0.05 after multiple corrections, using Bonferroni, were considered interesting for further investigation. Table 4.1 highlights the list of GO terms passing this criterion. Numerous terms related to cell cycle regulation were found after performing the GO term enrichment analysis on the 728 Refseq list but were also found in a similar analysis, using the initial 1,279 Refseq list as a control. This suggested that these genes, involved in cell cycle regulation, were differentially expressed in our SAGE results when comparing *Nr2e1^{frc/frc}* to Wt, but were most likely not significantly enriched for the presence of NR2E1 binding sites within their promoter regions. The term, “nervous system development” ($P < 0.01$), with 63 genes associated was found enriched only after performing the analysis on the 728 Refseq list, suggesting enrichment for the presence of NR2E1 binding

sites within the promoter regions of genes associated to this term (see Table 4.1 highlighted in green). We used the genes coming from this subsequent list for further investigations.

Term	Count	Bonferroni
GO:0006396~RNA processing	49	1.65E-07
GO:0016070~RNA metabolic process	62	6.97E-07
GO:0044267~cellular protein metabolic process	132	2.92E-05
GO:0015031~protein transport	56	1.04E-04
GO:0045184~establishment of protein localization	56	1.33E-04
GO:0051246~regulation of protein metabolic process	36	5.86E-04
GO:0046907~intracellular transport	41	5.94E-04
GO:0019538~protein metabolic process	149	1.86E-03
GO:0000280~nuclear division	24	2.10E-03
GO:0007067~mitosis	24	2.10E-03
GO:0030163~protein catabolic process	47	2.17E-03
GO:0044265~cellular macromolecule catabolic process	50	2.21E-03
GO:0007399~nervous system development	63	2.98E-03
GO:0000087~M phase of mitotic cell cycle	24	3.01E-03
GO:0000279~M phase	29	9.19E-03
GO:0006886~intracellular protein transport	28	1.64E-02
GO:0022403~cell cycle phase	31	2.08E-02
GO:0034613~cellular protein localization	29	2.55E-02
GO:0022900~electron transport chain	16	3.79E-02

Table 4.1 Gene Ontology Term Analysis Revealed Enrichment in Relevant Biological Processes.

The gene ontology (GO) term enrichment results are presented. The DAVID service (<http://david.abcc.ncifcrf.gov/summary.jsp>) was used to perform a GO term enrichment analysis for the category “biological process 4” on the list of 728 Refseq accession numbers obtained after performing the modified oPOSSUM analysis. The first column presents the identifiers and the terms related to the biological processes. The second column presents the number of genes, corresponding to the associated GO term categories that were counted from the initial submitted list. The third column present the *P*-values obtained for each terms after multiple correction (Bonferroni). The term highlighted in green represent the term used for further analyses.

4.3.3 LongSAGE Tags Expression Evaluation Results Suggested Distinct Roles for *Nr2e1* in Different Stages of Neocortex Development.

To understand the evolution of the role of *Nr2e1* on gene expression during neocortex development, we performed hierarchical clustering on the tag ratio values corresponding to each gene found in the “nervous system development” GO term category. The tag sequences, and corresponding tag numbers of the genes were retrieved for each SAGE libraries using the DiscoverySpace 4.0 application. Fold changes from tags statistically differentially expressed at least at one time-point between the *Nr2e1*^{frc/frc} and Wt libraries were retrieved, and hierarchical clustering was performed using the Gene Cluster software as described in Methods and Materials (de Hoon, Imoto et al. 2004). The clustering results were visualized in a heatmap display, using Java TreeView (Figure 4.3) (Saldanha 2004). The results demonstrated that the E13.5, and E15.5 time-points differential tag ratios clustered positively (* $r = 0.35$), whereas comparing these two time-points to E17.5 yielded a negative clustering value (** $r = -0.23$). This suggested that the differential tag ratio found between the E13.5, and E15.5 libraries were more similar than the one observed in the E17.5 library. This highlights the possibility of a distinct role for *Nr2e1*, separating the neurogenesis, and early gliogenesis stages during neocortex development.

Additionally, *Nr2e1* was found to be significantly downregulated at E13.5 when compared to Wt (-4.5 fold, $P < 0.05$), with 4 tags detected in the Wt library, and 0 tags detected in the *Nr2e1*^{frc/frc} library. This *Nr2e1*^{frc/frc} and Wt library comparison for *Nr2e1* validated our approach, and highlighted the low abundance of *Nr2e1* transcripts suggesting expression in a restricted cell population. The tag ratio results observed in the libraries corresponding to the E13.5, and E17.5 time-points highlighted a decline in *Nr2e1* expression

level in Wt embryos, culminating to 0 tags in both *Nr2e1*^{frc/frc} and Wt libraries at E17.5. These results correlated with previously reported expression analysis obtained using *in situ* hybridization (Monaghan, Grau et al. 1995). Interestingly, our prediction approach suggested enrichment for NR2E1 binding sites within the promoter regions of the *Nr2e1* gene. The enrichment of these predicted binding sites correlated with previously published observations demonstrating a self-regulating mechanism of *Nr2e1* (Zhao, Sun et al. 2009; Shimosaki, Zhang et al. 2012). *Nestin* (*Nes*), a common marker of progenitors cells appeared to be significantly downregulated in *Nr2e1*^{frc/frc} at E13.5 (-7.3 fold, $P < 0.05$) (Lendahl, Zimmerman et al. 1990; Reynolds, Tetzlaff et al. 1992). Our results not only correlated with the previously published observation in which a reduction of *Nestin* positive cells was found along the VZ of *Nr2e1*-null embryos at E14.5 but also suggested that this reduction might be due to a direct mechanism of regulation from *Nr2e1* (Li, Sun et al. 2008). Finally, two genes (*Neurod2*, *Sox11*) had more than one corresponding tag, suggesting the presence of alternative transcripts for these genes (Siddiqui, Khattra et al. 2005).

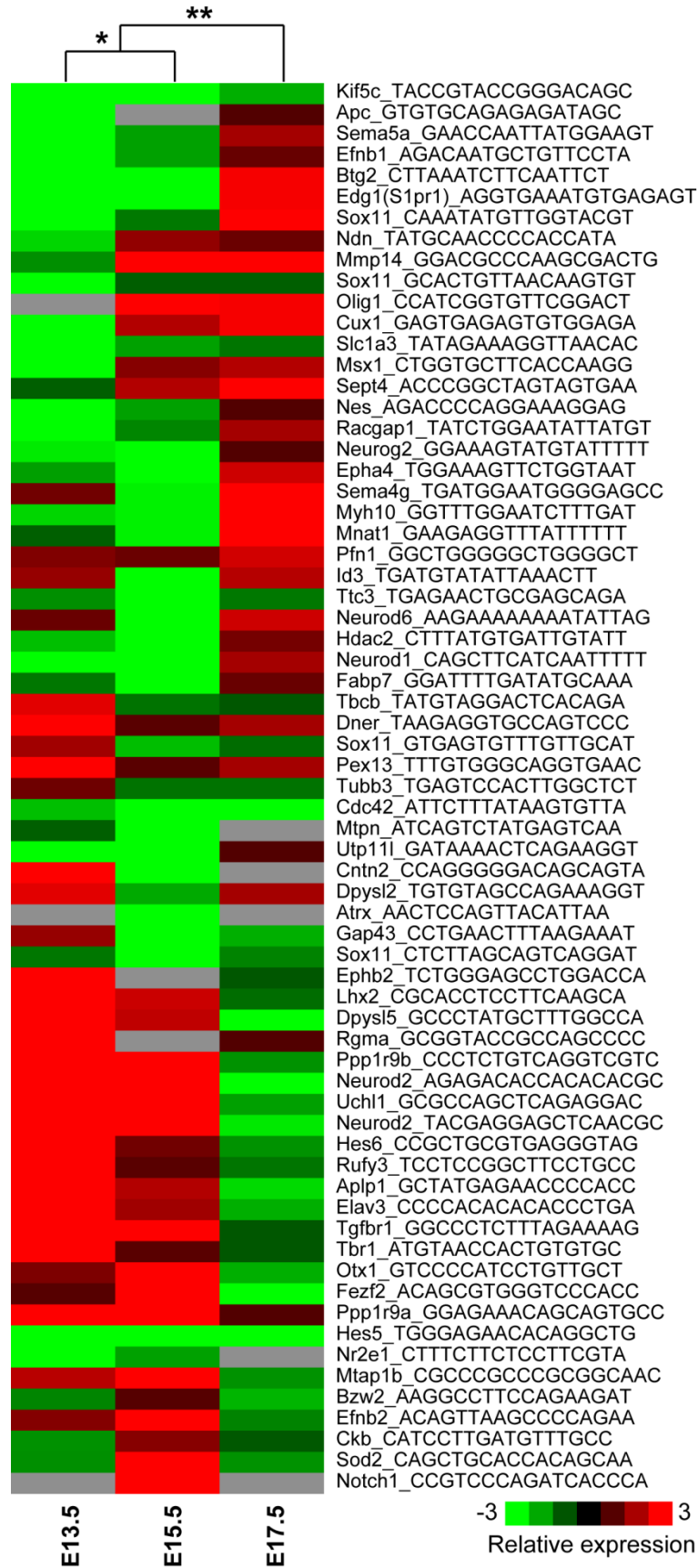


Figure 4.3 Gene Ontology Term Analysis Revealed Enrichment in “Nervous System Development” Biological Process Category.

Heatmap displaying differential tag ratios from the enriched GO term categories are presented. “Nervous system development” category was found to be highly relevant to the role of *Nr2e1* in brain development. Tag numbers for each gene were retrieved from DiscoverySpace and visualized in a heatmap displaying the differential tag ratios from *Nr2e1^{frc/frc}* vs Wt libraries. The fold changes were corrected according to the library sizes; (observed tag counts/total useful tags) X 100,000. Green: negative (low); Red: positive (high); black: no differences; grey: no expressed tags.

4.3.4 Transcription Factor Binding Sites Overrepresentation Analysis Revealed Novel Candidate NR2E1 Co-Interactors.

The tag ratio clustering results obtained across the three different time-points prompted us to evaluate the potential co-interactors of NR2E1 that could influence the underlying effect of *Nr2e1* on gene expression regulation. Transcription factors tend to interact together by forming regulatory complexes that binds specific DNA sequences. Hence we designed an experiment looking at identifying transcription factor binding sites within the vicinity of the predicted NR2E1 binding sites for each gene found in the GO term category “nervous system development”. The identified binding sites were then scored for their enrichment with a randomized list of genes used to compile both a Z-score and a Fisher score. Potential transcription factor binding sites having a Z-score value > 10, and a Fisher score value < 0.01 were considered enriched and kept for further characterization (Table 4.2).

We next looked at the relevance of these predicted enriched transcription factor binding sites according to the expression pattern of their corresponding genes using publicly available data from the Allen Mouse Brain Atlas (ABA) (<http://www.brain-map.org/>) at three different time-points (E13.5, E15.5, and E18.5) (Lein, Hawrylycz et al. 2007). The results suggested that the transcription factor binding sites associated to the member of the SRY-box family, SOX9, the transcription factor E2F1, and the nuclear receptor NR2F1, could have

