Direct programming of medium spiny neuron (MSN) differentiation from human pluripotent stem cells (hPSCs)

Sali Abubaker Ali Bagabir

A thesis submitted for the degree of Doctor of Philosophy

September 2015

Supervisor:

Professor Nicholas D. Allen

Cardiff University

Cardiff, Wales, UK

School of Biosciences

Table of Content

List of figures

List of tables

Abstract

Striatal medium spiny neurons (MSNs) are the main output from the striatum, a subcortical part of the forebrain, which is the main input of the basal ganglia (BG) system. 96% of the striatum is composed of MSNs. Huntington's disease (HD) is caused by a progressive loss of MSNs in the striatum. It is caused by polyglutamine expansion in the Huntingtin protein (HTT). This impairs cerebral cortex function and deregulates several genes that play a role in subpallium development.

The identification and use of transcription factors (TFs) to direct the differentiation of stem cells to MSNs is described. Microarray data analysis of MSNs, from data in NCBIs Gene Expression Omnibus (GEO), was performed to detect gene expression profiles involved in telencephalon development and striatum maturation. The genes Dlx2, Gsx2, Mash1, Pax6, Sox4 and Foxp1 were found to play roles in neurogenesis, forebrain neuron fate commitment, cell proliferation, anatomical structure morphology, maturation of MSNs and transcriptional activation and repression.

A differentiation protocol was developed in which three TFs, DLX2, GSX2 and MASH1, were selected and cloned into expression vectors, in different combinations, to direct the differentiation of stem cells into naïve rosette neural progenitor cells (nrNPCs). These were then terminally differentiated into striatal MSNs.

Expression of DLX2, GSX2 and MASH1 in human embryonic stem cell (hESC) and induced pluripotent stem cell (iPSC) lines successfully directed their differentiation into nrNPCs. iPSC-derived nrNPCs were successfully terminally differentiated into DARPP-32$_{+ve}$ MSNs. However, only overexpression of DLX2 and MASH1 in iPSC-derived nrNPCs yielded functionally active MSNs that expressed DARPP-32, CTIP2, FOXP1, EBF1, DRD1, DRD2, GAD2 and CALBIN-1. It was successful and, therefore, could provide a new cell source for disease modeling in vitro, transplantation studies and drug discovery approaches.

Disclaimer

The electophysiological characterisation in Chapter 6 was undertaken by electrophysiologist Dr. Vsevolod Telezhkin.

Acknowledgments

In the name of ALLAH, the most Gracious and the most Merciful, this thesis has been made possible and successfully completed.

I am especially indebted to my supervisor, Professor Nicholas D Allen and Professor Paul Kemp for the insights I gained by resting on this research topic, and for their help and support.

Although many other individuals have given generously of their valuable time, Dr Rania Abubaker Bagabir and Dr Farhatullah Syed were particularly instrumental in reviewing and proof reading this thesis. I am very thankful for both of them in providing me with valuable insights at the final stage of the thesis, and boosting my morale during this research.

Finally, I would like to name my father, Dr Abubaker Ali Bagabir, husband, Taha Mohammed Bahubayshi and son, Mohammed Taha Bahubayshi, as de facto co-authors of this study. They have supported me and encouraged me on numerous occasions when I was faltering and frustrated. My husband and son have forgiven me for the impossible, long hours I have kept away at the Laboratory. They are always loving and understanding husband and son; they have given me the motivation and determination to complete my PhD.

I would like to personally thank my friends in the laboratory, who became my second family as I stayed with them more than I stayed with my family. These friends include Dr Vsevolod Telezhkin, Dr Alex Harrison, Dr Belinda Thompson, Dr Shona Joy, Dr Susannah Williams, Dr Shun Ming Yuen, Dr Emma Cope, Dr Charlie Geater, Dr Rachel Steeg, Dr Julia Griffiths, Dr Sarah Brennan and Dr Maarab Alqurashi.

I would also like to thank my sisters: Dr Dalia Bagabir, Dr Hala Bagabir, my younger sisters Shahd and Rgad Bagabir, my brothers: Dr Shehad and Mohammed Bagabir.

Special thanks to my father, Dr Abubaker Bagabir, in editing the final draft.

DEDICATION

Dedicated

to

my lovely parents,

my sisters, my brothers,

my husband and

my sweet son

forever keeping me

motivated and happy.

List of abbreviations

Ω	Ohm
1° Ab	Primary antibody
2° Ab	Secondary antibody
3'	Three prime
5'	Five prime
A	Adenine
AA	Amino acid
AEP	Anterior entopedunuclar area
Ab	Antibody
AD	Alzheimer diseases
ANR	Anterior neural ridge
AP	Anteroposterior
APC	Adenomatous polyposis coli
ASD	Autism spectrum disorders
ASCL1	Achaete-scute complex homolog 1 (Drosophila)
BCA	Bicinchoninic acid
BCL11B	Striatum of B-cell lymphoma/leukemia 11B
BDNF	Brain-derived neurotrophic factor
BED	Bilaminar embryonic disc
BF1	Brain factor 1
β-Gal	β-Galactosidase
bHLH	Basic helix-loop-helix
BMP	Bone morphogenetic protein
BMPRI	BMP Receptor type I
BMPRII	BMP Receptor type II
bp	Base pair
BrdU	Bromodeoxyuridine
BSA	Albumin from bovine serum
CALBIN-1	Calcium-binding protein-1
CDKI	Cyclin-dependent kinase inhibitors
CIP	Calf intestinal alkaline phosphatase
ChAT	Choline acetyltransferase
CHIP	Chromatin immunoprecipitation study
CMV	Cytomegalovirus promoter
CNK1	Casein kinase 1 signalling
CNS	Central nervous system
CO_2	Carbon dioxide
CP	Commissural plate
CrCd	Craniocaudal
Co-Smad	Common partner Smad protein
CTIP2	COUP TF1-interacting protein 2

Cx	Cortex
DARPP-32	Dopamine- and cAMP-regulated phosphoprotein, 32 kDa
Dd	Differentiation day
dH2O	Distilled water
dLGE	Dorsal LGE
DLX2	Distal-less homeobox 2
DKK1	Dickkopf WNT signaling inhibitor-1
dLGE	Dorsal LGE
DM	Dorsomorphin
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Dinucleotide triple phosphotase
DP	Dorsal pallium
DRD1	Dopamine D1-like receptors
DRD2	Dopamine D2-like receptors
Dsh	Dishevelled
DTT	Dithiothreitol
DV	Dorsoventral
EB	Embryoid bodies
EBF1	Early B-cell factor 1
EGFP	Enhanced green fluorescent protein
EMSA	Electromobility shift assay
ESCs	Embryonic stem cells
FACS	Fluorescence activated cell sorting
FB	Forebrain
FBS	Foetal bovine serum
FC	Fold change
FDR	False discovery rate
FGF	Fibroblast growth factor
Fox	Forkhead box
FOXG1	Forkhead box G1
FOXP1	Forkhead box protein P1
GΩ	Gigaohm
GABA	Glutamic acid decarboxylase
GAD2	Glutamic acid decarboxylase 2
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GEO	Gene expression omnibus
GFP	Green fluorescent protein
GO	Gene ontology
G-P	Glycyl-prolyl
GP	Globus pallidus
GPe	External segment of globus pallidus
GPi	Internal segment of globus
Gsh2	Homeobox protein GSH-2

GSK-3	Glycogen synthase kinase 3
GSK-3β	Glycogen synthase kinase 3β
GSX2	GS homeobox 2
HB	Hindbrain
HD	Huntington's disease
HD-iPSCs	Huntington's disease specific induced pluripotent stem cells
HDFs	Human dermal fibroblasts
HD-NSCs	Huntington's disease specific neural stem cells
HEK293	Human embryonic kidney 293
hESCs	Human embryonic stem cells
HF-PCR	High Fidelity - Polymerase Chain Reaction
HLH	Helix-loop-helix
hPSCs	Human pluripoten stem cells
HT	Hypothalamus
HTT	Huntingtin gene
ICC	Immunocytochemistry
ICM	Inner-cell mass
IddU	Iododeoxyuridine
IHC	In situ hybridization and immunocytochemistry
iN	Induced neural
IRES	Internal ribosome entry site
iPSCs	Induced pluripotent stem cells
IVF	In vitro fertilization
JNK	c-jun N-terminal kinase
Kb	Kilobases
KCl	Potasium chloride
µl	Microliter
µg	Microgram
L-G	L-Glutamine
LGE	Lateral ganglionic emmencience
LIF	Leukaemia inhibitory factor
LP	Lateral pallium
LRP5/6	Low density lipoprotein receptors-related protein 5/6 co-receptors
LT	Lamina terminalis
MΩ	Milliohm
M	Molar
MACS	Magnetic activated cell sorting
MB	Midbrain
MCS	Multiple cloning site
mg	Megagram
$MgCl_2$	Magnesium chloride
MGE	Medial ganglionic emmencience
$MgSO_4$	Magnesium sulfate
MEFs	Mouse embryonic fibroblasts
mefi	Irradiated mouse embryonic fibroblasts

mESCs	Mouse embryonic Stem cells
ml	Mililiter
mM	Milimolar
mm	Master mix
MM	Mismatch
MP	Medial pallium
MSNs	Medium-sized spiny neuron cells
mT	Melting temperture
MZ	Mantle zone
NaCl	Sodium chloride
NaI	Sodium Iodide
NCBI	National center for biotechnology information
NCX	Neocortex
ND	Nucleofection day
NEB	Neurogenic embryiod bodies
ng	Nanogram
NG108-15	Neuroblatoma-glioma hybrid cell
Ngn1/2	Neurogenin1/2
NGS	Normal goat serum
Notch-IC	Intracellular domain of Notch
NPCs	Neuron precursor cells
nrNPCs	Naïve rosette stage neural progenitors cells
OC	Optic chiasm
Olfm1	Olfactomedin 1
ORF	Open-reading frame
P	Phosphorylated
PBS	Phosphate-buffered-saline
PC12	Pheochromocytoma
pCAGG	pCAGG-IRES-EGFP
PCR	Polymerase chain reaction
PCX	Paleocortex
PD	Parkinson’s disease
PdD	Platting down day
PDL	Poly-D-lysine
PGD	Pre-implantation genetic diagnosis
Pen/Strep	Penicillin/ Streptomycin
pH	Power of hydrogen-Scale measures for acidic or basic substance
PLL	Poly-L-lysine
PM	Perfect match
PS	Primitive streak
PSB	Pallial-subpallial boundary
PZ	Preventricular proliferation zone
Q-PCR	Quantitative - polymerase chain reaction
qRT-PCR	Real-time quantitative reverse transcription polymerase chain reaction
R-Smads	Receptor-associated Smad proteins

RA	Retinoic acid
RARs	RA Receptors
RIPA	Radio-Immunoprecipitation Assay
RNA	Ribonucleic acid
ROCK	Rho-associated kinase
RP	Roof plate
rpm	Revolutions per minute
RT	Room temperature
RXRs	Retinoid X receptors
S	Septum
SC	Spinal cord
Sey/Sey	Small eye
SFSC	Serum-free suspension culture
SHH	Sonic hedgehog
SN	Substantia nigra
SNc	SN pars compacta
SNr	SN pars reticulata
ST14A cell	Embryoid striatum with variation of temperature sensitive of large antigen named T
STN	Subthalamic nucleus
SVZ	Subventricular zone
T	Tymidine
TAE	Tris-acetate-EDTA
TCF	T-cell factor
TE	Tris-EDTA
TED	Trilaminal embryonic disc
TEMED	Tetramethylethylenediamine
TFs	Transcription factors
TGF-β	Transforming growth factor-β
TGF-βRI	Type I TGF-β receptor
TGF-βRII	Type II TGF-β receptor
vLGE	Ventral LGE
VP	Ventral pallium
VZ	Ventricular zone
WNT	The Wingless protein ligand family
WR	The BCA working reagent
WT-iPSCs	Wild type specific induced pluripotent stem cells
x g	Times gravity

Chapter 1: General introduction

In Huntington's disease (HD), medium spiny neurons (MSNs) in the striatum are the population of neurons most affected by the disease, and they are subsequently lost. The development of a direct differentiation protocol, to derive MSNs from human pluripotent stem cells (hPSCs) would result in the production of unlimited numbers of MSNs that can be used for disease modeling and cell transplantation in HD studies in order to achieve neural network repair. Therefore, the development of a differentiation protocol to derive MSNs from hPSCs is of significant importance. Currently, no efficient protocol for directing differentiation of hPSCs into MSNs *in vitro* or *in vivo* is available. Here, hPSCs (human embryonic stem cells (hESCs) and induced pluripotent stem cells (iPSCs)) were used to develop a model for direct differentiation into striatal MSNs. This thesis describes the identification and cloning of the transcription factors (TFs) required for ventral telencephalon commitment and specification towards lateral ganglionic eminence (LGE), the striatum primordium. These were then used to transform hPSCs in order to produce mature striatal MSNs.

1.1 Human pluripotent stem cells (hPSCs)

The first hESCs line was derived in 1998, 17 years after the derivation of mouse embryonic stem cells (mESCs) (Evans and Kaufman 1981; Martin 1981; Thomson *et al*. 1998). hPSCs, which include hESCs derived from the inner-cell mass (ICM) of a blastocyst, and iPSCs, which are reprogrammed from somatic cells, have the ability to differentiate into thousands of cell types of the three germ lines (ectoderm, mesoderm and endoderm) and preserve their capability of self-renewal (Martin 1981; Evans and Kaufman 1981; Thomson *et al*. 1998; Takahashi and Yamanaka 2006; Takahashi *et al*. 2007). Both of these cell types are discussed in the following sections.

1.1.1 hESCs

hESCs are derived from pre-implantation embryos and propagated *in vitro*. The important use of hESCs lies in their three fundamental properties, namely their unlimited proliferation capacity, the ability to be genetically manipulated and differentiated into functional specialised cell types. These characteristics of hESCs make them an attractive tool for tissue engineering and repair, or disease modeling, which would allow a detailed understanding of underlying disease mechanisms (Chang and Cotsarelis 2007).

1.1.2 iPSCs

Human fibroblast cells and a number of other somatic cells can be genetically induced into a stem cell-like state. This was first done by the ectopic expression of four defined transcription factors (TFs), namely OCT3/4, SOX2, C-MYC and KLF4. These reprogrammed cells were termed induced pluripotent stem cells (iPSCs) and can be generated by several methods, as shown in **Figure 1.1** (Takahashi and Yamanaka 2006; Takahashi *et al*. 2007). This successful technology opened new possibilities in pluripotent stem cell research, such as establishment of disease-specific iPSCs lines for disease modeling, cell-based therapies and tissue engineering.

While iPSCs and hESCs share key properties, such as morphological appearance, unlimited proliferation potential and differentiation capacity (Gao *et al*. 2012; Chang and Cotsarelis 2007), iPSCs are ethically more acceptable as the procedure for generating iPSCs does not involve the destruction of a human embryo (Gao *et al*. 2012). There is some evidence that iPSCs retain the memory of their parent cells (i.e. the somatic cells from which they were generated) and display some of the non-CG methylation characteristic of somatic cells in the regions of the centromeres and telomeres, resulting in changes in gene expression (Gao *et al*. 2012; Lister *et al*.

2011). As one example, human iPSCs carrying the mutant Huntington (m*HTT*) gene have been derived and used for cell modelling **(Section 1.2.1.1)**.

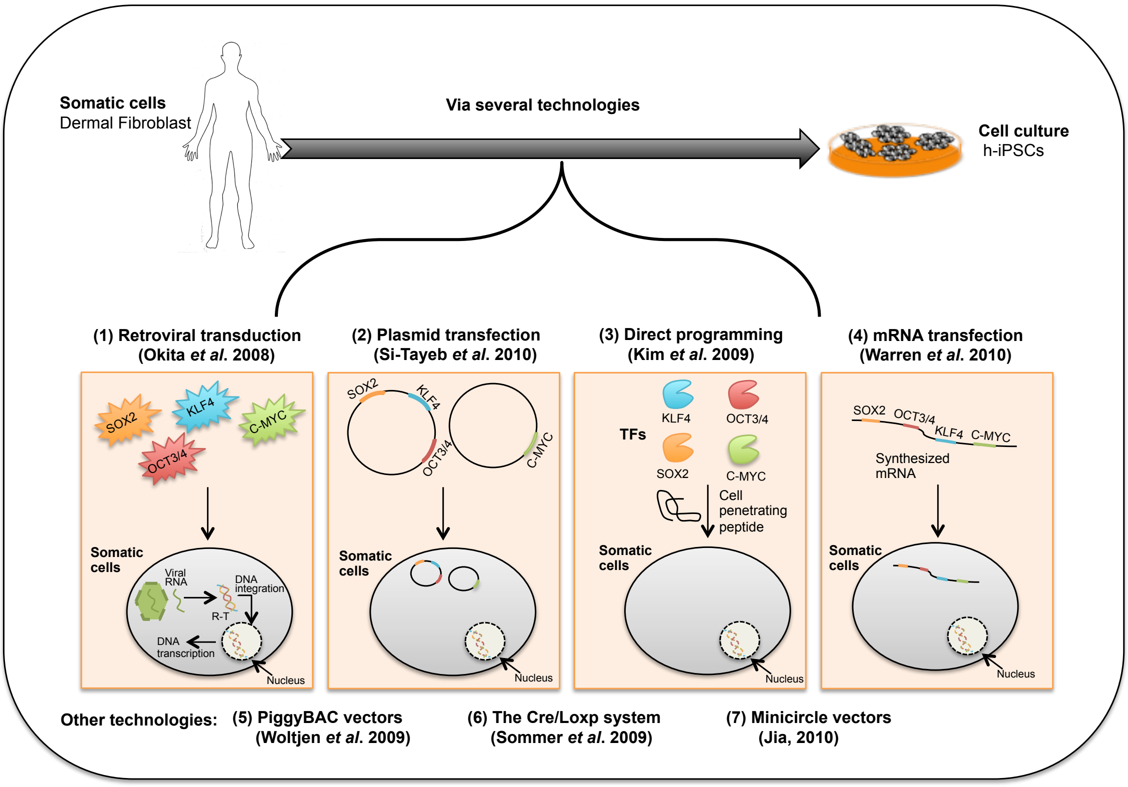


Figure 1.1: The established models for generating iPSCs through expression of TFs Oct3/4, Sox2, C-Mys and Klf4.

There are several ways to generate iPSCs from human somatic cells, such as (1) retroviral transduction, (2) plasmid transfection, (3) direct programming, (4) mRNA transfection, (5) piggyback vectors, (6) the Cre/Loxp system and (7) Minicircle vectors.

1.2 Huntington's disease (HD): symptoms and pathology

HD is an autosomal dominant inherited neurodegenerative disorder. The prevalence of HD is approximately 10 per 100 thousand in the UK (Ross and Tabrizi 2011). It is characterised by expansion of a tri-nucleotide repeat sequence (cytosine adenine guanine-CAG^n that encodes polyglutamine) in the Huntington (*HTT*) gene. Mutation occurs in the first exon of the Huntington gene (*IT15*), which encodes a 350 kDA HTT protein (Kelly *et al*. 2009; Connor 2011; Tauber *et al*. 2011; Benraiss and Goldman 2011). Clinically, this disorder is characterised by unconscious movement, cognitive impairment and psychological abnormalities (Walker 2007). Pathologically, it is characterised by loss of cortical neurons and striatal MSNs of the caudate nucleus and putamen, and abnormal growth of the ventricles and shrinkage of the overlying cortex **(Figure 1.2)** (Walker 2007; Kelly *et al*. 2009; Tauber *et al*. 2011; Benraiss and Goldman 2011). In the general population, the number of CAG tri-nucleotide repeats in the *HTT* gene is between 6 and 35. However, patients with HD have more than 35 CAG repeats (Tauber *et al*. 2011). The number of CAG repeats is linked to the age of disease onset. For example, in HD patients with 36 to 60 CAG repeats, the symptoms of HD manifest after the age of 35 years, whereas when the number of CAG repeats is more than 60, the symptoms of HD become apparent at a much younger age (Benraiss and Goldman 2011). In addition, when unaffected mothers and fathers have high numbers of CAG repeats, their offspring could be affected, as the CAG repeat is unstable on transmission (Ranen *et al*. 1995). Presently, there is no cure for HD. Current treatments are restricted to medications that decrease the symptoms of HD, such as, muscle relaxants, antidepressants and anticonvulsants (Ross and Tabrizi 2011).

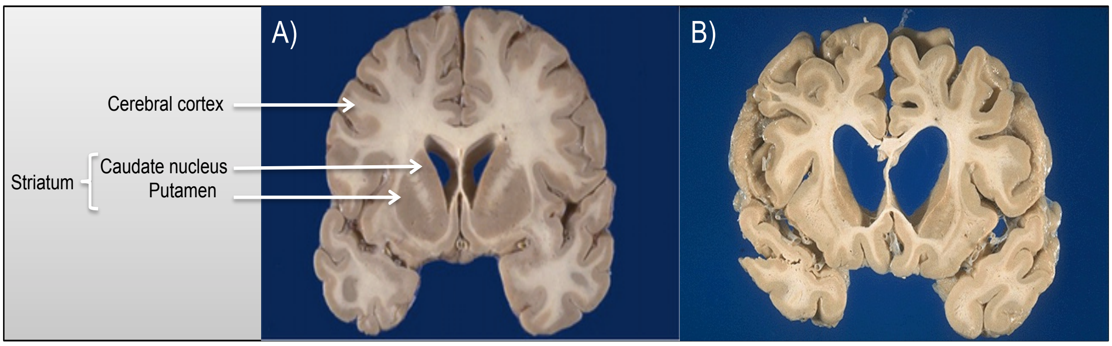


Figure 1.2: Comparison between a normal brain and HD brain.
Frontal section across a human brain, showing the normal adult brain (A), and the HD brain with deteriorate in the striatum (caudate nucleus and putamen) and the cerebral cortex (B) Figure taken from Benraiss and Goldman 2011.

The HTT protein is essential for normal brain function (Zheng and Diamond 2012). It undergoes post-translational modifications, such as acetylation and phosphorylation (Ross and Tabrizi 2011; Zheng and Diamond 2012). Although the exact function of HTT protein is still not fully known (Zheng and Diamond 2012), there is an increasing body of evidence describing its neuroprotective properties (Rigamonti *et al*. 2000; Cattaneo *et al*. 2005). The HTT protein can regulate RNA trafficking, gene transcription, intracellular trafficking, including membrane recycling, clathrin-mediated endocytosis, neuronal transport and postsynaptic signalling (Gutekunst *et al.* 1995; Ross and Tabrizi 2011; Sari 2011). However, the loss of function seen with wild-type HTT and the toxic gain of function with mHTT (which has neurotoxic properties) both contribute to HD pathogenesis (Cattaneo *et al*. 2001; Cattaneo *et al.* 2005). A large amount of evidence from both human HD (post-mortem) and animal HD models has revealed the different dysfunctional aspects of mHTT, including mitochondrial dysfunction (Cui *et al*. 2006; Kim *et al*. 2011), impaired axonal transport (Trushina *et al.* 2004), altered synaptic transmission, altered protein-protein interaction, glutamate- and dopamine-mediated excitotoxicity (Zeron *et al*. 2002), and most importantly, altered expression of transcription factors (Thomas *et al*. 2011). Such dysfunctional processes eventually lead to neuronal death or degeneration in HD (Johri and Beal 2012). Having said that a comprehensive understanding of the pathogenic mechanisms, with the aim of developing an effective therapy, is still emerging. In this respect, the development of cell and gene therapy is being considered as a new therapeutic approach to curing diseases such as HD (Zuccato e*t al*. 2010). Therefore, it is of importance to advance our knowledge of the cellular and molecular pathophysiology of HD, which becomes feasible through the establishment of human HD modelling using HD patient cells. This could consequently enhance future development of a cure.

1.2.1 Disease modeling of HD

Disease modeling can shed light on our understanding of the cellular and molecular pathogenesis resulting in clinical manifestation of HD. Importantly, better understanding of the disease mechanisms can facilitate the development of a potential cell-based therapy and also aid therapeutic drug screening.

Animal models of HD have been established by expressing whole m*HTT* or the N-terminal fragment that contains expanded polyglutamine (Sipione and Cattaneo 2001). Several animal models exist including, *Caenorhabditis elegans*, *Drosophila melanogaster*, mice, rats, sheep, pigs and monkeys (Sipione and Cattaneo 2001). In the past few decades, research on animal models of HD has added valuable information about HD pathogenesis. Nevertheless, animal models of HD have failed to provide a full understanding of HD pathogenesis, and hence it may not be possible to use those models to develop therapeutic approaches, which would prevent or slow down the HD progression in human subjects; therefore, cell-based models of HD could be preferable.

Several cellular models have been used previously as disease models to mimic the aspects of HD *in vitro*. The first approaches involved using transient expression of m*HTT* in Human Embryonic Kidney 293 (HEK293) or monkey kidney cells (Martindale *et al*. 1998). Later, models were developed that relied on neural–like cells rather than non-neuronal cells for long-term analysis. These include pheochromocytoma 12 (PC12), neuroblatoma-glioma hybrid cell (NG108-15) and embryoid striatum with variation of temperature sensitive large antigen named T (ST14A cell) (Li *et al*. 1999; Lunkes and Mandel 1998; Ehrlich *et al*. 2001). PC12 cells can be induced to differentiate into neuronal like cells by the presence of neuron growth factor. NG108-15 cells have the ability to differentiate into neuronal like cells, and ST14A cells are derived from E14 rat striatum and can display some characteristics of MSN subtype (Sipione and Cattaneo 2001; Ehrlich *et al*. 2001). From these HD cellular models, it was found that expansion of CAG repeats caused the *HTT* exon 1 protein to accumulate in the nucleus. This affected gene expression and resulted in abnormal cell morphology, a high rate of apoptosis and a deficiency in the development of neurites

(Li *et al*. 1999). In addition, the localisation of the m*HTT* in the nucleus occurs in a time- and CAG repeat unit number-dependent manner (Lunkes and Mandel 1998). Although these neurons were shown to mimic mechanisms of human cells in HD *in vitro*, more sophisticated models were required to recapitulate some characteristics of HD. Therefore, *in vitro* models of HD using disease-specific hPSCs were developed to fully recapitulate the pathogenesis of HD.

In 2009, two novel HD-hESCs were generated that carried mutant genes with 37 and 51 glutamate repeats, cell lines SI-186 and SI-187, respectively (Niclis *et al*. 2009). The cell lines with mutant alleles were derived from affected *in vitro* fertilization (IVF) embryos that were identified by pre-implantation genetic diagnosis (PGD) (Niclis *et al*. 2009). It was shown that the mutant *HTT* was expressed at the transcriptional and protein levels (Niclis *et al*. 2009). Therefore, both HD-hESC lines carrying the genetic mutation have the potential to exhibit HD pathology *in vitro*. In addition, it was demonstrated that HD-hESCs could be differentiated into the primary neurons and astrocytes that are associated with the pathology of HD, and were immunopositive to β-III Tubulin and GFAP, respectively (Niclis *et al*. 2009). The HD-hESC-derived primary neurons and astrocytes were frequently similar to those to HD-negative control hESC lines (HES2/HES3) (Niclis *et al*. 2009). The two HD lines were able to successfully differentiate into primary neurons and astrocytes (Niclis *et al*. 2009).

Although disease-specific hESCs provide a valuable knowledge of HD pathology, ethical and technical considerations make this a time-consuming and challenging disease model. The recent establishment of iPSCs in 2006 opened a new avenue for the generation and development of more sophisticated cell models to investigate and establish human therapies for complex diseases, such as HD, Parkinson’s disease (PD) and Alzheimer diseases (AD).

1.2.1.1 Human-cell modeling *in vitro* using iPSCs

Reprogramming of a patient's somatic cells, such as fibroblasts into iPSCs and their subsequent differentiation into target cells (neuronal cells in case of neurodegenerative disorders), would provide a wealth of information on disease pathology and potentially provide an unlimited source for autologous cell replacement therapy. The major advantage of iPSCs is their plasticity, the ability to acquire the morphological and functional properties of a wide range of cell types.

To date, modeling pathogenesis using disease-specific iPSCs has been established in diseases characterised by a single gene defect such as familial dysautonomia and autism spectrum disorders (ASD) (Lee *et al*. 2009; Marchetto *et al*. 2010), and rapid disease development in infants such as spinal muscular atrophy (Ebert *et al*. 2009; Gao *et al*. 2012).

Recently, direct reprogramming of monkey skin cells from HD monkey into iPSCs was successfully achieved by ectopic expression of the transcription factors Oct4, Kif4 and Sox2 (Chan *et al*. 2010). In addition, the first iPSCs derived from human HD patients were successfully produced by Park and colleagues in 2008. The HD-iPSCs were capable of differentiating into cells of the three germ layers, including neural cell types that demonstrated features of HD (Park *et al*. 2008).

In 2010 a human HD cell model was established. HD-iPSCs were successfully generated and induced to differentiate into neural stem cells (HD-NSCs) followed by differentiation into striatal neuronal precursors. The HD-NSCs were cultured with sonic hedgehog (SHH), dickkopf WNT signaling inhibitor-1 (DKK1) and brain-derived neurotrophic factor (BDNF) for 8-10 days, followed by treatment with BDNF, cAMP, valproic acid and Rho-associated kinase (ROCK) inhibitor (Y-27632), to prevent apoptosis. Further differentiation resulted in 10% of the cell population being positive for the most important MSN marker DARPP-32 (the dopamine and adenosine 3', 5'-cyclic monophosphate cAMP-regulated phosphoprotein of 32 kDa). In addition, the

CAG repeat was stable and there was an increase in caspase 3/7 activity, which is one characteristic of HD pathology; hence this human HD-cell model has the potential to use for drug screening (Zhang *et al*. 2010).

Camnasio and colleagues established HD-specific iPSCs using lentiviral technology to create an *in vitro* model of HD. A number of cell lines were derived from three patients, and three of these cell lines (two with a homozygous and one with a heterozygous genotype) were compared with wild type-iPSCs (WT-iPSCs) (Camnasio *et al*. 2012). The neural differentiation protocol used in this study is shown in **Figure1.3**. It was reported that the CAG repeats were stable in both HD and WT-iPSCs, and the caspase activity was the same in both HD and WT-iPSCs (Camnasio *et al*. 2012). Whereas, in the study of HD mouse models and HD patient brains, the CAG repeats were unstable (Gao *et al*. 2012). Hence, HD-iPSCs and HD-iPSC-derived neurons can reach maturation state and could be used as cell replacement therapy (Gao *et al*. 2012). Interestingly, the authors stated that the lysosomal activity in HD-iPSCs was four times greater than those in WT-iPSCs. They also suggest that the HD-iPSCs were capable of clearing out the mutant proteins using autophagosome-like structures (Camnasio *et al*. 2012).

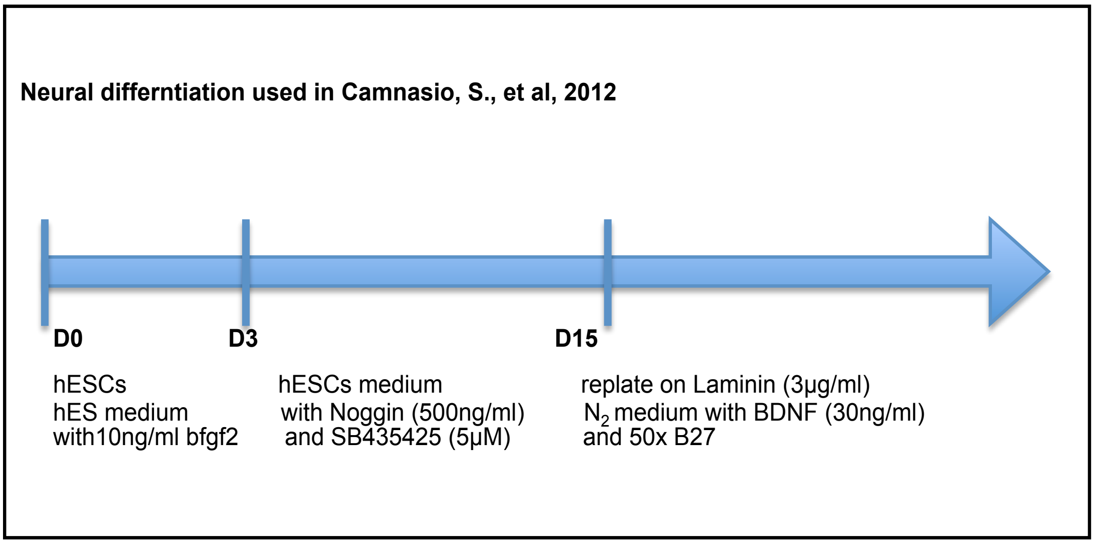


Figure 1.3: The neural differentiation protocol used by Camnasio and collaborators in 2012.

From Day (D)0 to D3, the hESCs were cultured with 10 ng/ml bfgf2, and between D3 and D15, the cells were neurally differentiated in culture media with 500 ng/ml Noggin and 5 μM SB435425. After D15, the cells were further differentiated with N_2 media with 30 ng/ml BDNF and 50x B27.

Joen *et al*. induced HD-iPSCs from an HD juvenile patient with 75 CAG repeats (HD75-h-iPSCs) and differentiated them into neurons by co-culture with PA6 stromal cells (Jeon *et al*. 2012; Jeon *et al*. 2014). The properties of HD75-iPSCs derived neural progenitors were examined and then transplanted into QA-Lesioned rat (Jeon *et al*. 2012) or YAC128 transgenic mice with 128 CAG repeats (Jeon *et al*. 2014) in order to investigate the role of HD75-h-iPSCs *in vivo* and analyse the motor behavior. Results showed that, a large proportion (75%) of the differentiated neurons derived from the HD75-h-iPSCs were immunopositive for striatal MSNs markers and exhibited the properties of functional GABAergic neurons (Jeon *et al*. 2012; Jeon *et al*. 2014). The HD75-h-iPSCs neural differentiation were transplanted into different models and showed similar outcomes. It was shown that the motor performance initiated to improve from 3 weeks following transplantation. At 12 weeks, the grafted cells were immunopositive for NESTIN and MAP2. It was also observed that the human cells differentiated into GABAergic MSNs that were DARPP-32^{+ve}, GABA^{+ve}, GAD6^{+ve} and SVP38^{+ve} (synaptic vesicle protein synaptophysin). The expression of SVP38 evidenced the ability of synapse formation in the human cells (grafted cells). In addition, the expression of aggregated mHTT protein (EM48) was not detected in the grafted cells, whereas it was expressed in the host cells. According to the double staining analysis at 12 weeks, the DARPP-32^{+ve} MSNs were affected by the immunopositive EM48, and hence it was suggested that the HD pathology was not developed in human cells or transmitted from host cells to the human cells (Jeon *et al*. 2014). However, at later cellular stages, the aggregates formation was detected (Jeon *et al*. 2012), as it would be expected to form in the transplanted YAC128 mice with HD75-h-iPSCs at later cellular stages (Jeon *et al*. 2014).

These data illustrate that HD cell models have the ability to differentiate into striatal MSNs and to have their HD phenotype corrected by genetic manipulation. However, the current protocols also generate non-specified neurons and the differentiated cells are not affected by HD to the same extent as the diseased cells (MSNs). Therefore, more reproducible and efficient protocols for MSN differentiation *in vitro* are required in order to investigate HD dysfunction at the cellular level.

1.3 Basal ganglia

The basal ganglion consists of the lateral ganglionic eminence (LGE), which gives rise to the striatum, the dorsal region, and the medial ganglionic eminence (MGE), which gives rise to the ventral region, the globus pallidus (GP) **(Figure 1.4 A, B & C)** (Pauly *et al*. 2013). Neocortical neurons migrate from the MGE through the LGE to their final targets (Sussel *et al*. 1999). For example, GABAergic (γ-aminobutyric acid) neurons, which are produced in the ventricular (VZ) and subventricular (SVZ) zones of the LGE and MGE, migrate throughout the SVZ to the cerebral cortex, olfactory bulb and hippocampus (Tamamaki *et al*. 1999; Wichterle *et al*. 1999).

Nkx2.1 and Gsx2 are two of the transcription factors that play a role in controlling and patterning the basal telencephalon, while Dlx1, Dlx2 and Mash1 have a role in differentiation. The interplay of these transcription factors regulates cortical GABAergic neuron output (Anderson 1997; Anderson *et al*. 1997; Szucsik *et al*. 1997; Casarosa *et al*. 1999; Horton *et al*. 1999; Sussel *et al*. 1999; Fode *et al*. 2000; Anderson *et al*. 2001; Pauly *et al*. 2013).

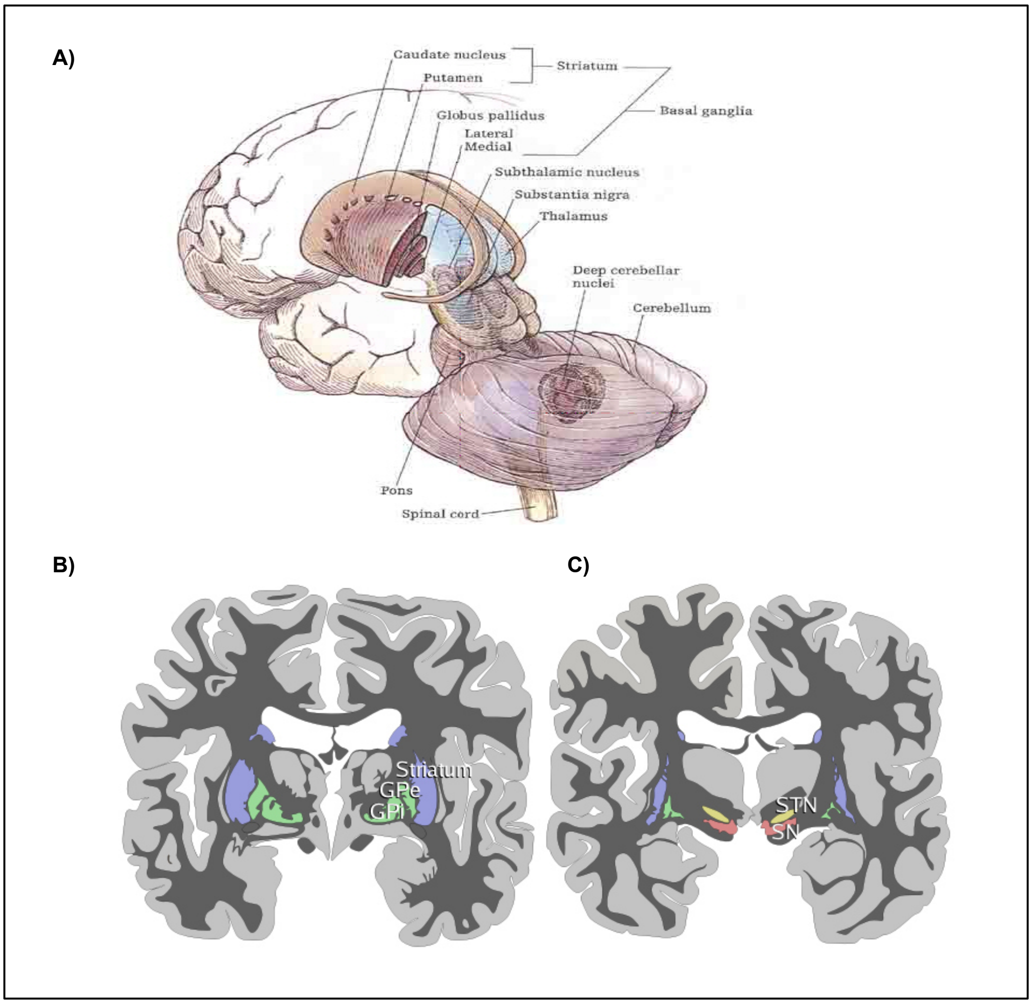


Figure 1.4: Structure of basal ganglion.

Human brain showing the structure of the basal ganglia. The basal ganglia consist of the LGE (striatum) and MGE (globus pallidus) (A). Coronal sections of human brain showing the basal ganglion components: striatum (caudate nucleus and putamen), globus pallidus (GPe and GPi) (B), SN and STN (C). Figure taken from Joseph 2011.

Abbreviations: LGE: Lateral ganglionic eminence, MGE: Medial ganglionic eminence, GPe: External segment of globus pallidus, GPi: Internal segment of globus pallidus. SN: Substantia nigra; STN: subthalamic nucleus.

1.3.1 Striatum

The striatum, known as the caudoputamen, is a complex of caudate nucleus and the putamen **(Figure 1.4 A & B)**. The striatum forms a major part of the basal ganglia along with the substantia nigra (SN), subthalamic nucleus (STN) and GP. It is organised into the striosome (patch) and matrix domains. The patch and matrix neurons receive input from deep and superficial parts of the neocortical layer, respectively. The patch cells pass their output to the SN pars compacta (SNc), whereas the matrix cells pass their output to the SN pars reticulata (SNr) (Gerfen 1992; Anderson *et al*. 1997; Feyder *et al*. 2011).

The functions of the striatum are to process information from the cerebral cortex, thalamus and SNc, and to send GABAergic outputs to the internal part of the GP (GPi) and SNr relating to voluntary movement, learning and cognition **(Figure 1.5)** (Gerfen 1992; Feyder *et al*. 2011).

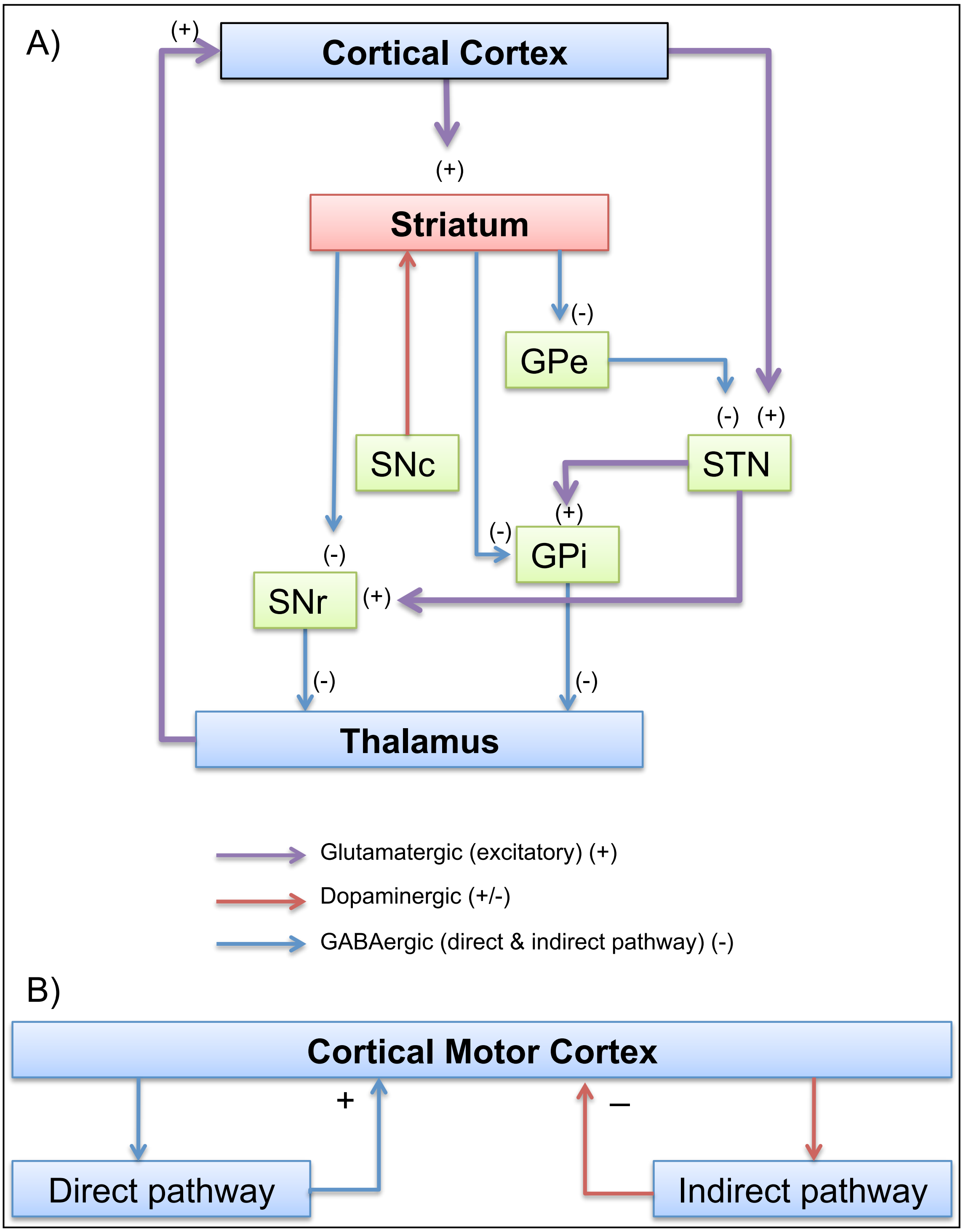


Figure 1.5: Medium spiny projection neurons in the basal ganglia.

This diagram shows the projection neurons of striatal MSN. Striatal MSN with DRD1 project neurons to GPi then to thalamus and go back to motor cortex (A) This pathway activates the motor activity (B). Striatal MSN with DRD2 project to GPe and STN then to GPi and thalamus then project back to the motor cortex (A). This pathway is indirect and decreases the motor activity (B). The red arrows indicate the dopaminergic pathway that acts as a positive feedback loop for the direct pathway and negative feedback loop for the indirect pathway to thalamus and then back to the cerebral cortex.

Abbreviations: GPe: external segment of globus pallidus, GPi: internal segment of globus pallidus, STN: the subthalamic nucleus, SNc: the pars compacta of substantia nigra, SNr: the pars reticulata substantia nigra, DRD1: dopamine D1-like receptors; DRD2: dopamine D2-like receptors.

1.3.2 Medium Spiny Neurons (MSNs)

The striatum's primary neurons are GABAergic medium-sized spiny neurons **(Figure 1.6)** that constitute around 90-95% of the striatal neurons in rats and over 85% in humans (Chang *et al*. 1982; Chang and Kitai 1985; Wictorin 1992; Kelly *et al*. 2009). These neurons originate from the LGE, while the small percentage of interstriatal neurons originate from the MGE (Olsson *et al*. 1998; Marin *et al*. 2000; Toresson and Campbell 2001; Arlotta *et al*. 2008). MSNs are characterised by complex dendritic arborisation and large dendritic spines receiving excitatory glutamatergic and modulatory dopaminergic inputs (Feyder *et al*. 2011; Penrod *et al*. 2011).

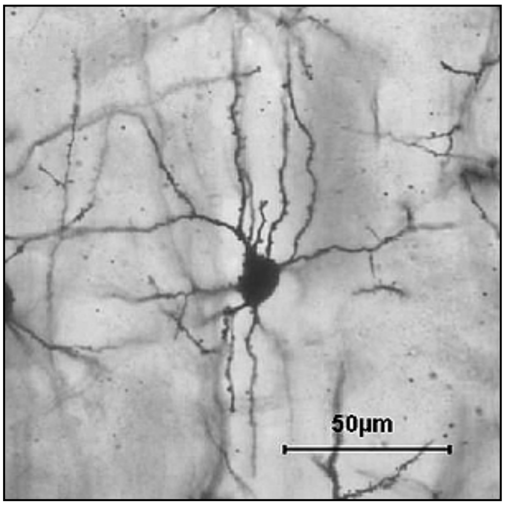


Figure 1.6: Morphology of medium-sized spiny neurons.
A Medium-sized spiny neuron with spiny dendrites. Figure taken from Churchill *et al*. 2004.

The MSNs in the striatum receive excitatory glutamatergic inputs from the cerebral cortex and dopaminergic input from the SNc. MSNs activity is regulated by dopamine, glutamate and acetylcholine (Ach). MSNs express two dopamine receptors, dopamine D1-like receptors (DRD1) and dopamine D2-like receptors (DRD2). These two groups of receptors have different expression patterns in the two projection pathways that connect the striatum toward the output of basal ganglion nuclei. For example, the MSNs with DRD1 project directly to the GPi and to the SNr, which then projects to the thalamus and sends signals back to the cortex **(Figure 1.5 A)**. Whereas the MSNs with DRD2 project indirectly to the GPi via a path along the external part of the globus pallidus (GPe) and the STN. The STN also receives input from the cortex and projects to the GPi which then sends signals to the thalamus and back to the cortex **(Figure 1.5 A)** (Nicola *et al*. 2000; Gong *et al*. 2003; Valjent *et al*. 2009).

The direct projection neuron pathway (striatonigral), with DRD1 MSNs, activates motor activity through disinhibition of thalamocortical neurons. Whereas activation of the indirect pathway (striatopallidal), with DRD2 from MSNs, decreases motor activity by increasing inhibition of thalamocortical neurons **(Figure 1.5 B)** (Feyder *et al*. 2011).

The MSNs can be distinguished by their expression of DARPP-32. DARPP-32 is enriched in the striatum and is the most commonly used marker of MSNs in the adult striatum, as it has been reported that more than 90% of MSNs express DARPP-32 (Anderson and Reiner 1991; Ouimet *et al*. 1984). Currently, DARPP-32 is the only available marker to distinguish MSNs from the other striatal neurons. The transcription factor COUP TF1-interacting protein 2 (CTIP2) was shown to be expressed in MSNs (Arlotta *et al*. 2008). It was observed that all the DARPP-32^{+ve} neurons expressed CTIP2 in the mouse striatum (Arlotta *et al*. 2008). Furthermore, it was confirmed that the CTIP2^{+ve} neurons co-expressed the TF forkhead box protein P1 (FOXP1). FOXP1 was detected in the developing mouse striatum from E13.5 by *in situ* hybridization and immunocytochemistry (IHC) and was expressed in the striatal projection neurons but not in the interneurons of the striatum (Tamura *et al*. 2004). It was also found that FOXP1 mRNA was expressed in the developing basal ganglia but not in the GP (Ferland *et al*. 2003; Tamura *et al*. 2004).

The DARPP-32 protein was first detected at E14 in the caudate nucleus of embryonic rat in a small percentage of cells by IHC (Foster *et al*. 1987). Over the next few days, the number of DARPP-32^{+ve} cells increased rapidly. By postnatal day 0 (P0), the developmental pattern of immunoreactivity of DARPP-32 was seen to be similar to that obtained in the adult striatum (Foster *et al*. 1987).

According to Ehrlich *et al*. (1990), Darpp-32 mRNA of was not detectable at E14 in the mouse brain. However, at P0, both Darpp-32 mRNA and protein were expressed in small amounts, and at postnatal weeks 3-4, Darpp-32 expression increased rapidly at both the transcriptional and protein levels. However, compared to mRNA, protein levels of DARPP-32 were decreased in the adult (Ehrlich *et al*. 1990). Moreover, it was suggested that the expression of Darpp-32 at transcriptional level occurs mostly during the last week of gestation, and subsequently Darpp-32 is present at P0. The induction of MSNs reaches its peak at E15 and is nearly complete by P1-2 in the rat and mouse (Marchand and Lajoie 1986; Ehrlich *et al*. 1990; Sturrock 1980). The MSNs originate from LGE (subpallium) and then migrate to the neocortex (NCX) (pallium) via the preventricular proliferation zone for motor activity.

1.3.3 Cerebral cortex

The cerebral cortex develops from the dorsal telencephalon. The projection neurones originate from the ventral telencephalon of the LGE and MGE in the cerebral cortex. Recently, it has been found that cerebral projection neurons originate from different proliferation zones. While some are derived from the cortical ventricular zone and then migrate into the cortical mantle, most of the interneurons and projection neurons of the cerebral cortex derive from the basal telencephalon and then transfer into the developing cortex (Anderson *et al*. 2001). In the early period of neurogenesis, at E11.5 to E14.5, cortical neurons cells migrate from the MGE through the piriform cortex (paleocortex–PCX) into the cerebral cortex. In the later stages of neurogenesis, at E.14.5 to E16.5, cells transfer from both the LGE and MGE via the preventricular proliferation zone (PZ) into the neocortex (NCX) **(Figure 1.7)** (Anderson *et al*. 2001).

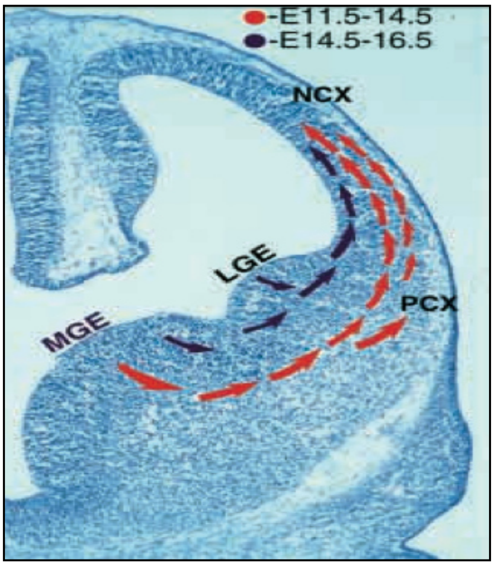


Figure 1.7: Cortical neuron migration from the MGE and LGE in the earlier and later stages of neurogenesis.

The red arrows show cortical neuron cell migration from the MGE into the neocortex (NCX) via the piriform cortex (PCX), in the early period of neurogenesis (E11.5 to E14.5). In the later period of neurogenesis (at E14.5 to E16.5), cortical cells migrate from the MGE and LGE into the NCX via the preventricular proliferation zone (blue arrows). Figure taken from Anderson *et al*. 2001.

1.4 Neural (brain) development

An efficient and reproducible protocol for directing the differentiation of hPSCs into MSNs *in vitro* is an important approach for disease modeling and drug screening. In order to establish an efficient method for direct differentiation of MSNs from hPSCs, an understanding of the TFs and signaling pathways that are essential in MSN development *in vivo* is fundamental. In this section, neural development, with a focus on forebrain (telencephalon) development, is reviewed.

In the first week of human embryonic development after fertilization, cleavage of the zygote, as it passes along the uterine tube, results in formation of the morula (Moore *et al*. 2011). The morula (12 to 16 cell stage) reaches the uterus 3-4 days after fertilization, and hence the formation of blastocyst occurs in the uterus **(Figure 1.8)**. The size of the developing embryo does not increase, although the number of blastomeres increases during the process of zygote cleavage, they are smaller than the parent cells. The zona pellucida is no longer present at the late stage of the blastocyst development **(Figure 1.8)**; hence the blastocyst starts to enlarge substantially. At day 6 after fertilization, the blastocyst attaches to the uterine endometrium, endometrial epithelium. At this stage the outer layer of the blastocyst, the trophoblast, starts to proliferate and differentiates into two layers, regulated by transforming growth factor-β (TGF-β). At day 7 after fertilization, the blastocyst is completely implanted in the endometrial epithelium. After implantation, the morphology of the ICM, known as the embryoblast, is changed and forms the bilaminar embryonic disc (BED), which consists of the epiblast and hypoblast. Extraembryonic structures, such as the yolk sac (umbilical vesicle), chorionic sac, amniotic cavity and amnion, start to form at the second week of human embryogenesis. The BED is located between the amniotic cavity and yolk sac and produces the germ layers, which form all the embryo cells, tissues and organs. Both the yolk sac and amniotic cavity produce morphogens for cell movements of the BED.

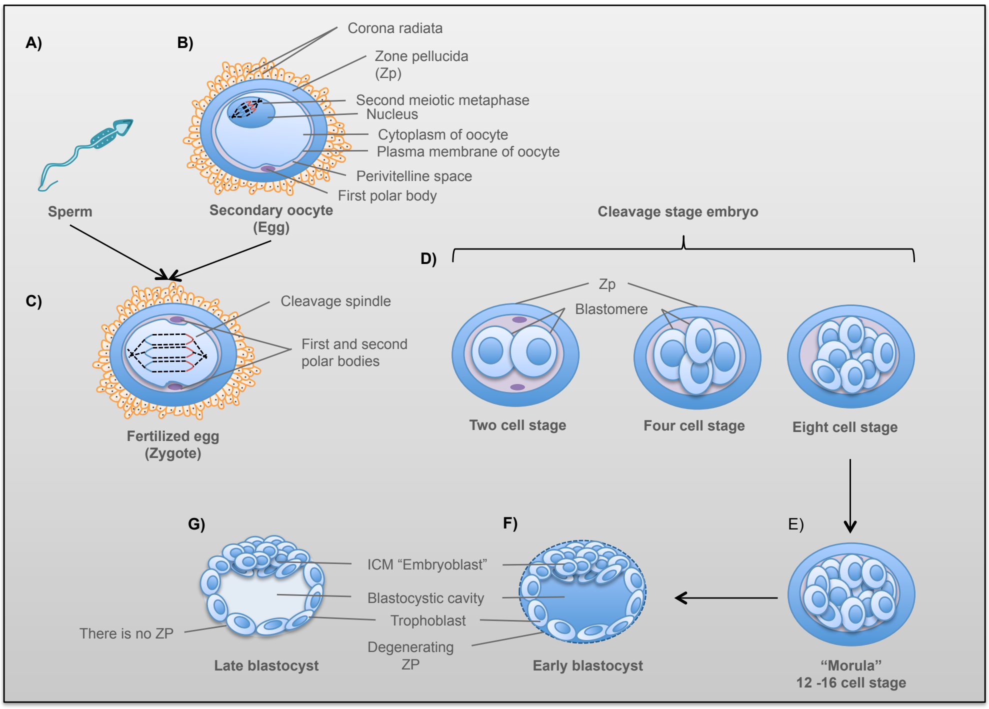


Figure 1.8: Formation of late blastocyst.

When the sperm (A) contacts the plasma membrane of secondary oocyte (B), the zygote (fertilized egg) is produced where the first mitotic division occurs containing the maternal and paternal chromosomes (C). This is followed by the cleavage stage of the zygote (D to E) as it travels along the uterine tube, and the formation of early blastocyst (F). When this reaches the uterus, the zona pellucida (zp) disappears and this is now known as the late blastocyst (G). The blastocyst enlarges considerably after the degeneration of zp.

The primitive streak (PS), notochord development and gastrulation take place during the third week of human embryogenesis (Moore *et al*. 2011). Gastrulation is the formation of the three germ layers, namely ectoderm, mesoderm and endoderm, and establishment of axial orientation. Each of the layers is responsible for production of particular cell types **(Figure 1.9)**. At the gastrulation stage, the BED is transformed into a trilaminal embryonic disc (TED). Morphogenesis starts in the third week of human embryonic development, driven by morphogens such as bone morphogenetic protein (BMP), fibroblast growth factors (FGFs), SHH and WNTs. The first step of morphogenesis results in the appearance of the PS located on the epiblast surface of the TED. The PS is established through proliferation and migration of the cells of the epiblast caudally to the middle level of the dorsal TED. Once it is formed, the craniocaudal (CrCd) and dorsoventral (DV) surface axes of the embryo are defined.

The mesenchymal cells are produced by PS and are located between the epiblast and hypoblast. The epiblast produces the embryonic ectoderm, the hypoblast produces the embryonic endoderm and the mesenchyme produces the embryonic mesoderm.

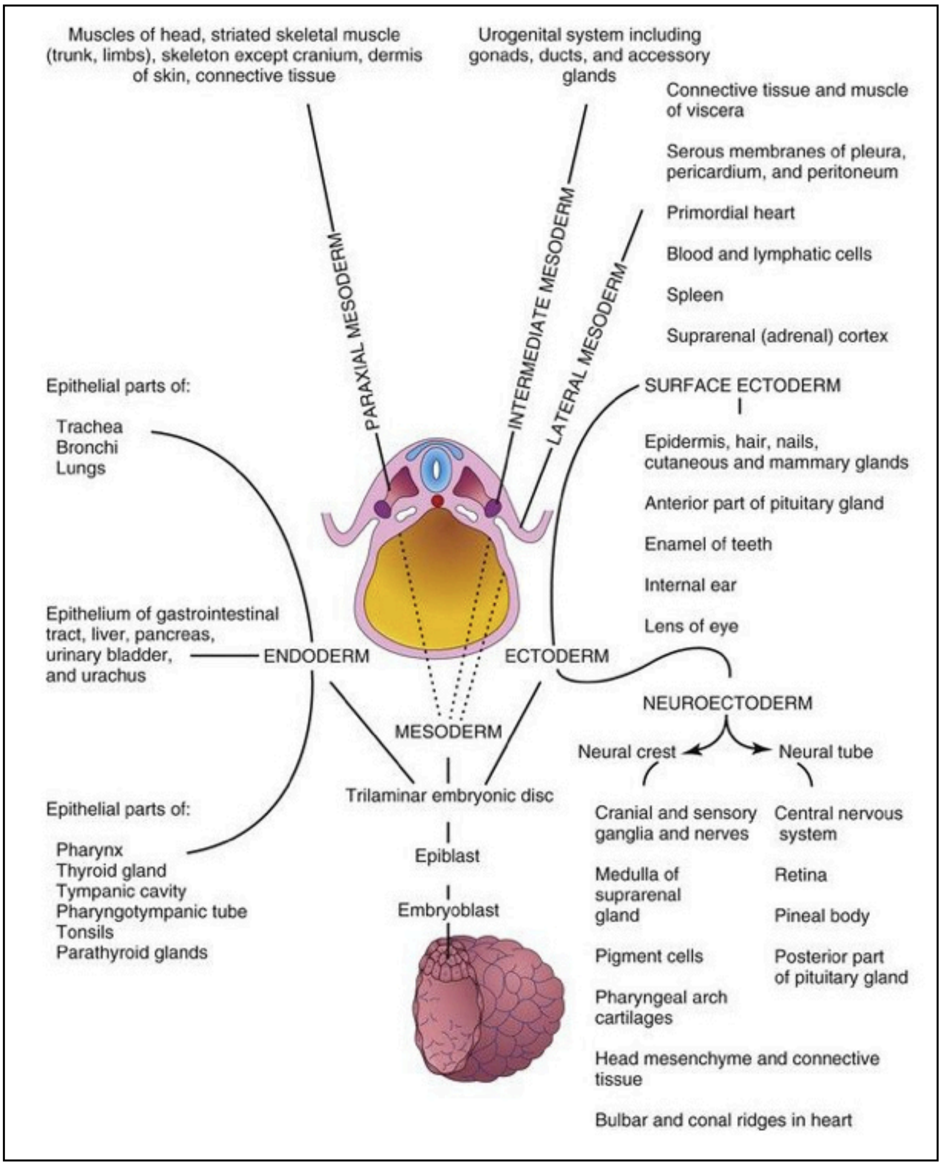


Figure 1.9: the three germ layers and their derivatives.

The inner cell mass (ICM), or embryoblast, gives rise to the embryo and forms the bilaminar embryonic disc (BED) which is transformed to the trilaminar embryonic disc (TED) at gastrulation stage. TED consists of endoderm, mesoderm and ectoderm. This diagram shows the derivatives of each layer in TED. Figure taken from Moore *et al*. 2011.

Following gastrulation, mesenchymal cells form between the ectoderm and endoderm form the notochord (Moore *et al*. 2011). The notochord promotes the thickening of the embryonic ectoderm by endogenous signals to form the neural plate. The neural plate folds and forms a central neural groove flanked by neural folds **(Figure 1.10)**. The neural folds fuse to form the neural tube in a process called neurulation **(Figure 1.10)**. The neural tube develops into the central nervous system (CNS).

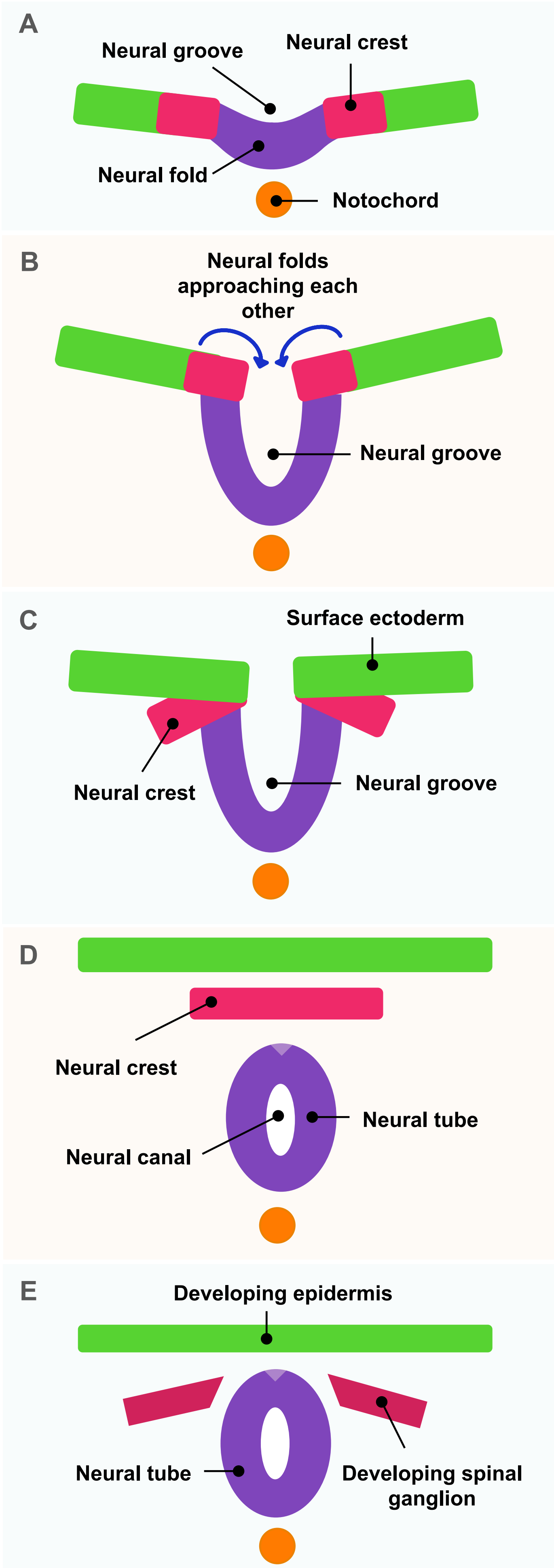

A
Neural groove
Neural crest
Neural fold
Notochord
B
Neural folds approaching each other
Neural groove
C
Surface ectoderm
Neural crest
Neural groove
D
Neural crest
Neural tube
Neural canal
E
Developing epidermis
Neural tube
Developing spinal ganglion

Figure 1.10: Neurulation stage.
This diagram shows the dorsal section of an embryo at 21 days of gestation (A), and the formation of the neural fold, neural groove, neural tube, neural crest and developing epidermis (B to F). Figure adapted from Moore *et al*. 2011.

The CNS in vertebrates comprises a large assortment of neurons and glial cells that are produced at specific times and in specific positions in the embryonic neural tube. The CNS is composed of three main cell types: neurons, astroglia and oligodendroglia. It consists of the brain (cranial end of neural tube) and spinal cord (caudal end of neural tube). During neural tube development, the tube differentiates into three parts: prosencephalon (forebrain), mesencephalon (midbrain) and rhombencephalon (hindbrain) **(Figure 1.11)**. Each of prosencephalon and rhombencephalon subdivides into two regions. The prosencephalon subdivides into the telencephalon and diencephalon, while the rhombencephalon divides into the metencephalon and myelencephalon, (Lumsden and Keynes 1989; Qiu *et al*. 1995; Rubenstein *et al*. 1998; Lumsden *et al*. 1991). The forebrain is the most rostral part of the neural tube; the telencephalon is the anterior part of the forebrain and is known as the most subdivided region of the CNS **(Figure 1.11)**. In the forebrain, there are several TFs that are expressed in limited patterns. These TFs are homeobox genes, 25 of which are found to be expressed in the forebrain (Rubenstein and Puelles 1994). These TFs include members of the Dlx, Emx, Otx and other families (Qiu *et al*. 1995).

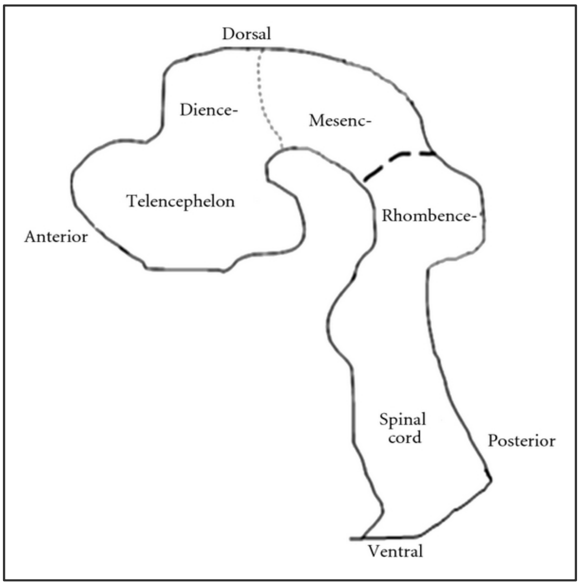


Figure 1.11: The neural tube differentiates into three parts.

Once the neural tube is formed, it differentiates into three domains. These are the prosencephalon (forebrain) consisting of the telencephalon and the diencephalon, the mesencephalon (midbrain) and the rhombencephalon (hindbrain). Figure taken from Evans *et al*. 2012.

During development, the brain is patterned by morphogens FGFs and retinoic acid (RA) along the anterior-posterior axis to form the forebrain (FB), midbrain (MB), hindbrain (HB), and spinal cord (SC). Each of these is further subdivided into several domains along the dorsal-ventral axis by specific morphogens such as WNTs, SHHs and BMPs **(Figure 1.12 A)**.

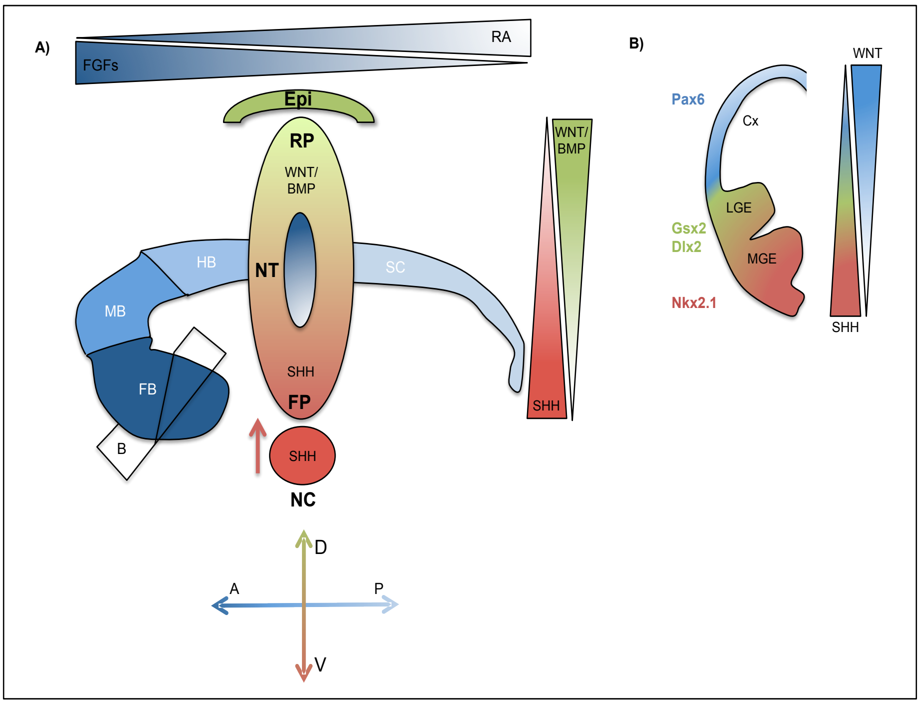


Figure 1.12: Patterning of brain according to the morphogenesis.

In the development of brain, NT, which differentiates into CNS, is patterned by morphogenesis of FGFs and retinoic acid (RA) along anterior to posterior axis to form forebrain (FB), midbrain (MB), hindbrain (HB), and spinal cord (SC), and each part of those are subdivided into several domains by dorsal to ventral axis by specific morphogenesis such as WNTs, BMPs (dorsal patterning) and SHHs (ventral patterning) (A). The structure of telencephalon that consists of pallium "Cortex" expressing the dorsal marker Pax6 and subpallium "basal ganglion: LGE and MGE" expressing the ventral markers Nkx2.1, Dlx2 and Gsx2 (B). The morphogenesis WNT and SHH have a role in the telencephalon patterning (B).

Abbreviations: NT: Neural tube, RP: Roof plate, FP: Floor plate, Epi: Epidermal development, NC: Notochord, SHH: Sonic hedgehog signaling, WNT/BMP: WNT and bone morphogenetic protein signaling, RA: Retinoic acid, FGF: Fibroblast growth factor, D: Dorsal, V: Ventral. LGE: Lateral ganglion eminence, MGE: Medial ganglion eminence; Cx: Cortex.

The telencephalon is subdivided into dorsoventral domains. The dorsal telencephalon, called "the pallium", develops into the cerebral cortex, and the ventral telencephalon, known as "the subpallium", develops into the basal ganglia. The homeobox genes, such as Pax6 and Emx2, play a significant role in the regulation of patterning and the proliferation of progenitor cells in the dorsal telencephalon (Warren *et al*. 1999; Bishop *et al*. 2003; Quinn *et al*. 2007; Pauly *et al*. 2013), whereas NKX2.1, Dlx2 and GSX2 have a role in the ventral telencephalon (Pauly *et al*. 2013; Carri *et al*. 2013) **(Figure 1.12 B)**.

In the following section, several factors that play an essential role in telencephalon development are discussed.

1.4.1 Factors involved in the development of telencephalon

Concerning the patterning of embryonic telencephalon, there are at least three patterning centres that are located in the developing telencephalon and provide definite signals, which have a role in the regulation of telencephalic morphogenesis and regional specification (Ohkubo *et al*. 2002; Storm *et al*. 2006; Crossley *et al*. 2001). The first patterning centre is the rostral patterning centre, which expresses certain genes of FGF family including FGF8, FGF18, FGF17 and FGF15 **(Figure 1.13)** (Storm *et al*. 2006). The second is the dorsal patterning centre, which expresses genes from WNT and BMP families **(Figure 1.13)** (Storm *et al*. 2006). Finally, the ventral patterning centre which expresses SHH **(Figure 1.13)** (Storm *et al*. 2006). During gastrulation, neurulation and rostral forebrain morphogenesis, FGF, SHH and BMP are located in the midline of the neural plate (Crossley *et al*. 2001; Storm *et al*. 2006). In addition, there is cross-regulation between them which regulates the patterning in the early development of embryonic telencephalon (Storm *et al*. 2006).

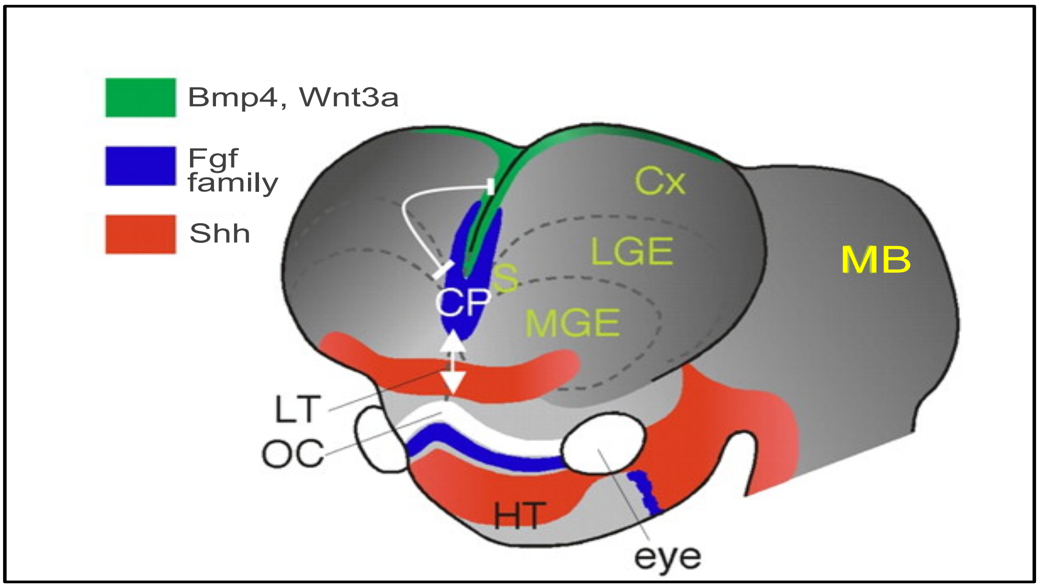


Figure 1.13: Schematic of patterning centers in the mouse telencephalon (frontolateral view).

This diagram shows the location of the three patterning centers in the telencephalon. It expresses the genes Bmp4/Wnt3a (green area), Fgf family (blue area) and SHH (red area). The cross regulation between the genes is also shown. Fgf family and SHH have a positive interaction, whereas Fgf family and Bmp4/Wnt3a have a negative interaction. Figure adapted from Storm *et al*. 2006.

Abbreviations: Cx: Cortex, LGE: Lateral ganglionic eminence, MGE: Medial ganglionic eminence, S: Septum, CP: Commissural plate, LT: Lamina terminalis, OC: Optic chiasm, HT: Hypothalamus; MB: Midbrain.

Numerous growth factors participate in the patterning of forebrain. One of them is FGF8 which has been shown to be expressed in the anterior neural ridge (ANR) and next to the tissues that express SHH and Bmp4 (Crossley *et al*. 2001; Storm *et al*. 2006). FGF8 has a function in cell proliferation, patterning of anterior-posterior (A-P) neural tube and regulating Foxg1 (forebrain marker) expression, which in turn promotes regionalisation and cell proliferation of telencephalic and optic vesicles (Shimamura and Rubenstein 1997; Storm *et al*. 2006). In a study of loss-of-function of FGF8, it was shown that the structure of LGE and MGE was lost and the expression of ventral genes Dlx2 and Nkx2.1 were reduced with an increase of dorsal gene expression Pax6 (Storm *et al*. 2006). It was therefore suggested that FGF8 has a role in fate commitment of telencephalon ventralisation (Storm *et al*. 2006).

The expression of FGF8 is maintained by the morphogen SHH (Ohkubo *et al*. 2002). SHH is produced from the notochord, located under the posterior area of the brain and prechordal plate, which in turn is located under the telencephalon domain in along a ventral to dorsal gradient (Rubenstein *et al*. 1998). SHH has a role in the patterning of the ventral neural tube, in embryonic development and in telencephalon development (Jessell 2000). It has been demonstrated that the concentration and duration of SHH signalling are essential for the fate commitment of ventral neural tube (Stamataki *et al*. 2005). According to Jessell (2000), SHH evokes its effect in a concentration-dependent manner. It was shown that in spinal cord patterning, SHH gradients generated five different classes of ventral neurons from ventral progenitor cells (four classes of ventral interneurons and one class of motor neurons) (Jessell 2000).

In the developing brain, it has been observed that SHH is expressed in the ventral area of telencephalon from E11.5 (Kohtz *et al*. 1998). It has a fundamental function in forebrain patterning, in particular the specification of ventral neuronal cell fate. It can also induce the expression of ventral markers in the telencephalon, such as Nkx2.1, Dlx and Islet-1/2 (Chiang *et al*. 1996; Ericson *et al*. 1995; Kohtz *et al*. 1998; Rallu *et al*. 2002). In a study of SHH knock out mice, it was observed that the structure of the ventral forebrain was lost with dorsalisation of the ventral telencephalon (Chiang *et al*. 1996). Conversely, another study showed that with overexpression of SHH, the expression domains of ventral definite genes, such as Nkx2.1, were expanded dorsally of the neural tube (Goodrich 1997). In addition, it was observed that when the telencephalon explant cultures were exposed to SHH, the ventral forebrain marker Nkx2.1 was expressed (Ericson *et al*. 1995). Furthermore, in a study of SHH misexpression in the cortex, it was found that ventral markers such as Nkx2.1 and Dlx2 were expressed ectopically (Kohtz *et al*. 1998).

The primary receptors for SHH are patched and smoothened. In the absence of SHH, patched inhibits downstream signalling from smoothened, whereas in the presence of SHH, it binds patched and blocks patched inhibition, and hence signalling from smoothened is activated. This activates glioma-associated oncogene homologs (Gli) 1, 2 and 3 (Gli1, Gli2 and Gli3), which translocate to the nucleus and induce transcriptional activation of target genes **(Figure 1.14)**.