TF	Background hits	Background non-hits	Target hits	Target non-hits	Z-score	Fisher score	Expression E13.5	Expression E15.5	Expression E18.5	Evidence
Gfi	77	423	24	39	18.45	4.32E-05	N/A	N/A	N/A	N/A
SRY	116	384	30	33	17.29	6.60E-05	Weak	Weak	Weak	+
Myb	74	426	23	40	23.24	7.08E-05	Weak, VZ/SVZ;	Weak, VZ/SVZ;	Weak	+
ZNF354C	141	359	33	30	18.82	1.33E-04	N/A	N/A	N/A	N/A
Lhx3	28	472	13	50	23.14	1.88E-04	Weak	Weak	Weak	+
SPIB	166	334	36	27	13.65	2.14E-04	Moderate, Ubiquitous	Moderate, Ubiquitous	Weak	+
SOX9	64	436	20	43	18.83	2.45E-04	Moderate, VZ/SVZ	Moderate, VZ/SVZ	Strong, VZ/SVZ	+++
Nkx2-5	139	361	32	31	28.66	2.54E-04	N/A	N/A	N/A	N/A
ELF5	107	393	27	36	14.15	2.99E-04	Weak, Ubiquitous	Weak, Ubiquitous	Weak, Ubiquitous	+
Cebpa	55	445	18	45	13.82	3.48E-04	Weak	Weak	Weak, Neocortex	+
Bapx1	96	404	25	38	20.08	3.76E-04	Weak	Weak	Weak	+
Pdx1	137	363	31	32	16.13	4.72E-04	Weak	Weak	Weak	+
Foxa2	63	437	19	44	20.12	5.53E-04	Weak, Ubiquitous	Weak, Ubiquitous	Weak, Ubiquitous	+
E2F1	13	487	8	55	12.18	9.96E-04	Strong, VZ/SVZ	Moderate, VZ/SVZ	Moderate, VZ/SVZ	+++
ELK1	56	444	17	46	15.89	1.12E-03	Weak, Ubiquitous	Weak, Ubiquitous	Weak, Ubiquitous	+
Prrx2	131	369	29	34	18.22	1.20E-03	Weak	Weak	Weak	+
Foxd3	68	432	19	44	19.17	1.26E-03	Moderate	Weak	Weak	+
ZEB1	167	333	34	29	19.65	1.29E-03	N/A	N/A	N/A	N/A
Arnt-Ahr	81	419	21	42	19.74	1.49E-03	Weak, Ubiquitous	Moderate, Ubiquitous	Moderate, Ubiquitous	++
MEF2A	37	463	13	50	11.34	1.66E-03	Weak	Weak	Moderate, Thalamus	+
NHLH1	19	481	9	54	11.9	1.93E-03	Strong, Neocortex	Weak, Neocortex	Weak	++
Nobox	103	397	24	39	21.92	2.26E-03	Moderate, Ubiquitous	Moderate, Ubiquitous	Weak	++
Myf	29	471	11	52	10.07	2.45E-03	N/A	N/A	N/A	N/A
MZF1_5-13	61	439	17	46	19.85	2.54E-03	N/A	N/A	N/A	N/A
NF-kappaB	21	479	9	54	14.35	3.32E-03	Moderate, Ubiquitous	Moderate, Ubiquitous	Moderate, Neocortex	++
PBX1	13	487	7	56	10.12	3.78E-03	Failed QC	Failed QC	Failed QC	N/A
Lhx3	70	430	18	45	12.64	4.00E-03	Weak	Weak	Weak	+
RXRA-VDR	1	499	3	60	22.98	4.94E-03	N/A	N/A	N/A	N/A
NKX3-1	67	433	17	46	16.79	6.00E-03	Moderate, Ubiquitous	Weak, Ubiquitous	Weak	+
NR2F1	11	489	6	57	17.01	7.14E-03	Strong, VZ/SVZ	Strong, VZ/SVZ	Strong, VZ/SVZ	+++
YY1	102	398	22	41	12.43	8.92E-03	N/A	N/A	N/A	N/A
FOXI1	64	436	16	47	12.81	8.92E-03	Moderate	Weak	Weak	+

Table 4.2 Overrepresentation Analysis Revealed Candidate Proximal Co-Interactors of *Nr2e1*.

Cofactor overrepresentation results are presented for gene ontology (GO) term “nervous system development” analysis. Column one presents the transcription factors name. Column two and three, present the details of the number of hits found in the background list of 500 genes. Column four and five present the details of the number of hits found in the list for the GO term “nervous system development”. Column six and seven present the Z-score and Fisher score associated with the

corresponding transcription factors, and columns eight to ten summarize the expression results coming from the Allen Mouse Brain Atlas (ABA). Column eleven highlights the score based on the evidence from the expression pattern where “+++” represent a strong evidence, and “+” represent a weak evidence (details see Methods and Material). Green highlights the most relevant transcription factors according to both the statistical scores and expression pattern. N/A means no expression data were available from the ABA.

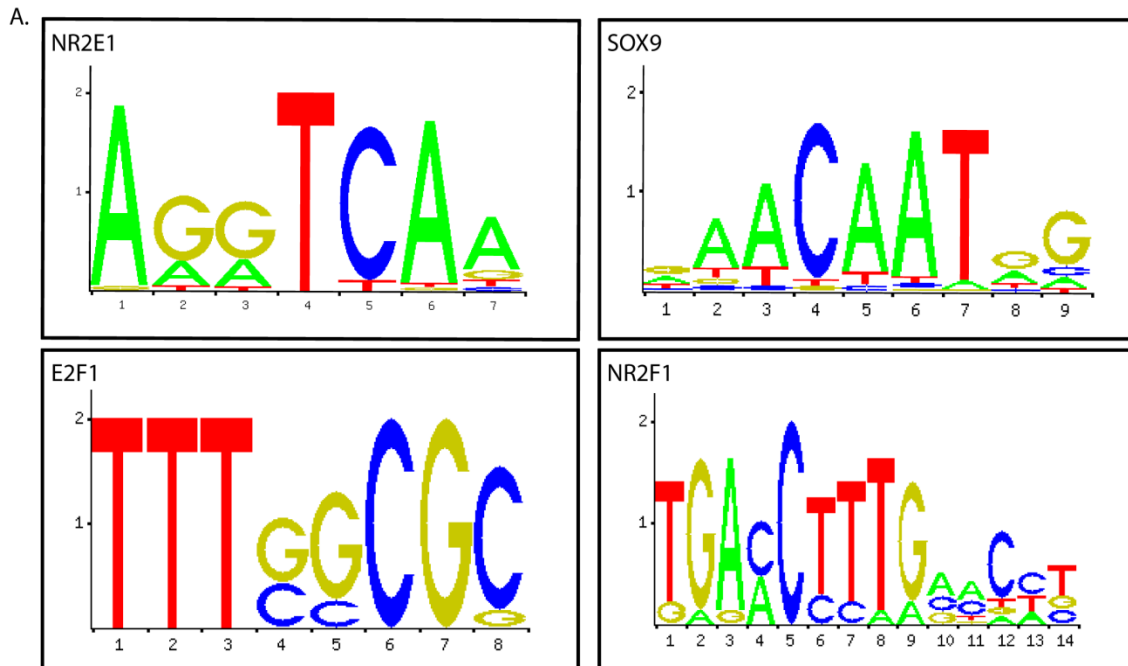
biological relevance as the expression pattern of their corresponding genes are found within the VZ/SVZ during development (moderate to strong expression, Table 4.2). These were compared with expression results obtained using the SAGE libraries generated for this project that demonstrated the presence of tags corresponding to these genes throughout the three different time-points for both genotypes (data not shown). Additionally, all of these transcription factors have been reported to have a role in neural stem/progenitor cells regulation, strengthening their potential role as co-interactors of NR2E1 (Cooper-Kuhn, Vroemen et al. 2002; Naka, Nakamura et al. 2008; Scott, Wynn et al. 2010). The corresponding genes and numbers of predicted transcription factor binding sites for NR2E1, SOX9, E2F1, and NR2F1 were retrieved, generating a list of 27 candidate genes (Table 4.3). Amongst these genes, the most interesting candidate was *Lhx2*, a gene encoding for a transcription factor that has ten NR2E1, twelve SOX9, one E2F1, and two NR2F1 predicted binding sites within its promoter regions (Table 4.3). Visualization of the distribution of the sites within the promoter regions of *Lhx2* revealed clustering of the predicted NR2E1 binding sites with SOX9, E2F1, and NR2F1 binding sites in highly conserved regions, suggesting a functional role for these sites (Figure 4.4A, Figure 4.4B). Evidences from the literature highlight a dynamic role for *Lhx2* in the developing forebrain of which the function appears to depend both on the developmental stage, and the regional specificity of expression. Early in development (E10.5-E11.5), *Lhx2* has been shown to work as a fate determinant of cortical identity in the developing forebrain while at later time-points (E14.5-E15.5), a role for this gene in the neurogenic to gliogenic switch in hippocampal development has been demonstrated (Chou, Perez-Garcia et al. 2009; Subramanian, Sarkar et al. 2011). Precocious

neurons formation in *Nr2e1*-null embryos from E9.5 to E14.5 has been reported, suggesting that *Nr2e1* could indirectly regulate neurogenesis via *Lhx2* (Roy, Kuznicki et al. 2004).

Gene name	NR2E1 binding sites	SOX9 binding sites	E2F1 binding sites	Nr2F1 binding sites
Fezf2	1	1	0	0
Epha4	1	1	0	0
Apc	1	1	0	0
Hes6	1	1	0	0
Tbr1	2	2	0	1
Mtap1b	1	1	0	0
Rgma	1	2	0	0
Utp11l	1	1	0	0
Otx1	1	1	0	0
Myh10	1	1	0	0
Neurog2	2	1	0	0
Msx1	2	2	1	0
Neurod1	1	2	0	0
Cntn2	1	1	0	0
Lhx2	10	12	1	2
Ppp1r9a	1	1	0	0
Nr2e1	3	2	0	1
Atrx	2	1	1	0
Ephb2	4	3	2	0
Cux1	4	5	0	0
Gap43	1	0	1	0
Bzw2	1	0	1	0
Elavl3	1	0	1	0
Tgfbr1	1	0	1	0
Sema4g	1	0	0	1
Sema5a	1	0	0	1
Efnb1	2	0	0	1

Table 4.3 Overrepresentation Analysis Gene List Revealed Candidate Direct Target Gene of *Nr2e1*.

Gene list having hits for NR2E1, SOX9, E2F1, and NR2F1 binding sites is presented. Column one presents the gene name, column two to five presents the number of binding sites for each transcription factors. Green highlights the most interesting candidate for biological validation.



B.

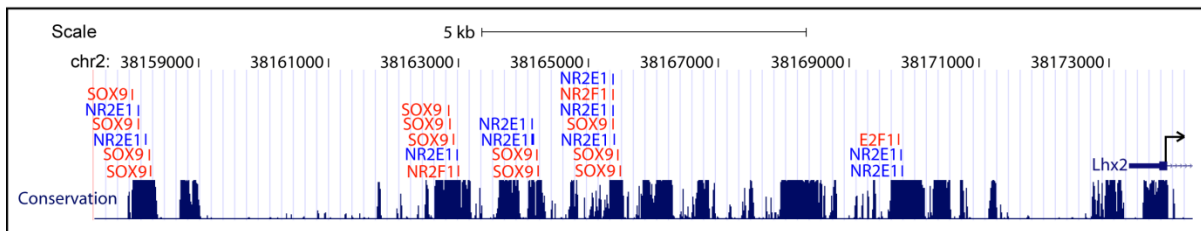


Figure 4.4 *Lhx2* Contains Enriched Overlapping Clusters of NR2E1 Binding Sites in Highly Conserved Regions.

A. The logos, representing the NR2E1, SOX9, E2F1, and NR2F1 binding matrices used in the cofactor overrepresentation analysis are presented. **B.** *Lhx2* was demonstrated to be an ideal candidate gene for binding sites prediction validation due to numerous overlapping clusters of NR2E1-SOX9 binding sites in highly conserved regions. Two predicted NR2F1, and one predicted E2F1 binding site found within the vicinity of predicted NR2E1 binding sites are also presented. The NR2E1 binding sites are highlighted in blue, and the other transcription factor binding sites are highlighted in red.

4.3.5 Transcription Factor Gene *Lhx2* Differential Expression Revealed Accurate Prediction Using our Bioinformatics Approach.

We sought to investigate the accuracy of the prediction approach used in this study by validating one of the potential direct targets of *Nr2e1*. The LongSAGE tag sequence, mapping to the *Lhx2* gene was first retrieved using the DiscoverySpace 4.0 application

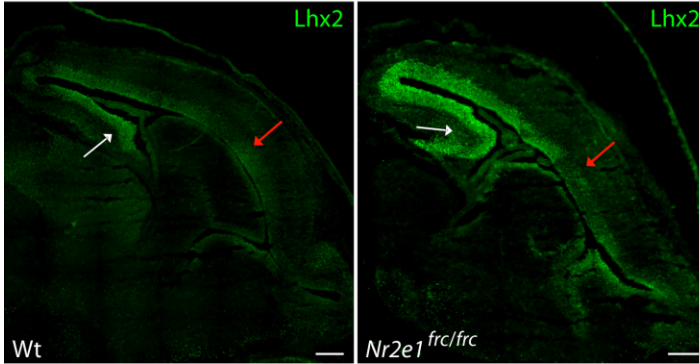
(Figure 4.5A). The corrected number of tags mapping to this gene were also retrieved from each library, and showed that *Lhx2* levels appeared to be significantly increased in *Nr2e1^{frc/frc}* libraries, at two different time-points (E13.5, and E15.5) (Figure 4.5A). We next looked at the expression pattern of the Lhx2 protein by immunofluorescence in both *Nr2e1^{frc/frc}*, and Wt E15.5 embryos. The results suggest a similar expression pattern for Lhx2 when comparing *Nr2e1^{frc/frc}*, and Wt embryos along the VS/SVZ of the developing forebrain, with expression level, extending from high in the medial pallium to low in the dorsal pallium (Figure 4.5B). Relative quantification of the Lhx2 level along the VZ/SVZ of the dorsal-lateral telencephalon revealed a significant increase in the *Nr2e1^{frc/frc}* embryos when compared to Wt ($P < 0.01$) (Figure 4.5C). These results demonstrated that the significant increase at the mRNA level for the *Lhx2* gene obtained using the SAGE approach also results in a significant increase at the protein level along the VS/SVZ of the dorsal-lateral telencephalon at E15.5. A method of neurogenesis from adherent-monoculture of embryonic stem cells (ESC) was also used to perform quantitative RT-PCR investigations (Gaspard, Bouschet et al. 2009). Using this method, we were able to demonstrate expression of the *Nr2e1* gene at 12 days of differentiation (d12) (Figure 4.5D) (Gaspard, Bouschet et al. 2008). Quantitative RT-PCR performed using a TaqMan[®] specific assay corresponding to *Lhx2* mRNA demonstrated a significant increase in the expression level of *Lhx2* in *Nr2e1^{frc/frc}* ESC at d12 of differentiation when compared to Wt ESC ($P < 0.01$) (Figure 4.5E). These results corroborated with the SAGE and immunofluorescence quantification, arguing in favour of the fact that *Nr2e1* might directly regulate *Lhx2* expression in the dorsal-lateral telencephalon during development.

A.

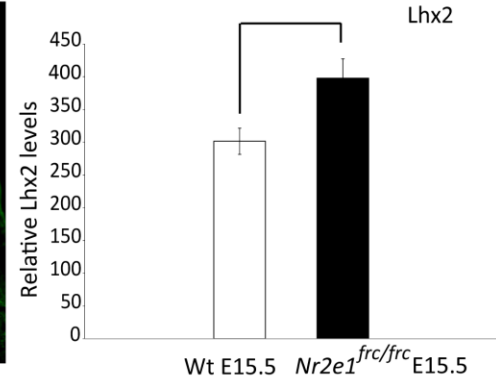
LongSAGE tag Sequence	Refseq accession number	Gene symbol	Embryonic days	*Tag numbers (Wt)	*Tag numbers (<i>Nr2e1^{frc/frc}</i>)	Fold change	P-value
			13.5	11	36	3.3	7.20E-04
CGCACCTCCTTCAAGCA	NM_010710	<i>Lhx2</i>	15.5	13	31	2.4	1.32E-02
			17.5	6	5	-1.3	7.15E-01

* Tag numbers were corrected as follow: (observed tag count/useful tags in library) X 100,000

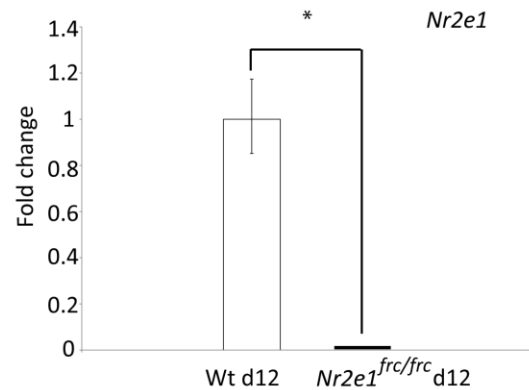
B.



C.



D.



E.

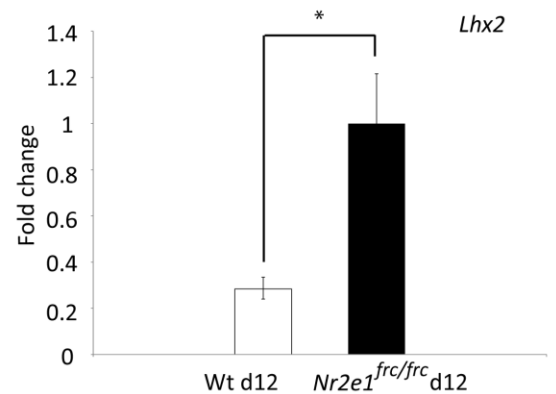


Figure 4.5 *Lhx2*, a Novel Target Gene of *Nr2e1* is Upregulated in *Nr2e1^{frc/frc}* when Compared to Wt.

A. The tag count results, mapping to *Lhx2* at the three different embryonic days are presented. Columns one to three present the tag sequence, the accession number, and the gene symbol corresponding to *Lhx2*. Columns four to six present the corrected tag numbers recovered from DiscoverySpace in both *Nr2e1^{frc/frc}* and Wt embryos at each time-point (E13.5, 15.5, and 17.5). Column seven present the fold change between the tag numbers corresponding to *Lhx2* found in *Nr2e1^{frc/frc}* and Wt embryos at each time-point. Column eight present the associated *P*-values obtained using the Audic-Claverie statistical method. According to this approach, *Lhx2* expression level is significantly upregulated at both E13.5 and E15.5 in *Nr2e1^{frc/frc}* embryos. **B.** Immunofluorescence using an anti-*Lhx2* antibody (green) demonstrated similar expression pattern for the *Lhx2* protein along the ventricular/subventricular zone (VZ/SVZ) of the lateral telencephalon in E15.5 *Nr2e1^{frc/frc}* embryos when compared to Wt. White arrows, medial pallium, red arrows, dorsal pallium (scale bar, 200 μ m) **C.** *Lhx2* level was increased by ~ 1.3 folds along the VZ/SVZ of the

lateral telencephalon in E15.5 *Nr2e1^{frcl/frc}* embryos when compared to Wt ($*P < 0.01$). **D.** Wt embryonic stem cells (ESC) submitted to a neurogenic differentiation protocol demonstrated expression of *Nr2e1* at 12 days of differentiation (d12) whereas *Nr2e1^{frcl/frc}* ESC did not express *Nr2e1* ($*P = 0.000$). **E.** *Lhx2* mRNA level is up regulated by ~ 3.6 folds in *Nr2e1^{frcl/frc}* ESC at d12 when compared to Wt ESC ($*P < 0.01$). **C, D, E.** Two Sample Student's *t*-test was performed on N = 3 samples. Error bars represent standard errors of the mean.

4.4 Discussion

In this manuscript, we report for the first time an approach integrating three different bioinformatics tools to unravel novel direct target genes affected by the neural stem cell fate determinant *Nr2e1*. We first evaluated the transcriptome of *Nr2e1^{frcl/frc}*, and Wt embryos by comparing SAGE libraries generated through LCM of the VZ/SVZ from the dorsal-lateral telencephalon. To better understand the biological role of *Nr2e1* during neocortex development, we chose two time-points that spanned the early to mid- neurogenesis stages (E13.5, E15.5), and one time-point corresponding to the early switch from neurogenesis to gliogenesis (E17.5). *Nr2e1* encodes an orphan nuclear receptor known to act as a transcription factor, recognizing the canonical DNA sequence AAGTCA (Yu, McKeown et al. 1994). Based on this, we developed a binding matrix predicted to represent the binding properties of the Nr2e-subfamily of transcription factors. We used this matrix in combination with the SAGE results in a high-throughput experiment, looking at identifying novel direct target genes of *Nr2e1*. The resulting list of GO terms coming from this analysis was predicted to contain genes differentially regulated by SAGE comparison that were enriched for NR2E1 binding sites within their promoter regions. The GO terms list obtained was shown to contain 63 genes overrepresented in the biological process “nervous system development”, a term that was not enriched when using only the genes coming from the SAGE analysis, suggesting a direct implication for *Nr2e1* in regulating this biological

process. A co-factor analysis performed on these genes, revealed enrichment for binding sites predicted to be bound by the member of the SRY-box family, SOX9, the transcription factor E2F1, and the nuclear receptor NR2F1 within the vicinity of the predicted NR2E1 binding sites. This suggested potential co-interactions between these transcription factors and the NR2E1 protein product in regulating the expression of these 63 genes involved in nervous system development.

A careful review of the numbers of tags found at each time-point suggested that only a small number of cells obtained by LCM expressed *Nr2e1*. In fact, E13.5, the developmental time-point known to be the peak of expression of *Nr2e1*, was the only time-point demonstrating significant down regulation of the *Nr2e1* gene when comparing the *Nr2e1^{frc/frc}* and Wt libraries. To our surprise, *Pten* and *P21 (Cdkn1a)*, two genes demonstrated to be direct targets of *Nr2e1* were not found differentially regulated in *Nr2e1^{frc/frc}* libraries when compared to Wt using our SAGE approach (Sun, Yu et al. 2007; Li, Sun et al. 2008; Yokoyama, Takezawa et al. 2008). *Nr2e1* has been shown to act as a repressor on these genes, and this effect is mediated by the interaction with the histone demethylase *Lsd1 (Kdm1a)*, and the histone deacetylase *Hdac5* (Sun, Alzayady et al. 2010). A review of the number of tags for both these histone modifying enzymes in our libraries suggested normal expression at the three different time-points for both genotypes. Thus, we hypothesize that the absence of significant differences in expression for both *Pten*, and *P21* in the *Nr2e1^{frc/frc}* libraries when compared to Wt is due to the detection efficiency of the SAGE approach.

Nevertheless, our SAGE analyses demonstrated that for genes expressed at mid to high levels such as *Nestin*, a common marker of proliferating neural progenitors, accurate recapitulation of previously published results was possible. In our case *Nestin* appeared to be

down regulated in *Nr2e1^{frc/frc}* at E13.5 when compared to Wt. These results correlated with previously published observation in which a reduction in the number of *Nestin*-positive cells were observed in the VZ of *Nr2e1*-null embryos at E14.5 (Li, Sun et al. 2008). Our prediction approach suggests that this phenomenon might involve direct regulation from *Nr2e1* as *Nestin* was found within the list of genes enriched for predicted NR2E1 binding sites. Previously published observations also revealed that co-expression of Nr2e1 protein occurred in *Nestin*-positive cells along the VZ of E14.5 embryos (Li, Sun et al. 2008). This argues in favor of the possibility of a direct interaction of *Nr2e1* in *Nestin* gene regulation.

Our approach revealed a potential distinct role for *Nr2e1* during neocortex development. Analyses performed on the differential tag ratio for the genes found in the GO term category “nervous system development”, and retrieved from the *Nr2e1^{frc/frc}*, and Wt libraries revealed positive correlation at E13.5, and E15.5 whereas negative correlation was observed when compared to E17.5. This suggested that the differential tag ratio found at E13.5, and E15.5 were more similar than when compared to E17.5. From E13.5 to E17.5, the neocortex undergoes drastic changes, including the formation of the SVZ, a layer of cells being fed by the VZ, and a switch from neurogenesis to gliogenesis also starts to occur. Our results tend to highlight that *Nr2e1* appears to have a greater effect in the early stages of neurogenesis (E13.5, and E15.5) than where the switch from neurogenesis to gliogenesis start to appear. These results correlate with expression evidences from the literature, where the peak of expression of *Nr2e1* has been reported at E13.5, and gradually decreases until birth (Monaghan, Grau et al. 1995). In *Nr2e1*-null embryos, premature neurogenesis has been reported to occur from E9.5 to E14.5 in both the dorsal and ventral telencephalon (Roy, Kuznicki et al. 2004). Our prediction approach suggested that *Lhx2*, a gene encoding for a

transcription factor involved in neurogenesis might be regulated via an *Nr2e1* dependent regulation mechanism. Overexpression of *Lhx2* has been reported to prolong neurogenesis in hippocampal development, resulting in generation of neurons from progenitors that would normally produce astrocytes (Subramanian, Sarkar et al. 2011). However, inactivation of this gene in neocortical progenitors does not seem to affect the fate of these cells, suggesting a regional effect of *Lhx2*. Hence, the premature neurogenesis phenotype observed in the neocortex of *Nr2e1*-null embryos cannot solely be explained by the differential regulation of *Lhx2*. Presumably, the concerted effect of deregulation of other transcription factors having roles in neocortex neurogenesis would be required to explain this phenotype. Transcription factors such as *Neurog2*, *Tbr1*, and *Fezf2* have already been reported to be involved in regulating neurogenesis in the central nervous system (CNS) and are found to be differentially regulated in our SAGE analysis (Schuurmans, Armant et al. 2004; Roybon, Deierborg et al. 2009; McKenna, Betancourt et al. 2011; Mendez-Gomez, Vergano-Vera et al. 2011). The implication of these transcription factors in regulating neurogenesis in *Nr2e1*^{frc/frc} animals during neocortex development will require further validation.