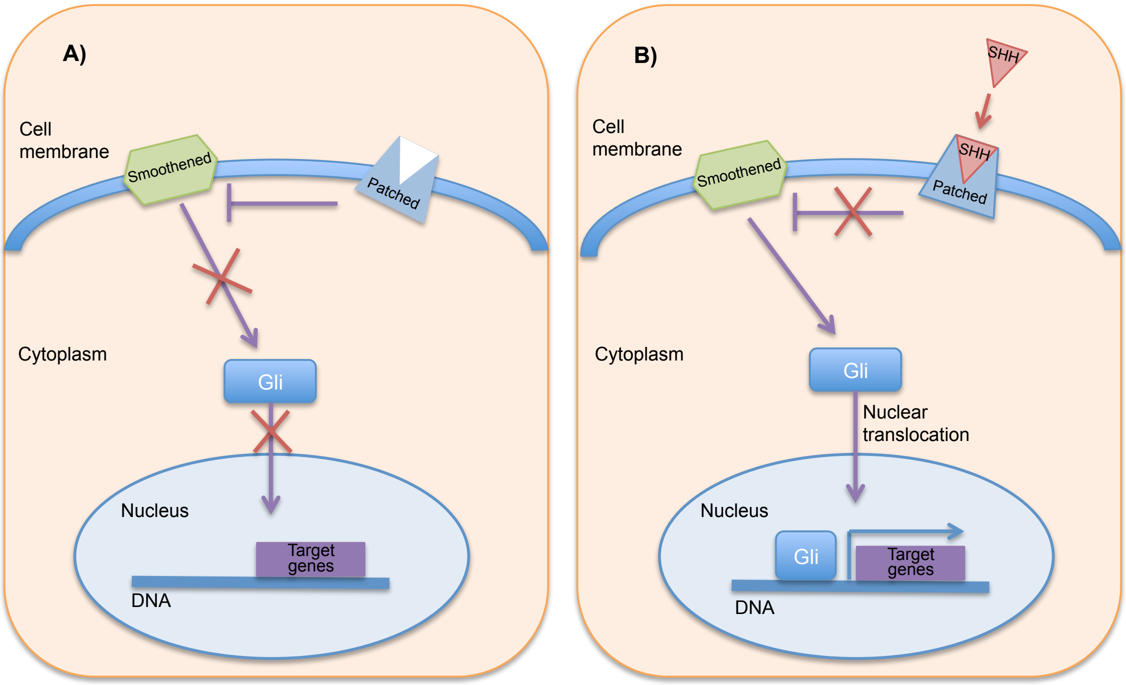
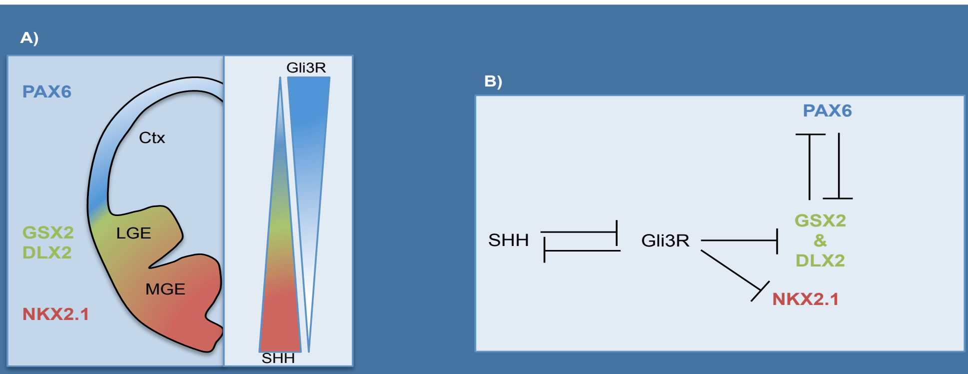


Figure 1.14 The SHH pathway of gene expression.
In the absence of SHH (A) patched inhibits downstream signaling from smoothened. However, in the presence of SHH (B) patched inhibition is blocked, and smoothened is free to activate Gli. Activated Gli is free to translocate to the nucleus and activates the transcriptional target gene expression.

It was shown that the SHH–dependent gene expression is regulated by the three members of Gli family of transcription factors, Gli1, Gli2 and Gli3 (Gulacsi and Anderson 2006). According to Stamataki and collaborators (2005), different concentration levels of SHH generate a gradient of Gli activity. It was generally accepted that Gli1 and Gli2 proteins stimulate the patterning of ventral telencephalon in response to SHH signalling. At a low concentration of SHH, the Gli3 protein is transformed from an activator form into a repressor form (Gli3R), which stimulates the patterning of dorsal telencephalon. Subsequently, in the patterning of telencephalon, the main function of SHH is to inhibit the formation of Gli3R and indirectly up-regulate expression of the ventral gene Nkx2.1, and the lateral ventral genes Dlx2 and Gsx2 **(Figure 1.15 A & B)**. In the study of double mutants of Gli1/Gli2, a normal patterning of telencephalon was observed **(Figure 1.15 C)**. Furthermore, in the study of double homozygous mutants SHH/Gli3 and Gli3$^{-/-}$, the patterning of ventral telencephalon was largely preserved with Nkx2.1 gene expression **(Figure 1.15 C)** (Gulacsi and Anderson 2006; Rallu *et al*. 2002).



C)

Loss of function models	Observation	References
Gli1$^{-/-}$ / Gli2$^{-/-}$	• Normal patterning of telencephalon	Rallu, M. *et al*. 2002
Gli3$^{-/-}$	• The boundary Pax6/Gsx2 was affected. • Gsx2 and Dlx2 was expanded dorsally	Rallu, M. *et al*.(2002
SHH$^{-/-}$	• Dlx2 and Gsx2 expressions (at low levels) maintained but displaced ventrally • Nk2.1 expression was lost	Rallu, M. *et al*. 2002
SHH$^{-/-}$ / Gli3$^{-/-}$	• Ventral telencephalon persisted • Dlx2, Gsx2 and Mash1 expressions were the same expressions as in wild type • The Pax6/Gsx2 boundary was rescued • Nkx2.1 expression was at low expression	Rallu, M. *et al*. 2002

Figure 1.15: Role of SHH and Gli3R in patterning of mouse telencephalon.

A diagram of a coronal section of mouse telencephalon (A) highlighting the three deferent domains via the dorso-ventral axis, which are cortex (ctx) (blue), LGE (green) and MGE (red). The homoedomain genes are shown on the left side of the diagram and colour-coded to indicate their expression in these domains and their fundamental role in the telencephalic patterning. The gradient of SHH and Gli3R are shown on the right of this figure. The SHH has a role in the telencephalic patterning by inhibition of Gli3R activity and promotes the expression of Nkx2.1, Dlx2 and Gsx2 indirectly (B). The gene regulation of Nkx2.1 is shown to be different from Dlx2 and Gsx2 gene expressions as it was expressed normally in the Gli3$^{-/-}$ mouse model (C) Gli3R promotes the dorsal telencephalon patterning (A, B) and expression of dorsal markers such as Pax6. The boundary of Pax6 and Gsx2 expression repress each other, where the Pax6 expression represses the expression of Gsx2 which is regulated by Gli3 as it was shown that, in the mouse model of Gli3$^{-/-}$, the expression of Pax6 was down-regulated whereas the expression of Gsx2 was expanded dorsally (C).

BMPs are a member of the secreted growth factor superfamily of TGFβ. BMPs are expressed from the roof plate (RP) of the neural tube and spread ventrally to promote the fate commitment of dorsal neurons in a concentration-dependant gradient (Dale and Wardle 1999). In neural induction of the neural tube, FGF signalling is required to repress the expression of BMP (Wilson and Houart 2004). It was observed that in the gain of function of BMP4, the expression of SHH and FGF8 were reduced in the telencephalon (Ohkubo *et al*. 2002). Moreover, in a culture of a telencephalon explant in the presence of BMPs, it was shown that the specification of forebrain was inhibited because the expression of Foxg1 was repressed as well as the expression of ventral expression markers, such as Nkx2.1 and Dlx2 (Furuta *et al*. 1997; Ohkubo *et al*. 2002).

In the BMP pathway, BMP ligands bind to their heterotetrameric receptors. These receptors consist of BMP receptor type I (BMPRI) and BMP receptor type II (BMPRII). BMPRI is an inactive domain whereas BMPRII is a phosphorylated and constitutively active receptor. Binding of the ligand to the transmembrane receptors triggers intracellular signalling. The BMPRI is transphosphorylated and becomes active hence phosphorylating the receptor-associated Smad proteins (R-Smads), consisting of Smad 1,2,3,5 and 8, which bind to common-partner Smad, Smad4 (Smad4). The phosphorylated R-Smad/Smad4 complex translocates to the nucleus and activates the transcriptional process with cofactors **(Figure 1.16).**

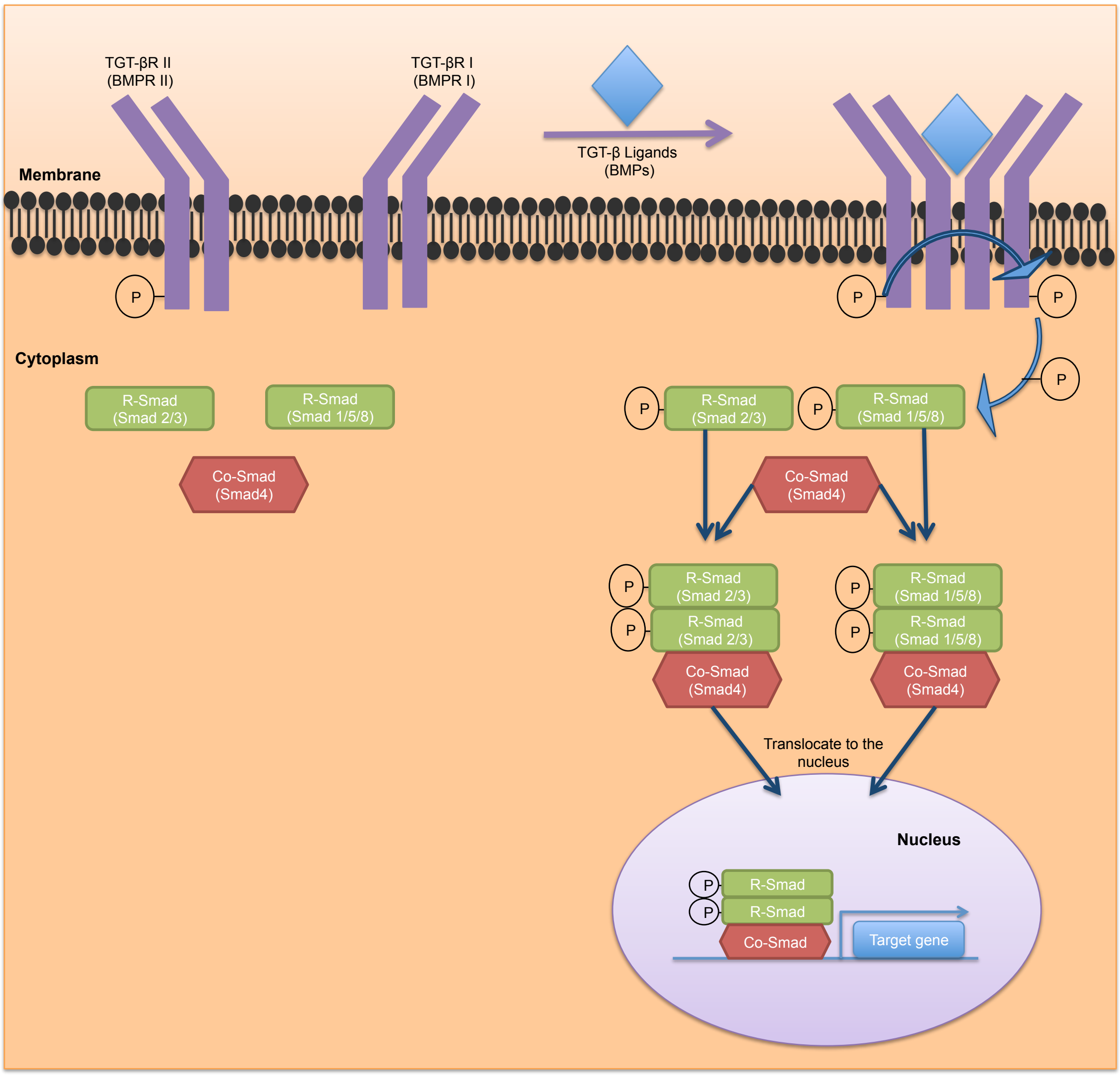


Figure 1.16: TGT-β/Samd signalling pathway

When TGT-β ligands, such as BMPs, bind to the BMP receptors, which include BMPRI and BMPRII, BMPRI is recruited and transphosphorylated. BMPRI becomes active and phosphorylates R-Smad, which includes Smad 1-2/5/8. The phosphorylated R-Smad binds to the Co-Smad and then translocates to the nucleus and activates transcription of target genes.

Abbreviations: TGT-β: Transforming growth factor-β, BMPs: Bone morphogenetic protein, TGF-βRI: Type I TGF-β receptor, TGF-βRII: Type II TGF-β receptor, BMPRI= Type I BMP receptor, BMPRII: Type II BMP receptor, P: Phosphorylated, R-Smad: Receptor associated Smad protein; Co-Smad: Common partner Smad protein.

Retinoic acid (RA) is the biologically active (bioactive) form of vitamin A. During forebrain development, RA has roles in cell survival, proliferation, differentiation and specification during development of the forebrain (Haskell and LaMantia 2005). RA is produced in the retina by the retinal aldehyde dehydrogenase enzyme **(Figure 1.17)** (Liao *et al*. 2005). It carries out its function via two receptors: RA receptors (RARs) and retinoid X receptors (RXRs). Each of these have α, β and γ subtypes, which can be recognised in the developing striatum (Liao *et al*. 2005). Both of RARβ and RXRγ are expressed selectively and mainly in the LGE domain. RXRγ is expressed at low levels in the MGE domain but it is not present in the CTX domain (Liao *et al*. 2005). In a study of a RXRγ homozygous mutant, expression of the choline acetyltransferase (ChAT) gene was reduced in the striatum region; however, GABA gene expression was not affected (Saga *et al*. 1999). Furthermore, in a study of $RAR\beta^{-/-}$ mice, striatal enriched tyrosine phosphatase mRNA, which is regulated by RA, was found to be reduced, (Liao *et al*. 2005). The number of Darpp-32^{+ve} neuron cells in the dorsal part of the striatum was also reduced (Liao *et al*. 2008).

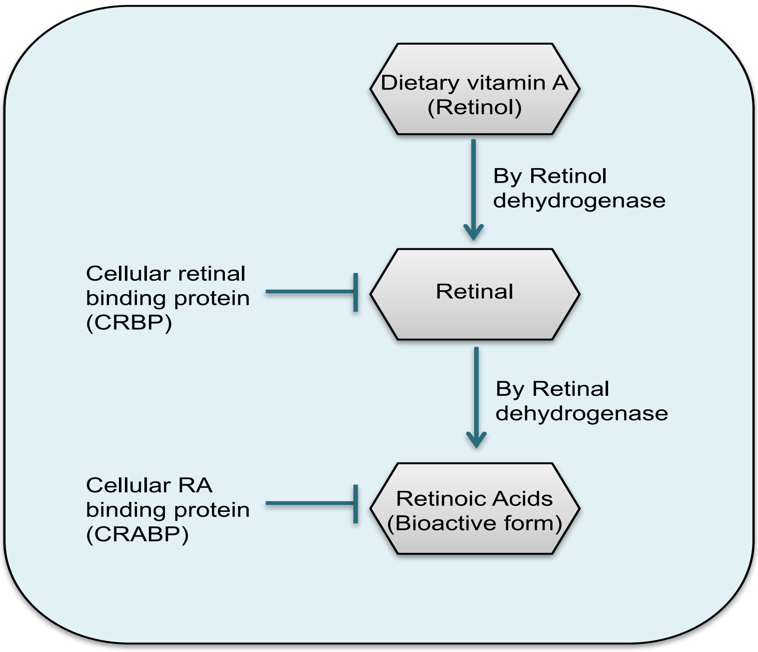


Figure 1.17: The RA synthetic pathway

Retinal is produced from retinol by the enzyme Retinol dehydrogenase. RA is then generated through further processing by retinal dehydrogenase. Two enzymes: CRBP and CRABP, block the formation of retinal and RA respectively.

It has been reported that RA is essential for development of the forebrain (Schneider *et al*. 2001). In the adult, RA expression is conserved in the domain of forebrain (Haskell and LaMantia 2005). In addition, it was found that retinoid signalling maintains expression of FGF8 and SHH in the forebrain (Haskell and LaMantia 2005; Schneider *et al*. 2001). When retinoid signalling is not present, loss of FGF8 and SHH expression was observed (Schneider *et al*. 2001).

In a study of Gsx2^{-}/$^{-}$ mice, it was observed that forebrain development was disturbed and Darpp-32 expression was reduced when the RA synthesis enzyme levels in LGE were reduced. However, it was reported that when the exogenous RA was added to the mutant mice, the expression of Darpp-32 increased. Consequently, it has been suggested that RA plays a role in regulation of DARPP-32 gene expression in the developing forebrain (Waclaw *et al*. 2004).

The wingless (WNT) protein ligand family is a large group of secreted glycoproteins that play a role in embryogenesis. WNT regulates the activity of β-catenin, thereby regulating the target gene expression (Huelsken and Behrens 2002). There are three WNT intracellular signalling pathways: (i) the canonical pathway that regulates the activity of β-catenin, (ii) the non-canonical planner cell polarity pathway that regulates the c-jun N-terminal kinase (JNK), and (iii) the non-canonical WNT/calcium pathway that regulates calcineurin (Huelsken and Behrens 2002). The canonical pathway is that which is involved in the development of telencephalon.

In the WNT canonical signalling pathway **(Figure 1.18)**, the WNT- protein ligand binds to the seven-transmembrane receptor Frizzled that activates intracellular signalling, leading to phosphorylation of Dishevelled (Dsh) (Huelsken and Behrens 2002; Moore *et al*. 2011). The phosphorylated Dsh blocks formation of a multiprotein complex of axin, glycogen synthase kinase 3β (GSK-3β) and adenomatous polyposis coli (APC) as well as phosphorylation of GSK-3β. The phosphorylated GSK-3β becomes inactive and subsequently cannot phosphorylate the cytoplasmic β-catenin, therefore preventing degradation of β-catenin. Hence, β-catenin accumulates in the cytoplasm and is subsequently translocated into the nucleus where it activates WNT target gene expression with T-cell factor (TCF). In the absence of WNT, the formation of multiprotein complex is initiated so that GSK-3β can phosphorylate the cytoplasmic β-catenin leading to its degradation. Consequently, the accumulation of β-catenin in the cytoplasm and its translocation to the nucleus is prevented and the TCF represses the target gene of WNT.

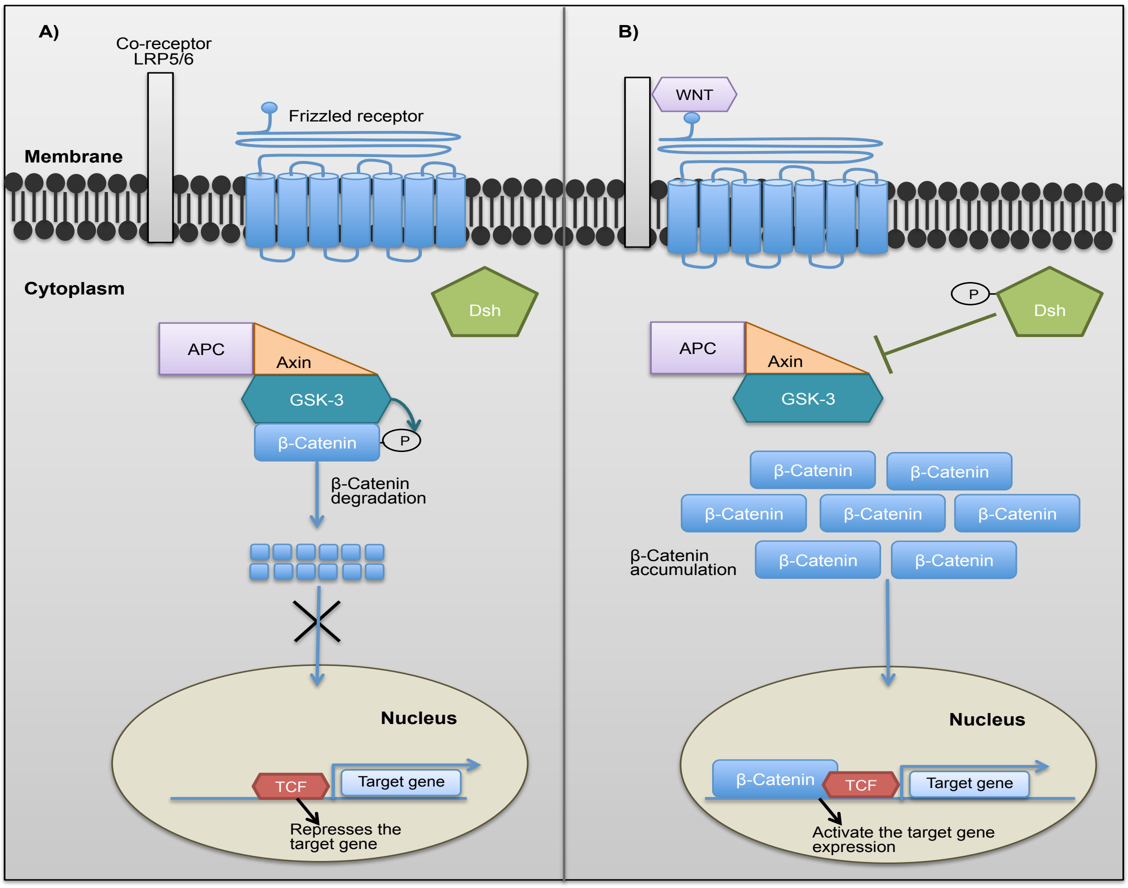


Figure 1.18: WNT/β-Catenin signaling pathway

When WNT is absent, the multiprotein complex is formed and GSK-3 phosphorylates the cytoplasmic β-Catenin, which is then degraded in the cytoplasm. TCF in the nucleus represses transcription of the target gene (A). When WNT is present, it binds to the seven-transmembrane frizzled receptor and co-receptor LRP5/6 is recruited and activates the intracellular signaling. Dsh is phosphorylated and blocks the formation of the complex such that GSK-3 cannot phosphorylate β-Catenin. Subsequently, β-Catenin accumulates in the cytoplasm and freely translocates to the nucleus where it activates the transcription of the target gene with the transcription factor TCF.

Abbreviations: LRP5/6: Low density lipoprotein receptors-related protein 5/6 co-receptors, APC: Adenomatous polyposis coli, GSK-3: Glycogen synthase kinase-3, Dsh: Dishevelled; TCF: T-Cell factor.

During murine development, canonical WNT signalling was observed in the dorsal telencephalon but not in the ventral domain at E11.5 and E16.5. This was demonstrated using a reporter line with the expression of lacZ gene under the control of responsive elements of β-catenin and TCF (Maretto *et al*. 2003; Backman *et al*. 2005). It was reported that inhibition of the WNT pathway is essential for generation of the telencephalon (Houart *et al*. 2002). Using a Cre-loxP system, Backman and colleagues examined the influence of canonical WNT signals before and after the onset of neurogenesis (Backman *et al*. 2005). Before neurogenesis, it was observed that the inactivation of β-catenin in the dorsal mouse telencephalon down-regulates the expression of dorsal markers such as Emx1/2 and Nng2, with expansion of ventral markers such as Dlx2, Mash1 and Gsx2. Conversely, when β-catenin was activated in the ventral mouse telencephalon, the ventral expression markers were down-regulated and the dorsal markers were expressed in the ventral telencephalon. However, after neurogenesis, it was observed that the canonical WNT signalling has no role in the dorsoventral fate commitment shift (Backman *et al*. 2005). Consequently, it was suggested that the role of canonical WNT is to specify the telencephalic cells by repressing expression of the ventral markers (Backman *et al*. 2005).

It is, therefore, clear that the exact levels of the certain factors in specific sections are essential in neural development. The specification of cells according to their position determines the appropriate phenotype through differentiation.

There are numerous classes of TFs that play a role in different stages of neuronal differentiation and the determination of different neuronal subtypes (Helms *et al*. 2005; Guillemot 2007). These include patterning proteins, progenitor proteins, proneural proteins, neuronal differentiation basic helix-loop-helix (bHLH) proteins, neuronal homeodomain proteins and inhibitory HLH proteins, which repress the expression of proneural genes (Guillemot 2007).

1.4.2 The TF expression of FOXG1 in the telencephalon development

Foxg1 was formally known as brain factor 1 (BF1) and was the first telencephalon marker to be identified (Tao and Lai 1992). Foxg1 is an evolutionarily conserved TF of the forkhead–box (Fox) (also known as winged–helix) family, and is expressed in the forebrain (Hanashima *et al*. 2002). In mammals, Foxg1 was first detected in the developing rat brain, and its expression was restricted to the telencephalon domain of the forebrain during embryogenesis (Tao and Lai 1992). In a study of rat forebrain development, it was observed that the expression of Foxg1 was at high level in the telencephalon domains from E10, and absent in the adjacent diencephalon domains as shown by *in situ* hybridization (Tao and Lai 1992). It was therefore concluded that Foxg1 has a role in development of the telencephalon and formation of the boundaries between the telencephalon and diencephalon.

A mouse model with a null mutation of Foxg1 was produced by replacing the majority of the coding sequence of Foxg1 with a LacZ gene and neomycin antibiotic resistance cassette (Xuan *et al*. 1995). The expression of the enzyme β-galactosidase (β-Gal) was controlled by the Foxg1 promoter (Xuan *et al*. 1995). The expression of Foxg1 was detected by X-Gal histochemistry in the neural tube from E8.5 and E9, and there was no difference between the wild type and the homozygous mutant mice (Foxg1^{-}/$^{-}$). However, the differences between them emerged from E10.5, the size of the telencephalon was tremendously reduced and even further impaired in E12.5 Foxg1 null embryos (Xuan *et al*. 1995; Martynoga *et al*. 2005).

A study of Foxg1^{-}/$^{-}$ mice revealed that the subpallium of telencephalon was more affected than the pallium of telencephalon compared to the wild type mice (Xuan *et al*. 1995). At E12.5, the Foxg1^{-}/$^{-}$ mice expressed the pallium markers, Pax6, Emx1 and Emx2; however, the subpallium markers Dlx1 and Dlx2, were not detected (Xuan *et al*. 1995). Moreover, morphological changes were observed; the structure of ganglionic eminences failed to form at E12.5 in the Foxg1^{-}/$^{-}$ mice (Martynoga *et al*.

2005). A similar outcome was observed in the study of mouse telencephalon, when the ventral telencephalon markers, such as Nkx2.1 and Mash1, were absent in the Foxg1^{-}/$^{-}$ mutant mice compared with the wild type mice at E9.5, as was the TF Gsx2 at E10.5 (Martynoga *et al*. 2005). Furthermore, the dorsal marker Pax3 was overexpressed in the Foxg1^{-}/$^{-}$ mutant compared with the wild type. The dorsal marker Pax6, which is normally expressed in the pallium telencephalon, was also overexpressed throughout the telencephalon (Martynoga *et al*. 2005).

In another study, cell proliferation in the Foxg1^{-}/$^{-}$ mouse model was analysed by bromodeoxyuridine (BrdU) labeling. It was found that while the precursor cells in the dorsal telencephalon were actively proliferating, the ventral telencephalic precursors cells were not (Xuan *et al*. 1995). Therefore, the ventral telencephalon failed to develop in the Foxg1^{-}/$^{-}$ mutant because the precursor cells were not proliferating (Xuan *et al*. 1995). In a more recent study, cells were double labelled with BrdU and iododeoxyuridine (IddU) in order to estimate the duration of the S phase in the cell cycle and the entire cell cycle. It was observed that the duration of cell cycle increased in the Foxg1^{-}/$^{-}$ telencephalon but remained unchanged in the ventral telencephalon domain (Martynoga *et al*. 2005). In addition, the rate of proliferation was reduced in the Foxg1^{-}/$^{-}$ mutant, consistent with results from the earlier study (Xuan *et al*. 1995).

Recently, it was shown that the TF Foxg1 coordinates the signaling pathway of SHH, which is independently essential for the development of subpallium telencephalon, and WNT/β-catenin, which is independently essential for the development of pallium telencephalon (Danesin *et al*. 2009). It has been established that Foxg1 represses the identity of the dorsal telencephalon by inhibition of the activity of WNT/β-catenin, confirming Foxg1 requirement for the induction of ventral telencephalon. However, SHH and Foxg1 independently play a role in the induction of the ventral telencephalon (Danesin *et al*. 2009).

1.4.3 The organisation of Dorsoventral (DV) pattern in the developing telencephalon

As mentioned earlier, the telencephalon is subdivided into two domains: the dorsal telencephalon that develops into cortex and the ventral telencephalon that develops into LGE and MGE. In terms of gene expression, there are well-defined boundaries between the DV domains and between the two subdivisions of the ventral telencephalon **(Figure 1.19)**. The TFs Pax6, Emx1/2 and Ngn1/2 are expressed in the dorsal telencephalon, while the TFs Mash1, Gsx2, Dlx1/2 and Nkx2.1 are expressed in the ventral telencephalon. The TF Nkx2.1 is specific to the MGE region while Mash1, Gsx2 and Dlx1/2 are expressed mainly (but not exclusively) in the LGE.

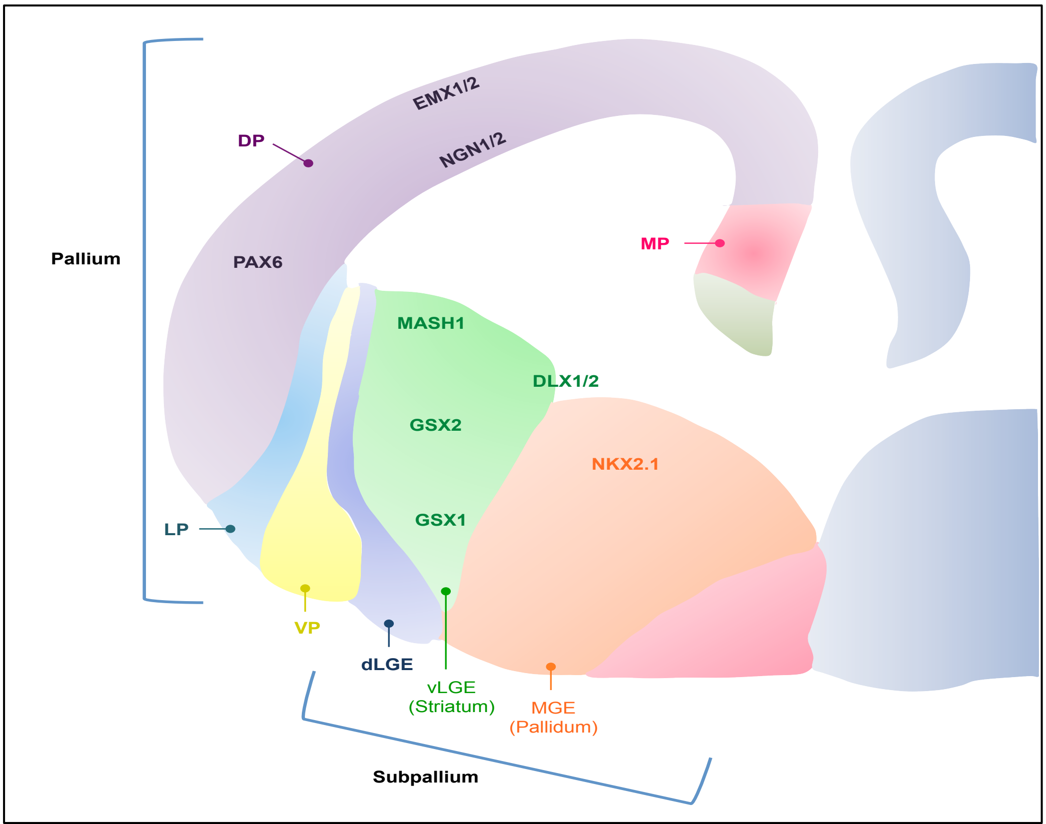


Figure 1.19: Schematic coronal section of the developing telencephalon at E12.5.

The pallium and subpallium telencephalon are shown and defined by specific gene expression patterns. The pallium telencephalon expresses the TFs Pax6, Emx1/2 and Ngn1/2, while the subpallium telencephalon expresses the TFs Mash1, Dlx1/2 and Gsx2 mainly in LGE domain and the TF Nkx2.1 specifically in the MGE region. Figure adapted from Schuurmans and Guillemot 2002.

Abbreviations: DP: Dorsal pallium, VP: Ventral pallium, MP: Medial pallium, LP: Lateral pallium, dLGE: Dorsal LGE; vLGE: Ventral LGE.

The TF expression profile for the pallium telencephalon includes Pax6, Emx1/2 and Ngn1/2. Pax6, a paired homeodomain gene, is essential for the proliferation of precursor cells in the developing cerebral cortex and for the development of cortical progenitors (Toresson *et al*. 2000). Another marker of the pallium telencephalon is Emx1. Its expression is restricted to the dorsal part of cortex and, therefore, it is absent in the ventral area of the cortex **(Figure 1.19)**. Neurogenin1/2 (Ngn1/2) is known as neurogenic basic-helix-loop-helix (bHLH) TF and is expressed throughout the pallium alongside Pax6. It is essential for determining the phenotype of pallium telencephalon.

The expression of Pax6 was observed at E8 in the developing forebrain (Stoykova and Gruss 1994). In $Pax6^{-/-}$ mutant mice, known as the small eye phenotype (Sey/Sey), it was observed that the patterning of forebrain and the development of cortex were defective. The pallial-subpallial boundary (PSB) was also shifted (Toresson *et al*. 2000). Moreover, it was observed that the expression of dorsal markers such as Ngn1/2 and Emx1/2 were down-regulated, whereas the LGE ventral markers such as Gsx2, Mash1 and Dlx1/2 were ectopically expressed throughout the pallium region of the mutant telencephalon. Nkx2.1 expression, which is normally restricted to the MGE domain, was expanded into the LGE domain, thereby shifting the boundary between LGE and MGE regions (Toresson *et al*. 2000; Stoykova *et al*. 2000).

The TFs Pax6 and Gsx2 play a role in the dorsoventral patterning of telencephalon and the formation of the PSB (Yun *et al*. 2001; Toresson *et al*. 2000). The PSB is defined as the border of the ventral pallium and the dorsal LGE (dLGE) (Yun *et al*. 2001). In the region of the dLGE, expression of both Pax6 and Gxs2 were overlapping at E10.5 of murine development (Yun *et al*. 2001). A previous study showed that the PSB is comprised of a subgroup of progenitor pools of olfactory bulb interneurons and cortical neurons. In addition to this, expression of both Pax6 and Gsx2 was required for specification of PSB progenitors (Carney *et al*. 2009).

In a study of $Gsx2^{-/-}$, Sey/Sey and double mutant Gsx2/Sey mice models, it was shown that Gsx2 repressed the expression of Pax6 in the subpallium telencephalon and maintained expression of the ventral markers Mash1 and Dlx1/2. Meanwhile, Pax6 inhibited the expression of Gsx2 in the pallium telencephalon and maintained the expression of the dorsal markers Emx1/2 and Ngn1/2 (Toresson *et al*. 2000). In addition, Gsx2 repressed the dorsal markers Emx1/2 and Ngn1/2, whereas Pax6 inhibited the ventral markers Mash1 and Dlx1/2. According to Fode and colleagues, Ngn1/2 repressed the expression of Mash1 and Dlx1/2 in the developing pallium (Fode *et al.* 2000). It has, therefore, been suggested that Pax6 inhibits the expression of ventral markers Mash1 and Dlx1/2 via the expression of Ngn1/2 (Toresson *et al*. 2000). However, in mice model when Gsx2 was present, the expression of Mash1 and Dlx1/2 was not required for the inhibition of Ngn1/2 (Toresson *et al*. 2000). Consequently, it was concluded that the function of Gsx2 in the development of telencephalon was essential to repress Pax6 and to maintain the identity of subpallium domain (Toresson *et al*. 2000). Pax6 and Gsx2 have complementary roles in generating the pallium-subpallium boundary in the developing telencephalon and in specification of precursor cells in the cortex and striatum (Toresson *et al*. 2000).

1.4.3.1 GSX2 (GS homeobox 2)

The GSX2 gene is located on chromosome 4 at 4q12. It was formerly known as homeobox protein GSH-2 (GSH2) (Hsieh-Li *et al.* 1995). Gsx2 belongs to the homeobox TF family and is expressed beginning at E9 and E10 in the developing forebrain (Corbin *et al*. 2000). GSX2 and GSX1, another homeobox gene located on chromosome13q.2, are the earliest TFs expressed in the LGE progenitor cells (Pei *et al*. 2011).

The homeobox Gsx genes are involved in the initial specification of the neural progenitors of the LGE (Pei *et al*. 2011). Gsx2 and Gsx1 have a similar function in LGE patterning. However, they play different roles in the balance between proliferation and differentiation in LGE progenitor cells (Pei *et al*. 2011). Gsx2 expression controls maintenance of the undifferentiated phase of the neural progenitors of the LGE, while Gsx1 expression supports the maturation of the progenitor cells via down-regulation of Gsx2 (Pei *et al*. 2011). At embryonic day E12.5, Gsx2 is more highly expressed in the neuronal progenitors of dorsal LGE at the boundary of VZ, than in the ventral LGE and MGE, while Gsx1 is mainly expressed in the ventral LGE and MGE progenitor cells at the boundary of VZ and SVZ (Toresson *et al*. 2000; Yun *et al*. 2003; Pei *et al*. 2011). Recently, it has been found that Gsx1 is expressed in areas where the expression of Gsx2 is low, such as in ventricular LGE and MGE. In the Gsx1 mutant mice, during the late phases of neurogenesis, the expression of Gsx2 is increased in the ventral LGE. Also, when Gsx1 is overexpressed, the expression of Gsx2 ceases. Therefore, it was suggested that Gsx1 could be a repressor of Gsx2 expression (Pei *et al*. 2011). Consequently, the Gsx2 gene expression gradient along the dorsal to ventral axis of telencephalic LGE goes from high (dorsal) to low (ventral), and is thought to be controlled by Gsx1 expression (Pei *et al*. 2011).

Furthermore, the homeobox Gsx genes have a role in the development of striatal pyramidal neurons and interneurons of the olfactory bulb (Toresson *et al*. 2000; Yun *et al*. 2003). In the early stages of neurogenesis, Gsx2 is highly expressed in the progenitor cells of the ventral LGE, whereas in later stages it is highly expressed in progenitor cells of the dorsal LGE (Waclaw *et al*. 2009). During LGE neurogenesis, Gsx2 plays a fundamental role in cell fate commitment of striatal projection neurons and olfactory bulb interneurons at distinct time points. In the early stages of telencephalic development, Gsx2 is highly expressed and specifies ventral LGE and its main derivatives, namely the striatum, and the dorsal LGE and its derivatives, such as the olfactory bulb (Waclaw *et al*. 2009).

Following the loss of Gsx2, both ventral and dorsal LGE and their derivatives are acutely reduced (Yun *et al*. 2001; Yun *et al*. 2003; Waclaw *et al*. 2004; Waclaw *et al*. 2006). When Gsx2 is mutated in the early stages of telencephalon development, the number of striatal projection neurons is reduced. Whereas, when the mutation of Gsx2 is delayed, the olfactory interneurons are defective (Waclaw *et al*. 2009). Therefore, development of the striatum depends on the early expression of Gsx2, and vice versa for the olfactory bulb.

Interestingly, knocking out Gsx2 in mice leads to misspecification of the neuronal progenitors of the LGE and its derivatives, but only in early precursor cells. However, when Gsx2 is knocked out at a later stage, it is compensated for as the expression of Gsx1 is increased in Gsx2 mutant LGE (Toresson and Campbell 2001). However, the resulting striatum is less than half of the size of the striatum in wild type LGE (Pei *et al*. 2011). The double homozygous mutants Gsx2/Gsx1 have a more acute misspecification of LGE than the Gsx2 single mutant. Overexpression of Gsx2 causes a reduction in telencephalon progenitor cell maturation (neurogenesis) both *in vivo* and *in vitro* (Pei *et al*. 2011). However, overexpression of both Gsx2 and Gsx1 has different effects on the maturation of neuronal progenitors (Pei *et al*. 2011).

1.4.3.2 DLX2 (Distal-less homeobox)

The Dlx gene family contains homeobox genes homologous to Drosophila Distal-less, which are expressed in the developing head and limbs. Dlx genes are present in the genome in three clusters with each pair sharing common enhancers (Stock *et al*. 1996; Eisenstat *et al*. 1999). Pairs Dlx1/2 and Dlx5/6 are expressed in the developing brain in the telencephalon and diencephalon (Poitras *et al*. 2007). In addition to having a definite role in ventral forebrain patterning and neuronal subtype specification, they have functions in craniofacial development (Panganiban and Rubenstein 2002).

Dlx gene expression in the telencephalon is confined to the differentiating γ-aminobutyric acid (GABA)-expressing neurons (Stühmer *et al*. 2002b). Dlx expression in the MGE is associated with GABA interneuron development, whereas Dlx expression in LGE progenitors is associated with striatal and olfactory bulb GABA neurogenesis (Poitras *et al*. 2007). The expression of Dlx1 is localised to the VZ and the SVZ of the LGE and MGE; it is also expressed in the mantle zone (MZ) (Poitras *et al*. 2007). The expression of the Dlx2 is found in two zones of the telencephalon: the VZ and SVZ of mouse ventral telencephalon of embryos at E12.5 where early differentiation arises (Panganiban and Rubenstein 2002). Dlx5 and Dlx6 expression is restricted to migrating neurons that are further differentiated, and these are located in the SVZ and MZ (Poitras *et al*. 2007).

Mice lacking Dlx1 and Dlx2 do not exhibit migration of the GABAergic interneurons from the telencephalon of the SVZ and the VZ of the LGE and MGE to the cerebral cortex. This leads to a fourfold reduction in the number of GABAergic expressing cells in the cerebral cortex, striatum and olfactory bulb, the final destinations of the GABAergic interneurons (Anderson 1997). In addition, the development of striatal SVZ and differentiation of late born striatal neurons is disrupted (Anderson *et al*. 1997; Anderson 1997; Marin *et al*. 2000). When both Dlx1 and Dlx2 are knocked out in mice, there is a reduced expression of the bigene cluster Dlx5/Dlx6. In 2004, chromatin immunoprecipitation (CHIP) studies showed that the Dlx2 protein binds the Dlx5/Dlx6 intergenic enhancer known as the I56i (Zerucha *et al*. 2000).

1.4.3.2.1 Cis-acting regulatory elements separate the two Dlx genes

Transgenic and phylogenetic footprinting analyses have shown that there are a minimum of two cis-acting regulatory elements that separate the two Dlx genes in the intergenic region. For Dlx1 and Dlx2, these cis-acting regulatory elements are I12a and I12b (Ghanem *et al*. 2003; Park *et al*. 2004; Poitras *et al*. 2007). The I12b cis-acting regulatory element was analysed to understand the genetic pathways that control Dlx gene family expression in the prosencephalon (forebrain). DNase I footprint analysis of the I12b enhancer followed by transgenic enhancer assays revealed that the Dlx proteins auto-regulate and cross-regulate the expression of DLX1/2 in the telencephalon and diencephalon. Furthermore, it was discovered that the expression of the DLX1/2 is regulated by the bHLH transcription factor ASCL1, also known as MASH1 (Poitras *et al*. 2007).

In transgenic mice the I12b enhancer directs expression of reporter genes to the forebrain (Ghanem *et al*. 2003). I12b-lacZ reporter transgene expression is detectable in the diencephalon and basal telencephalon from E10, and in the VZ, SVZ and MZ of the LGE, the MGE, the anterior entopedunuclar area (AEP) of the telencephalon and the frontonasal prominence at E11.5 in a mouse embryo. The expression of I12b-lacZ reporter transgene was detectable in cells migrating to the dorsal telencephalon, or pallium which develops into the cerebral cortex. Moreover, after birth, expression of the I12b-lacZ reporter transgene was found in the neocortex, and at P25 was detected in the olfactory bulb that contains GABAergic neurons (Poitras *et al*. 2007). Therefore, it was concluded that the enhancer in the region of the Dlx1/2 has a role in differentiation of GABAergic interneurons and projection neurons (Poitras *et al*. 2007).

1.4.3.3 ASCL1 (achaete-scute complex homologue 1 (Drosophila))

The ASCL1 gene, also known as MASH1, ASH1, HASH1 or bHLHA46, is a member of the TFs of the basic helix-loop-helix (bHLH) family. It activates transcription by binding to the E-box sequence, 5'-CANNTG-3'. For DNA binding, Mash1 is dimerized with other bHLH proteins (Ross *et al*. 2003; Poitras *et al*. 2007; Henke *et al*. 2009). Moreover, Mash1 is one of the proneural transcription factors that regulates neurogenesis in the embryonic brain (Castro *et al*. 2011). It is expressed in the ventral regions of the telencephalon, more specifically, the proliferation zones of the MGE and LGE, which determine GABAergic neural differentiation (Parras *et al*. 2004). It is highly expressed in the SVZ, VZ and MZ of the ventral telencephalon of the LGE and MGE at E12.5 (Castro *et al*. 2011).

1.4.3.3.1 MASH1 activates Notch signaling

The timing of cell fate specification and differentiation in the nervous system of vertebrates is regulated by a lateral inhibition process that is mediated by Notch signalling (Chitnis and Kintner 1996; Lewis 1996; Henrique *et al*. 1997). Mash1 indirectly influences the activation of Notch signalling by controlling the expression of the Notch ligands Delta and Jagged (Dll1, Dll3, Dll4, Jag1 and Jag2) (Lindsell *et al*. 1996; Castro *et al*. 2006; Henke *et al*. 2009). Notch-ligand binding results in cleavage of the intracellular domain of Notch (Notch-IC) and translocation of Notch-IC to the nucleus where it regulates the expression of neurogenic TFs. Notch signalling represses differentiation of neurons and inhibits proneural bHLH expression, including Mash1 (Artavanis-Tsakonas *et al*. 1995; de la Pompa *et al*. 1997; Robey 1997).

It has been shown that while Mash1 gene expression is required in the early stages of neurogenesis, Dlx2 is required in the late stages of neurogenesis to down-regulate Notch signalling during the specification and differentiation steps. Cell fate commitment is therefore regulated by the coordinated function of Mash1 and Dlx1/2, via their distinct influence of the Notch signalling pathway (Yun *et al*. 2002).

1.4.3.3.2 MASH1 directly regulates DLX1/2 expression

Several groups have demonstrated that Mash1 is an upstream regulator of Dlx2 (Porteus *et al*. 1994; Casarosa *et al*. 1999; Fode *et al*. 2000; Letinic *et al*. 2002; Yun *et al*. 2002). CHIP and electromobility shift assay (EMSA) analysis have shown that

Mash1 binds to the E-box sequence at FP5, which is a functional bHLH binding site present in the I12b enhancer. This enhancer is located upstream of the bigene cluster Dlx1/2. Binding of Mash1 to the E-box site of the I12b enhancer, activates transcription, thereby regulating the Dlx1/2 bigene directly (Poitras *et al*. 2007).

Both Dlx1/2 and Mash1 have common expression patterns in the ventral telencephalon region of the proliferation zone of the LGE and MGE (Porteus *et al*. 1994). Further evidence comes from Mash1 knockout mice, which show a reduction in Dlx gene expression in the SVZ of the MGE and the LGE at E12.5 (Horton *et al*. 1999). Moreover, when Mash1 is ectopically expressed in neocortical neurons, Dlx1/2 expression is up-regulated (Fode *et al*. 2000).

1.4.3.3.3 MASH1 regulates a large number of other target genes which promote neurogenesis

Mash1 plays a significant role in controlling neurogenesis by regulating neural progenitor processes including cell cycle progression, proliferation and differentiation. It also directly regulates the early and late phases of neurogenesis (Castro *et al*. 2011). Castro and colleagues performed a genome-wide study with CHIP on chip with an antibody against Mash1 to the microarrays promoter (chip) from the embryonic ventral telencephalon at E12.5, in order to understand the genetic programme that is activated by Mash1 in telencephalon development. Mash1 directly regulates a considerable number of genes that are associated with all the main phases of neurogenesis, including distinct biological processes, molecular functions and cellular processes. Biological processes of target genes activated by Mash1 include the early steps of inhibition processes (Notch signalling), cell fate specification, regulation of cell proliferation and neuronal differentiation (Castro *et al*. 2011). Several molecular processes are regulated by Mash1; for example, 48% of the target genes are involved in the regulation of transcription, 36% in signal transduction, 64% in nucleic acid binding, and small percentages in kinase activity (19%), enzyme activity (13%), transporter activity (14%) and cytoskeletal activity (11%) (Castro *et al*. 2011). In addition to this, Mash1 directly regulates a number of positive cell cycle regulators that promote cell cycle exits (Castro *et al*. 2011).

In Mash1 knock out mice, differentiation of the earlier stages of LGE and MGE is obstructed. Furthermore, there is evidence of a reduction in the number of cortical GABAergic neurons (Casarosa *et al*. 1999; Horton *et al*. 1999). As Mash1 regulates the expression of Dlx1/2, it can be concluded that Mash1 and Dlx1/2 direct the differentiation of GABAergic neurons (Petryniak *et al*. 2007; Long, Cobos, *et al*. 2009).

Inhibition of Mash1 results in a reduction in the number of cell divisions, and a division failure in intermediate progenitor cells in the ventral telencephalon (Castro *et al*. 2011). In the adult telencephalon, deletion of Mash1 results in acute failure of dividing neural progenitors and stem cells in the proliferation zone of the SVZ in the telencephalon (Castro *et al*. 2011).

When Mash1 is overexpressed in neural stem cells, it causes rapid differentiation of neuronal cells into operative neurons and it has an outstanding capability to control the entire sequence of phases of neurogenesis (Berninger, Guillemot, *et al*. 2007; Geoffroy *et al*. 2009; Vierbuchen *et al*. 2010). This is because, as a proneural TF, it promotes cell cycle exit and differentiates neurons into a distinct progenitor population (Bertrand *et al*. 2002; Ross *et al*. 2003). Moreover, most of the positive cell cycle regulators are up-regulated when Mash1 is overexpressed (Castro *et al*. 2011). On the other hand, loss of Mash1 results in acute failure of basal ganglia neurons in the telencephalon, as well as loss of cortical projection neurons (Casarosa *et al*. 1999; Horton *et al*. 1999; Marin *et al*. 2000; Yun *et al*. 2002; Castro *et al*. 2011).

1.5 Stages of striatal GABAergic neurons differentiation

1.5.1 Direct differentiation of hPSCs into neural lineage

In vitro mouse ESCs (mESCs) are maintained in their state of pluripotency and self-renewal by the presence of the cytokine leukaemia inhibitory factor (LIF), without LIF the mESCs start to differentiate and lose their pluripotent state. *In vitro* mESCs can be induced to differentiate into different lineages by changing the culture conditions (Bain *et al*. 1995). Differentiation of mESCs is promoted by culturing them in bacteriological non-adhesive substrate petri dishes where they proliferate, in suspension, as multicellular aggregates called embryoid bodies (EB) (Bain *et al*. 1995). Following 8-10 days of suspension culture, EBs are plated onto an adhesive substrate (Bain *et al*. 1995).

There are some differences between the maintenance of mESCs and hESCs. To sustain the pluripotent state *in vitro*, mESCs require the presence of LIF and serum replacement. hESCs are unresponsive to LIF, and instead require a feeder layer, such as a mouse embryonic fibroblast (MEF) feeder layer, to maintain their multilineage differentiation capacities (Thomson *et al*. 1998). In spite of this, hESCs and mESCs share many similarities including high telomerase activity, expression of pluripotency marker Oct3/4 and the ability to form teratomas composed of the derivatives of the three germ layers (Thomson *et al*. 1998). The growth rate of hESCs is slower than that of mESCs, and they are more susceptible to apoptosis upon dissociation. This issue can be resolved by the application of ROCK inhibitor, which functions as an apoptosis inhibitor (Y-27632). Using this inhibitor, it was observed that the rate of apoptosis during dissociation of hESCs was significantly decreased compared to untreated hESCs (Watanabe *et al*. 2007).

For neural induction of hESCs, a novel method was established that uses dual inhibitors of SMAD signaling, which is activated in the signaling pathway of BMP **(Figure 1.16)**. This is achieved through treatment with SB431542 (10 µM) and noggin

(300 ng/ml) in adherent culture or SB431542 (10 μM) and Dorsomorphin (DM) under stromal cell co-culture and EB culture (Smith *et al*. 2008; Chambers *et al*. 2009; Morizane *et al*. 2011). Dual SMAD inhibition using SB431542 and noggin, was shown to achieve an increase in expression of neuroectoderm marker Pax6 (80%) and other neural markers, such as Foxg1, epiblast marker Otx2 and Sox1, by day seven, with a decline in expression of ES cell marker Oct-4 by day five (Chambers *et al*. 2009). The recombinant protein noggin was replaced with small molecule DM at different concentrations (the most optimal concentration was 2 μM); cell survival was determined by measuring the number of colonies formed (Morizane *et al.* 2011). Cells cultured with DM had increased cell survival, whereas cells cultured with noggin failed to proliferate and form colonies. Furthermore, it is important to consider the advantages of using small molecules; they are more stable and cost effective than recombinant proteins, and also pose a lower risk of infection (Morizane *et al*. 2011). However, a study by Surmacz and colleagues showed that DM above 5 μM was toxic to cells in culture. They also showed that replacement with small molecule LDN193189 (1 μM), with SB431542 (10 μM), was more efficient at inducing the expression of neuroectoderm marker Pax6 (Surmacz *et al*. 2012). Nevertheless, the neural induction protocol often generates heterogeneous culture (multiple cell lineages) that is not exclusively differentiated into neural cells. This is a matter of huge importance in the field of transplantation as the undifferentiated cells in culture have ability to form teratomas, and therefore, could be a risk to patients. However, this risk can be reduced by using techniques such as fluorescence activated cell sorting (FACS), magnetic activated cell sorting (MACS) and DNA plasmid integration linked to a specific gene carrying antibiotic resistance, which can be used for cell selection or increased differentiation of hESCs before transplantation to reduce or eliminate the undifferentiated cells.

In neural patterning, regional commitment of anteroposterior (AP) identity and expansion specification of DV identity are dependent on the morphogenic factors WNT and SHH, respectively. It was shown that *in vitro* WNT signaling inhibits neural induction of EBs (Aubert *et al*. 2002). However, for regional commitment, WNT signaling is both essential and sufficient for determining AP patterning of the neuraxis in a dose-dependent manner (Kiecker and Niehrs 2001; Houart *et al*. 2002; ten Berge *et al*. 2008; Paek *et al*. 2012). It was shown that exogenous WNT signaling initiated development of the characteristics of the posterior structure with the differentiation of mesendoderm. On the other hand, inhibition of WNT signals was required for establishment of anterior structure with the differentiation of neuroectoderm (ten Berge *et al*. 2008).

In the patterning of DV, addition of SHH to serum-free suspension culture (SFSC) resulted in an increased expression of the ventral marker Nkx2.1 with decreased expression of the dorsal markers Pax6 and Emx1. Foxg1 expression was not affected (Watanabe *et al*. 2005). Consequently, SHH has an effect on the patterning of DV but not in AP identities in forebrain population. Meanwhile, inhibition of SHH, using SHH antagonist in chemically defined serum-free media, resulted in an increase in expression of Emx1 and Pax6 with reduced expression of the ventral telencephalon markers Dlx2, Gsx2 and Nkx2.1 (Gaspard *et al*. 2008).

1.5.2 Differentiation into striatal medium spiny neurons

Over the last 4 years, several studies have differentiated hPSCs into MSN-like cells and have used them for cell replacement therapy in the rodent brain (some of these studies are summarized in **Table 1.1**). The studies' approach was to employ developmental cues (also called morphogens), such as SHH, to control and stimulate the transcriptional networks that regulate sequential neuron progenitor fate (Carri *et al*. 2013; Nicoleau *et al*. 2013; Ma *et al*. 2012) This strategy generates a mixture of cell types, including LGE and MGE progenitor cells (reviewed in Soldner and Jaenisch 2012), and so far, no-one has succeeded in generating a pure population of striatal

MSNs. The development of a protocol for inducing disease-specific cell types *in vitro* is a pressing need in order to produce iPSC-disease-specific cell types with high efficiency to be employed in potential cell replacement therapy.

Table 1.1: Summary of published papers that differentiate hPSCs into MSN-like cells and transplantation into rodent.

Authors	Cell line	Neuron differentiation protocol	Host striatum	Number of cells transplant	Summary of outcomes
Kallur *et al.* 2006	Human striatal NSCs (6-9 weeks post-fertilization)	Expanded in neurosphere cell culture with: 20ng/ml EGF, 10ng/ml FGF, and 10ng/ml LIF.	• Neonatal rat (2-3 postnatal days) • Survival at 4 and 16 weeks. • Immune-suppression was not used.	150,000 cells	• In the early months, few cells had survived. • Some cells had migrated to the cortex, GP and corpus callosum, but the majority of cells were in the striatum. • At 4 months, the number of positive NESTIN cells had decreased, while the number of Doublecortin (DCX) and NeuN cells had increased. • Few cells were positive for astrocyte markers, while, most of the cells were parvalbumin positive. • Morphology: The grafted cells had mature neuron-like morphology, while others had more astrocyte/oligodendrocyte - like characteristics.
Song *et al.* 2007	hESCs (Miz-hES1)	Co-culture with PA6 stromal cells.	• QA-Lesioned rat • Survival at 3 weeks. • Use immune-suppression daily (Cyclosporine).	20,000 cells	• At 3 weeks, the grafted cells were found in the striatum and cortex, even though the cells had been transplanted into the striatum. • No formation of tumours occurred. • Cells in the cortex formed aggregates ($NESTIN^{+ve}$, $NeuN^{-ve}$ and $MAP2^{-ve}$). • The striatal cells were DCX^{+ve}, $GAD6^{+ve}$ and $DARPP\text{-}32^{-ve}$ and they migrated to the lesion core. • Few $GFAP^{+ve}$ (astrocyte cells) were found.
Aubry *et al.* 2008	hESCs (H9, SA-01)	200ng/ml SHH, 100ng/ml DKK1, and 20ng/ml BDNF (46-59 days).	• QA-Lesioned nude rat (A) 21-30 days of neural progenitors differentiation. (B) 46-59 days of striatal precursors differentiation. • Survival at 4-6 weeks.	50,000-200,000 cells	• At A: in the grafted cells, there was no DARPP-32 expression was seen, and tetratomas were formed. • At B: a cluster of DARPP-32 expression was found. • 13-15 weeks after transplantation, overgrowth

			• Immune-suppression was not used.		of grafted cells was observed. • Functional phenotyping was not performed. • Morphology: bi-polar, medium-sized cell bodies and expansive neurite outgrowth.
(Jeon *et al.* 2012)	HD75-h-iPSCs-derived neural progenitors.	Co-culture with PA6 stromal cells.	• QA-Lesioned rat. • Survival at 12 weeks. • Use immune-suppression daily (Cyclosporine).	100,000 cells	• Behavioural recovery started to improve after 3 weeks, as seen using the Rotarod test. • At 12 weeks, no aggregates were observed. • The grafted cells produced GABAergic neurons. • No HD transmission from the mice cells to the human grafted cells was observed, and HD pathology was not seen in human cells.
Ma *et al.* 2012	hESCs (H9, H1)	200ng/ml SHH, or 0.65μM Purmorphamine (which is in equivalence of 200ng/ml SHH).	• QA-Lesioned mice (Day 40 of differentiation protocol) • Survival at 4 months. • Immune-suppression was not used.	100,000 cells	• The grafted cells were DARPP-32^{+ve}, CTIP2^{+ve}, MEIS2^{+ve} GABA MSNs, and enkephalin or substance P^{+ve}. • Few populations were found to express vGLUT1, CHAT, 5-HT, TH and CALBINDIN. • The neurons projected to the substantia nigra. • The neurons received dopaminergic and glutamatergic inputs. • Correct motor asymmetry was observed, as seen using Trendscan.
Carri *et al.* 2013	hESCs (H9, H540)	200ng/ml SHH, and 100ng/ml DKK1.	• QA-Lesioned rat (Day 38 of differentiation protocol). • Survival at 3, 6 and 9 weeks. • Use immune-suppression daily (Cyclosporine).	500,000 cells	• At 3 weeks, no striatal markers were present. • At 6 & 9 weeks, the grafted cells were OTX2^{+ve}, FOXG1^{+ve}, MAP2^{+ve} and β-Tubulin III^{+ve}. • At 9 weeks, the cells were FOXP1^{+ve}, FOXP2^{+ve}, DARPP-32^{+ve}, but expression was not quantified. • A projection of NESTIN-positive fibers into the intact striatum was observed, illustrating an interaction between the grafts and the rat host tissue. • A decline in motor abnormality was observed.

Nicoleau *et al*. 2013	hESCs (H9)	50ng/ml SHH, and WNT antagonist (100ng/ml DKK1 or 1μM XAV-939).	• QA-Lesioned nude rat (Day 25 of differentiation protocol) • Survival at 5 months. • Immune-suppression was not used	150,000 cells	• At 5 months, the grafted cells were FOXG1^{+ve}, FOXP1^{+ve}, CTIP2^{+ve} and DARPP-32^{+ve}. • No behavioural recovery.
(Jeon *et al*. 2014)	HD75-h-iPSCs-derived neural progenitors.	Co-culture with PA6 stromal cells.	• YAC128 transgenic mice with 128 CAG repeats (12 months old). • Survival at 12 weeks • Use immune-suppression daily (Cyclosporine).	100,000 cells	• Signs of HD pathology observed at later cellular age (33 weeks). However, at early cellular age (12 weeks), no aggregate formation was evident. • Behavioural tests were performed using stepping, staircase and apomorphine-induced rotation tests. Behavioural recovery began at week 4 for the stepping test and at week 6 for the other two tests.
(Arber *et al*. 2015)	hESCs (H7)	ACTIVIN A.	• QA-Lesioned rat (Day 20 of differentiation protocol). • Survival at 4-16 weeks. • Use immune-suppression daily (Cyclosporine).	500,000 cells	• At 16 weeks, the grafted cells were GABAergic neurons that were FOXP2^{+ve}, CTIP2^{+ve}, DARPP-32^{+ve} and CALBINDIN^{+ve}. • There was not overgrowth of the grafted cells.

1.5.3 Direct differentiation into a specific differentiated cell type by ectopic expression of transcription factors

In 1988, the first direct differentiation strategy using TFs was used by Tapscott *et al*. (1988) using the protein of MyoD1 to direct reprogramed fibroblasts into myogenic cells. Recently, the forced expression of TFs was used to direct differentiate human and mouse fibroblasts into induced neural (iN) cells and neuron precursor cells (NPCs) that was summarized in **Table1.2** as Group 1 and 2 respectively. Also, direct reprogramming into different cell types such as cardiomyocytes **(Table1.2 Group 3)**, haematopoietic fate precursor cells **(Table1.2 Group 4)** and induced hepatocyte-like cells **(Table1.2 Group 5).**

Table 1.2: Summary of some published papers used the direct reprogramming strategy to differentiate somatic cells into a specific cell type.

	Cell line	Direct differentiation into cell type	Ectopic expression of TFs	References
Group 1: direct differentiation into iN	Astroglia	Functional iN cells	Ngn2 & Mash1	Berninger *et al*. 2007
	Postnatal cerebral cortical astroglia	Functional glutamatergic neurons or GABAergic neurons	Ngn2 or Dlx2	Heinrich *et al*. 2010
	mPSCs and postnatal fibroblast	Functional iN cells	Brn2, Mash1 & Myt1l	Vierbuchen *et al*. 2010
	Human fibroblasts	Functional iN cells or dopaminergic neurons	BRN2, MASH1 & MYT1L or BRN2, MASH1, MYT1L, LMX1A & FOXA2	Pfisterer *et al*. 2011
	Adult human primary dermal fibroblasts	Functional human iN cells	miRNA-124, MYT1L & BRN2	Ambasudhan *et al*. 2011
	Human and mouse fibroblasts	Functional iDA cells	Mash1, Nurr1 & Lmx1a	Caiazzo *et al*. 2011
	Human fibroblasts including neonatal	Functional neuronal cells	miRNA-9/9-124, NEUROD1, MASH1 & MYT1L	Yoo *et al*. 2011

	adult dermal cells.			
	Human and mouse fibroblasts	Functional iMNs	Brn2, Mash1, Myt1l, Lhx3, Isl1, Hb9 & Ngn2	Son *et al.* 2011
	Mouse hepatocytes	Functional iN	Brn2, Mash1 & Myt1l	Marro *et al.* 2011
	hPSCs	Functional iN cells	BRN2, MASH1 & MYT1L	Pang *et al.* 2012
	Mouse fibroblasts	Functional iNSC	Brn2, Sox2, Klf4, c-Myc & E47	Han *et al.* 2012
	Mouse fibroblasts	Functional iNSC	Sox2, Klf4 & c-Myc	Thier *et al.* 2012
	Human and mouse fibroblasts	Functional multipotent iNSCs	Sox2	Ring *et al.* 2012
	Adipocyte precursor cells	Functional iN cells	Brn2, Mash1 & Myt1l	Yang *et al.* 2013
	hPSCs	Functional iN cells	NGN2 OR NEUROD1	Zhang *et al.* 2013
	Human fibroblast	Functional iN cells with high efficiency	MASH1, BRN2 & MYT1L vector then GFP vector with 4x miRNA-124 target sequences	Lau *et al.* 2014
Group 2: direct differentiation into iNPCs	Mouse fibroblasts	Functional midbrain iDA progenitors cells	Pitx3, Nurr1, Lmx1a, Lmx1b, En1, Mash1, Myt1l, Brn2, Ngn2, Sox1 & Pax6	Kim *et al.* 2011
	Mouse fibroblasts	1- Functional iNPCs with glia and neural morphologies 2- Bi-potent iNPCs (differentiated into astrocyte and functional neurons), 3- tri-potent iNPCs (into astrocytes, neurons and oligodendrocytes) and 4- tri-potent iNPCs (into astrocytes, oligodendrocytes and less mature neurons)	1-11 TFs; Foxg1, Sox2, Brn2, Mash1, Lhx2, Dlx1, Zic1, Olig2, Pax6, ID4 & Rfx4, 2-Foxg1 & Sox2, 3-Foxg1, Sox2 & Brn2 and 4-Foxg1 & Brn2	Lujan *et al.* 2012
Group 3: direct differentiation into cardiac muscle cells	Mouse fibroblasts	Cardiomyocytes	Oct4, Sox2, C-Myc & Klf4	Efe *et al.* 2011
Group 4: direct differentiation into blood cells	Human fibroblasts	Mature haematopoietic precursors cells	Oct4	Szabo *et al.* 2010
Group 5: direct differentiation into liver cells	Mouse fibroblasts	Functional induced hepatocyte-like cells	Gata4, Hnf1α, Foxa3 & $p19^{Arf}$ inactivation	Huang *et al.* 2011

Abbreviations: iN cells: induced neuronal cells, iMNs: induced motor neurons, iDA cells: induced dopaminergic (iDA) cells, miRNA: MicroRNA, iNSC: induced neural stem cells; iNPCs: Induced neural precursor cells.

1.6 Working hypothesis and aims

Direct differentiation of mouse and human somatic cells into induced neuronal cells has been successfully achieved through the forced expression of TFs, that have a role in neurogenesis, such as MASH1, Sox2, NGN2 and DLX2 **(see details in Table 1.2 Group 1&2).** Nonetheless, this strategy has not been performed yet, to direct differentiate hPSCs into the functional MSNs through forced expression of major TFs, with a specific role in ventral telencephalon development, such as MASH1, DLX2 and GSX2 (Pauly *et al*. 2013). However, indirect ectopic expression of morphogens, such as SHH, has been evaluated; this approach has resulted in low percentage of DARPP-32^{+ve} MSNs (Kallur e*t al.* 2006; Aubry *et al*. 2008; Jeon *et al*. 2012;Carri *et al*. 2013; Nicoleau *et al*. 2013; Jeon *et al*. 2014; Arber *et al*. 2015). In fact, these studies have failed to generate a pure population of striatal MSNs. Only one study has observed in high yield of (up to 80%) DARPP-32^{+ve} MSNs (Ma *et al*. 2012). Another disadvantage of some of these studies was the lack of comprehensive function analysis of striatal neurons using electrophysiology (Kallur *et al*. 2006;Aubry *et al*. 2008;Nicoleau *et al*. 2013) Therefore, based on the above hypothesis, the aim of this study was to express the TFs ectopically, that have role in ventral telencephalon development into hPSCs to direct reprogram into specific neuron cell type, i.e. striatal GABAergic MSNs. In addition, a comprehensive approach was also employed to assess the phenotype and functionality of GABAergic MSNs in this project.

1.7 Objectives of the project

The main objectives of the project were to predict and confirm the desired TFs that have a role in the MSN development using existing microarray public database with the help of bioinformatics tool. Then, the identified TFs were used to express ectopically in the hPSCs-derived forebrain-like neural progenitors in order to directly differentiate them into MSNs-like cells. In addition, the reprogrammed cells were characterized for the ectopic expression of nucleofected TFs, and further validated for differentiation into ventral forebrain commitment towards LGE neuronal progenitors, eventually towards functional GABAergic MSNs using ventral markers and electrophysiology, *in vitro*.

Chapter 2: Bioinformatics analysis to predict novel transcription factors and regulators that have a role in differentiation and specification of medium spiny neurons

2.1 Introduction

The striatum forms a major part of the basal ganglia that is located in the subpallium domain of telencephalon (Pauly *et al*. 2013). The striatum's primary and major neurons are GABAergic medium-sized spiny neurons (MSNs), originating from the lateral ganglionic eminence (LGE) (Feyder *et al*. 2011). In Huntington's disease (HD) patients, the most affected tissue is the striatum, where the degeneration of GABAergic MSNs takes place (Vonsattel *et al*. 1985). Several mechanisms have been identified to promote neurodegeneration in HD, including mitochondrial dysfunction (Cui *et al*. 2006), impaired axonal transport (Trushina *et al*. 2004), altered synaptic transmission, altered protein-protein interactions, glutamate- and dopamine-mediated excitotoxicity (Zeron *et al*. 2002) and, most importantly, altered transcription factor (TF) expression (Thomas *et al.* 2011). In HD, dysregulation of gene expression is based on a loss of function in Huntington protein (HTT)-mediated regulation of transcription (Gardian *et al.* 2005; Bithell *et al.* 2009; Soldati *et al.* 2013). A growing number of studies from both HD patients (post-mortem) and animal models have indicated a widespread changes in gene expression, which possibly trigger a cascade of several intracellular pathways and subsequently cause loss of neuronal identity and neurodegeneration (Augood *et al.* 1997; Gardian *et al.* 2005; Bithell *et al*. 2009; Soldati *et al.* 2013). More specifically, previous investigations using the HD model in mice have shown that mutant Huntington protein (mHTT) interacts with TFs and reduces histone acetylation such that TFs cannot access specific regions of DNA to initiate transcription, producing a TF impairment and contributing to HD pathogenesis (Gardian *et al*. 2005). Transcriptional changes in HD also involve transcriptional repressor dysregulation (Landles and Bates 2004; Hodges *et al*. 2006; Bithell *et al*. 2009). By way of example, disrupted interactions between mHTT and RE1 (repressor element 1)-silencing transcription factor (REST) and neuron-restrictive silencing factor (NRSF), a TF repressor, in HD promotes migration of REST into the nucleus, resulting in aberrant repression of coding target genes and non-coding RNAs (Zuccato *et al.* 2001; Zuccato *et al.* 2003). This reduced expression of coding genes, such as the brain-derived neurotrophic factor (BDNF) genes, leads to loss of neuronal trophic support and leaves striatal neurons vulnerable to degenerative changes in HD (Altar *et al*. 1997; Zuccato *et al.* 2001; Zuccato *et al.* 2003; Zuccato *et al.* 2007). On the other hand, the increase in nuclear REST consequences in repression of miRNA, e.g. *mir*-124, and in a concomitant increase of its target genes, driving the loss of neuronal identity (Wu and Sun 2006; Johnson *et al*. 2008).

It is becoming increasingly essential to understand the molecular dysregulation of HD and its underlying pathogenic mechanism, the knowledge of which is still in its infancy. To achieve this goal, a human disease model derived from HD patients' actual cells is required instead of using animal models. With this, the first step in establishing a human disease model is to successfully differentiate HD patient-derived stem cells to MSNs using the right combination of TFs to drive neurogenesis and MSN differentiation. To gain further insight into the TF network involved in normal forebrain neurogenesis and striatal differentiation, bioinformatics tools can be used to import genomic datasets from Gene Expression Omnibus (GEO) and analyze the statistical significance of candidate genes using GeneSpring software (Genetics 2003). Importantly, GeneSpring allows microarray data to be more easily understood relative to its biological function (Genetics 2003). Such an approach has been employed previously to explore the genetic circuits of Parkinson's Disease (PD) (Hu 2011), where significant gene expression profiles were retrieved using two datasets, GSE6613 and GSE7621, from the GEO website (Genetics 2003).

The aim of this study was to validate and identify TFs and their target genes that have fundamental roles in ventral telencephalic fate commitment, regulation of neurogenesis, and MSN differentiation and maturation. The only available microarray data from the NCBI that elucidates TF networks in brain development and neurogenesis are derived from B-cell lymphoma/leukemia 11B (BCL11B, also named

Ctip2) null mutant mice and Ctip2$^{-/+}$ mutant mice (Arlotta *et al*. 2008). In the present work, the MSN microarray repository dataset analysed here were derived from the striatum of Ctip2$^{-/-}$ and Ctip2$^{-/+}$ mutant mice at postnatal day 0 (P0), (series reference GSE9330) (Arlotta *et al*. 2008), obtained from the GEO datasets of the NCBI. Then, differentially expressed TFs and their target genes associated with neurogenesis identified using GeneSpring software could be manipulated to generate a gene pool of MSNs *in vitro*, which may then be used for the disease model of HD.

2.2 Experimental Strategy

Global transcriptome comparison was conducted between two different mouse models and the wild type. GeneSpring software was used to import experiment data sets from GEO and to generate expression datasets to assess statistically significant genes from a total of eight samples. The control datasets included three independent replicates. The test samples included datasets representing four independent replicates of the Ctip2$^{-/-}$ homozygous mouse and one dataset from the Ctip2$^{-/+}$ heterozygous mouse with no replicates (Arlotta *et al*. 2008).

Analysis of the microarray data sets using GeneSpring software involved four steps: importing the microarray datasets to the software; creating the experiment; quality control of the microarray data sets; and statistical analysis of the data.

2.2.1 Importing the microarray data set to GeneSpring software

Eight Affymetrix files were downloaded from the tools menu by selecting an option “imported NCBI GEO experiment”, as the datasets were originally from the GEO database. The experimental data were generated using an Affymetrix mouse 430-2 chip.

2.2.2 Creating the experiment

The experiment was created using three steps as follows: (i) normalization of the datasets; (ii) definition of parameters; and (iii) interpretation. Firstly, the CEL scanned image files were converted into values, which represented intensity values associated with probes, and the values were grouped into probe sets. The MAS.5 algorithm was then used to transform these probe sets into expression values. The advantage of the MAS.5 algorithm is that it defines the mismatch positions and counts the number of nonspecific bindings for a given object. All samples were baseline normalised to a median. An example of mismatch (MM) and perfect match (PM) probes used in the Affymetrix platform is shown in **Figure 2.1.** The Affymetrix approach assembly probes into probe set pairs comprising a MM and a PM. The PM probe is a 25 base oligonucleotide that is complementary to a transcript, and the MM probe has the same sequence as the PM; however, at the 13th base position is hybridized to the PM probe set **(Figure 2.1)** (Rouchka *et al*. 2008). The data generated from MM probes enables recognition of cross-hybridization. Whether or not to include a probe set is based on the ratio of the intensity values of the corresponding PM and MM. For example, in **Figure 2.2,** which shows the probe set 206055-at, the intensity values for the PM probes are higher than the corresponding MM probes, and hence this probe set is included (Rouchka *et al*. 2008). However, in **Figure 2.2** for the probe set 219820-at, the MM probes have higher intensity values than the PM probes, which may be due to cross hybridization, and hence this probe set may be excluded (Rouchka *et al*. 2008).

Transcript	**A**	**T**	**C**	**G**	**A**	**T**	**C**	**G**	**A**	**T**	**C**	**G**	**A**	**---- To 25 Base**
PM	T	A	C	G	T	A	G	C	T	A	G	C	**T**	---- To 25 Base
MM	T	A	C	G	T	A	G	C	T	A	G	C	**A**	---- To 25 Base

Cross-hybridization at 13th position

Figure 2.1: An example of complementary of PM versus MM to the transcript in the Affymetrix platform.

The diagram shows an example of hybridization of PM and MM to the transcript. At the 13th position, MM is cross-hybridize with the position in PM probe.

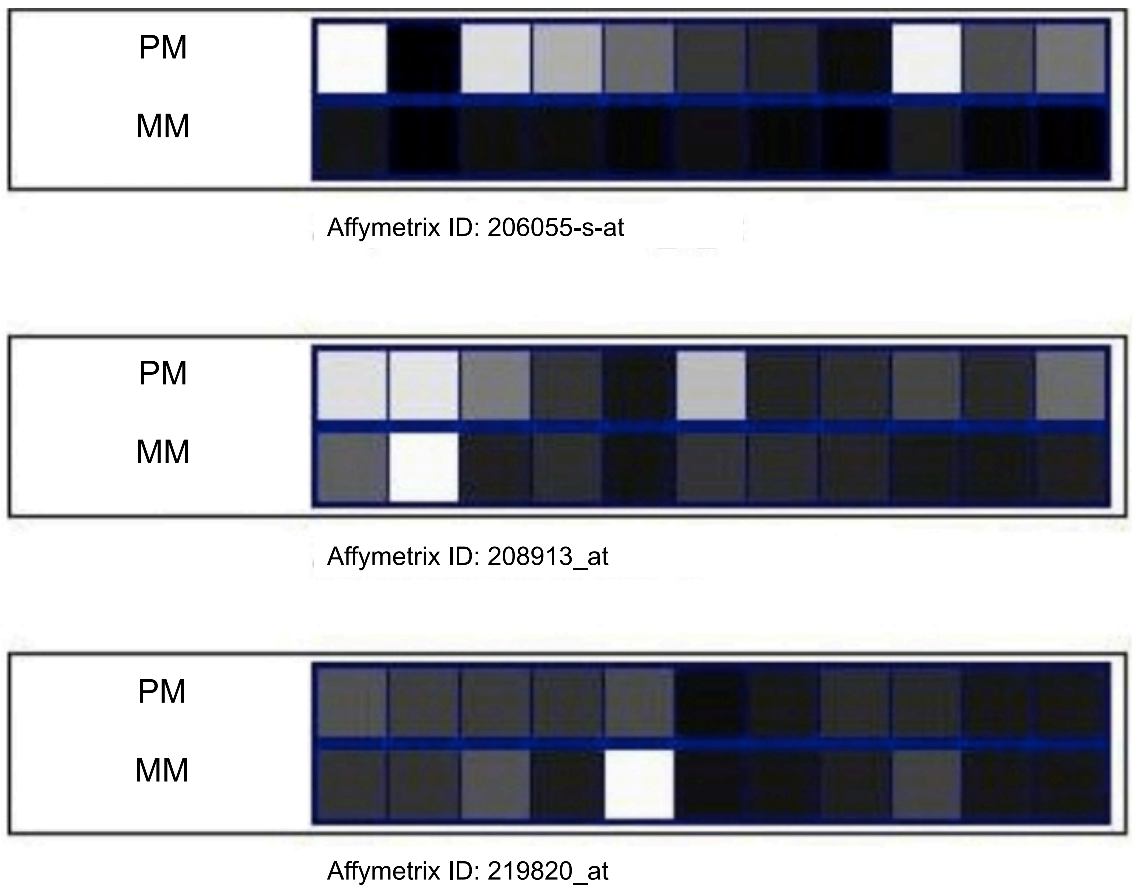


Figure 2.2: An example of intensity levels for three Affymetrix probe set PM and MM pairs.

Three different probe set pairs of PM/MM that show the intensity levels of PM and the corresponding MM that located directly below the PM probe. Each probe set represents eleven probe set pairs and a greyscale is used to depict intensity levels. Figure adapted from (Rouchka *et al*. 2008).

Abbreviations: PM: Perfect match; MM: Mismatch.

In the following step, a new parameter, such as tissue type (the control wild type mouse, Ctip2$^{-/-}$ homozygous mouse and Ctip2$^{-/+}$ heterozygous mouse), was added. Importantly, the tissue type parameter was chosen as a predictor in this experiment. Following construction of the experiment, a list of probe set IDs were generated with additional information, such as the Gene Symbol, Entrez Gene and Gene Ontology, specifically the biological process, cellular components and molecular functions.

2.2.3 Quality control and statistical analysis of the data sets

The MAS.5 algorithm was used to detect the hybridized genes and establish if they were present, marginal or absent. The data was then filtered using flags to determine if they were present, marginal or absent and below a condition of one out of eight, this being shown as 12.5%. Each gene had a signal value with a detection p-value to indicate if the transcript was detected as present, marginal or absent. In addition, where the data flagged as present marginal or absent flags, additional filtering was performed via a differential expression using statistics (unpaired t-test) with a p-value of less than or equal to 0.05, with correction for multiple comparisons using the Benjamini-Hochberg false discovery rate (FDR). An unpaired T-test was used for the three cultures, the control and the two mutant cultures with different replicates. The profiles of differentially expressed genes were then filtered again with a cut-off fold change (FC) of less than 1.3 in order to retrieve the differentially expressed genes with a large magnitude of FC. This meant that there were changes in expression between the mutant and control tissues which revealed the genes that passed the t-test with a p-value of less than 0.05 and also showed changes in expression of more than 1.3.