Our analysis predicted that the transcription mechanism regulated by NR2E1 involves the possible interaction with SOX9, E2F1, and NR2F1 co-interactors. Two of these candidate co-interactors, SOX9, and NR2F1 are involved in the acquisition of gliogenic competence of neural stem/progenitor cells during CNS development, and E2F1 has been demonstrated to be involved in newborn neurons production in adult CNS (Cooper-Kuhn, Vroemen et al. 2002; Stolt, Lommes et al. 2003; Naka, Nakamura et al. 2008). Hence, our results highlight for the first time a possible pathway by which *Nr2e1* regulates neurogenesis of which *Lhx2* is

one of the possible target genes, and this regulation may also include other co-interactors such as SOX9, NR2F1, and E2F1.

Chapter 5: General Discussion

High-throughput is the linking factor of the three projects presented in this thesis. With that in mind, I focus my discussion on the utility of the HuGX strategy for the generation of site-specific humanized mouse models (HuMM) in a high-throughput manner. Firstly, I compare the HuGX strategy to the efforts of other large-scale groups (Pleiades Promoter Project, GENSAT) and discuss how we can refine our knowledge of regulatory elements for brain-specific gene expression. Secondly, with *NR2E1* as an example, I discuss the use of the HuGX strategy in an effort to understand the role of candidate human regulatory mutations. With the increasing reports of novel candidate mutations in patient populations, and the efforts from the international knock-out mouse consortium (IKMC) which is developing null-alleles for every mouse protein-coding genes, I discuss the applicability of the HuGX strategy for the study of human genes. Finally, I comment on the applications of large-scale mRNA profiling experiments, used to unravel networks of genes affecting biological pathways.

5.1 Challenges for the Future

5.1.1 Deciphering the Regulatory Sequence Code; a Gene Expression Approach

The milestone accomplished by the Human Genome Project more than a decade ago provided the blueprints of the human genome which allowed a better understanding of the underlying roles of DNA mutations, and their implication in disease development (Lander, Linton et al. 2001; Venter, Adams et al. 2001; Collins, Green et al. 2003). Protein-coding

mutations have been the scope of most investigations but a greater challenge awaits the field of research, understanding the transcriptional network.

The mammalian central nervous system (CNS) comprises numerous cell types with diverse spatial organization. Large scale projects such as the Allen Mouse Brain Atlas (ABA), and the Brain Gene Expression Map (BGEM) have aimed at increasing our knowledge of genes with brain specific expression using *in situ* hybridization (Magdaleno, Jensen et al. 2006; Lein, Hawrylycz et al. 2007). Other groups such as the Pleiades Promoter Project, and the Gene Expression Nervous System Atlas (GENSAT) used reporter mouse model generation to refine our knowledge of regulatory regions driving brain-specific expression (Gong, Zheng et al. 2003; Portales-Casamar, Swanson et al. 2010). The Pleiades Promoter Project used phylogenetic footprinting as a predictive approach to generate HuMM carrying human promoters (≤ 4 kb), driving gene expression in specific brain regions. This project avoided complex gene structures containing multiple transcription start sites, as well as genes with poor resolution of candidate regulatory regions. The GENSAT project, working with large mouse bacterial artificial chromosomes (BAC) could use more complex gene structures. This project was successful in identifying important potential regulatory regions by comparing the expression pattern resulting from two different BAC-harboring strains for the gene encoding choline acetyltransferase (ChAT) (Gong, Zheng et al. 2003). The GENSAT project, was based on random insertion of BACs in the mouse genome, requiring the characterization of multiple mouse strains, and reducing the throughput efficiency of the method. An approach using site-specific insertion in the mouse genome would reduce the number of strains to characterize, as well as reduce the variability in terms

of expression due to the nature of the genomic environment surrounding the different sites of insertion (position effect).

In the second chapter of my thesis, using the HuGX strategy, we tested a novel approach to generate site-specific BAC harbouring HuMM in a high-throughput manner. The genes chosen were deemed too complex for MiniPromoter design and instead the expression pattern resulting from the genes were evaluated using MaxiPromoter BAC constructs (> 100kb). This approach was used to test our capacity to identify BACs containing key regulatory sequences. Like the Pleiades Promoter Project, the HuGX strategy included the use of the *Hprt* locus as a docking site, allowing control for both the site of insertion and copy numbers of the BAC constructs inserted in the genome (Heaney, Rettew et al. 2004; Portales-Casamar, Swanson et al. 2010). Our results demonstrated that of the ten genes selected, eight were successfully retrofitted to comprise both the *HPRT* complementation cassette and a reporter gene (*lacZ* or *EGFP*). Of these eight, five were positive for expression, and four of them were characterized in this chapter (*AMOTL1*, *MAOA*, *NOV*, *NR2F2*). The results obtained highlighted that each of these constructs were expressed in the predicted adult brain regions, suggesting that the proper regulatory elements were included in the selected BAC constructs, and that these elements were functionally conserved from human to mouse.

Inspired by pioneer works from GENSAT, and the Pleiades Promoter Project, one can envision two possible approaches to decipher the regulatory sequence code from BAC-based analyses. One possibility is to use the knowledge acquired from various MaxiPromoter-harboring mouse strains and compare it to the expression pattern resulting from the orthologous mouse sequence in the GENSAT project. This would be done to better predict

and isolate specific regulatory regions through the subsequent generation of BAC deletion mouse strains. BAC deletion approaches have proven successful in identifying cis-regulatory elements for various genes such as *Leptin*, *Bmp4*, and *Musashi1* (Chandler, Chandler et al. 2009; Kawase, Imai et al. 2011; Wrann, Eguchi et al. 2012). The knowledge obtained from this approach would be applicable in designing MiniPromoters to drive expression in brain specific regions. This approach, although instructive, has the disadvantage of slow speed due to the laborious process of generating BAC harbouring mouse strains. A second possibility would be to use established MiniPromoter strains to refine our knowledge of the regulatory network controlling gene expression. For example, recently 22 transgenic mouse strains generated through the GENSAT project were used to isolate subgroups of transcription factors expressed in retinal cells by mean of specificity of the target reporter genes (Siegert, Cabuy et al. 2012). To date 17 MiniPromoter strains generated through the Pleiades Promoter Project have been demonstrated to direct retinal specific expression (de Leeuw et al. In preparation). Using the cell type specific transcription factor knowledge acquired from the GENSAT mice in combination with the relative simplicity of the human MiniPromoter constructs (≤ 4 kb), together with the power of transcription factor binding sites predictions; one could envision an efficient approach to understand the interaction between cell type specific transcription factors in the retina and the binding to their underlying DNA elements. In time, the knowledge and technology developed using the retina as a model, which comprise six major cell types, could be applied to a larger scale in the brain.

5.1.2 Deciphering the Regulatory Sequence Code; a Functional Approach

Understanding regulatory sequences provides insight into the potential role of DNA mutations in gene regulation. This is especially important considering the growth in the literature of genome wide association studies reporting novel candidate regulatory variants (Bosma, Chowdhury et al. 1995; Nakamura, Kugiyama et al. 2002; Sugatani, Yamakawa et al. 2002; Ono, Ezura et al. 2003; Jinnai, Sakagami et al. 2004; Marzec, Christie et al. 2007; Anttila, Stefansson et al. 2010; Dubois, Trynka et al. 2010; Speliotes, Willer et al. 2010). Furthermore, with the decreasing cost of sequencing technologies, we envision the acceleration of this trend. The HuGX strategy was originally designed as a platform to test the relevance of candidate *NR2E1* regulatory variants found in patient populations suffering from various brain disorders (Kumar, Leach et al. 2007; Kumar, McGhee et al. 2008). Inspired by the efforts from large scale projects such as the IKMC, we recently highlighted the fact that the HuGX strategy could be used for every human gene for which a mouse ortholog of similar function exist, and for which a human BAC construct can be derived (Schmouth, Bonaguro et al. 2012). The third chapter of this thesis highlights an experimental approach, using the HuGX strategy to generate two different HuMM; one containing a functional BAC construct for *NR2E1*, and the other containing the same BAC construct fused to a *lacZ* reporter gene (Schmouth, Banks et al. 2012). The functional allele allowed testing of the complementation capacity of the *NR2E1* human gene as a single copy on the *Nr2e1*-null background (*Nr2e1^{frc/frc}*). The BAC reporter construct allowed for the identification of the cell types in which the *NR2E1* human gene was expressed, results of which parallel the phenotype observed in the mouse strain harbouring the functional BAC. Using this approach, we uncovered novel potential regulatory regions for *NR2E1*, directing brain specific

expression in both the developing and adult forebrain (Schmouh, Banks et al. 2012). Furthermore, we have been able to demonstrate that a single copy of human *NR2E1* was sufficient to rescue the *Nr2e1*-null retinal phenotype, suggesting that this gene is a candidate in eye disorders. The results obtained using these two mouse models, highlight the need for novel potential regulatory regions to be included in future studies.

The knowledge acquired with the HuGX strategy can be applied as a hypothesis-generating approach, guiding future patient population studies. Obviously, the veracity of the predicted regulatory regions found in this project will have to be tested. Novel HuMM for *NR2E1*, generated through the HuGX strategy with a BAC construct harbouring the predicted regulatory regions will be useful in complementation analyses to test the functionality of these regions *in vivo*. Additionally, the development of differentiation protocols using mouse ESC that recapitulate forebrain development are allowing researchers to get more insights on the mechanisms affecting gene expression regulation (Barberi, Klivenyi et al. 2003; Gaspard, Bouschet et al. 2008). With this aspect in mind, the ESC used in the process of generating future HuMM for *NR2E1* will serve as valuable reductionist model, applicable in demonstrating if the predicted transcription factors truly bind to the corresponding regulatory sequences. The knowledge acquired using both a functional approach, with the humanized mouse models of *NR2E1* and a reductionist approach, using the ESC differentiation assays, will give insight on the potential role of *NR2E1* in neurodevelopmental diseases.

The utility of the HuGX strategy as an approach to other genes has been discussed in the literature (Schmouh, Bonaguro et al. 2012). Considering size as the only limitative factor, and based on the efficiency numbers obtained from the GENSAT project; which reported that 85% of their 1,000 mouse BAC random-insertion transgenics harbouring

constructs ≤ 100 kb had endogenous like expression, and based on the fact that 86% of the known human genes are ≤ 100 kb (Ensemble assembly, February 2009, GRCh37/hg19), we estimate that $\sim 75\%$ of the genes of the human genome could be amenable to analysis with the HuGX strategy. Furthermore, advancement in recombineering technologies, allowing the fusion of two BAC constructs, or the deletion of selected sequences (e.g. neighbouring genes), have made it a powerful approach for manipulating large DNA fragments in a higher throughput manner (Sopher and La Spada 2006). The HuGX strategy, using BAC harbouring constructs targeted at the *Hprt* locus, is well suited for genes ≤ 100 kb in size. Hence, for human genes spanning more than one megabase (e.g. dystrophin) the HuGX strategy remains unsuitable (den Dunnen, Bakker et al. 1989; Shizuya, Birren et al. 1992; Ioannou, Amemiya et al. 1994). Yeast artificial chromosomes (YAC) can accommodate such large genomic DNA, and site-specific mutagenesis can be readily applied using the homologous recombination system of yeast (Duff and Huxley 1996). However, at the moment, site-specific docking of YAC constructs is not possible making YAC unsuitable for high-throughput single-copy approaches. The *Hprt* locus has been used to dock the largest DNA fragment in the genome (> 200 kb), but its presence on the X chromosome is not ideal as neither in males or females can you obtain two functional copies of the construct inserted. The generation of an alternative autosomal docking site that allows insertion of large-DNA fragments up to the size of YAC would be ideal.

5.1.3 The Future of Large mRNA Profiling Experiments

Nr2e1 encodes an orphan nuclear receptor involved in maintaining the neural stem cell population during forebrain development (Roy, Kuznicki et al. 2004; Li, Sun et al. 2008). In chapter 4 of this thesis, we sought to identify novel biological pathways regulated

by *Nr2e1* using a large scale mRNA expression approach. One of the biggest challenges coming from this type of approach is to identify target genes relevant to potential biological pathways of interest. We took advantage of the intrinsic nature of *Nr2e1*, known to act as a transcription factor, and the advancement in prediction algorithms to unravel potential novel direct target genes of *Nr2e1* (Yu, McKeown et al. 1994). The analysis highlighted 63 potential target genes involved in nervous system development. We also performed an analysis to identify potential co-interactors of *Nr2e1* using the list of 63 genes. Our results suggest that *Nr2e1* regulation of genes implicated in nervous system development involves interactions with other transcription factors, such as SOX9, E2F1, and NR2F1, with roles in neural stem cell processes (Tsai, Hu et al. 1998; Naka, Nakamura et al. 2008; Scott, Wynn et al. 2010). An interesting target gene predicted from this analysis was *Lhx2*, a transcription factor encoding gene that suppresses astrogliogenesis and promotes neurogenesis during hippocampal development (Subramanian, Sarkar et al. 2011). These results correlate with the fact that *Nr2e1*-null embryos demonstrate premature neurogenesis during development (Roy, Kuznicki et al. 2004). Furthermore, studies involving mutant mouse models of *Sox9*, and *Nr2f1* have demonstrated an involvement of these genes in cell fate decisions during central nervous system development (Stolt, Lommes et al. 2003; Naka, Nakamura et al. 2008). The analysis highlights, for the first time, a potential interaction network involving *Nr2e1*, *Sox9*, *E2f1*, and *Nr2f1* in regulating the expression of *Lhx2*. Based on these findings, *Nr2e1* may regulate cell fate decision during forebrain development.

Although promising, the results obtained in chapter 4 are based on predictions that will require extensive validation. To uncover the mechanism of gene regulation by *Nr2e1* will require minimalist approaches such as neural stem cells cultures isolated from

embryonic forebrains. Extensive protocols for the culturing of these cells are readily available and several successful examples in identifying novel direct target genes of *Nr2e1* using these approaches exist (Sun, Yu et al. 2007; Iwahara, Hisahara et al. 2009; Zhao, Sun et al. 2009; Zhao, Sun et al. 2010). In this case, the relative ease of transfection of neural stem cells in culture will be ideal to study the relevance of predicted binding sites, using a chromatin immunoprecipitation (ChIP) PCR approach, and the predicted co-interactors of NR2E1, using “pull-down” experiments. Alternatively, electroporation of siRNA in the telencephalon of developing embryos for genes predicted to be upregulated in our SAGE analysis would allow temporal evaluation of the biological implication of this deregulation. As an example, our SAGE results demonstrated that *Lhx2* appears to be upregulated in the dorsal-lateral telencephalon of *Nr2e1^{frc/frc}* embryos at E13.5, and E15.5. Hence it would be interesting to assess whether precocious neuron formation occurs when electroporating a siRNA recognizing *Lhx2* specifically in the lateral telencephalon of *Nr2e1^{frc/frc}* embryos. Presumably, if *Lhx2* truly is involved in this phenotype, the number of neurons formed close to the site of electroporation should be reduced in comparison to a control siRNA.

To date, *Nr2e1* transcriptome studies have used protocols involving isolation and purification of cell populations from either the developing retina or the adult forebrain (Zhang, Zou et al. 2006; Zhang, Zou et al. 2008). These approaches have allowed access to a relatively large amount of material as well as allowed the amplification, and freezing of used cells for further validation experiments. However, the isolation of these cells requires culturing, the conditions of which can alter expression profiles. Our approach involved direct RNA purification coming from tissues obtained by laser capture microdissection of the lateral telencephalon, minimizing the potential impact of culture conditions on expression

profiles. However, this laborious approach produced insufficient tissue for further validation experiments. This reality demonstrates the compromise that sometimes has to be made between the abundance of the starting material, and the possibilities of including an artifact when using culture conditions when a large scale mRNA expression profile experiment is designed.

The method used for RNA isolation is an important consideration. In our case, the RNeasy Micro Kit (Qiagen) was used which allows efficient purification of RNA molecules ≥ 200 bps in size but does not allow purification of micro-RNA (miRNA) molecules. A recent publication has demonstrated that a miRNA gene, *miR-9* regulates *Nr2e1* gene expression. This regulation involved an inhibitory feedback loop in which *Nr2e1* represses the expression of the precursor of *miR-9* (Zhao, Sun et al. 2009). The growing body of research in the literature, demonstrating the roles of miRNA genes in biological processes, is another important aspect that needs to be taken into account when designing a large scale mRNA profiling approach (Ebert and Sharp 2012).

To date, two major technologies are being used in the literature for large scale mRNA profiling experiments; microarrays which rely on the use of hybridization to a chip, and serial analysis of gene expression (SAGE) which rely on sequencing. Microarrays have been the technique of choice, but the decreasing cost of sequencing is changing the field. In RNA-seq, direct sequencing of the RNA content of a cell or tissue is performed. Novel protocols, which allow the generation of RNA-seq libraries with submicrogram amounts of RNA material (~10 pg), and a reduced number of purification steps are supplanting both SAGE and microarray technologies in the field of transcriptome analyses (Gertz, Varley et al. 2012). Advancements in large scale ChIP technologies, which study the binding of proteins to

genomic DNA, are useful for studying transcription factors. Therefore, a combination of both large scale profiling mRNA analysis and ChIP (ChIP-seq or ChIP-ChIP) would reliably uncover the interaction of a transcription factor within the regulatory regions of a target gene. Such use of complementary technologies has been applied for deciphering the role of the transcription factor gene *Pax6*, demonstrating during brain development that a dynamic balance for this gene is required in order to maintain proper neural stem cell self-renewal, and neurogenesis (Sansom, Griffiths et al. 2009). Combining the results obtained from chapter 4 with a large scale ChIP experiment using different models of *Nr2e1* would increase our knowledge of potential biological pathways affected by this gene, as well as allow a large scale evaluation of our prediction results. This in turn would help us to develop better transcription factor binding site prediction analyses for *Nr2e1*.

5.2 Conclusions

The work presented in this thesis has applied the HuGX strategy for humanized mouse model generation. The results in chapter 2 demonstrated that specific brain regional expression of human genes matched the predictions made based on orthologous mouse genes. The findings suggest a functional conservation of gene regulatory regions allowing a delineation of the expression of the human genes in the predicted brain regions. Future studies should focus on designing and testing MiniPromoter constructs based on the minimal genomic DNA boundaries highlighted by these humanized mouse models. Further testing of the successful constructs in higher mammals such as primates should be considered, with the hope that, one day, they could be applied in human gene therapy. Chapter 3 highlights how the HuGX strategy can help in understanding the biological role of the human *NR2E1* gene in both brain, and retina development. The results obtained from this study highlight the

potential role of *NR2E1* in eye disorders, and will influence the design of future sequencing studies, looking at unraveling the role of this gene in potential brain disorders. We can also add that the novel *NR2E1* forebrain specific candidate regulatory regions identified in this chapter were excluded in previous human population analyses. Hence performing a re-sequencing experiment, with emphasis on the novel candidate regulatory regions highlighted in chapter 3 should be considered, considering the potential for the discovery of novel candidate regulatory mutations. Finally, chapter 4 highlights the use of a large scale mRNA profiling approach looking at identifying novel pathways in which *Nr2e1* acts during forebrain development. This chapter also highlights the use of large scale transcription factor prediction approaches to unravel novel interacting partners of *Nr2e1* possibly involved in gene expression regulation. Future studies related to this project will require further validation of the candidate target genes identified using the large scale mRNA profiling experiment as well as the transcription factor prediction approach to clearly reveal novel biological pathway(s) affected by *Nr2e1* during forebrain development.

References

- Abe, T., H. Kiyonari, et al. (2011). "Establishment of conditional reporter mouse lines at ROSA26 locus for live cell imaging." Genesis **49**(7): 579-590.
- Abrahams, B. S., M. C. Kwok, et al. (2005). "Pathological aggression in "fierce" mice corrected by human nuclear receptor 2E1." J Neurosci **25**(27): 6263-6270.
- Abrahams, B. S., G. M. Mak, et al. (2002). "Novel vertebrate genes and putative regulatory elements identified at kidney disease and NR2E1/fierce loci." Genomics **80**(1): 45-53.
- Anderson, S. A., D. D. Eisenstat, et al. (1997). "Interneuron migration from basal forebrain to neocortex: dependence on Dlx genes." Science **278**(5337): 474-476.
- Angevine, J. B., Jr. and R. L. Sidman (1961). "Autoradiographic study of cell migration during histogenesis of cerebral cortex in the mouse." Nature **192**: 766-768.
- Anttila, V., H. Stefansson, et al. (2010). "Genome-wide association study of migraine implicates a common susceptibility variant on 8q22.1." Nature genetics **42**(10): 869-873.
- Ashburner, M., C. A. Ball, et al. (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." Nature genetics **25**(1): 25-29.
- Audic, S. and J. M. Claverie (1997). "The significance of digital gene expression profiles." Genome Res **7**(10): 986-995.
- Bailey, T. L., M. Boden, et al. (2009). "MEME SUITE: tools for motif discovery and searching." Nucleic Acids Res **37**(Web Server issue): W202-208.

- Barberi, T., P. Klivenyi, et al. (2003). "Neural subtype specification of fertilization and nuclear transfer embryonic stem cells and application in parkinsonian mice." Nat Biotechnol **21**(10): 1200-1207.
- Bayer, S. A. and J. Altman (1991). Neocortical development, New York: Raven Press.
- Bosma, P. J., J. R. Chowdhury, et al. (1995). "The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert's syndrome." The New England journal of medicine **333**(18): 1171-1175.
- Bratt, A., W. J. Wilson, et al. (2002). "Angiomotin belongs to a novel protein family with conserved coiled-coil and PDZ binding domains." Gene **298**(1): 69-77.
- Brinster, R. L., H. Y. Chen, et al. (1981). "Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs." Cell **27**(1 Pt 2): 223-231.
- Bronson, S. K., E. G. Plaehn, et al. (1996). "Single-copy transgenic mice with chosen-site integration." Proc Natl Acad Sci U S A **93**(17): 9067-9072.
- Buchholz, F., P. O. Angrand, et al. (1998). "Improved properties of FLP recombinase evolved by cycling mutagenesis." Nature biotechnology **16**(7): 657-662.
- Bulyk, M. L., E. Gentalen, et al. (1999). "Quantifying DNA-protein interactions by double-stranded DNA arrays." Nature biotechnology **17**(6): 573-577.
- Butler, A. B. (1994). "The evolution of the dorsal pallium in the telencephalon of amniotes: cladistic analysis and a new hypothesis." Brain research. Brain research reviews **19**(1): 66-101.
- Butler, A. B. (1994). "The evolution of the dorsal thalamus of jawed vertebrates, including mammals: cladistic analysis and a new hypothesis." Brain research. Brain research reviews **19**(1): 29-65.

- Cases, O., I. Seif, et al. (1995). "Aggressive behavior and altered amounts of brain serotonin and norepinephrine in mice lacking MAOA." Science **268**(5218): 1763-1766.
- Chandler, J., P. Hohenstein, et al. (2001). "Human BRCA1 gene rescues the embryonic lethality of Brca1 mutant mice." Genesis **29**(2): 72-77.
- Chandler, K. J., R. L. Chandler, et al. (2009). "Identification of an ancient Bmp4 mesoderm enhancer located 46 kb from the promoter." Developmental biology **327**(2): 590-602.
- Chen, H., M. Centola, et al. (1998). "Characterization of gene expression in resting and activated mast cells." The Journal of experimental medicine **188**(9): 1657-1668.
- Chen, J., A. Rattner, et al. (2005). "The rod photoreceptor-specific nuclear receptor Nr2e3 represses transcription of multiple cone-specific genes." J Neurosci **25**(1): 118-129.
- Chen, K., O. Cases, et al. (2007). "Forebrain-specific expression of monoamine oxidase A reduces neurotransmitter levels, restores the brain structure, and rescues aggressive behavior in monoamine oxidase A-deficient mice." The Journal of biological chemistry **282**(1): 115-123.
- Cheng, A., A. L. Scott, et al. (2010). "Monoamine oxidases regulate telencephalic neural progenitors in late embryonic and early postnatal development." The Journal of neuroscience : the official journal of the Society for Neuroscience **30**(32): 10752-10762.
- Chou, S. J., C. G. Perez-Garcia, et al. (2009). "Lhx2 specifies regional fate in Emx1 lineage of telencephalic progenitors generating cerebral cortex." Nat Neurosci **12**(11): 1381-1389.
- Collins, F. S., E. D. Green, et al. (2003). "A vision for the future of genomics research." Nature **422**(6934): 835-847.