Clustering of the data for the hierarchal gene tree was performed with entities of clusters; these had particular conditions (Ctip2$^{-/-}$ homozygous mouse, Ctip2$^{-/+}$ heterozygous mouse and control). For this analysis, the metric distance of the differentiation and the linkage role were centroid.

Finally, analysed data and figures were exported. Further data analysis utilized bioinformatics databases, including the David Bioinformatics Resources 6.7 (National Institute of Allergies and Infectious Diseases (NIAID), NIH), the GeneCards Human Gene Database v.3 (Weizmann Institute of Science), and the UniProtKG (Protein Knowledgebase, UniProt Consortium). The probe set IDs were pasted into the David Bioinformatics database under functional annotation clustering in order to retrieve more information about the biological role of the differentially expressed significant genes.

2.3 Results

2.3.1 Differentially expressed genes identified in Ctip2$^{-/-}$heterozygous, Ctip2$^{-/+}$ heterozygous and wild-type mice

A total of 45,101 genes were analysed for the Ctip2$^{-/-}$ homozygous, Ctip2$^{-/+}$ heterozygous and control mice. In order to obtain a nearly complete knowledge about TFs association with forebrain development and specific striatal medium spiny neurons (MSNs) differentiation, all present, marginal, and absent genes detected were included in the analyses. The results showed 2,791 differentially expressed genes with a p-value less than 0.05 and a fold change more than 1.3. The top 20 up-regulated and down-regulated genes with a wide range of functions are listed in **Table 2.1**. Next, Gene ontology (GO) analyses were performed to define only dysregulated genes involved in forebrain development.

Table 2.1: Dysregulated genes between Ctip2$^{-/-}$ homozygous, Ctip2$^{-/+}$ heterozygous and wild-type.

Probe Set ID	p-value	Regulation	Fold change	Gene symbol
1449470_at	4.44E-06	up	4.25	Dlx1
1416302_at	2.07E-05	up	2.33	Ebf1
1448789_at	4.98E-05	up	6.02	Aldh1a3
1416561_at	1.32E-04	up	2.26	Gad1
1457072_at	2.62E-04	up	2.53	Ctip1
1426637_a_at	3.73E-04	up	10.24	Six3
1438194_at	5.57E-04	up	1.48	Slc1a2
1434023_at	8.78E-04	up	3.07	Cep120
1428939_s_at	0.002	up	2.04	Gnaq
1427523_at	0.003	up	2.65	Six3
1419271_at	0.004	up	4.18	Pax6
1448877_at	0.004	up	3.49	Dlx2
1416855_at	0.007	up	2.80	Gas1
1422165_at	0.007	up	2.05	Pou3f4
1427703_at	0.007	up	2.07	Pafah1b1
1425094_a_at	0.008	up	2.23	Lhx6
1438232_at	0.010	up	1.5	Foxp2
1421978_at	0.012	up	2.17	Gad2
1449863_a_at	0.020	up	1.36	Dlx5
1448893_at	0.026	up	1.43	Ncor2
1422206_at	1.58E-07	down	1.73	B3galt1
1438784_at	8.27E-06	down	-3.59	Ctip2
1436868_at	3.17E-05	down	2.07	Rtn4rl1
1431091_at	7.23E-05	down	1.62	Pygo1
1417399_at	9.26E-05	down	2.41	Gas6
1435227_at	1.51E-04	down	2.68	Ctip2
1416221_at	1.58E-04	down	2.15	Fstl1
1435649_at	1.99E-04	down	7.04	Nexn
1450339_a_at	2.17E-04	down	-3.58	Ctip2
1448978_at	4.02E-04	down	3.50	Ngef
1429485_a_at	4.30E-04	down	1.77	Utp11l
1449465_at	0.002	down	-3.9	Reln
1446633_at	0.003	down	-4.83	Atg7
1456051_at	0.007	down	-2.74	Drd1a
1421140_a_at	0.010	down	1.42	Foxp1
1437086_at	0.011	down	-2.44	Mash1
1433602_at	0.021	down	1.62	Gabra5
1427044_a_at	0.021	down	1.40	Amph
1424601_at	0.026	down	-1.36	Xrcc4
1422285_at	0.031	down	-4.19	Otp

2.3.2 Identification of dysregulated genes related to brain development and neurogenesis using GO tree

To focus on differentially expressed genes relevant to brain development and neurogenesis in Ctip2$^{-/-}$ homozygous, Ctip2$^{-/+}$ heterozygous compared to control mice, GO tree analyses were performed. As the initial results revealed in **Figure 2.3**, over 100 transcription factors, transcription repressors, target genes or effector genes associated with neuronal differentiation and development were dysregulated. Further, GO analysis to determine the individual genes involved was demonstrated and listed in **Tables 2.2 to 2.4**. Among the significant genes, both Dlx2 and Mash1 (also known as Ascl1) transcription factors are significantly dysregulated in Ctip2$^{-/-}$ homozygous and Ctip2$^{-/+}$ heterozygous mice in comparison to controls, and they have an important role in ventral forebrain fate commitment and development **(Tables 2.2 and 2.3)**. The GO analysis shows significant upregulation of Dlx2 (fold change of 3.49), but downregulation of Mash1 (fold change of -2.44). The key roles of these genes are consistent with previous findings (Yun *et al*. 2002; Petryniak *et al*. 2007; Long *et al*. 2009b; Wang *et al*. 2013). Therefore, these two transcription factors are considered a complementary combination to differentiate stem cells into MSNs in this study. Interestingly, Gsx2, which also promotes early ventral telencephalon development through induction of Mash1, Olig2, and Dlx2 expression (Szucsik *et al*. 1997; Corbin *et al*. 2000; Toresson *et al*. 2000; Wang *et al*. 2013), is not expressed any different within this dataset (data not shown). Therefore, these two transcription factors, along with Gsx2, are considered a suitable combination to differentiate stem cells into MSNs.

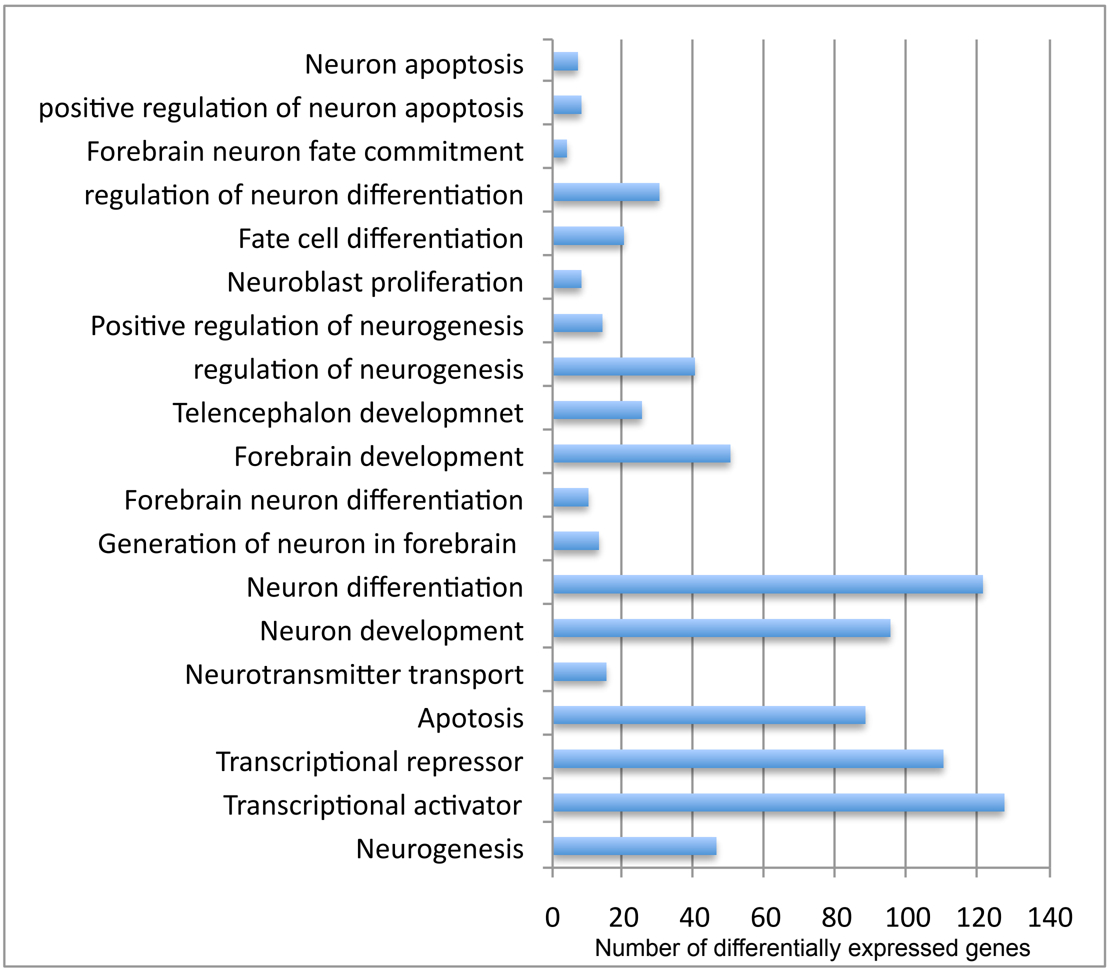


Figure 2.3: Initial GO analysis of differentially expressed genes with a significant role in forebrain development.

Table 2.2: Gene Ontology (GO) analysis for the genes identified in the development of telencephalon.

Probe Set ID	p-value	Regulation	Fold change	Gene symbol	GO terms			
1438231_at	0.014	up	1.56	Foxp2	Caudate nucleus development	Striatum development	Subpallium development	Telencephalon development
1421140_a_at	0.01	down	1.42	Foxp1				
1438232_at	0.01	up	1.5	Foxp2	Putamen development			
1421140_a_at	0.01	down	1.42	Foxp1				
1448789_at	4.98E-05	up	6.02	Aldh1a3	Nucleus accumbens development			
1456051_at	0.01	down	-2.74	Drd1a				
1437086_at	0.01	down	-2.44	Mash1	Subpallium neuron fate commitment			
1417086_at	0.02	up	1.61	Pafah1b1	Telencephalon cell migration			
1427703_at	0.01	up	2.07	Pafah1b1				
1425094_a_at	0.01	up	2.23	Lhx6				
1422262_a_at	0.04	up	1.84	Lhx6				
1435577_at	0.04	up	1.71	Dab1				
1449465_at	0.001	down	-3.9	Reln				
1451086_s_at	0.04	up	1.31	Rac1				
1456051_at	0.01	down	-2.74	Drd1a				
1416561_at	1.32E-04	up	2.255	Gad1				
1421978_at	0.01245	up	2.172	Gad2				
1448789_at	4.98E-05	up	6.02	Aldh1a3	Pallium development / Olfactory Lobe (OL) development / Corpus callosum development			
1438231_at	0.014	up	1.56	Foxp2				
1438232_at	0.01	up	1.5	Foxp2				
1456051_at	0.01	down	-2.74	Drd1a				

Table 2.3: Gene Ontology (GO) analysis for genes identified in the forebrain generation of neurons.

Probe Set ID	p-value	Regulation	Fold change	Gene symbol	GO terms		
1416967_at	0.04	up	1.29	Sox2			Forebrain neuron differentiation
1419271_at	0.01	up	4.18	Pax6			
1420995_at	0.04	up	2.34	Plxna3			
1422165_at	0.01	up	2.05	Pou3f4			
1425094_a_at	0.01	up	2.23	Lhx6			
1422262_a_at	0.04	up	1.84	Lhx6			
1422285_at	0.03	down	-4.19	Otp			
1448877_at	0.01	up	3.49	Dlx2			
1449470_at	4.44E-06	up	4.25	Dlx1			
1428938_at	0.02	up	1.79	Gnaq			
1428939_s_at	0.002	up	2.04	Gnaq			
1428940_at	0.01	up	1.91	Gnaq			
1429559_at	0.02	up	1.72	Gnaq			
1455729_at	2.00E-03	up	1.83	Gnaq			
1446633_at	0.002	down	-4.83	Atg7			
1419271_at	0.01	up	4.18	Pax6	Commitment of neuronal cell to specific neuron type in forebrain	Forebrain neuron fate commitment	
1448877_at	0.01	up	3.49	Dlx2	Commitment of multipotent stem cells to neuronal lineage in forebrain		
1449470_at	4.44E-06	up	4.25	Dlx1			
1425094_a_at	0.01	up	2.23	Lhx6			
1422262_a_at	0.04	up	1.84	Lhx6			
1420995_at	0.04	up	2.34	Plxna3	Pyramidal neuron development		Forebrain neuron development
1446633_at	0.002	down	-4.83	Atg7			
1420995_at	0.04	up	2.34	Plxna3			
1425094_a_at	0.01	up	2.23	Lhx6			
1422262_a_at	0.04	up	1.84	Lhx6			
1428938_at	0.02	up	1.79	Gnaq			
1428939_s_at	0.002	up	2.04	Gnaq			
1428940_at	0.01	up	1.91	Gnaq			
1429559_at	0.02	up	1.72	Gnaq			
1455729_at	2.00E-03	up	1.83	Gnaq			
1446633_at	0.002	down	-4.83	Atg7			
1458560_at	0.04	down	-2.23	Aspm	Forebrain neuroblast division		

Table 2.4: Gene Ontology (GO) analysis for genes identified in the development of the forebrain.

Probe Set ID	p-value	Regulation	Fold change	Gene symbol	GO terms	
1448893_at	0.03	up	1.43	Ncor2		Cell proliferation in forebrain
1434023_at	8.78E-04	up	3.07	Cep120	Interkinetic nuclear migration	
1419271_at	1.00E-02	up	4.18	Pax6	Anterior/posterior pattern formation	Forebrain regionalisation
1426637_a_at	3.73E-04	up	10.24	Six3		
1427523_at	0.003	up	2.65	Six3		
1419271_at	0.004	up	4.18	Pax6	Dorsal/ventral pattern formation	
1416967_at	0.04	up	1.29	Sox2	Forebrain morphogenesis	

2.3.3 Identification of DLX2 and MASH1 target gene interactions involved in forebrain neuron generation using pathway analysis

We subjected differentially expressed genes with potential roles in forebrain development to cellular and molecular pathway analyses in order to identify target genes associated with DXL2 and MASH1 TFs and other potential effector genes.

Fifty-three target genes were found to play critical roles in neurogenesis and anatomical structure morphologies **(Figure 2.4)**. As shown in **Figure 2.4**, Mash1 regulates Dlx1 and Dlx2 expression while Dlx1/2 regulates Arx, Dlx5, Uba2, and Spg7 expression. Furthermore, the results allude to important candidate effector genes, such as Ctip2, Ebf1, Foxp1/2, DRD1/2, and GAD1/2, normally promoting striatal development and differentiation. Moreover, these genes including Foxp1, Ebf1, Drd1/2, Gad1/2 are also dysregulated as indicated by GO tree in previous **Tables 2.2 and 2.3**. Clearly, what is shown is that these target and effector genes are potential candidates for assessing the success of the differentiation program for stem cells into MSNs after transfecting stem cells with major TFs, i.e. Mash1, Dlx2 and Gsx2.

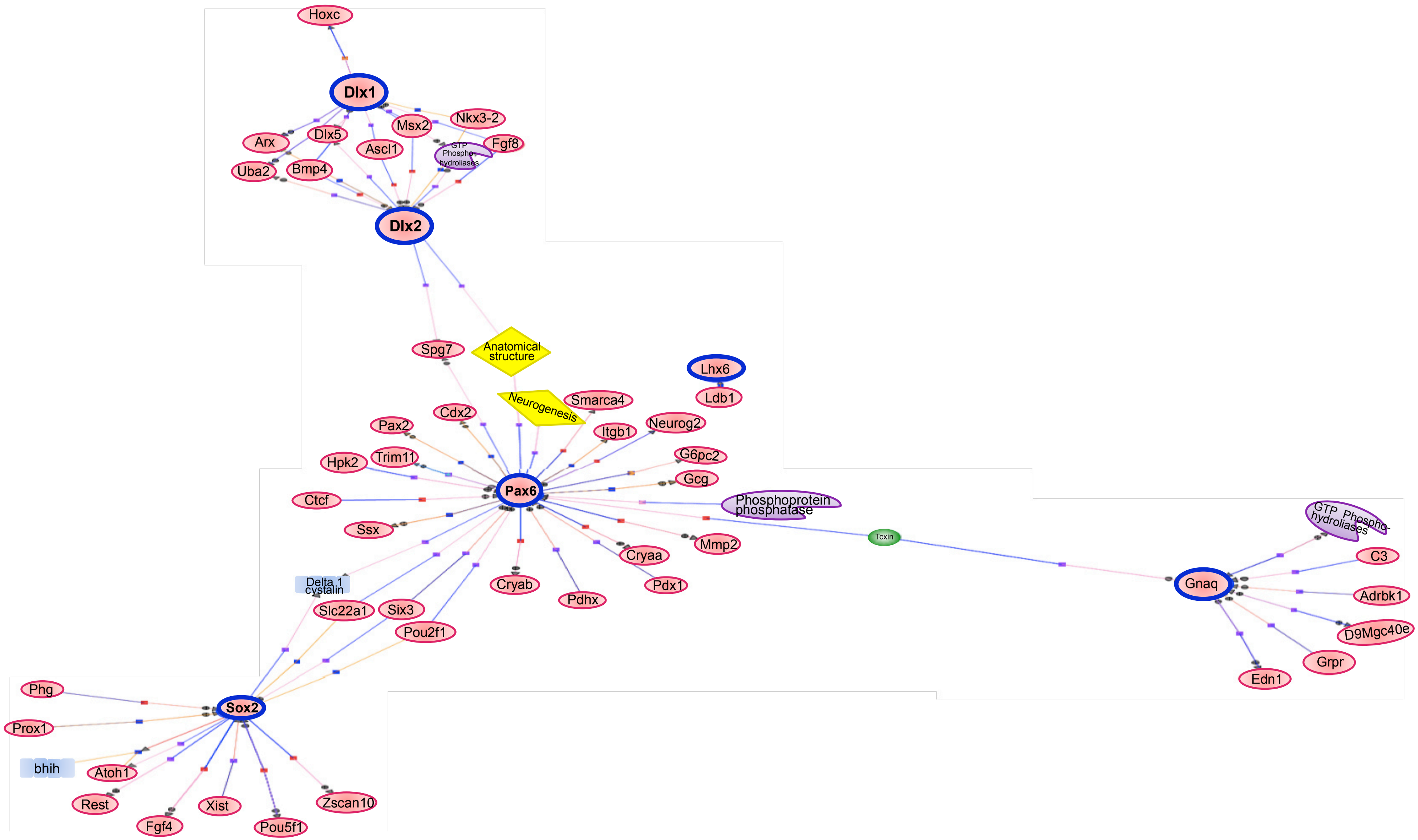

Hoxc
Dlx1
Nkx3-2
Msx2
Arx
Dlx5
Fgf8
Ascl1
GTP Phospho-hydroliases
Uba2
Bmp4
Dlx2
Spg7
Anatomical structure
Lhx6
Neurogenesis
Smarca4
Ldb1
Cdx2
Pax2
Neurog2
Itgb1
Hpk2
Trim11
G6pc2
Gcg
Pax6
Ctcf
Phosphoprotein phosphatase
Ssx
Toxin
Mmp2
Cryaa
GTP Phospho-hydroliases
C3
Delta 1 cystalin
Cryab
Pdx1
Gnaq
Pdhx
Adrbk1
Slc22a1
Six3
Pou2f1
D9Mgc40e
Grpr
Edn1
Phg
Sox2
Prox1
bhih
Atoh1
Rest
Xist
Zscan10
Fgf4
Pou5f1

Figure 2.4: The expanded interaction for the target genes involved in the forebrain neuron generation.

The advanced interaction between the target genes identified as significantly dysregulated in the forebrain neurons generation.

2.4 Discussion

While the aetiology of HD is the “mHTT toxic gain of function” and “HTT loss of function”, the comprehensive understanding of pathogenic mechanisms with respect to altered gene expression is still developing. It is of importance to have a human HD model using a patient’s own stem cells that will differentiate into MSNs to understand the molecular and biological aspects of HD pathogenesis and to develop a future for gene therapy. To this end, we identified statistically dysregulated TFs and their target genes in the only NCBI available dataset (Ctip2$^{-/-}$ mutant mouse microarray) using bioinformatics tools such as “GeneSpring”. These dysregulated TFs and their target genes play a major part in ventral telencephalon (forebrain) development, differentiation and fate commitment. Here, we have demonstrated that significantly dysregulated TFs, Dlx2 and Mash1, are appropriate and complementary to use in combination to differentiate stem cells into GABAergic neurons. In addition, their dysregulated target genes Dlx1/2 by TF Mash1 and target genes Dlx5 and Arx by TF Dlx1/2 as well as other effector genes, like Foxp1/2, Drd1/2, Ebf1, and Gad1/2, can be used as striatal or GABAergic neuron markers to validate the success of stem cell differentiation into GABAergic neurons and maturation.

From the gene expression profile using GeneSpring, Mash1 and Dlx2 were among the significantly dysregulated genes **(Table 2.1)**. Mash1 and Dlx1/2 have a fundamental role in ventral telencephalon (forebrain) development, activation of transcription factors, regulation of neurogenesis, LGE differentiation, and neuronal fate commitment in the forebrain. Furthermore, Mash1 induces neuroblast proliferation, particularly in the ventricular zone (VZ). These functions are in line with the previously reported studies conducted in Mash1$^{-/-}$ and Dlx1/2$^{-/-}$ mutant mice and Mash1$^{-/-}$;Dlx1/2$^{-/-}$ triple mutant mice models (Long, Swan, *et al*. 2009; Pauly *et al*. 2013; Yun *et al*. 2001; Yun *et al*. 2002; Yun *et al*. 2003; Toresson and Campbell 2001).

Many studies have revealed that alternative GABAergic cell fate is controlled by the coordinated functions of Mash1 and Dlx1/2. During early neurogenesis, the increased expression of Mash1 in the ventral telencephalon (Lo *et al.* 1991, Guillemot *et al*. 1993, Horton *et al*. 1999) activates Notch signaling and enhances the expression of Dll1, Dll3, and Notch signalling's target gene i.e. Hes5. Consequently, this gene profile triggers adjacent progenitor proliferation and inhibits differentiation (Horton *et al*. 1999, Yun *et al*. 2002, Poitras 2007). Expression failure of these candidate effectors was observed in Mash1$^{-/-}$ mutant mice, causing loss of proliferative progenitor in the subventricular zone (SVZ) of the medial ganglionic eminence (MGE). In addition, the ventricular zone (VZ) progentiors precociously acquired prematurely the SVZ progenitors' property of lateral ganglionic eminence (LGE), exhibited by an increased ectopic expression of GAD1 and Dlx5 in VZ (Casarosa *et al*. 1999). However, during late neurogenesis, domination by Dlx1/2 expression represses the expression of Mash1 and the Notch signaling pathway, subsequently promoting differentiation, specification and maturation of striatal neurons (GABAergic neurons), evidenced by increased expression of Drd2, Gad1/2 in SVZ and mantle zone (MZ) (Anderson *et* al. 1997a, Yun *et al*. 2002, Cobos *et* al. 2007). Therefore, corroborating with the above results, the inhibitory negative feedback of Dlx1/2 on Mash1 expression explains the current observation of Dlx1/2 over-expression and Mash1 diminished expression in Ctip2$^{-/-}$ mutant mice, when compared to wild-type mice in this chapter. Migration of GABAergic neurons to the cerebral cortex is also mediated by Dlx1/2 through the induction of Arx expression and inhibition of p21-activated serine/theronine kinase (PAK3), as demonstrated by Arx$^{-/-}$ and/or Dlx1/2$^{-/-}$ mutant mice (Anderson *et al*. 1997a, Stühmer *et al*. 2002b, Cobos *et al*. 2005a, Yoshihara *et al*. 2005, Cobos *et al*. 2007, Colombo *et al*. 2007). Thus, it has been proposed that Mash1 is necessary for the early development of the subpallium (ventral) telencephalon, while Dlx1/2 is critical for late neurogenesis (Yun *et al*. 2002; Long *et al*. 2009a; Long *et al*. 2009b).

On the other hand, Mash1 and Dlx1/2 also have a parallel yet redundant manner of directing the neurogenesis program. This is supported by the partial blockage of striatal development in Dlx1/2$^{-/-}$ mutant mice, represented by preserved expression of DRD1/2 and GAD1/2 (Long *et al*. 2009a). These mutant mice presented with a clear defect of striatal neuron differentiation in dLGE, but vLGE neuronal identity is partially maintained, suggesting the presence of a parallel pathway to Dlx1/2 function. Such a preservation of vLGE development could be maintained by the expression of Mash1, Gsh1/2 (also named Gsx1/2), and Tlx (Long *et al*. 2009a). Another line of evidence giving credence to the parallel function of Mash1 is the sustained expression of the Dlx1/2 target gene, i.e. Gad1/2 in Dlx1/2$^{-/-}$ mutant mice (Stühmer *et al.* 2002a*;* Poitras *et al*. 2007). This sustained expression is accomplished by MASH1-induced Dlx1/2 expression through binding to the enhancer, I12b (Fode *et al*. 2000). Such substantiation highlights the critical role of Mash1 and Dlx1/2 in early and late neurogenesis in the dLGE, and Mash1 alone in the septum and vLGE as well as MGE neurogenesis (Casarosa *et al*. 1999, Long *et al*. 2009a). These characteristic anatomical functions correspond with the expression pattern of Dlx1/2 and Mash1, both highly co-expressed in most dLGE VZ and SVZ progenitors, whereas there is much less expression of Dlx1/2 in vLGE and MGE (Porteus *et al*. 1994, Casarosa *et al*. 1999, Yun *et al*. 2002, Long *et al*. 2009a). These studies reveal the profound role of Dlx2 and Mash1 to induce efficient and successive progress of neurogenesis in striatum. Based on the parallel and overlapping function of Dlx1/2 and Mash1 in regulating neurogenesis of GABAergic neurons (MSNs), this combination of TFs will be used in the subsequent chapters for cloning and differentiating human pluripotent stem cell (hPSC)-derived naïve rosette neural progenitor cells (nrNPCs) into MSNs.

Gsx2 is a further important TF in promoting the early identity of the ventral domain of the forebrain, with great emphasis on VZ in dLGE and dCGE (Hsieh *et al*. 1995, Corbin *et al*. 2000, Toresson *et al*. 2000, Wang *et al*. 2013). The loss-of-function mouse model (Gsx2$^{-/-}$ mutant mice) has presented with a profound defect of LGE, particularly dLGE, and reduction of Mash1, Dlx1/2, Ebf1, and GAD1 expression, while the ectopic expression of Gsx2 retained expression of all these genes (Szucsik *et al*. 1997, Corbin *et al*. 2000, Toresson *et al*. 2000, Wang *et al*. 2013). Despite Gsx2 expression not being dysregulated in the current bioinformatics analysis, the parallel function of Gsx2 with Mash1 and Dlx1/2 TFs in programming striatal progenitor development cannot be denied (Long *et al*. 2009a). Hence, in this project, Gsx2 was combined with Mash1 and Dlx1/2 TFs to differentiate hPSC-derived nrNPCs into GABAergic neurons.

The other aim of this study was to validate previously described neuronal differentiation and maturation markers (Drd1/2, Gad2, Ebf1 and Foxp1) as well as to detect neuronal phenotypes in the SVZ and MZ, some of which are target genes for Dlx1/2. In this chapter, target genes of Dlx1/2 such as Drd1/2 and Gad2 were dysregulated in Ctip2$^{-/-}$ striatum (Stühmer *et al*. 2002a; Cobos 2005a; Yoshihara *et al*. 2005; Cobos *et* al. 2007; Colombo *et* al. 2007; Long *et al*. 2009a). The striatum is a major part of the brain that controls inputs and outputs for motor and cognitive functions (Albin *et al.* 1989; Moyer *et al.* 2007). This is largely accomplished by the dopaminergic actions of Drd1 and Drd2. It has been shown that Drd1 provides projections to control direct pathways, whereas Drd2 provides projections to control indirect pathways. The expression of Drd1/2 is higher in the striatum than the frontal cortex in mice (Araki *et al.* 2007). In fact, expression of Drd1/2 is localized to the SVZ and MZ in the striatum, along with Gad1/2 expression (Long *et al.* 2009b). The dominant expression of Drd1/2 in the SVZ and MZ of the striatum is indicative of the correspondence between its expression pattern and striatal MSNs' maturation and phenotype. In addition, Gad1/2 genes encode the glutamic acid decarboxylase 1/2 enzyme, which converts glutamate into GABA (Pinal and Tobin 1998). This allows neurons to gain the GABAergic phenotype, which is the dominant neuron type in the striatum (Feyder *et al*. 2011). Gad1 and Gad2 are localized in the neuronal cytoplasm and the nerve terminal, respectively (Pinal and Tobin 1998). A growing number of studies have been used Drd1/2 and Gad2 to closely examine LGE striatal differentiation and maturation status in SVZ and MZ in loss-of-function mouse models, such as Mash1$^{-/-}$ and Dlx1/2$^{-/-}$ (Anderson *et* al. 1997b; Casarosa *et al*. 1999; Garel *et al.* 1999; Horton *et al*. 1999; Zerucha *et al*. 2000; Stühmer *et al*. 2002a; Yun *et al*. 2002; Long *et al*. 2007; Poitras *et al*. 2007;Colasante *et al*. 2008). Taken together, Drd1/2 and Gad1/2 are major biomarkers for the differentiation and maturation stages that take place in the SVZ and MZ as described in previous mice models. The dysregulation of Drd1/2 and Gad1/2 observed in the current bioinformatics analysis and hence these biomarkers were utilized for further in in vitro analysis in the LGE striatal differentiation and maturation.

The current microarray analysis of the Ctip2$^{-/-}$ striatum has also predicted dysregulation of effector TFs with a role in MSN differentiation. Some examples of dysregulated effectors are Ebf1, Foxp1/2, Drd1/2, Gad2 and Ctip2, which are normally expressed in the SVZ and MZ of the LGE (Long, Swan, *et al*. 2009; Pauly *et al*. 2013). Ctip2 TF is uniquely expressed in striatal MSNs during early post-mitotic maturation, and controls patch-matrix compartmentalisation of MSNs (Arlotta *et al.* 2008). Lack of Ctip2 in a mutant mouse exhibited defective organization of MSNs into striatal patches (Arlotta *et al.* 20008). In addition, Ctip2 is likely to be a downstream gene of Dlx1/2, Mash1, Gsx2, and Islet1 (Anderson *et* al. 1997a; Casarosa *et al*. 1999; Yun *et al*. 2002; Stenman *et al*. 2003). Foxp1 is a preferential marker for striatal projection neurons in the matrix compartment of the striatum, cortex, and hippocampus, while Foxp2 is a marker for striosomal compartment (Tamura *et al.* 2004; Ibanez *et al*. 2012). Ebf1, which is a target gene for GSX2 (Wang *et al.* 2013), is preferentially expressed in striatonigral neurons, and also involved in regulating striatal projection neuronal development and differentiation (Garel *et al*. 1999; Garcia-Dominguez *et al*. 2003; Lobo *et al*. 2006; Ibanez *et al*. 2012). The importance of Ebf1 is documented in mutant mice presented with defective projections of neurons to the substantia nigra (Lobo *et al*. 2006). These various forms of evidence support the use of the target markers (Ebf1, Foxp1/2, and Ctip2) in this study not only to assess phenotype of neurons, but also to examine the progress of the direct differentiation of hPSC-nrNPCs through the expression of critical TFs and effector genes.

Overall, the TFs Dlx2, Mash1, and Gsx2 have a dramatic role in neurogenesis in striatum as evident by growing number of literature. Hence, Dlx2, Mash1, and Gsx2 were selected for ectopic expression in the hPSC-nrNPCs to test the hypothesis that ectopic expression of TFs involved in MSN specification and differentiation could trigger differentiation of hPSC-nrNPCs into mature striatal GABAergic MSNs. In addition, candidate target and effector genes identified here as interacting genes involved in forebrain neuron generation were used to validate the success of combining TFs to re-program stem cell differentiation and maturation into GABAergic MSNs and to assess the phenotype of generated neurons. In addition, these target and effector biomarkers (Foxp1/2, Ebf1, Gad1/2, Drd1/2 and Ctip2) were used to determine the phenotype of neuronal differentiation.

Chapter 3: Materials and methods.

3.1 PCR gene amplification and cloning

3.1.1 GoTaq Flexi DNA Polymerase PCR Amplification

DNA was amplified using different commercial polymerase chain reaction kits, including GoTaq Flexi DNA Polymerase (PCR) (Cat No. M8295, Promega, Southampton, UK) and Platinum *Taq* DNA high-fidelity polymerase (Cat No. 11304-029, Invitrogen, Paisley, Scotland, UK).

PCR amplification conditions were optimized by varying the $MgCl_2$ concentration and the annealing temperature. The set-up used in this project was as follows:

Reagent	$MgCl_2$ concentration				Negative (-ve) control
	1.5 mM	2 mM	2.5 mM	3 mM	
5× GoTaq green buffer	5 µl	5 µl	5 µl	5 µl	5 µl
25 mM $MgCl_2$	---------	0.5 µl	1 µl	1.5 µl	0.5-1.5 µl
10 mM Nucleotide mix (dNTP)	0.5 µl	0.5 µl	0.5 µl	0.5 µl	0.5 µl
Primers*	1 µl	1 µl	1 µl	1 µl	1 µl
GoTaq DNA polymerase	0.13 µl	0.13 µl	0.13 µl	0.13 µl	0.13 µl
Template DNA	< 250 ng/25 µl				--------------
dH_2O (distilled water)	Up to 25 µl				
Total	**25 µl**	**25 µl**	**25 µl**	**25 µl**	**25 µl**

* The primers (forward and reverse) were diluted with dH_2O from 100 pmol/µl to 10 pmol/µl (20 µl of the forward primer + 20 µl of the reverse primer in 160 µl of dH_2O = a total of 200 µl).

The thermal cycling conditions for GoTaq DNA polymerase are as follows:

Steps	Temperature	Time	Number of cycles
Initial Denaturation	95°C	5 min	1 cycle
Denaturation	95°C	1 min	35 cycles
Annealing	62–65°C*	1 min	
Extension	72°C	1 min/kb	
Final Extension	72°C	5 min	1 cycle

* Each gene required for cloning has a different annealing temperature. DLX2, MASH1 and GSX2 have annealing temperatures of 63°C, 65°C and 62°C, respectively.

The primers of each specific gene were designed to amplify and clone the four selected transcription factors. These are shown below:

Primer name	Sequence (5'-3')	Melting temperature (Tm) (°C)
***Bgl*II-GSX2-F**	AACAGATCTATGTCGCGCTCC TTCTATGTCGA (32 bp)	68.2
***Bgl*II-GSX2-R**	AACAGATCTTAAGGGGGAAAT CTCCTTGTCATCG (34 bp)	68.3
***Nhe*I-DLX2-F**	AACGCTAGCATGACTGGAGTC TTTGACAGTC (31 bp)	68.2
***Nhe*I-DLX2-R**	AACGCTAGCGAAAATCGTCCC CGCGCTCAC (30 bp)	72.2
***Xba*I-FOXGI-F**	AACTCTAGAATGCTGGACATG GGAGATAGGAAAG (34 bp)	68.3
***Xba*I-FOXGI-R**	AACTCTAGAATGTATTAAAGG GTTGGAAGAAGACCC (36 bp)	67.2
***Bam*HI-MASH1-F**	AACGGATCCATGGAAAGCTCT GCCAAGATGG (31 bp)	69.5
***Bam*HI-MASH1-R**	CTGGATCCGAACCAGTTGGTG AAGGCGA (28 bp)	68.0

3.1.2 Platinum *Taq* DNA Polymerase High Fidelity PCR (HF PCR)

The Platinum *Taq* DNA high-fidelity polymerase used in this project was supplied by Invitrogen. This enzyme is a mixture of a recombinant enzyme (*Taq* DNA polymerase), an anti-*Taq* polymerase antibody and *Pyrococcus* species GB-D polymerase, which has DNA proofreading activity. This mixture helps to increase fidelity by about a factor of six, compared with *Taq* DNA polymerase alone. Platinum *Taq* DNA high-fidelity polymerase contains a 10× high-fidelity PCR buffer, magnesium sulphate ($MgSO_4$), as well as high-fidelity Platinum *Taq* DNA polymerase.

The following reagents were added to DNase/RNase-free microcentrifuge PCR tubes:

Reagent	Volumes	Negative (-ve) control	Final concentration
10× high-fidelity PCR buffer	2.5 µl	2.5 µl	1×
50 mM $MgSO_4$	1 µl	1 µl	2 mM
10 mM nucleotide mix (dNTP)	0.5 µl	0.5 µl	0.2 mM
Primer	1 µl	1 µl	0.4 mM
Platinum *Taq* DNA high-fidelity polymerase	0.1 µl	0.1 µl	1 unit
Template DNA	1 µl	-------------------	< 250 ng
dH_2O (distilled water)	Up to 25 µl		-------------------
Total	**25 µl**	**25 µl**	**-------------------**

The thermal cycle condition for Platinum Tag DNA polymerase is as follows:

Step	Temperature	Time	Number of cycles
Initial Denaturation	95°C	5 min	1 cycle
Denaturation	95°C	30 s	30 – 32 cycles*
Annealing	62–65°C*	30 s	
Extension	68°C	1 min/kb	

*The three genes, MASH1, DLX2 and GSX2, required different annealing temperatures as well as different numbers of cycles, these being 65°C – 32 cycles, 63°C – 32 cycles and 62°C – 30 cycles, respectively.

3.1.3 Agarose Gel Electrophoresis

Using electrophoresis, DNA fragments can easily be separated on the basis of size. In an electric field, DNA migration is relative to its mass. Here, an agarose medium was used to retain the DNA sample and a loading dye (10×) was added to the DNA sample prior to loading. The reagents for DNA gel electrophoresis were as follows:

0.8–2% agarose (in Tris-acetate-EDTA (TAE) buffer)
10 mg/ml ethidium bromide
0.1 M of guanosine (only for digestion and ligation steps)

Ethidium bromide-stained DNA bands were visualized via UV tranilluminator at a wavelength of 254 nm. The required amount of guanosine solution (Cat No. G6752, Sigma-Aldrich, Gillingham, Dorset, UK) was added to the TAE buffer to protect the digested DNA from UV damage prior to gel extraction purification (Grundemann and

Schomig 1996). Gels were electrophoresed in 1 × TAE buffer at 80–100 V for 1 h and 30 min. DNA fragments were sized using a 1-kb or 100-bp ladder.

3.1.4 DNA Gel Extraction

Following gel electrophoresis, DNA fragments were extracted either using the GeneClean II Kit (Cat No. 1001-400, MP Biomedicals, Cambridge, UK), or the Qiaquick Gel Extraction Kit (Cat. No. 28704, Qiagen, West Sussex, UK) following the manufacturer's instructions. DNA fragments separated in agarose gels were excised with a razor blade and visualized under UV light (365 nm).

For the GeneClean kit, gel slices were dissolved in three times the volume of sliced gel of 6 M sodium Iodide (NaI) by incubation for 5 min at 55°C with periodic agitation. DNA was then bound to GlassMilk, an aqueous suspension of silica, using 1 µl of GlassMilk per 1–2 µg of DNA. The DNA/GlassMilk suspension was incubated at room temperature (RT) for 5 min. DNA bound silica was pelleted by centrifugation at 14,000×*g* for 5 s and the supernatant was removed. The pellet was washed three times with 500 µl of prepared New Wash solution (solution of NaCl, Tris, EDTA, ethanol and H_2O), and after the final centrifugation, the pellet was air dried to remove any residual ethanol. DNA was eluted from the silica by re-suspension in 10 mM Tris (pH 7.6) and 1 mM EDTA (pH 8.0) (TE) buffer. Eluted DNA was separated from the silica matrix by centrifugation at 14,000×*g* for 30 s.

For Qiaquick gel extractions, gel slices were dissolved in three times the gel volumes of QG buffer and 10 µl of 3M sodium acetate by incubation for approximately 10 min at 50°C. DNA was then bound to a Qiaquick column by adding QG buffer, followed by centrifugation at 13,000×*g* for 1 min. The flow-through was removed, and 500 µl of QG buffer was added to the Qiaquick column and centrifuged for 1 min to eliminate any residual gel. The Qiaquick column was washed twice with 750 µl of PE buffer and allowed to stand for 2–5 minutes at RT to eliminate residual ethanol, followed by centrifugation at 13,000×*g* for 1 min. The Qiaquick column was then placed into a new 1.5 ml tube. DNA was eluted from the membrane of the Qiaquick column by incubating for 1 min with 30 µl of EB buffer at RT. Eluted DNA was separated from the membrane of the Qiaquick column by centrifugation at 13,000×*g* for 1 min.

3.1.5 DNA Cloning

3.1.5.1 Gene amplification by PCR and insertion into the TOPO vector pENTR5' TOPO

PCR products for each gene were cloned into the TOPO vector pENTR5' TOPO (Cat No. K591-10, K591-20 and K5910-00, Invitrogen) using the pENTR5' TOPO TA cloning kit. This kit contained the linearized pENTR5' TOPO vector, which had 3' thymidine (T) overhangs, and topoisomerase I. Because the PCR products contained polyA overhangs and the linearized vector had 3' T residues, this enabled the efficient ligation of the PCR products with the linearized vector. At the same time, the topoisomerase I bound at specific sites to the TOPO vector (CCCTT) and cleaved the phosphodiester bond between two nucleotides in one strand where the vector was subsequently linearized. Here, the covalent bond between the 3' phosphate and the tyrosyl residue of the enzyme was formed because of the energy released when the phosphodiester bond was broken. The bond between DNA and enzyme (phospho-tyrosyl) was identified by the presence of a 5' hydroxyl from the original, cleaved strand. At this point, the covalent bond between phospho-tyrosyl was broken and the topoisomerase 1 enzyme was released (Invitrogen, 2007).

3.1.5.2 One-Shot TOP10 Chemically Competent *Escherichia coli*

A PCR product (DNA of interest) was inserted into an entry vector called pENTR5'-TOPO, which was supplied in the pENTR5'-TOPO TA cloning kit. Here, the recombinant vector can be transformed into chemically competent cells of *E. coli*. The following paragraphs describe the procedure for the TOPO cloning reaction and the transformation used in this project.

TOPO cloning reaction

Reagent	Volume
DNA (PCR product)	2 µl
Salt solution	0.5 µl
pENTR5' - TOPO	0.5 µl
Total	**3 µl**

The reaction was mixed gently and incubated at RT for 5 min. Afterwards, it was placed on ice while preparing for the next step, namely, the transformation of One-shot competent *E. coli*.

Transformation of One-shot TOP10 chemically competent *E. coli*

One vial (50 µl) of competent cells for each transformation was placed on thawing ice. Then, 1–5 µl of the TOPO cloning reaction was gently added to the vial of competent cells and mixed. Then, the vial was incubated on ice for 30 min. A heat shock was then applied for 30 s in a 42°C water bath, after which the cells were returned to the ice for 2 min. Next, 250 µl of pre-warmed S.O.C. medium (20 mM glucose, 10 mM NaCl, $MgCl_2$ and $MgSO_4$, 2.5 mM KCl, 2% tryptone and 0.5% yeast extract) (Cat. No. 15544-034, Invitrogen) was added to the vial, which was incubated in a shaking (225 rpm) incubator at 37°C for 1 h. Each of the transformation mixtures (20μl) was spread on a pre-warmed LB agar plate containing an appropriate antibiotic. Finally, the plate was incubated overnight at 37°C in an inverted position.

Subsequently, colonies from the selective plate were selected and analysed to determine whether they were positive transformants. This was done by isolating the plasmid DNA and analysing it via restriction digest analysis; this was then followed by DNA sequencing. The correct sequences of the DNAs of interest were confirmed prior to subcloning.

3.1.5.3 Transformation of DH5α Competent Cells

DH5α competent cells (Cat. No. 18265-017, Invitrogen) were put on thawing ice, gently mixed with the tip of a pipette, and a 50 μl aliquot was used for each transformation (unused competent cells were placed in an ethanol-dry ice bath for 5 min and stored at -80°C). Next, 1–5 μl of DNA was added. The cells were incubated on ice for 30 min and then heat shocked at 42°C for 20 s. Cells were returned to the ice for 2 min and 950 μl of pre-warmed S.O.C medium was added. Cells were then incubated in a shaking (225 rpm) incubator at 37°C for 1 h. Following the incubation, 20 μl of cells were spread on a pre-warmed LB agar plate containing an appropriate antibiotic. Lastly, the plates were placed in an inverted position in an incubator at 37°C and left overnight.

3.1.6 DNA sequencing

After cloning the PCR-amplified genes into the pENTR5’ TOPO vector, plasmid DNA was purified, eluted with RNase free water, and sent to Eurofins MWG operon for DNA sequencing. Sanger sequencing was performed using M13 forward and reverse primers. DNA sequences were analysed using BLAST on the NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

3.1.7 Plasmid Extraction & Purification

3.1.7.1 Plasmid Miniprep

Plasmid DNA was prepared by the alkaline lysis method (Sambrook *et al*. 1989). Bacterial colonies were picked and grown in 2 ml of LB medium containing the appropriate antibiotic (50 mg/ml kanamycin or 100 mg/ml ampicillin) by incubation in a shaker at 37°C overnight. Each bacterial culture 500 µl was transferred into an Eppendorf tube with 100% sterile glycerol and stored at -80°C to preserve the culture for future experiments.

Bacterial cultures were transferred to sterile 1.5 ml Eppendorf tubes and cells were pelleted by centrifugation at 12,500×*g* for 5 min. Bacterial pellets were re-suspended in 100 µl of ice-cold Solution I, which contained 50 mM glucose (Cat No. 1011747-500 G, BDH Laboratory Suppliers, Poole, UK), 10 mM EDTA (Cat. No. E5139-500 G, Sigma-Aldrich) and 25 mM Tris (pH 8.0) (Cat. No. 10708968001-500 G, Roche, Applied Science, West Sussex, UK), and then lysed by the addition of 200 µl of freshly made Solution II, which contained 0.2 M NaOH (sodium hydroxide) (Cat. No. S5881-1Kg, Sigma-Aldrich) and 1% SDS (Cat. No. L3771-500 G, Sigma-Aldrich), followed by incubation for 5 min at RT. Ice-cold neutralization Solution III (150 µl), which contained potassium acetate (Cat. No. P1190-500 G, Sigma-Aldrich) and glacial acetic acid (Cat. No. 64-19-7, Thermo Fisher Scientific, Loughborough, Leicestershine, UK), was added and mixed by inversion and incubated on ice for 3 to 5 min until a white, downy precipitate was seen.

Plasmids were isolated by centrifuging the lysates at 12,500×*g* for 5 min. At this point, a clear solution was observed (supernatant), which was then transferred into a clean 1.5-ml microcentrifuge tube. An equal volume of phenol-chloroform was added, mixed well and centrifuged at 12,000×*g* for 2 min. The top aqueous layer containing the DNA plasmid was transferred into a new microcentrifuge tube. Then, two volumes of 100% ethanol were added, mixed thoroughly and incubated at RT for 2 min, followed by centrifugation at 12,000×*g* for 5 min. The pellet was washed with 70% ethanol (200 µl) and centrifuged at 12,000×*g* for 5 min. The pellet was incubated at RT for 10 min to dry the ethanol. Then, the DNA pellet was re-suspended in 49 µl of TE buffer containing RNase (20 mg/µl) to eliminate any RNA molecules.

3.1.7.2 Endotoxin Free Maxiprep

The Endofree plasmid maxi kit (Cat. No. 12362, Qiagen) was used for large scale isolation of endotoxin-free plasmid DNA according to the manufacturer's instructions. A colony was either picked from a master plate or a stab from a glycerol stock and grown in a 3 ml starter culture of LB containing an appropriate antibiotic, and incubated overnight on a shaker (220 rpm) at 37°C. The starter culture was diluted with 100 or 250 ml of LB for high- or low-copy plasmid, respectively, containing an appropriate selective antibiotic. It was then incubated overnight under the same conditions for 12–16 h. Next, the resulting bacterial suspension was harvested via centrifugation at 4°C, 6,000×*g* for 15 min. Then, plasmids were isolated using the Endofree plasmid maxi kit according to the manufacturer's instructions.

3.1.8 Glycerol Stocks

Stocks of competent cells containing the plasmid of interest were made by adding glycerol to 500 μl of a bacterial suspension from an overnight culture at a ratio of 1:1, mixed well and stored at -80°C.

3.1.9 Analysis by Restriction Digestion

For the plasmid miniprep or maxiprep purifications, restriction digestion analysis was carried out as follows:

Solution	Volumes
DNA	5 μl
Enzyme	0.5 μl
Buffer (×10)	2.5 μl
BSA (×100)*	0.25 μl
dH_2O	up to 25 μl

*Some enzymes require bovine serum albumin (BSA).

The normal protocol for restriction digestion is as follows:

Solution	Concentration // Volume
DNA	0.5–1 μg
Enzyme	5 units
Buffer (×10)	1×
BSA (×100)	1×
dH_2O	up to 25 μl

Restriction digestion analysis was used to check the ligations and for subcloning purposes.

Enzymes were purchased from New England Biolabs (NEB, Hitchin, UK). The incubation temperature and the time depended on the DNA concentration, as well as the enzyme. The completion of the digestion was then checked by gel electrophoresis.

If a ligation reaction was being conducted, the vector was treated with 0.5 units of calf intestinal alkaline phosphatase (CIP) (Cat. No. M0290S, NEB) and then incubated at 37°C for 20 min, followed by purification and ligation.

The enzymes used in this project are listed below:

*Ahd*I	*Ase*I	*Bam*HI-HF
*Bgl*II	*Bsp*MI	*Bst*XI
*Eco*RI-HF	*Kpn*I-HF	*Nde*I
*Nhe*I-HF	*Not*I-HF	*Pst*I
*Sac*II	*Sal*I-HF	*Xba*I
*Xmn*I		

3.1.10 Ligation

A T4 DNA ligation kit was used (Cat. No. M0202, NEB). T4 DNA ligase forms a phosphodiester bond between two nucleotides: between the 5’ phosphate and the 3’ hydroxyl group. This ligase can be used to ligate both blunt and cohesive ends. The incubation time and depends on the type of DNA ends. For cohesive ends, the incubation is typically at RT for 10 min (2 h for blunt ends). However, in this project, the ligation mixture was incubated in a thermocycler for different lengths of time and at different temperatures.

T4 DNA ligase buffer (pH 7.5) contains 50 mM Tris - HCl, 10 mM $MgCl_2$, 1 mM ATP and 10 mM dithiothreitol (DTT). The ratio of vector-insert (v:i) used in this project varied in each experiment, starting from a 1:3 ratio. Here, the amount of vector was 100 ng, and the amount of insert was calculated from the following equation:

$$100 \times \frac{\textit{the size of insert}}{\textit{the size of vector}} \times \textit{the ratio of } \frac{i}{v} = \textit{amount of insert}\,(ng)$$

The following table shows the setup for the ligation experiment:

Reagent	Volume	Vector only
10x T4 DNA ligase buffer	2 µl	2 µl
T4 DNA ligase	1 µl	1 µl
Insert	*	------------
Vector	*	*
dH_2O	Up to 20 µl	
Total	**20 µl**	**20 µl**

* The calculated amount of insert and vector depends on a ratio of v:i

In each ligation experiment, a vector-only control was used. After incubation, 1–5 µl of the ligation sample was used for bacterial transformations.

3.1.11 Construct the expression vectors

The open-reading frame (ORF) of each TF was cloned by PCR using specific primers that were designed with appropriate sequence and a specific restriction enzyme recognition sequence that is required for cloning into the expression vector **(Table 3.1)**. In addition, the generation of the TF expression vectors required the subcloning of the ORF for each TF into the interim vector p3X-2A, without initiation and stop codons, as both of these sequences were already present in the p3X-2A vector between the restriction site *Sal*I, where the cloned TFs would be inserted **(Figure 3.1)**. The initiation codon (ATG) was located before the *Bam*HI restriction site that was used for the first fragment insertion, and the stop codon was located after the fourth fragment insertion in the *Nhe*I restriction site **(Figure 3.1)**.

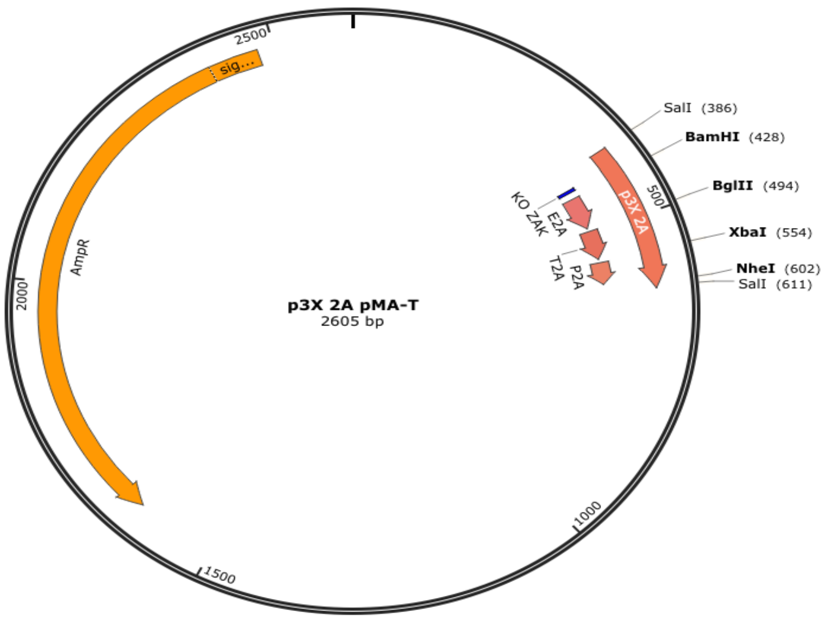
Table 3.1: The PCR primers used for PCR cloning to subclone the three desired TFs

Restriction enzyme name-Primer name	Forward Sequence (5'-3')	Reverse Sequence (5'-3')	Product size (bp)
***Bam*HI**-MASH1	AAC **GGA TCC** ATG GAA AGC TCT GCC AAG ATG G	CT**G GAT CC**G AAC CAG TTG GTG AAG GCG A	725
***Bgl*II**-GSX2	AAC **AGA TCT** ATG TCG CGC TCC TTC TAT GTC GA	AAC **AGA TCT** TAA GGG GGA AAT CTC CTT GTC ATC G	930
***Nhe*I**-DLX2	AAC **GCT AGC** ATG ACT GGA GTC TTT GAC AGT C	AAC **GCT AGC** GAA AAT CGT CCC CGC GCT CAC	1,002

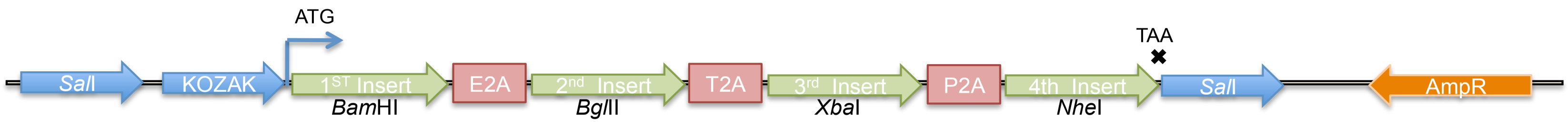
Abbreviations: *Bam*HI-MASH1: primer for Achaete-scute complex homologue 1 with restriction enzyme site *Bam*HI, *Bgl*II–GSX2: primer for GS homeobox 2 with restriction enzyme site *Bgl*II, *Nhe*I–DLX2: primer for Distal-less homeobox with restriction enzyme site *Nhe*I; bp: base pair.

The p3X-2A vector was designed to insert cloned genes using restriction enzyme sites located between three 2A peptide linked sequences **(Figure 3.1)**. In addition, a Kozak sequence (GCC GCC) is present upstream of the start codon (ATG) to enable efficient translation initiation (Kozak 1987) **(Figure 3.1)**. Genes inserted into the p3X-2A vector could then be released as a *Sal*I restriction fragment and inserted into the unique *Sal*I site in pCAGG-IRES-EGFP **(Figure 3.2)**.

A)



B)



p3X 2A pMA-T
2605 bp

C)

```
                                                                                                     SalI
301  ACGGCCAGTG AGCGCGACGT AATACGACTC ACTATAGGGC GAATTGGCGG AAGGCCGTCA AGGCCACGTG TCTTGTCCAG AGCTCGTCGA CGAATTCAGC
     TGCCGGTCAC TCGCGCTGCA TTATGCTGAG TGATATCCCG CTTAACCGCC TTCCGGCAGT TCCGGTGCAC AGAACAGGTC TCGAGCAGCT GCTTAAGTCG
                     *          **   BamHI                     E2A                                          BglII
401  GCTCTCGAGA CCGGTGCCGC CATGGGAGGA TCCCAGTGTA CTAATTATGC TCTCTTGAAA TTGGCTGGAG ATGTTGAGAG CAACCCAGGT CCCAGATCTG
     CGAGAGCTCT GGCCACGGCG GTACCCTCCT AGGGTCACAT GATTAATACG AGAGAACTTT AACCGACCTC TACAACTCTC GTTGGGTCCA GGGTCTAGAC
                      T2A                                        XbaI                    P2A
501  AGGGCAGAGG AAGTCTTCTA ACATGCGGTG ACGTGGAGGA GAATCCCGGC CCTTCTAGAG CCACGAAGCA AGCAGGAGAT GTTGAAGAAA ACCCCGGTCC
     TCCCGTCTCC TTCAGAAGAT TGTACGCCAC TGCACCTCCT CTTAGGGCCG GGAAGATCTC GGTGCTTCGT TCGTCCTCTA CAACTTCTTT TGGGGCCAGG
        NheI  **  SalI
601  TGCTAGCTAA GTCGACGGTA CCTGGAGCAC AAGACTGGCC TCATGGGCCT TCCGCTCACT GCCCGCTTTC CAGTCGGGAA ACCTGTCGTG CCAGCTGCAT
     ACGATCGATT CAGCTGCCAT GGACCTCGTG TTCTGACCGG AGTACCCGGA AGGCGAGTGA CGGGCGAAAG GTCAGCCCTT TGGACAGCAC GGTCGACGTA
```

p3X 2A pMA-T
DNA sequencing

Figure 3.1: Map of p3X-2A pMA-T vector for subcloning

The circular map of the vector (A), the linear map of the vector (B), and the DNA sequencing between the *Sal*I restriction sites (C). The restriction sites *Bam*HI I and *Nhe*I were required for ligation of the TFs, MASH1 and DLX2, respectively.

*Kozak sequences.

** Start codon (ATG) and stop codon (TAA).

The pCAGG-IRES-EGFP (pCAGG) vector was constructed from the vector “pIRES2EGFP” (from Clontech) that contains the cytomegalovirus (CMV) promoter, a multiple cloning site (MCS), internal ribosome entry site (IRES), coding sequences for enhanced green fluorescent protein (EGFP) and a gene conferring kanamycin resistance **(Figure 3.2)**. Because the CMV promoter is known to be silenced in hESCs, this promoter was replaced by the artificial CAGG promoter, which has been shown to drive high levels of transgene expression in hESCs (Alexopoulou *et al*. 2008) **(Figure 3.2).**

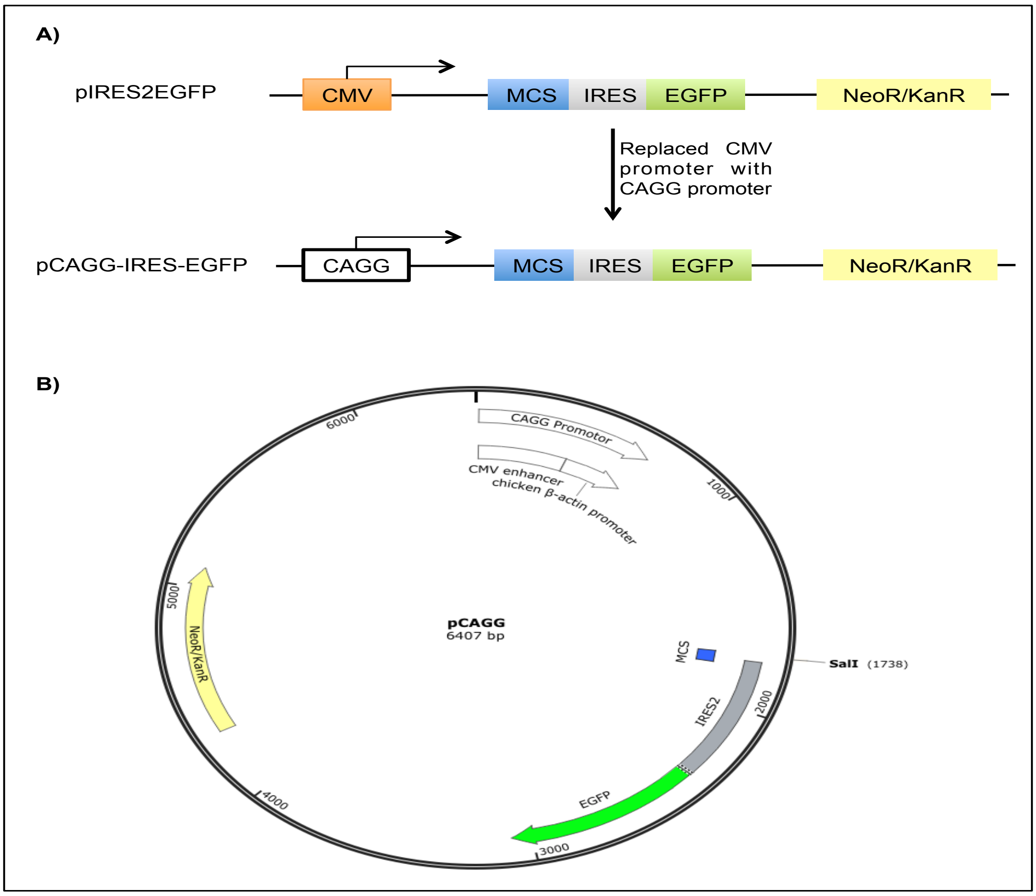


Figure 3.2: Circular and linear map of optimized pIRES2EGFP to pCAGG–IRES–EGFP.

The linear map of the vector pCAGG in which the CMV promoter was replaced with CAG promoter (A). The backbone of the expression vector was CAGG-MCS-IRES–EGFP with kanamycin resistance gene. The circular map of pCAGG (6.4 kb) (B).

Abbreviations: CMV: Cytomegalovirus (CMV) promoter, CAGG: CMV early enhancer/chicken beta actin (CAG) promoter, MCS: Multiple cloning site, IRES: Internal ribosome entry site, EGFP: Enhanced green fluorescent protein; NeoR/KanR: Neomycin resistance/Kanamycin resistance.

The specific order of cloning was essential in this study, as all the cloned TFs have overlapping restriction enzyme sites. Therefore, the MASH1 gene was first subcloned into the p3X-2A vector, and then the DLX2 gene was inserted, yielding p3X-2A-MASH1/DLX2 **(Figure 3.3 A)**. In addition, each gene was cloned separately into the vector p3X-2A to construct p3X-2A-MASH1 and p3X-2A-DLX2 **(Figure 3.4 A and Figure 3.5 A)**. The GSX2 gene was not added directly to the p3X-2A vector, as this cloned gene has the restriction enzyme site for *Sal*I, which was used to subclone the polyprotein from the p3X-2A vector into the pCAGG expression vector. Therefore, GSX2 was inserted into the *Bgl*II restriction site into the pCAGG vector after the insertion of the polyprotein coding sequence into pCAGG using the *Sal*I restriction site.

The sequence encoding the polyprotein, comprising the cloned genes of MASH1 and DLX2, separately or together, from the p3X-2A vector that was flanked by *Sal*I restriction sites, was digested and inserted into the expression vector pCAGG to construct the expression vectors pCAGG-MASH1/DLX2, pCAGG-MASH1 and pCAGG-DLX2 **(Figure 3.3 B, Figure 3.4 B and Figure 3.5 B)**. Then, the cloned gene GSX2 was inserted into the expression vector pCAGG using *Bgl*II sites. There is only one *Bgl*II restriction site in this vector, located upstream of the T2A peptide, for the insertion of GSX2. GSX2 was inserted into pCAGG-MASH1/DLX2 and pCAGG-DLX2. The insertion of GSX2 yielded the expression vectors pCAGG-MASH1/DLX2/GSX2 and pCAGG-DLX2/GSX2 **(Figure 3.3 B and Figure 3.5 B)**. The sequences are shown **in Appendix 3.2.**

TF expression vectors were validated through the transient nucleofection of human embryonic kidney 293 (HEK293) cells. The cloned TF expression vectors, with different combinations of TFs and the empty vector (pCAGG), were validated before nucleofecting them into hPSCs. To do so, the TF expression vectors were nucleofected (transient nucleofection) into HEK293 cells for western blotting analysis to confirm that the appropriate combinations of TFs were co-expressed via the self-cleavage of 2A peptides from the polycistronic vector. Gene expression from the HEK293 nucleofected cells was analysed by western blotting 48 h after nucleofection. As the percentage of the GFP expressing cells was high, protein lysates were harvested and western blotting was performed.

HEK293 cells were used for the TF vector validation, since they are commonly used, and the AMAXA nucleofection efficiency is very high (93%) and the toxicity is low (Maurisse *et al*. 2010).

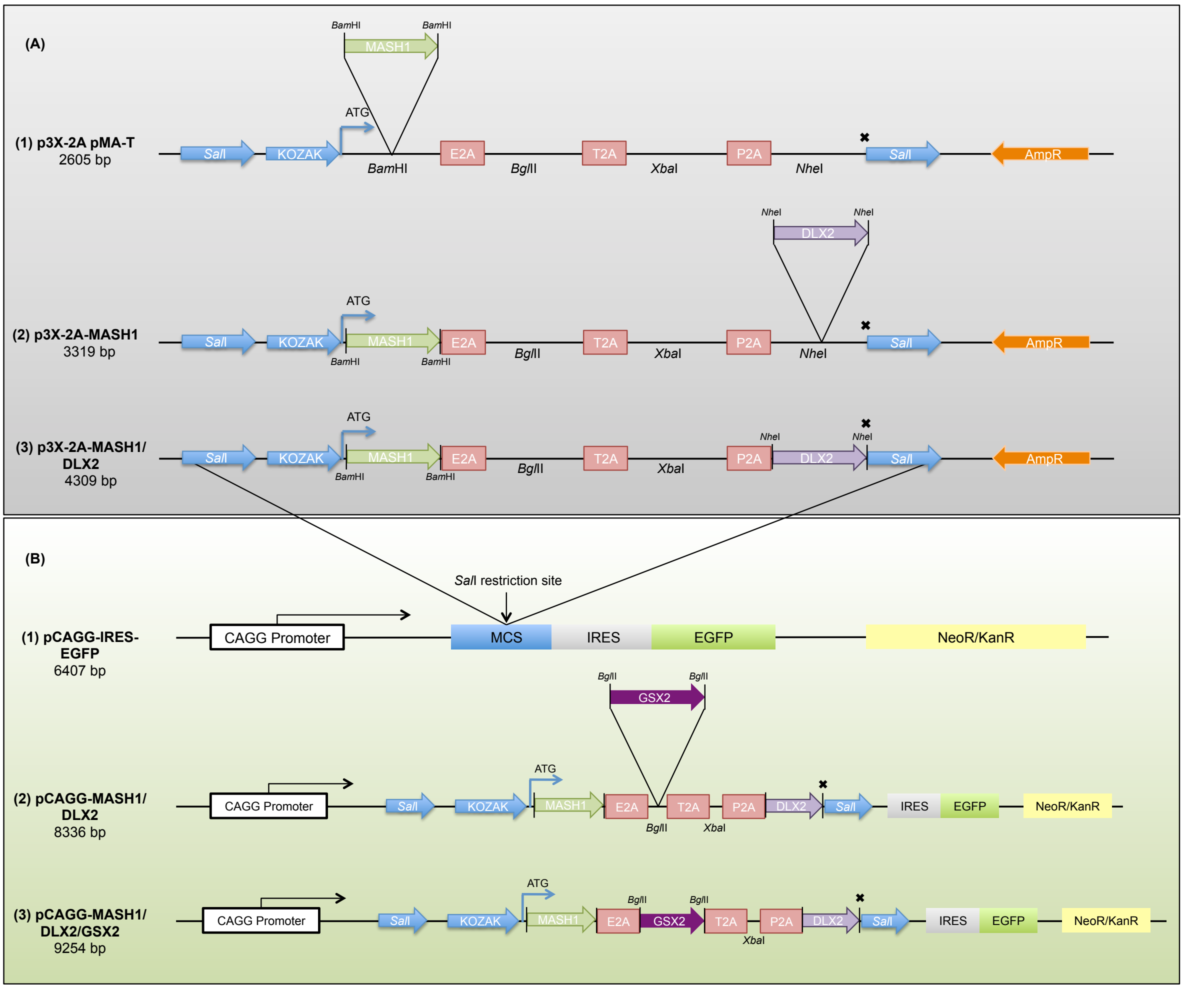


Figure 3.3: The cloning of TFs MASH1, DLX2 and GSX2 into the p3X-2A vector, and subsequent subcloning into the expression vector pCAGG via the *Sal*I restriction site.

The TF MASH1 was first inserted into the p3X-2A vector using the *Bam*HI restriction site (A-1). Afterwards, the TF DLX2 was inserted using *Nhe*I to yield the construct p3X-2A-MASHI/DLX2 (A-2 and A-3). Next, the polycistronic construct containing both DLX2 and MASH1 was subcloned into the pCAGG expression vector using the *Sal*I restriction site to yield the construct pCAGG-MASH1/DLX2 (B-1). At this point, the TF GSX2 was inserted to yield pCAGG-MASHI/DLX2/GSX2 (B-2 and B-3).

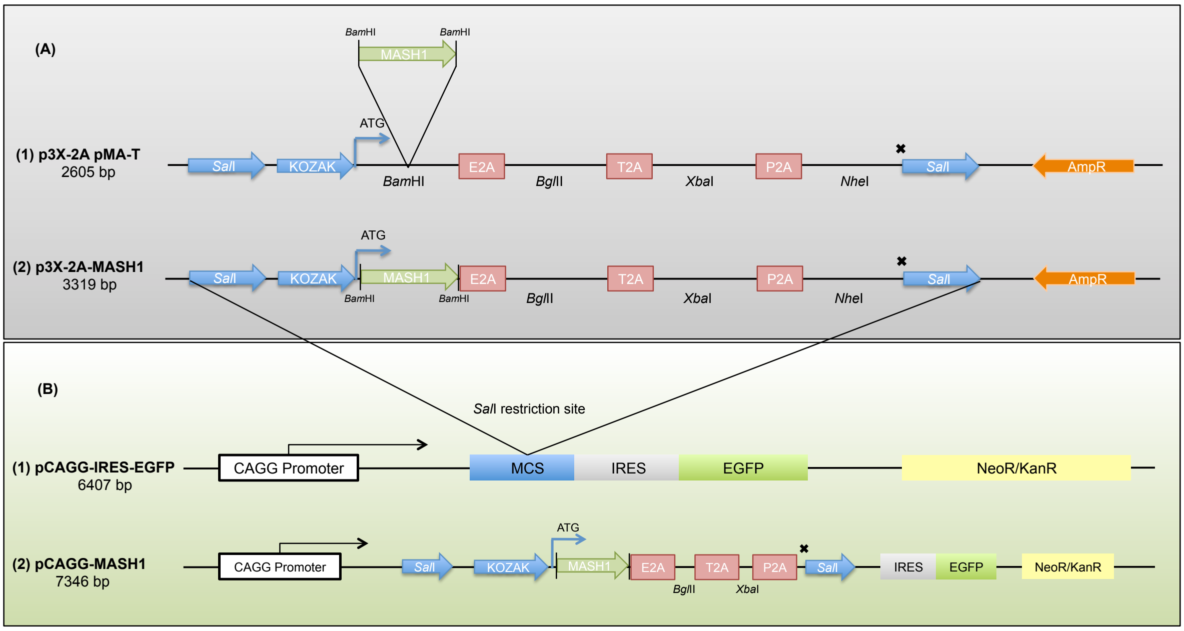


Figure 3.4: The cloning of TF MASH1 into the p3X-2A vector, and subsequent subcloning into the expression vector pCAGG using the *Sal*I restriction site.

The TF MASH1 was inserted into the p3X-2A vector using the *Bam*HI restriction site (A-1 and A-2). Next, the polycistronic construct containing the TF MASH1 was subcloned into the pCAGG expression vector using the *Sal*I restriction site to yield the construct pCAGG-MASH1 (B-1 and B-2).

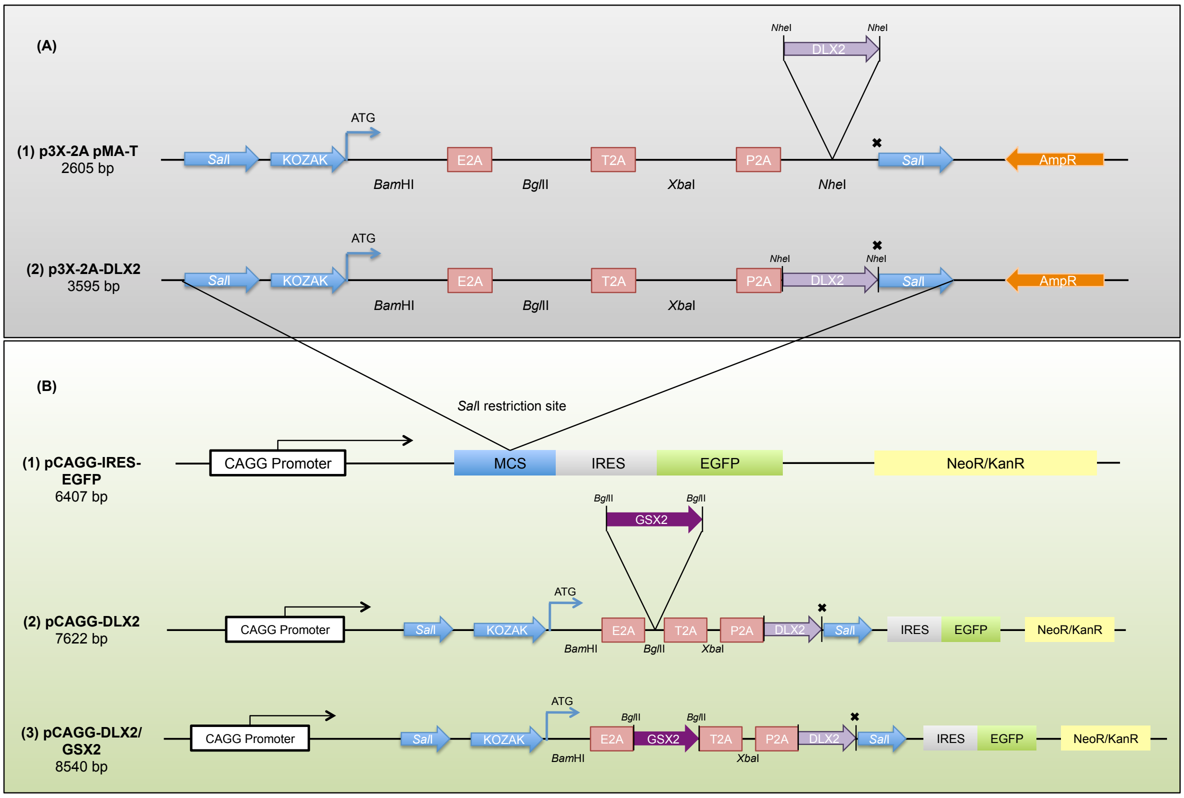


Figure 3.5: The cloning of TFs DLX2 and GSX2 into the p3X-2A vector, and subsequent subcloning into the expression vector pCAGG using the *Sal*I restriction site.

The TF DLX2 was inserted via the *Nhe*I restriction site into p3X-2A to yield the construct p3X-2A-DLX2 (A-1 and A-2), and then subcloned into the pCAGG expression vector via the *Sal*I restriction site (B-1); at this point, GSX2 was inserted using the *Bgl*II restriction site into pCAGG-DLX2 to construct the expression vector pCAGG-DLX2/GSX2 (B-2 and B-3).

3.2 Cell culture techniques

3.2.1 Maintenance of cell lines in culture

3.2.1.1 H9 human embryonic stem cells (hESCs)

H9 cells were grown in the H9 medium. KnockOut Dulbecco's modified Eagle's medium (DMEM) (Cat. No. 10829018, Invitrogen), was supplemented with 15% knockout serum replacement (KSR) (Cat. No. 10828028, Invitrogen), 1% non-essential amino acids (NEAA) (100 ×) (Cat. No. 11140035, Invitrogen), 1% L-Glutamine (L–G) (200 mM) (Cat. No. 25030024, Invitrogen), 1% penicillin (100 U/ml) streptomycin (100 µg/ml) (Pen/Strep) (Cat. No. 15070063, Invitrogen), and 0.68% beta-mercaptoethanol (55 mM) (Cat. No. 21985-023, Invitrogen). H9 cells were grown in 6- or 10-cm Nunc tissue culture plates (Cat. No. TKT–110-010S, TKT–110-170T, respectively, Thermo Fisher Scientific) in a humidified incubator at 37°C, 5% carbon dioxide (CO_2).

H9 cells were passaged 1:4. Prior to passaging, areas of differentiated cells were pruned from culture plates by scraping with a P20 Gilson pipette tip. Plates were then washed once with phosphate-buffered-saline (PBS) without calcium and magnesium (Cat. No. 100-10-056, Invitrogen). hESC colonies were then lifted from their matrix by digestion with pre–warmed, filter sterilised, collagenase type IV (1mg/ml) (Cat. No. 17104019, 1 g/U, Invitrogen) in DMEM containing 10 µM Y-27632 Rho-associated protein kinase (ROCK) inhibitor (Cat. No. ab120129, Abcam Biochemicals, Cambridge, UK) and incubated for a maximum of 20 min in a humidified incubator at 37°C, 5% CO_2. Lifted colonies were gently collected using P1000 Gilson tips with pre-warmed H9 medium (without fibroblast growth factor 2 (FGF2), Cat. No. 10018B, 1 mg/U, PeproTech, Rocky Hill, NJ, USA) in a 15-ml polypropylene tube (Falcon tubes) and pelleted by centrifugation at 1,000×*g* for 3 min. The pellet was re-suspended in pre-warmed H9 medium containing 5 ng/ml FGF2 and 10 µM Y-27632 ROCK inhibitor. Cells were plated into freshly prepared, irradiated mouse embryonic fibroblasts (mefi) plates. Cultures were incubated in a humidified incubator at 37°C, 5% CO_2 and were fed with fresh medium daily.

3.2.1.1.1 Preparing irradiated mouse embryonic fibroblasts (mefi) for plating

E13 mouse embryos were dissected to remove extra-embryonic tissues, heads and viscera. The carcasses were minced in Hank’s balanced salt solution (HBSS) with a sterile blade and incubated with trypsin – EDTA (Cat. No. 253-00-054, 100ml/U, Invitrogen) for 5 min in a humidified incubator at 37°C, 5% CO_2. Digested tissue was collected with MEF medium (DMEM, 10% foetal bovine serum (FBS), 1% Pen/Strep and 1% Antibiotic/Antimycotic (Anti/Anti) (Cat. No. 15240062, 100ml/U, Invitrogen). The cell suspension was centrifuged at 1,000×*g* for 3 min and the cell pellet was washed three times by re-suspension in MEF medium. Cells from one embryo were plated on a 14-cm Nunc plate (Cat. No. TKT-110-130Y, Thermo Fisher Scientific). Cells were fed every other day with fresh medium.

When MEFs reached confluence, they were passaged using trypsin (P1), 80% of the cells were irradiated by γ-radiation (see below) and stored at –80°C, while the other 20% was sub-cultured at a 1:5 ratio. For the second passage (P2), all the cells were irradiated and stored at -80°C. Here, the mefi (irradiated mef) cells were frozen with the freezing medium, 10% dimethyl sulfoxide (DMSO) (Cat. No. D2650, Sigma-Aldrich) and mef medium; this was done at a density of 1×10^6 cells/vial.

At this point, when the mef cells were ready to be irradiated, the cells were dissociated by adding trypsin, and then the cells were collected in a 50-ml Falcon tube, after which they were centrifuged at 1,000×*g* for 3 min. Then, the pellet was re-suspended in 20 ml of mef medium, and irradiated by γ-radiation for 30 min.

Plating mefi cells on gelatin-coated Nunc plate dishes for H9–hESCs maintenance

Eight 6-cm Nunc plates were coated with gelatin (ultrapure water with 0.1% gelatin, Cat. No. ES006B, Millipore, Hertfordshire, UK) and incubated in a humidified incubator for 30 min. One vial of mefi cells was then thawed with 1 ml of pre-warmed mef media and mixed. All the cells were transferred into a 15-ml Falcon tube containing 4 ml of mef medium and centrifuged at 1,000×*g* for 3 min. The pellet was re-suspended in pre-warmed mef medium and plated into a gelatin-coated 6-cm Nunc plate. Cultures were incubated in a humidified incubator. The mefi cells were prepared one day prior to use and used within 7 days of plating.