- Consortium, T. G. O. (2001). "Creating the gene ontology resource: design and implementation." Genome research **11**(8): 1425-1433.
- Cooper-Kuhn, C. M., M. Vroemen, et al. (2002). "Impaired adult neurogenesis in mice lacking the transcription factor E2F1." Molecular and cellular neurosciences **21**(2): 312-323.
- Copeland, N. G., N. A. Jenkins, et al. (2001). "Recombineering: a powerful new tool for mouse functional genomics." Nat Rev Genet **2**(10): 769-779.
- Cvetkovic, B., B. Yang, et al. (2000). "Appropriate tissue- and cell-specific expression of a single copy human angiotensinogen transgene specifically targeted upstream of the HPRT locus by homologous recombination." The Journal of biological chemistry **275**(2): 1073-1078.
- D'Souza, C. A., V. Chopra, et al. (2008). "Identification of a set of genes showing regionally enriched expression in the mouse brain." BMC Neurosci **9**: 66.
- de Carlos, J. A., L. Lopez-Mascaraque, et al. (1996). "Dynamics of cell migration from the lateral ganglionic eminence in the rat." J Neurosci **16**(19): 6146-6156.
- de Hoon, M. J., S. Imoto, et al. (2004). "Open source clustering software." Bioinformatics **20**(9): 1453-1454.
- den Dunnen, J. T., E. Bakker, et al. (1989). "The DMD gene analysed by field inversion gel electrophoresis." British medical bulletin **45**(3): 644-658.
- Dennis, G., Jr., B. T. Sherman, et al. (2003). "DAVID: Database for Annotation, Visualization, and Integrated Discovery." Genome biology **4**(5): P3.

- Dermitzakis, E. T. and A. G. Clark (2002). "Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover." Molecular biology and evolution **19**(7): 1114-1121.
- Ding, S., X. Wu, et al. (2005). "Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice." Cell **122**(3): 473-483.
- Doetschman, T., R. G. Gregg, et al. (1987). "Targetted correction of a mutant HPRT gene in mouse embryonic stem cells." Nature **330**(6148): 576-578.
- Dubois, P. C., G. Trynka, et al. (2010). "Multiple common variants for celiac disease influencing immune gene expression." Nature genetics **42**(4): 295-302.
- Duff, K. and C. Huxley (1996). "Targeting mutations to YACs by homologous recombination." Methods in molecular biology **54**: 187-198.
- Dunnett, S. B., D. J. Sirinathsinghji, et al. (1989). "Monoamine deficiency in a transgenic (Hprt-) mouse model of Lesch-Nyhan syndrome." Brain research **501**(2): 401-406.
- Ebert, M. S. and P. A. Sharp (2012). "Roles for MicroRNAs in Conferring Robustness to Biological Processes." Cell **149**(3): 515-524.
- Evans, V., A. Hatzopoulos, et al. (2000). "Targeting the Hprt locus in mice reveals differential regulation of Tie2 gene expression in the endothelium." Physiological genomics **2**(2): 67-75.
- Falcon, S. and R. Gentleman (2007). "Using GOSTats to test gene lists for GO term association." Bioinformatics **23**(2): 257-258.
- Finger, S., R. P. Heavens, et al. (1988). "Behavioral and neurochemical evaluation of a transgenic mouse model of Lesch-Nyhan syndrome." Journal of the neurological sciences **86**(2-3): 203-213.

- Frazer, K. A., D. G. Ballinger, et al. (2007). "A second generation human haplotype map of over 3.1 million SNPs." Nature **449**(7164): 851-861.
- Friedrich, G. and P. Soriano (1991). "Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice." Genes & development **5**(9): 1513-1523.
- Gao, Q., G. E. Reynolds, et al. (2007). "Telomeric transgenes are silenced in adult mouse tissues and embryo fibroblasts but are expressed in embryonic stem cells." Stem Cells **25**(12): 3085-3092.
- Garrick, D., S. Fiering, et al. (1998). "Repeat-induced gene silencing in mammals." Nature genetics **18**(1): 56-59.
- Gaspard, N., T. Bouschet, et al. (2009). "Generation of cortical neurons from mouse embryonic stem cells." Nat Protoc **4**(10): 1454-1463.
- Gaspard, N., T. Bouschet, et al. (2008). "An intrinsic mechanism of corticogenesis from embryonic stem cells." Nature **455**(7211): 351-357.
- Gertz, J., K. E. Varley, et al. (2012). "Transposase mediated construction of RNA-seq libraries." Genome research **22**(1): 134-141.
- Gong, S., C. Zheng, et al. (2003). "A gene expression atlas of the central nervous system based on bacterial artificial chromosomes." Nature **425**(6961): 917-925.
- Gordon, J. W. and F. H. Ruddle (1981). "Integration and stable germ line transmission of genes injected into mouse pronuclei." Science **214**(4526): 1244-1246.
- Gordon, J. W., G. A. Scangos, et al. (1980). "Genetic transformation of mouse embryos by microinjection of purified DNA." Proceedings of the National Academy of Sciences of the United States of America **77**(12): 7380-7384.

- Graham, V., J. Khudyakov, et al. (2003). "SOX2 functions to maintain neural progenitor identity." Neuron **39**(5): 749-765.
- Guerin, K., C. Y. Gregory-Evans, et al. (2008). "Systemic aminoglycoside treatment in rodent models of retinitis pigmentosa." Exp Eye Res **87**(3): 197-207.
- Guillot, P. V., L. Liu, et al. (2000). "Targeting of human eNOS promoter to the Hprt locus of mice leads to tissue-restricted transgene expression." Physiological genomics **2**(2): 77-83.
- Harris, M. A., J. Clark, et al. (2004). "The Gene Ontology (GO) database and informatics resource." Nucleic acids research **32**(Database issue): D258-261.
- Haverkamp, S. and H. Wassle (2000). "Immunocytochemical analysis of the mouse retina." The Journal of comparative neurology **424**(1): 1-23.
- Hawes, N. L., R. S. Smith, et al. (1999). "Mouse fundus photography and angiography: a catalogue of normal and mutant phenotypes." Mol Vis **5**: 22.
- Heaney, J. D., A. N. Rettew, et al. (2004). "Tissue-specific expression of a BAC transgene targeted to the Hprt locus in mouse embryonic stem cells." Genomics **83**(6): 1072-1082.
- Hisahara, S., S. Chiba, et al. (2008). "Histone deacetylase SIRT1 modulates neuronal differentiation by its nuclear translocation." Proceedings of the National Academy of Sciences of the United States of America **105**(40): 15599-15604.
- Ho Sui, S. J., D. L. Fulton, et al. (2007). "oPOSSUM: integrated tools for analysis of regulatory motif over-representation." Nucleic Acids Res **35**(Web Server issue): W245-252.

- Ho Sui, S. J., J. R. Mortimer, et al. (2005). "oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes." Nucleic Acids Res **33**(10): 3154-3164.
- Hodgson, J. G., N. Agopyan, et al. (1999). "A YAC mouse model for Huntington's disease with full-length mutant huntingtin, cytoplasmic toxicity, and selective striatal neurodegeneration." Neuron **23**(1): 181-192.
- Hodgson, J. G., D. J. Smith, et al. (1996). "Human huntingtin derived from YAC transgenes compensates for loss of murine huntingtin by rescue of the embryonic lethal phenotype." Human molecular genetics **5**(12): 1875-1885.
- Hooper, M., K. Hardy, et al. (1987). "HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells." Nature **326**(6110): 292-295.
- Horan, K., C. Jang, et al. (2008). "Annotating genes of known and unknown function by large-scale coexpression analysis." Plant Physiol **147**(1): 41-57.
- Hosack, D. A., G. Dennis, Jr., et al. (2003). "Identifying biological themes within lists of genes with EASE." Genome Biol **4**(10): R70.
- Huang da, W., B. T. Sherman, et al. (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nature protocols **4**(1): 44-57.
- Huang da, W., B. T. Sherman, et al. (2007). "DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists." Nucleic Acids Res **35**(Web Server issue): W169-175.
- Inoue, M., A. Iida, et al. (2010). "COUP-TFI and -TFII nuclear receptors are expressed in amacrine cells and play roles in regulating the differentiation of retinal progenitor cells." Experimental eye research **90**(1): 49-56.

- Ioannou, P. A., C. T. Amemiya, et al. (1994). "A new bacteriophage P1-derived vector for the propagation of large human DNA fragments." Nature genetics **6**(1): 84-89.
- Iwahara, N., S. Hisahara, et al. (2009). "Transcriptional activation of NAD⁺-dependent protein deacetylase SIRT1 by nuclear receptor TLX." Biochemical and biophysical research communications **386**(4): 671-675.
- Jackson, A., P. Panayiotidis, et al. (1998). "The human homologue of the *Drosophila* tailless gene (TLX): characterization and mapping to a region of common deletion in human lymphoid leukemia on chromosome 6q21." Genomics **50**(1): 34-43.
- Jamsai, D., F. Zaibak, et al. (2005). "A humanized mouse model for a common beta0-thalassemia mutation." Genomics **85**(4): 453-461.
- Jinnah, H. A., F. H. Gage, et al. (1991). "Amphetamine-induced behavioral phenotype in a hypoxanthine-guanine phosphoribosyltransferase-deficient mouse model of Lesch-Nyhan syndrome." Behavioral neuroscience **105**(6): 1004-1012.
- Jinnai, N., T. Sakagami, et al. (2004). "Polymorphisms in the prostaglandin E2 receptor subtype 2 gene confer susceptibility to aspirin-intolerant asthma: a candidate gene approach." Human molecular genetics **13**(24): 3203-3217.
- Job, C. and S. S. Tan (2003). "Constructing the mammalian neocortex: the role of intrinsic factors." Developmental biology **257**(2): 221-232.
- Kamachi, Y., M. Uchikawa, et al. (2001). "Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development." Genes Dev **15**(10): 1272-1286.
- Kanatani, S., M. Yozu, et al. (2008). "COUP-TFII is preferentially expressed in the caudal ganglionic eminence and is involved in the caudal migratory stream." The Journal of

- neuroscience : the official journal of the Society for Neuroscience **28**(50): 13582-13591.
- Kawase, S., T. Imai, et al. (2011). "Identification of a novel intronic enhancer responsible for the transcriptional regulation of musashi1 in neural stem/progenitor cells." Molecular brain **4**: 14.
- Khattra, J., A. D. Delaney, et al. (2007). "Large-scale production of SAGE libraries from microdissected tissues, flow-sorted cells, and cell lines." Genome Res **17**(1): 108-116.
- Kim, J. J., J. C. Shih, et al. (1997). "Selective enhancement of emotional, but not motor, learning in monoamine oxidase A-deficient mice." Proceedings of the National Academy of Sciences of the United States of America **94**(11): 5929-5933.
- King, M. C. and A. C. Wilson (1975). "Evolution at two levels in humans and chimpanzees." Science **188**(4184): 107-116.
- Kitambi, S. S. and G. Hauptmann (2007). "The zebrafish orphan nuclear receptor genes nr2e1 and nr2e3 are expressed in developing eye and forebrain." Gene Expr Patterns **7**(4): 521-528.
- Kohlhepp, R. L., L. F. Hegge, et al. (2001). "The ROSA26 LacZ-neo(R) insertion confers resistance to mammary tumors in Apc(Min/+) mice." Mammalian genome : official journal of the International Mammalian Genome Society **12**(8): 606-611.
- Kuhn, A., D. R. Goldstein, et al. (2007). "Mutant huntingtin's effects on striatal gene expression in mice recapitulate changes observed in human Huntington's disease brain and do not differ with mutant huntingtin length or wild-type huntingtin dosage." Human molecular genetics **16**(15): 1845-1861.

- Kumar, R. A., D. B. Everman, et al. (2007). "Absence of mutations in NR2E1 and SNX3 in five patients with MMEP (microcephaly, microphthalmia, ectrodactyly, and prognathism) and related phenotypes." BMC Med Genet **8**(1): 48.
- Kumar, R. A., S. Leach, et al. (2007). "Mutation and evolutionary analyses identify NR2E1-candidate-regulatory mutations in humans with severe cortical malformations." Genes Brain Behav **6**(6): 503-516.
- Kumar, R. A., K. A. McGhee, et al. (2008). "Initial association of NR2E1 with bipolar disorder and identification of candidate mutations in bipolar disorder, schizophrenia, and aggression through resequencing." Am J Med Genet B Neuropsychiatr Genet **147B**(6): 880-889.
- Land, P. W. and A. P. Monaghan (2003). "Expression of the transcription factor, tailless, is required for formation of superficial cortical layers." Cereb Cortex **13**(9): 921-931.
- Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.
- Lane, T. F., C. Lin, et al. (2000). "Gene replacement with the human BRCA1 locus: tissue specific expression and rescue of embryonic lethality in mice." Oncogene **19**(36): 4085-4090.
- Lash, A. E., C. M. Tolstoshev, et al. (2000). "SAGEmap: a public gene expression resource." Genome research **10**(7): 1051-1060.
- Lee, C. T., L. Li, et al. (2004). "The nuclear orphan receptor COUP-TFII is required for limb and skeletal muscle development." Mol Cell Biol **24**(24): 10835-10843.
- Lein, E. S., M. J. Hawrylycz, et al. (2007). "Genome-wide atlas of gene expression in the adult mouse brain." Nature **445**(7124): 168-176.