3.2.1.2 Human-induced pluripotent stem cells (h-iPSCs)

The iPSCs were the 34D6 cell line (Bilican *et al*. 2012); these cells were cultured with mTeSR1 media (Cat. No. 05850, StemCell Technologies, Manchester, UK). The medium was complete and serum-free. It was supplemented with 1% Pen/Strep and 5× supplement (Cat. No. 05852, StemCell Technologies), which contains recombinant human basic fibroblast growth factors (rhbFGF), as well as recombinant human transforming growth factor β (rhTGFβ). 34D6 cells were grown in a 6–10 cm Nunc tissue culture plate in a humidified incubator at 37°C, 5% CO_2.

The 34D6 cells were sub-cultured when they reached a dense state; when the colonies were large and starting to merge, the centers of the colonies were dense with a bright phase in comparison to the edges. Prior to passaging, 10-cm petri dishes were coated with BD Matrigel (Cat. No. 7341440, VWR, Leicestershire, UK) and incubated at 37°C. h-iPSCs colonies were then lifted from their matrix by digestion with pre–warmed dispase solution, which is a protease that is inhibited by EDTA, at a concentration of 1 mg/ml (Cat. No. 07923, StemCell Technologies) with 10 µM Y-27632 ROCK inhibitor and incubated at 37°C for 8 to 20 min depending on when the edges of the colonies started to fold back slightly. After incubation, the dispase was aspirated, and each plate was gently rinsed 2–3 times with pre–warmed DMEM/F12 (5 ml/plate) to eradicate any remaining dispase. Lifted colonies were gently dislodged with P1000 Gilson tips or a cell scraper in pre-warmed DMEM/F12 or mTeSR1 and collected in 15-ml Falcon tubes. If the cells were detached with mTeSR1, then the volume of the mTeSR1 medium was adjusted and the cells were plated onto Matrigel-coated plates. However, if the cells were removed using DMEM/F12 medium, they were pelleted by centrifugation for 3 min at 1,000×*g*. Next, the pellet was re-suspended with an accurate amount of pre-warmed mTeSR1 and plated on Matrigel-coated plates. Cultures were incubated in a humidified incubator at 37°C, 5% CO_2 and were fed with fresh medium daily.

3.2.1.3 Human embryonic kidney 293 (HEK293) cells

HEK293 cells were grown in DMEM/F12 (1:1) medium (Cat. No. 21331020, Invitrogen) supplemented with 10% FBS (Cat. No. 10106-169, Invitrogen), 1% L-G and 1% Pen/Strep. Medium was sterilized by filtration through a 0.2-μm sterile filter before use. HEK293 cells were cultured in T25 tissue culture flasks (TKT-130-150-L, Thermo Fisher Scientific) in a humidified incubator at 37°C, 5% CO_2.

When the cells became dense, they were passaged with 3 ml trypsin for 5 min at 37°C. The cells were lifted off by flicking the flask with HEK293 medium and collected in a 15-ml Falcon tube. Then, the cells were pelleted by centrifugation for 3 minutes at 1,000×*g*. The pellet was re-suspended with an appropriate amount of HEK293 medium and re-plated into T25 flasks at a ratio of 1:8 or 1:10. Cultures were incubated in a humidified incubator at 37°C, 5% CO_2 and fed with fresh medium every other day.

3.2.2 Preparation of frozen cells

The cells were preserved at -80°C or in liquid nitrogen for long-term storage. Prior to freezing, H9 and 34D6 cells were detached from plates by collagenase **(Section 3.2.1.1)** or dispase **(Section 3.2.1.2),** respectively. Then, the cells were centrifuged at 1,000×*g* for 3 min and re-suspended in freezing media containing 10% DMSO in FBS for H9 cells and Cryostor CS10 (Cat. No. 7930, StemCell Technologies) for 34D6 cells. Then, the cells were divided into aliquots using1-ml cryopreservation vials (Cat. No. 4000198, 100ml/U, Thermo Fisher Scientific) and stored at 80°C in a Mr. Frosty freezing container (Cat. No. 5100-0036, Thermo Fisher Scientific) overnight, and then transferred to a liquid nitrogen tank.

3.2.3 Thawing frozen cells

3.2.3.1 H9 and 34D6 cells

Using a P1000 pipette, a vial of H9 or 34D6 cells was quickly thawed to a semi-solid state and returned to a pre-warmed medium, such as H9 medium, in a 15-ml Falcon tube. Tubes were centrifuged at 1,000×*g* for 3 min to remove the DMSO. After centrifugation, the pellets were re-suspended in 5 ml of H9 medium for H9 cells and DMEM/F12 medium for 34D6 cells and centrifuged again. Subsequently, the cells were re-plated.

For H9 cells, the mefi cells plates were prepared by removing the mef medium and washed once with PBS. Then 5 ml of H9 media containing FGF2 and Y-27632 ROCK inhibitor was added to the mefi plate. After a second centrifugation step, the cell pellet was re-suspended in 1 ml of H9 medium, and the cells were transferred to the mefi plate.

For 34D6 cells, after the second centrifugation step, the pellet was re-suspended in mTeSR1 medium and re-plated on Matrigel-coated plates.

Feeding was initiated 24 h after thawing, and the cells were fed every day.

3.2.3.2 HEK293 cells

One vial of HEK293 cells was quickly thawed using a 37°C water bath. Then, 1 ml of DMEM/F12 medium was gently added to the cells (in drops) and transferred to a 15-ml Falcon tube, which was centrifuged at 1,200×*g* for 4 min. The pellet was washed again with DMEM/F12 and re-suspended in HEK293 medium (**Section 3.2.1.3)**, then re-plated in a T25 flask.

3.2.4 Neural induction (Neurogenic Embryiod Bodies – NEBs)

Neural differentiation was performed in either embryoid body (EB) culture or monolayer culture. For EB culture, hPSC colonies were washed 3× with PBS and whole colonies lifted by collagenase (for H9 hESCs) or dispase (for 34D6 h-iPSCs) treatment, as described for cell passaging above (**Sections 3.2.1.1 and 3.2.1.2**, respectively). After collection, colonies were re-suspended in Advanced DMEM/F12 (ADF) medium supplemented with 700 μl of 10 mg/ml insulin (Cat. No. 407709, Millipore), 300 μl of 12.5 mg/ml transferrin (Cat. No. T8158, Sigma–Aldrich), 1% lipid concentrate (Cat. No. 11905031, Invitrogen), 1% L-G, 1% Pen/Strep and 0.68% beta–mercaptoethanol, Y-27632 ROCK inhibitor (for the first day), 0.5 μM LDN193189 (Cat. No. 04-0074-10, StemGent, Cambridge, UK) and 10 μM SB431542 (for the first four days) (Cat. No. ab120163, Abcam Biochemicals) and plated in a sterile dish plate (Cat. No. PDS140050F, Thermo Fisher Scientific) to obtain suspension cells in ADF media. The SB431542 was then removed on the fourth day. The NEBs were fed every other day with half a quantity of fresh media **(Figure 3.6)**.

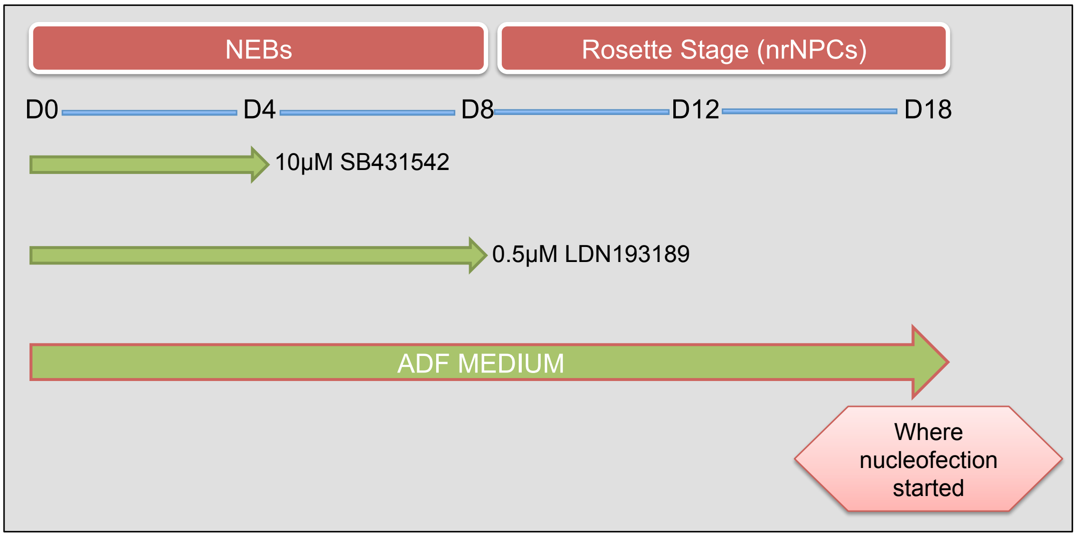


Figure 3.6: The differentiation protocol, from the undifferentiated state to NEBs to Nr-NPCs

The NEBs were cultured with SB431542 and LDN193189 and then they were removed on days 4 and 8, respectively. Subsequently, the rosettes state was performed and cultured with ADF medium until day 18 (D18).

NEB differentiation

Differentiation of hESCs generates naïve rosette-stage neural progenitors (Nr-NPCs). The day before plating, a 24-well plate was prepared, coated with 2 μg/ml poly-L-Lysine (PLL) (diluted with dH_2O) (Cat. No. P5899-5 MG, Sigma-Aldrich) and incubated at 37°C for 30 min, washed with dH_2O (Cat. No. 15230188, Invitrogen), and incubated overnight inside a hood. The next day, the plate was coated with 10 μg/ml laminin (Cat. No. 130095602, Miltenyi Biotec, Surrey, UK) and incubated at 37°C for 30 min. The NEBs were centrifuged at low speed (500×*g* for 2 min) and mixed with Accutase for 5 min to obtain a single cell suspension (Accutase, Cat. No. L11007, PAA Laboratory, Farnborough, Hamsphire, UK). Subsequently, they were centrifuged at 1,000×*g* for 3 min and 1 ml of ADF media was added to conduct cell counts. An appropriate amount of ADF was added to the cells, which were re-plated onto coated PLL/Laminin plates at a concentration of 50 x 10^3 cells/well; cells were fed every other day.

3.2.5 Cell counts

From a 1 ml cell suspension, 10 μl of the cells were added to 10 μl of trypan blue (Cat. No. T8154, Sigma-Aldrich), after which they were added to a haemocytometer with a glass cover slip. The cells in the central square (5 × 5 grid) and the four corner squares were then counted, and the number of cells per ml was calculated using the following equation:

Total number of cells × (dilution factor / number of squares) × 10^4 cell/ml = cells/ml

3.2.6 AMAXA Nucleofection

The cells were treated with ROCK inhibitor prior to nucleofection. Lonza AMAXA kits (AMAXA Mouse NSC (MNSC) with the Nucleofector kit, Cat. No. VPG-1004 with the optimized programme A-023, Lonza Biologics plc, Cambridge, UK) were used for hESCs, and the AMAXA cell line kit V, Cat. No. VCA-1003, with the programme HEK293, was used for HEK293 cells.

The medium was aspirated and cells were washed once with PBS. Then, Accutase with ROCK inhibitor was added for 2–5 min. Then, the cells were collected into a 15-ml Falcon tube and centrifuged at 1,000×*g* for 3 min, and washed again with PBS. At this point, the pellet was re-suspended with nucleofection solution, which was composed of 78 µl of MNSC solution plus 22 µl of nucleofector supplement, and plasmid DNA (up to 20 µg/ml) was added to the nucleofection solution. The whole mixture of cells, plasmid DNA solution and the supplement was transferred into a new cuvette, and subjected to nucleofection with the appropriate programme. Immediately after this, 1 ml of pre–warmed media (ADF or HEK293 medium with ROCK inhibitor) was added to the nucleofected cells, and this process was followed by a cell count **(see section 3.2.5)** and re-plating into Matrigel-coated, 6-well Nunc plates at a concentration of 1 x 10^6 cells/well. Green fluorescent protein (GFP) expression was checked after 24 and 48 h, and cells were fed every other day. The selection was started after 48 h.

3.2.7 Neomycin selection

Because they were neomycin-resistant, 48 h after nucleofection, the cells were selected using G418 sulphate, (Cat. No. 10131-027, 50 mg/ml, Invitrogen). Different concentrations of G418 with different incubation times were tested to find the most suitable concentration of G418 and duration time to select GFP^{+ve} cells.

Hence, the surviving cells were the ones that expressed the cloned transcription factors (TFs), and they were, therefore, selected. After selection, the surviving cells were washed and fed with ADF media for one day. Then, the selected cells were re-plated into 24-well plates **(Section 3.2.7.1 below)** at a density of 30×10^3 cells/well for the differentiation experiments. Before re-plating, cover slips placed onto the 24-well plates were subjected to a specific treatment as described below, to ensure strong adherence and subsequent growth on the cover slips for further downstream applications.

For the H9 differentiation experiments, cover slips were placed onto 24-well plates and coated with 10 mg/ml poly-L-lysine (PLL) (diluted with dH_2O) and incubated at 37°C for 30 min. Next, the cover slips were washed three times with dH_2O and incubated overnight inside a sterile hood to dry. The next day, the cover slips were coated with 10 μg/ml laminin and incubated at 37°C for 30 min. Approximately 5 ml of dH_2O was added between the wells (inter-well space) to avoid drying of the re-plated cells (H9 neural progenitors).

For the 34D6 differentiation experiments, poly-D-lysine (Cat. No. 27964-99-4, 5 mg/ml, Sigma-Aldrich) (PDL)/Matrigel-coated cover slips were used. In addition, the cover slips were treated with nitric acid. One hundred cover slips were placed into a

50-ml Falcon tube containing 25 ml of nitric acid and rocked overnight. Then, the cover slips were washed 3-5 times with ddH_2O. Then, they were washed once with absolute ethanol. Finally, the cover slips were placed on a glass petri dish and baked at 150°C overnight.

Subsequently, the cover slips were placed onto 24-well plates and coated with 100 µl of 100 µg/ml PDL in borate buffer, at pH 8.4 for 1 h at RT. The borate buffer was prepared with 1.24 g of boric acid, 1.90 g sodium tetraborate and 400 ml H_2O at pH 8.4. Each well was washed three times with dH_2O to remove the borate buffer. The plates were placed inside a hood to dry off the cover slips. Then, the cover slips were coated with 50 µl of Matrigel, and incubated for 1 h. The Matrigel was diluted (1:25) in cold medium using Knockout DMEM. Approximately 5 ml of H_2O was added between the wells (inter-well space) to avoid drying the Matrigel and the re-plated cells (34D6 neural progenitors).

3.2.7.1 Re-plating the selected nucleofected cells into 24-well plates containing treated cover slips

The media was aspirated off the 24-well plate, which was washed once with PBS without calcium and magnesium. The cells were dissociated with Accutase with Y-27632 ROCK inhibitor for 2 to 5 min, and the pellet was collected by centrifugation at 1000×g for 3 min. The pellet was re-suspended in 1 ml of ADF to conduct cell counts **(Section 3.2.5)**. After cell counting, 30×10^3 cells were re-plated as a droplet onto the treated cover slips and incubated at 37°C for 30 min to allow the selected nucleofected cells to attach to the cover slips. Then, each well of the 24-well plates was flooded with ADF media for 3 to 5 h. Subsequently, the ADF media was aspirated off and replaced with differentiation media. The cells were fed every 2 days.

3.3 Immunocytochemistry (ICC)

The cells were washed once with PBS and fixed with fresh 4% paraformaldehyde (PFA) (Cat. No. P6148, Sigma-Aldrich) at pH 7.8 for 10 min at 4°C. Next, they were washed three times for 5 min with PBS. In the next stage, cells were permeabilized using PBS with 0.1% Triton X-100 (for internal antibody (Ab) only) (Cat. No. T8787, Sigma-Aldrich) for 20 min at RT or 100% ethanol for 20 minutes at RT. The solution was aspirated, and the cells were washed three times with PBS for 5 min. Next, 1 M glycine (Cat. No. 67419, Sigma-Aldrich) was added for 20 min at RT. Next, blocking was initiated using 2% normal goat serum (NGS) (Cat. No. S-1000, Vector Laboratory, Peterborough, UK), according to the secondary Ab, 3% BSA (Cat. No. A8531, Sigma-Aldrich) and 0.1% Triton X-100 (for internal Ab only). Blocking was carried out at RT for 1 h, followed by incubation with the primary Ab (1° Ab) overnight at 4°C; cells were then washed three times with PBS, followed by secondary Ab (2° Ab) incubation for 1 h in the dark at RT. At this point, cells were washed again. For nuclear staining and mounting, VECTASHIELD Mounting Medium with DAPI (Cat. No. H1200, Vector Laboratories) was used.

The Abs used in this study and their dilutions are shown in the table below

Table 3.2: Antibodies and dilutions used for ICC

ABs	Primary Antibody (1°Ab) Species	1°Ab Dilution
β – TUBULIN III	Anti-mouse (Cat. No. T8660, Sigma-Aldrich) and Anti-rabbit (Cat. No. T2200, Sigma-Aldrich)	1:800 (for anti-mouse) 1:400 (for anti-rabbit)
CTIP2	Anti-rat (Cat. No. ab18465, Abcam)	1:500
DARPP-32	Anti-rabbit (Cat. No. sc-11365, Santa Cruz	1:100

	Biotechnology, Heidelberg, Germany)	
DLX2	Anti-rabbit (Cat. No. ab18188, Abcam)	1:800
FOXG1	Anti-rabbit (Cat. No. 518-694-8188, NeuraCell, Rensselaer, NY, USA)	1:1000
GFP	Anti-rabbit (Cat. No. ab290, Abcam)	1:4,000
GSX2	Anti-rabbit (Cat. No. ABN162, Millipore).	1:500
2A	Anti-rabbit (Cat. No. ABS31, Millipore).	1:500
Ki67	Anti-mouse (Cat. No. VP-K451, Vector Laboratory).	1:100
MASH1	Anti-mouse (Cat. No. 556604, BD Biosciences, Oxford, UK)	1:500
MAP2	Anti-rabbit (Cat. No. ab24640, Abcam). Anti-mouse (Cat. No. MAB3418, Millipore).	1:1000 (for anti-rabbit) 1:500 (for anti-mouse)
Human ZO-1	Anti-mouse (Cat. No. 610966, BD Biosciences)	1:250

3.4 RNA/DNA-related techniques

3.4.1 RNA extraction

RNA extraction was performed using the RNeasy mini kit (Cat No. 74104, Qiagen) according to the manufacturer's guidelines.

The media was removed from the cell-culture dish, which was washed once with PBS, and then 350 µl of RLT buffer was added to disrupt the cells. The lysate was collected into an Eppendorf tube and mixed by vortexing or pipetting to ensure that there were no clumps. The lysate was homogenized using a QIAshredder spin column and centrifuged for 2 min at 13,000×*g*. One volume of 70% ethanol was added to the homogenized lysate, and cells were mixed by pipetting. Then, the samples were transferred into an RNeasy mini column and centrifuged for 15 s at 10,000×*g*. Afterwards, the supernatant was removed and 350 µl of RW1 buffer (contains guanidine thiocyanate) was added to the RNeasy mini column, followed by centrifugation for 15 s. A mixture of RNase-Free DNase I stock and RNase-Free DNase buffer (RDD) was made by adding 10 µl and 70 µl of each component, respectively (Cat. No. 79254, Qiagen). The mixture (80 µl) was added to the RNeasy mini column and incubated at RT for 15 min. Then, 350 µl of RW1 was added, followed by centrifugation for 15 s. Then 500 µl of RPE buffer, which contained 80% of ethanol, was added to the RNeasy mini column, followed by centrifugation for 15 s. The last step was repeated by adding the RPE buffer and centrifuging for 2 min to dry the RNeasy silica-gel membrane. Subsequently, the collection tube was changed, followed by centrifugation at 13,000×*g* for 1 min. Afterwards, the RNeasy mini column was transferred to a new 1.5 ml collection tube and 30 µl of Rnase-free water was added. Finally, the column was centrifuged for 1 min and the eluted RNA was stored at -80°C. The RNA concentration was determined using a NanoDrop spectrophotometer.

3.4.2 Complementary DNA (cDNA) synthesis by reverse transcriptase polymerase chain reaction (RT-PCR)

After RNA extraction, 1 µg mRNA was reverse transcribed into cDNA using SuperScript™ II Reverse Transcriptase (Cat. No. 18064-022, Invitrogen). The protocol was conducted according to the manufacturer's guidelines.

RNA (1 µg in a total volume of 10 µl) was added to 1.5 ml, nuclease-free Eppendorf tubes and kept on ice. Firstly, a master mix (mm) was prepared (1 µl of 10 mM dNTPs (10 mM each of dCTP, dATP, dTTP and dGTP at neutral pH) and 1 µl of random primers (Cat. No. 48190-011, Invitrogen). Two mRNA samples were prepared and labelled as RT+ and RT-. In the RT- sample, H_2O was added instead of SuperScript reverse transcriptase, followed by the addition of 2 µl of mm to the RT+ and RT- samples. The samples were then incubated at 65°C for 5 min and quickly chilled on ice. Then, a mixture of 0.1 M DTT (2 µl), 5× first standard buffer (250 mM Tris-HCl, pH 8.3 at RT; 375 mM KCL; 15 mM $MgCL_2$) (4 µl) and RNaseOUT™ Recombinant Ribonuclease Inhibitor (1 µl) was added. Then, the samples were gently mixed by flicking the tubes, and centrifuged for 2 min at 25°C. Finally, 1 µl of Superscript II reverse transcriptase was added, but only to the RT+ tube, and samples were incubated at 25°C, 42°C and 70°C for 10, 50 and 15 min, respectively. The samples were incubated at 70°C for 15 min to inactivate the reverse transcriptase. The samples were stored at -20°C.

3.4.3 Quantitative polymerase chain reaction (Q-PCR)

Real-time quantitative reverse transcription polymerase chain reactions (qRT-PCR) were conducted using a standard protocol. The cDNA was diluted in TE or Rnase-free water for use in the Q-PCR reactions. mm was prepared as follows:

Components	Volume
Primer (10 pmol/µl)	1 µl
dH_2O	8 µl
mm (DyNAmo HS SYBR Green 5×; F410L)	10 µl

mm (19 µl) was pipetted into each well of a 96-well plate. Then, 1 µl of template DNA was added to each well. All reactions were conducted in triplicate. The Q-PCR reaction was performed in a CFX Connect real-time PCR system machine (Cat. No. 185-5200, Bio-Rad Laboratory, Hertfordshire, UK) using the following conditions:

Cycle repeat	Purpose	Temperature	Time
1x	Initial denaturation	95°C	15 min
40x	Denaturation	95°C	30 s
	Annealing	60°C	30 s
	Extension	72°C	30 s
Plate read			
1x	Melting curve	53°C–95°C	Read every 0.5°C, and hold 00:00:02
End			

The primers used in this project are shown in **Table 3.3**. New primers were also tested **(Appendix 3.1)**.

Table 3.3: Lists of primers used in this project

Gene/Primer	Sequences 5'-3'	Annealing temp.	Amplicon size (bp)
ARX-F	GCTGAAACGCAAACAGAGGC (20 bp)	59.4°C	114
ARX-R	AGTTCCTCCCTGGTGAAGACGT (22 bp)	62.1°C	
B-ACTIN-F	CCCAGCACAATGAAGATCAA (20 bp)	55.3°C	103
B-ACTIN-R	ACATCTGCTGGAAGGTGGAC (20 bp)	59.4°C	
CALBIN-1-F	TGT GGA TCA GTA TGG GCA AAG (21 bp)	57.9°C	96
CALBIN-1-R	CGG AAG AGC AGC AGG AAA T (19 bp)	56.7°C	
CTIP2-F	CCATCCTCGAAGAAGACGAG (20 bp)	57.5°C	106
CTIP2-R	ATTTGACACTGGCCACAGGT (20 bp)	59.8°C	
DARPP-32-F	CTCCAGAGAACGGCATTGTT (20 bp)	58.2°C	116
DARPP-32-R	TCCTGCTCCTGACTTGGATT (20 bp)	58.3°C	
DRD1-F	TGC CAT AGA GAC GGT GAG TA (20 bp)	57.3°C	116
DRD1-R	CAG CAT GTG GGA TCA GGT AAA (21 bp)	57.9°C	
DRD2-F	CAC TCC TCT TCG GAC TCA ATA AC (23 bp)	60.6°C	107
DRD2-R	GAC AAT GAA GGG CAC GTA GAA (21 bp)	57.9°C	
EBF1-F	GTGGAGATCGAGAGGACAGC (20 bp)	59.6°C	99
EBF1-R	AAGCTGAAGCCGGTAGTGAA (20 bp)	59.3°C	
EMX2-F	ACCTTCTACCCCTGGCTCAT (20 bp)	57.8°C	85
EMX2-R	AAAGGAAACTCTCGGGGCTA (20 bp)	55.8°C	
Endo-DLX2-F	TCACCACCACCACCATCAC (19 bp)	58.8°C	96
Endo-DLX2-R	CTCTGCTCTCAGTCTCTGGC (20 bp)	61.4°C	
Endo-MASH1-F	CCCCCAACTACTCCAACGAC (20 bp)	61.4°C	173
Endo-MASH1-R	TCCAAAGTCCATTCGCACCA (20 bp)	57.3°C	
Endo-GSX2-F	CTCCGAGGATGAGGACTC (18 bp)	60.5°C	100
Endo-GSX2-R	AGGAGCGGGGGATGTGAG (18 bp)	58.2°C	
Exo-DLX2-F	ATGTTGAAGAAAACCCCGGTCCT (23 bp)	60.6°C	74

Exo-DLX2-R	GGTCGAGTGCATATCAGCCACTA (23 bp)	62.4°C	
Exo-MASH1-F	AACTACTCCAACGACTTGAACTCCAT (26 bp)	61.6°C	97
Exo-MASH1-R	AGAGCATAATTAGTACACTGGGATCC (26 bp)	61.6°C	
Exo-GSX2-F	AGTGAAGCACAAGAAGGAGGGG (22 bp)	62.1°C	129
Exo-GSX2-R	GACTTCCTCTGCCCTCAGATCT (22 bp)	62.1°C	
FOXP1-F	CGATCCCTTCTCTGATTTGC (20 bp)	56.3°C	103
FOXP1-R	CATGCATAATGCCACAGGAC (20 bp)	57.2°C	
GAD2-F	GCT CTG GCG ATG GGA TAT TT (20 bp)	57.3°C	105
GAD2-R	CCA TTC CTT TCT CCT TGA CTT CT (23 bp)	58.9°C	
GAPDH-F	TGCACCACCAACTGCTTAGC (20 bp)	58.3°C	87
GAPDH-R	GGCATGGACTGTGGTCATGAG (21 bp)	58.1°C	
NKX2.1-F	AGGACACCATGAGGAACAGC (20 bp)	57.1°C	88
NKX2.1-R	CCCATGAAGCGGGAGATG (18 bp)	55.8°C	
PAX6-F	AGGCCAGCAACACACCTAGT (20 bp)	61.4°C	108
PAX6-R	AGCCAGATGTGAAGGAGGAA (20 bp)	58.3°C	

3.5 Western blotting analysis

The compositions of the stock solutions and buffers for protein analysis by SDS-PAGE and western blotting are shown in **Table 3.4** below.

Table 3.4: Composition of solutions used for western blotting

Solutions	Composition
Lysis Buffer	10 ml RIPA Buffer (Cat. No. R0278, Sigma-Aldrich), 1 tablet of PhosStop (Cat. No. 04906845001, Roche) and 1 tablet of Complete Mini 1× solution (Cat. No. 11836153001, Roche).
10x Running Buffer	0.25 M Tris base (Cat. No. T1503, Sigma-Aldrich), 1.92 M glycine (Cat. No. G8898, Sigma-Aldrich), 0.1% sodium dodecyl sulphate (SDS) (Cat. No. L3771, Sigma-Aldrich), and pH 8.3 (diluted 10× before use).
10% SDS	10 g SDS in 100 ml.
1% SDS	1 g SDS in 100ml.
Transfer Buffer	0.25 M Tris base, 1.92 M glycine, and 20% methanol (Cat. No. 34860, Sigma-Aldrich).
Sample Loading Buffer	2% SDS, 10% glycerol (Cat. No. G5516, Sigma-Aldrich), 60 mM Tris base, pH 6.8, 0.005% bromophenol blue (Cat. No. B0126, Sigma-Aldrich) and 500 mM DL-dithiothreitol (DTT) (Cat. No. D0632, Sigma-Aldrich).
Resolving Gel	10% acrylamide (Cat. No. A8887, Sigma-Aldrich), 0.37 M Tris, 0.1% SDS, 0.1% ammonium persulphate (APS) (Cat. No. A3678, Sigma-Aldrich) and 0.06% tetramethylethylenediamine (TEMED) (Cat. No. T9281, Sigma-Aldrich).
Stacking Solution	5% acrylamide, 0.125 M Tris, 0.1% SDS, 0.05% APS and 0.5% TEMED.
Ponceau S	0.1% in 5% acetic acid (Cat. No. P3504, Sigma-Aldrich).
Wash Buffer	1× PBS (Cat. No. 70011-044, Invitrogen), 0.1% Tween 20 (Cat. No. P1379, Sigma-Aldrich)
Blocking Solution	5% skimmed milk powder (Marvel, Lincolnshire, UK) in 1× PBS and 0.1% Tween 20.
Antibody blocking solution	5% milk powder in PBS/Tween + 100 µl sodium azide (Cat. No. S2002, Sigma-Aldrich) (1 M) / 20 ml.
SuperSignal West Dura chemiluminescent substrate	1:1 ratio of SuperSignal West Dura Liminol:Enhancer solution (Thermo scientific, Massachusetts, USA, Cat. No. 34075)

3.5.1 Protein extraction from monolayer cells using RIPA buffer

A confluent, 10-cm dish was washed once with cold PBS then placed on ice. Ice-cold lysis buffer **(see Table 3.4)** and 1× PhosStop solution were added, and the cells were scraped using a rubber policeman. Next, the solution was incubated for 30 min at 4°C with intermittent agitation and centrifuged at 13,000×*g* for 20 min at 4°C. The supernatant was then transferred to a new tube and the pellet was discarded. Samples were stored as aliquots at -80°C.

3.5.2 Protein assay

After the protein extraction from the cells, the concentration of total protein was measured by the Pierce Bicinchoninic Acid (BCA) Protein Assay Kit (Cat. No. 23227, Thermo Fisher Scientific).

Nine standards were prepared using a dilution series of BSA (Cat. No. A9418-100-G, Sigma-Aldrich) ranging from 20 to 2,000 µg/ml. The standards and unknown samples (25 µl each) were plated in triplicate in a 96-well plate. Then, the BCA working reagent (WR) was prepared, as recommended by the manufacturer:

(Number of standards + number of unknown samples) × (number of triplicates) x (volume of WR per samples) = total volume of WR required.

The BCA WR consisted of BCA reagent A and BCA reagent B. The mixing ratio was 1:50 (B:A). WR (200 µl) was added to each well of the standards - in addition to the unknown samples in the 96-well plate - and mixed thoroughly on a plate shaker for

30 s. Then, the plate was incubated at 37°C for 30 min, and allowed to cool to RT. The absorbance was measured at 590 nm using a Bio-Tek plate reader (Epson, California, UK).

3.5.3 SDS polyacrylamide gel electrophoresis

A resolving gel **(Table 3.4)** was prepared and left to set with a layer of 1% SDS on the surface. The SDS was removed; the stacking gel **(Table 3.4)** was added and then left to set with an appropriate comb size. While waiting for the gel to set, the protein samples were prepared by adding a 1:1 ratio of sample loading buffer and protein samples. These were then incubated at 95°C for 5 min to denature the protein samples. Once the gel was set, the samples (20-40 µg/lane) were loaded into the gel wells alongside a rainbow protein ladder (10 µl/lane) (The Novex® Sharp Pre-Stained Protein Standard, Cat. No. LC5800, Invitrogen). Running buffer **(Table 3.4)** was added to the running tank and the stacking gel was run at 80 V, and the resolving gel was run at 120 V for 2 h.

3.5.4 Western blotting

After SDS-PAGE, the gel was removed from its casing. A Hybond ECL membrane (0.2 µm pore size nitrocellulose) (Cat. No. RPN3032D, GE Healthcare Life Science, Belfast, UK) was soaked briefly in dH_2O, and then placed in transfer buffer for 10 min. The gel, sponges and filter paper were soaked in transfer buffer before being layered with filter paper, thus sandwiching the gel and the membrane. They were arranged in the sandwich cassette as follows:

Cathode (-): sponge : filter paper : **Gel : Nitrocellulose** : filter paper : sponge : **Anode (+)**

Next, transfer buffer **(Table 3.3)** was added to a tank containing a magnetic spinner and left in a cold room overnight at 25 V or at 100 V for 1 h. The following day, the membrane was carefully removed and washed once with dH_2O and then with wash buffer **(Table 3.3)**. Finally, it was washed with a wash buffer on a shaker to remove the stain.

3.5.5 Immuno-detection of proteins

Following the blotting, the membrane was removed from the sandwich and washed, as previously described **(section 3.5.4)**. The wash was then removed and replaced with a blocking solution **(Table 3.4)** for 1 h with shaking. At this point, the blocking solution was replaced with a specific primary antibody (a dilution of primary antibody in blocking solution, **Table 3.5**) and left for 2 h at RT, and then left overnight at 4°C. At this stage, the membrane was washed three times for 5 min in wash buffer. Afterwards, the secondary antibody in blocking solution (diluted in blocking solution, **Table 3.5**) was added for 1 h at RT. Then, the membrane was washed again under the same conditions.

Table 3.5: Antibodies and dilutions used for western blotting

Protein	Primary Antibody (1°Ab) Species	1°Ab Dilution	Secondary Antibody (2°Ab) Species	2°Ab Dilution	Protein Size
β-ACTIN	Mouse (Cat. No. A2228, Sigma)	1:10,000	Mouse	1:10,000	42 kDa
GSX2	Rabbit (Cat. No. ARP32208-P050, Aviva System Biology, San Diego, USA)	1:1000	Rabbit	1:10,000	≈ 32 kDa
MASH1	Mouse (BD Biosciences)	1:100	Mouse	1:10,000	≈ 34 kDa
DLX2	Rabbit (Abcam)	1:400	Rabbit	1:10,000	≈ 34 kDa
DARPP-32	Rabbit (Santa Cruz)	1:100	Rabbit	1:10,000	≈ 32 kDa

3.5.6 Detection of chemiluminescence

After the nitrocellulose had been probed with the appropriate antibodies, the membrane was placed in the developing cassette and incubated with SuperSignal West Dura buffer (0.5 ml) **(Table 3.4)** for 5 min in dark. Then, the solution was poured off, and film (Cat. No. 11666657001, Roche) was placed over the membrane and exposed for various time intervals.

3.6 Electrophysiology studies

3.6.1 Whole-cell patch

Standard whole cell patch clamp and analysis was carried out by Dr Vsevolod Telezhkin (Cardiff University), following methods previously reported (Telezhkin *et al*. 2010). Briefly, nucleofected hPSC-derived nrNPCs were mounted on an inverted microscope (Olumpus CK40) and the whole cell patch clamp configuration obtained. Internal solution consisted of 117 mM KCl, 10 mM NaCl, 11 mM HEPES, 11 mM EGTA, 2 mM $MgCl_2$, 1 mM $CaCl_2$ and 2 mM Na_2-ATP and external consisted of 135 mM NaCl, 5 mM KCl, 5 mM HEPES, 10 mM Glucose, 1.2 mM $MgCl_2$ and 1.25 mM $CaCl_{2.}$

Upon successful cell access, zero current injection continuous recordings were made in current clamp mode for measurement of resting membrane potential and spontaneous synaptic activity. Offline data was reviewed and analysed by Axon Laboratory’s Clampfit and Microsoft Office Excel.

3.7 Statistical analysis of data

GraphPad PRISM version 6.0d software was used to analyse the data. All datasets were tested for normality using the D'Agostino-Pearson test.

For single comparisons of the data, a Student's t-test (two-tailed, paired or unpaired) was used. For multiple comparisons, however, two-way analysis of variance (ANOVA) was used with Bonferroni correction, which is one of the multiple-comparison corrections. The results were regarded as significant if the p-value was equal or less than 0.05 ($p\text{-value} \leq 0.05$).

Chapter 4: Generation and validation of vectors for the ectopic expression of the transcription factors DLX2, MASH1 and GSX2.

4.1 Introduction

The transcription factors (TFs) DLX2, MASH1 and GSX2 have proven roles in the specification of the Lateral Ganglionic Eminence (LGE) and medium spiny neuron (MSN) fate determination. This chapter describes the construction of plasmid expression vectors for the three TFs. The open reading frames of the genes were cloned, from cDNA, into polycistronic vectors to enable the co-expression of linked TFs from a single pro-protein (O'Malley *et al*. 2009).

The ability to express more than one protein from a single vector that contains 2A self-cleaving peptides could be used for gene therapy of diseases, such as HD, in which it is unknown whether one or more genes are essential for the efficient production of MSNs. Therefore, in this thesis, the co-expression of different combinations of the three desired TFs was examined to investigate which combination of TFs had a significant role in the production of MSNs. For example, in Parkinson's disease, using a polycistronic vector three genes encoding catecholarminergic synthetic enzymes (tyrosine hydroxylase (TH), aromatic amino acid L-3,4-dihydroxyphenylalanine (DOPA) decarboxylase (AADC) and GTP cyclohydrolase (CH1)) were found to be essential for the efficient generation of dopamine (Azzouz *et al*. 2002; Radcliffe and Mitrophanous 2004).

4.1.1 Self-cleavage 2A peptide

The polycistronic vector used in this study contains three self-cleaving 2A peptide sequences, E2A, T2A and P2A. The 2A peptide sequences was initially reported by Ryan and collaborators (1991) in one genus of the picornavirus family, the foot-and-mouth disease virus (FMDV) (Ryan *et al*. 1991; J. H. Kim *et al*. 2011). The self-cleavage of 2A peptides takes place during translation, thereby releasing each protein (Donnelly *et al*.

2001). Self-cleavage occurs by a process known as ribosome skipping at the C-terminus of the 2A peptide, where the ribosome skips the synthesis of the glycyl-prolyl (G-P) peptide bond. As a result of the cleavage, two proteins will be formed: the upstream peptide (that contains glycine (G) at its C-terminus) and the downstream peptide (that contains proline (P) at its N-terminus) **(Figure 4.1)** (Donnelly *et al*. 2001; J. H. Kim *et al*. 2011).

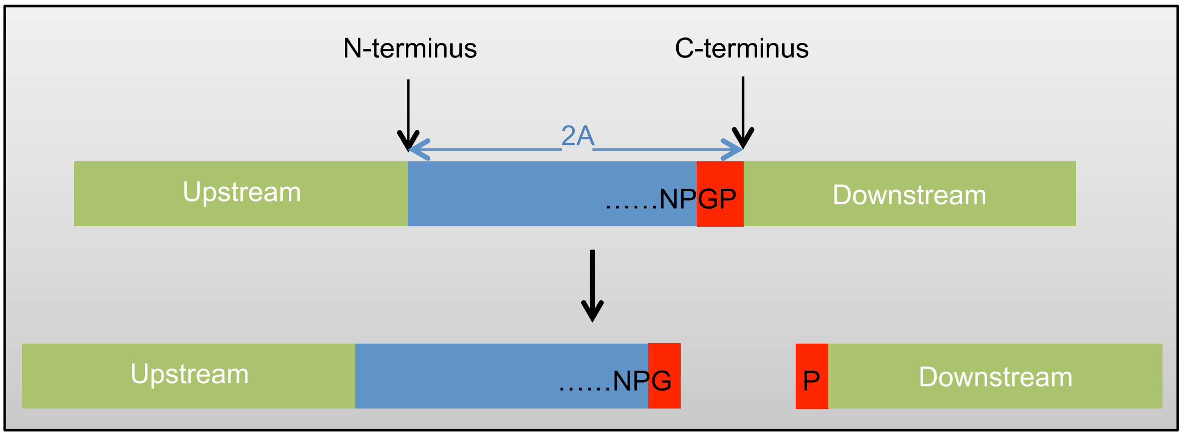


Figure 4.1: Construction of self-cleaving 2A peptide by ribosome skipping.
Diagrammatic representation of self-cleavage of the 2A peptide in the FMDV during translation. The red area indicates where the cleavage site starts, which releases the 2A peptide that contains a G residue at the N-terminus and the downstream peptide that contains a P residue at its C-terminus.

Abbreviations: G: Glycine, P: Proline; NPGP: amino acid sequence of asparagyl-prolyl-glycyl-prolyl.

4.1.2 DLX2

Dlx genes expressed in the telencephalon are confined to the differentiating, γ-aminobutyric acid (GABA)-expressing neurons (Stühmer *et al*. 2002b). The expression of Dlx1 is localised in the telencephalon of the ventricular zone (VZ) and sub-ventricular zone (SVZ) of the LGE and MGE; Dlx1 is also expressed in the mantle zone (MZ) (Poitras *et al*. 2007). The expression of Dlx2 is found in the ventricular and sub-ventricular telencephalon zones of mouse embryos at E12.5, where early differentiation occurs (Panganiban and

Rubenstein 2002). Recently, it was found that the DLX2 gene was expressed in the developing human foetal forebrain at late embryonic stages, and was localised to the SVZ of the entire ventral telencephalon domain (Pauly *et al*. 2013). Dlx expression in the MGE is associated with GABA interneuron development, whereas Dlx expression in LGE progenitors is associated with striatal and olfactory bulb GABA neurogenesis (Poitras *et al*. 2007).

Dlx2 has a role in ventral forebrain patterning and neuronal subtype specification, and it also plays a significant role in striatal and olfactory bulb GABA neurogenesis (Poitras *et al*. 2007). Mice lacking Dlx1 and Dlx2 lack migrating GABAergic interneurons from the telencephalon of the SVZ and VZ of the LGE and MGE to the cerebral cortex, which then results in a four-fold reduction in the number of GABAergic expressing cells in the cerebral cortex, striatum and olfactory bulb, which are the final destinations of the GABAergic interneurons (Anderson 1997). In addition, a mutation in Dlx2 causes abnormalities in the differentiation of late-born striatal neurons (Anderson *et al*. 1997; Anderson 1997; Marin *et al*. 2000).

4.1.3 MASH1

The MASH1 is highly expressed in the region of the SVZ, VZ and MZ of the ventral telencephalon of the LGE and MGE at E12.5 (Parras *et al*. 2004; Castro *et al*. 2011). MASH1 has a major function in regulating neurogenesis in the brain during embryogenesis (Castro *et al*. 2011). The loss of MASH1 results in the acute failure of the basal ganglia neurons in the telencephalon, as well as the loss of cortical projection neurons (Casarosa *et al*. 1999; Horton *et al*. 1999; Marin *et al*. 2000; Yun *et al*. 2002; Castro *et al*. 2011).

MASH1 also regulates a large number of other target genes that promote neurogenesis and have roles in distinct biological processes, molecular functions and cellular processes, as determined by microarray data (Castro *et al*. 2011). In the biological processes, the target genes activated by MASH1 are involved in the early steps of inhibition processes (Notch signaling), cell fate specification, the regulation of cell proliferation and neuronal differentiation in the later steps of neurogenesis (Castro *et al*. 2011). Moreover, in the molecular processes, 48% of the target genes are involved in the regulation of transcription, 36% in signal transduction, 64% in nucleic acid binding, and small percentages in kinase activity (19%), enzyme activity (13%), transporter activity (14%) and cytoskeletal activity (11%) (Castro *et al*. 2011). In addition, MASH1 directly regulates a number of positive cell cycle regulators that promote cell cycle exit (Castro *et al*. 2011).

MASH1 plays a role as a direct regulator of DLX1/2 expression. The human DLX1&2 are orientated in an inverted convergent pattern and named as bigene cluster DLX1/2 **(Figure 4.2)** (Simeone *et al*. 1994; McGuinness *et al*. 1996). It has been reported that MASH1 is an upstream regulator of DLX2 (Porteus *et al*. 1994; Casarosa *et al*. 1999; Fode *et al*. 2000; Letinic *et al*. 2002; Yun *et al*. 2002). Chromatin immunoprecipitation (ChIP) and electromobility shift assay (EMSA) analyses have shown that Mash1 binds to the E-box sequence at FP5, which is a functional basic helix–loop–helix (bHLH) binding site present in the I12b enhancer. This enhancer is located upstream of the bigene cluster Dlx1/2, in which Mash1 binds to the E-box site of the I12b enhancer and activates transcription, and, hence, regulates the Dlx1/2 bigene directly **(Figure 4.2)** (Poitras *et al*. 2007).

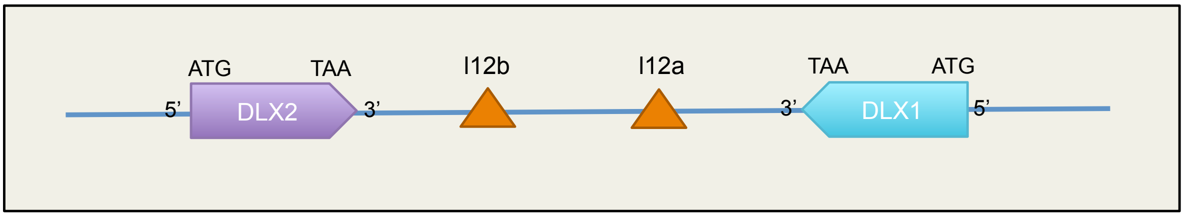


Figure 4.2: Schematic representation of the orientation of the human DLX1&2 loci

The DLX1&2 are closely linked and located in an inverted transcribed manner. The enhancer I2b and I12a are located upstream of the bigene cluster DLX1/2.

Both DLX1/2 and MASH1 have similar expression patterns in the ventral telencephalon region of the proliferation zone of the LGE and MGE (Porteus *et al*. 1994). Further evidence from Mash1 knockout mice shows a reduction in Dlx gene expression in the SVZ of the MGE and the LGE at E12.5 (Horton *et al*. 1999). Moreover, when Mash1 is ectopically expressed in neocortical neurons, Dlx1/2 expression is up-regulated (Fode *et al*. 2000).

The timing of cell fate specification and differentiation in the nervous system of vertebrates is regulated by a lateral inhibition process mediated by Notch signaling (Chitnis and Kintner 1996; Henrique *et al*. 1997; Lewis 1996). MASH1 indirectly influences the activation of this signaling pathway by controlling the expression of Notch ligands, such as Delta and Jagged (Lindsell *et al*. 1996; Castro *et al*. 2006; Henke *et al*. 2009). Notch signaling represses the differentiation of neurons and inhibits proneural bHLH expression, including that of MASH1 (Artavanis-Tsakonas *et al*. 1995; de la Pompa *et al*. 1997; Robey 1997).

It has been suggested that the MASH1 gene is required in the early stages of neurogenesis, and that DLX2 is needed in the late stages of neurogenesis, during the specification and differentiation steps, to down-regulate Notch signaling. Hence, cell fate commitment is regulated by the coordinated functioning of MASH1 and DLX1/2 via the distinct influence on the Notch signaling pathway (Yun *et al*. 2002).

4.1.4 GSX2

GSX1 and GSX2 are the earliest TFs expressed in the LGE progenitor cells. Here, the homeobox GSX genes are involved in the initial specification of the neural progenitors of the LGE (Pei *et al*. 2011). They also play a role in the development of striatal pyramidal neurons and interneurons of the olfactory bulb (Toresson and Campbell 2001; Yun *et al*. 2003). In the early stages of neurogenesis, GSX2 is highly expressed in the progenitors of the ventral LGE, whereas in the later stages, GSX2 is expressed in the progenitor cells of the dorsal LGE (Waclaw *et al*. 2009). During LGE neurogenesis, GSX2 plays a fundamental role in the cell fate commitment of striatal projection neurons, as well as olfactory bulb interneurons, at distinct time points. In the early stages of telencephalic development, GSX2 is highly expressed in the ventral LGE and its main derivatives, such as the striatum. Meanwhile, during the later stages, it is mainly expressed in the dorsal LGE and its derivatives, such as the olfactory bulb (Waclaw *et al*. 2009).

With the loss of Gsx2 in mice, both the ventral and dorsal LGE, coupled with their derivatives, are acutely reduced (Yun *et al*. 2001; Yun *et al*. 2003; Waclaw *et al*. 2004; Waclaw *et al*. 2006). However, when Gsx2 is mutated in the early stages of telencephalon development, the number of striatal projection neurons is reduced. This notwithstanding, when the mutation of Gsx2 is delayed, the olfactory interneurons are defective (Waclaw *et al*. 2009). Hence, the development of the striatum depends on the early expression of Gsx2, and this is also true for the olfactory bulb.

4.2 Aims

The aim of this section was to clone different combinations of the TFs of interest into expression vectors of and to validate vector function before utilizing these vectors in cell differentiation studies with the hPSCs reported in Chapter 5.

4.3 Experimental design

In this chapter, to address the above aim, several experimental strategies were undertaken. The detailed methodology has been described in the Materials and Methods section **(Section 3.1.11)** in Chapter 3.

4.4 Results

4.4.1 DLX2, MASH1 and GSX2 expression vectors

The generation of the TF expression vectors required the subcloning of the ORF for each TF into the interim vector p3X-2A. To achieve this, the ORFs were cloned by PCR using specific primers. The primers used for PCR cloning are shown in Materials and Methods section **(Section 3.1.11, Table 3.1)** in Chapter 3.

PCR cloning was performed from cDNA, which cloned into a plasmid, stocks obtained from the Harvard PlasmID Repository **(Table 4.1)**.

Table 4.1: The plasmidID list from Harvard.

Clone ID	Clone Type	Gene Symbol	Gene Name	Reference Sequence	Vector	Selection Markers
HsCD00345838	cDNA	GSX2	GS homeobox 2	BC075089	pCR4-TOPO	Bacterial: ampicillin; bacterial: kanamycin;
HsCD00338128	cDNA	DLX2	Distal-less homeobox 2	BC032558	pCMV-SPORT6	Bacterial: ampicillin
HsCD00076006	cDNA	MASH1	Achaete-scute complex-like 1	BC002341	pDONR221	Bacterial: kanamycin;

The table above illustrates the three plasmid cDNAs from Harvard used for PCR cloning. Each plasmid contains the gene of interest, (GSX2, DLX2 and MASH1), and a selection marker for growth in *E. coli*. The reference sequence is the GenBank number (NCBI website) that has the information about the insertion of the genes of interest.

4.4.1.1 PCR and pENTR5' TOPO TA cloning

DNA was amplified using a proofreading polymerase (Platinum *Taq* DNA high-fidelity polymerase). The three genes, DLX2, MASH1 and GSX2, were amplified under different conditions. The PCR primers used amplified 930-bp, 725-bp and 1,002-bp products which corresponded to the GSX2, MASH1 and DLX2 gene products, respectively **(Figure 4.3),** and the products were inserted into the TOPO vector pENTR5' TOPO (Invitrogen, UK: Paisley, 2007) using the pENTR5' TOPO TA cloning kit.

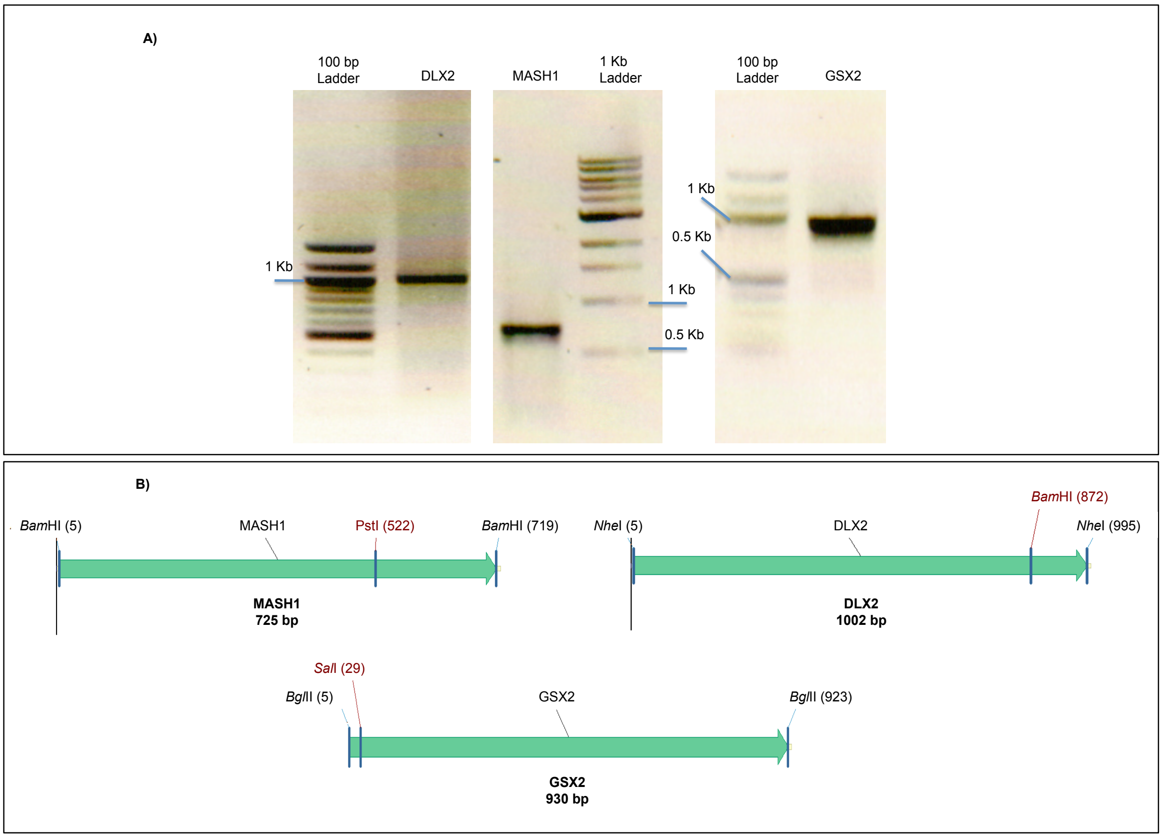


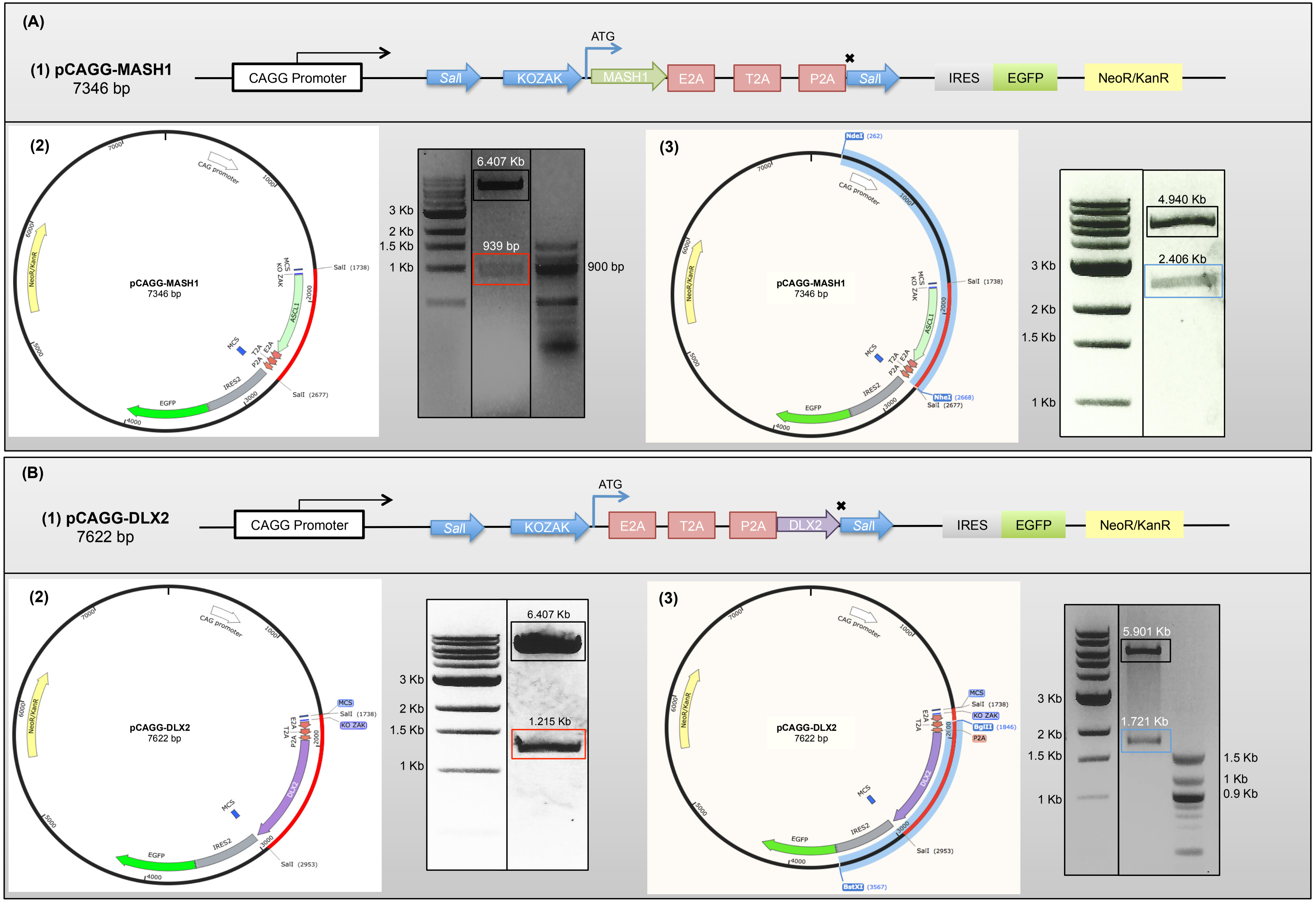
Figure 4.3: Full-length amplification of MASH1, DLX2, GSX2 ORFs

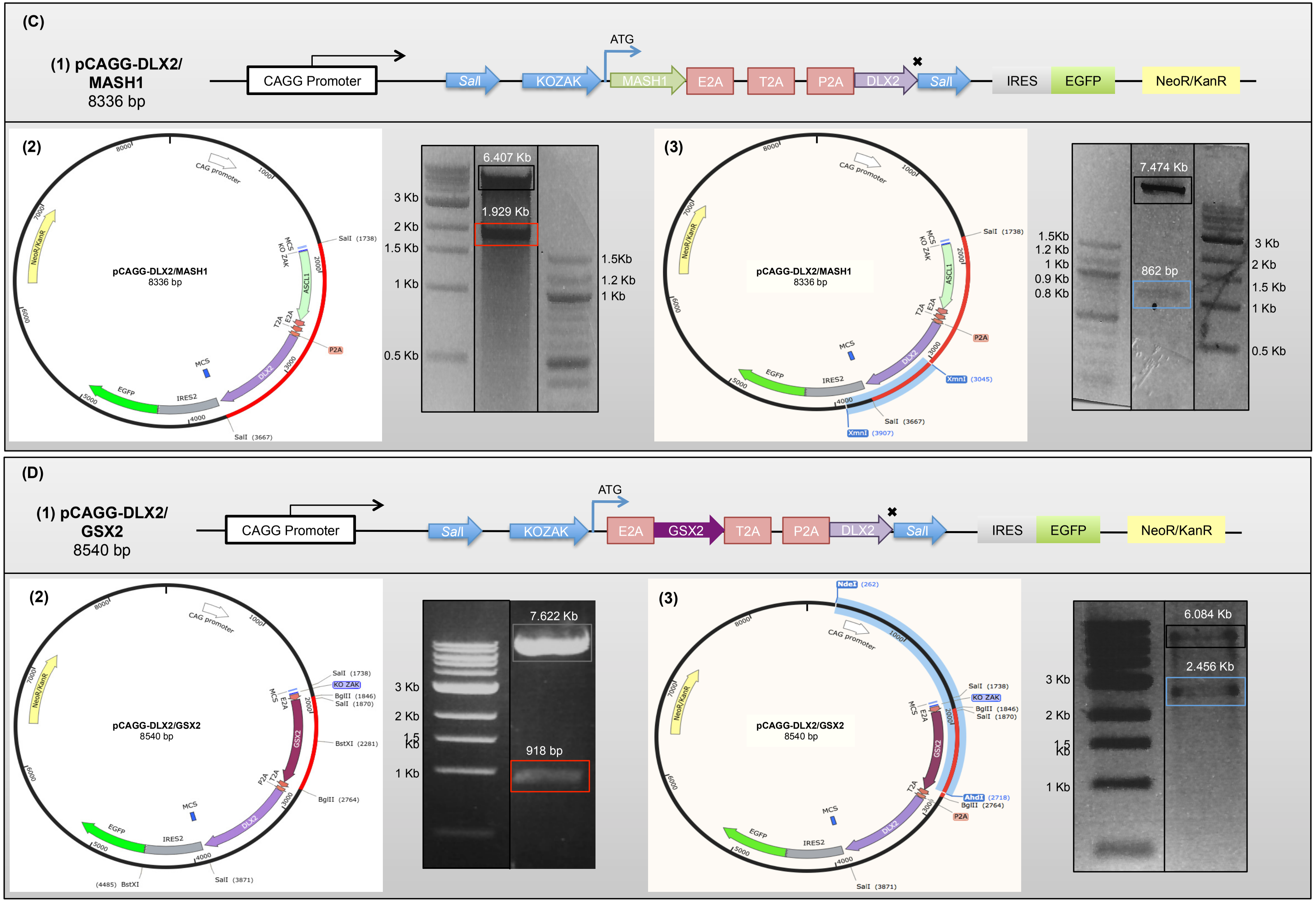
PCR products of the amplified genes (A). Linear map of the PCR products for the three genes: MASH1, DLX2 and GSX2, with restriction sites. (B) In the diagram, it can be seen that there is an overlap of restriction sites between the cloned genes, which is important to take into a count for subcloning into the interim vector p3X-2A and the expression vector pCAGG.

4.4.1.2 Subcloning into the p3X-2A vector designed to insert the cloned genes with three 2A peptide linkers, followed by the transfer of these genes into the expression vector pCAGG-IRES-EGFP

After confirming the proper cloning of TFs by DNA sequencing, the genes, MASH1 and DLX2, were subcloned sequentially into the p3X-2A vector.

Initially, the MASH1 gene was digested with *Bam*HI and ligated with the p3X-2A vector before the DLX2 gene was introduced; this is because this particular gene has the restriction site for *Bam*HI **(Figure 4.3 B)**. Subsequently, p3X-2A-MASH1/DLX2 was digested with *Sal*I and the two cloned genes, MASH1 and DLX2, were inserted into pCAGG-IRES–EGFP to generate the construct pCAGG–DLX2/MASH1. The GSX2 ORF was then inserted into the pCAGG-DLX2/MASH1 vector at the *Bgl*II restriction site. Then, the MASH1 and DLX2 genes were inserted separately into the p3X-2A vector. In the next phase, the gene GSX2 ORF was inserted into the pCAGG-DLX2 vector, thus yielding the pCAGG-DLX2/GSX2 vector. Together, five combinations of the three cloned genes were made for this project. Restriction digestion analyses of these constructs, as well as the gene orientations, is shown in **Figure 4.4**.


(A)
(1) pCAGG-MASH1
7346 bp
CAGG Promoter
SalI
KOZAK
ATG
MASH1
E2A
T2A
P2A
SalI
IRES
EGFP
NeoR/KanR
(2)
pCAGG-MASH1
7346 bp
3 Kb
2 Kb
1.5 Kb
1 Kb
6.407 Kb
939 bp
900 bp
(3)
pCAGG-MASH1
7346 bp
3 Kb
2 Kb
1.5 Kb
1 Kb
4.940 Kb
2.406 Kb
(B)
(1) pCAGG-DLX2
7622 bp
CAGG Promoter
SalI
KOZAK
ATG
E2A
T2A
P2A
DLX2
SalI
IRES
EGFP
NeoR/KanR
(2)
pCAGG-DLX2
7622 bp
3 Kb
2 Kb
1.5 Kb
1 Kb
6.407 Kb
1.215 Kb
(3)
pCAGG-DLX2
7622 bp
3 Kb
2 Kb
1.5 Kb
1 Kb
5.901 Kb
1.721 Kb
1.5 Kb
1 Kb
0.9 Kb


(C)
(1) pCAGG-DLX2/ MASH1
8336 bp
CAGG Promoter
SalI
KOZAK
ATG
MASH1
E2A
T2A
P2A
DLX2
SalI
IRES
EGFP
NeoR/KanR
(2)
pCAGG-DLX2/MASH1
8336 bp
6.407 Kb
1.929 Kb
3 Kb
2 Kb
1.5 Kb
1 Kb
0.5 Kb
1.5Kb
1.2 Kb
1 Kb
(3)
7.474 Kb
862 bp
0.9 Kb
0.8 Kb
(D)
(1) pCAGG-DLX2/ GSX2
8540 bp
GSX2
pCAGG-DLX2/GSX2
8540 bp
7.622 Kb
918 bp
6.084 Kb
2.456 Kb

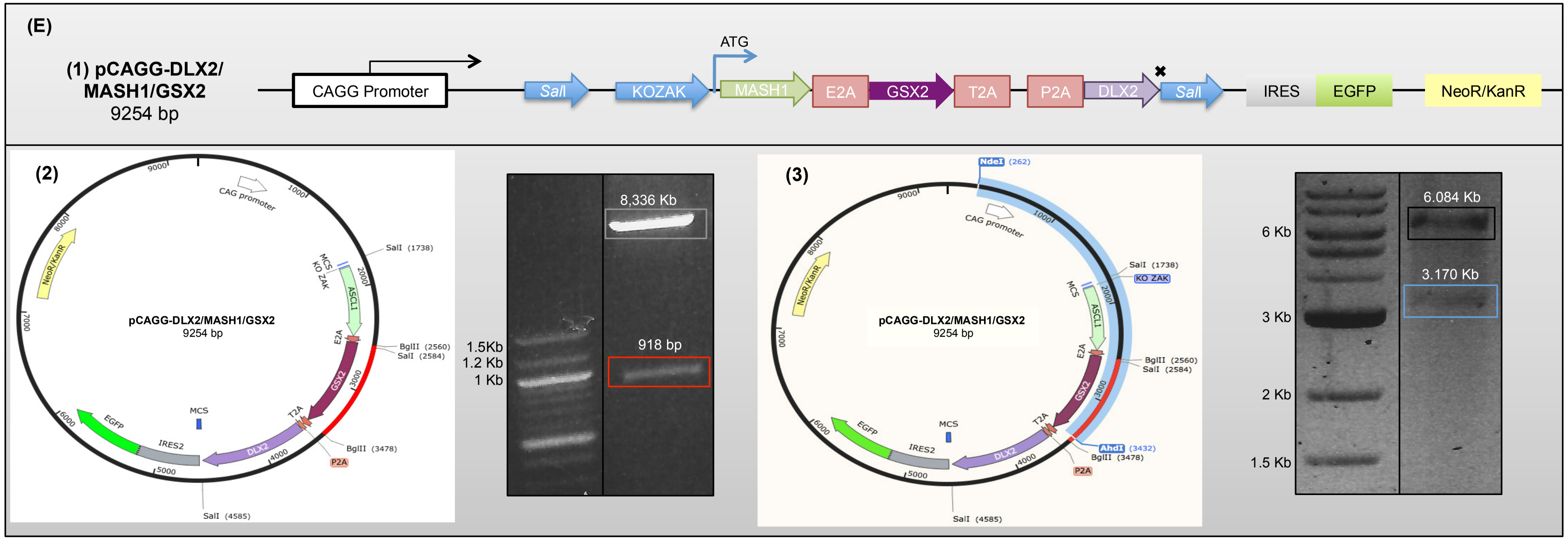


Figure 4.4: Analysis of the insertion and orientation of the inserts in the pCAGG vector by restriction digestion.

The diagram above shows the restriction digestion analysis for the insertion of TFs by *Sal*I (the red line in the constructs) into pCAGG vectors (A-1, B-1 and C-1), with fragment sizes of 6.407 and 0.939 kb, 6,407 and 1,215 kb and 6.407 and 1.929 kb for pCAGG-MASH1, pCAGG-DLX2 and pCAGG-DLX2/MASH1 constructs, respectively (A-2, B-2 and C-2). Meanwhile, the insertions of GSX2 via *Bgl*II into the pCAGG-DLX2 and pCAGG/DLX2/MASH1 vectors were performed, resulting in fragment sizes of 7.622 and 0.918 kb and 8.336 and 0.918 kB, respectively (D-1, D-2, E-1 and E-2). The orientations of the insertions were then analysed. For pCAGG-DLX2/GSX2 and pCAGG-DLX2/MASH1/GSX2, the orientations of GSX2 were analysed by double digestion with *Ahd*I and *Nde*I, which yielded fragment sizes of 6.084 and 2.456 kb and 6.084 and 3.170 kb, respectively (D-3 and E-3). In addition, for the pCAGG-MASH1 construct, the double digestion analysis for the orientation was obtained with the restriction enzymes *Nde*I and *Nhe*I, which yielded fragment sizes of 4.940 and 2.406 kb (A-3). Additionally, for pCAGG-DLX2, the restriction enzymes *Bstx*I and *Bgl*II were used for the orientation analysis, and yielded fragment sizes of 5.901 and 1.721 kb (B-3). Lastly, for pCAGG-DLX2/MASH1, the digestion analysis for the orientation was performed using the restriction enzyme *Xmn*I, which yielded fragment sizes of 7.474 and 0.862 kb (C-3).

Abbreviations: ATG: Start codon; X: Stop codon.

4.5 TF expression vectors validation through transient nucleofection of HEK293 cells

To monitor the transfection efficiency, GFP reporter expression in transfected HEK293 cells was observed and measured. Western blotting was then used to assess the appropriate expression, as well as self-cleavage, for each TF.

Forty eight hours after nucleofection, GFP expression was observed, and the efficiency of the transfection calculated using ImageJ software, which ranged from 80–90% (using number of GFP^{+ve} cells as a percentage of total number of cells in bright field) **(Figure 4.5)**. In addition, lysates were prepared from harvested cells and the cellular proteins were analysed by western blotting. The nitrocellulose membranes were probed with antibodies specific for each TF. The correct size of each protein, DLX2, MASH1 and GSX2, was observed, 36, 35 and 34 kDa, respectively **(Figure 4.6)**. β-actin was used to confirm the level of protein in each sample and to normalize the expression levels of the desired TFs.

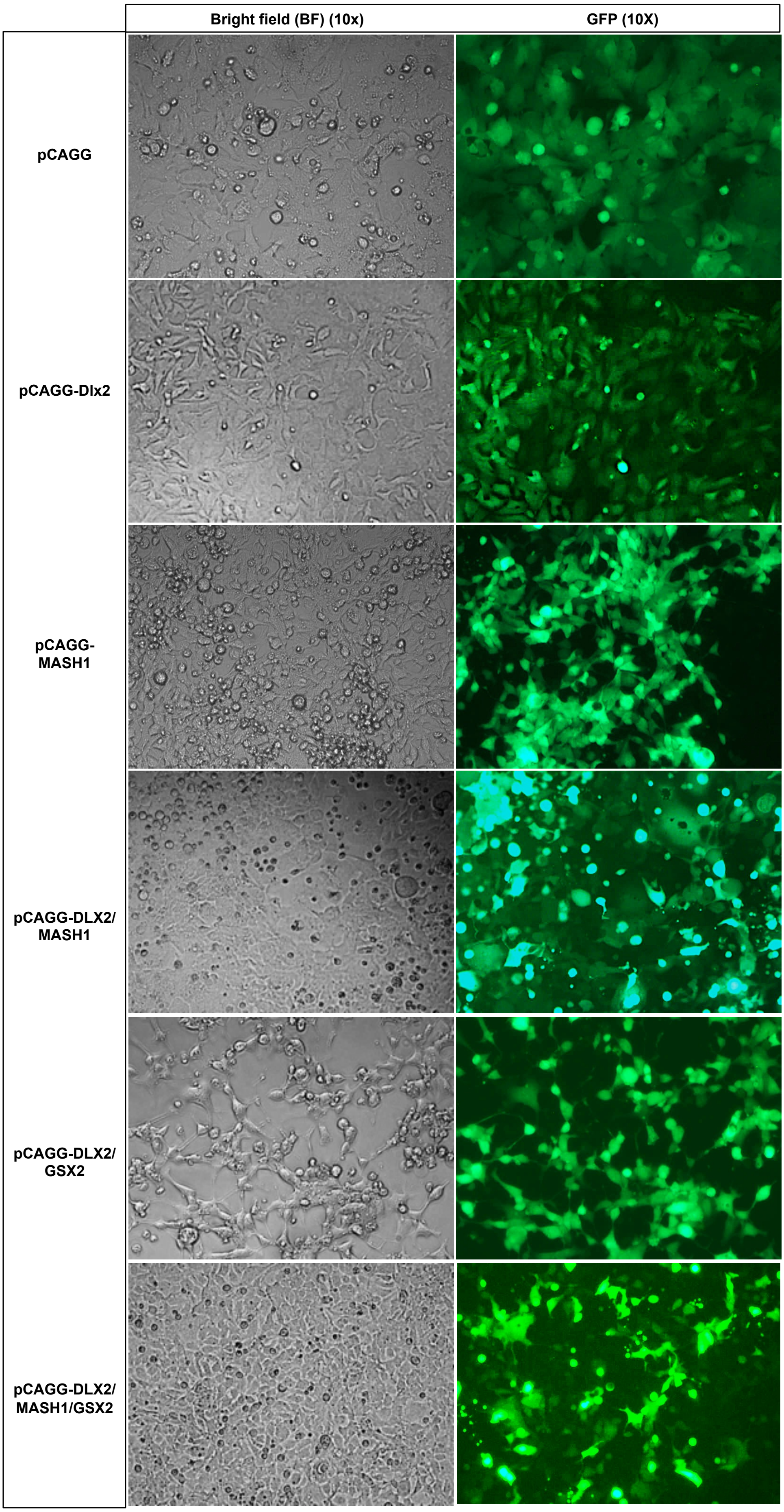

Bright field (BF) (10x)
GFP (10X)
pCAGG
pCAGG-Dlx2
pCAGG-MASH1
pCAGG-DLX2/MASH1
pCAGG-DLX2/GSX2
pCAGG-DLX2/MASH1/GSX2

Figure 4.5: Transiently transfected HEK293 cells with the four cloned -polycistronic expression vectors, pCAGG-DLX2, pCAGG-DLX2/MASH1, pCAGG-DLX2/GSX2, and pCAGG-DLX2/MASH1/GSX2, plus the control, which is the empty vector pCAGG.

The five cloning expression vectors were nucleofected into HEK293 cells. After 48 h, GFP expression for each vector, as shown in the above graph, was calculated using ImageJ software. The efficiency of the nucleofection was approximately 80–90%.

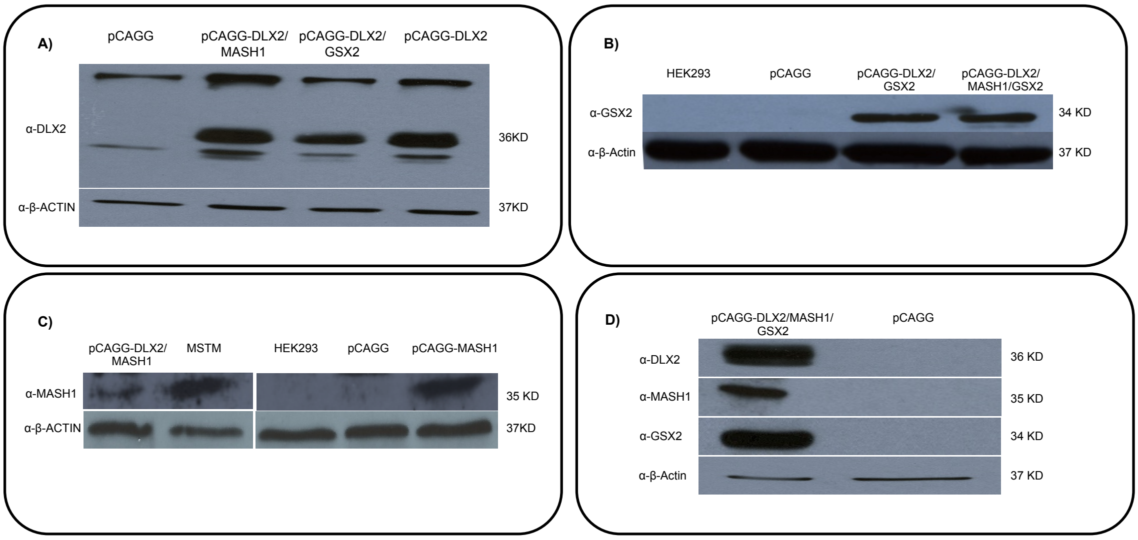


Figure 4.6: Western blotting of the cloned TFs (DLX2, MASH1 and GSX2) expressed by the pCAGG vector

The transfected HEK293 cells with the five expression vectors were lysed for western blotting analysis. HEK293 cells transiently transfected with the expression vector pCAGG-DLX2/MASHI/GSX2 were harvested and analysed with anti-DLX2, anti-MASHI, anti-GSX2 and anti-β-actin antibodies (D); HEK293 cells transfected with the expression vectors pCAGG-DLX2, pCAGG-DLX2/MASHI and pCAGG-DLX2/GSX2 were lysed and analysed with anti-DLX2 and β-actin antibodies (A). GSX2 expression for the pCAGG-DLX2/GSX2 and pCAGG-DLX2/MASH1/GSX2 vectors was obtained (B). In addition, MASH1 expression by the constructs pCAGG-DLX2/MASH1 and pCAGG-MASH1 was analysed (C).

Abbreviations: MSTM: Mouse striatum; HEK293: Human embryonic kidney 293 cells.

4.6 Conclusion

The role of the DLX2, MASH1 and GSX2 TFs in LGE specification

The TFs DLX2, MASH1 and GSX2 are required for the specification of LGE progenitor cells. DLX2 and MASH1 are expressed in the ganglia regions of the VZ and SVZ of LGE and MGE, while GSX2 is strongly expressed in the dorsal LGE and is weakly expressed in the ventral LGE and MGE. MASH1 regulates DLX2, and together they regulate the differentiation of GABAergic neurons, as well as neurogenesis, by activating the Notch signaling pathway. MASH1 regulates neurogenesis in telencephalic development, as well as the patterning and specification of LGE, while MASH1 and GSX2 play a role in striatum development. When LGE progenitor cells are produced, MASH1, DLX2 and GSX2 are required for these precursors to develop to striatal complex (Anderson *et al*. 1997; Casarosa *et al*. 1999; Toresson and Campbell 2001). Because of the interaction of these genes and their essential functions in LGE development, they were selected as candidates to drive the differentiation of hESC–naïve rosette-stage neural progenitors (nrNPCs) towards an LGE fate via their ectopic expression (Chapter 5). This chapter described the successful cloning of these genes into a novel vector designed for high-level, transient, TF co-expression **(Table 4.2)**.

Table 4.2: Summary of completed constructs

TFs combination	Completed constructs in p3X-2A vector	Completed constructs in pCAGG vector
DLX2	p3X-2A-DLX2	pCAGG-DLX2
MASH1	p3X-2A–MASH1	pCAGG–MASH1
DLX2/MASH1	p3X-2A-DLX2/MASH1	pCAGG-DLX2/MASH1
DLX2/GSX2	--------------------	pCAGG-DLX2/GSX2
DLX2/MASH1/GSX2	--------------------	pCAGG-DLX2/MASH1/GSX2

Both the MASH1 and DLX2 TF genes were first inserted into the polyprotein vector p3X-2A and then subcloned into the expression vector pCAGG. However, the TF GSX2 gene was not inserted into p3X-2A. Instead, it was inserted into the expression vector pCAGG, after subcloning the other TFs into the pCAGG vector.

Using alternatives to the 2A peptide strategy

There are alternatives to express more than one gene in one vector, such as using an internal ribosome entry site (IRES) between TFs or different promoters located upstream of a gene ORF (Radcliffe and Mitrophanous 2004; J. H. Kim *et al*. 2011). However, there are some disadvantages of using these approaches, such as their sizes and their expression efficiency (Radcliffe and Mitrophanous 2004; J. H. Kim *et al*. 2011). When using an IRES between genes, the expression of a gene located after the IRES is lower than that of a gene located before the IRES, and IRES sizes are more than 500 nucleotides in length (J. H. Kim *et al*. 2011). Additional promoters can result in different amounts of encoded proteins (Radcliffe and Mitrophanous 2004). Therefore, in this study, self-cleaving 2A peptide sequences between the TFs were used, as it has a high cleavage efficiency among TFs, and their sizes are small (J. H. Kim *et al*. 2011). However, it was stated that the use of the 2A peptide is not common in biomedical studies, as it is not yet established; 2A peptide of the four, E2A, P2A, T2A and F2A, has the highest cleavage productivity, and there is not any commercially available 2A technology for the expression of more than one gene in a single vector (J. H. Kim *et al*. 2011).

TFs validation by transient transfection of HEK293

The high efficiency of the HEK293 nucleofection with the expression vectors containing the four TFs and the control vectors was useful for validating the vector function prior to using the vectors in neural stem cells.

Significantly, all of the cloned genes, together with the GFP reporter, were expressed. Secondly, the western blotting analysis indicated that a polyprotein of the predicted size, ~132 kDa (GSX2 (34 kDa), DLX2 (36 kDa) and MASH1 (36 kDa)) was generated. This indicates that the 2A peptides within the construct were successfully cleaved during translation. Therefore, the expression vectors with the TFs and the control vectors were nucleofected into hPSCs at day 18, followed by the characterisation of TF expression and differentiation into MSNs. These experiments will be described in the next chapter.

Chapter 5: Characterisation of DLX2, MASH1 and GSX2 expression in nucleofected 34D6 and H9 cells.

5.1 Introduction

The TFs, DLX2, MASH1 and GSX2, showed appropriate transgene expression in HEK293 cells (Chapter 4). In this study, ectopic expression of TFs, i.e. DLX2, MASH1 and GSX2 in hESC or iPSC-derived naïve rosette stage neural progenitors cells (nrNPCs), was conducted to investigate whether the hESC or iPSC derived nrNPCs could differentiate into LGE-specific cell type, and subsequently into MSNs. To further validate their potential use for differentiation in hESC- or iPSC-derived nrNPCs, expression of the TFs in the nrNPCs of interest was examined in this chapter. Transgene expression of DLX2, MASH1, and GSX2, as well as 2A self-cleavage peptide were confirmed by immunocytochemistry (ICC). In addition, the expression of exogenous and endogenous TFs was also confirmed by qRT-PCR.

34D6 and H9 nrNPCs (at platting down day (PdD) 18) used for nucleofection were FOXG1 positive. FOXG1 plays a role in the development of the telencephalon, as well as in the expansion of the forebrain, as it promotes the proliferation of progenitor cells and suppresses the differentiation of those cells during neurogenesis (Regad *et al.* 2007; Hanashima *et al.* 2002). It also plays a major role in regulating the timing of neurogenesis in the telencephalon (Hanashima *et al.* 2002). Interestingly, its function is important because of its restricted expression pattern. Its expression is nuclear in progenitor cells, but cytoplasmic in differentiating cells (Regad *et al.* 2007). Hence, investigating FOXG1 expression in nrNPCs is important for this study.

Neurons are generated from two proliferative populations, which are in the VZ and SVZ. In these regions, cell proliferation continues throughout life. In these two populations, the neuron progenitors are in the cell cycle; once they exit the cell cycle, they differentiate and migrate from to the periphery of the telencephalic vesicles and complete their differentiation (Casarosa *et al.* 1999). For example, striatal progenitors or GABAergic interneurons that are generated in the LGE proliferative VZ and SVZ migrate tangentially to the cortical intermediate zone (IZ) and marginal zone (MZ) (Anderson 1997b; Tamamaki *et al*. 1997; Casarosa *et al*. 1999; Dehay and Kennedy 2007). A similar concept also applies to stem cells. For example, in early development, embryonic stem cells are characterised by rapid proliferation and the production of daughter cells, which either remain stem cells or differentiate into a specific cell-type (Takahashi and Yamanaka 2006). As the cells undergo differentiation, the rate of cell proliferation decreases, and when fully differentiated, cell proliferation ceases. Therefore, proliferation and differentiation are regulated by a balance of intrinsic and extrinsic cues that direct progenitors to enter the cell cycle and proliferate, or exit the cell cycle and begin differentiation (Dehay and Kennedy 2007). It is believed that such a balance is maintained by the parallel, overlapping and/or sequential function of TFs such as Dlx2 and Mash1 (Anderson 1997; Horton *et al*. 1999; Yun *et al*. 2002; Cobos *et al*. 2007; Colombo *et al*. 2007; Poitras *et al*. 2007; Long e*t al.* 2009b).

At the proteomic level, there are many proteins that program the cell cycle upon expression. There are two categories of cell cycle proteins: one group drives the cell cycle, while the other inhibits cell cycle **(Figure 5.1)**. In the developing cortex during neurogenesis, neuroepithelial cells divide in proliferative VZ and express phospho-histone H3, which is an M-phase marker, and cyclin E and Ki67, which are the markers for G1-, S-, G2- and M- phases (Herrup and Yang 2007). Brdu/^{3}H-T is the S-phase marker. The mitotic activity takes place in the SVZ (Herrup and Yang 2007). When cyclin-dependent kinase inhibitors (CDKI), such as CDK5, P27, P21 and p57, are expressed, cells exit the cell cycle, start the process of early differentiation and migrate from the proliferative VZ and SVZ to the IZ, cortical plate and MZ. In the cortical plate, the cells are fully differentiated to mature neurons driven by the presence of CDKI (Herrup and Yang 2007) **(Figure 5.2)**. Therefore, in this chapter, the functional consequences of transient expression of the TFs, DLX2, MASH1, GSX2 and downstream effector genes, on cell proliferation was assessed by looking at the cell cycle protein Ki67.

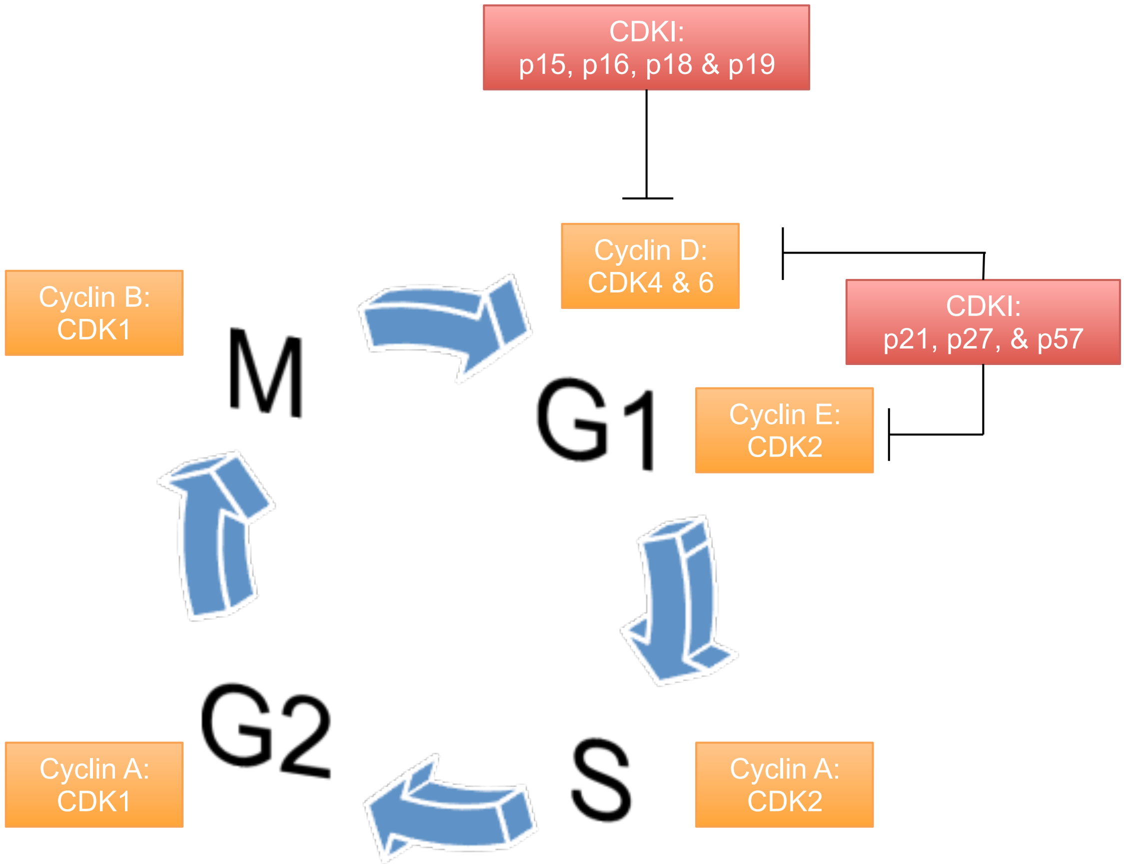


Figure 5.1: The phases of cell cycle and the proteins involved in cell cycle regulation.

The cell cycle phases include: G1 phase, where the cells commit to divide or exit from the cell cycle due to responses to extracellular signals, S phase, where DNA synthesis and process of replication take place, G2 phase, where the completion of DNA replication is checked and M phase, where two daughter cells are generated. Following the M phase, the new daughter cells can re-enter the cell cycle and proliferate or exit the cell cycle and start to differentiate. The cell cycle proteins in orange boxes drive progression through the different stages of the cell cycle while the ones in red inhibit progression of the cell cycle.

Abbreviations: G1: Gap 1 or growth 1, S: DNA synthesis, G2: Gap 2 or growth 2, M: Mitosis, CDK: Cyclin-dependent kinases; CDKI: Cyclin-dependent kinases inhibitor.

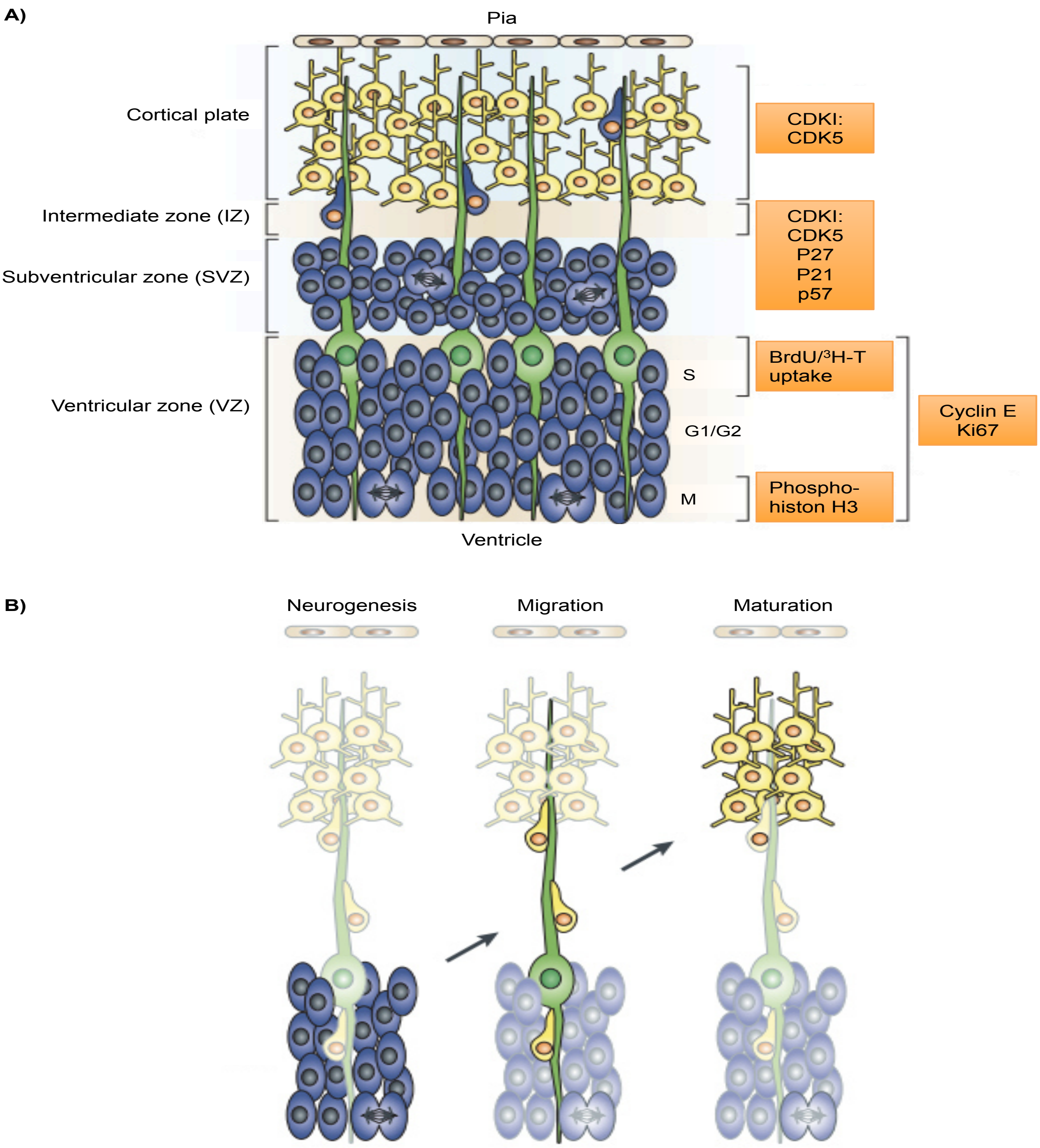


Figure 5.2: Cell cycle proteins involved in neuronal development.

This schematic represents the developing cortex and the cell cycle proteins that are involved in the development of each layer (A). The neuroepithelial cells (shown in blue) divide in the VZ and their mitotic activity continues in the SVZ. The protein phospho-histone H3 is associated with the M phase, BrdU/^{3}H-T uptake is associated with the S phase and Cyclin E and Ki67 are associated with cell cycle phases (G1, S, G2 and M). CDKI expression drives neuronal cell exit from the cell cycle and migration along the radial glia (shown in green) from VZ and SVZ to IZ and cortical plate, where they are fully differentiated (shown in yellow). Cell cycle proteins are associated in the three stages of neuronal development: neurogenesis, migration and maturation stages (B). Figure taken from Herrup and Yang 2007.

Previously, it has been shown that the three TFs - DLX2, MASH1 and GSX2 - play a role in neuronal fate in the subpallium **(Table 2.3)** (Pauly *et al.* 2013; Wang *et al.* 2013). Furthermore, the TFs together play a role in the transcriptional network that drives striatal and MSN development **(Figure 5.3)** (Stühmer *et al.* 2002b; Poitras *et al.* 2007; Colasante *et al.* 2008; Wang *et al.* 2013). GSX2 regulates the expression of MASH1, DLX2 and EBF1 (Wang *et al.* 2013). MASH1 regulates the expression of DLX2 by binding to the I12b enhancer that is located downstream of DLX2 (Poitras *et al.* 2007). DLX2 regulates the expression of ARX, which triggers migration, through binding to the UAS3 enhancer downstream of ARX (Colasante *et al.* 2008). DLX2 also induces the expression of DLX5, and together they regulate the expression of GAD1/2 (Stühmer *et al.* 2002b). These interactions aid in proliferation, differentiation and migration in neurogenesis (Yun *et al*. 2002; Long *et al*. 2009b; Wang *et al*. 2013). In this chapter, the target genes of DLX2, MASH1 and GSX2 are examined at transcriptional level (qRT-PCR) to validate the progress of the stem cell differentiation program. We address the following question: does ectopic expression of the TFs DLX2, MASH1 and GSX2 in 34D6 nrNPCs cause them to differentiate into ventral telencephalic neuronal progenitors and LGE-specific progenitor cells?

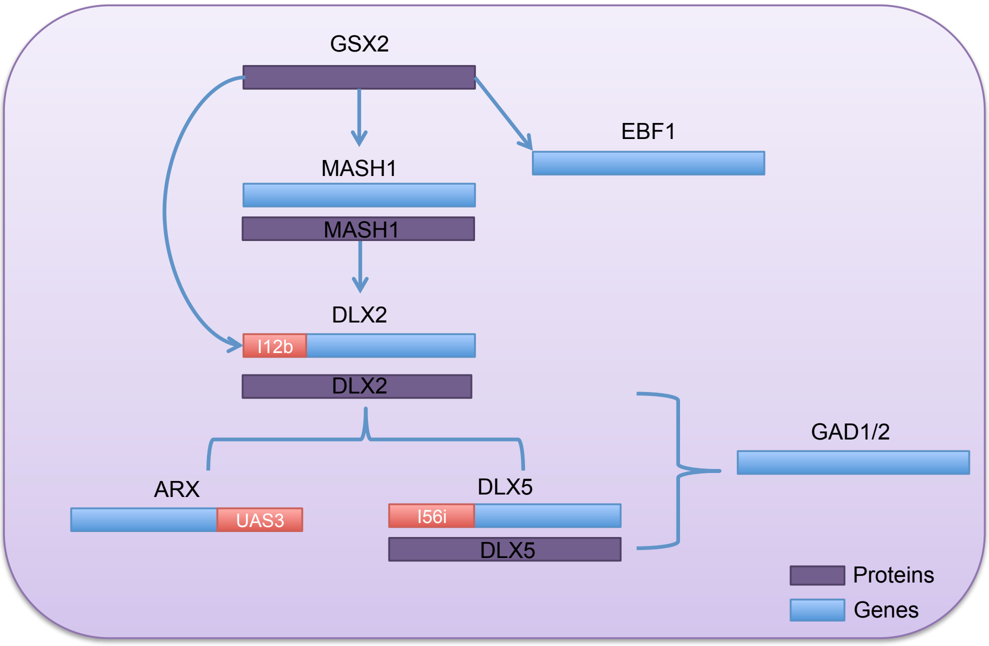


Figure 5.3: Transcriptional network of the TFs that play a role in striatal and MSN differentiation.

The GSX2 regulates the expression of EBF1, MASH1 and DLX2. MASH1 regulates the expression of DLX2 by binding to the I12b enhancer. In addition, DLX2 regulates the expression of ARX by binding to the UAS3 enhancer, and DLX5 by binding to I56i enhancer. Together, DLX2 and DLX5 regulate the expression of GAD1/2.

5.2 Aims

The main aim of this study was to determine whether the protocol of ectopic expression for the different combinations of TFs (DLX2, GSX2 and MASH1) drives differentiation of hPSC-derived nrNPCs into LGE-like specific progenitors cells. For this purpose, the strategy was initially optimized for the expression of TFs using GFP as a marker that was fused with the TFs using a small molecule inhibitor (G418) to eliminate those cells not expressing GFP in the 34D6 cell line model. Then, the nucleofected TFs of various combinations in H9 and 34D6 cell line models were assessed for their expression profiles at the molecular level. Finally, to test the functionality of the ectopically expressed TFs, nucleofected nrNPC FOXG1^{+ve} progenitor cells were assessed for their ability to re-programme into LGE-like neuronal progenitors using PAX6, EMX2 (dorsal marker) and NKX2.1 (ventral-specific marker for MGE).

5.3 Experimental design

The experimental design used, in this chapter, for cell maintenance is illustrated in **Figure 5.4**. The detailed methodology has also been described in the Materials and Methods section **(section 3.2)** in Chapter 3.

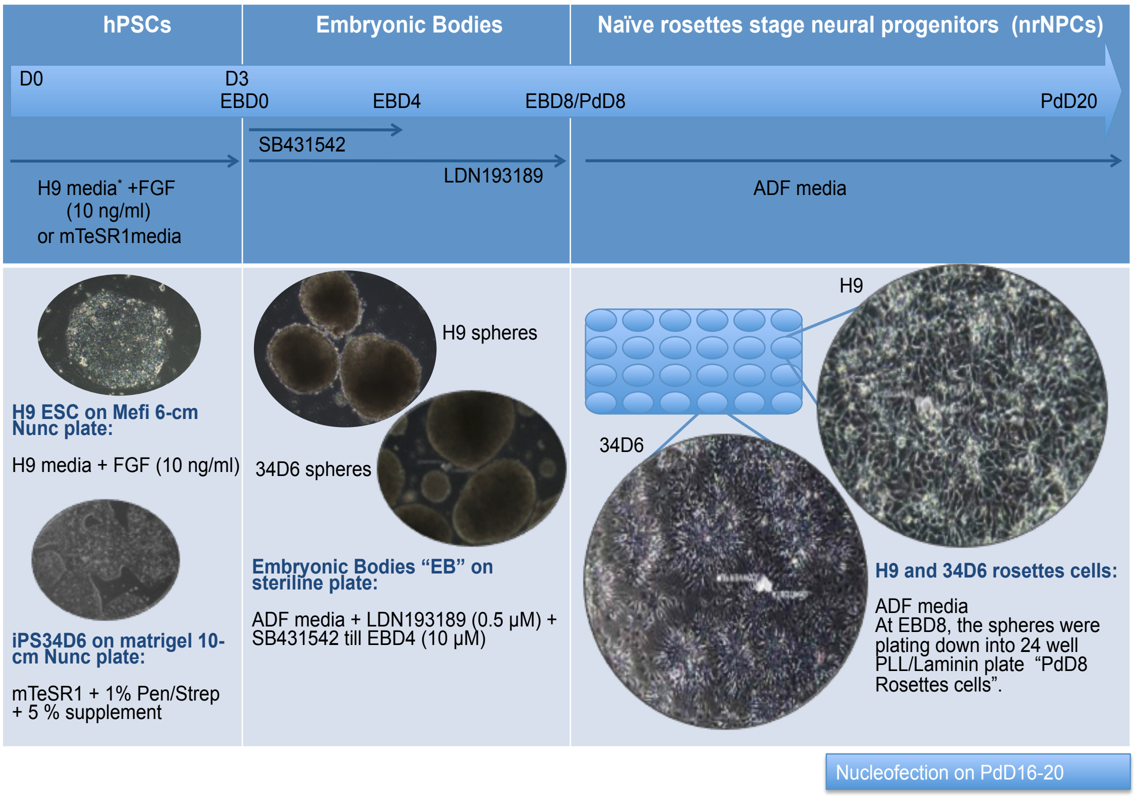


Figure 5.4: Experimental design.

The experimental protocol used in this chapter for cell maintenance from the human pluripotent stem cells (hPSCs) stage (D0) to naïve rosette stage neural progenitors (nrNPCs) at PD 20.

Abbreviations: D: Day, nrNPC: Naïve rosette stage neural progenitors, EB: Embryonic body, PdD8: Platting down Day 8, Pen/Strep: Penicillin (100 U/ml)/Streptomycin (100 μg/ml); PLL: poly-L-Lysine.

5.4 Results

5.4.1 Characterisation of TF vector expression in transiently nucleofected H9 and 34D6 cells

5.4.1.1 Quality control prior to nucleofection: nrNPCs at day 18 were positive for FOXG1, human ZO.1 and NESTIN [multipotent neural stem cells (NSCs)], and were negative for OCT4 (pluripotency marker).

Quality control prior to nucleofection of hPSC-derived nrNPCs at platting down day 18 (PdD18) was performed for the protein and transcriptome levels using ICC and qRT-PCR. Prior to nucleofection of hPSCs-derived nrNPCs at PdD18, nrNPC quality was determined by assessing the FOXG1 and human ZO.1 expression by ICC. These cells stained positively for forebrain marker (FOXG1^{+ve}) and tight junctions (ZO.1^{+ve}) **(Figure 5.5 A and B).** The neural precursors of H9 and 34D6 were examined by qRT-PCR for expression of NESTIN, a multipotent neural stem cell (NSC) marker, and OCT4, pluripotency marker. The nrNPCs were NESTIN^{+ve} and OCT4^{-ve} at PdD12; however at D0, the undifferentiated 34D6 and H9 cells were NESTIN^{-ve} and OCT4^{+ve} **(Figure 5.5 C)**. The negative control for FOXG1 staining is shown in **Appendix 5.1.** FOXG1 expression was clearly evident, suggesting that these cells were forebrain neuron precursors and hence could be used to differentiate into LGE-like progenitor cells through ectopic expression of the TFs MASH1, DLX2 and GSX2.

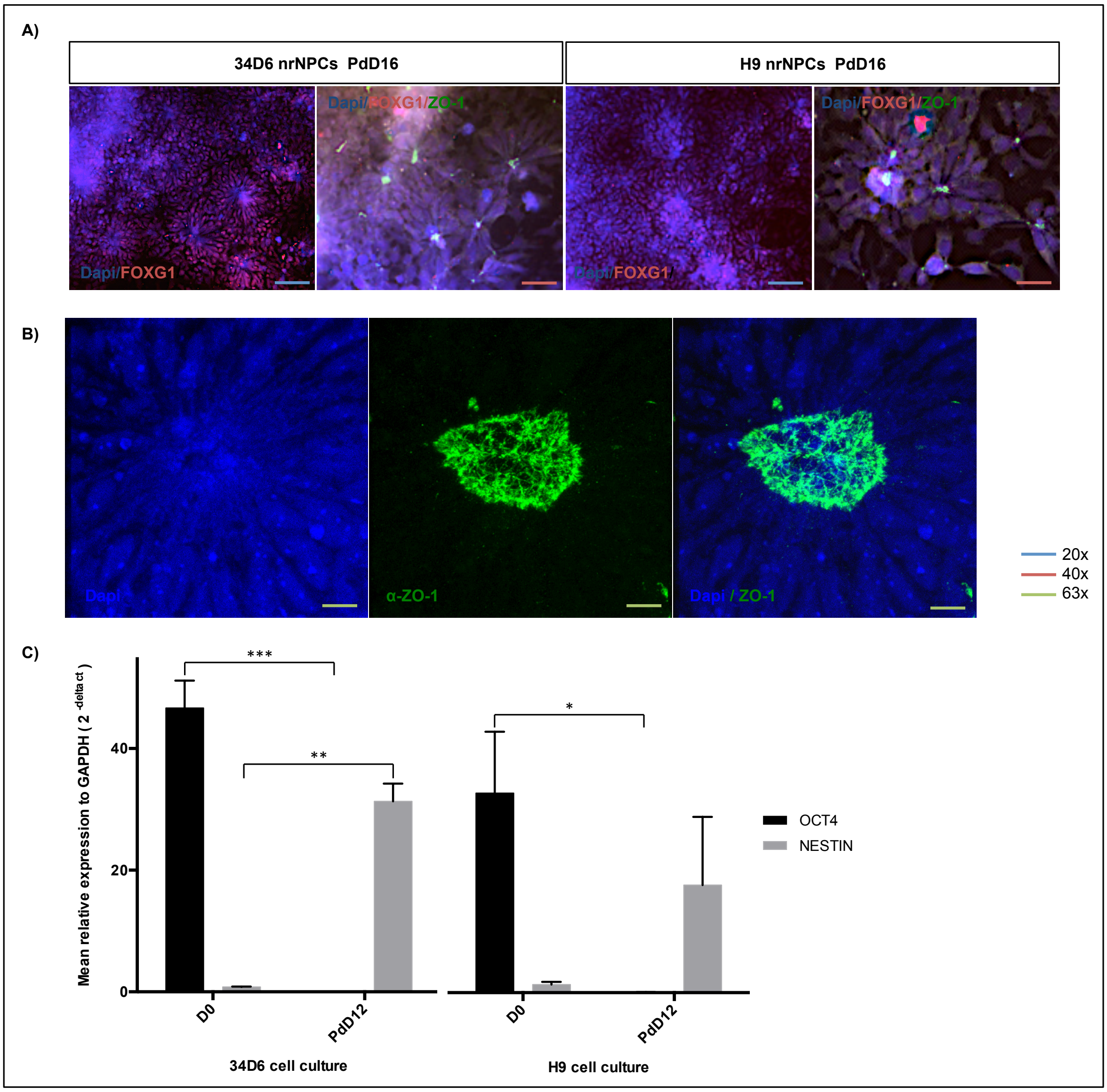


Figure 5.5: Quality control prior to the nucleofection of nrNPCS using ICC with FOXG1 and human ZO.1 antibodies, and qRT-PCR expression analysis of the pluripotency marker, OCT4, and neural marker, NESTIN.

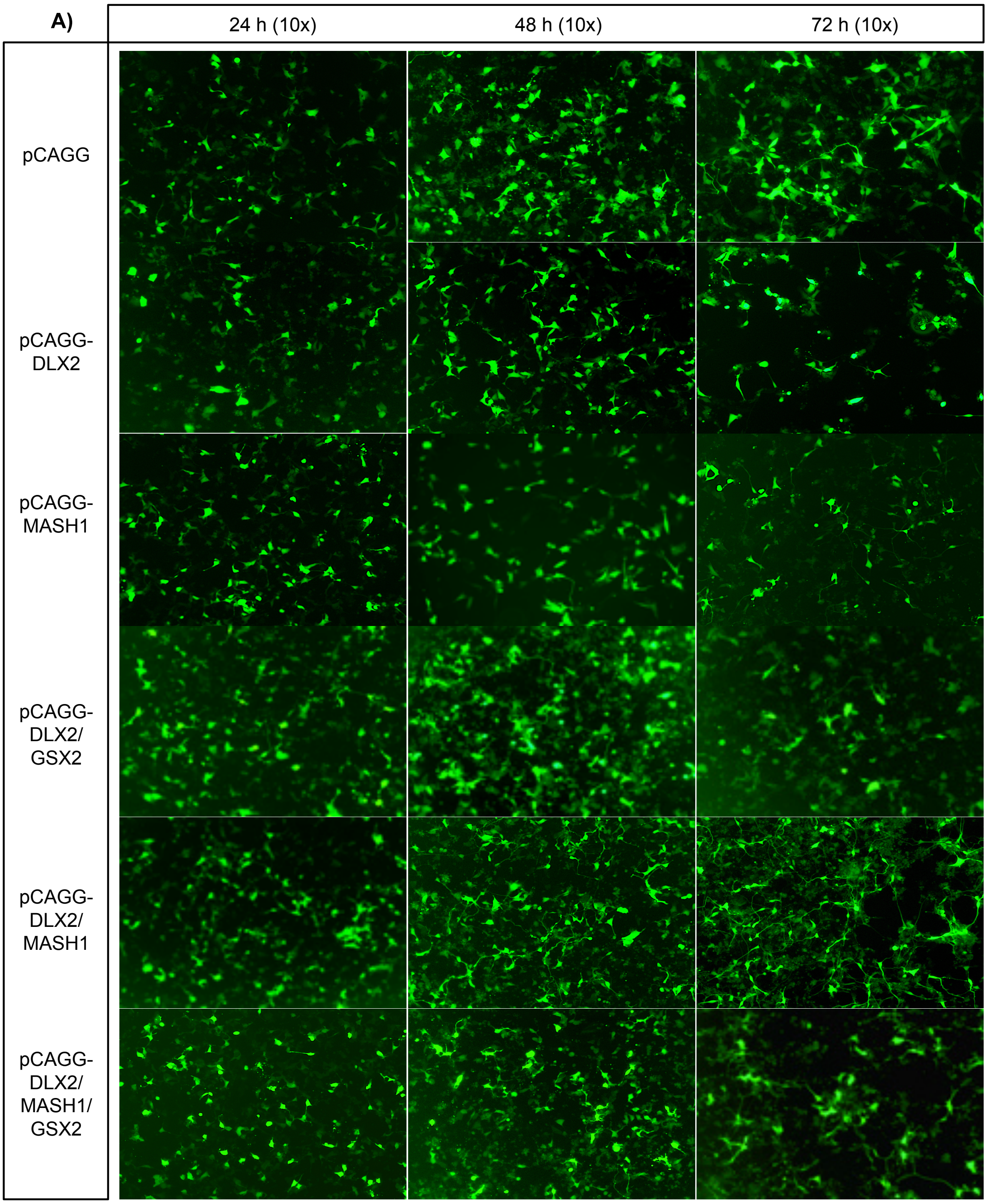
Before nucleofection, some of the H9 and 34D6-derived nrNPCs at PdD16 were fixed and stained for forebrain marker, i.e. FOXG1 (Primary dilution: 1:1000, NeuraCell) with an Alexa Fluor® 594 labelled anti-rabbit IgG secondary antibody (red), Abcam (A). The cells were also stained for tight junction marker, i.e. human ZO-1 (Primary dilution: 1:250, BD Biosciences) with an Alexa Fluor® 488 labelled anti-mouse IgG1 secondary antibody (green), Abcam (A). The confocal microscope images of human ZO-1 (green) and Dapi/nuclear (blue) staining (Hoechst) were obtained (B). qRT-PCR of OCT4 and NESTIN at D0 and PDd12 for H9 and 34D6 cell culture (C).

The blue scale bar indicates 100 µm, the red scale bar indicates 50 µm, and the green scale bar indicates 36 µm.

5.4.1.2 The efficiency of TF expression was approximately 45% higher following acute G418 selection at 48h post-nucleofection, as compared to non-selected cells.

The PdD18 nrNSCs were nucleofected with either the empty vector (pCAGG), or one of the pCAGG-DLX2, pCAGG-DLX2/MASH1, pCAGG-DLX2/GSX2 and pCAGG-DLX2/MASH1/GSX2 vectors. The efficacy of nucleofection was determined by GFP expression, using ImageJ software, after 24, 48, 72 and 96 hours (h) post-nucleofection **(Figure 5.6 A and B)**. GFP was highly expressed from all the vectors at 48 h post-nucleofection; however, the percentage of GFP expressing cells had declined at 72 and 96 h time points **(Figure 5.6 B)**. This decline could be due to transient expression from a non-integrated plasmid. Since DLX2, MASH1 and GSX2 were only required in a restricted window of time during the differentiation of cells from progenitor through to mature neuron, a transient G418 selection strategy was used to select against non-transfected cells from the mixed population. To achieve this, different concentrations of G418 (200, 400, 600 and 800 µg/ml) with different incubation times (1 day to 1 week) were tested to find the most suitable concentration of G418 and duration time to select GFP^{+ve} cells **(Figure 5.7)**.

At a low concentration of G418 (200 µg/ml), GFP expression was low (15%); this was increased (40%) at a higher concentration of G418 (400 µg/ml) at week 1. At the 800 µg/ml concentration of G418, an increase in GFP^{+ve} cells (45%) was only seen at day one; surprisingly the cells did not show similarly high GFP expression at week 1, suggesting that G418 causes toxicity at higher doses with prolonged incubation.


A)
24 h (10x)
48 h (10x)
72 h (10x)
pCAGG
pCAGG-DLX2
pCAGG-MASH1
pCAGG-DLX2/GSX2
pCAGG-DLX2/MASH1
pCAGG-DLX2/MASH1/GSX2

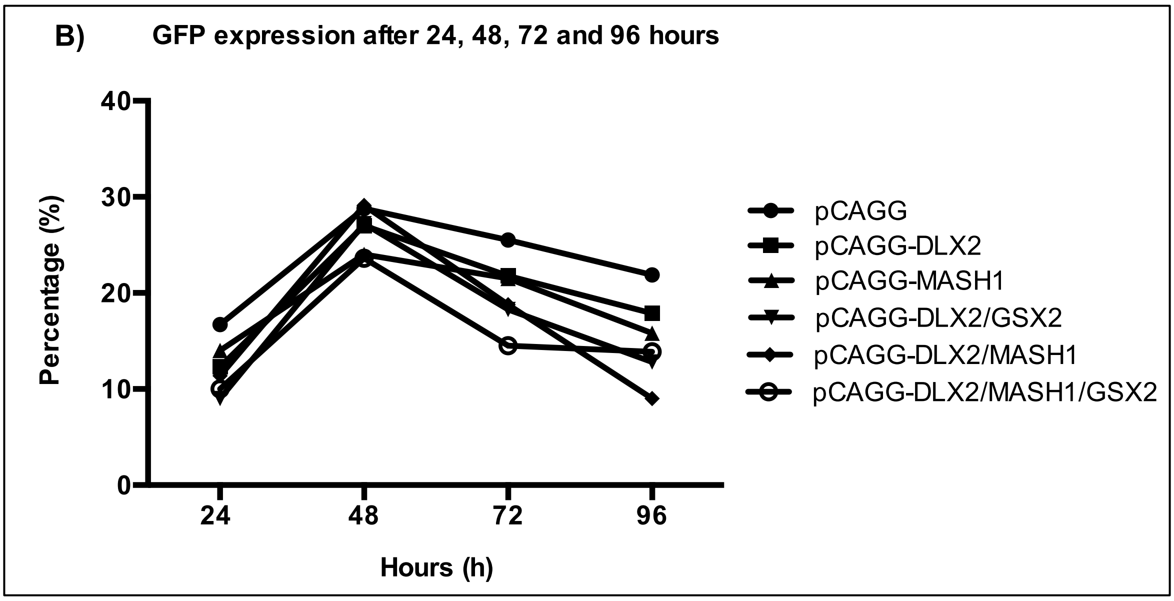


Figure 5.6: GFP expression in nucleofected PdD18 nrNSCs.

GFP expression in nrNSCs 24, 48 and 72 h after nucleofection with the vectors shown (left column) (A). The percentage of GFP expressing cells after each nucleofection was determined by ImageJ (B).

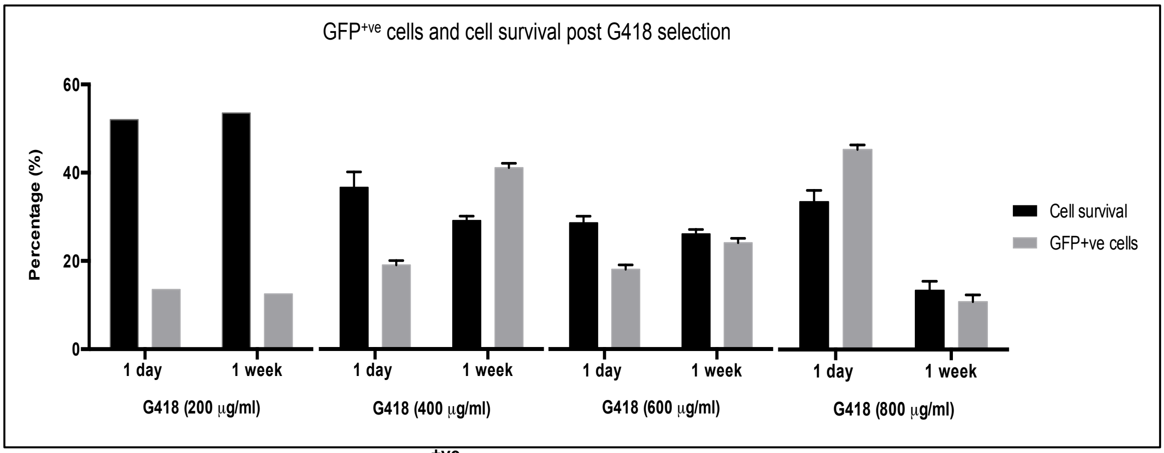


Figure 5.7: Percentage of GFP^{+ve} cells and cell survival post G418 selection of nucleofected cells, with different concentrations and incubation times

Twenty four hours post-nucleofection of nrNPCs, the cells were incubated for 1 day to 1 week in different concentration of G418 (200, 400, 600 or 800 μg/ml). At low concentrations of G418, the percentage of GFP^{+ve} cells was low. This percentage increased at 400 μg/ml G418 concentration with an incubation time of 1 week. In comparison with 400 μg/ml of G418, GFP^{+ve} cells were increased at 800 μg/ml of G418 on day one, exhibiting a cell survival of around 35% and a nucleofection efficiency of 45% (N = 2).

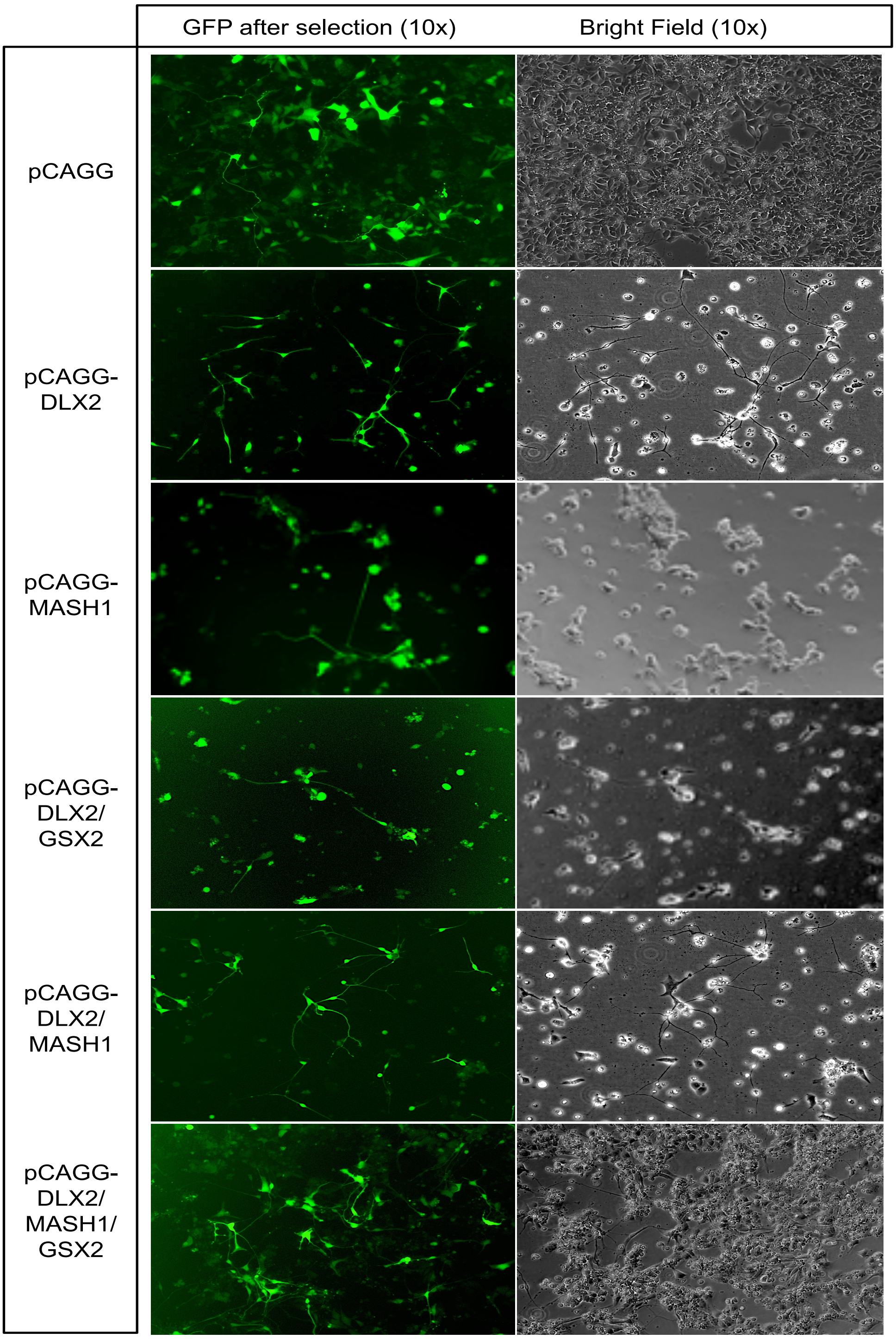


Figure 5.8: GFP expression in nucleofected nrNSCs after G418 selection.

G418 sulfate selection was started 48 h after nucleofection with the vectors. GFP expression and bright field of each group of nucleofected cells after G418 selection.

5.4.1.3 Successful expression of DLX2, MASH1, GSX2 and self-cleavage peptides 2A into H9 and 34D6 cells from all six expression vectors at nucleofection day 4 (ND4).

At PdD18 of nrNPCs, nucleofection with the six expression vectors (pCAGG-DLX2, pCAGG-MASH1, pCAGG-DLX2/GSX2, pCAGG-DLX2/MASH1 and pCAGG-DLX2/MASH1/GSX2) and the control (pCAGG) was performed. Following acute selection with G418, TF transgene expression in selected cells was assessed by ICC. At ND4, the nucleofected H9 cells (not shown) and 34D cells were fixed for ICC to determine the expression of DLX2, MASH1, GSX2 genes and the 2A self-cleavage peptide tag. Self-cleavage of the 2A peptide allows the release of cloned TFs in nucleofected cells (Donnelly *et al.* 2001). The negative control for MASH1, DLX2, GSX2 and GFP staining is shown in **Appendix 5.1.**

All nrNPCs nucleofected with the TFs expressing pCAGG-DLX2, pCAGG-DLX2/GSX2, pCAGG-DLX2/MASH1 and pCAGG-DLX2/MASH1/GSX2 were immunopositive for the 2A peptide and for DLX2, whilst cells transfected with the empty vector showed no expression for 2A peptide and DLX2 **(Figures 5.9 and 5.10)**. Similarly, MASH1 and GSX2 expression was only evident in cells transfected with the MASH1 and/or GSX2 containing vectors (pCAGG-DLX2/GSX2, pCAGG-DLX2/MASH1, and pCAGG-DLX2/MASH1/GSX2) **(Figures 5.11 and 5.12)**.

H9 and 34D6 transient nucleofected nrNPCs at PdD18 had the same characteristics, such as, the morphology and the viability of the cells. In addition, TF characterisation yielded equivalent results in both H9 and 34D6 nrNPCS PdD18 cell lines, therefore, any one of the cell lines could be carried forward for further analysis. The 34D6 was chosen for further experiments as the differentiation protocol can be used with HD-iPS cell lines for disease modeling studies. These results clearly demonstrate successful expression of all cloned TFs in 34D6-derived nrNPC cells at the protein level using ICC.

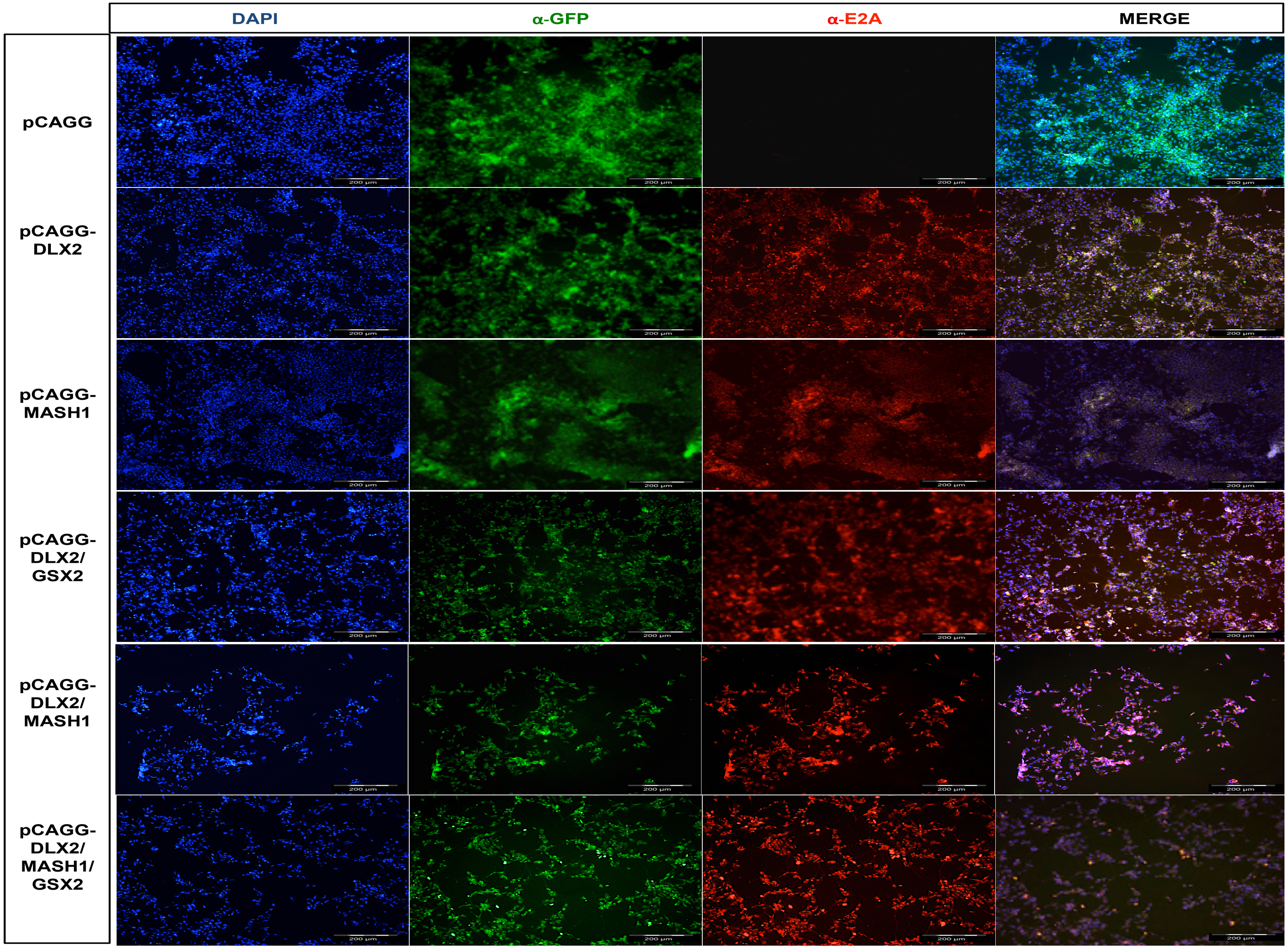

DAPI
α-GFP
α-E2A
MERGE
pCAGG
pCAGG-DLX2
pCAGG-MASH1
pCAGG-DLX2/GSX2
pCAGG-DLX2/MASH1
pCAGG-DLX2/MASH1/GSX2
200 µm

Figure 5.9: 2A peptide expression in 34D6 nrNPCs four days after nucleofection with the TF expressing vectors.

The six expression vectors (left column) were nucleofected into 34D6-derived neural progenitors. Four days later, the cells were fixed and analysed by ICC with anti-E2A, which is an antibody against 2A peptide (primary dilution: 1:500, Millipore), the cells were stained with Hoechst (nuclear staining) and GFP expression was also analysed (primary dilution: 1:4,000, Abcam).

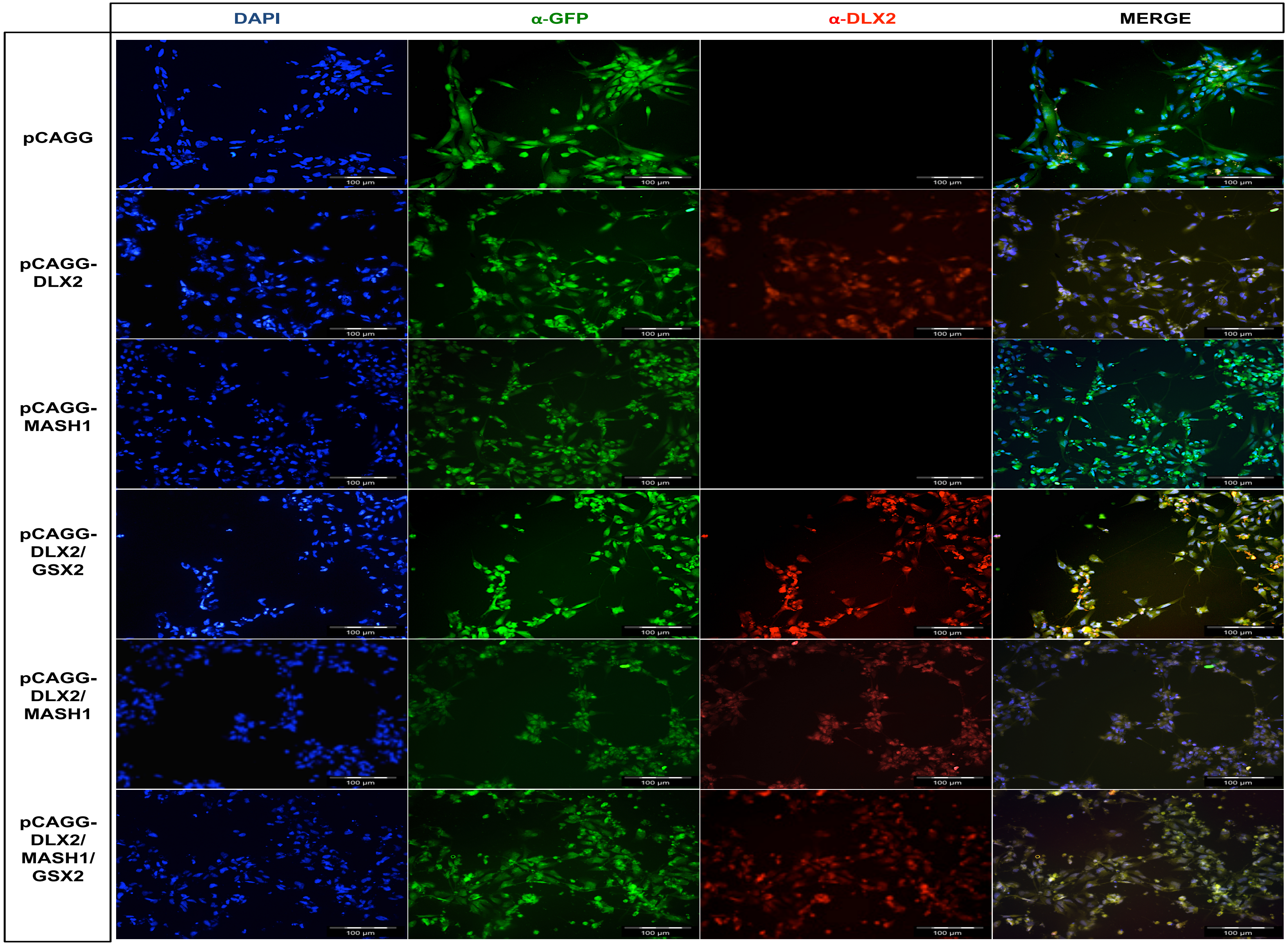

DAPI
α-GFP
α-DLX2
MERGE
pCAGG
pCAGG-DLX2
pCAGG-MASH1
pCAGG-DLX2/GSX2
pCAGG-DLX2/MASH1
pCAGG-DLX2/MASH1/GSX2
100 µm

Figure 5.10: DLX2 transgene expression in 34D6 nrNPCs four days after nucleofection with the TF expressing vectors.

The six expression vectors (left column) were nucleofected into 34D6-derived neural progenitors. Four days later, the cells were fixed and analysed by ICC with anti-DLX2 (primary dilution: 1:800, Abcam), the cells were stained with Hoechst (nuclear staining) and GFP expression was also analysed (primary dilution: 1:4,000, Abcam).

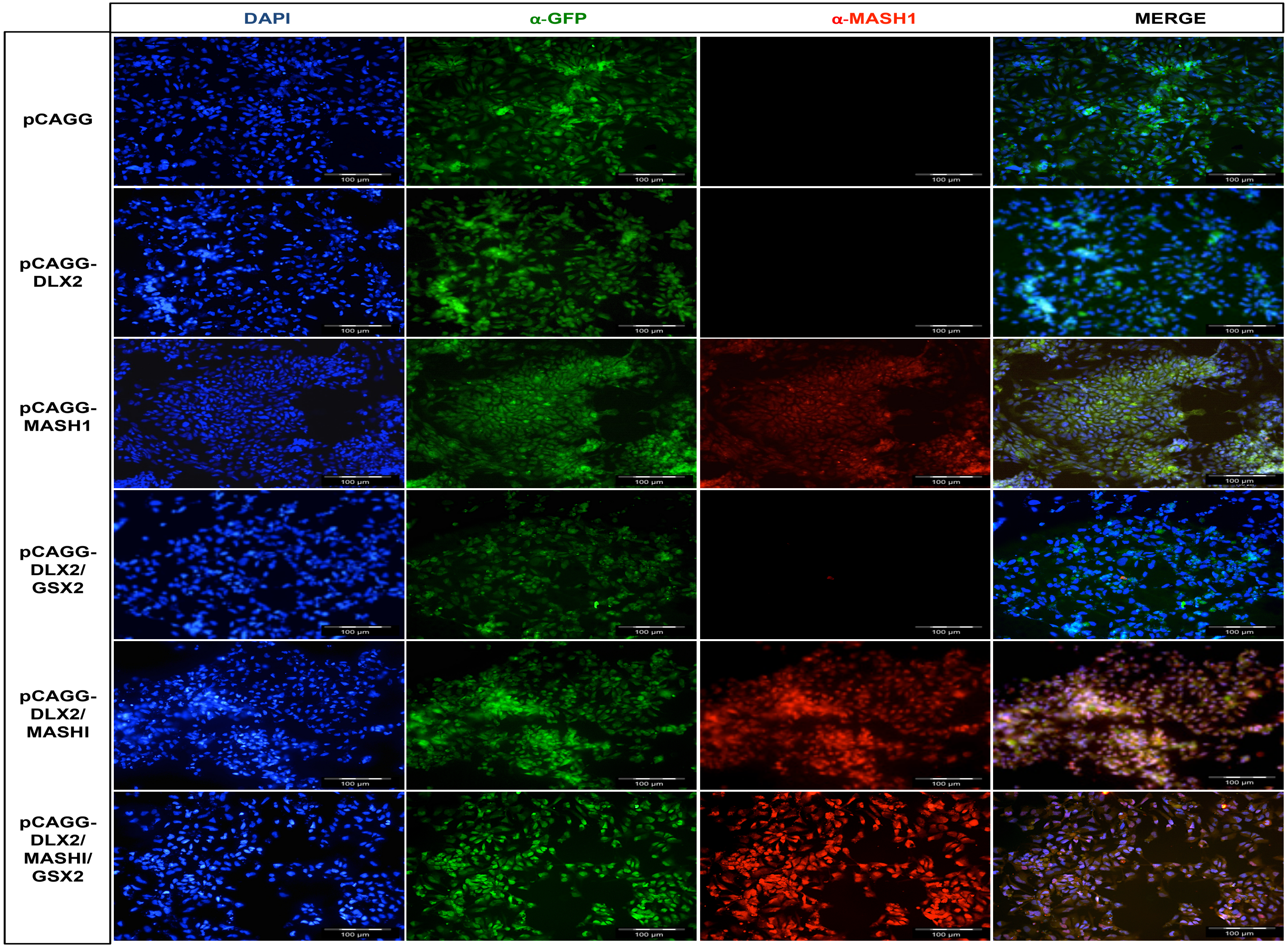

DAPI
α-GFP
α-MASH1
MERGE
pCAGG
pCAGG-DLX2
pCAGG-MASH1
pCAGG-DLX2/GSX2
pCAGG-DLX2/MASHI
pCAGG-DLX2/MASHI/GSX2
100 µm

Figure 5.11: MASH1 transgene expression in 34D6 nrNPCs four days after nucleofection with the TF expressing vectors.

The six expression vectors (left column) were nucleofected into 34D6-derived neural progenitors. Four days later, the cells were fixed and analysed by ICC with anti-MASH1 (primary dilution: 1:500, BD Pharmingen), the cells were stained with Hoechst (nuclear staining) and GFP expression was also analysed (primary dilution: 1:4,000, Abcam).

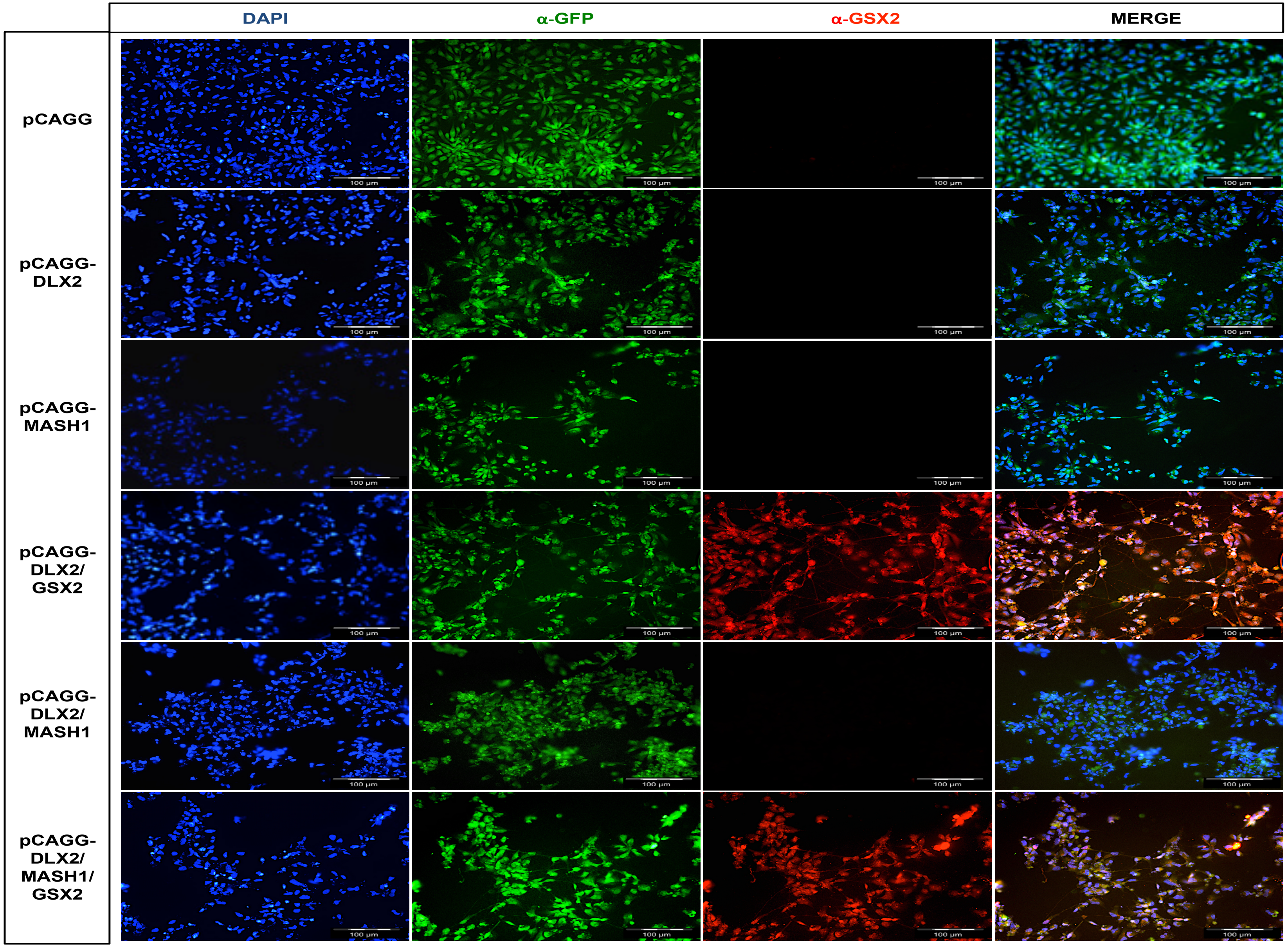

DAPI
α-GFP
α-GSX2
MERGE
pCAGG
pCAGG-DLX2
pCAGG-MASH1
pCAGG-DLX2/GSX2
pCAGG-DLX2/MASH1
pCAGG-DLX2/MASH1/GSX2
100 µm

Figure 5.12: GSX2 transgene expression in 34D6 nrNPCs four days after nucleofection with the TF expressing vectors.

The six expression vectors (left column) were nucleofected into 34D6-derived neural progenitors. Four days later, the cells were fixed and analysed by ICC with anti-GSX2 (primary dilution: 1:500 Millipore), the cells were stained with Hoechst (nuclear staining) and GFP expression was also analysed (primary dilution: 1:4,000, Abcam).

5.4.2 Both endogenous and exogenous expression of MASH1, GSX2 and DLX2 were examined by qRT-PCR in 34D6 nrNPCs using TF expression vectors.

To corroborate the expression of cloned TFs and to further confirm whether the expression of TFs is endogenous or exogenous, qRT-PCR was performed. nrNPCs were nucleofected with the different TF expression vectors and cultured in ADF media. The cells were then harvested at nucleofection day (ND) 0, ND2, ND3, ND5 and ND7. Their RNA was extracted and qRT-PCR was performed to measure the expression of the DLX2, MASH1 and GSX2 transgenes and their endogenous gene counterparts. The primers used to distinguish exogenous and endogenous genes for the expression studies are shown in the Materials and Methods section **(section 3.4.3)** in Chapter 3 and the characterisation of the primers is shown in **Figure 5.13.**

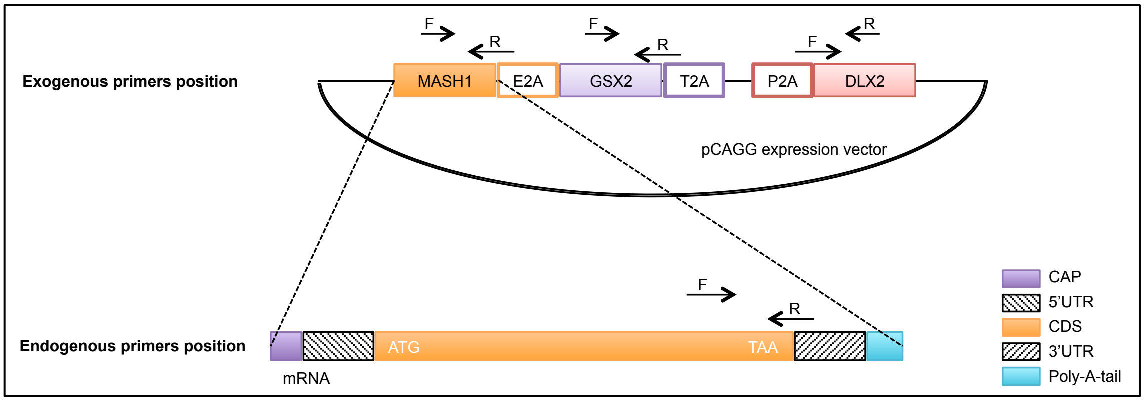


Figure 5.13: Schematic showing primers used for qRT-PCR to distinguish between exogenous and endogenous TF expression.

The position of the exogenous and endogenous primers (F or R) for the TFs DLX2, MASH1 and GSX2 are shown. The exogenous primers contain a few base pairs of the 2A peptides and therefore will only detect a transcript derived from the expression vectors. The endogenous primers contain a few base pairs of the 3'UTR in the R primers and therefore will only detect endogenously derived transcripts.

Abbreviations: CAP: RNA 7-methyl-guanosine cap, 5'UTR: 5 prime untranslated region, CDS: Coding sequence, 3'UTR: 3 prime untranslated region, F: forward primer and R: Reverse primer, ATG: Methionine - start codon; TAA: Stop codon.

5.4.2.1 The expression pattern of endogenous MASH1 was altered in a time-dependent manner.

Forty-eight hours after nucleofection, the mean relative expression of exogenous MASH1 peaked in all cells that contained the MASH1 transgene **(Figure 5.14 A, B and C)**. Its expression was considerably higher in the pCAGG-MASH1 nucleofected cells compared to pCAGG-DLX2/MASH1 and pCAGG-DLX2/MASH1/GSX2 nucleofected cells **(Figure 5.14 A, B and C)**. The expression of endogenous MASH1 was increased even after expression of MASH1 transgene declined after ND2 in both pCAGG-MASH1 and pCAGG-DLX2/MASH1/GSX2 nucleofected cells; between ND2 to ND5, it remained constant and, at ND5, it started to increase gradually **(Figure 5.14 A and C)**. Meanwhile, in the pCAGG-DLX2/MASH1 nucleofected cells, the mean relative expression of endogenous MASH1 was low in comparison to endogenous expression of DLX2, beginning to increase slightly after ND5 **(Figure 5.14 B)**. However, in the same nucleofected cells, the expression of endogenous DLX2 reached its maximum level at ND7 relative to the expression of endogenous MASH1 **(Figure 5.14 B and E)**. Therefore, from these results, it was demonstrated that the expression of endogenous MASH1 increases slightly in a time-dependent manner, while DLX2 increases rapidly, indicating the presence of an alternative mode of action between these two TFs in the neuronal differentiation program.

5.4.2.2 The expression of endogenous DLX2 was altered by the co-expression of the TFs MASH1 and GSX2

In the 34D6 nrNPCs nucleofected with four constructs (pCAGG-DLX2, pCAGG-DLX2/MASH1, pCAGG-DLX2/GSX2, and pCAGG-DLX2/MASH1/GSX2), the mean relative expression of exogenous DLX2 transgene peaked after 48 h as compared to control group (pCAGG). The expression of exogenous DLX2 was maximal at 48 h in the pCAGG-DLX2 nucleofected cells **(Figure 5.14 D)**. On the other hand, the mean relative expression of endogenous DLX2 gradually increased in pCAGG-DLX2/MASH1 nucleofected cells from ND3 to ND7 **(Figure 5.14 E)**, but decreased in pCAGG-DLX2/GSX2 **(Figure 5.14 F)**, and pCAGG-DLX2/MASH1/GSX2 nucleofected cells **(Figure 5.14 G)**. In the pCAGG-DLX2/MASH1/GSX2 nucleofected cells, the endogenous expression started to rise slightly at ND7 **(Figure 5.14 G)**. In the pCAGG-DLX2 nucleofected cells, endogenous DLX2 expression was constant from ND5 to ND7.

Interestingly, when the gene DLX2 was co-expressed with MASH1, the expression of endogenous DLX2 increased dramatically (ND3 - ND7) **(Figure 5.14 D and E)**, whereas, the endogenous expression of MASH1 was very low **(Figure 5.14 B and E)**. Furthermore, when DLX2 was co-expressed with MASH1 and Gsx2, the endogenous DLX2 expression was initially low at ND5 and then started to increase slightly at ND7, whereas the expression of endogenous GSX2 was dramatically elevated at ND7. The expression of endogenous MASH1 was also visibly increased at ND7 **(Figure 5.14 C, G and I)**. Together, these results show that expression of endogenous DLX2 gene was transiently affected by ectopic expression of MASH1 and GSX2. The expression of DLX2 was reduced when co-expressed with GSX2, but increased when co-expressed with MASH1. However, co-expression of GSX2 and MASH1 caused a decrease in DLX2 expression followed by an increase **(Figure 5.14 D, E, F and G)**. These results clearly show that TFs interact to drive the level of endogenous gene expression.

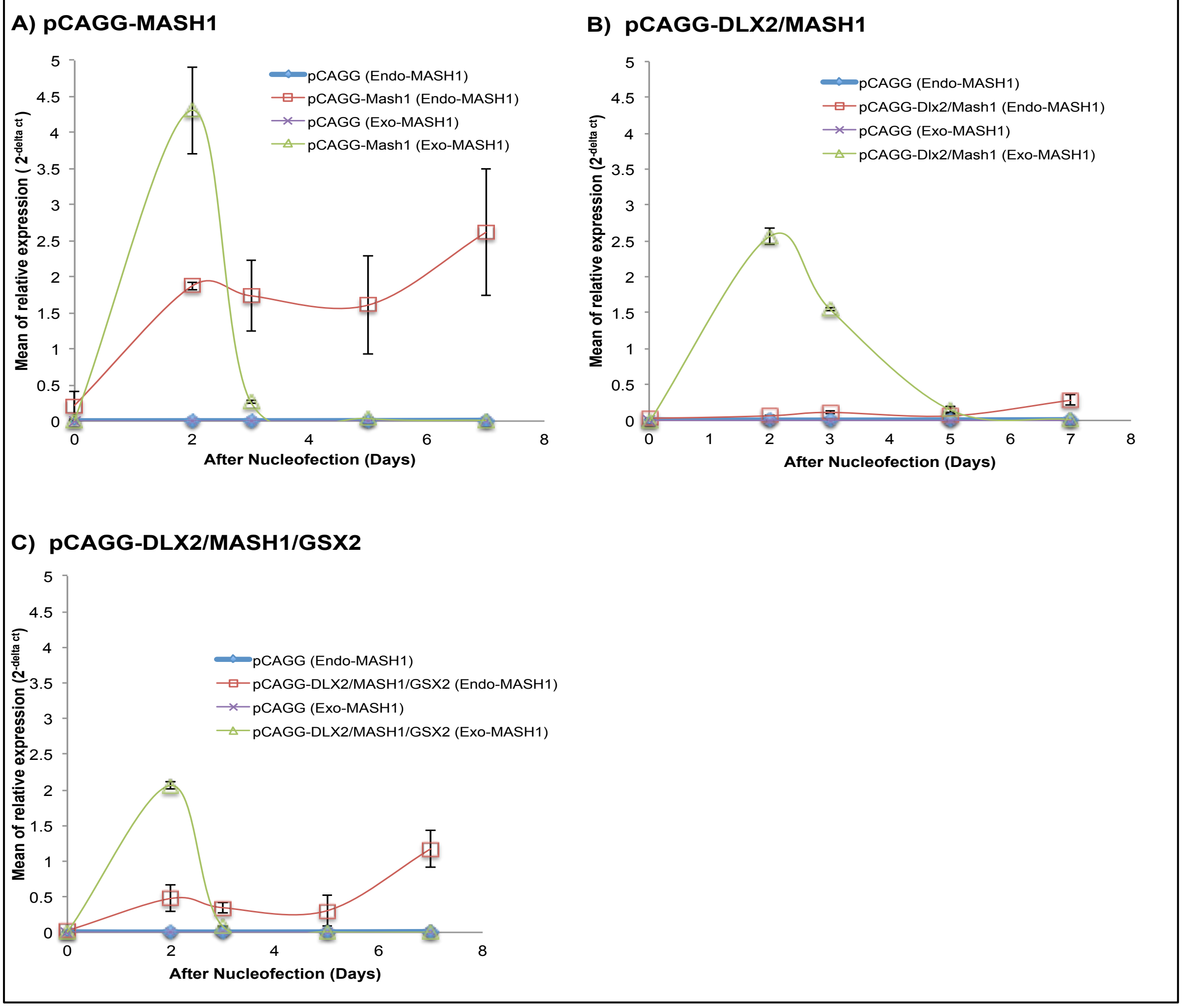
5.4.2.3 MASH1 co-expression increases GSX2 starting at ND5

The mean relative expression of exogenous GSX2 peaked two days after nucleofection with pCAGG-DLX2/GSX2 or pCAGG-DLX2/MASH1/GSX2 **(Figure 5.14 H and I)**. The mean relative expression of endogenous GSX2 differed depending on the constructs expressed in the 34D6 cells. In the pCAGG-DLX2/GSX2 cells, the expression of endogenous GSX2 reached its maximum level at ND5 then declined at ND7 **(Figure 5.14 H)**. However, when the GSX2 was co-expressed with DLX2 and

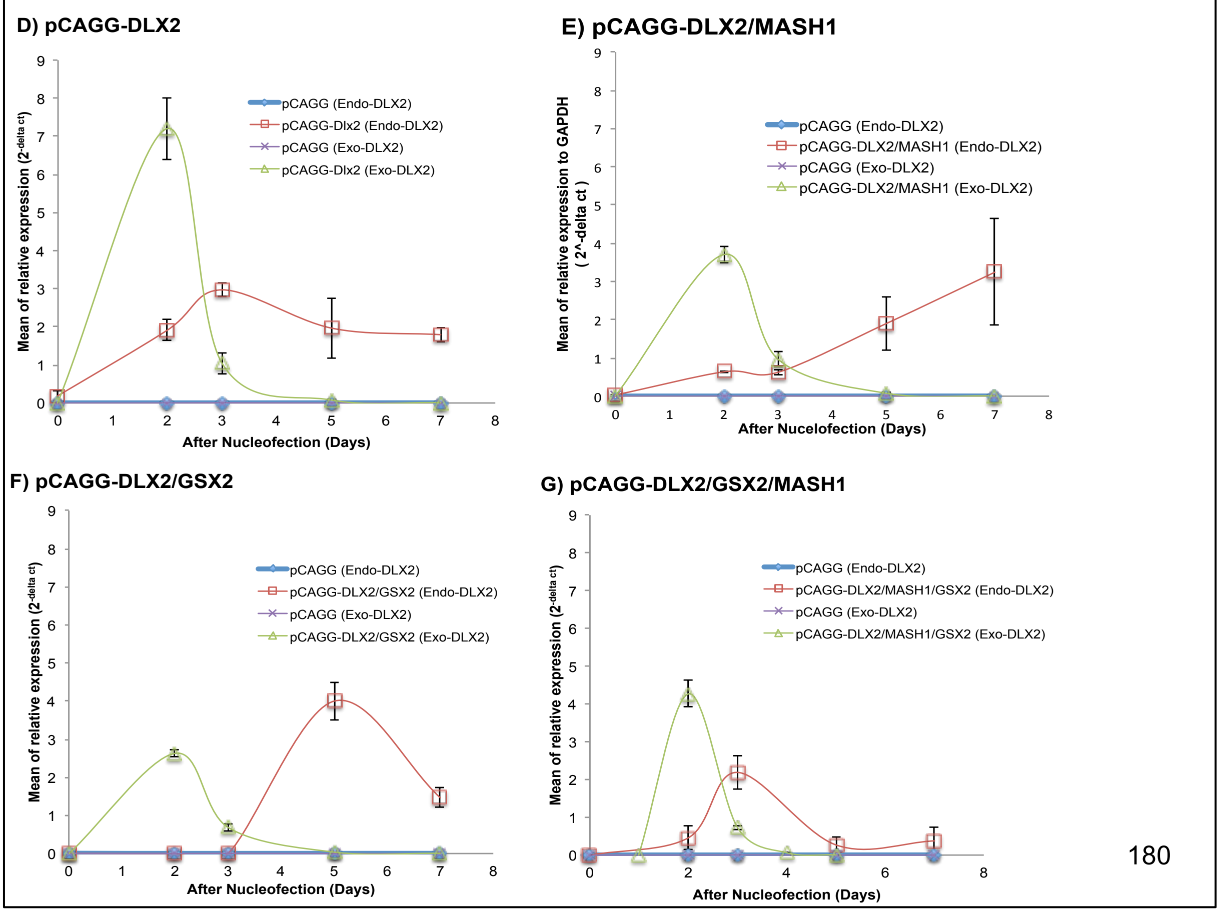
MASH1, the endogenous expression of GSX2 started to increase at ND5 reaching its peak at ND7 **(Figure 5.14 I),** whereas the expression of endogenous DLX2 and MASH1 was only increased slightly at ND7 in the pCAGG-DLX2/MASH1/GSX2 nucleofected cells **(Figure 5.14 C, G and I)**. Section 5.4.2 demonstrates the complexity of TF interaction and expression, which likely influence the proliferation-differentiation drive in neurogenesis in this model.

This experiment was done in order to examine whether or not the exogenous expression of TFs become integrated (endogenous expression of TFs) in nucleofected cells. The results show that the post-nucleofection expression of endogenous TFs is increased when the expression of exogenous TFs is decreased.

Relative Expression of Endogenous and Exogenous Mash1



Relative Expression of Endogenous and Exogenous DLX2



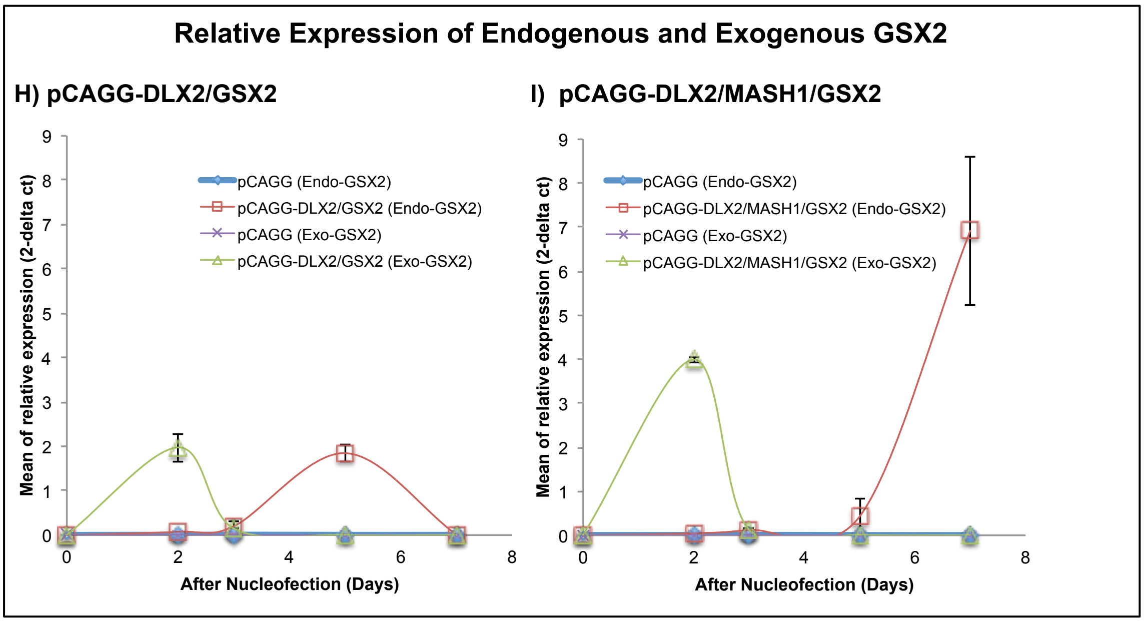


Figure 5.14: Endogenous and exogenous expression of the TFs, DLX2, MASH1 and GSX2, in the nucleofected 34D6 cells compared to the control (pCAGG empty vector) nucleofected cells.

The endogenous and exogenous expression of MASH1 in pCAGG-MASH1, pCAGG-DLX2/MASH1 and pCAGG-DLX2/MASH1/GSX2 nucleofected 34D6 cells (A, B, C). The endogenous and exogenous expression of DLX2, in all nucleofected cells (D, E, F, G). The endogenous and exogenous expression of GSX2 in pCAGG-DLX2/GSX2 and pCAGG-DLX2/MASH1/GSX2 nucleofected cells (H, I).