- Lendahl, U., L. B. Zimmerman, et al. (1990). "CNS stem cells express a new class of intermediate filament protein." Cell **60**(4): 585-595.
- Lenhard, B., A. Sandelin, et al. (2003). "Identification of conserved regulatory elements by comparative genome analysis." Journal of biology **2**(2): 13.
- Li, W., G. Sun, et al. (2008). "Nuclear receptor TLX regulates cell cycle progression in neural stem cells of the developing brain." Mol Endocrinol **22**(1): 56-64.
- Liskay, R. M. and R. J. Evans (1980). "Inactive X chromosome DNA does not function in DNA-mediated cell transformation for the hypoxanthine phosphoribosyltransferase gene." Proc Natl Acad Sci U S A **77**(8): 4895-4898.
- Liu, C., X. J. Liu, et al. (1999). "Nephroblastoma overexpressed gene (NOV) codes for a growth factor that induces protein tyrosine phosphorylation." Gene **238**(2): 471-478.
- Liu, H. K., T. Belz, et al. (2008). "The nuclear receptor tailless is required for neurogenesis in the adult subventricular zone." Genes Dev **22**(18): 2473-2478.
- Liu, H. K., Y. Wang, et al. (2010). "The nuclear receptor tailless induces long-term neural stem cell expansion and brain tumor initiation." Genes Dev **24**(7): 683-695.
- Lois, C., E. J. Hong, et al. (2002). "Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors." Science **295**(5556): 868-872.
- Luque, J. M., S. W. Kwan, et al. (1995). "Cellular expression of mRNAs encoding monoamine oxidases A and B in the rat central nervous system." The Journal of comparative neurology **363**(4): 665-680.
- Lyznik, L. A., J. C. Mitchell, et al. (1993). "Activity of yeast FLP recombinase in maize and rice protoplasts." Nucleic acids research **21**(4): 969-975.

- Madisen, L., T. A. Zwingman, et al. (2010). "A robust and high-throughput Cre reporting and characterization system for the whole mouse brain." Nature neuroscience **13**(1): 133-140.
- Magdaleno, S., P. Jensen, et al. (2006). "BGEM: an in situ hybridization database of gene expression in the embryonic and adult mouse nervous system." PLoS biology **4**(4): e86.
- Mao, X., Y. Fujiwara, et al. (1999). "Improved reporter strain for monitoring Cre recombinase-mediated DNA excisions in mice." Proceedings of the National Academy of Sciences of the United States of America **96**(9): 5037-5042.
- Marzec, J. M., J. D. Christie, et al. (2007). "Functional polymorphisms in the transcription factor NRF2 in humans increase the risk of acute lung injury." The FASEB journal : official publication of the Federation of American Societies for Experimental Biology **21**(9): 2237-2246.
- Mates, L., M. K. Chuah, et al. (2009). "Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates." Nature genetics **41**(6): 753-761.
- McKenna, W. L., J. Betancourt, et al. (2011). "Tbr1 and Fezf2 regulate alternate corticofugal neuronal identities during neocortical development." The Journal of neuroscience : the official journal of the Society for Neuroscience **31**(2): 549-564.
- Mendez-Gomez, H. R., E. Vergano-Vera, et al. (2011). "The T-box brain 1 (Tbr1) transcription factor inhibits astrocyte formation in the olfactory bulb and regulates neural stem cell fate." Molecular and cellular neurosciences **46**(1): 108-121.

- Miller, F. D. and A. S. Gauthier (2007). "Timing is everything: making neurons versus glia in the developing cortex." Neuron **54**(3): 357-369.
- Milot, E., J. Strouboulis, et al. (1996). "Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription." Cell **87**(1): 105-114.
- Minami, T., D. J. Donovan, et al. (2002). "Differential regulation of the von Willebrand factor and Flt-1 promoters in the endothelium of hypoxanthine phosphoribosyltransferase-targeted mice." Blood **100**(12): 4019-4025.
- Miyawaki, T., A. Uemura, et al. (2004). "Tlx, an orphan nuclear receptor, regulates cell numbers and astrocyte development in the developing retina." J Neurosci **24**(37): 8124-8134.
- Mizutani, R., K. Nakamura, et al. (2011). "Developmental expression of sorting nexin 3 in the mouse central nervous system." Gene expression patterns : GEP **11**(1-2): 33-40.
- Monaghan, A. P., D. Bock, et al. (1997). "Defective limbic system in mice lacking the tailless gene." Nature **390**(6659): 515-517.
- Monaghan, A. P., E. Grau, et al. (1995). "The mouse homolog of the orphan nuclear receptor tailless is expressed in the developing forebrain." Development **121**(3): 839-853.
- Monuki, E. S., F. D. Porter, et al. (2001). "Patterning of the dorsal telencephalon and cerebral cortex by a roof plate-Lhx2 pathway." Neuron **32**(4): 591-604.
- Myers, R. M., J. Stamatoyannopoulos, et al. (2011). "A user's guide to the encyclopedia of DNA elements (ENCODE)." PLoS biology **9**(4): e1001046.
- Nadarajah, B., J. E. Brunstrom, et al. (2001). "Two modes of radial migration in early development of the cerebral cortex." Nat Neurosci **4**(2): 143-150.

- Naka, H., S. Nakamura, et al. (2008). "Requirement for COUP-TFI and II in the temporal specification of neural stem cells in CNS development." Nat Neurosci **11**(9): 1014-1023.
- Nakamura, S., K. Kugiyama, et al. (2002). "Polymorphism in the 5'-flanking region of human glutamate-cysteine ligase modifier subunit gene is associated with myocardial infarction." Circulation **105**(25): 2968-2973.
- Narayanan, K., R. Williamson, et al. (1999). "Efficient and precise engineering of a 200 kb beta-globin human/bacterial artificial chromosome in E. coli DH10B using an inducible homologous recombination system." Gene therapy **6**(3): 442-447.
- Natarajan, D., E. Andermarcher, et al. (2000). "Mouse Nov gene is expressed in hypaxial musculature and cranial structures derived from neural crest cells and placodes." Developmental dynamics : an official publication of the American Association of Anatomists **219**(3): 417-425.
- Nery, S., G. Fishell, et al. (2002). "The caudal ganglionic eminence is a source of distinct cortical and subcortical cell populations." Nature neuroscience **5**(12): 1279-1287.
- Nishimura, M., M. Kakizaki, et al. (2002). "JEAP, a novel component of tight junctions in exocrine cells." The Journal of biological chemistry **277**(7): 5583-5587.
- O'Doherty, A., S. Ruf, et al. (2005). "An aneuploid mouse strain carrying human chromosome 21 with Down syndrome phenotypes." Science **309**(5743): 2033-2037.
- Ono, S., Y. Ezura, et al. (2003). "A promoter SNP (-1323T>C) in G-substrate gene (GSBS) correlates with hypercholesterolemia." Journal of human genetics **48**(9): 447-450.
- Paigen, K. (1995). "A miracle enough: the power of mice." Nat Med **1**(3): 215-220.

- Park, H. J., J. K. Kim, et al. (2010). "The neural stem cell fate determinant TLX promotes tumorigenesis and genesis of cells resembling glioma stem cells." Molecules and cells **30**(5): 403-408.
- Parsons, D. W., S. Jones, et al. (2008). "An integrated genomic analysis of human glioblastoma multiforme." Science **321**(5897): 1807-1812.
- Pedram, M., C. N. Sprung, et al. (2006). "Telomere position effect and silencing of transgenes near telomeres in the mouse." Mol Cell Biol **26**(5): 1865-1878.
- Pereira, F. A., Y. Qiu, et al. (1999). "The orphan nuclear receptor COUP-TFII is required for angiogenesis and heart development." Genes & development **13**(8): 1037-1049.
- Peters, D. G., A. B. Kassam, et al. (1999). "Comprehensive transcript analysis in small quantities of mRNA by SAGE-lite." Nucleic acids research **27**(24): e39.
- Phillips, H. S., S. Kharbanda, et al. (2006). "Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis." Cancer Cell **9**(3): 157-173.
- Pignoni, F., R. M. Baldarelli, et al. (1990). "The Drosophila gene tailless is expressed at the embryonic termini and is a member of the steroid receptor superfamily." Cell **62**(1): 151-163.
- Pollock, R. and R. Treisman (1990). "A sensitive method for the determination of protein-DNA binding specificities." Nucleic acids research **18**(21): 6197-6204.
- Popova, N. K., Y. A. Skrinskaya, et al. (2001). "Behavioral characteristics of mice with genetic knockout of monoamine oxidase type A." Neuroscience and behavioral physiology **31**(6): 597-602.

- Portales-Casamar, E., D. Arenillas, et al. (2009). "The PAZAR database of gene regulatory information coupled to the ORCA toolkit for the study of regulatory sequences." Nucleic Acids Res **37**(Database issue): D54-60.
- Portales-Casamar, E., S. Kirov, et al. (2007). "PAZAR: a framework for collection and dissemination of cis-regulatory sequence annotation." Genome biology **8**(10): R207.
- Portales-Casamar, E., D. J. Swanson, et al. (2010). "A regulatory toolbox of MiniPromoters to drive selective expression in the brain." Proceedings of the National Academy of Sciences of the United States of America **107**(38): 16589-16594.
- Qin, J., X. Chen, et al. (2010). "COUP-TFII regulates tumor growth and metastasis by modulating tumor angiogenesis." Proceedings of the National Academy of Sciences of the United States of America **107**(8): 3687-3692.
- Qin, J., X. Chen, et al. (2010). "Nuclear receptor COUP-TFII controls pancreatic islet tumor angiogenesis by regulating vascular endothelial growth factor/vascular endothelial growth factor receptor-2 signaling." Cancer research **70**(21): 8812-8821.
- Qiu, Y., A. J. Cooney, et al. (1994). "Spatiotemporal expression patterns of chicken ovalbumin upstream promoter-transcription factors in the developing mouse central nervous system: evidence for a role in segmental patterning of the diencephalon." Proceedings of the National Academy of Sciences of the United States of America **91**(10): 4451-4455.
- Reynolds, B. A., W. Tetzlaff, et al. (1992). "A multipotent EGF-responsive striatal embryonic progenitor cell produces neurons and astrocytes." The Journal of neuroscience : the official journal of the Society for Neuroscience **12**(11): 4565-4574.

- Reynolds, L. E., A. R. Watson, et al. (2010). "Tumour angiogenesis is reduced in the Tc1 mouse model of Down's syndrome." Nature **465**(7299): 813-817.
- Robertson, N., M. Oveisi-Fordorei, et al. (2007). "DiscoverySpace: an interactive data analysis application." Genome Biol **8**(1): R6.
- Romanuik, T. L., G. Wang, et al. (2009). "Identification of novel androgen-responsive genes by sequencing of LongSAGE libraries." BMC Genomics **10**: 476.
- Rossant, J. and C. McKerlie (2001). "Mouse-based phenogenomics for modelling human disease." Trends Mol Med **7**(11): 502-507.
- Roy, K., K. Kuznicki, et al. (2004). "The Tlx gene regulates the timing of neurogenesis in the cortex." J Neurosci **24**(38): 8333-8345.
- Roy, K., E. Thiels, et al. (2002). "Loss of the tailless gene affects forebrain development and emotional behavior." Physiol Behav **77**(4-5): 595-600.
- Roybon, L., T. Deierborg, et al. (2009). "Involvement of Ngn2, Tbr and NeuroD proteins during postnatal olfactory bulb neurogenesis." Eur J Neurosci **29**(2): 232-243.
- Sachidanandam, R., D. Weissman, et al. (2001). "A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms." Nature **409**(6822): 928-933.
- Saldanha, A. J. (2004). "Java Treeview--extensible visualization of microarray data." Bioinformatics **20**(17): 3246-3248.
- Sansom, S. N., D. S. Griffiths, et al. (2009). "The level of the transcription factor Pax6 is essential for controlling the balance between neural stem cell self-renewal and neurogenesis." PLoS Genet **5**(6): e1000511.