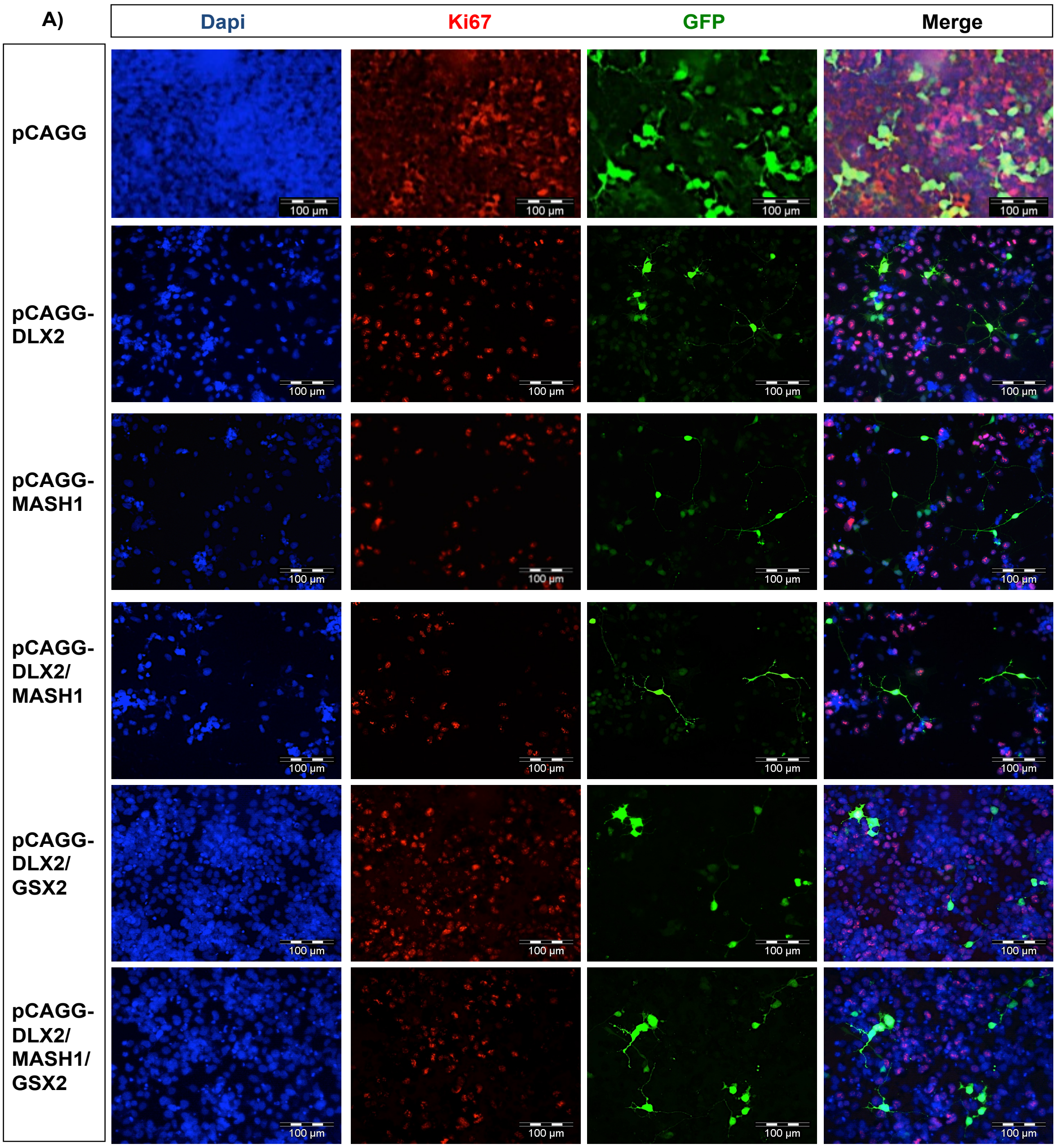
5.4.3 Transient ectopic expression of DLX2, MASH1 and GSX2 resulted in cell cycle exit leading to neuronal differentiation, as observed by downregulation of the proliferation marker Ki67

The previous section described the nucleofection and validation of expression of DLX2, MASH1 and GSX2 TFs in nrNPCs. In this section, in order to determine when the cells exit the cell cycle and start to differentiate, the nucleofected cells were analysed for expression of GFP and Ki67, a proliferation marker. Two different populations of nucleofected cells were involved in this analysis: GFP^{+ve} cells populations indicated the transient ectopic expression of TFs, meanwhile, GFP^{-ve} cells populations indicated the non-ectopic expression of TFs. Therefore, for this analysis, nucleofected cells were not placed under G418 selection in order to examine the Ki67 expression in the GFP^{-ve} cell population.

Between ND3 and ND4, the nucleofected cells were fixed and stained with the proliferation marker Ki67 and GFP **(Figure 5.15 A)**. The percentage of GFP^{+ve} and GFP^{-ve} cells that were Ki67^{+ve} was calculated using ImageJ software from 4 replicates. Statistically, there was a significant (p-value < 0.0001) decrease in the percentage of Ki67^{+ve} cells in the GFP^{+ve} populations as compared to the GFP^{-ve} populations in all the TF vector nucleofected groups **(Figure 5.15 B)**. The p–value of Ki67^{+ve} cells between GFP^{+ve} and GFP^{-ve} cells for pCAGG-DLX2, pCAGG-MASH1, pCAGG-DLX2/MASH1, pCAGG-DLX2/GSX2 and pCAGG-DLX2/MASH1/GSX2 nucleofected cells, was <0.0001, 0.01, 0.0001, 0.001 and 0.01, respectively (N= 4 for each nucleofection). However, the empty vector (pCAGG) nucleofected cells did not show any significant difference in the percentage of Ki67 cells between GFP^{+ve} and GFP^{-ve} cells.

In the GFP^{+ve} population, the pCAGG-DLX2/MASH1 nucleofected cells showed the lowest Ki67 expression (p-value = 0.0001) compared with the control pCAGG, and pCAGG-DLX2, pCAGG-DLX2/GSX2 and pCAGG-DLX2/MASH1/GSX2 nucleofected cells with p-value = 0.001 **(Figure 5.15 B)**.

These data suggest that the 34D6 nrNPCs ectopically expressing the different TFs constructs (with emphasis on pCAGG-DLX2/MASH1) experience decreased proliferation, as evidenced by the decline in cell cycle marker Ki67. This, therefore, indicates that these cells have exited the cell cycle and have committed to differentiation.


A)
Dapi
Ki67
GFP
Merge
pCAGG
pCAGG-DLX2
pCAGG-MASH1
pCAGG-DLX2/MASH1
pCAGG-DLX2/GSX2
pCAGG-DLX2/MASH1/GSX2
100 µm

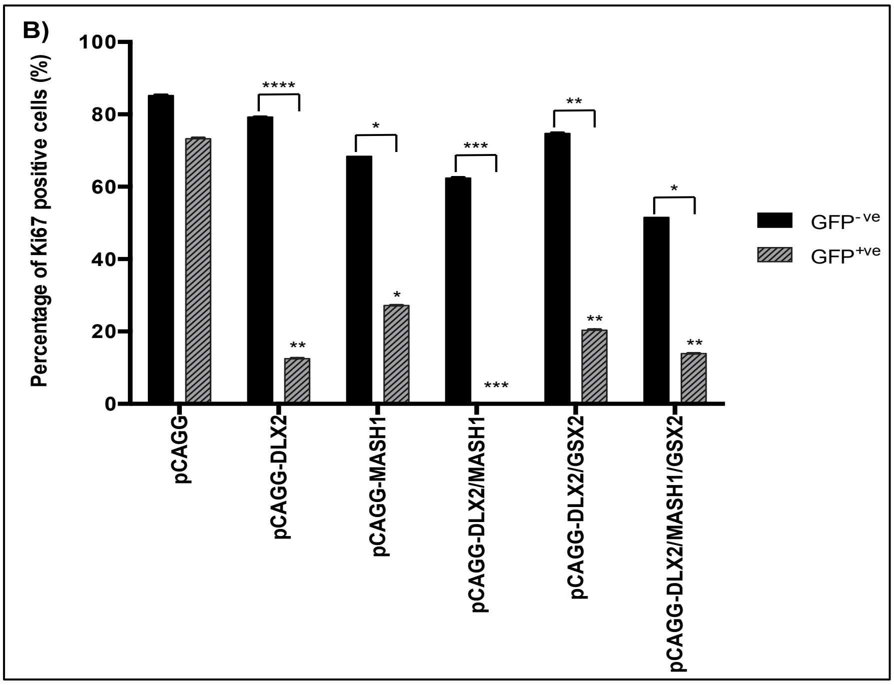

B)
Percentage of Ki67 positive cells (%)
100
80
60
40
20
0
****
*
***
**
*
**
*
***
**
**
GFP-ve
GFP+ve
pCAGG
pCAGG-DLX2
pCAGG-MASH1
pCAGG-DLX2/MASH1
pCAGG-DLX2/GSX2
pCAGG-DLX2/MASH1/GSX2

Figure 5.15: Analysis of cell proliferation, using Ki67, in 34D6 nrNPCs four days after nucleofection with TF vectors.

The six expression vectors were nucleofected into 34D6 nrNPCs, and at ND4 the cells were fixed and immuno-stained against Ki67 (Cell cycle marker-Red), GFP (Green) and Dapi (Nuclear staining-Blue) (A). The scale bar shows 100μm. The percentage of Ki67^{+ve} cells, in each nucleofection group, was calculated between GFP^{+ve} and GFP^{-ve} populations (N = 4 for each nucleofection) (B).

* indicates p-value = 0.01, ** indicates p-value = 0.001, *** indicates p-value = 0.0001 and **** indicates p-value < 0.0001 (Two-way ANOVA with Bonferroni correction).

5.4.4 Ectopic expression of TFs induces an LGE-like progenitor fate from 34D6-derived forebrain nrNPCs, as assessed by dorsal-specific markers (PAX6 and EMX2) and a ventral-specific marker for MGE (NKX2.1).

In order to understand and demonstrate the functional effect of the five TFs expression constructs along with empty vector (control group), on the 34D6 nrNPCs, RNA samples were analysed by qRT-PCR for expression of the dorsal markers, EMX2 and PAX6, and the ventral MGE specific marker, NKX2.1 **(Figures 5.16, 5.17 and 5.18)**. This analysis was repeated at ND0, ND2, ND3, ND5 and ND7 for EMX2 and NKX2.1 expressions, plus ND21 and ND42 for PAX6 expression. For this analysis, nucleofected cells were placed under G418 selection.

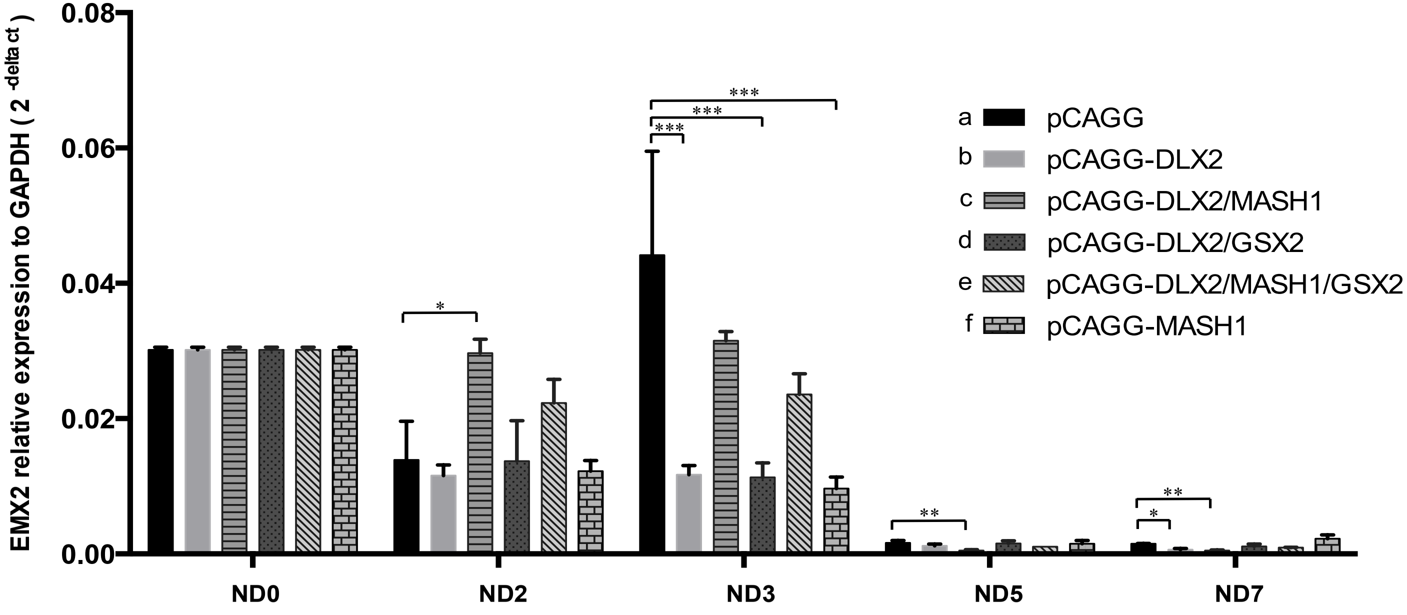
It was observed that there was a significant difference, in the expression of the dorsal marker, EMX2, between the nucleofected 34D6 nrNPCs (p = 0.0126) and between different time points (p < 0.0001). Moreover, the interaction of nucleofected 34D6 nrNPCs and incubation time was significant (p < 0.0001). From ND0 to ND2, only four groups of nucleofected cells displayed significantly decreased EMX2 expression (p = 0.0126). Those were pCAGG-DLX2, pCAGG-DLX2/GSX2, pCAGG-MASH1 and also the control **(Figure 5.16)**. From ND1 to ND3, there was a steady expression of EMX2 in all nucleofected cells in the control; however, there was a significant increase (p = 0.001) **(Figure 5.16)**. From ND3 to ND5 and ND7, there was dramatic decrease in EMX2 expression in all nucleofected cells with a p-value of less than 0.0001 **(Figure 5.16)**. Between ND5 and ND7, EMX2 expression in the pCAGG-DLX2/MASH1 nucleofected cells decreased significantly (p = 0.001) compared to the control group (pCAGG nucleofected cells) **(Figure 5.16)**. Furthermore, EMX2 expression in pCAGG-DLX2 nucleofected cells declined significantly (p-value = 0.0126) at ND7 compared to the control group **(Figure 5.16)**.

There were significant differences between PAX6 expression among the nucleofected 34D6 nrNPCs (p-value 0.0102) and also at the different time points (p-value 0.0001). In addition, the interaction of nucleofected 34D6 nrNPCs and incubation time was significant ($p = 0.0168$). Interestingly, the expression of PAX6 in the 34D6 nrNPCs nucleofected with pCAGG and pCAGG-DLX2/MASH1 were the same from ND0 to ND5 **(Figure 5.17)**. Expression of PAX6, after ND5, in the pCAGG-DLX2/MASH1 nucleofected cells was reduced statistically significant compared with ND0; however, in control (pCAGG nucleofected) cells, expression increased from ND5 to ND7 **(Figure 5.17)**. Moreover, there was a further decrease in PAX6 expression in pCAGG-DLX2/MASH1 nucleofected cells from ND7 to ND21 ($p = 0.0102$); no statistically significant reduction in PAX6 expression was observed in the other nucleofected cells over the same period **(Figure 5.17)**. From ND3 to ND21 and ND42, the most significant reduction in PAX6 expression was the pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs (p-value < 0.0001) **(Figure 5.17)**. In contrast, expression of PAX6 in pCAGG-MASH1 nucleofected cells was noticeably increased at ND21 and ND42 compared with control ($p = 0.0102$ and 0.001, respectively) **(Figure 5.17)**. In summary, pCAGG-DLX2/MASH1 nucleofected cells showed significant downregulation in the PAX6 expression from ND0 to ND42 (p-value < 0.0001) **(Figure 5.17)**.

There was a statistical difference in the expression of NKX2.1 between the nucleofected 34D6 nrNPCs (p-value < 0.0001) and at different time points (p-value < 0.001) **(Figure 5.18)**. At ND2, only the pCAGG-MASH1 nucleofected cells showed dramatic reduction of NKX2.1 levels ($p = 0.001$) compared with the control group (pCAGG) **(Figure 5.18)**. At ND3, ND5 and ND7, there was a significant decrease in the expression of NKX2.1 in all nucleofected cells compared to the control group (pCAGG) (p-value < 0.0001) **(Figure 5.18)**.

NKX2.1 expression decreased considerably in pCAGG-DLX2/MASH1 ($p < 0.0001$) and pCAGG-DLX2/MASH1/GSX2 ($p = 0.001$) nucleofected cells from ND3 to ND5 **(Figure 5.18)**. From ND3 to ND7, there was a dramatic reduction of NKX2.1 expression in pCAGG-DLX2/MASH1, pCAGG-DLX2/GSX2 and pCAGG-DLX2/MASH1/GSX2 nucleofected cells ($p = 0.001$, 0.01 and 0.001, respectively) **(Figure 5.18)**. From these results, we noticed that in all the cells nucleofected with vectors containing different combinations of DLX2, MASH1 or GSX2, there was a significant decrease in NKX2.1 expression from ND3 to ND7

The above experimental data suggest that, in combination, the TFs DLX2, MASH1 and GSX2 are capable of inducing an LGE-like progenitor fate from 34D6-derived forebrain nrNPCs.



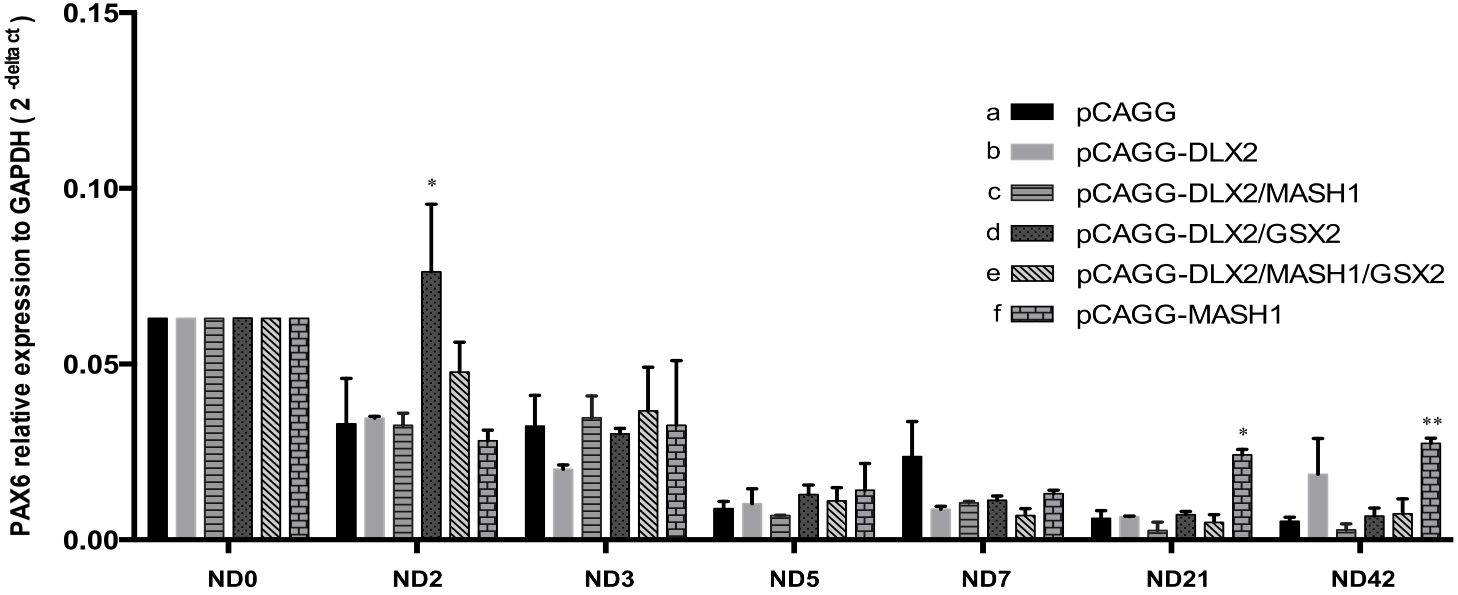
	ND2						ND3						ND5						ND7					
	a	b	c	d	e	f	a	b	c	d	e	f	a	b	c	d	e	f	a	b	c	d	e	f
ND0	*	*		*		*		*		*		**	****	****	****	****	****	****	****	****	****	****	****	****
ND2							**						****	****	****	****	****	****	****	****	****	****	****	****
ND3													****	****	****	****	****	****	****	****	****	****	****	***

Nucleofection day (ND)

Figure 5.16: EMX2 expression in 34D6 nrNPCs ectopically expressing various TFs.

All nucleofected nrNPCs were analysed by qRT-PCR for expression of the dorsal-ventral marker EMX2 at ND0, ND2, ND3, ND5 and ND7. The table below the bar graph illustrates the statistical analysis of EMX2 expression in the nucleofected 34D6 nrNPCs (a, b, c, d, e and f) at different time points. Statistically significant changes compared to the control (pCAGG nucleofected cells) are indicated as follows:

p-value summary: (Two-way ANOVA with Bonferroni correction)
* = 0.0126, ** = 0.001, *** = 0.0001, and **** < 0.0001



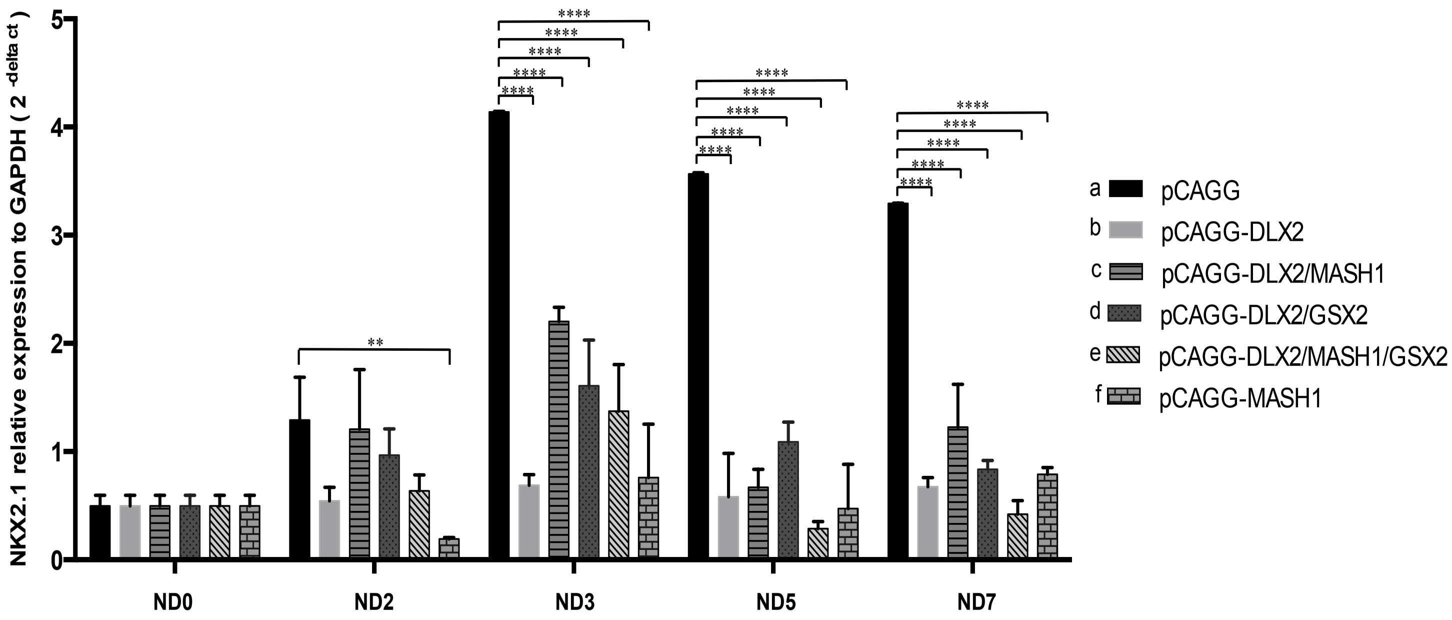
	ND5						ND7						ND21						ND42					
	a	b	c	d	e	f	a	b	c	d	e	f	a	b	c	d	e	f	a	b	c	d	e	f
ND0	***	***	***	**	***	**		***	***	**	****	**	****	****	****	****	****		****	*	****	****	****	
ND2			*	***	*			*		***	**		*	*	****	****	***		**		***	****	***	
ND3			*								*		*		****	*	**		**		****	*	**	
ND7															*									

Nucleofection day (ND)

Figure 5.17: PAX6 expression in 34D6 nrNPCs ectopically expressing various TFs.

All nucleofected nrNPCs were analysed by qRT-PCR for expression of the dorsal-ventral marker PAX6 at ND0, ND2, ND3, ND5, ND7, ND21 and ND42. The table below the graph shows the statistical analysis of PAX6 expression in the nucleofected 34D6 nrNPCs (a, b, c, d, e and f) between the time points. Statistically significant changes compared to the control (pCAGG nucleofected cells) are indicated as follows:

p-value summary: (Two-way ANOVA with Bonferroni correction)
* = 0.0102, ** = 0.001, *** = 0.0001, and **** < 0.0001



	ND5						ND7					
	a	b	c	d	e	f	a	b	c	d	e	f
ND3			****		**				**	*	**	

Nucleofection day (ND)

Figure 5.18: NKX2.1 expression in 34D6 nrNPCs ectopically expressing various TFs.

All nucleofected nrNPCs were analysed by qRT-PCR for expression of NKX2.1, which is a ventral specific marker for MGE. This was done at time points ND0, ND2, ND3, ND5 and ND7. The table below the graph shows the statistical analysis of NKX2.1 expression in the nucleofected 34D6 nrNPCs (a, b, c, d, e and f) between the time points. Statistically significant changes compared to the control (pCAGG nucleofected cells) are indicated as follows:

p-value summary: (Two-way ANOVA with Bonferroni correction)
* = 0.01, ** = 0.001, and **** < 0.0001

5.4.4.1 Overexpression of TFs DLX2, GSX2 and MASH1 in nrNPCs have an effect on endogenous target genes

In this section, the target genes for the TFs DLX2, MASH1 and GSX2 were analysed at the transcriptional level by qRT-PCR at different time points (ND0, ND2, ND3, ND5 and ND7). The DLX2 target genes are ARX **(Figure 5.19)** (Colasante *et al.* 2008) and GAD2 **(Figure 5.20)** (Stühmer *et al.* 2002b). The GSX2 target genes are EBF1 **(Figure 5.21)**, DLX2 (See previous **Figure 5.14 F and G**) and MASH1 (See **Figure 5.14 C** for pCAGG-DLX2/MASH1/GSX2 nucleofected cells) (Wang *et al.* 2013). The MASH1 target genes are DLX2 (Poitras *et al.* 2007) (See **Figure 5.14 E and G** for pCAGG-DLX2/MASH1 and pCAGG-DLX2/MASH1/GSX2 nucleofected cells).

The pallial-subpallial boundary (PSB) is determined by the related repression of PAX6 and GSX2, and both of these TFs play a role in the patterning of DV of telencephalon (Yun *et al.* 2001; Pauly *et al.* 2013). When GSX2 was ectopically expressed in pCAGG-DLX2/GSX2/MASH1 and pCAGG-DLX2/GSX2 nucleofected 34D6 nrNPCs, the expression of PAX6 was affected **(Table 5.1).** In pCAGG-DLX2/MASH1/GSX2 nucleofected cells, ectopic expression of endogenous GSX2 was increased and reached its peak at ND7, and the expression of PAX6 was reduced at ND7 compared to the control group (pCAGG) **(Table 5.1)**. Moreover, in pCAGG-DLX2/GSX2 nucleofected cells, the ectopic expression of endogenous GSX2 was increased from ND3 to ND5, while the expression of PAX6 was reduced **(Table 5.1).** These findings suggest that ectopic expression of GSX2 in the nucleofected 34D6 nrNPCs repressed the expression of PAX6.

Table 5.1: Summary of the outcome of PAX6 and GSX2 expression, which determines the boundary of pallial-subpallial (PSB) of telencephalon.

Expression / Construct expression	PAX6	GSX2
pCAG-DLX2/MASH1/GSX2	The expression was decreased at ND7 in the pCAGG-DLX2/MASH1/GSX2 nucleofected 34D6 nrNPCs.	The expression was increased and reached its peak at ND7 in the pCAGG-DLX2/MASH1/GSX2 nucleofected 34D6 nrNPCs.
pCAG-DLX2/GSX2	The expression was decreased from ND3-ND5 in the pCAGG-DLX2/GSX2 nucleofected 34D6 nrNPCs.	The expression was increased from ND3 to ND5 in the pCAGG-DLX2/GSX2 nucleofected 34D6 nrNPCs.

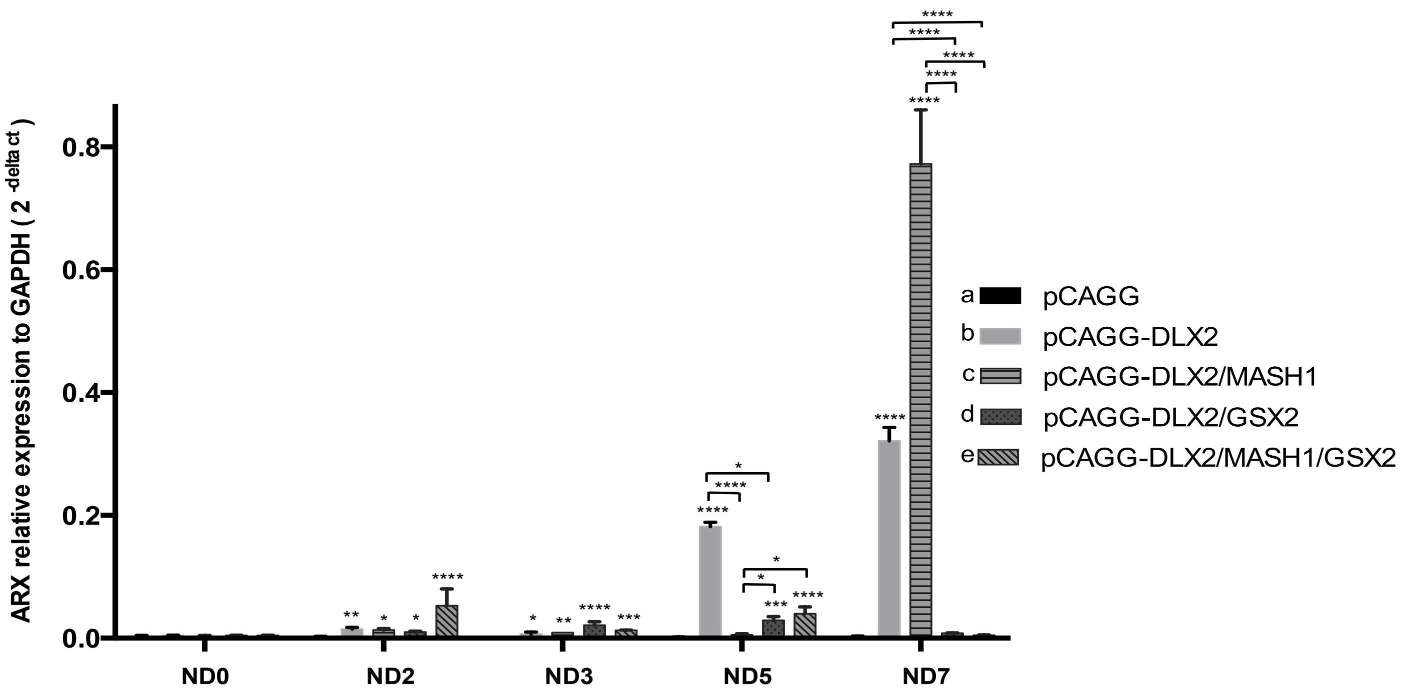
There was a significant difference in the expression of ARX in all nucleofected nrNPCs, compared to control (pCAGG nucleofected) cells, across the time points ($p < 0.0001$) **(Figure 5.19)**. Furthermore, there were significant differences in the expression of ARX between the different groups of nucleofected cells ($p < 0.0001$). In the pCAGG-DLX2/MASH1 nucleofected cells, there was an 8 fold increase in the expression of ARX at ND7 as compared to ND0 ($p < 0.0001$) **(Figure 5.19)**. Additionally, compared to the control at ND7, the increase in expression of ARX, in the pCAGG-DLX2/MASH1 nucleofected nrNPCs, was statistically significant ($p < 0.0001$) **(Figure 5.19)**. The pCAGG-DLX2 nucleofected cells showed a significant increase in ARX expression from ND0 to ND7 (p-value < 0.0001) **(Figure 5.19)**. Compared to pCAGG-DLX2 nucleofected cells, the expression of ARX at ND7 was significantly increased in pCAGG-DLX2/MASH1 nucleofected cells ($p < 0.0001$) **(Figure 5.19)**. Consequently, it was strongly indicated that overexpression of DLX2 and MASH1 is associated with intracellular accumulation of ARX.

GAD2 is one of DLX2's target genes (Stühmer *et al.* 2002b) and is commonly used as a striatal neuron marker in SVZ and MZ (Anderson 1997b; Casarosa *et al*. 1999; Horton *et al*. 1999; Yun *et al*. 2002; Long *et al*. 2009a). Results also reveal a significant difference in the expression of GAD2 in all nucleofected nrNPCs, compared to control (pCAGG nucleofected) cells, across the time points ($p < 0.0001$) **(Figure**

5.20). Moreover, there was a significant difference in the expression of GAD2 between the different groups of nucleofected cells ($p < 0.0001$). In the pCAGG-DLX2 nucleofected cells, the expression of GAD2 at ND2 was significantly increased ($p = 0.001$) compared to the control. Then, from ND2 to ND3, there was a sharp decline, and then start to increase from ND5 to ND7 **(Figure 5.20)**. This can be explained by the outcome of exogenous and endogenous expression of DLX2. The exogenous expression of DLX2 was reached its peak at ND2 and then declined sharply from ND2 to ND3. In this duration time (ND2 to ND3), the endogenous expression of DLX2 was initiated to increase and stayed constant from ND5 to ND7 **(Figure 5.14 D)**. In the pCAGG-DLX2/MASH1 nucleofected cells, the expression of GAD2 from ND3 to ND5 was increased significantly ($p < 0.0001$) compared to the control **(Figure 5.20)**. Then, there was a slight decreased but was not significant from ND5 to ND7. However, compared to the control at ND7, the increase in expression of GAD2, in the pCAGG-DLX2/MASH1 nucleofected cells, was statistically significant ($p = 0.01$) **(Figure 5.20)**. In addition, in the pCAGG-DLX2/GSX2 nucleofected cells, the expression of GAD2 was increased significantly at ND3, ND5 and ND7 compared to the control ($p = 0.0001$, $p = 0.01$ and $p = 0.01$, respectively) **(Figure 5.20)**. Furthermore, in the pCAGG-DLX2/GSX2/MASH1 nucleofected cells, the increase in expression of GAD2 was in accordance with the increase of endogenous expression of DLX2 **(Figures 5.14 G and 5.20)**. The expressions of DLX2 and GAD2 were increased from ND2 to ND3, then declined from ND3 to ND5 and started to increase from ND5 to ND7 **(Figures 5.14 G and 5.20)**.

EBF1 is a target gene for Gsx2 (Wang *et al.* 2013) that is also used as a marker for LGE-specific and striatal projection neurons (Garel *et al.* 1999; Lobo *et al.* 2006; Garcia-Dominguez *et al.* 2003). EBF1 expression in the cells nucleofected with different TF vectors **(Figure 5.21)** was subjected to ANOVA analysis. There was a significant difference between the various groups of nucleofected cells ($p < 0.0001$). Furthermore, both the ectopically expressed TFs and the time post-nucleofection interacted together to significantly affect EBF1 expression ($p = 0.0218$) **(Figure 5.21)**. Ectopic expression of DLX2 and GSX2 resulted in a significant increase in EBF1 expression from ND0 to ND5 ($p = 0.001$.) At ND5 EBF1 expression in these cells was increased compared to pCAGG nucleofected nrNPCs ($p < 0.0001$) **(Figure 5.21)**. However, when MASH1 was expressed with the other TFs in the same cells, it caused a sharp reduction of EBF1 expression at ND5 ($p < 0.0001$) **(Figure 5.21)**. From ND5 to ND7, there was a reduction in the expression of EBF1 in pCAGG-DLX2/GSX2 nucleofected nrNPCs **(Figure 5.21)**, and this same time point at which endogenous GSX2 expression was decreased **(Figure 5.14 H)**. However, expression of EBF1 in pCAGG-DLX2/GSX2 nucleofected cells was significantly increased compared to the control group (pCAGG nucleofected cells) at ND7 ($p = 0.0218$) **(Figure 5.21)**.

In conclusion, the ectopic expressions of the TFs MASH1, DLX2 and GSX2 have an effect on their target genes, which in turn regulate striatal differentiation and define the striatal phenotype. Hence, the ectopic expression of TFs can trigger differentiation into LGE-like cells via the fundamental functions of their target genes.



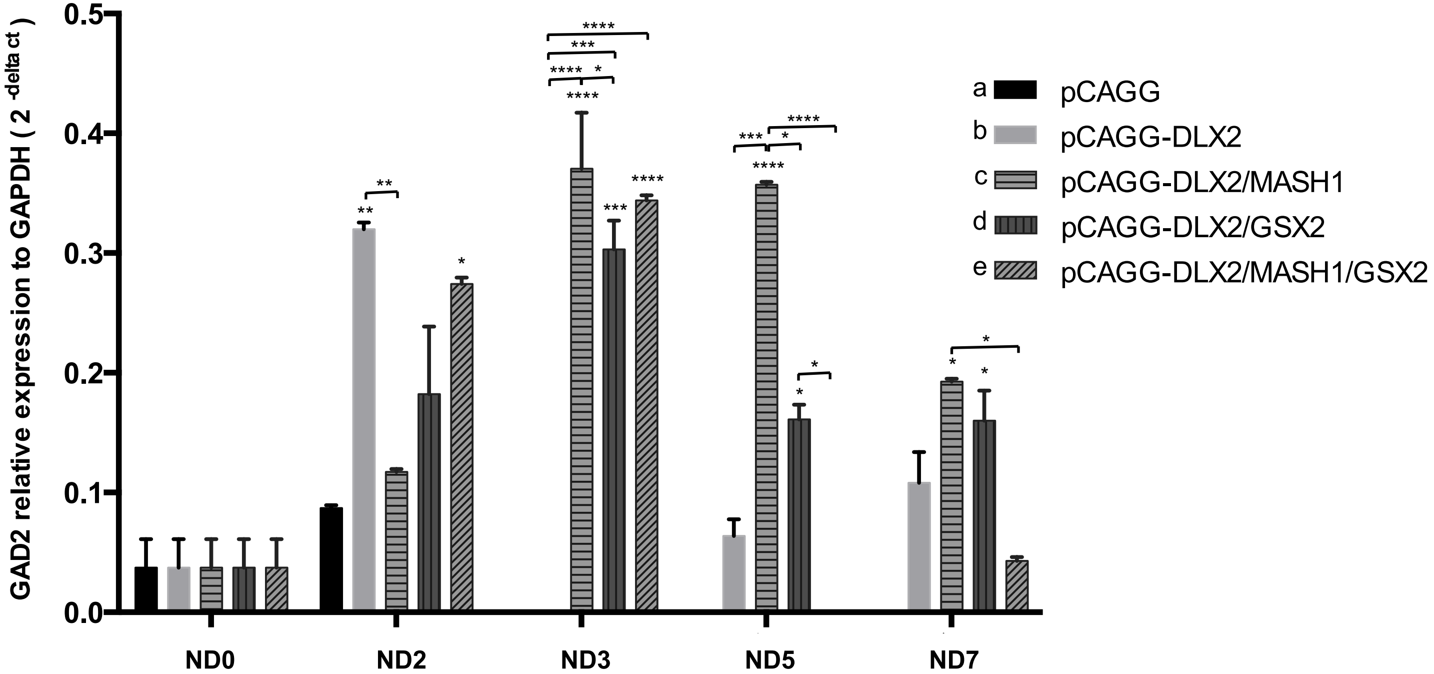
	ND2					ND3					ND5					ND7				
	a	b	c	d	e	a	b	c	d	e	a	b	c	d	e	a	b	c	d	e
ND 0			**		**			*				****		*	**		****	****		
ND 2												**					****	****		**
ND 3												****					****	****		
ND 5																		****		**

Nucleofection day (ND)

Figure 5.19: DLX2 targets expression of ARX in the nucleofected 34D6 nrNPCs.

All nucleofected nrNPCs were analysed by qRT-PCR for expression of ARX at ND0, ND2, ND3, ND5 and ND7. The table below the graph shows the statistical analysis of ARX expression in the nucleofected 34D6 nrNPCs (a, b, c, d and e) between the time points. Statistically signficant changes compared to the control (pCAGG nucleofected cells) are indicated as follows:

p-value summary: (Two-way ANOVA with Bonferroni correction)
* = 0.0218, ** = 0.001, *** = 0.0001, and **** < 0.0001.



	ND2					ND3					ND5					ND7				
	a	b	c	d	e	a	b	c	d	e	a	b	c	d	e	a	b	c	d	e
ND0		**			*			***	**	***			**							
ND2							***	**				**	**		**		*			**
ND3															***					***

Nucleofection day (ND)

Figure 5.20: DLX2 targets expression of GAD2 in the nucleofected 34D6 nrNPCs.

All nucleofected nrNPCs were analysed by qRT-PCR for the expression of GAD2 at ND0, ND2, ND3, ND5 and ND7. The table below the graph shows the statistical analysis of GAD2 expression in the nucleofected 34D6 nrNPCs (a, b, c, d and e) between the time points. Statistically significant changes compared to the control (pCAGG nucleofected cells) are indicated as follows:

p-value summary: (Two-way ANOVA with Bonferroni correction)

* = 0.01, ** = 0.001, *** = 0.0001, and **** < 0.0001

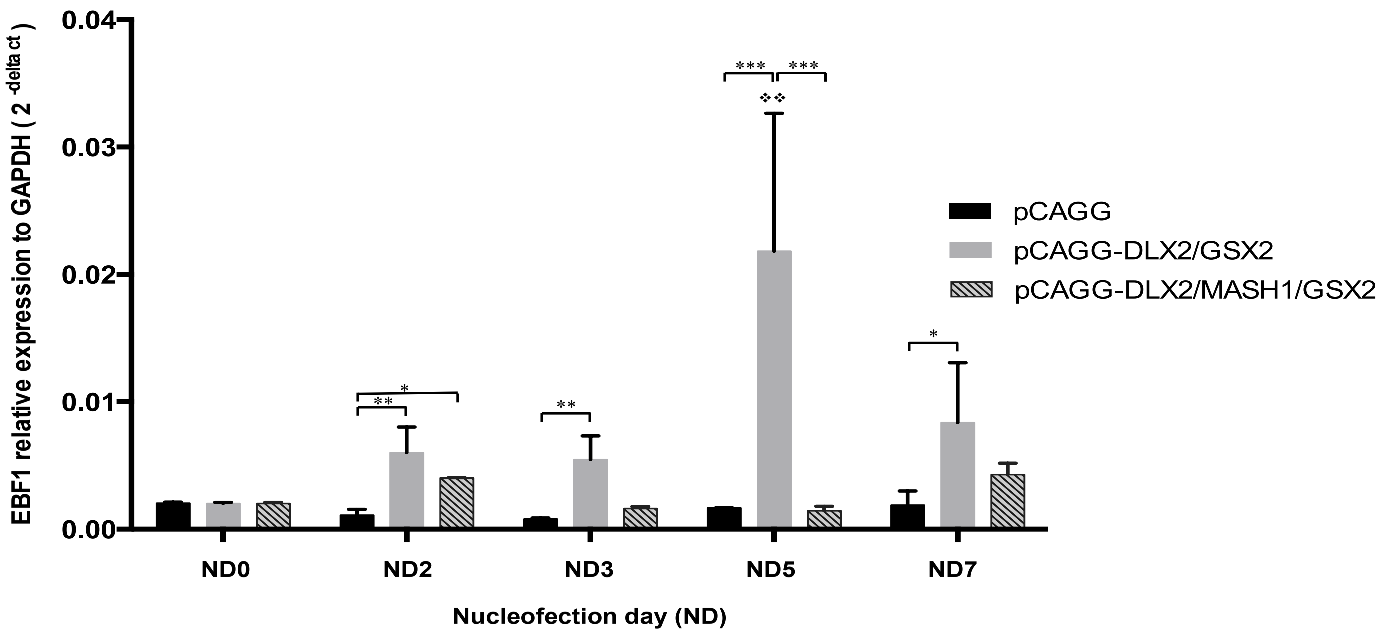


Figure 5.21: GSX2 targets expression of EBF1 in the nucleofected 34D6 nrNPCs.

All nucleofected nrNPCs were analysed by qRT-PCR for the expression of EBF1 at ND0, ND2, ND3, ND5 and ND7. The symbols on the bar graph indicate the statistically significant changes compared to the control (pCAGG nucleofected cells) within each time point. This symbol ❖ represents the statistical analysis between different time points within each group of nucleofected cells.

p-value summary: (Two-way ANOVA with Bonferroni correction)
* = 0.0218, *** = 0.0001, and ❖❖ = 0.001 (ND5 vs. ND0 of pCAGG-DLX2/GSX2).

5.5 Discussion

In this chapter, H9 hESCs and 34D6 iPSCs were differentiated using the differentiation protocol outlined in section 3.2.4 for neural induction and neural rosette formation (nrNPCs). Subsequently NSC specification towards a LGE progenitor phenotype was promoted by ectopic expression of the TFs DLX2, MASH1 and GSX2. The experimental protocol was designed to elicit transient expression of nucleofected transcription factor vectors. This was for two reasons. First, in order to investigate if cell proliferation was decreased as a result of nucleofection with control and TF expression vectors, thus, two populations of cells (GFP^{-ve} and GFP^{+ve}) were analysed to determine the percentage of Ki67 positive cells. Second, to compare the exogenous mRNA expression of TFs DLX2, MASH1 and GSX2 in the TF vector nucleofected 34D6 nrNPCs versus control (pCAGG) nucleofected cells. In this way, the time point of peak expression of exogenous mRNA (transient expression) could be observed. If the G418 selection was initiated, the mRNA expression of the transgenes would be maintained; therefore, it would not be possible to determine the time point at which transgene expression reaches its peak. Determining the time point of peak expression, after nucleofection, was necessary in order to determine the ideal time to initiate G418 selection.

ICC analysis of the nucleofected nrNPCs showed that the specific TF expression was only apparent in cells transfected with the specific cloned TF containing vector. For example, DLX2 was only expressed in the nucleofected nrNPCs with the DLX2 containing vectors (pCAGG-DLX2, pCAGG-DLX2/MASH1, pCAGG-DLX2/GSX2 and pCAGG-DLX2/MASH1/GSX2). Previous studies have established the use of polycistronic-vector-containing 2A self-cleaving peptides to express more than one gene from a single vector *in vitro* (Sommer *et al.* 2009) and *in vivo* (Szymczak *et al.* 2004). In this study, the expression of 2A peptides was evident in cells transfected with all five expression vectors (pCAGG-DLX2, pCAGG-MASH1, pCAGG-

DLX2/MASH1, pCAGG-DLX2/GSX2 and pCAGG-DLX2/MASH1/GSX2) compared to the empty vector. These observations point out to successful self-cleavage of 2A peptides during translation and processing of the polycistronic vector, to release the cloned TFs DLX2, GSX2 and MASH1. Therefore, the cloned TFs were successfully expressed in the 34D6 and H9 derived nrNPCs.

Cell proliferation on the nucleofected 34D6 nrNPCs at day 4 post-nucleofection (ND4)

In neural tube, neurogenesis is characterised by the generation of different cell types at specific periods, locations and production numbers. The complexity of neurogenesis involves the regulation and organization of following processes, such as, cell proliferation, neuron differentiation, neuron specification and migration (Toresson *et al.* 2000; Yun *et al*. 2002; Suh *et al.* 2009; Jones and Connor 2012). However, TFs that play a role in these processes are not fully characterised.

A growing number of studies have shown that TFs play a role in neurogenesis and also play a role in regulating cell proliferation (Casarosa *et al.* 1999; Toresson *et al.* 2000; Yun *et al*. 2002; Suh *et al*. 2009; Jones and Connor 2012). Interestingly, DLX2, MASH1 and GSX2, particularly DLX/Mash1 combination, seemed to have an effect on cell proliferation in this study. The amount of cell proliferation was significantly reduced in pCAGG-DLX2, pCAGG-DLX2/GSX2 and pCAGG-DLX2/MASH1/GSX2 nucleofected 34D6 nrNPCs, and almost absent in pCAGG-DLX2/MASH1 nucleofected progenitors, as indicated by decreased Ki67. Indeed, this indicates that these cells are no longer proliferative progenitors, i.e. they are exiting the cell cycle and have the potential to differentiate. TFs, particularly DLX2 and MASH1, which have parallel and complementary roles in neurogenesis (Casarosa *et al*. 1999; Yun *et al*. 2002; Long *et al*. 2009a) regulate the interface between proliferation and cell cycle exit/differentiation through different mechanisms. One of these mechanisms is proneural Mash1-stimulatory and DLX2-inhibitory effect on the Notch signalling pathway, which controls proliferation (Casarosa *et al*. 1999; Horton *et al*. 1999; Yun *et al.* 2002; Poitras *et al.* 2007; Castro *et al.* 2011; Jones and Connor 2012). The loss of proliferative progenitors is evident in the SVZ of MGE in the Mash1$^{-/-}$ mutant mice model (Casarosa *et al*. 1999). Epidermal growth factor receptor (EGFR)-mediated reduction of the Notch signalling pathway also contributes to reduced proliferation (Aguirre *et al*. 2010). Gil is another TF that controls the cell cycle through EGFR^{+} progenitors or MASH11^{+ve} or DLX2^{+ve} transient-amplifying precursors (TAPs) in the SVZ (Doetschman *et al.* 1985; Suh *et al.* 2009). Furthermore, the interaction of basic helix-loop-helix factor (bHLH) with cyclin-dependent kinase inhibitors (CDKIs), such as p27Kip1, induces cell cycle exit and simultaneously promotes differentiation of neurons by increasing the number of DLX2^{+ve} TAPs (Doetschman *et al*. 1985). In this respect, MASH1 is also a bHLH gene, which could suggest an interaction between MASH1 and CDKI to halt neuronal proliferation and initiate differentiation, lineage fate commitment and neuron specification through DLX2 and GSX2 (Casarosa *et al.* 1999; Jones and Connor 2012).

Dorsal-ventral marker expression in the nucleofected nrNPCs at PdD18

As described earlier in this chapter, only pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs showed a significant and constant reduction of dorsal markers, i.e. PAX6 and EMX2, compared to the control group (pCAGG), from ND3 to ND21. However, all ectopic expression of cloned TFs in pCAGG vector (DLX2, MASH1, DLX2/MASH1, DLX2/MASH1/GSX2, and DLX2/GSX2) showed significant reduction in the expression of the ventral MGE marker NKX2.1, as compared to control group, in 34D6 nrNPCs.

This reduction of dorsal and ventral MGE markers points toward differentiation of progenitor cells into the LGE phenotype. The DLX2 and MASH1 play a critical role in maintaining subpallium telencephalon fate commitment, specification and eventually LGE development, i.e. GABAergic (Anderson *et al.* 1997; Casarosa *et al.* 1999; Horton *et al.* 1999; Fode *et al.* 2000; Yun *et al.* 2002; stuhmer *et al.* 2002b). These roles are demonstrated further in Dlx$^{-/-}$ mutant mice, which show a profound defect in LGE development and the differentiation of striatal matrix neurons (Anderson *et al*. 1997a).

In the same mutant mouse model, an increase in Pax6 has also been observed (Long *et al.* 2009a; Long *et al.* 2009b), suggesting that neuron differentiation is diverted into the dorsal part of the brain. Therefore, it has been concluded that the functional Dlx1&2 are essential for supressing some LGE progenitor TFs, including Pax6, and ventral cortical TFs, such as Nkx2.1 (cortical interneurons), and cortical TFs, thus maintaining the identity of LGE (Long *et al.* 2009a; Long *et al.* 2009b).

It was observed that, compared with the single ectopic expression of DLX2 or MASH1, co-expression of different combinations of DLX2, MASH1 or GSX2 in 34D6 nrNPCs resulted in a dramatic reduction in the expression of NKX2.1. These findings could indicate that the combinations of TFs DLX2, MASH1 or GSX2 play a critical role in LGE specific neurons. Many studies have reported that MASH1, DLX2 and GSX2 were required for LGE differentiation (Casarosa *et al*. 1999; Horton *et al*. 1999; Corbin *et al*. 2000; Toresson *et al*. 2000; Toresson and Campbell 2001; Yun *et al*. 2002; Yun *et al*. 2003). In the Dlx1&2$^{-/-}$ mutant LGE mouse, ectopic expression of cortical and MGE TFs was observed, indicating that Dlx1&2 regulate the identity of LGE by repression of some MGE TFs, such as Nkx2.1 and cortical TFs (Long *et al.* 2009a; Long *et al.* 2009b). Moreover, it was shown that the Dlx1/2$^{-/-}$:Mash1$^{-/-}$ mutants showed similar results in the expression of Nkx2.1 and more severe deficiency in the phenotype of LGE progenitors than in the single mutants of Dlx1/2$^{/-}$ or Mash1$^{-/-}$ as shown previously by other group (Long *et al.* 2009a; Long *et al.* 2009b; Wang *et al.* 2013). The same findings were observed in the Gsx2 and Mash1 double mutant (Wang *et al.* 2009).

The data presented in this chapter are in line with the previous studies suggesting that the reduction of dorsal and ventral MGE markers confirms correct route of differentiation into an LGE-like progenitor fate through the combination of ectopic expression of TFs in 34D6-derived forebrain nrNPCs.

The expression of target genes is regulated by ectopic expression of DLX2

DLX2 targets the ARX gene expression, which means that DLX2 has a role in regulating ARX expression (Colasante *et al.* 2008). In this study, it was found that the mRNA expression of ARX was increased in the pCAGG-DLX2 and pCAGG-DLX2/MASH1 nucleofected cells. However, the expression of ARX in pCAGG-DLX2/MASH1 nucleofected cells was higher than those in pCAGG-DLX2 nucleofected cells at ND7. The same observation has been made with endogenous DLX2 expression in pCAGG-DLX2/MASH1 and pCAGG-DLX2 nucleofected cells at ND7. Therefore, the overexpression of DLX2 has an effect in the expression of ARX. These outcomes were in agreement with earlier findings regarding the influence of Dlx2 activity on endogenous expression of ARX. It was previously observed that overexpression of Dlx2 in the forebrain of E13.5 mouse models regulated the endogenous expression of Arx by the activation of the cis-regulatory element UAS3 enhancer, which is located downstream of Arx (Colasante *et al.* 2008). This evident in Dlx1/2$^{-/-}$ mutant knockout mice, where Arx expression was decreased (Colasante *et al.* 2008).

The TF Arx is known as aristaless related homeobox (Colasante *et al.* 2008). It is located in the developing subpallium telencephalon in the LGE and MGE in the proliferating cells in subventricular zone (SVZ) and mantle zone. In the cerebral cortex, it is expressed in the proliferating cells in the ventricular zone (VZ), as well as in other parts of brain (Colasante *et al.* 2008). It has a role in forebrain development and in late born striatal projection neuron migration to the pallium (Colasante *et al.* 2008; Colombo *et al.* 2007). Mutation of the TF ARX leads to neuropathological diseases, such as, motor deficiency, mental retardation and epilepsy (Colasante *et al.* 2008; Nawara *et al.* 2006). According to Colasante *et al.* (2008), in a study of loss and gain of function models, Arx is essential for stimulation of Dlx-dependent interneuron migration but it is not required for GABAergic cell fate specification, which is regulated by the TF Dlx. Therefore, TF Arx is essential for migration of GABAergic neurons.

GAD2 is the enzyme synthesize the GABA neurotransmitter (Stühmer *et al*. 2002b). In this study, it was found that GAD2 expression, which is a Dlx target gene, was in accordance with the expression of exogenous or endogenous Dlx2. The same outcome was found in the study of embryonic mouse forebrain, where the pattern of Dlx2 expression was closely identical to the pattern of GAD2 expression (Stühmer *et al*. 2002b). In addition, in the gain of function study, it was found that when Dlx2 was ectopic expressed into coronal slices of embryonic mouse forebrain, more than 85% of the Dlx2 transfected cells expressed GAD2 (Stühmer *et al*. 2002b). However, the knock out studies of Dlx2, showed a decrease in GAD2 expression (Stühmer *et al*. 2002b; Petryniak *et al*. 2007; Long *et al*. 2009b). In addition, at early stages of embryogenesis, the ventral telencephalon expresses GAD1&2 and Dlx1/2/5&6, but not in the dorsal telencephalon (Stühmer *et al*. 2002b; Fode *et al*. 2000). When the tangential migration from subpallium to pallium take place, the expression of Dlx2 and Gad2 has similar expression in pallium (Stühmer *et al*. 2002b). Taken together, these data indicate that DLX2 has a fundamental role in regulating the expression of GAD2, hence stimulating the GABAergic neuronal phenotype. Its expression, along with ARX genes, indicates the successful differentiation of stem cells into striatal MSNs in this study.

The expression of these target genes, as indicated through the above research, clearly suggests the active role of TFs in proliferation, differentiation and cell fate commitment.

Endogenous DLX2 expression was affected by ectopic expression of other TFs cloned in the same expression vector.

The pCAGG-DLX2 and pCAGG-DLX2/MASH1, two days post nucleofection in nrNPCs, showed decreased exogenous DLX2 transgenic expression at ND3 in both the constructs, while the endogenous DLX2 expression increased from ND0 to ND3 and then remained constant from ND5 to ND7 in pCAGG-DLX2 construct only. Interestingly, when DLX2 and MASH1 were co-expressed, the endogenous DLX2 expression was increased sharply from ND3 to ND7. However in the same vector construct, exogenous MASH1 expressed until ND5, whereas endogenous expression of MASH1 increased slightly from ND5 to ND7, when co-expressed with DLX2. These findings strongly support the regulatory role of MASH1 in DLX2 expression. Corroborating these findings, the potential importance of Mash1 on Dlx2 expression has previously been reported in $Dlx^{-/-}$ mutant mice, where striatal histogenesis (but not that of the dLGE) was partially preserved due to expression of Mash1 (Long *et al*. 2009a; Long *et al.* 2009b). It can be speculated that the expression of MASH1 might maintain low expression of DLX2, thus preserving vLGE. In fact, it has previously been shown that MASH1 regulates the expression of DLX2 by binding to the I12b enhancer in the E-box sequences located upstream of DLX2 (Poitras *et al.* 2007).

The findings presented here also suggest that DLX2 represses MASH1, since the expression of endogenous MASH1 was lower in pCAGG-DLX2/MASH1 than pCAGG-MASH1 nucleofected cells. This is further supported by findings that DLX2-mediated down-regulation of Mash1 suppresses Notch downstream target genes (Dll1 and Hes5), leading to differentiation of nrNPCs (Anderson *et* al. 1997a, Yun *et al*. 2002). It has been also shown that Dlx1/2 causes progenitors to exit the cell cycle in order to begin differentiating into GABAergic neurons, particularly in the late stage of neurogenesis, as evidenced through increased expression of DRD2, GAD1/2 in the SVZ and mantle zone (MZ) (Anderson *et* al. 1997a, Yun *et al*. 2002, Cobos *et* al. 2007). This effect has been further supported in $Dlx1/2^{-/-}$ mutant mice, where upregulation of MASH1 and Mash1-mediated Notch signalling blocks differentiation (Yun *et al*. 2002). From this and previous studies, it has been concluded that coordination of Dlx2 and Mash1 plays a role in the balance between proliferation and differentiation, as well as in defining cell fate commitment to aid the development of the subpallial telencephalon (Yun *et al.* 2002; Castro *et al.* 2011). Moreover, it was shown that expression of Mash1 only promotes the subpallial progenitor state via Notch signaling, and co-expression of Mash1 with Dlx2 promotes subpallial differentiation

(Wang *et al.* 2013). It is believed that Mash1 is required in early neurogenesis, whereas Dlx2 is needed by late progenitors for differentiation and specification through repression of Notch signalling (Yun *et al.* 2002). Future studies of Notch ligand expression in cells ectopically expressing DLX2 and MASH1 together or on their own would facilitate a better understanding of the relationship between DLX2 and MASH1 in relation to the Notch signaling pathway.

All of the above findings clearly show that DLX2 and MASH1 co-expression plays an important role in neurogenesis and differentiation into LGE-like progenitor fate from hPSC-derived forebrain nrNPCs, which were required in this study for further differentiation into MSN-like cells.

The expression of EBF1 is regulated by ectopic expression of GSX2 and DLX2

One of the GSX2 target genes is EBF1. Expression of EBF1 mRNA was increased in pCAGG-DLX2/GSX2 nucleofected cells and reached its maximum at ND5. The expression of endogenous DLX2 and GSX2 also reached its peak at ND5. When MASH1 was co-expressed with DLX2 and GSX2, endogenous GSX2 mRNA expression started to increase after ND5, and it was observed that the EBF1 expression increased after ND5 in pCAGG-DLX2/MASH1/GSX2 nucleofected cells. From these findings, it was shown that GSX2 regulates the expression of EBF1, as previously reported by Wang and colleagues. They also reported that in the mouse homologous mutant Gsx2$^{-/-}$, expression of EBF1 was reduced compared to the wild type (Wang *et al.* 2013).

As described earlier in this chapter, it was found that at ND7, EBF1 expression was higher in pCAGG-DLX2/GSX2 nucleofected cells than in pCAGG-DLX2/MASH1/GSX2 nucleofected cells. This could be due to the differences in the endogenous expression of DLX2 in both pCAGG-DLX2/GSX2 and pCAGG-DLX2/MASH1/GSX2 nucleofected cells. The expression of endogenous DLX2 was higher in pCAGG-DLX2/GSX2 nucleofected cells than in pCAGG-DLX2/MASH1/GSX2 nucleofected cells. Therefore, it could be concluded that the expression of EBF1 was regulated by both DLX2 and GSX2. This finding is supported by $Dlx2^{-/-};Gsx2^{-/-}$ double mutant and single $Dlx2^{-/-}$ and $Gsx2^{-/-}$ homologous mutants mice models, which demonstrated that EBF1 expression was reduced in all mutants ($Dlx2^{-/-};Gsx2^{-/-}$, $Dlx2^{-/-}$ and $Gsx2^{-/-}$) compared to the wild type mouse (Wang *et al.* 2013).

Overall, these findings suggest that the complex molecular regulation seen in *in vitro* models successfully stimulates proliferation and differentiation of nrNPCs into LGE-like progenitors. Furthermore, in *in vitro* models, DLX2/MASH1 nucleofected cells seem to closely regulate the neurogenesis process to promote LGE-like progenitor fate, which can lead to further differentiation into MSNs.

Chapter 6: Direct programming of medium spiny neuron differentiation from hPSCs via ectopic expression of different combinations of the transcription factors DLX2, GSX2 and MASH1.

6.1 Introduction

Many TFs are known to have a parallel or in-series role in ventral telencephalic development and differentiation (Porteus *et al*. 1994; Anderson 1997b; Horton *et al.* 1999; Casarosa *et al*. 1999; Fode *et al*. 2000; Yun *et al*. 2002; Long *et al*. 2009b). One of these TFs is Dlx1/2, which acts in concert with Mash1 to specify GABAergic MSN differentiation and fate commitment in the LGE (Porteus *et al*. 1994; Yun *et al*. 2002; Long *et al.* 2009b). The distinct role of Dlx1/2 and Mash1 is further supported by the findings of reduced GABAergic interneurons in many mutant mice models, including Dlx1/2$^{-/-}$, Mash1$^{-/-}$, double Gsx2$^{-/-}$;Mash1$^{-/-}$ mutant and triple Dlx1/2$^{-/-}$;Mash1$^{-/-}$ mutant (Yun *et al*. 2002; Long, Swan, *et al*. 2009; Wang *et al*. 2009). For example, there is a partial preservation of vLGE identity, but a severe defect in the dLGE of Dlx1/2$^{-/-}$ mice, as evidenced by striatal markers such as Drd1/2 and Gad1/2 (Long *et al*. 2009b). This outcome indicates that Dlx1/2 has a profound role in dLGE development and the subsequent fate commitment of GABAergic neurons, whereas the development of the LGE's remaining parts is maintained by the expression of Mash1 and other TFs such as Gsx1/2 and Tlx (Long *et al*. 2009b). The parallel action of Dlx2 and Mash1 is further supported by the induced expression of ventral biomarkers Dlx1/2/5 and Gad2 when Mash1 is ectopically expressed in the dorsal telencephalon of Ngn$^{-/-}$ mutant mice (Fode *et al.* 2000; Wilson and Rubenstein 2000).

Furthermore, Dlx1/2 and Mash1 have been shown to regulate the complex proliferation, differentiation and maturation phases of neurogenesis (Anderson *et al*. 1997a; Horton *et al.* 1999; Stühmer *et al*. 2002a; Yun *et al*. 2002; Yoshihara e*t al*. 2005; Colombo *et al*. 2007; Cobos *et al*. 2007; Poitras *et al.* 2007; Long e*t al.* 2009b). More specifically, Mash1 drives the proliferative phase of neurogenesis, while Dlx1/2 triggers the differentiation process (Casarosa *et al*. 1999; Yun *et al*. 2002; Castro *et al.* 2011). Gsx2 is another important TF with a role in promoting the early identity of the telencephalic ventral domain (Hsieh-Li *et al*. 1995; Corbin *et al.* 2000; Toresson *et al.*

2000; Wang *et al*. 2013), as documented by a profound decrease in LGE size, reduction in Mash1, Dlx1/2, Ebf1 and Gad1/2 expression at E12.5, and subsequent reduction of DARPP-32^{+ve} MSNs in GSX2$^{-/-}$ mutant mice (Corbin *et al.* 2000; Toresson *et al*. 2000; Wang *et al.* 2013).

The critical role of TFs in neurogenesis has encouraged researchers to pursue the differentiation of mainly somatic cells into relatively high-yield induced neuronal cells (Berninger e*t al*. 2007; Heinrich *et al*. 2010; Yoo *et al*. 2011; Son *et al*. 2011; Lujan *et al*. 2012; Yang *et al*. 2013; Lau *et al*. 2014). In fact, quite a few studies have succeeded in differentiating human or mouse stem cells into induced neurons (Vierbuchen *et al*. 2010; Pang *et al*. 2012; Zhang *et al*. 2013). Interestingly, forced expression of different TF constructs, including Dlx2 and Neurog2 in cortical astroglial cells (non-neuronal cells), has resulted in a relatively high percentage of mature glutamatergic neurons (excitatory neurons) with Neurog2 alone (58%) and a low percentage of mature GABAergic neurons (inhibitory neurons) with Dlx2 alone (6%) (Heinrich *et al.* 2010). To date, however, a protocol for differentiating hPSCs into MSNs via forced expression of key ventral TFs (such as Dlx2, Mash1 and Gsx2) has not been attempted. Instead, ectopic expression of morphogens, such as SHH and DKK1, in hPSCs has been used to induce the indirect expression of key ventral TFs (such as Gsx2), which has then triggered differentiation into a low number of pure MSNs (Carri *et al.* 2013). Hence, it is necessary to identify an effective combination of TFs that will trigger the cascade of intracellular pathways leading to cell proliferation and differentiation into mature GABAergic MSNs.

To induce the differentiation of human stem cells into a specific neuron type, two stages of development are crucial. The first stage of development is the differentiation of human stem cells into naïve rosette-stage neural progenitors (nrNPCs), a process dependent on the addition of instructive factors, such as fibroblast growth factor (FGF), or inhibitors, such as bone morphogenetic protein-antagonists (BMPa) (Nat and

Hovatta 2004; Dhara and Stice 2008). The second stage is the differentiation of nrNPCs into a target neuron type (Martinez *et al.* 2012). This approach was recently used by Carri *et al.* (2013). Briefly, hPSCs were directly differentiated into ventral progenitors (FOXG1^{+ve} and GSX2^{+ve}) via dual inhibition of BMP/TGFβ, and then terminally differentiated upon treatment with sonic hedgehog/dickkopf WNT signaling pathway inhibitor 1 (SHH/DKK1) into functional MSNs. However, among these MSNs, few were DARPP-32^{+ve} neurons (20%) (Carri *et al*. 2013).

A similar approach is employed in this study. The major goal of this research is to successfully differentiate human pluripotent stem cells (hESCs- and iPSCs) into LGE-specific progenitors, as discussed in Chapter 5. Subsequently, these progenitors (H9 or 34D6 nrNPCs) will be differentiated further into MSNs through ectopic expression of DLX2, MASH1, GSX2, and DLX2/MASH1 with or without external GSX2 treatment, as these TFs are well known for playing a critical role in subpallial fate commitment and striatal development (Yun *et al*. 2002; Long *et al*. 2009a; Pauly *et al*. 2013) MSN fate commitment is examined using biomarkers that are widely utilized for detecting neurons, MSNs, and GABAergic neurons, as described in **Table 6.1**. Moreover, maturity is investigated using biomarkers **(Table 6.1)** and electrophysiological analysis. This approach will establish a cellular disease model using the end-point mature GABAergic MSNs to investigate molecular and cellular pathogenesis of HD and subsequently develop a new therapeutic intervention.

Table 6.1: Biomarkers to identify neuron cells and GABAergic striatal MSNs.

Biomarker / Target gene	Biomarker for
β-Tubulin III	Neurons
DARPP-32	Striatal MSNs
CTIP2	Striatal MSNs
EBF1	Striatal neuron (LGE SVZ and MZ)
FOXP1	Striatal neuron (LGE SVZ and MZ)
DRD1	Striatonigral neurons type of MSN
DRD2	Striatopallidal neurons type of MSN
GAD2	GABAergic neurons
CALBIN-1	GABAergic neuron

6.2 Aims

The objective of this chapter was to establishing a successful protocol for directly differentiating ventral telencephalon-like progenitor cells (described in Chapter 5) into the mature GABAergic MSN-like phenotype using novel differentiation media plus different TFs constructs. In addition, the maturity and functionality of GABAergic MSNs generated in *in vitro* was investigated using different means of assessment.

6.3 Experimental design

The protocol used for characterisation of TFs in the previous chapter (Chapter 5) was also used here. The experimental design conducted to maintain cells is illustrated in **Figure 6.1**. This method is described in greater detail in the Materials and Methods chapter (Chapter 3) (see **section 3.11, 3.12 & 3.13**).

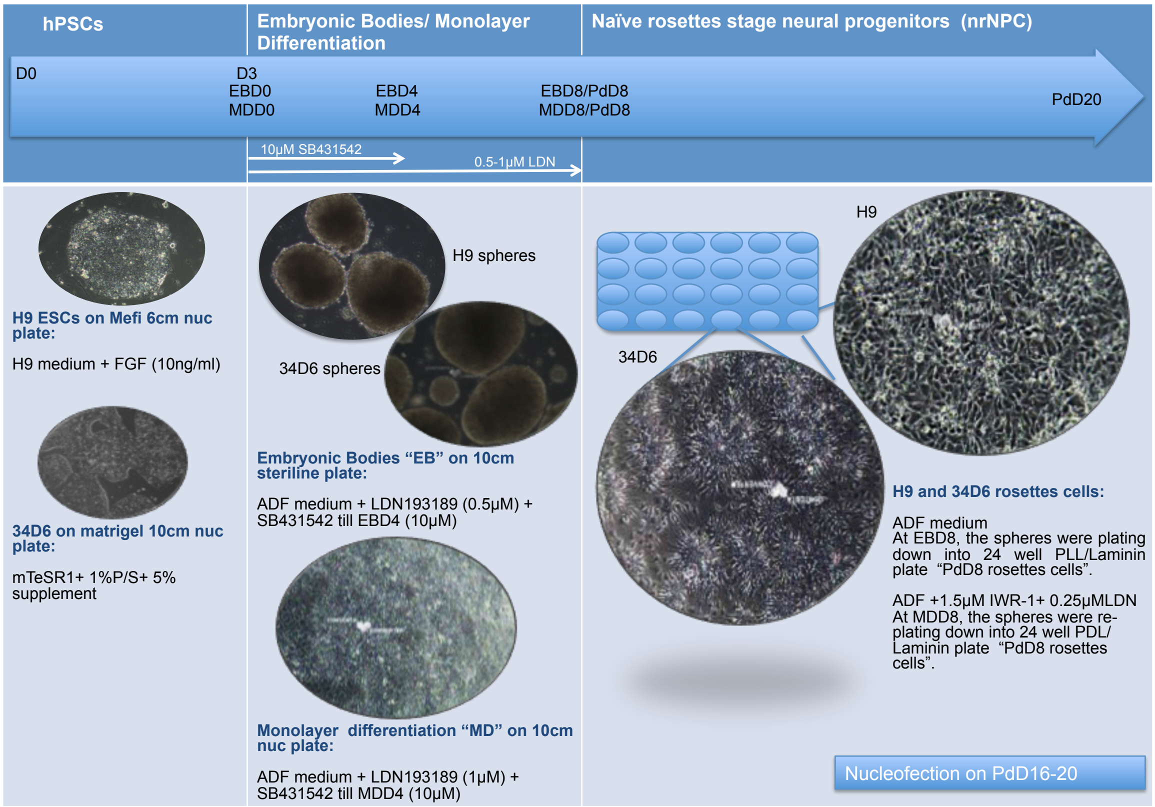


Figure 6.1: Experimental design.

The experimental protocol used in this chapter to maintain the human pluripotent stem cells (hPSCs) stage (D0) to naïve rosette stage neural progenitors (nrNPCs) at PD 20.

Abbreviations: D: Day, P/S: Pen/Strep, nrNPC: Naïve rosettes stage neural progenitors, EB: Embryonic body, EBD: embryonic body day, MD: Monolayer differentiation, MDD: Monolayer differentiation day; PdD8: Platting down Day 8.

The nucleofection of H9 or 34D6 nrNPCs with different TF expression vectors was initiated at PdD18. The cells were checked for GFP expression 48 hours after nucleofection; then selection of nucleofected cells was initiated with G418 sulfate. Cells that survived selection were washed with PBS and cultured in ADF medium for one day in order to recover. These cells were then re-plated onto treated cover slips in 24-well plates for differentiation experiments (described in **Section 3.2.7.1**). Cover slips were subjected to specific treatment, as described in the Materials and Methods chapter (**Section 3.2.7**), to ensure strong adherence and subsequent growth on the cover slips for downstream applications.

The nucleofected H9 or 34D6 nrNPCs were maintained in differentiation medium and allowed to differentiate for varying time periods. Three differentiation experiments were performed in this chapter, which are described briefly in **Table 6.2.** In each experiment, there were slight differences in the neural induction protocol, G418 selection, coating of cover slips and differentiation medium.

Table 6.2: The summary of the experimental design.

Experiment number	Human cell lines	Neural induction protocol	G418 selection	Coating on cover slips in 24-well plates	Differentiation medium
1	H9-derived nrNPCs (Nucleofected with pCAGG-DLX2/GSX2 and the control pCAGG)	Embryonic bodies (EB) (0.5µm LDN193189 + 10µM SB431542)	400µg/ml for 1 week	PLL/Laminin	ADF medium, 2% B27+A, 1% P/S, 1% L-G and 1% FBS.
2	34D6-derived nrNPCs (Nucleofected with one of five expression vectors or the control pCAGG)		800µg/ml for 1 day	Cover slips treated with nitric acids, washed and coated with PDL/Matrigel **(see section 3.2.7)**.	**Week 1:** ADF medium, 2% B27+A, 1% P/S, 1% L-G, 10ng/ml BDNF, 200µM AA, 0.5µM dbcAMP, 5µM DAPT and 0.5µM VPA. **Week2:** Addition of ACM to give an ADF:ACM ratio of 1:1, and 1mM $Ca^{+}Cl_2$.
Only the nucleofected cells that were mature DARPP-32^{+ve} neurons were used					
3	34D6-derived nrNPCs (Nucleofected with pCAGG-DLX2/MASH1 or the control pCAGG)	Monolayer differentiation (MD) with IWR-1 (WNT antagonist) (1µm LDN193189 + 10µM SB431542)	800µg/ml for 1 day	Cover slips treated with nitric acids, washed and coated with PDL/Matrigel **(see section 3.2.7).**	**Week 1:** ADF medium, 2% B27+A, 1% P/S, 1% L-G, 10ng/ml BDNF, 200µM AA, 10µM DAPT, 1.8mM $Ca^{+}Cl_2$, 2µM PD332991, 10µM Forskolin, 3µM CHIR99021, 300µM GABA. **Week 1:** (1:1) ADF medium: Neurobasal A, 2% B27+A, 1% P/S, 1% L-G, 10ng/ml BDNF, 200µM AA, 1.8mM $Ca^{+}Cl_2$, 3µM CHIR99021.

Abbreviations: B27+A: Serum-free supplement with retinoic acid, P/S: Pen/Strep, L/G: L/Glutamate, FBS: Fetal bovine serum, ACML: Astrocyte conditioned media, BDNF: Brain-derived neurotrophic factor, AA: Ascorbic acid, dbcAMP: Dibutyryl cyclic adenosine 3', 5'-monophosphate, DAPT: as Notch signaling inhibitor, chemical name is N-[(3,5-Difluorophenyl)acetyl]-L-alanyl-2-phenyl]glycine-1,1-dimethylethyl ester, VPA: Valproic acid, $Ca^{+}Cl_2$: Calcium chloride, PD332991: CDK4/6 inhibitor, Forskolin: elevates cAMP; CHIR99021: Glycogen synthase kinase 3 (GSK3) inhibitor and WNT agonist.

6.3.1 Strategy for the analysis of mature MSNs

The nucleofected nrNPCs were analysed at the following time points: week 0 (W0), which is equivalent to Differentiation day 0 (Dd0), W3 (Dd21) and W6 (Dd42) for 34D6 cell line and W0, W2, W4 and W6 for H9 cell line. RNA samples harvested at different time points and protein samples at Dd42 were analysed by qRT-PCR and western blotting, respectively. This was in order to examine DARRP-32 and CTIP2 mRNA, and DARPP-32 protein expression levels.

For the mRNA expression of DARPP-32, two-way ANOVA was used to test the following hypotheses:

H_0: Nucleofection of cells with different constructs has no significant effect on the expression of DARPP-32.

H_0: Incubation time with differentiation media has no significant effect on the expression of DARPP-32.

H_0: Nucleofection of cells with different constructs and the incubation time with differentiation media together have no significant effect on the expression of DARPP-32.

Cells expressing DARPP-32 and CTIP2 were then subjected to further analysis by qRT-PCR. The expression of some of the mature striatum markers, such as FOXPI and EBF1, dopamine receptors, such as DRD1 and DRD2 (Garel *et al.* 1999; Tamura *et al.* 2004; Martín-Ibáñez *et al*. 2012; Lobo *et al*. 2006), and markers of GABAergic neurons, such as GAD2 and CALBIN-1 (Kiyama *et al.* 1990; Gerfen 1992; Pickel and Heras 1996; Pinal and Tobin 1998; Pan 2012; Lin *et al.* 2015), was examined at different time points.

6.4 Results

6.4.1 Nucleofection of H9 nrNPCs with pCAGG-DLX2/GSX2 or the control pCAGG vector.

6.4.1.1 Increased expression of β-Tubulin III from W0 to W2 in pCAGG-DLX2/GSX2 nucleofected H9 nrNPCs.

H9 nrNPCs at PdD18 were nucleofected with pCAGG-DLX2/GSX2 or pCAGG, and maintained in differentiation media. These were then analysed for expression of the neuron specific marker, β-Tubulin III, by ICC at W0 and W2.

At W0, β-Tubulin III was expressed at low levels in both pCAGG and pCAGG-DLX2/GSX2 nucleofected H9 nrNPCs **(Figure 6.2)**. At W2, β-Tubulin III was highly expressed in both of pCAGG and pCAGG-DLX2/GSX2 nucleofected H9 nrNPCs **(Figure 6.2)**. In addition, the development of neuron morphology in both populations of H9 nucleofected cells was more advanced at W2 than W0, indicating that β-Tubulin III^{+ve} neuron cells are generated upon nucleofection of H9 nrNPCs **(Figure 6.2)**. The negative control for β-Tubulin III and GFP is shown in **Appendix 5.1.**

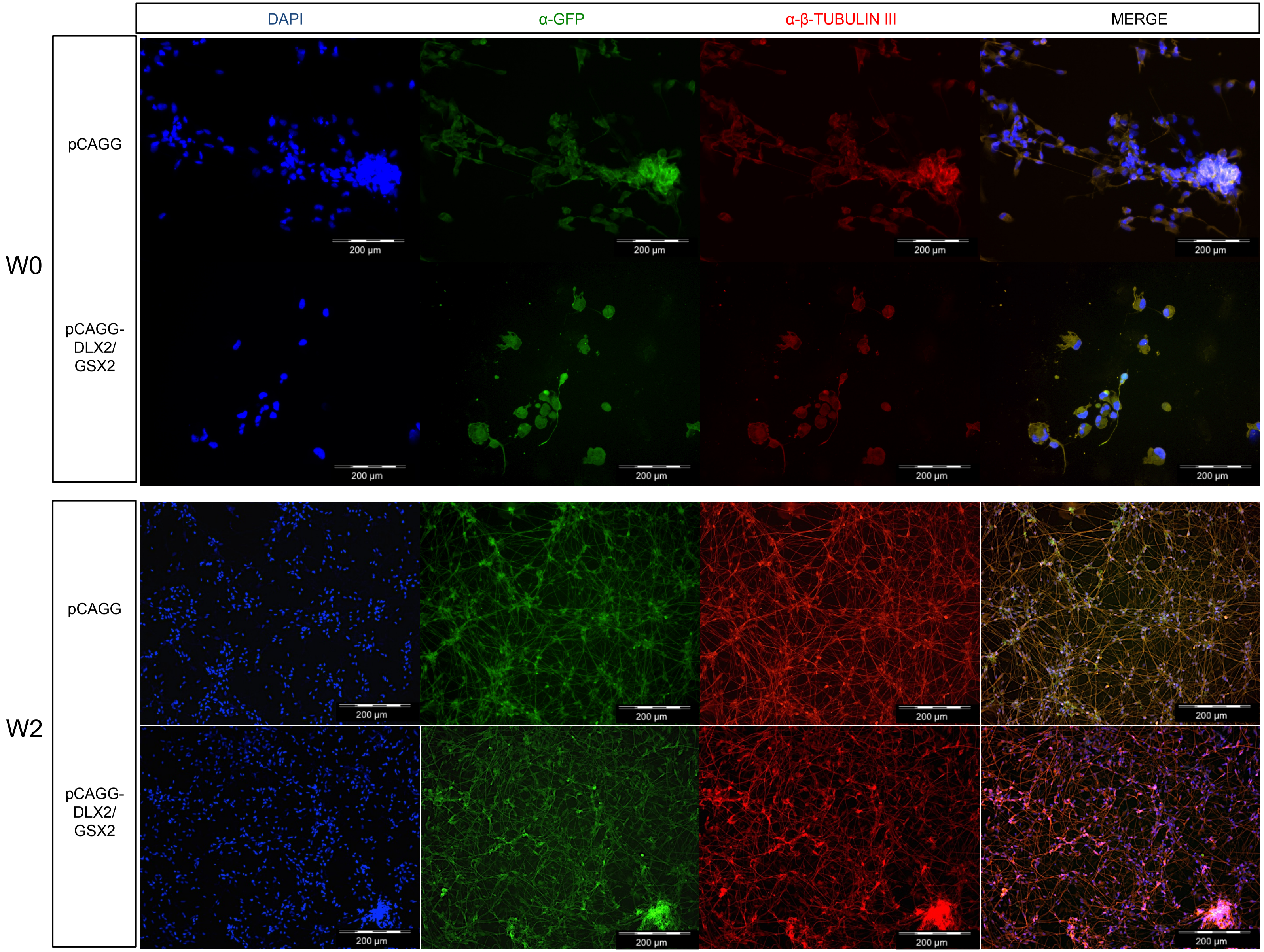

DAPI
α-GFP
α-β-TUBULIN III
MERGE
W0
pCAGG
pCAGG-
DLX2/
GSX2
W2
pCAGG
pCAGG-
DLX2/
GSX2
200 μm

Figure 6.2: β-Tubulin III expression in pCAGG and pCAGG-DLX2/GSX2 nucleofected H9 nrNPCs at W0 and W2 differentiation time points.

pCAGG and pCAGG-DLX2/GSX2 nucleofected H9 nrNPCs were stained with antibodies against β-Tubulin III (red) and DAPI (Blue). GFP expression (Green) was also shown and the micrographs were superimposed to visualize co-localisation of the different proteins. Scale bar indicates 200μm.

6.4.1.2 Failure of pCAGG-DLX2/GSX2 nucleofected H9 nrNPCs to generate mature MSNs despite increased expression of DARPP-32.

DARPP-32 expression in both pCAGG and pCAGG-DLX2/GSX2 nucleofected H9 nrNPCs increased dramatically and significantly from W2 to W6 at transcriptome level **(Figure 6.3 A)**. This gene expression corresponded with the increased DARPP-32^{+ve} immunoreactivity in ICC (**Figure 6.3 B and C)** and western blotting **(Figure 6.4)** at W6 relative to negative control (HEK293) and/or pCAGG. These results confirm the generation of MSNs, as evidenced by DARPP-32^{+ve} immunoreactivity at the transcriptome and protein levels.

Next, the electrophysiological characterisation was carried out to assess the functionality of the differentiated neurons. This was undertaken by Dr. Vsevalod Telezhkin, as described in the Materials and Methods chapter. Unfortunately, electrophysiological analysis of the pCAGG-DLX2/GSX2 nucleofected H9 nrNPCs indicated that DARPP-32^{+ve} MSNs were not functionally mature MSNs (data not shown). As a result, hPSC-derived nrNPCs were nucleofected with all the TF expression vectors (pCAGG-DLX2, pCAGG-DLX2/GSX2, pCAGG-DLX2/MASH1, pCAGG-DLX2/GSX2/MASH1, pCAGG-MASH1 plus the control pCAGG) in parallel in order to determine if any of them would differentiate into functionally mature DARPP-32^{+ve} MSNs. An iPSCs line, 34D6, was used for these experiments. This was performed in order to determine and obtain HD-iPSC lines, which consequently differentiate into mature MSNs for future experiments.

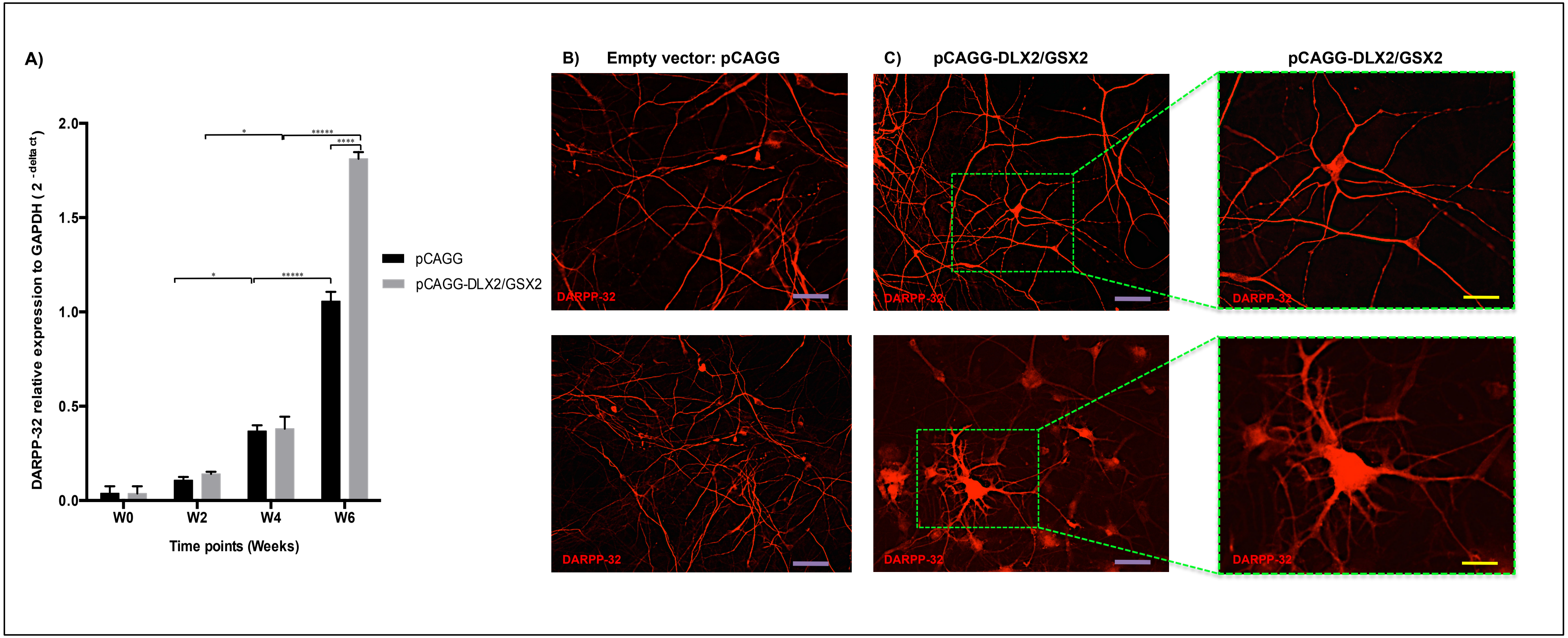


Figure 6.3: DARPP-32 mRNA expression and immunoreactivity in pCAGG and pCAGG-DLX2/GSX2 nucleofected H9 nrNPCs.

Empty vector (pCAGG) or pCAGG-DLX2/GSX2 nucleofected H9 nrNPCs at PdD18 were analysed for the expression of DARPP-32 using qRT-PCR at several time points (W0, W2, W4 and W6) (N = 2) (A). ICC analysis of pCAGG (B) and pCAGG-DLX2/GSX2 (C) nucleofected H9 cells at W6. Cells were stained with anti-DARPP-32 antibody. The micrographs and inserts were captured at 100μm (B and C) and 50μm magnifications (C). Purple scale bar indicates 100μm. Yellow scale bar indicates 50μm.

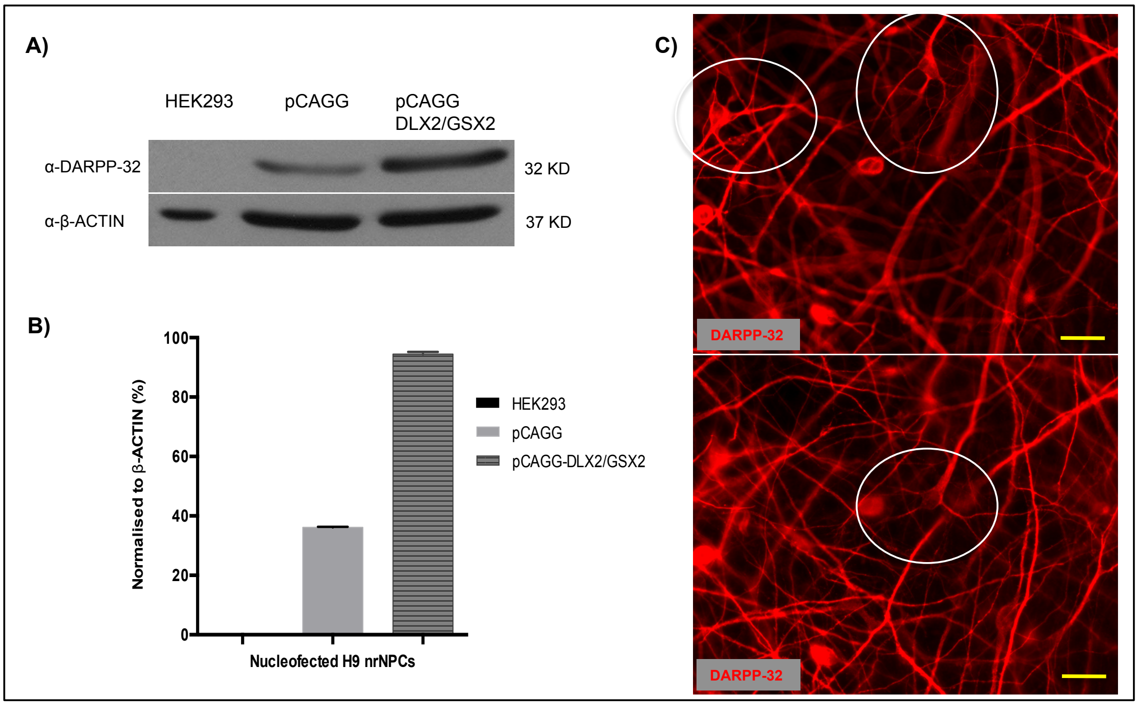


Figure 6.4: DARPP-32 protein expression in pCAGG and pCAGG-DLX2/GSX2 nucleofected H9 nrNPCs at W6.

Western blotting analysis of pCAGG and pCAGG-DLX2/GSX2 nucleofected H9 nrNPC along with negative control lysates (A). HEK293 cells were used as a negative control for DARPP-32 gene expression. DARPP-32 gene expression was normalised to β-ACTIN for each nucleofection and quantified as determined by densitometric analysis ($N = 2$, $p < 0.0001$) (B). Micrographs of pCAGG-DLX2/GSX2 nucleofected H9 nrNPCs stained with anti-DARPP-32 antibodies were captured at 50μm (C). The two panels show focusing on different areas (white circles) of the same image, to highlight morphology. Yellow scale bar indicates 50μm.

6.4.2 Nucleofection of 34D6 nrNPCs with different transcription factor expressing vectors.

34D6 nrNPCs were nucleofected at PdD18 with one of the following five constructs (pCAGG-DLX2, pCAGG-DLX2/MASH1, pCAGG-DLX2/GSX2, pCAGG-DLX2/MASH1/GSX2, pCAGG-MASH1) or a control vector (pCAGG). A large number of studies have used β-Tubulin III to detect neuron cells and double DARPP-32 and CTIP2 to confirm the presence of MSNs (Arlotta *et al.* 2008; Carri *et al.* 2013; Ding *et al.* 2014). Therefore, in this study, the nucleofected cells were maintained in the differentiation media **(as described in Table 6.2)** and examined for the expression of neuron specific marker, i.e. β-Tubulin III, and MSNs specific markers, i.e. DARPP-32 and CTIP2 at molecular and gene levels.

6.4.2.1 Ectopic expression of DLX2 and MASH1 promotes differentiation of iPSCs into DARPP-32^{+ve} and CTIP2^{+ve} functional MSNs.

All differently nucleofected vectors, including control, are expressing β-Tubulin III at W3(Dd21), confirming their neuronal identity as demonstrated in **Figure 6.5.** Expression of DARPP-32 and CTIP2, at the transcriptional level, was statistically significant in all the nucleofected 34D6 nrNPCs from W0(Dd0) to W6(Dd42) (**Figure 6.6 A and B**). Also, DARPP-32 gene expression correlated well to protein expression data obtained by Western blotting in all nucleofected 34D6 nrNPCs at W6(Dd42) **(Figure 6.7 A and B)**.

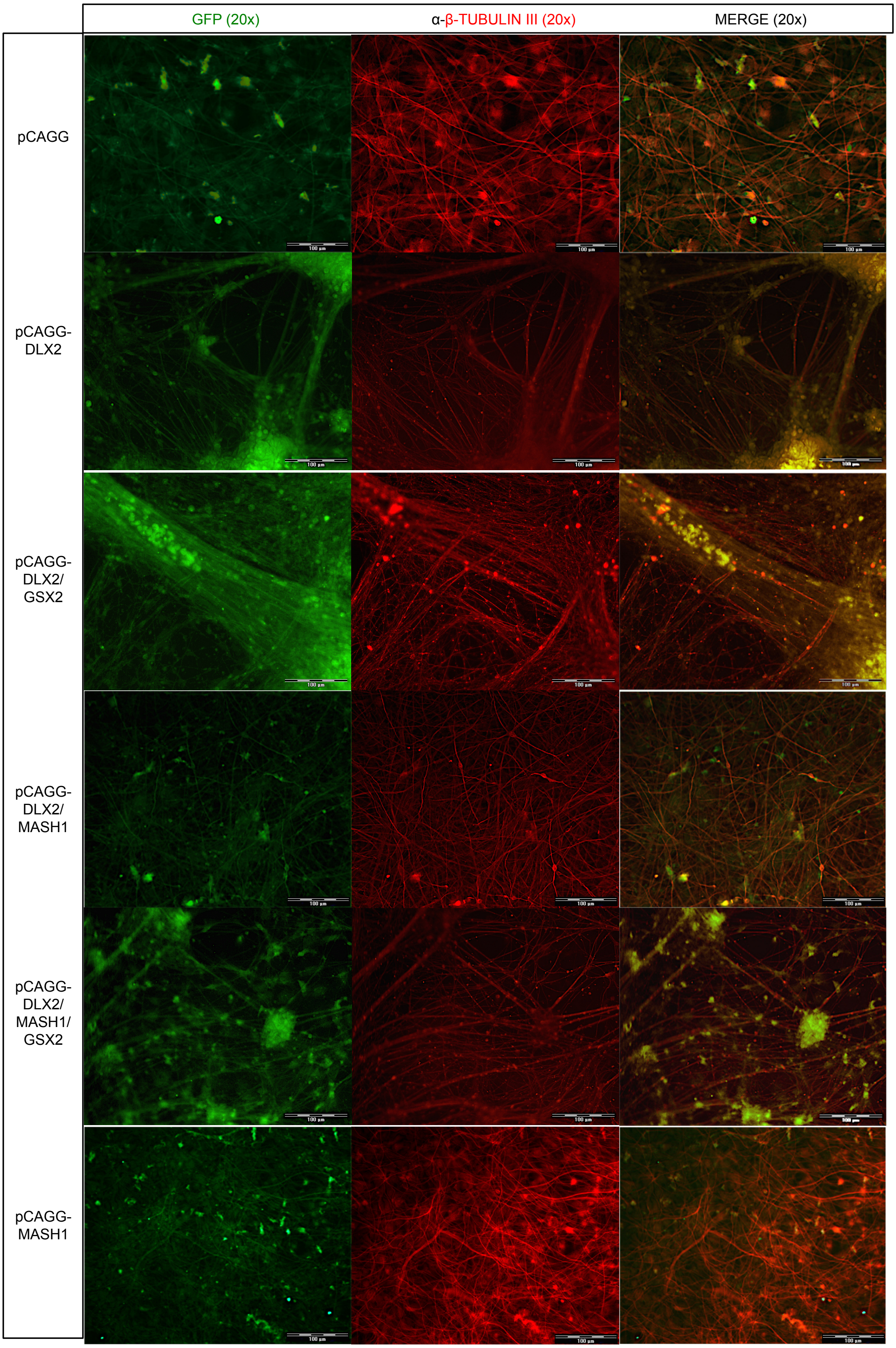

GFP (20x)
α-β-TUBULIN III (20x)
MERGE (20x)
pCAGG
pCAGG-DLX2
pCAGG-DLX2/GSX2
pCAGG-DLX2/MASH1
pCAGG-DLX2/MASH1/GSX2
pCAGG-MASH1

Figure 6.5: β-Tubulin III and GFP expression in differently nucleofected 34D6 nrNPCs at W3.

ICC images of 34D6 nrNPCs nucleofected with the indicated transcription factor expressing vectors (left column) and stained with antinuclear antibody (Dapi: Blue), and anti-β-Tubulin III (Red). GFP expression (Green) was also shown and the micrographs were superimposed to visualize co-localisation of the different proteins. Scale bar indicates 100μm.

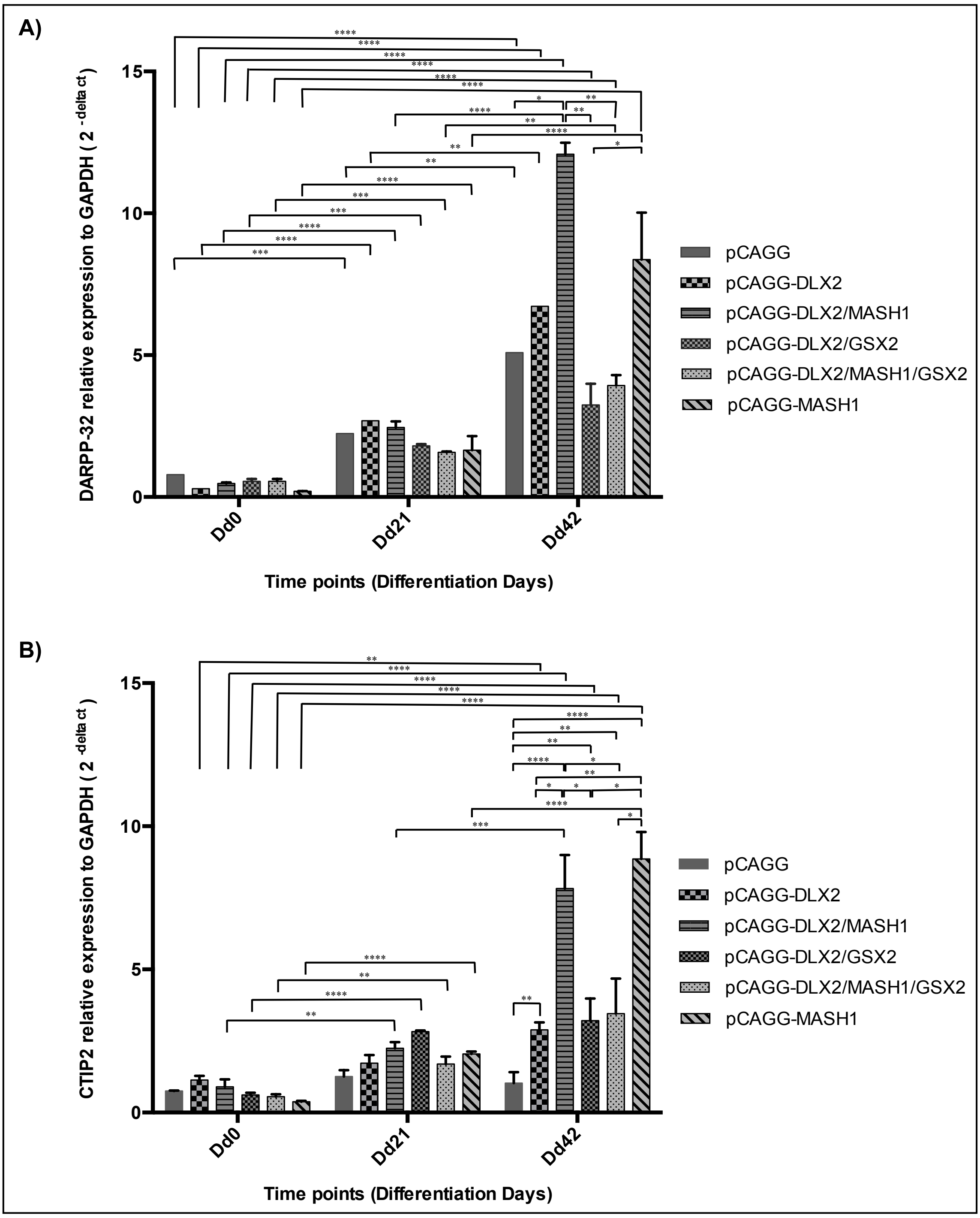


Figure 6.6: Expression of DARPP-32 and CTIP2 mRNA in nucleofected 34D6 nrNPCs at different time points.

Expression of DARPP-32 (A) and CTIP2 (B) in nucleofected 34D6 nrNPCs was assessed by qRT-PCR at various time points and normalised to GAPDH ($2^{-delta\ ct}$). p-value was calculated using Two-way ANOVA with Bonferroni correction. Summary of p-value: * = 0.01, ** = 0.0029, *** = 0.0001 and **** < 0.0001.

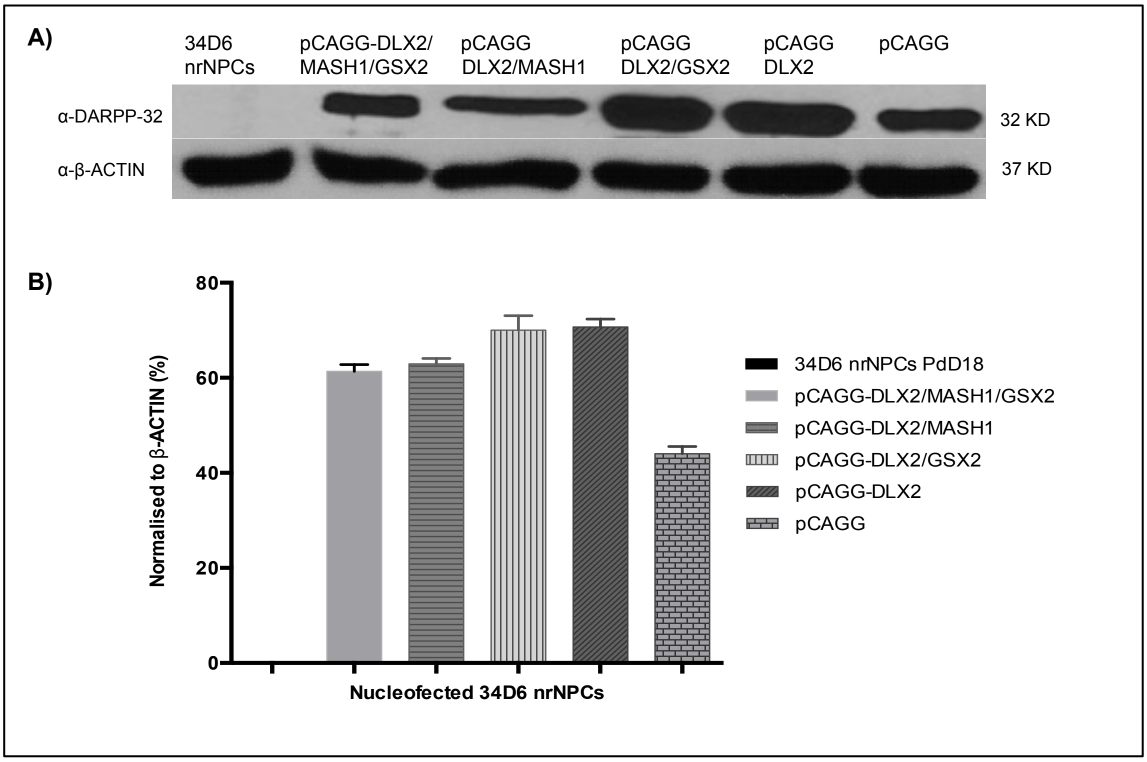


Figure 6.7: DARPP-32 protein expression at W6 in the nucleofected 34D6 nrNPCs. DARPP-32 expression in the nucleofected cells was assessed at W6(D42) by Western blotting (A). This was quantified and normalized to β-ACTIN for each nucleofection and quantified as determined by densitometric analysis (N = 2; $p < 0.0001$) (B).

Ectopic expression of different TFs in 34D6 nrNPCs resulted in expression of DARPP-32 and CTIP2. This effect, however, differed on a time dependent manner in differentiation medium. The expression level of DARPP-32 and CTIP2 in the different groups of nucleofected cells incubated for varying lengths of time in differentiation medium, was subjected to ANOVA analysis. There was a significant difference in the expression level of DARPP-32 and CTIP2 among different groups of nucleofected 34D6 nrNPCs (p-value for DARPP-32 = 0.0029 and p-value for Ctip2 = 0.0001), and between cells incubated for different lengths of time ($p < 0.0001$ for both markers). Moreover, the interaction of 34D6 nrNPC nucleofection group and incubation time was significant for both DARPP-32 and Ctip2 ($p \leq 0.0001$).

The expression level of DARPP-32 and CTIP2 increased noticeably in all nucleofected 34D6 nrNPCs from W0(Dd0) to W6(Dd42) ($p < 0.0001$) compared to pCAGG nucleofected 34D6 nrNPCs **(Figure 6.6 A and B)**. The expression of DARPP-

32 was increased in the 34D6 nrNPCs nucleofected with pCAGG-MASH1, pCAGG-DLX2 and pCAGG-DLX2/MASH1 at W6(Dd42). The pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs were the only group that showed the highest dramatic increase in DARPP-32 expression compared with the control (pCAGG) nucleofected cells at W6(Dd42) ($p = 0.01$). It was also shown that there was a significant increase in DARPP-32 expression between Dd21 and Dd42 ($p < 0.0001$) **(Figure 6.6 A)**. Similarly, CTIP2 expression was significantly increased at W6 (Dd42) in the pCAGG-MASH1, pCAGG-DLX2 and pCAGG-DLX2/MASH1 nucleofected cells, but only when compared to the other nucleofected 34D6 nrNPCs ($p < 0.0001$) **(Figure 6.6 B)**. In addition, electrophysiological studies showed that only the pCAGG-DLX2/MASH1 nucleofected cells differentiated into more neuron-like cells that generated spontaneous action potentials from W3 (Dd21) (Experiment conducted by Dr. Vsevolod Telezhkin, a postdoc in the neuroscience physiology group at Cardiff University) **(Figure 6.8 A and B)**. The electrophysiology showed the development of the resting membrane potential, which dropped from -30 to -40 over time, indicating the progression towards more neuron-like cells **(Figure 6.8 A and B)**. *In vivo*, the resting membrane potential is -70, and hence the neurons were not fully mature. However, they were progressing in the right direction.

Together, these results indicate that among all the vectors tested, only pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs are associated with increased DARPP-32 and CTIP2 expression at the molecular and protein levels. Furthermore, the DLX2/MASH1 combination generated functional and mature MSNs that elicited trains of action potentials **(Figure 6.8 A and B).** Therefore, the pCAGG-DLX2/MASH1 nucleofected cells were subjected to further analysis for studies of mature striatal MSNs.

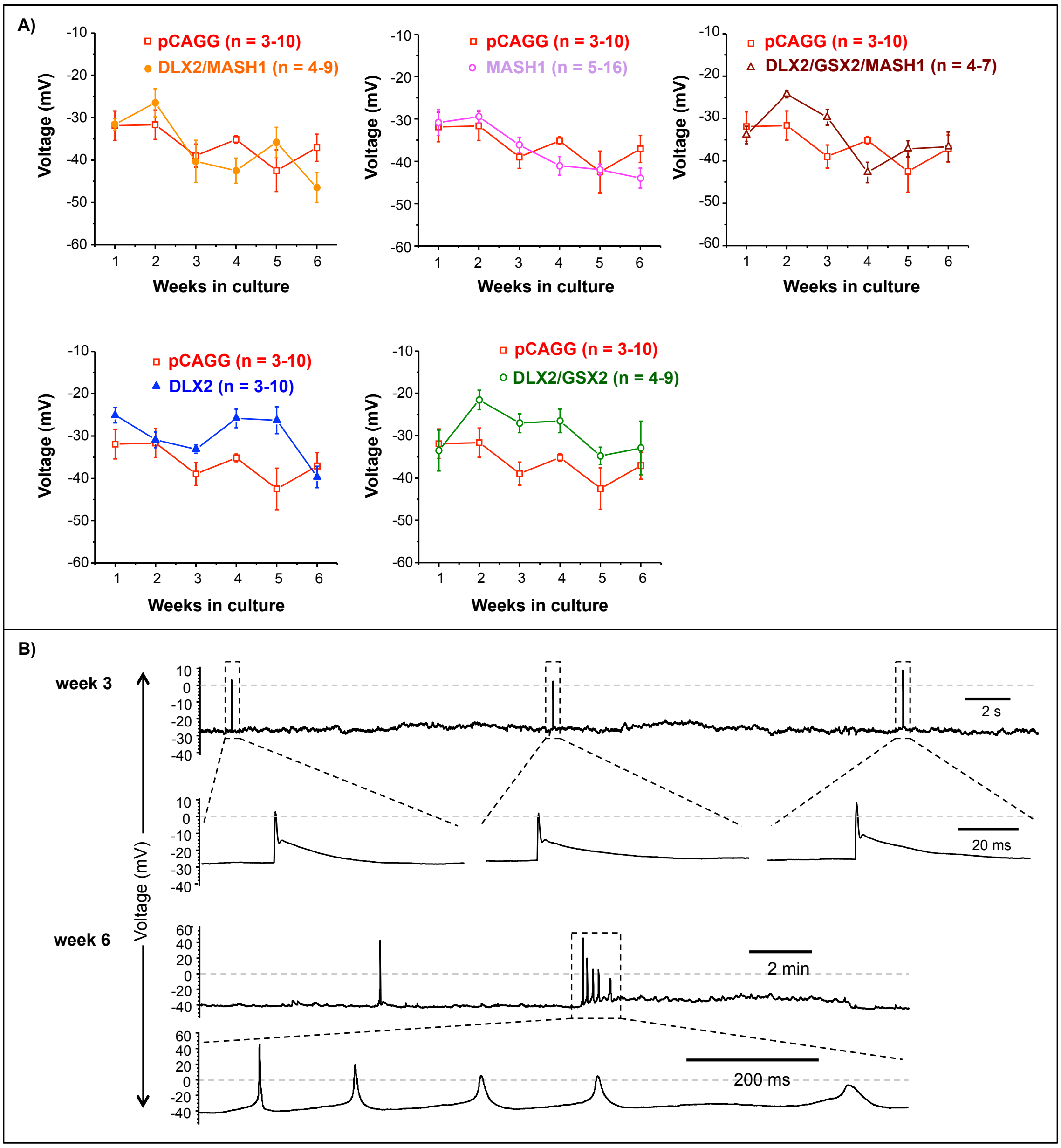


Figure 6.8: Development of membrane potential in 34D6 nrNPCs expressing different combinations of TFs

The membrane potential of TF expressing (or control) vector nucleofected 34D6 cells incubated for varying lengths of time in differentiation media (A). In pCAGG-DLX2/MASH1 nucleofected cells, the development of the resting membrane potential over time is dropping from -30 to -40. pCAGG-DLX2/MASH1 nucleofected cells show spontaneous activity at week 3 and week 6 (B).

6.4.2.2 Increased DARPP-32 and CTIP2 immunoreactivity in pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs

pCAGG and pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs, at W6(Dd42), were double stained with anti-DARPP-32 and anti-CTIP2 and analysed using ICC **(Figure 6.9 A and B)**. Around 80% of the pCAGG-DLX2/MASH1 nucleofected cells were DARPP-32^{+ve}; however, only 12% of the control (pCAGG) nucleofected cells were

DARPP-32^{+ve} **(Figure 6.9 C)**. Meanwhile, 85% of the pCAGG-DLX2/MASH1 nucleofected cells, and 69% of the pCAGG nucleofected cells, were CTIP2^{+ve} **(Figure 6.9 B)**. Most of the pCAGG-DLX2/MASH1 nucleofected cells were both DARPP-32^{+ve} and CTIP2^{+ve} as shown in the confocal microscope images of DARPP-32 and CTIP2 double staining **(Figure 6.9 B)**. This result support MSNs phenotype in concordance to previous results in **section 6.4.2.1**.

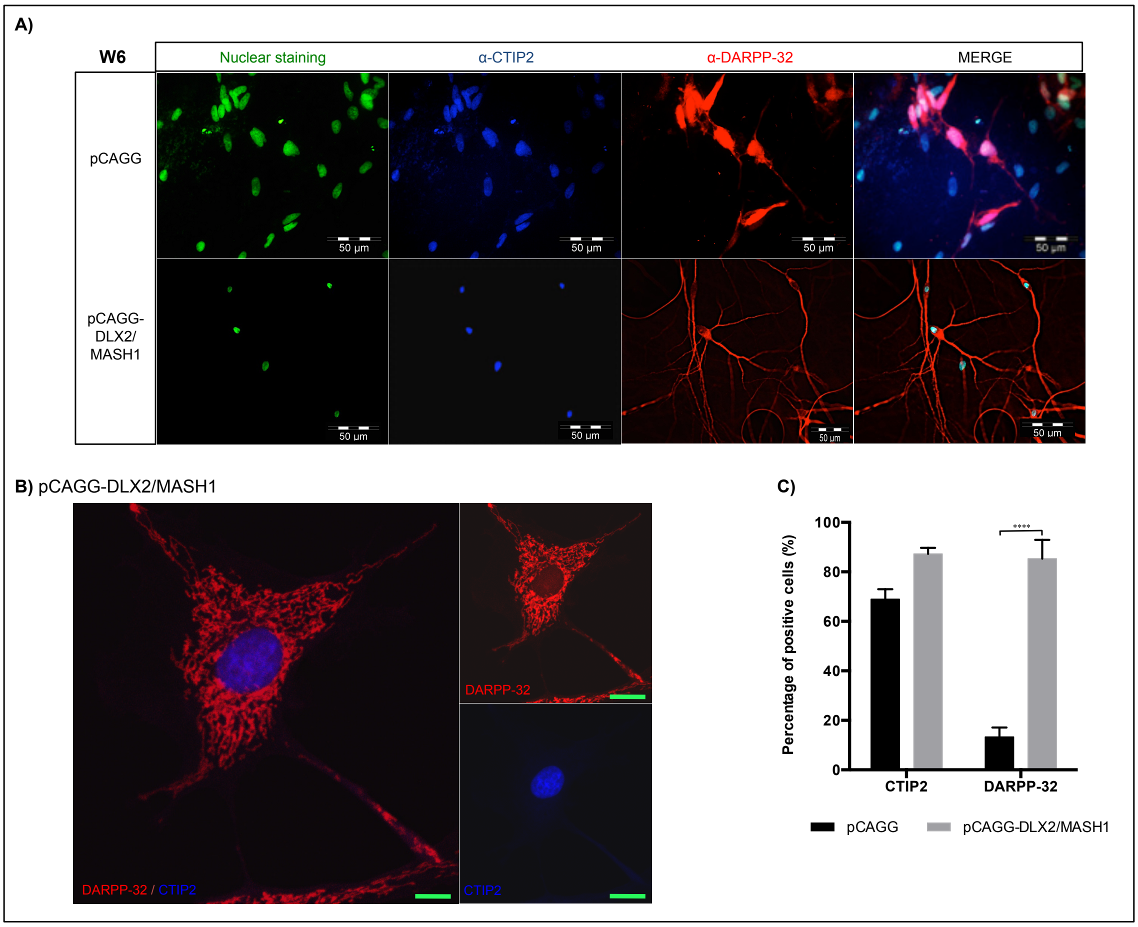


Figure 6.9: Expression of DARPP-32 and CTIP2 at W6 in pCAGG and pCAGG-DLX2/MASH1 nucleofected cells.

pCAGG and pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs were labelled with anti-DARPP-32 antibodies (primary dilution; 1:100, Santa Cruz) and detected with an Alexa Fluor® 594 labelled anti-rabbit IgG secondary antibody (red), and anti-CTIP2 antibody (primary dilution; 1:500, Abcam) and detected with an Alexa Fluor® 350 anti-Rat IgG secondary antibody (blue) and Nuclear staining with SYTOX green fluorescent counterstain (primary dilution; 1:300, Invitrogen. Cat. No. S33025) (A). Confocal microscopy images of pCAGG-DLX2/MASH1 nucleofected cells double stained against anti-DARPP-32 antibody (red) and anti-CTIP2 antibody (blue) (B). Green scale bar indicates 36µm. The quantification of DARPP-32 and CTIP2 expression in pCAGG and pCAGG-DLX2/MASH1 nucleofected cells was measured by cell profiler program (N = 3) (C).

6.4.2.3 Increased gene expression of FOXP1, EBF1, DRD1 and DRD2 in pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs provides an evidence of mature striatal MSNs.

In the terminal differentiation phase, FOXP1 and EBF1 are TFs that are expressed by MSNs (Arlotta *et al.* 2008). These were used as markers to identify LGE determinants, as they are expressed in the SVZ and MZ of the LGE (Long *et al.* 2009b), and are also known as markers for striatal projection neurons (Martín-Ibáñez *et al.* 2012). Therefore, these markers were to assess the maturity of the MSNs. There was a gradual increase in FOXP1 and EBF1 mRNA expression from Dd0 to Dd21 and from Dd0 to Dd42 in pCAGG-DLX2/MASH1 nucleofected cells when compared to control (pCAGG) **(Figures 6.10 and 6.11)**. This increase was statistically significant for FOXP1 expression ($p < 0.01$) and for EBF1 expression ($p < 0.001$). No differences between the expression of FOXP1 and EBF1 from Dd21 to Dd42 were observed. The expression of both FOXP1 and EBF1 in pCAGG-DLX2/MASH1 nucleofected cells is indicative of mature MSNs.

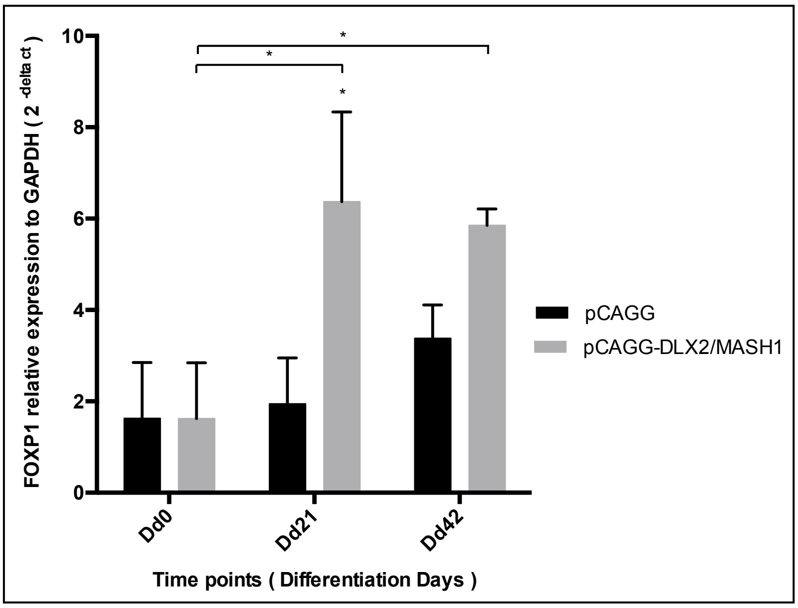


Figure 6.10 FOXP1 mRNA expression in pCAGG-DLX2/MASH1 and pCAGG nucleofected 34D6 nrNPCs at different time points.

Expression of FOXP1, relative to GAPDH ($2^{-delta\ ct}$), in pCAGG and pCAGG-DLX2/MASH1 nucleofected cells, was measured by qRT-PCR. The start at Dd21 indicates the significance differences between pCAGG-DLX2/MASH1 nucleofected cells and the control at this time point. Statistical test was performed using Two-way ANOVA with Bonferroni correction. p-value summary: * = 0.01.

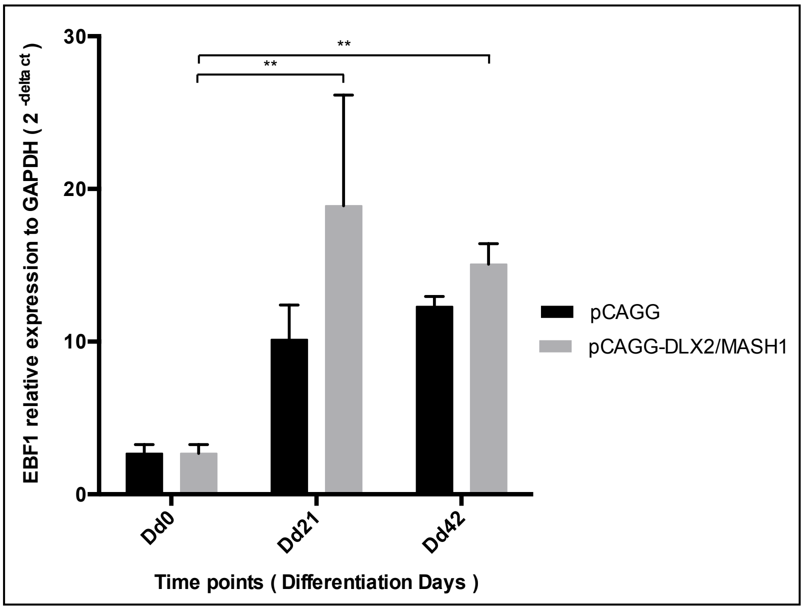


Figure 6.11: EBF1 mRNA expression in pCAGG-DLX2/MASH1 and pCAGG nucleofected 34D6 nrNPCs at different time points.

Expression of EBF1, relative to GAPDH ($2^{-delta\ ct}$), in pCAGG and pCAGG-DLX2/MASH1 nucleofected cells, was measured by qRT-PCR. Statistical test was performed using Two-way ANOVA with Bonferroni correction. p-value summary: ** = 0.001.

pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs expressed DRD1 and DRD2 **(Figure 6.12 A and B)**. Compared to control (pCAGG) nucleofected cells, there was a significant increase in the expression of both of DRD1 ($p = 0.0282$) and DRD2 ($p < 0.0001$) in pCAGG-DLX2/MASH1 nucleofected cells. The difference in DRD1 and DRD2 expression at different time points was also significant ($p = 0.0073$ and $p < 0.0001$, respectively). In addition, the interaction between nucleofection of cells and incubation time had a significant effect on DRD1 and DRD2 expression ($p = 0.0367$ and $p < 0.0001$, respectively). Moreover, expression of both of DRD1 and DRD2 in pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs, relative to GAPDH ($2^{-\text{delta ct}}$), increased sharply from W0(Dd0) to W3(Dd21), and these increases were statistically significant ($p = 0.001$ and $p < 0.001$ for DRD1 and DRD2 respectively) **(Figure 6.12 A and B)**. This was followed by a gradual decrease in DRD1 and DRD2 expression from W3(Dd21) to W6(Dd42) **(Figure 6.12 A and B)**. However, compared to W0(Dd0), the expression of both DRD1 and DRD2, at W6(Dd42), was increased. This increase was statistically significant for DRD2 ($p = 0.01$) **(Figure 6.12 A and B)**. At Dd21, the expression of both of DRD1 and DRD2 was significantly elevated in pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs compared to control (pCAGG) nucleofected 34D6 nrNPCs ($p = 0.001$ and $p < 0.001$ for DRD1 and DRD2 respectively) **(Figure 6.12 A and B)**. Data also showed that the expression of DRD2, at Dd21, was higher than the expression of DRD1 in the pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs **(Figure 6.12 A and B)**. The expression of DRD1 and DRD2 appears when neuron cells become mature striatum as it is known as neurochemical characteristic of endogenous MSNs (Lobo *et al.* 2006; Carri *et al.* 2013). This expression is mediated by DLX2 as evident in $Dlx1/2^{-/-}$ mutant mice (Long *et al*. 2009a). Together, these data indicate that the overexpression of DLX2 and MASH1 in 34D6 nrNPCs generates $DRD1^{+ve}$ and $DRD2^{+ve}$ mature MSNs by W3 (Dd21) via increased expression of FOXP1 and EBF1.

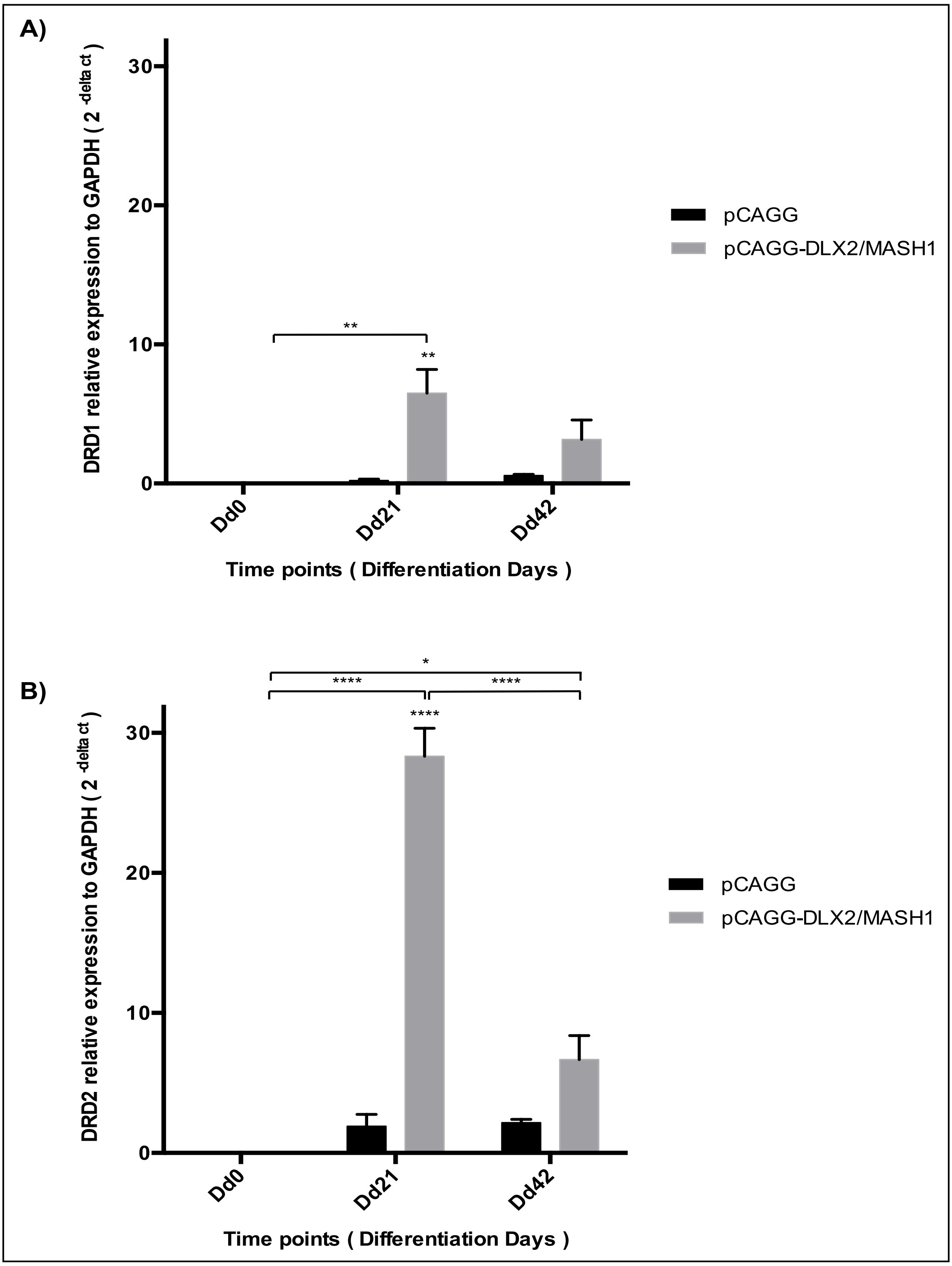


Figure 6.12: DRD1 and DRD2 mRNA expression in pCAGG-DLX2/MASH1 and pCAGG nucleofected 34D6 nrNPCs, at different time points.

Expression of DRD1 (A) and DRD2 (B) dopamine receptors, relative to GAPDH ($2^{-\text{delta ct}}$), in pCAGG and pCAGG-DLX2/MASH1 nucleofected cells. Statistical test was performed using Two-way ANOVA with Bonferroni correction. p-value summary: * = 0.01, ** = 0.001, and **** < 0.0001.

6.4.2.4 Characterisation of mature GABAergic MSNs through the CALBIN-1 and GAD2 expression

The development of GABAergic neuron, which is the major cell type in the striatum, is induced by glutamic acid decarboxylases enzyme (GADs). Hence, its expression is a specific calibrate of GABAergic neuron (Pinal and Tobin 1998; Pan 2012). Another widely used marker for GABAergic neuron is CALBIN-1 (Kiyama *et al.* 1990; Gerfen 1992; Pickel and Heras 1996; Lin *et al.* 2015). In this study, pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs exhibited GABAergic properties, such as expression GAD2 and CALBIN-1 **(Figures 6.13 and 6.14)**. This effect was influenced by the incubation time in differentiation medium. The ANOVA analysis of variance tested the significance of differences between the expression level of both of GAD2 and CALBIN-1 in pCAGG-DLX2/MASH1 and pCAGG nucleofected cells harvested after different incubation times in differentiation media. There was a significant difference in the expression level of GAD2 and CALBIN-1 between pCAGG and pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs ($p = 0.0005$ and $p = 0.036$ respectively), and between the incubation times ($p = 0.0002$ and $p = 0.0308$ respectively). Moreover, the interaction of nucleofected 34D6 nrNPCs and incubation time was significant only for GAD2 expression ($p = 0.0013$).

Expression of both GAD2 and CALBIN-1 was elevated from Dd0 to Dd21 ($p = 0.0001$). Their expression declined gradually from Dd21 to Dd42. However, compared to Dd0, the increase in expression of both GAD2 and CALBIN-1, at Dd42, was statistically significant ($p = 0.001$) **(Figures 6.13 and 6.14)**. The increase in expression of GAD2, in pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs, compared to control cells, at Dd21 and at Dd42, was statistically significant ($p = 0.0001$ and $p = 0.01$ respectively) **(Figure 6.13)**. CALBIN-1 expression in pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs was significantly increased when compared to control cells, at Dd21 ($p = 0.0001$) **(Figure 6.14)**.

Interestingly, dopamine receptor expression in pCAGG-DLX2/MASH1 and pCAGG nucleofected cells, at different incubation times, correlated with GABAergic phenotypic expression. Therefore, it can be concluded that co-expression of DLX2 with MASH1 in 34D6 nrNPCs at Pd18 has the ability to differentiate progenitor neurons into functional and mature GABAergic MSNs that express all the relevant markers (i.e. DARPP-32, CTIP2, FOXP1, EBF1, DRD1, DRD2, GAD2 and CALBIN-1), from Dd21 onward.

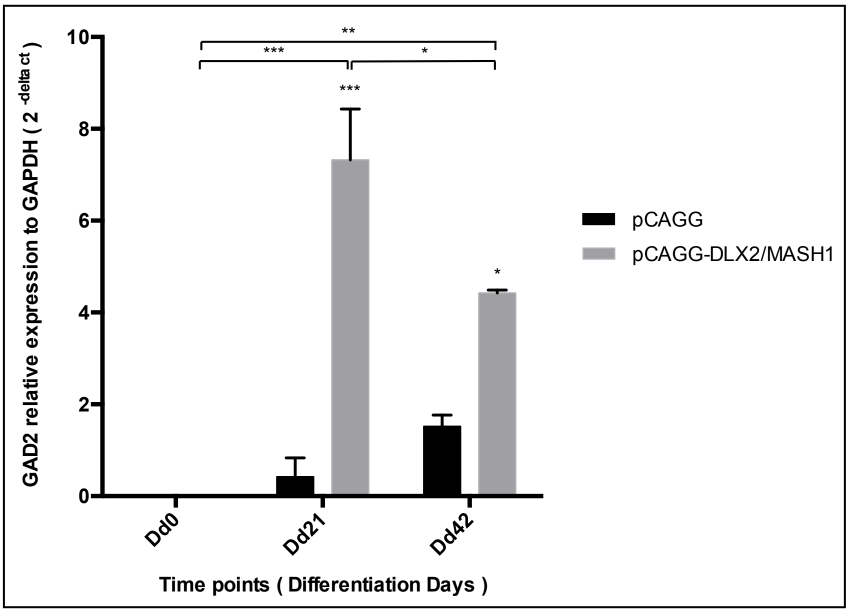


Figure 6.13: GAD2 mRNA expression in pCAGG-DLX2/MASH1 and pCAGG nucleofected 34D6 nrNPCs at different time points.

Expression of GAD2, relative to GAPDH ($2^{-\text{delta ct}}$), in pCAGG and pCAGG-DLX2/MASH1 nucleofected cells was measured by qRT-PCR. Statistical test was performed using Two-way ANOVA with Bonferroni correction. p-value summary: * = 0.01, ** = 0.001, and *** = 0.0001.

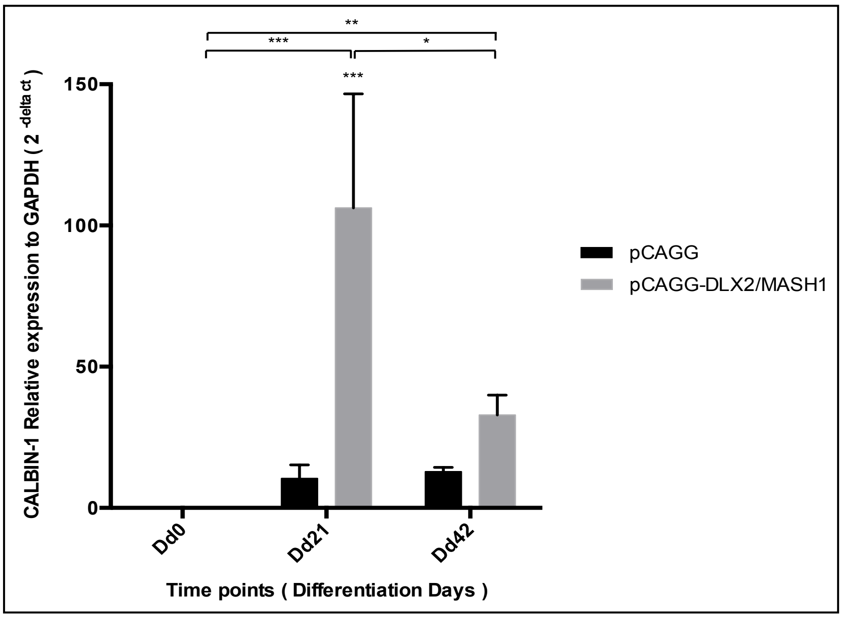


Figure 6.14: CALBIN-1 mRNA expression in pCAGG-DLX2/MASH1 and pCAGG nucleofected 34D6 nrNPCs at different time points.

Expression of CALBIN-1, relative to GAPDH ($2^{-\text{delta ct}}$), in pCAGG and pCAGG-DLX2/MASH1 nucleofected cells was measured by qRT-PCR. Statistical test was performed using Two-way ANOVA with Bonferroni correction. p-value summary: * = 0.01, ** = 0.001, and *** = 0.0001.

6.4.3 IWR-1 pre-treated 34D6 nrNPCs induced GSX2 upon nucleofection of pCAGG-DLX2/MASH1 leads to direct programming of functional striatal GABAergic MSN-like cells.

It had been observed previously that treatment of iPSCs with the small molecule IWR-1 at PdD8 resulted in expression of GSX2 in the iPSC-derived nrNPCs (unpublished data from the Allen lab, Cardiff University, Cardiff, UK). Having demonstrated that ectopic expression of DLX2 and MASH1 in 34D6 nrNPCs leads to their differentiation into mature MSNs; as illustrated earlier, it is investigated further whether adding and external GSX2 in these cells would affect cell differentiation.

34D6 nrNPCs were initially treated with IWR-1, followed by combined DLX2/MASH1 and control (pCAGG) nucleofection. These pCAGG-DLX2/MASH1 and pCAGG nucleofected 34D6 nrNPCs were maintained in differentiation media containing PD332991, a CDK4/6 inhibitor, to inhibit cell proliferation **(Table 6.1)**, in order to sustain the expression of GFP^{+ve} in 34D6 nrNPCs. A GFP^{+ve} population was clearly observed after G418 selection in both, pCAGG-DLX2/MASH1 and pCAGG nucleofected 34D6 nrNPCs at week two differentiation **(Figure 6.15)**, indicating the ectopic expression of integrated DLX2 and MASH1 in 34D6 nrNPCs at W2(Dd14). The differentiated cells were then examined for functionality (electrophysiology) and the presence of neuronal and its maturation markers, as described previously.

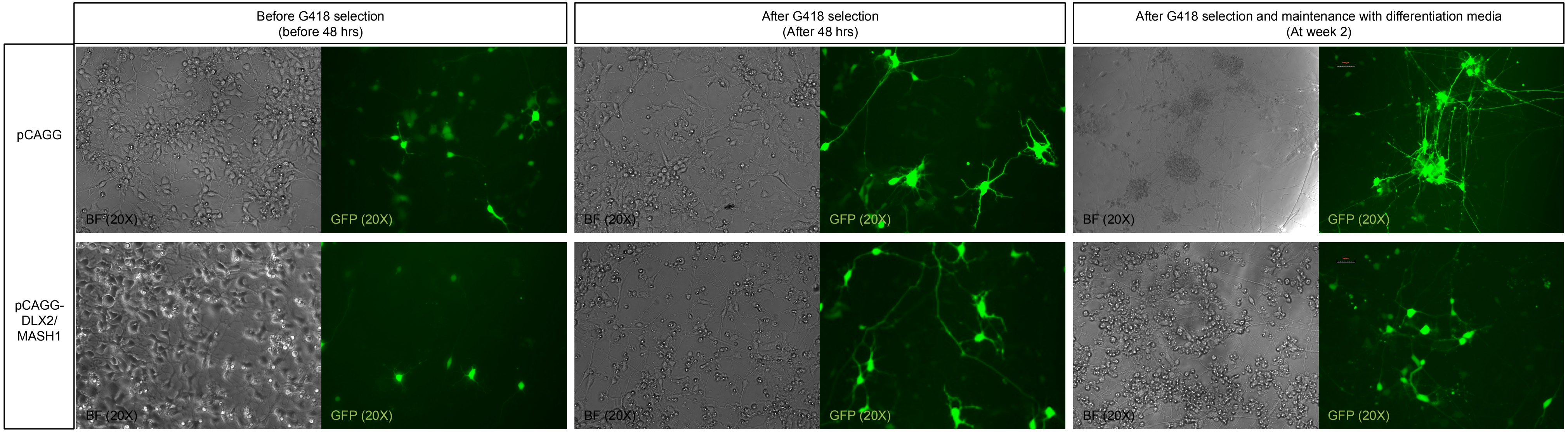


Figure 6.15: GFP expression in IWR-1-treated, pCAGG and pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs before and after G418 selection and at W2 of differentiation.

GFP-expressing IWR-1-treated pCAGG-DLX2/MASH1 or pCAGG nucleofected 34D6 cells were visualized by light microscopy both before and after G418 selection and at week two of differentiation (W2). The grey micrographs show the bright field (BF) and the green fluorescence is GFP. Scale bar indicates 100μm.

Ectopic expression of DLX2 and MASH1 directed the differentiation of IWR-1-treated 34D6 nrNPCs towards MSN phenotype expressing DARPP-32 and CTIP2 at molecular level **(Figures 6.16 and 6.17)**. In addition, ICC analysis showed that the differentiated cells were both DARPP-32^{+ve} and CTIP2^{+ve} **(Figure 6.17 A)**. DARPP-32 was expressed in 20% of pCAGG-DLX2/MASH1 nucleofected cells compared to 12% of pCAGG nucleofected cells. CTIP2 was expressed in 48% of pCAGG-DLX2/MASH1 nucleofected cells and in 52% of pCAGG nucleofected cells **(Figure 6.17 B).**

Importantly, IWR-1-treated, pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs showed increased expression of LGE-derived mature striatal GABAergic MSNs markers at Dd21, i.e. FOXP1 and EBF1 compared to the control (pCAGG nucleofected cells) group as demonstrated **in Figures 6.18 and 6.19**. In addition, the phenotype of LGE-derived mature striatal GABAergic MSNs was also supported through the expression of other specific markers including DRD1, DRD2 **(Figure 6.20 A and B)**, GAD2 **(Figure 6.21)** and CALBIN-1 **(Figure 6.22)**. The functionality and maturity of these GABAergic MSNs was also validated through the electrophysiological analysis, which showed a high percentage of functional neurons **(Figure 6.23).** The expression of GSX2 was observed in IWR-1-treated 34D6 cells (data not shown). It was found that expression of GSX2 in pCAGG or pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs re-programmed their differentiation such that 100% of these differentiated cells were capable of generating spontaneous action potentials starting at week 4 **(Figure 6.23)**. The IWR-1-containing differentiation media improved neuronal maturation dramatically in pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs. A higher proportion of pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs (68.4%), compared to control (pCAGG) nucleofected 34D6 nrNPCs (46.2%), had differentiated into functionally active neurons, capable of generating spontaneous action potentials, at week 3 **(Figure 6.23)**. In pCAGG-DLX2/MASH1 nucleofected cells, nearly 70% fire action potentials, and 20% attempt to generate spontaneous action potential; while in pCAGG nucleofected cells, 46% fire action potential, and 46% are not generating spontaneous action potential, and hence are not being active neurons **(Figure 6.23)**.

These data together suggest that the GSX2-induced expression by IWR-1 pre-treatment in pCAGG-DLX2/MASH1 nucleofected cells were direct differentiated into functional and mature MSN-like neurons with striatal GABAergic phenotypes expressing the above genes (EBF1, FOXP1, DRD1/2, GAD2, and CALBIN-1). The induction of GSX2 using IWR-1 pre-treatment has effectively differentiated 34D6 nrNPCs into GABAergic neurons.

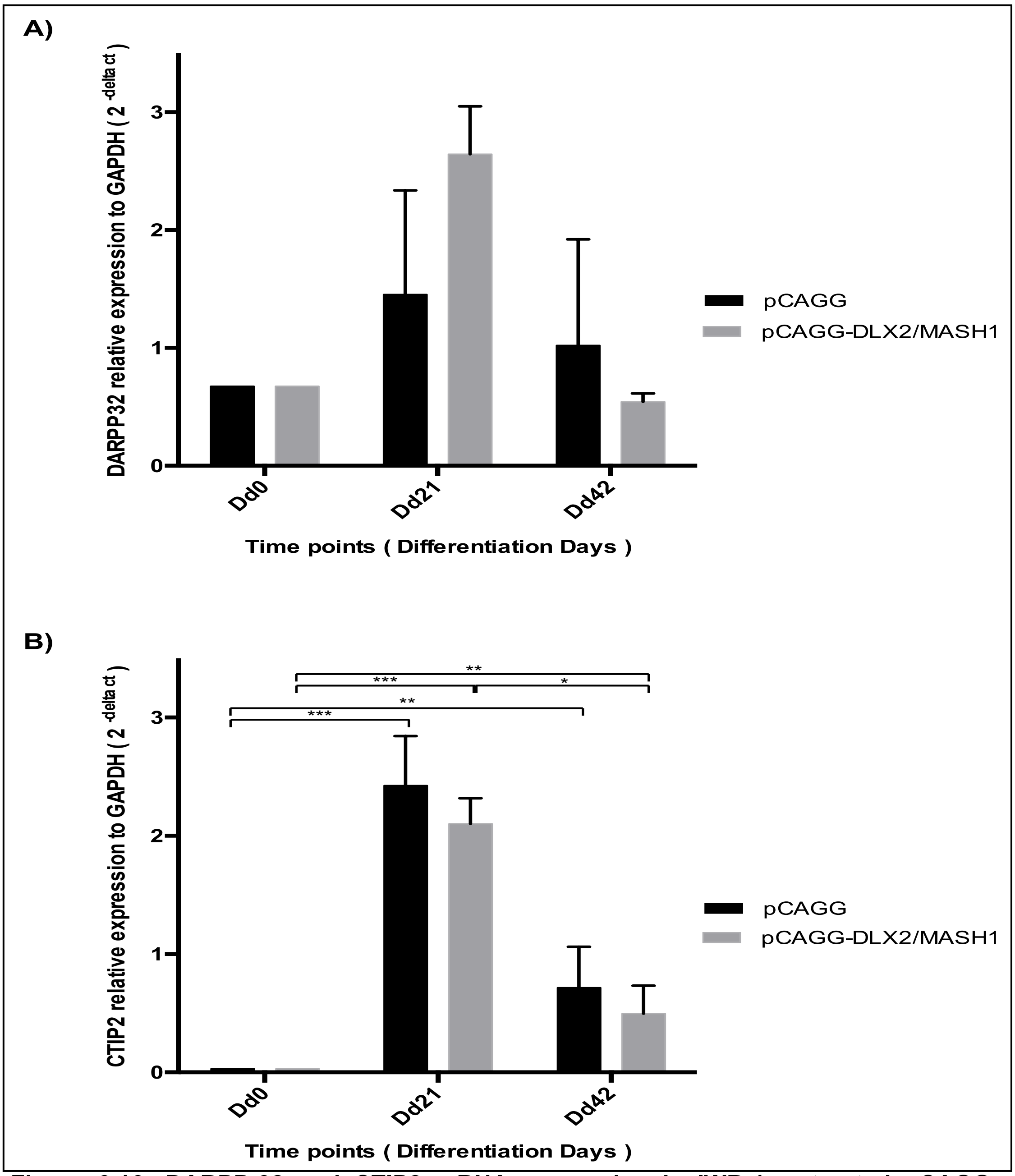


Figure 6.16: DARPP-32 and CTIP2 mRNA expression in IWR-1-pretreated pCAGG-DLX2/MASH1 and pCAGG nucleofected 34D6 nrNPCs at different time points.

DARP32 (A) and CTIP2 (B) expression, relative to GAPDH ($2^{-\text{delta ct}}$), in pCAGG and pCAGG-DLX2/MASH1 nucleofected cells was measured by qRT-PCR. Statistical test was performed using Two-way ANOVA with Bonferroni correction. p-value summary: * = 0.01, ** = 0.001, and *** = 0.0001.

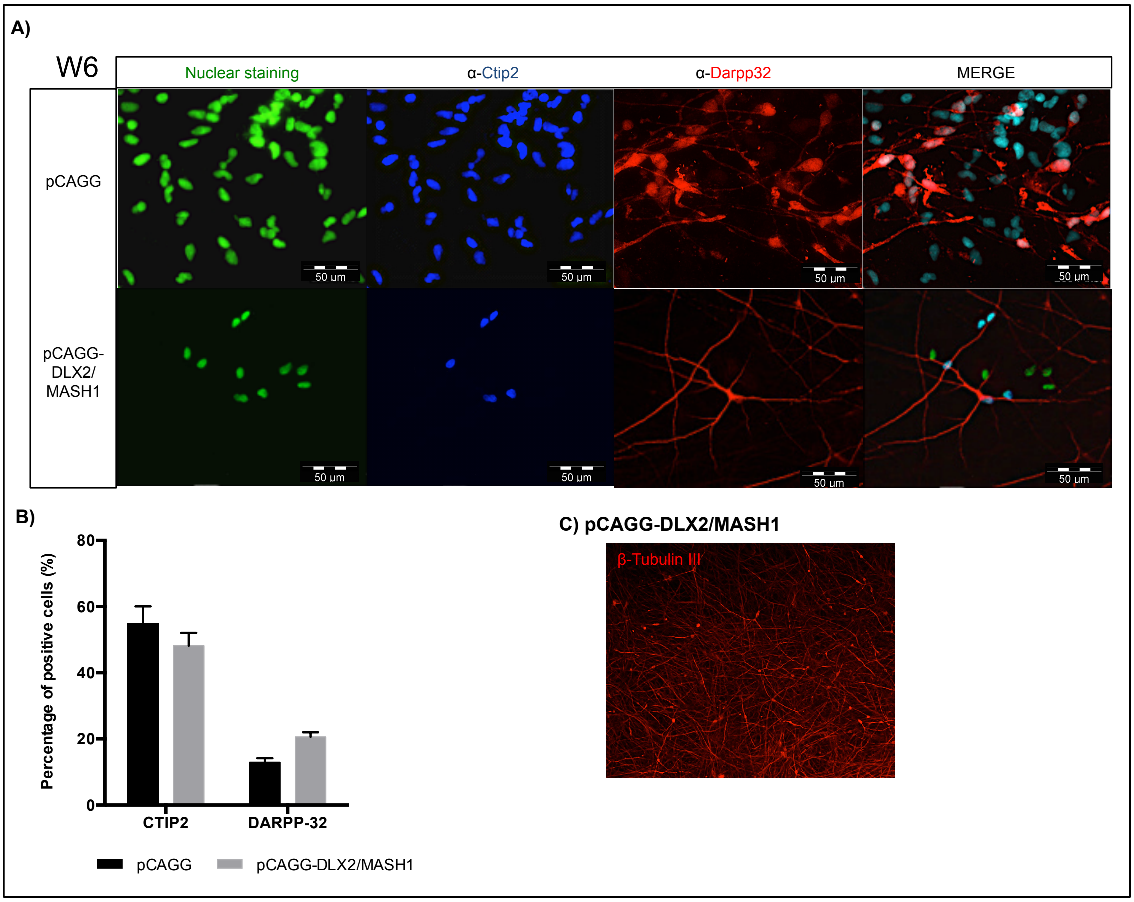


Figure 6.17: Expression of DARPP-32 and CTIP2 at W6 in IWR-1-pretreated pCAGG and pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs.

IWR-1-treated, pCAGG and pCAGG-DLX2/MASH1 nucleofected cells were labeled with antibodies against DARPP-32 (primary dilution; 1:100, Santa Cruz) with an Alexa Fluor® 594 labelled anti-rabbit IgG secondary antibody (red), and CTIP2 (primary dilution; 1:500, Abcam) with an Alexa Fluor® 350 anti-Rat IgG secondary antibody (blue). Nuclear staining was done with SYTOX green fluorescent counterstain (primary dilution; 1:300, Invitrogen. Cat. No. S33025) (A). DARPP-32 and CTIP2 gene expression in pCAGG and pCAGG-DLX2/MASH1 nucleofected cells was measured by cell profiler program (N = 3) (B). pCAGG-DLX2/MASH1 nucleofected cells expressed the neuronal marker β-Tubulin III (C).

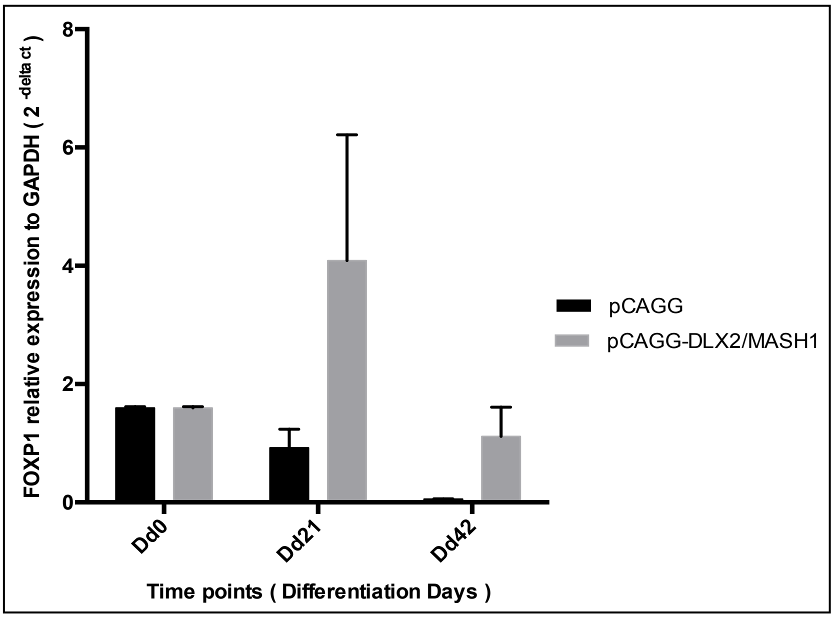


Figure 6.18: FOXP1 mRNA expression in IWR-1-treated pCAGG-DLX2/MASH1 and pCAGG nucleofected 34D6 nrNPCs at different time points.

Expression of FOXP1, relative to GAPDH ($2^{-\text{delta ct}}$), in IWR-1-treated pCAGG and pCAGG-DLX2/MASH1 nucleofected cells was measured by qRT-PCR. The expression of FOXP1 in IWR-1-treated pCAGG-DLX2/MASH1 nucleofected cells was increased from Dd0 to Dd21. Also, at Dd21 and Dd42, the expression of FOXP1 was increased in these cells compared to the control.

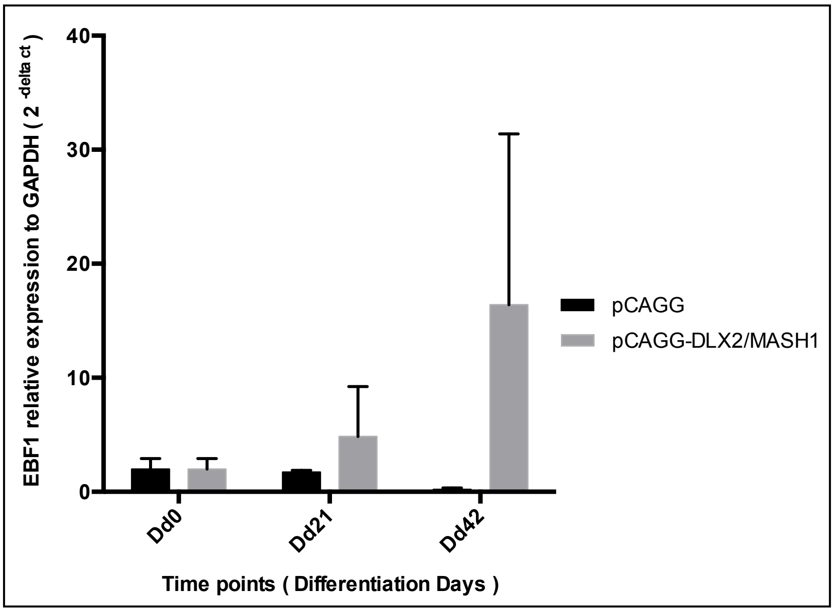


Figure 6.19: EBF1 mRNA expression in IWR-1-treated pCAGG-DLX2/MASH1 and pCAGG nucleofected 34D6 nrNPCs at different time points.

Expression of EBF1, relative to GAPDH ($2^{-\text{delta ct}}$), in IWR-1-treated pCAGG and pCAGG-DLX2/MASH1 nucleofected cells, was measured by qRT-PCR. The expression of EBF1 in IWR-1-treated pCAGG-DLX2/MASH1 nucleofected cells was increased from Dd0 to Dd42. Also, at Dd21 and Dd42, the expression of EBF1 was increased in these cells compared to the control.

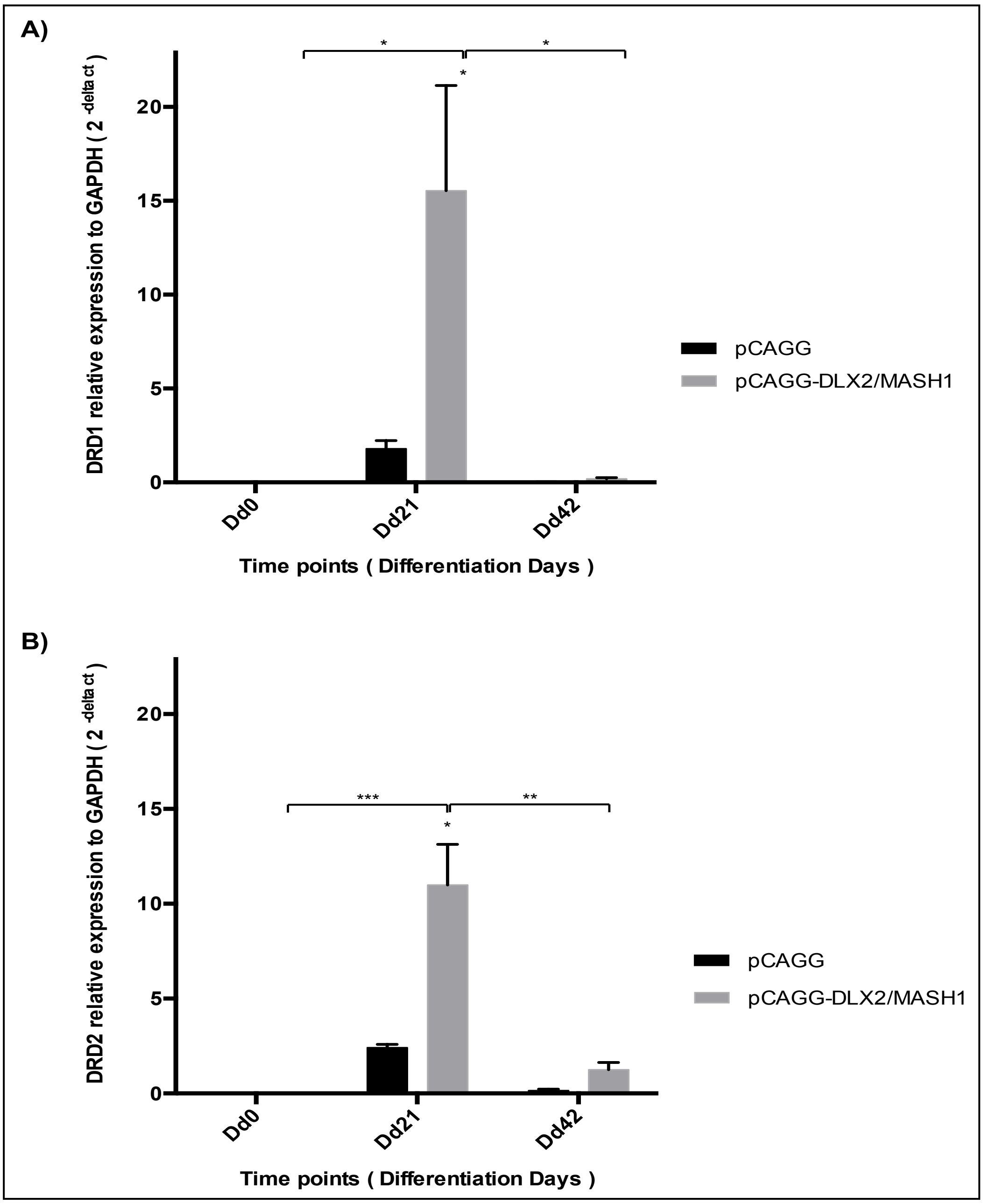


Figure 6.20: DRD1 and DRD2 mRNA expression in IWR-1-treated pCAGG-DLX2/MASH1 and pCAGG nucleofected 34D6 nrNPCs at different time points.

Expression of DRD1 (A) and DRD2 (B) dopamine receptors, relative to GAPDH ($2^{-\text{delta ct}}$), in IWR-1-treated pCAGG and pCAGG-DLX2/MASH1 nucleofected cells, was measured by qRT-PCR. The expression of DRD1 and DRD2 in IWR-1-treated pCAGG-DLX2/MASH1 nucleofected cells was increased significantly from Dd0 to Dd21. Compared to the control, the expression of DRD1 and DRD2 was increased significantly in treated pCAGG-DLX2/MASH1 cells at Dd21. Statistical test was performed using Two-way ANOVA with Bonferroni correction. p-value summary: * = 0.01, ** = 0.001, and *** = 0.0001.