- Schatz, O., E. Golenser, et al. (2005). "Clearing and photography of whole mount X-gal stained mouse embryos." Biotechniques **39**(5): 650, 652, 654
- Schlake, T. and J. Bode (1994). "Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci." Biochemistry **33**(43): 12746-12751.
- Schmouh, J. F., K. G. Banks, et al. (2012). "Retina restored and brain abnormalities ameliorated by single-copy knock-in of human NR2E1 in null mice." Molecular and cellular biology **32**(7): 1296-1311.
- Schmouh, J. F., R. J. Bonaguro, et al. (2012). "Modelling human regulatory variation in mouse: finding the function in genome-wide association studies and whole-genome sequencing." PLoS genetics **8**(3): e1002544.
- Schuurmans, C., O. Armant, et al. (2004). "Sequential phases of cortical specification involve Neurogenin-dependent and -independent pathways." Embo J **23**(14): 2892-2902.
- Scott, C. E., S. L. Wynn, et al. (2010). "SOX9 induces and maintains neural stem cells." Nature neuroscience **13**(10): 1181-1189.
- Seth, P., I. Krop, et al. (2002). "Novel estrogen and tamoxifen induced genes identified by SAGE (Serial Analysis of Gene Expression)." Oncogene **21**(5): 836-843.
- Sherman, B. T., W. Huang da, et al. (2007). "DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis." BMC Bioinformatics **8**: 426.
- Shi, Y., D. Chichung Lie, et al. (2004). "Expression and function of orphan nuclear receptor TLX in adult neural stem cells." Nature **427**(6969): 78-83.

- Shimozaki, K., C. L. Zhang, et al. (2012). "SRY-box-containing gene 2 regulation of nuclear receptor tailless (Tlx) transcription in adult neural stem cells." The Journal of biological chemistry **287**(8): 5969-5978.
- Shizuya, H., B. Birren, et al. (1992). "Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector." Proceedings of the National Academy of Sciences of the United States of America **89**(18): 8794-8797.
- Siddiqui, A. S., J. Khattri, et al. (2005). "A mouse atlas of gene expression: large-scale digital gene-expression profiles from precisely defined developing C57BL/6J mouse tissues and cells." Proc Natl Acad Sci U S A **102**(51): 18485-18490.
- Siegert, S., E. Cabuy, et al. (2012). "Transcriptional code and disease map for adult retinal cell types." Nature neuroscience **15**(3): 487-495, S481-482.
- Siegert, S., B. G. Scherf, et al. (2009). "Genetic address book for retinal cell types." Nature neuroscience **12**(9): 1197-1204.
- Sim, F. J., H. M. Keyoung, et al. (2006). "Neurocytoma is a tumor of adult neuronal progenitor cells." J Neurosci **26**(48): 12544-12555.
- Singaraja, R. R., K. Huang, et al. (2011). "Altered palmitoylation and neuropathological deficits in mice lacking HIP14." Human molecular genetics **20**(20): 3899-3909.
- Slow, E. J., J. van Raamsdonk, et al. (2003). "Selective striatal neuronal loss in a YAC128 mouse model of Huntington disease." Human molecular genetics **12**(13): 1555-1567.
- Sopher, B. L. and A. R. La Spada (2006). "Efficient recombination-based methods for bacterial artificial chromosome fusion and mutagenesis." Gene **371**(1): 136-143.

- Soriano, P. (1999). "Generalized lacZ expression with the ROSA26 Cre reporter strain." Nature genetics **21**(1): 70-71.
- Speliotes, E. K., C. J. Willer, et al. (2010). "Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index." Nature genetics **42**(11): 937-948.
- Stenman, J., R. T. Yu, et al. (2003). "Tlx and Pax6 co-operate genetically to establish the pallio-subpallial boundary in the embryonic mouse telencephalon." Development **130**(6): 1113-1122.
- Stenman, J. M., B. Wang, et al. (2003). "Tlx controls proliferation and patterning of lateral telencephalic progenitor domains." J Neurosci **23**(33): 10568-10576.
- Stolt, C. C., P. Lommes, et al. (2003). "The Sox9 transcription factor determines glial fate choice in the developing spinal cord." Genes Dev **17**(13): 1677-1689.
- Stormo, G. D. (2000). "DNA binding sites: representation and discovery." Bioinformatics **16**(1): 16-23.
- Stoykova, A., D. Treichel, et al. (2000). "Pax6 modulates the dorsoventral patterning of the mammalian telencephalon." J Neurosci **20**(21): 8042-8050.
- Subramanian, L., A. Sarkar, et al. (2011). "Transcription factor Lhx2 is necessary and sufficient to suppress astrogliogenesis and promote neurogenesis in the developing hippocampus." Proceedings of the National Academy of Sciences of the United States of America **108**(27): E265-274.
- Sugatani, J., K. Yamakawa, et al. (2002). "Identification of a defect in the UGT1A1 gene promoter and its association with hyperbilirubinemia." Biochemical and biophysical research communications **292**(2): 492-497.

- Sun, G., K. Alzayady, et al. (2010). "Histone demethylase LSD1 regulates neural stem cell proliferation." Molecular and cellular biology **30**(8): 1997-2005.
- Sun, G., R. T. Yu, et al. (2007). "Orphan nuclear receptor TLX recruits histone deacetylases to repress transcription and regulate neural stem cell proliferation." Proceedings of the National Academy of Sciences of the United States of America **104**(39): 15282-15287.
- Swaminathan, S., H. M. Ellis, et al. (2001). "Rapid engineering of bacterial artificial chromosomes using oligonucleotides." Genesis **29**(1): 14-21.
- Tamamaki, N., K. E. Fujimori, et al. (1997). "Origin and route of tangentially migrating neurons in the developing neocortical intermediate zone." J Neurosci **17**(21): 8313-8323.
- Tang, L. S., H. M. Alger, et al. (2005). "Dynamic expression of COUP-TFI and COUP-TFII during development and functional maturation of the mouse inner ear." Gene expression patterns : GEP **5**(5): 587-592.
- Taranova, O. V., S. T. Magness, et al. (2006). "SOX2 is a dose-dependent regulator of retinal neural progenitor competence." Genes Dev **20**(9): 1187-1202.
- Tasic, B., S. Hippenmeyer, et al. (2011). "From the Cover: Site-specific integrase-mediated transgenesis in mice via pronuclear injection." Proceedings of the National Academy of Sciences of the United States of America **108**(19): 7902-7907.
- Tsai, K. Y., Y. Hu, et al. (1998). "Mutation of E2f-1 suppresses apoptosis and inappropriate S phase entry and extends survival of Rb-deficient mouse embryos." Molecular cell **2**(3): 293-304.

- Uemura, A., S. Kusuhara, et al. (2006). "Tlx acts as a proangiogenic switch by regulating extracellular assembly of fibronectin matrices in retinal astrocytes." J Clin Invest **116**(2): 369-377.
- Upton, A. L., N. Salichon, et al. (1999). "Excess of serotonin (5-HT) alters the segregation of ipsilateral and contralateral retinal projections in monoamine oxidase A knock-out mice: possible role of 5-HT uptake in retinal ganglion cells during development." The Journal of neuroscience : the official journal of the Society for Neuroscience **19**(16): 7007-7024.
- Vadolas, J., H. Wardan, et al. (2005). "Transgene copy number-dependent rescue of murine beta-globin knockout mice carrying a 183 kb human beta-globin BAC genomic fragment." Biochimica et biophysica acta **1728**(3): 150-162.
- Vencio, R. Z., H. Brentani, et al. (2004). "Bayesian model accounting for within-class biological variability in Serial Analysis of Gene Expression (SAGE)." BMC bioinformatics **5**: 119.
- Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science **291**(5507): 1304-1351.
- Vitalis, T., C. Alvarez, et al. (2003). "Developmental expression pattern of monoamine oxidases in sensory organs and neural crest derivatives." The Journal of comparative neurology **464**(3): 392-403.
- Vitalis, T., C. Fouquet, et al. (2002). "Developmental expression of monoamine oxidases A and B in the central and peripheral nervous systems of the mouse." The Journal of comparative neurology **442**(4): 331-347.

- Wahl, M. B., U. Heinzmann, et al. (2005). "LongSAGE analysis significantly improves genome annotation: identifications of novel genes and alternative transcripts in the mouse." Bioinformatics **21**(8): 1393-1400.
- Wallace, H. A., F. Marques-Kranc, et al. (2007). "Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence." Cell **128**(1): 197-209.
- Wang, L., H. Rajan, et al. (2006). "Histone deacetylase-associating Atrophin proteins are nuclear receptor corepressors." Genes Dev **20**(5): 525-530.
- Waterston, R. H., K. Lindblad-Toh, et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-562.
- Weyler, W., Y. P. Hsu, et al. (1990). "Biochemistry and genetics of monoamine oxidase." Pharmacology & therapeutics **47**(3): 391-417.
- Williams, A., N. Harker, et al. (2008). "Position effect variegation and imprinting of transgenes in lymphocytes." Nucleic acids research **36**(7): 2320-2329.
- Wilson, M. D., N. L. Barbosa-Morais, et al. (2008). "Species-specific transcription in mice carrying human chromosome 21." Science **322**(5900): 434-438.
- Wong, B. K., S. M. Hossain, et al. (2010). "Hyperactivity, startle reactivity and cell-proliferation deficits are resistant to chronic lithium treatment in adult Nr2e1(frc/frc) mice." Genes, brain, and behavior **9**(7): 681-694.
- Wrann, C. D., J. Eguchi, et al. (2012). "FOSL2 promotes leptin gene expression in human and mouse adipocytes." The Journal of clinical investigation **122**(3): 1010-1021.

- Yang, G. S., K. G. Banks, et al. (2009). "Next generation tools for high-throughput promoter and expression analysis employing single-copy knock-ins at the Hprt1 locus." Genomics **93**(3): 196-204.
- Yang, Y., S. Swaminathan, et al. (2003). "Aberrant splicing induced by missense mutations in BRCA1: clues from a humanized mouse model." Hum Mol Genet **12**(17): 2121-2131.
- Yokoyama, A., S. Takezawa, et al. (2008). "Transrepressive function of TLX requires the histone demethylase LSD1." Molecular and cellular biology **28**(12): 3995-4003.
- Young, K. A., M. L. Berry, et al. (2002). "Fierce: a new mouse deletion of Nr2e1; violent behaviour and ocular abnormalities are background-dependent." Behav Brain Res **132**(2): 145-158.
- Yozu, M., H. Tabata, et al. (2005). "The caudal migratory stream: a novel migratory stream of interneurons derived from the caudal ganglionic eminence in the developing mouse forebrain." The Journal of neuroscience : the official journal of the Society for Neuroscience **25**(31): 7268-7277.
- Yu, D., H. M. Ellis, et al. (2000). "An efficient recombination system for chromosome engineering in Escherichia coli." Proc Natl Acad Sci U S A **97**(11): 5978-5983.
- Yu, R. T., M. Y. Chiang, et al. (2000). "The orphan nuclear receptor Tlx regulates Pax2 and is essential for vision." Proc Natl Acad Sci U S A **97**(6): 2621-2625.
- Yu, R. T., M. McKeown, et al. (1994). "Relationship between Drosophila gap gene tailless and a vertebrate nuclear receptor Tlx." Nature **370**(6488): 375-379.
- Yu, Y. and A. Bradley (2001). "Engineering chromosomal rearrangements in mice." Nat Rev Genet **2**(10): 780-790.

- Yurchenko, E., H. Friedman, et al. (2007). "Ubiquitous expression of mRFP-1 in vivo by site-directed transgenesis." Transgenic research **16**(1): 29-40.
- Zambrowicz, B. P., A. Imamoto, et al. (1997). "Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells." Proc Natl Acad Sci U S A **94**(8): 3789-3794.
- Zhang, C. L., Y. Zou, et al. (2008). "A role for adult TLX-positive neural stem cells in learning and behaviour." Nature **451**(7181): 1004-1007.
- Zhang, C. L., Y. Zou, et al. (2006). "Nuclear receptor TLX prevents retinal dystrophy and recruits the corepressor atrophin1." Genes Dev **20**(10): 1308-1320.
- Zhao, C., G. Sun, et al. (2010). "MicroRNA let-7b regulates neural stem cell proliferation and differentiation by targeting nuclear receptor TLX signaling." Proceedings of the National Academy of Sciences of the United States of America **107**(5): 1876-1881.
- Zhao, C., G. Sun, et al. (2009). "A feedback regulatory loop involving microRNA-9 and nuclear receptor TLX in neural stem cell fate determination." Nature structural & molecular biology **16**(4): 365-371.
- Zheng, Y., S. Vertuani, et al. (2009). "Angiomotin-like protein 1 controls endothelial polarity and junction stability during sprouting angiogenesis." Circulation research **105**(3): 260-270.