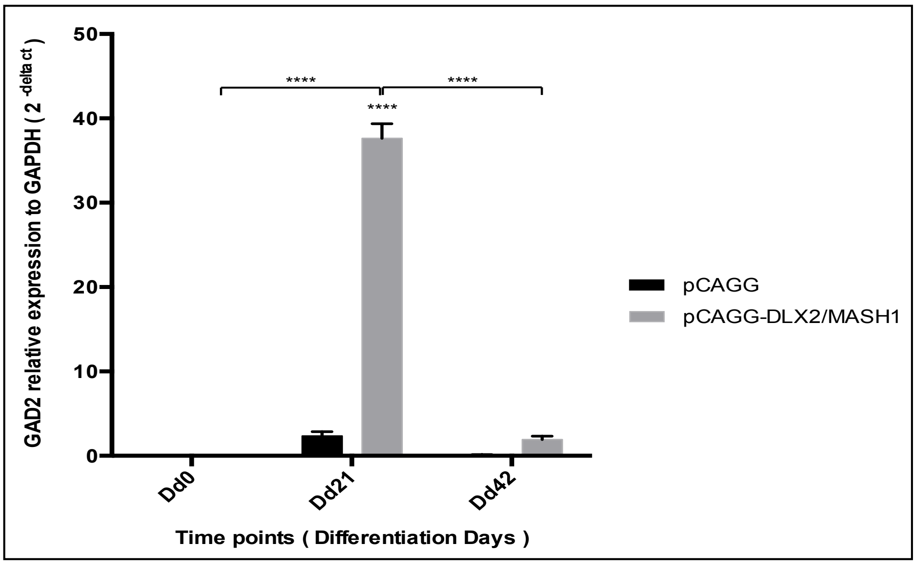


Figure 6.21: GAD2 mRNA expression in IWR-1-treated pCAGG-DLX2/MASH1 and pCAGG nucleofected 34D6 nrNPCs at different time points.

Expression of GAD2, relative to GAPDH ($2^{-\text{delta ct}}$), in IWR-1-treated pCAGG and pCAGG-DLX2/MASH1 nucleofected cells, was measured by qRT-PCR. The expression of GAD2 in IWR-1-treated pCAGG-DLX2/MASH1 nucleofected cells was increased significantly from Dd0 to Dd21. Compared with the control, the expression of GAD2 was increased significantly in IWR-1-treated pCAGG-DLX2/MASH1 cells. Statistical test was performed using Two-way ANOVA with Bonferroni correction. p-value summary: **** < 0.0001.

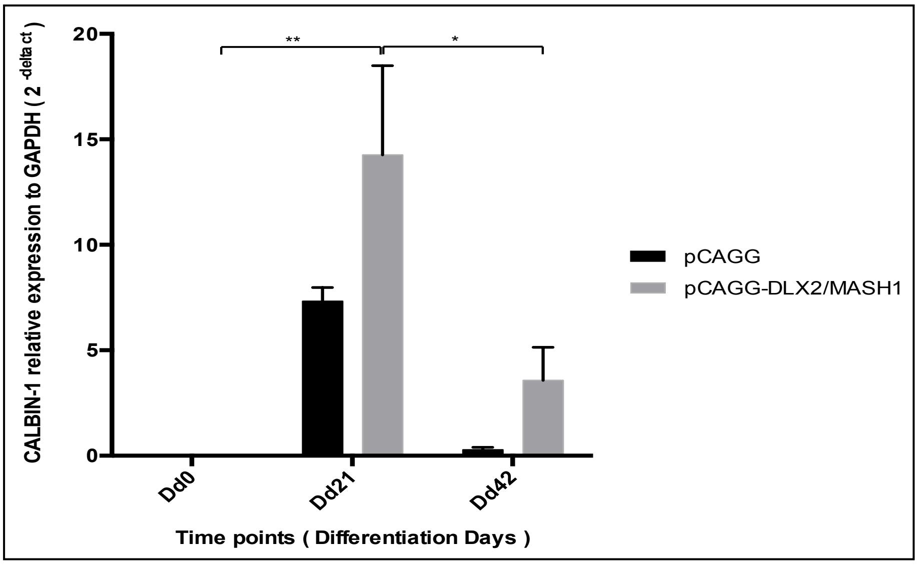


Figure 6.22: CALBIN-1 mRNA expression in IWR-1-treated pCAGG-DLX2/MASH1 and pCAGG nucleofected 34D6 nrNPCs at different time points.

Expression of CALBIN-1, relative to GAPDH ($2^{-\text{delta ct}}$), in IWR-1-treated pCAGG and pCAGG-DLX2/MASH1 nucleofected cells, was measured by qRT-PCR. The expression of CALBIN-1 in IWR-1-treated pCAGG-DLX2/MASH1 nucleofected cells was increased significantly from Dd0 to Dd21. Compared with the control, the expression of CALBIN-1 was increased in IWR-1-treated pCAGG-DLX2/MASH1 cells at Dd21 and Dd42. Statistical test was performed using Two-way ANOVA with Bonferroni correction. p-value summary: * = 0.01, and ** = 0.001.

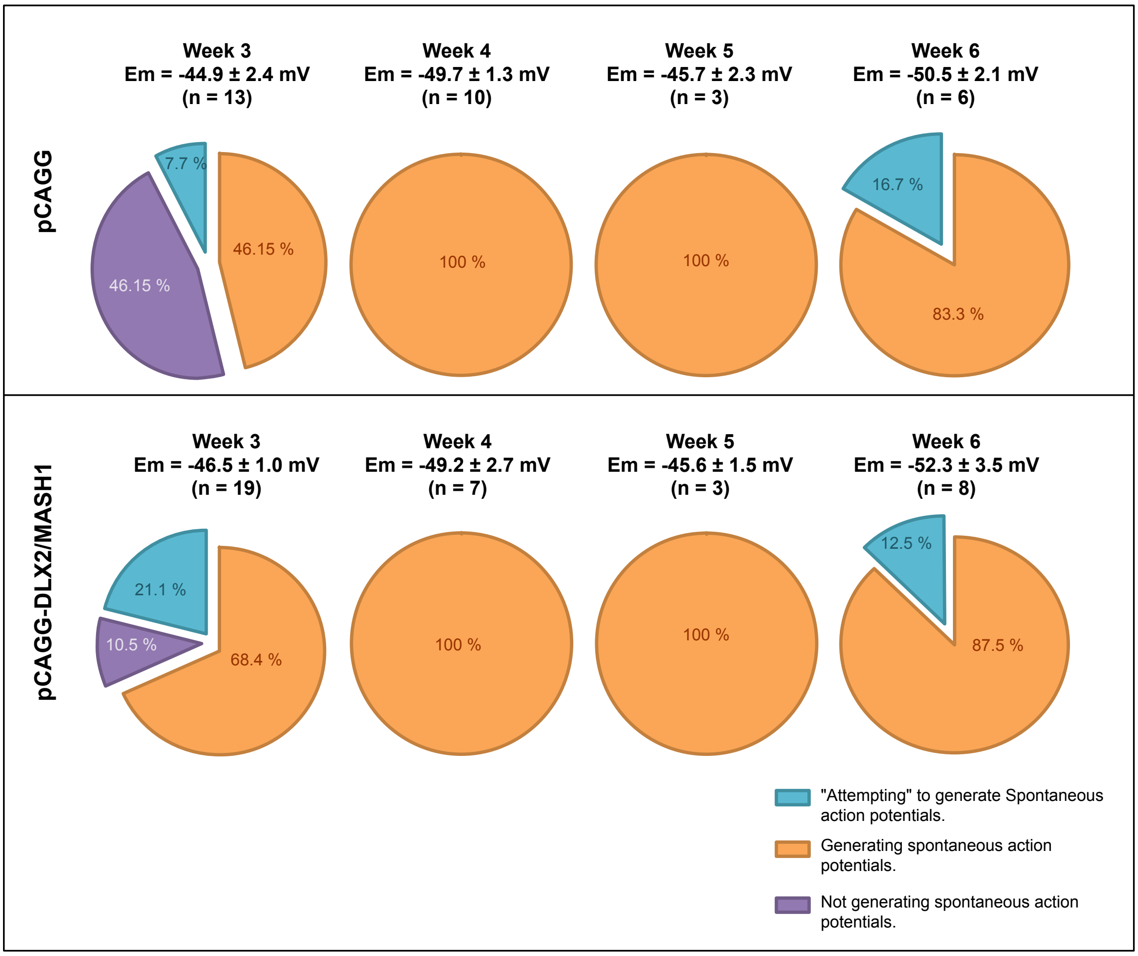


Figure 6.23: Percentage of spontaneously active IWR-1-treated pCAGG-DLX2/MASH1 and pCAGG nucleofected 34D6 nrNPCs after several weeks of differentiation in culture media.

The pie charts above show the percentage of cells generating spontaneous action potentials, attempting to generate spontaneous action potentials or not generating any spontaneous action potentials, in nucleofected IWR-1-treated cells from week 3 to week 6. Overexpression of DLX2 and MASH1 in IWR-1-treated cells resulted in an increase in the proportion of cells generating spontaneous action potentials at week 3 (20%) and at week 6 (4%).

6.5 Discussion

The present study has established a novel and highly effective *in vitro* protocol to generate a high yield of functional GABAergic MSN-like cells via forced co-expression of DLX2 and MASH1 in 34D6 nrNPCs. These GABAergic neurons were FOXP1^{+ve} / EBF1^{+ve} / GAD2^{+ve} / CALBIN-1^{+ve} / DRD1^{+ve} / DRD2^{+ve} / DARPP-32^{+ve} and CTIP2^{+ve}. We have demonstrated that the remaining combinations of TFs (DLX2, MASH1, DLX2/GSX2, and DLX2/MASH1/GSX2) in H9 nrNPCs led to differentiation into DARPP-32^{+ve}/CTIP^{+ve} MSNs, but not functionally active. In addition, pre-treatment of 34D6 nrNPCs with IWR-1 prior to nucleofection with pCAGG-DLX2/MASH1 also caused differentiation into mature and functional GABAergic MSN-like cells.

The data presented here shows that all of the differently nucleofected cells were undergoing neurogenesis, as evidenced by expression of neuronal biomarker, β-Tubulin III^{+ve} at W3. These β-Tubulin III^{+ve} neurons gained the striatal MSN phenotype by expressing DARPP-32^{+ve} and CTIP2^{+ve} at W3. The results are consistent with previous studies that employed the widely used biomarkers β-Tubulin III, Darpp-32, Ctip2 and Darpp-32/Ctip2 co-expression to detect striatal MSNs (Ouimet and Greengard 1990; Ouimet *et al*. 1984; Arlotta *et al*. 2008; Carri *et al.* 2013; Ding *et al.* 2014) The data also clearly shows that single MASH1, DLX2 and DLX2/MASH1 nucleofection in H9 nrNPCs are associated with the highest increase in the DARPP-32 and CTIP2 biomarkers at W6. Interestingly, of all the nucleofected constructs, only pCAGG-DLX2/MASH1 showed action potential firing starting at W3 in both, H9 and 34D6 nrNPCs. In addition, DLX2/MASH1 co-expression in 34D6 nrNPCs was associated with a high yield of pure MSN-like cells and approximately 85% of these cells were CTIP2^{+ve} and DARPP-32^{+ve}. These findings suggest that the TF combination of DLX2/MASH1 is the most efficient for inducing the expression of neuron-related biomarkers and subsequently generating a higher yield of functional MSN-like cells.

At the stage of neurogenesis, increased expression of CTIP2 and DARPP-32 in pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs has a functional impact in MSNs. Previous studies have shown increased expression of Ctip2 during embryogenesis at E12.5 through to adulthood, and MSN neurogenesis occurs at E13.5, revealing importance of Ctip2 in striatal development (Arlotta *et al.* 2008). In addition, expression of Ctip2 increases during MSN migration in the mantle zone (MZ) and in post-mitotic neurons (Arlotta *et al.* 2008). Ctip2 is exclusively expressed in the striatum and plays a fundamental role in MSN differentiation and striatal architecture (Arlotta *et al.* 2008). DARPP-32 acts as an integrator at dopaminoceptive neurons and is a regulator of neuronal excitability, electrophysiology, and transcriptional as well as behavioral responses to physiological plus pharmacological stimuli, as evidenced by profound functional deficits in mice lacking the Darpp-32 gene (Fienberg *et al.* 1998; Bibb *et al.* 1999; Svenningsson *et al.* 2004). Many studies have also shown that Darpp-32 acts via phosphorylation and dephosphorylation of its various threonine and serine locations, which in turn regulate phosphatase and kinase function (Nishi *et al.* 1997; Hemmings *et al.* 1984; Bibb *et al.* 1999; Svenningsson *et al.* 2004; Fernandez *et al.* 2006). For example, phosphorylation of threonine 34 (thr^{34}) on Darpp-32 by cAMP-dependent protein kinase (PKA), upon activation of DRD1, results in the conversion into potent protein phosphatase-1 inhibitor (PP1) (Hemmings *et al.* 1984; Nishi et al. *1997*). Subsequently, PP1 inhibitors dephosphorylates many downstream physiological effectors, including NMDA glutamate receptors, voltage-gated ion channels, kinases and transcription factors (Greengard *et al.* 1999; Bibb *et al.* 1999; Svenningsson *et al.* 2004). However, phosphorylation of thr^{75} by cell division protein kinase 5 (CDK5) modulates PP1's inhibitory effect through inhibition of PKA (Bibb *et al.* 1999; Nishi *et al.* 2000). These functions may have aided in both, the characterisation of GABAergic MSN function and the successful differentiation in the current *in vitro* models.

Following successful generation of functional neurons, the striatal MSN phenotype was assessed by examining the FOXP1, EBF1 and DRD1/2 biomarkers.

Many studies have demonstrated the expression of these biomarkers in striatal MSNs in the supraventricular zone (SVZ) and the MZ (Garel *et al.* 1999; Yun *et al.* 2002; Garcia-Dominguez *et al.* 2003; Tamura *et al*. 2004; Long *et al.* 2009b). Hence, they were used to detect striatal MSNs in many mutant mouse models when studying the role of TFs in neurogenesis. FOXP1 has a key role in up-regulating glutamate and GABA receptor signaling genes, which are required in order for striatal MSNs to receive glutamatergic input from the cortex (Jones *et al.* 1977; Royce 1982; Ferino *et al.* 1987; Wilson 1987; Tang *et al.* 2012). In addition, FOXP1 is highly expressed in striatal projection neuron development (Tamura *et al.* 2004; Martín-Ibáñez *et al.* 2012). Similarly, EBF1 has been found to play a role in striatal projection neuron development and differentiation (Garel *et al.* 1999; Tamura *et al.* 2004; Martín-Ibáñez *et al*. 2012; Lobo *et al*. 2006). Consistent with the previous findings, this study has showed significantly increased expression of FOXP1 and EBF1 at W3, confirming a striatal MSN phenotype in neurons derived from pCAGG-DLX2/MASH1 nucleofected nrNPCs. The expression of these biomarkers also suggests the important role in the development of the striatum and MSNs. Therefore, both FOXP1 and EBF1 are principal biomarkers for striatal projection neurons.

The importance of the striatum in controlling motor and cognitive functions has been documented extensively (Albin *et al.* 1989; Moyer *et al.* 2007). This role is accomplished, either direct or indirect pathways through dopaminergic action via Drd1/2. Drd1 provides direct neuronal projections to the substantia nigra (SNr) and entopeduncular nucleus. Conversely, Drd2 provides indirect projections to the external segment of the globus pallidus (GPe) and subthalamic nucleus (STN) (Albin *et al.* 1989; Moyer *et al.* 2007). The counterbalance between Drd1 and Drd2 modulates the responsiveness of the direct and indirect pathways to cortical signals that is the key model of basal ganglia (Albin *et al.* 1989). In addition, several lines of evidence have demonstrated that Drd1 and Drd2 receptors have opposing effects on cAMP and PKA activity in neostriatal neurons, thus regulating dendritic excitability and glutamatergic signaling in MSNs (Nishi *et al*. 1997). Activation of Drd1 receptors increases cAMP and PKA phosphorylation and, subsequently, increases Darpp-32 phosphorylation of thr^{34}. Drd2 receptor activation, on the other hand, decreases cAMP and PKA activity and dephosphorylates Darpp-32 at thr^{34} (Stoof and Kebabian 1981). A number of studies have demonstrated increased expression of both Drd1 and Drd2 in the striatum, with Drd2 expression being higher than Drd1 in mice and rat (Schambra *et al.* 1994; Jung and Bennett 1996; Sillivan and Konradi 2011). This pattern is observed in the striatum during neurogenesis, in embryogenesis through adulthood (Jung and Bennett 1996; Araki *et al.* 2007; Sillivan and Konradi 2011). However, the predominance of Drd2 over Drd1 is reversed in the cortex (Araki *et al.* 2007; Sillivan and Konradi 2011). Many studies have shown the presence of two types of MSNs that are segregated based on location: Drd1-dominated (striatonigral) and Drd2-dominated (striatopallidal) (Gerfen *et al.* 1990). In line with these studies, current study demonstrates increased expression of both the DRD1 and DRD2 genes in pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs compared to control group. Interestingly, the results also reveal greater up-regulation of DRD2 over DRD1. Therefore, these findings confirm the generation of striatal MSNs from iPSCs in our *in vitro* model.

Moreover, the data demonstrate that functional MSNs derived from pCAGG-DLX2/MASH1 nucleofected 34D6 nrNPCs gain the GABAergic phenotype, as seen through increased CALBIN-1 and GAD2 expression at W3. A growing number of studies are utilizing these two biomarkers to detect GABAergic MSNs (Kiyama *et al.* 1990; Gerfen 1992; Pickel and Heras 1996; Pinal and Tobin 1998; Pan 2012; Lin *et al.* 2015). In mice studies, Calbin-1 is known to be localised in GABA-dominant regions of the brain (e.g. dorsal striatum and caudate-putamen nuclei) (Kiyama *et al.* 1990; Gerfen 1992; Pickel and Heras 1996; Lin *et al.* 2015). The Gad2 gene encodes the glutamic acid decarboxylase enzyme, which decarboxylates glutamate into GABA (Pinal and Tobin 1998; Pan 2012). The distinct role of Gad2 is supported by its restricted expression in nerve terminals and synapses, where GABA acts as a neurotransmitter (Pinal and Tobin 1998).

Many mutant mice models have been established to understand the role of different TFs in neurogenesis in the ventral telencephalon (Casarosa *et al.* 1999; Stühmer *et al*. 2002b; Yun *et al*. 2002; Cobos *et al*. 2007; Poitras *et al*. 2007; Long *et al*. 2009b; Castro *et al*. 2011). Double and triple TF mutant mice, such as Mash1$^{-/-}$;Dlx1/2$^{-/-}$ and Gsx2$^{-/-}$;Mash1$^{-/-}$ have been associated with severe defects in striatal development, unlike single TF mutant mice models such as Mash1$^{-/-}$ and Gsx2$^{-/-}$. The exception was the single Dlx2 mutant mouse model, where severe defects were observed in the dLGE, with the vLGE and septum preserved. These findings shed light on the importance of TFs interaction in regulating and driving the consecutive processes of neurogenesis. Therefore, the use of a combination of TFs to generate GABAergic MSN-like cells in this work was the fundamental objective. In fact, the results of the current study clearly demonstrated the successful use of TFs combination (i.e. DLX/MASH1) to generate functional GABAergic MSN-like cells *in vitro*.

The expression of endogenous DLX2 and MASH1 in *in vitro* has initiated proliferation and differentiation of stem cells into functional GABAergic MSNs via induction of target and effector genes. Many studies have shown the significance of both DLX2 and MASH1 and their target genes in neuronal differentiation and development. DLX2 plays an essential role in promoting the differentiation of striatal projection neurons in the BG (Long *et al*. 2009b; Lobo *et al*. 2006). Its importance was revealed through the Dlx1/2$^{-/-}$ mutant mouse model, where the expression of striatal differentiation markers such as Drd1/2 and Gad1/2 was reduced in the striatum (Yun *et al*. 2002; Long *et al.* 2009b). These striatal differentiation genes were linked to the dLGE; hence it was suggested that the function of Dlx1/2 was critical for the dLGE regions (olfactory bulb and striatum interneurons) (Long *et al*. 2009b). Dlx2 partially supports the GABAergic phenotype and differentiation phases during the late stage of neurogenesis via repression of Mash1 and Notch signaling (Casarosa *et al.* 1999; Yun *et al*. 2002; Poitras *et al*. 2007; Long *et al*. 2009b). In addition, Dlx2 triggers GABAergic neuron differentiation through up-regulation of Gad1/2 expression and vGat (Long *et al*. 2009b; Stühmer *et al*. 2002b). Dlx2 also plays a role in neurite maturation by repressing the p21-activated serine/threonine kinase, PAK3, and migration of GABAergic neurons to the neocortex by promoting Arx expression (Colasante *et al*. 2008; Cobos *et al*. 2007). Impaired neuronal migration was evident in Arx mutant mice and Dlx1/2 double homozygous mutants (Long *et al.* 2009b; Colombo *et al*. 2007; Colasante *et al*. 2008). Furthermore, Dlx2 has a role in specification of progenitors in the SVZ of LGE by repressing some TFs, such as MGE TFs (Gsx1, Gbx1/2), a diencephalon TFs (Otp) and ventral cortical TFs (such as Ebf3) (Long *et al*. 2009b).These knowledge have indicated that Dlx2 was required at later steps in the development of LGE to regulate differentiation, migration and maturation of LGE (Yun *et al*. 2002).

Mash1 is believed to exhibit a parallel and overlapping role with Dlx2. In other words, Mash1 plays several roles at both the early and late stages of neurogenesis including (i) promoting the expression of neural markers such as Map2 and Sox1, which are not expressed by TF Dlx2 (Yun *et al*. 2002; Long *et al*. 2009b; Cobos *et al*. 2007), (ii) repressing differentiation of adjacent progenitors through upregulation of Notch signaling, (iii) cell fate specification and (vi) cell proliferation (Castro *et al*. 2011; Yun *et al*. 2002). Overexpression of Mash1 was observed in the Dlx1/2$^{-/-}$ double homozygous mutant, which maintained some characteristics of striatal differentiation (Long *et al*. 2009b). Moreover, in Dlx1$^{-/-}$;Dlx2$^{-/-}$;Mash1$^{-/-}$ triple mutant, it was demonstrated that the majority of LGE differentiation relied on their combined function (Long *et al*. 2009b; Yun *et al*. 2002; Poitras *et al*. 2007). Consequently, it was suggested that both Dlx1/2 and Mash1 play a parallel role in regulating LGE differentiation and specification (Long *et al*. 2009b). However, it was found that the development of dLGE was more reliant on Dlx1 and Dlx2 than Mash1, while the opposite was true for the development of vLGE and septum (Long *et al*. 2009b). This variation is due to their different expression levels in the dLGE and vLGE. For example, it was found that Dlx2 and Mash1 were mostly co-expressed in the VZ and SVZ progenitors of dLGE. However, Dlx2 was expressed to a lesser extent in the vLGE than the dLGE region (Yun *et al*. 2002). Hence, the function of Dlx2 was more critical in the dLGE rather than the vLGE, while the function of Mash1 was more critical in the vLGE and septum (Long *et al*. 2009b). Therefore, the interaction between Dlx2 and Mash1 was identified to play a major role in the transcriptional hierarchies regulating LGE and GABAergic neuron differentiation and specification (Long *et al*. 2009b). This is further supported by current work showing that the combination of DLX2 and MASH1 was sufficient to direct the differentiation of iPSCs into functionally mature MSNs, unlike the other TFs constructs.

In the past, while some progress had been made in disease modeling, the pathogenesis of neurodegenerative diseases was not yet fully understood, and hence treatments for such diseases were not developed. The recent establishment of iPSCs, by Takahashi and Yamanaka *et al*., in 2006, has opened new avenues for scientists to generate and develop more sophisticated cell models for investigating and developing treatments for diseases such as, Huntington's, Parkinson's and Alzheimer's diseases. A recent study succeeded in grafting hPSCs-derived striatal precursors into the striatum of quinolinic acid (QA)-lesioned rats after they were treated with SHH/DKK1 (Carri *et al*. 2013). *In vivo*, these precursor cells differentiated further into DARPP-32^{+ve} MSNs (Carri *et al*. 2013). Subsequently, motor neuron deficit symptoms in these rats were improved (Carri *et al*. 2013). This study highlights the possibility that hPSC-derived nrNPCs could, in the future, be used in regenerative medicine to cure neurological diseases. However, this study has demonstrated that ectopic expression of striatum-specific TFs such as DLX2 and MASH1, rather than morphogens (e.g. SHH), could efficiently drive the differentiation of iPSCs into a high yield of functionally mature MSN-like cells. These iPSC-derived MSNs establish a foundation for the future differentiation of HD-specific patient iPSCs into mature and functional neurons, for use in HD disease modeling and in regenerative medicine or gene therapy.

Chapter 7: General discussion.

To date, HD pathophysiology remains poorly understood. The use of human induced pluripotent stem cells (h-iPSCs) for disease modeling, affords an opportunity to understand disease pathophysiology. HD-specific cell-based models, derived by the differentiation of HD-related cell types such as MSN from HD-specific patient iPSCs (HD-iPSCs), offer an opportunity to gain a deeper understanding of HD pathophysiology. The main aim of this study was to develop an efficient and reproducible protocol for direct programming of MSN differentiation from h-iPSCs. This is currently an important approach for disease modeling, drug screening and gene therapy. However, the best protocols available for direct programming of MSN differentiation from hPSCs have been unsatisfactory. Nevertheless, some innovative work has been achieved recently by Carri *et al*. (2013), Nicoleau *et al*. (2013) and Arber *et al*. (2014) in the generation of MSN from hPSCs using DKK1 & SHH (Carri *et al*. 2013; Nicoleau *et al*. 2013) or ACTIVIN A (Arber *et al*. 2014). Their approach was to use developmental signals to control and stimulate transcriptional networks that regulate sequential neuron progenitor fate.

This thesis is the first to develop an alternative protocol, and it has been shown that this protocol can successfully, efficiently and reproducibly generate functional MSN by the ectopic expression of key fate defining TFs that play a role in subpallium and MSNs specification and differentiation.

7.1 The three TFs DLX2, MASH1 and GSX2 were chosen for ectopic expression in hPSCs to direct differentiation into MSN.

The TFs DLX2, MASH1 and GSX2 have been shown to be expressed in the ventral telencephalon and to have a role in the development of ventral telencephalon and striatum (Chapter 1) (Porteus *et al*. 1994; Horton *et al.* 1999; Anderson 1997b; Casarosa *et al*. 1999; Panganiban and Rubenstein 2002; Yun *et al*. 2002; Cobos *et al.* 2007; Long *et al.* 2009b; Wang *et al*. 2013). Further, microarray analysis, detailed in

Chapter 2, has shown DLX2, MASH1 and GSX2 to be involved in subpallium development. Thus, microarray analysis is a valuable tool for validation and elucidation of the transcriptional network. Ventral telencephalon deficiency could be induced by knockout of these TFs (Anderson *et al.* 1997a; Toresson *et al.* 2000; Long *et al*. 2009b; Wang *et al.* 2009) and, therefore, combinations of the TFs DLX2, MASH1 and GSX2 were chosen to be cloned and sub-cloned into the expression vector pCAGG.

MASH1, DLX2 and GSX2 are similarly expressed in both mouse and human ventral telencephalon domains (Fode *et al.* 2000; Carri *et al.* 2013; Pauly *et al*. 2013). However, loss-of-function studies have demonstrated that each of DLX2, MASH1 and GSX2 has different functions in striatal neuron development, and they differ in their response to several signals (Anderson *et al.* 1997a; Anderson 1997b; Casarosa *et al*. 1999; Toresson *et al.* 2000; Corbin *et al.* 2000; Stühmer *et al.* 2002a; Yun *et al*. 2002; Yun *et al*. 2003; Woltjen *et al.* 2009; Wang *et al.* 2013). Consequently, this raised the question: which gene targets of these TFs promote MSN differentiation? To address this question, different constructs of these TFs in the pCAGG vector (pCAGG-DLX2, pCAGG-MASH1, pCAGG-DLX2/MASH1, pCAGG-DLX2/GSX2 and pCAGG-DLX2/MASH1/GSX2) were successfully produced and expressed in hPSCs.

The successful and efficient cloning of expression vectors with the appropriate expression of each of the TFs and the self-cleavage of 2A peptides were achieved. Nucleofection of the expression constructs (pCAGG-DLX2, pCAGG-DLX2/MASH1, pCAGG-DLX2/GSX2, pCAGG-DLX2/GSX2/MASH1, pCAGG-MASH1 and pCAGG) into HEK293 have resulted in 80-90% GFP expression, and the 36, 35 and 34 kDa proteins (DLX2, MASH1 and GSX2 respectively) have been detected. These data confirm co-expression of TFs together from one vector (pCAGG polycistronic vector), essential for the efficient production of MSN. Thus, this vector can achieve efficient expression of multiple TFs and can be used for gene therapy. A similar outcome was reported by Szymczak *et al*. (2004), showing that a polycistronic vector with 2A peptides produced an efficient translation of multiple CD3 genes that stimulated T-cell differentiation in CD3 knockout mice (Szymczak *et al*. 2004).

GSX2 is enriched in the LGE (Carri *et al*. 2013) and is required for the up-regulation of ventral telencephalon genes that are essential for LGE specification and differentiation, such as MASH1 and DLX2 (Corbin *et al.* 2000; Wang and Steinbeisser 2009; Wang *et al*. 2013). In Gsx2 knockout studies in mouse, it was shown that the expression of Mash1 and Dlx2 were reduced compared to the wild type, and the striatal neuronal phenotype that originates from LGE was lost while the cortical interneurons were unaffected (Toresson *et al.* 2000; Corbin *et al.* 2000; Yun *et al.* 2003; Wang *et al*. 2013). It seems that Gsx2 is an upstream gene for Dlx2 and Mash1; hence Gsx2 was chosen as a TF to direct differentiation *in vitro* in this study.

DLX2 is located in the two domains of ventral telencephalon: the LGE and to some extent in MGE (Pauly *et al*. 2013). Fate-mapping studies showed that DLX2 was expressed in cells co-expressing GABA (Stühmer *et al*. 2002a). Dlx2 contributes profoundly in different stages of neurogenesis: proliferation and differentiation (Lobo *et al*. 2006; Long, Swan, *et al*. 2009). Dlx2 is required to inhibit Mash1 and Notch signaling to repress proliferation and initiate differentiation (Casarosa *et al.* 1999; Yun *et al.* 2002; Poitras *et al*. 2007; Long *et al*. 2009b). In addition, Dlx2 up-regulates Arx and suppresses Pax3 to induce migration of striatal neurons (Cobos *et al*. 2007; Colasante *et al*. 2008). These are supported further by $Dlx2^{-/-}$ mice, where both cortical interneurons and striatal matrix GABAergic neurons were decreased (Anderson 1997a; Anderson *et al*. 1997b). Another TF of importance in ventral telencephalon is MASH1. Its expressing cells are also located in the LGE and MGE (Fode *et al*. 2000). In the studies of mice lacking Mash1, there was a reduction in the expression of cortical interneurons. There was also a reduction in the expression of Foxp1, Darpp-32, and Calbindin, which are expressed in striatal neuronal progenitors, MSN and striatal matrix neurons respectively (Casarosa *et al*. 1999; Wang *et al*. 2009). The loss-of-function model provides evidence of role of Mash1 in neurogenesis. A growing number of studies have revealed that Mash1 exhibits its function through Notch-dependent (in LGE) and independent (in MGE and septum) mechanisms (Horton *et al*. 1999; Casarosa *et al*. 1999; Yun *et al*. 2002).

Together, the demonstrated roles of GSX2, DLX2 and MASH1 encourage use of these key TFs to direct stem cell programming in vitro. In fact, triple and double mutant mice model: MASH1;DLX1/2 and GSX2;MASH1, respectively, have further drawn the line of importance of these key TFs in striatal development (Wang and Steinbeisser 2009; Long *et al.* 2009). Therefore, these Key TFs were selected to conduct the aims and objectives of this thesis.

7.2 Ectopic expression of different combinations of MASH1, DLX2 and GSX2 in hPSCs induced direct programming of sequential LGE fate specification and eventual differentiation into mature MSNs.

Ectopic expression of different combinations of DLX2, MASH1 and GSX2 in 34D6 nrNPCs has been shown to induce direct LGE fate specification, the striatum primordium (Chapter 5). All combinations of TF nucleofected into 34D6 nrNPCs, with the exception of individual MASH1 or DLX2 constructs, reduced expression of PAX6 from ND3 to ND42. In addition, different combinations of TF reduced the expression of MGE NKX2.1 from ND3 in all nucleofected 34D6 cells. In this study, the reduction of PAX6 and NKX2.1 excludes formation of MGE and dorsal neuronal phenotype, which is mediated by different combinations of DLX2, MASH1 and GSX2. The endogenous expression of these TFs were efficient to drive hPSC-derived nrNPCs into LGE neuronal phenotype. Determination of LGE phenotype mediated by DLX2, MASH1 and GSX2 has been explored in many studies. It has been reported that when the subpallium markers, such as Dlx2, Mash1 and Gsx2, were mutated within telencephalon, the pallium markers, such as Pax6, were increased and expanded into

LGE (Casarosa *et al*. 1999; Wilson and Rubenstein 2000; Yun *et al*. 2001; Yun *et al*. 2003). In the Gsx2$^{-/-}$ study, a reduction in the striatal projection neurons, originating from LGE, and misspecification of ventral and dorsal LGE was observed (Yun *et al*. 2001; Yun *et al*. 2003). In the Mash1$^{-/-}$ the differentiation of the early stages of LGE was obstructed (Casarosa *et al*. 1999). In the Dlx2$^{-/-}$, the differentiation of subpallium was deficient (Wang *et al*. 2013), and the number of differentiations of late born LGE neurons was reduced (Anderson *et al*. 1997; Marin *et al*. 2000; Cobos *et al*. 2007). Also, in the mice lacking Dlx2, Nkx2.1 accumulated in the mutant MGE and expanded dorsally into the LGE region and PSB (Marin *et al*. 2000). These data, taken together with the findings of this study and the literature, confirm the importance of DLX2 and MASH1 co-expression or/and GSX2 in the development of LGE, and provide evidences of the function of these genes in the development of LGE within telencephalon.

Based on the positive results reported in Chapter 5, further work was carried out to determine if the 34D6-derived LGE fate specific cells could differentiate further into mature MSN-like neurons. It was demonstrated that 34D6 nrNPCs expressing different combinations of TFs expressed the MSN marker DARPP-32 as well as the striatal MSN marker CTIP2. However, whole cell-patch clamp analysis showed that only the pCAGG-DLX2/MASH1 nucleofected cells were functional and mature MSNs as evident by electrophysiology. These outcomes indicate that ectopic expression of DLX2, MASH1 and GSX2 in 34D6 nrNPCs induced direct programming of differentiation into MSNs. However, in iPS34D6 cells, ectopic expression of DLX2 and MASH1 only was sufficient to direct differentiation to functional MSNs.

DARPP-32 is the most commonly used marker for detection of terminally differentiated striatal GABAergic MSNs (Aubry *et al*. 2008). There are two destinations for striatal projection neurons, which are related to the embryonic stages. The early-born neurons are destined to the patch compartment of striatum, which forms 15% of striatum, while the mid-late neurons migrate to the matrix compartment of striatum, which forms 85% of striatum (Kooy and Fishell 1987). Consequently, the MSNs locate in the two striatum compartments and are combinations of early, mid and late born neurons. DARPP-32 is only expressed in the late born neurons. In this study a large percentage of pCAGG-DLX2/MASH1 nucleofected 34D6nrNPCs expressed DARPP-32, and hence, it indicates the differentiation of late-born MSNs. In fact, DLX2 and MASH1 ectopic expression in 34D6-nrNPCs resulted in an increase of DARPP-32 from around 12% to 85% and CTIP2 from around 69% to 85%. In concordance, the association of DARP-32 expression and late-born neuron generation mediated by DLX2 and MASH1 was illustrated in many mice mutant models. In the mice lacking DLX2, the late-born striatal neurons accumulated in LGE, and the migration from LGE to the striatal MZ was obstructed and DARPP-32 was not expressed (Anderson *et al*. 1997). Moreover, in mice lacking MASH1, the expression of DARPP-32 was reduced (Wang *et al*. 2009). Also, it was reported that ectopic co-expression of Dlx2 and Mash1 differentiates a higher percentage of cortical astroglia cells into GABAergic neurons than sole ectopic expression of Dlx2 (Berninger, Costa, *et al*. 2007; Heinrich *et al*. 2010). These findings indicate that DLX2 and MASH1 provide a greater regulator of genes related to late-born neuron production in the development of striatal MSNs.

This project provides the first example of the use of ectopic delivery or expression of TFs (such as DLX2 and MASH1) in h-iPSCs to direct the reprogramming of terminal differentiation of functional MSNs expressing a variety of TFs markers for striatal GABAergic MSNs. Overexpression of MASH1 and DLX2 in 34D6 nrNPCs, generated GFP^{+ve} / DARPP-32^{+ve} / CTIP^{+ve} / FOXP1^{+ve} / EBF1^{+ve} / DRD1&2^{+ve} / GAD2^{+ve} / CALBIN-1^{+ve} MSNs. The direct differentiated MSNs, in this study, show dopaminergic markers (DRD1 and DRD2), GABAergic markers (GAD2 and CALBIN-1), LGE determinants (FOXP1, EBF1 and CTIP2) and terminal differentiation of striatal MSN (DARPP-32 and CTIP2) similar to those seen in developing human subpallium MSNs (Carri *et al*. 2013). In addition, co-expression of CTIP2 with the MSN marker DARPP-

32 was observed both in this study and *in vivo* human studies (Carri *et al*. 2013). In the caudate nucleus, CTIP2 was co-expressed with the majority of MSN (95%) and was not co-expressed with other cell types (Arlotta *et al*. 2008; Carri *et al*. 2013). Co-expression of DARPP-32 with CTIP2 in cells expressing both of MASH1 and DLX2 suggests that ectopic co-expression of these two genes is able to drive neurons towards a striatal MSN type. Moreover, it was found that FOXP1 and FOXP2 were expressed in LGE of both an 11 week old human fetus (Carri *et al*. 2013) and 50-54 days post fertilization and that FOXP1 and FOXP2 identified striatal progenitors and differentiated MSN (Pauly *et al*. 2013). FOXP2, EBF1 and CTIP2 were reported to be expressed in the LGE SVZ and MZ but not in the MGE (Long *et al*. 2009a) while reduced expression of Calbindin was seen in mice lacking MASH1 (Wang *et al*. 2009), and the expression of GAD67 and DRD2 were also reduced (Casarosa *et al*. 1999). These outcomes support the hypothesis underpinning this thesis and demonstrate the ability of DLX2 and MASH1 to promote the conversion of h-iPSCs into striatal MSN. Also, the role of DLX2 and MASH1 is linked to the development of striatal MSN, dopaminergic receptors of MSN and GABAergic MSN phenotypes, and hence DLX2 and MASH1 together drive the LGE progenitors towards striatal GABAergic MSN marker.

LGE precursor cells induced to differentiate into DARPP-32^{+ve} MSN through both overexpression of GSX2 and DLX2 in H9 nrNPCs and overexpression of DLX2 alone or DLX2, GSX2 and MASH1 in 34D6 nrNPCs. However, the MSNs did not elicit action potentials, and hence they were generated immature MSNs. Also, the co-expression of GSX2 with DLX2 and/or MASH1 was not able to push the DARPP-32^{+ve} MSN into mature neurons. The constrained action of GSX2 on maturity has no definite explanation. However, It has been reported that in the Gsx2$^{-/-}$ mutant striatum the expression of Darpp-32 was reduced, while the expression of Calbindin, the marker for mature matrix striatal MSNs, was increased (Wang *et al*. 2009). These data speculate that overexpression of GSX2 has a role in the development of MSNs but these are not of mature phenotype. Meanwhile, the co-expression of DLX2 and MASH1 was able to generate mature, yet functional DARPP-32^{+ve} MSN. Therefore, IWR-1 (GSX2) pre-treatment of stem cells was used to generate normal expression of GSX2 in cell culture and then the cells were nucleofected with MASH1 and DLX2 construct to investigate the effect of external GSX2-mediated expression on MSN differentiation and the neuronal maturation.

34D6 cultured with or without IWR-1 differentiated into MSNs. IWR-1 cultures in 34D6 nrNPCs induced the expression of GSX2 in cell culture (unpublished data from the Allen lab). The 34D6 cells cultured with IWR-1 were more mature as evident by fired action potential than those cultured without IWR-1. However, the generation of functional MSN-like cells upon IWR-1 pre-treated 34D6 nrNPCs was merely for a short time at W3. This was supported by a sharp decrease of biomarkers at W6. This phenomenon could be as a result of cell toxicity or conflict in exogenously induced GSX2 expression with endogenous DLX2/MASH1 expression. Therefore, an extra work is required to understand what causes hindered development of long lasting MSNs in this model.

Neuronal maturation in the nucleofected 34D6 nrNPCs with DLX2 and MASH1 construct was investigated through assessment of functional properties by the whole cell patch clamp. Via use of ectopic expression of DLX2 and MASH1 to mature 34D6-derived MSNs generated cells is able to fire action potentials and having a depolarized resting membrane potential (mean -46 mV at week 6). Furthermore, it was observed that 68.4, 100 and 87.5% of patched neurons show spontaneous generating of action potentials at week 3, 4 and 6 respectively in the nucleofected 34D6 with DLX2 and MASH1 plus treatment of 34D6 with IWR-1. Also, the resting potential membrane was -46.5±1.0 mV, -49.2±2.7 mV and -52.3±3.5 mV at week 3, 4 and 6. It was shown that the hESC-derived MSNs by using SHH and DKK1 have -43±4.9 mV of resting membrane potential (Carri *et al*. 2013). These data suggest that the DLX2 and MASH1 construct, with or without IWR-1 in the culture, has the ability to generate mature MSNs that can produce an action potential.

The successive steps in differentiating hPSCs that generate GABAergic neurons are efficiently regulated by the dual action of MASH1 and DLX2 applying the current protocol as evident by the expression of various biomarkers and electrophysiology.

7.3 An alternative protocol was performed in this study compared with morphogen strategy

Developmental cues, such as WNT and SHH, to differentiate hPSCs directly towards MSN were previously performed (Carri *et al*. 2013; Nicoleau *et al*. 2013; Ma *et al*. 2012). However, this strategy generates a mixture of cell types (Review in Soldner and Jaenisch 2012). According to previous papers (Carri *et al*. 2013; Nicoleau *et al*. 2013; Ma *et al*. 2012), their protocol could generates a mixture of LGE and MGE specific progenitors which was carried out for terminally differentiated into MSNs. It was found that 200 ng/ml of SHH promoted LGE-like neurons (Carri *et al*. 2013). Meanwhile, in a paper published by Nicoleau *et al*. (2013), it was shown that 50 ng/ml of SHH in hPSCs promoted the LGE-like neurons that expressed GSX2^{+ve}/NKX2.1^{-ve}, and that the expression level of NKX2.1 was increased when 200 ng/ml of SHH was used in culture (Nicoleau *et al*. 2013). Therefore, a more-defined and efficient route for neuron differentiation into MSNs is required. The requirements of a protocol for the induction of uniformed disease-specific cell types *in vitro*, that can be used for iPSCs-disease modeling and cell replacement therapy, are to produce and highly specific iPSC-disease-specific cell types with high efficiency.

In this study, the strategy of direct reprogramming by ectopic expression of TF was used to introduce multiple candidate genes into hPSC-derived nrNPCs and to allow the selection and reproducibility of a disease-specific cell type (MSN). This protocol also generated PAX6^{-ve}/DLX2^{+ve}/MASH1^{+ve}/NKX2.1^{-ve} LGE-specific progenitors that differentiated further into mature MSNs. This is in line with Pauly *et al*. (2013) finding showing that the LGE domains, *in vivo*, were DLX2^{+ve}/NKX2.1^{-ve}/ PAX6^{-ve}, while the MGE domains were DLX2^{+ve}/NKX2.1^{+ve}. These data indicate that the direct differentiated MSNs, in this study, was well-defined and efficient.

In 2014, a paper by Victor *et al*. 2014 was published using essentially the same protocol as in this study, but with slight differences in the cell line, type and the genes that were ectopically expressed. MicroRNA (mRNA)-9/9 and mRNA-124 were used for direct programming, as they have a role, during neural development, in an ATP-dependent chromatin-remodeling regulation process that is essential for functional neuronal differentiation (Wu *et al*. 2009; Yoo *et al*. 2009; Staahl *et al*. 2013). As the neuronal progenitors exit cell cycle, migrate and differentiate, the process of chromatin-remodeling is taking place where the neural precursor Brg/Brm-associated factor switches through a change in conformation to neuron specific Brg/Brm-associated factor. This process typically is driven by mRNA-9/9 and mRNA-124 (Staahl *et al*. 2013). TFs MASH1, NEUROD2 and MYT1L (MNM), used in conjunction with mRNA-9/9 and mRNA-142 (mRNA9/9-142) can direct differentiation into neurons (Yoo *et al*. 2011). However, the outcome of this procedure was a mixture of inhibitory and excitatory neurons. Victor *et al*. (2014) used mRNA9/9-142 with TFs CTIP2, DLX1, DLX2 (CDM) and MYT1L to replace the TFs with brain enriched genes that characterised the MSN so as to promote mRNA9/9-142 to mediate the neuronal differentiation into more specific cell types such as striatal MSNs (Victor *et al*. 2014). This work directly reprogrammed MSNs from human postnatal cells by lentiviral transduction (Victor *et al*. 2014). The differentiated cells expressed 70% of DARPP-32 in this case, whereas in our study, 85% of DARPP-32 and CTIP2 were expressed. In addition, it was found that when TFs MNM and CDM were expressed individually in human postnatal cells without the mRNA-9/9-142, the cells were not direct differentiated into neurons and were immuno-negative to MAP2 (Yoo *et al*. 2011; Victor *et al*. 2014). However, the reprogramming of hPSCs by TFs DLX2 and MASH1 has been shown in this work to convert cells successfully into LGE-like neurons that then terminally differentiate to mature MSN.

7.4 Limitation of study

It is generally known that the efficiency of neuronal nucleofection with plasmids is poor. In this study, it was shown that the efficiency of expression of vector nucleofection into hPSC-derived nrNPCs before G418 selection was around 26% that was increased to around 80% after selection. While a lentiviral system has advantages, screening of multiple constructs by creating them as plasmids is more straightforward. Promising candidates emerging from this primary test can then be cloned further to create the lentiviral constructs that can increase efficiency of transduction. Thus, lentiviral transduction, as used by Victor *et al*. (2014) to differentiate human postnatal cells, can be used to obtain high yield, stable TF transduction into hPSCs-derived nrNPCs.

7.5 Future work

The successful development of a protocol to direct the differentiation of iPSCs into functionally mature MSNs opens new avenues for future research. Future work will include establishing a disease model for HD using the following methods. Skin fibroblasts from HD patients can be collected to generate stable iPSCs derived using established methods from Takahashi, Yamanaka and their colleagues, who used the retroviral introduction of ESC TFs, such as, OCT3/4, SOX2, KLF4 and c-MYC (Takahashi *et al*. 2007; Takahashi and Yamanaka 2006). Other options for generating HD patient-derived iPSCs involve using non-retroviral vectors such as polycistronic vectors (Sommer *et al*. 2009), piggyBac transposons (Woltjen *et al*. 2009), transient episomal delivery (Okita *et al*. 2008), RNA (Warren *et al*. 2010) and even protein delivery (Kim *et al*. 2009). These HD patient-derived iPSCs can then be differentiated into neural progenitor cells to generate NSC that characterised by self-renewal and proliferation. Finally, these NSC can be differentiated into specific neuron types, such as mature MSNs, using the protocol described in this thesis, which is the ectopic expression of DLX2 and MASH1. Future successful generation of MSN-like cells from HD-patient specific iPSCs, that recapitulate the phenotype of HD neuropathology, can facilitate HD disease modeling *in vitro* to understand HD progression, cell replacement therapy for human medical trials and therapeutic drug screening **(Figure 7.1)**. In addition, generating HD-iPSCs with different CAG repeat lengths is recommended as it was reported that HD cell lines with longer CAG repeats were most vulnerable to BDNF withdrawal and cellular stressors (Consortium 2012).

In addition, the use of an isogenic genetically modified iPSCs model could be used to replace the mutated HTT gene with a healthy sequence gene using zinc finger-mediated gene transfer or BAC-mediated homologous recombination. According to An *et al*. (2012), the BAC technique was used to correct the mutated HTT in iPSCs derived from an HD patient, and the characterisation of iPSCs were maintained (An *et al*. 2012). Therefore, this technique could be used to generate an isogenic control model from HD-iPSCs, and then this study protocol could be used to direct differentiation into MSNs. The generation of successful MSNs from genetically modified iPSCs would be used for HD disease modeling and hence for cell replacement therapy **(Figure 7.1)**.

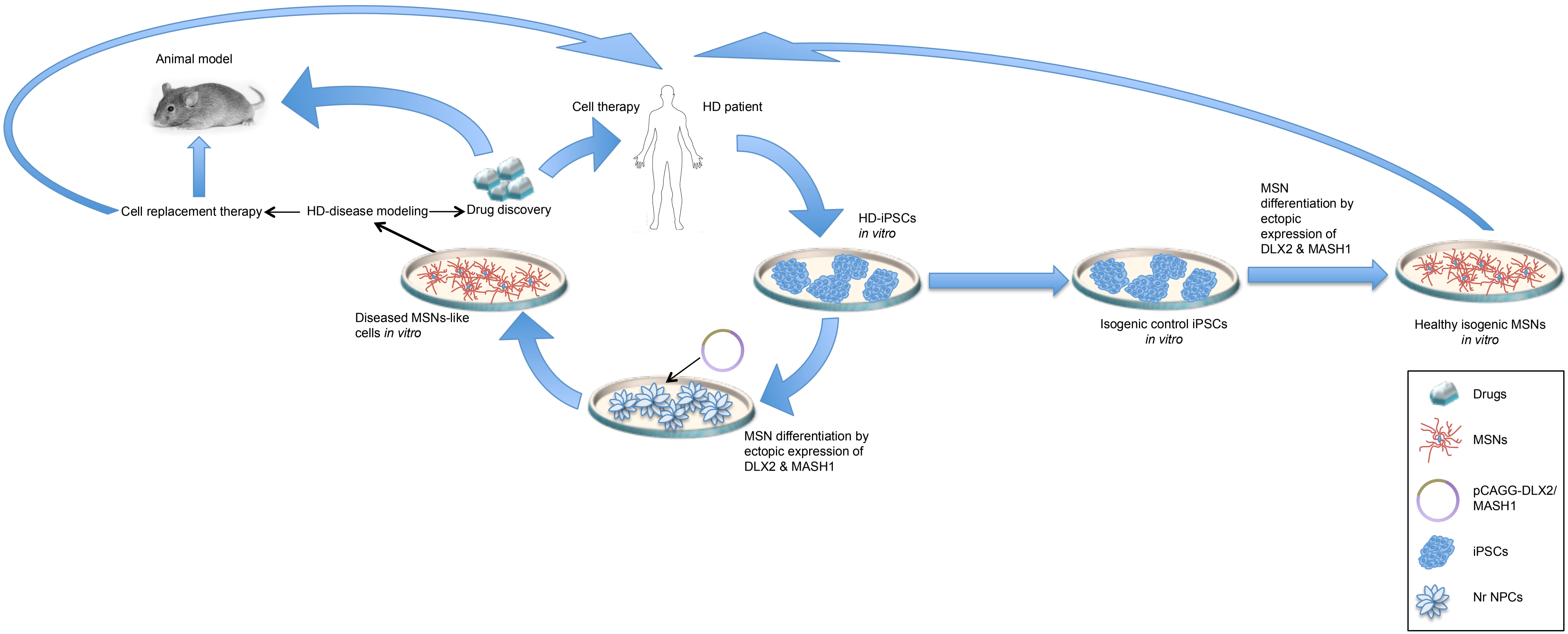


Figure 7.1: Schematic diagram of future work

From HD-patients, there are two approaches to direct differentiate HD-iPSC derived nrNPCs into diseased MSNs-like cells or isogenic (engineered) MSNs-like cells. Direct differentiation of HD-patient specific iPSCs into diseased MSNs-like cells can be generated by ectopic expression of DLX2 and MASH1. This protocol can be used for HD disease modeling *in vitro* to examine the characteristics of diseased cells, by ion channel activity, caspase activity, gene expression and protein processing modification to discover drugs that can then be used in animal models as well as clinical trials. Moreover, the MSNs, derived from HD-iPSCs, could be injected into animals to improve the current achievements of cell replacement therapy. The other approach is to generate isogenic control iPSCs from HD-iPSCs *in vitro* which then can be differentiated into healthy MSNs-like cells by the ectopic expression protocol of this study, and hence it can be used for cell replacement therapy.

Over the last 4 years, there have been several studies that differentiated hPSCs into MSN-like cells and used these for cell replacement therapy in the rodent brain. However, functional studies, such as the motor and cognitive roles as well as electrophysiological connectivity, were not fully carried out. Consequently, functional effectiveness in further complex screening is required, such as the characterisation of motor and cognitive roles as well as electrophysiological connectivity in more detail. Examples of some tests for functional studies include: IntelliCage (Krackow *et al*. 2010), Giant Analysis System that includes Trendmill and GaintScan software (Malone *et al*. 1998), Conditioned Place Preference (CPP) (Rosecrans *et al*. 2009) and Ultrasonic Vocalisation Analysis System (UVAS) (Branchi *et al*. 2001).

7.6 Weakness and strength of the thesis

7.6.1 Weakness of the thesis

All combinations of TFs with the exception of MASH1 and GSX2 (pcagg-MASH1/GSX2), and each TF alone apart from GSX2 (pCAGG-GSX2) were sub-cloned to the expression vector pCAGG. The overexpression of MASH1 and GSX2 in hPSCs, and their effects on MSN reprogramming should be investigated. It was shown that GSX2 (Méndez-Gómez and Vicario-Abejón 2012) and MASH1 (Fode *et al*. 2000) antagonize cortical fates, and they are both essential in determining the genetic network for the development of post mitotic MSNs, since cells migrate and differentiate in the MZ. In addition, overexpression of GSX2 in hPSCs and its role in gene expression of dorso-ventral markers, especially the ventral markers such as DLX2 and MASH1, in comparison with other expression vectors such as pCAGG-DLX2, pCAGG-MASH1, pCAGG-DLX2/MASH1, PCAGG-DLX2/GSX2 and pCAGG-DLX2/MASH1/GSX2 should be considered since the expression of ventral markers was up-regulated by GSX2. However, the sub-cloning of GSX2 in pCAGG-MASH1 to construct (pCAGG-MASH1/GSX2), or in pCAGG to construct (pCAGG-GSX2) was not finished on time.

7.6.2 Strength of the thesis

One of the strengths of this thesis is the use of hPSCs to differentiate into MSNs. The protocol achieved in this study will enable differentiation of HD-patient specific iPSCs into MSN. Generate of an HD-iPSCs disease model would lead to an improved understanding of the pathophysiology of HD.

Two cell lines were used for direct differentiation into MSNs: hESCs (H9) and h-iPSCs (34D6). Consequently, in this study it has been shown that the strategy for direct reprogramming of MSN by ectopic expression is effective in two different cell lines. The resulting hPS-derived MSNs are abundant, express the profile of several striatal GABAergic markers and most notably are able to generate spontaneous action potentials. Consequently the strategy used in this study has the potential to be transferred to HD-patient specific iPSCs and to produce diseased MSNs *in vitro*.

In conclusion, this study provides the first example of ectopic expression of TFs (such as DLX2 and MASH1) in h-iPSCs to direct their differentiation into MSN-like cells. Overexpression of MASH1 and DLX2 in h-iPSCs-derived nrNPCs promoted generation of fully differentiated MSN that expresses FOXP2/EBF1/CTIP2/DARPP-32/DRD1&2/GAD2 and CALBIN-1. These MSN can be used for HD-modeling to understand the mechanisms of neurodegeneration in human HD, facilitate the development of a potential cell-based therapy and aid in therapeutic drug screening.

Bibliography

Aguirre, A., Rubio, M.E. and Gallo, V. 2010. Notch and EGFR pathway interaction regulates neural stem cell number and self-renewal. *Nature* 467(7313), pp. 323–327.

Albin, R.L., Young, a B. and Penney, J.B. 1989. The functional anatomy of basal ganglia disorders. *Trends in neurosciences* 12(10), pp. 366–375.

Alexopoulou, A.N., Couchman, J.R. and Whiteford, J.R. 2008. The CMV early enhancer/chicken beta actin (CAG) promoter can be used to drive transgene expression during the differentiation of murine embryonic stem cells into vascular progenitors. *BMC cell biology* 9, p. 2. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2254385&tool=pmcentrez&rendertype=abstract.

Altar, C. a, Cai, N., Bliven, T., Juhasz, M., Conner, J.M., Acheson, a L., Lindsay, R.M. and Wiegand, S.J. 1997. Anterograde transport of brain-derived neurotrophic factor and its role in the brain. *Nature* 389(6653), pp. 856–860.

Ambasudhan, R., Talantova, M., Coleman, R., Yuan, X., Zhu, S., Lipton, S. a and Ding, S. 2011. Direct reprogramming of adult human fibroblasts to functional neurons under defined conditions. *Cell stem cell* 9(2), pp. 113–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21802386.

An, M.C., Zhang, N., Scott, G., Montoro, D., Wittkop, T., Mooney, S., Melov, S. and Ellerby, L.M. 2012. Genetic correction of Huntington's disease phenotypes in induced pluripotent stem cells. *Cell stem cell* 11(2), pp. 253–63. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3608272&tool=pmcentrez&rendertype=abstract.

Anderson, K.D. and Reiner, a 1991. Immunohistochemical localization of DARPP-32 in striatal projection neurons and striatal interneurons: implications for the localization of D1-like dopamine receptors on different types of striatal neurons. *Brain research* 568(1-2), pp. 235–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1839966.

Anderson, S. a, Marín, O., Horn, C., Jennings, K. and Rubenstein, J.L. 2001. Distinct cortical migrations from the medial and lateral ganglionic eminences. *Development (Cambridge, England)* 128(3), pp. 353–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11152634.

Anderson, S. a, Qiu, M., Bulfone, A., Eisenstat, D.D., Meneses, J., Pedersen, R. and Rubenstein, J.L. 1997a. Mutations of the homeobox genes Dlx-1 and Dlx-2 disrupt the striatal subventricular zone and differentiation of late born striatal neurons. *Neuron* 19(1), pp. 27–37. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9247261.

Anderson, S. a. 1997b. Interneuron Migration from Basal Forebrain to Neocortex: Dependence on Dlx Genes. *Science* 278(5337), pp. 474–476. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.278.5337.474.

Araki, K.Y., Sims, J.R. and Bhide, P.G. 2007. Dopamine receptor mRNA and protein expression in the mouse corpus striatum and cerebral cortex during pre- and postnatal development. *Brain Research* 1156(1), pp. 31–45.

Arber, C., Preclous, S. V., Cambray, S., Rlsner-Janlczek, J.R., Kelly, C., Noakes, Z., Fjodorova, M., Heuer, A., Umgless, M.A., Rodriguez, T.A., Rosser, A.E., Dunnet, S.B.

and Li, M. 2015. Activin A directs striatal projection neuron differentiation of human pluripotent stem cells. *Development* 142, pp. 1375–1386.

Arlotta, P., Molyneaux, B.J., Jabaudon, D., Yoshida, Y. and Macklis, J.D. 2008. Ctip2 controls the differentiation of medium spiny neurons and the establishment of the cellular architecture of the striatum. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28(3), pp. 622–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18199763.

Artavanis-Tsakonas, S., Matsuno, K. and Fortini, M.E. 1995. Notch signaling. *Science (New York, N.Y.)* 268(5208), pp. 225–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7716513.

Aubert, J., Dunstan, H., Chambers, I. and Smith, A. 2002. Functional gene screening in embryonic stem cells implicates Wnt antagonism in neural differentiation. *Nature biotechnology* 20(12), pp. 1240–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12447396.

Aubry, L., Bugi, A., Lefort, N., Rousseau, F., Peschanski, M. and Perrier, A.L. 2008. Striatal progenitors derived from human ES cells mature into DARPP32 neurons in vitro and in quinolinic acid-lesioned rats. *Proceedings of the National Academy of Sciences of the United States of America* 105(43), pp. 16707–12. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2575484&tool=pmcentrez&rendertype=abstract.

Augood, S.J., Faull, R.L. and Emson, P.C. 1997. Dopamine D1 and D2 receptor gene expression in the striatum in Huntington's disease. *Annals of neurology* 42(2), pp. 215–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9266732.

Azzouz, M., Martin-Rendon, E., Barber, R.D., Mitrophanous, K.A., Carter, E.E., Rohll, J.B., Kingsman, S.M., Kingsman, A.J. and Mazarakis, N.D. 2002. Multicistronic lentiviral vector-mediated striatal gene transfer of aromatic L-amino acid decarboxylase, Tyrosine hydroxylase, and GTP cyclohydrolase I induces sustained transgene expression, dopamine production, and functional improvement in a Rat model. *The Journal of neuroscience* 22(23), pp. 10302–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12451130.

Backman, M., Machon, O., Mygland, L., van den Bout, C.J., Zhong, W., Taketo, M.M. and Krauss, S. 2005. Effects of canonical Wnt signaling on dorso-ventral specification of the mouse telencephalon. *Developmental biology* 279(1), pp. 155–68. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15708565.

Bain, G., Kitchens, D., Yao, M., Huettner, J.E. and Gottlieb, D.I. 1995. Embryonic stem cells express neuronal properties in vitro. *Developmental biology* 168(2), pp. 342–57. Available at: http://www.sciencedirect.com/science/article/pii/S0012160685710858.

Benraiss, A. and Goldman, S. a 2011. Cellular therapy and induced neuronal replacement for Huntington's disease. *Neurotherapeutics : the journal of the American Society for Experimental NeuroTherapeutics* 8(4), pp. 577–90. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3250300&tool=pmcentrez&rendertype=abstract.

Ten Berge, D., Koole, W., Fuerer, C., Fish, M., Eroglu, E. and Nusse, R. 2008. Wnt signaling mediates self-organization and axis formation in embryoid bodies. *Cell stem* *cell* 3(5), pp. 508–18. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2683270&tool=pmcentrez&rendertype=abstract.

Berninger, B., Costa, M.R., Koch, U., Schroeder, T., Sutor, B., Grothe, B. and Götz, M. 2007. Functional properties of neurons derived from in vitro reprogrammed postnatal astroglia. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27(32), pp. 8654–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17687043.

Berninger, B., Guillemot, F. and Götz, M. 2007. Directing neurotransmitter identity of neurones derived from expanded adult neural stem cells. *The European journal of neuroscience* 25(9), pp. 2581–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17561834.

Bertrand, N., Castro, D.S. and Guillemot, F. 2002. Proneural genes and the specification of neural cell types. *Nature reviews. Neuroscience* 3(7), pp. 517–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12094208.

Bibb, J. a, Snyder, G.L., Nishi, a, Yan, Z., Meijer, L., Fienberg, a a, Tsai, L.H., Kwon, Y.T., Girault, J. a, Czernik, a J., Huganir, R.L., Hemmings, H.C., Nairn, a C. and Greengard, P. 1999. Phosphorylation of DARPP-32 by Cdk5 modulates dopamine signalling in neurons. *Nature* 402(6762), pp. 669–671.

Bilican, B., Serio, A., Barmada, S.J., Nishimura, A.L., Sullivan, G.J., Carrasco, M., Phatnani, H.P., Puddifoot, C. a, Story, D., Fletcher, J., Park, I.-H., Friedman, B. a, Daley, G.Q., Wyllie, D.J. a, Hardingham, G.E., Wilmut, I., Finkbeiner, S., Maniatis, T., Shaw, C.E. and Chandran, S. 2012. Mutant induced pluripotent stem cell lines recapitulate aspects of TDP-43 proteinopathies and reveal cell-specific vulnerability. *Proceedings of the National Academy of Sciences of the United States of America* 109(15), pp. 5803–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3326463&tool=pmcentrez&rendertype=abstract.

Bishop, K.M., Garel, S., Nakagawa, Y., Rubenstein, J.L.R. and O'Leary, D.D.M. 2003. Emx1 and Emx2 cooperate to regulate cortical size, lamination, neuronal differentiation, development of cortical efferents, and thalamocortical pathfinding. *The Journal of comparative neurology* 457(4), pp. 345–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12561075.

Bithell, A., Johnson, R. and Buckley, N.J. 2009. Transcriptional dysregulation of coding and non-coding genes in cellular models of Huntington's disease. *Biochemical Society transactions* 37(Pt 6), pp. 1270–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19909260.

Branchi, I., Santucci, D. and Alleva, E. 2001. Ultrasonic vocalisation emitted by infant rodents: a tool for assessment of neurobehavioural development. *Behavioural Brain Research* 125(1-2), pp. 49–56. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0166432801002777.

Caiazzo, M., Dell'Anno, M.T., Dvoretskova, E., Lazarevic, D., Taverna, S., Leo, D., Sotnikova, T.D., Menegon, A., Roncaglia, P., Colciago, G., Russo, G., Carninci, P., Pezzoli, G., Gainetdinov, R.R., Gustincich, S., Dityatev, A. and Broccoli, V. 2011. Direct generation of functional dopaminergic neurons from mouse and human fibroblasts. *Nature* 476(7359), pp. 224–7. Available at: http://dx.doi.org/10.1038/nature10284.

Camnasio, S., Carri, A.D., Lombardo, A., Grad, I., Mariotti, C., Castucci, A., Rozell, B., Riso, P. Lo, Castiglioni, V., Zuccato, C., Rochon, C., Takashima, Y., Diaferia, G., Biunno, I., Gellera, C., Jaconi, M., Smith, A., Hovatta, O., Naldini, L., Di Donato, S., Feki, A. and Cattaneo, E. 2012. The first reported generation of several induced pluripotent stem cell lines from homozygous and heterozygous Huntington's disease patients demonstrates mutation related enhanced lysosomal activity. *Neurobiology of Disease* 46, pp. 41–51.

Carney, R.S.E., Cocas, L. a, Hirata, T., Mansfield, K. and Corbin, J.G. 2009. Differential regulation of telencephalic pallial-subpallial boundary patterning by Pax6 and Gsh2. *Cerebral cortex (New York, N.Y. : 1991)* 19(4), pp. 745–59. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2651477&tool=pmcentrez&rendertype=abstract.

Carri, A.D., Onorati, M., Lelos, M.J., Castiglioni, V., Faedo, A., Menon, R., Camnasio, S., Vuono, R., Spaiardi, P., Talpo, F., Toselli, M., Martino, G., Barker, R. a, Dunnett, S.B., Biella, G. and Cattaneo, E. 2013. Developmentally coordinated extrinsic signals drive human pluripotent stem cell differentiation toward authentic DARPP-32+ medium-sized spiny neurons. *Development (Cambridge, England)* 140, pp. 301–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23250204.

Casarosa, S., Fode, C. and Guillemot, F. 1999. Mash1 regulates neurogenesis in the ventral telencephalon. *Development (Cambridge, England)* 126(3), pp. 525–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9876181.

Castro, D.S., Martynoga, B., Parras, C., Ramesh, V., Pacary, E., Johnston, C., Drechsel, D., Lebel-Potter, M., Garcia, L.G., Hunt, C., Dolle, D., Bithell, A., Ettwiller, L., Buckley, N. and Guillemot, F. 2011. A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. *Genes & development* 25(9), pp. 930–45. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3084027&tool=pmcentrez&rendertype=abstract.

Castro, D.S., Skowronska-Krawczyk, D., Armant, O., Donaldson, I.J., Parras, C., Hunt, C., Critchley, J. a, Nguyen, L., Gossler, A., Göttgens, B., Matter, J.-M. and Guillemot, F. 2006. Proneural bHLH and Brn proteins coregulate a neurogenic program through cooperative binding to a conserved DNA motif. *Developmental cell* 11(6), pp. 831–44. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17141158.

Cattaneo, E., Rigamonti, D., Goffredo, D., Zuccato, C., Squitieri, F. and Sipione, S. 2001. Loss of normal huntingtin function: New developments in Huntington's disease research. *Trends in Neurosciences* 24(3), pp. 182–188.

Cattaneo, E., Zuccato, C. and Tartari, M. 2005. Normal huntingtin function: an alternative approach to Huntington's disease. *Nature reviews. Neuroscience* 6(12), pp. 919–930.

Chambers, S.M., Fasano, C. a, Papapetrou, E.P., Tomishima, M., Sadelain, M. and Studer, L. 2009. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature biotechnology* 27(3), pp. 275–80. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2756723&tool=pmcentrez&rendertype=abstract.

Chan, A.W.S., Cheng, P.-H., Neumann, A. and Yang, J.-J. 2010. Reprogramming Huntington monkey skin cells into pluripotent stem cells. *Cellular reprogramming* 12, pp. 509–517.

Chang, H. and Kitai, S. 1985. Projection neurons of the nucleus accumbens: an intracellular labeling study. *Brain research* 347, pp. 112–16. Available at: http://www.sciencedirect.com/science/article/pii/0006899385908947.

Chang, H.T., Wilson, C.J. and Kitai, S.T. 1982. A Golgi study of rat neostriatal neurons: light microscopic analysis. *The Journal of comparative neurology* 208(2), pp. 107–26. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6181102.

Chang, H.Y. and Cotsarelis, G. 2007. Turning skin into embryonic stem cells. *Nature medicine* 13, pp. 783–784.

Chiang, C., Litingtung, Y., Lee, E. and Young, K. 1996. Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* 383, pp. 407–413. Available at: http://www.nature.com/nature/journal/v383/n6599/abs/383407a0.html.

Chitnis, a and Kintner, C. 1996. Sensitivity of proneural genes to lateral inhibition affects the pattern of primary neurons in Xenopus embryos. *Development (Cambridge, England)* 122(7), pp. 2295–301. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8681809.

Churchill, J.D., Tharp, J. a, Wellman, C.L., Sengelaub, D.R. and Garraghty, P.E. 2004. Morphological correlates of injury-induced reorganization in primate somatosensory cortex. *BMC neuroscience* 5, p. 43. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=529444&tool=pmcentrez&rendertype=abstract.

Cobos, I., Borello, U. and Rubenstein, J.L.R. 2007. Dlx transcription factors promote migration through repression of axon and dendrite growth. *Neuron* 54(6), pp. 873–88. Available at: http://www.sciencedirect.com/science/article/pii/S0896627307003819.

Colasante, G., Collombat, P., Raimondi, V., Bonanomi, D., Ferrai, C., Maira, M., Yoshikawa, K., Mansouri, A., Valtorta, F., Rubenstein, J.L.R. and Broccoli, V. 2008. Arx is a direct target of Dlx2 and thereby contributes to the tangential migration of GABAergic interneurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28, pp. 10674–10686.

Colombo, E., Collombat, P., Colasante, G., Bianchi, M., Long, J., Mansouri, A., Rubenstein, J.L.R. and Broccoli, V. 2007. Inactivation of Arx, the murine ortholog of the X-linked lissencephaly with ambiguous genitalia gene, leads to severe disorganization of the ventral telencephalon with impaired neuronal migration and differentiation. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27, pp. 4786–4798.

Connor, B. 2011. Adult neural progenitor cells and cell replacement therapy for huntington disease Appasani, K. and Appasani, R. K. eds. *stem cells and regenerative medicien*, pp. 299–314. Available at: http://link.springer.com/10.1007/978-1-60761-860-7.

Consortium, Th. 2012. Induced pluripotent stem cells from patients with Huntington's disease show CAG-repeat-expansion-associated phenotypes. *Cell stem cell* 11(2), pp. 264–78. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3804072&tool=pmcentrez&rendertype=abstract.

Corbin, J.G., Gaiano, N., Machold, R.P., Langston, a and Fishell, G. 2000. The Gsh2 homeodomain gene controls multiple aspects of telencephalic development. *Development (Cambridge, England)* 127(23), pp. 5007–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11060228.

Crossley, P., Martinez, S., Ohkubo, Y. and Rubenstein, J.L. 2001. Coordinate expression of Fgf8, Otx2, Bmp4, and Shh in the rostral prosencephalon during development of the telencephalic and optic vesicles. *Neuroscience* 108(2), pp. 183–206. Available at: http://www.sciencedirect.com/science/article/pii/S030645220 1004110.

Cui, L., Jeong, H., Borovecki, F., Parkhurst, C.N., Tanese, N. and Krainc, D. 2006. Transcriptional Repression of PGC-1α by Mutant Huntingtin Leads to Mitochondrial Dysfunction and Neurodegeneration. *Cell* 127(1), pp. 59–69.

Dale, L. and Wardle, F.C. 1999. A gradient of BMP activity specifies dorsal-ventral fates in early Xenopus embryos. *Seminars in cell & developmental biology* 10(3), pp. 319–26. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10441546.

Danesin, C., Peres, J.N., Johansson, M., Snowden, V., Cording, A., Papalopulu, N. and Houart, C. 2009. Integration of telencephalic Wnt and hedgehog signaling center activities by Foxg1. *Developmental cell* 16(4), pp. 576–87. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19386266.

Dehay, C. and Kennedy, H. 2007. Cell-cycle control and cortical development. *Nature reviews. Neuroscience* 8(6), pp. 438–50. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17514197.

Dhara, S.K. and Stice, S.L. 2008. Neural differentiation of human embryonic stem cells. *Journal of Cellular Biochemistry* 105(3), pp. 633–640.

Ding, D., Xu, L., Xu, H., Li, X., Liang, Q., Zhao, Y. and Wang, Y. 2014. Mash1 efficiently reprograms rat astrocytes into neurons. *Neural regeneration research* 9(1), pp. 25–32. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4146312&tool=pmcentrez&rendertype=abstract.

Doetschman, T.C., Eistetter, H., Katz, M., Schmidt, W. and Kemler, R. 1985. The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *Journal of embryology and experimental morphology* 87, pp. 27–45. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3897439.

Donnelly, M.L., Luke, G., Mehrotra, a, Li, X., Hughes, L.E., Gani, D. and Ryan, M.D. 2001. Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. *The Journal of general virology* 82(Pt 5), pp. 1013–25. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11297676.

Ebert, A.D., Yu, J., Jr, F.F.R., Mattis, V.B., Christian, L., Thomson, J.A. and Svendsen, C.N. 2009. Induced pluripotent stem cells from a spinal muscular atrophy patient. 457(7227), pp. 277–280.

Efe, J.A., Hilcove, S., Kim, J., Zhou, H., Ouyang, K., Wang, G., Chen, J. and Ding, S. 2011. Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy. *Nature cell biology* 13(3), pp. 215–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21278734.

Ehrlich, M.E., Conti, L., Toselli, M., Taglietti, L., Fiorillo, E., Taglietti, V., Ivkovic, S., Guinea, B., Tranberg, A., Sipione, S., Rigamonti, D. and Cattaneo, E. 2001. ST14A cells have properties of a medim-size spiny neuron. *Experimental neurology* 167, pp. 215–226.

Ehrlich, M.E., Rosen, N.L., Kurihara, T., Shalaby, I.A. and Greengard, P. 1990. DARPP-32 development in the caudate nucleus is independent of afferent input from the substantia nigra. *Brain research. Developmental brain research* 54, pp. 257–263.

Eisenstat, D.D., Liu, J.K., Mione, M., Zhong, W., Yu, G., Anderson, S.A., Ghattas, I., Puelles, L. and Rubenstein, J.L. 1999. DLX-1, DLX-2, and DLX-5 expression define distinct stages of basal forebrain differentiation. *The Journal of comparative neurology* 414(2), pp. 217–37. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10516593.

Ericson, J., Muhr, J., Placzek, M., Lints, T., Jessell, T.M. and Edlund, T. 1995. Sonic hedgehog induces the differentiation of ventral forebrain neurons: a common signal for ventral patterning within the neural tube. *Cell* 81(5), pp. 747–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7774016.

Evans, a E., Kelly, C.M., Precious, S. V and Rosser, a E. 2012. Molecular regulation of striatal development: a review. *Anatomy research international* 2012, p. 106529. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3335634&tool=pmcentrez&rendertype=abstract.

Evans, M.J. and Kaufman, M.H. 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292, pp. 154–156.

Ferino, F., Thierry, a M., Saffroy, M. and Glowinski, J. 1987. Interhemispheric and subcortical collaterals of medial prefrontal cortical neurons in the rat. *Brain research* 417(2), pp. 257–266.

Ferland, R.J., Cherry, T.J., Preware, P.O., Morrisey, E.E. and Walsh, C. a 2003. Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. *The Journal of comparative neurology* 460(2), pp. 266–79. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12687690.

Fernandez, É., Schiappa, R., Girault, J.A. and Le Novère, N. 2006. DARPP-32 is a robust integrator of dopamine and glutamate signals. *PLoS Computational Biology* 2(12), pp. 1619–1633.

Feyder, M., Bonito-Oliva, A. and Fisone, G. 2011. L-DOPA-Induced Dyskinesia and Abnormal Signaling in Striatal Medium Spiny Neurons: Focus on Dopamine D1 Receptor-Mediated Transmission. *Frontiers in behavioral neuroscience* 5(October), pp. 1–11. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3199545&tool=pmcentrez&rendertype=abstract.

Fienberg, a a, Hiroi, N., Mermelstein, P.G., Song, W., Snyder, G.L., Nishi, a, Cheramy, a, O'Callaghan, J.P., Miller, D.B., Cole, D.G., Corbett, R., Haile, C.N., Cooper, D.C., Onn, S.P., Grace, a a, Ouimet, C.C., White, F.J., Hyman, S.E., Surmeier, D.J., Girault, J., Nestler, E.J. and Greengard, P. 1998. DARPP-32: regulator of the efficacy of dopaminergic neurotransmission. *Science (New York, N.Y.)* 281(5378), pp. 838–842.

Fode, C., Ma, Q., Casarosa, S., Ang, S.L., Anderson, D.J. and Guillemot, F. 2000. A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons. *Genes & development* 14, pp. 67–80.

Foster, G. a, Schultzberg, M., Hökfelt, T., Goldstein, M., Hemmings, H.C., Ouimet, C.C., Walaas, S.I. and Greengard, P. 1987. Development of a dopamine- and cyclic adenosine 3':5'-monophosphate-regulated phosphoprotein (DARPP-32) in the prenatal rat central nervous system, and its relationship to the arrival of presumptive dopaminergic innervation. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 7(7), pp. 1994–2018. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2886563.

Furuta, Y., Piston, D.W. and Hogan, B.L. 1997. Bone morphogenetic proteins (BMPs) as regulators of dorsal forebrain development. *Development (Cambridge, England)* 124(11), pp. 2203–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9187146.

Gao, a, Peng, Y., Deng, Y. and Qing, H. 2012. Potential therapeutic applications of differentiated induced pluripotent stem cells (iPSCs) in the treatment of neurodegenerative diseases. *Neuroscience* (October). Available at: http://www.ncbi.nlm.nih.gov/pubmed/23069758.

Garcia-Dominguez, M., Poquet, C., Garel, S. and Charnay, P. 2003. Ebf gene function is required for coupling neuronal differentiation and cell cycle exit. *Development (Cambridge, England)* 130(24), pp. 6013–6025.

Gardian, G., Browne, S.E., Choi, D.-K., Klivenyi, P., Gregorio, J., Kubilus, J.K., Ryu, H., Langley, B., Ratan, R.R., Ferrante, R.J. and Beal, M.F. 2005. Neuroprotective effects of phenylbutyrate in the N171-82Q transgenic mouse model of Huntington's disease. *The Journal of biological chemistry* 280(1), pp. 556–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15494404.

Garel, S., Marín, F., Grosschedl, R. and Charnay, P. 1999. Ebf1 controls early cell differentiation in the embryonic striatum. *Development (Cambridge, England)* 126(23), pp. 5285–5294.

Gaspard, N., Bouschet, T., Hourez, R., Dimidschstein, J., Naeije, G., van den Ameele, J., Espuny-Camacho, I., Herpoel, A., Passante, L., Schiffmann, S.N., Gaillard, A. and Vanderhaeghen, P. 2008. An intrinsic mechanism of corticogenesis from embryonic stem cells. *Nature* 455(7211), pp. 351–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18716623.

Genetics, S. 2003. GeneSpring User Manual. *Genetics* 346 Pt 2.

Geoffroy, C.G., Critchley, J. a, Castro, D.S., Ramelli, S., Barraclough, C., Descombes, P., Guillemot, F. and Raineteau, O. 2009. Engineering of dominant active basic helix-loop-helix proteins that are resistant to negative regulation by postnatal central nervous system antineurogenic cues. *Stem cells (Dayton, Ohio)* 27(4), pp. 847–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19350686.

Gerfen, C. 1992. The neostriatal mosaic: multiple levels of compartmental organization. *Trends in Neurosciences* 15(4), pp. 133–139. Available at: http://link.springer.com/chapter/10.1007/978-3-7091-9211-5_4.

Gerfen, C., Engber, T., Mahan, L., Susel, Z., Chase, T., Monsma, F. and Sibley, D. 1990. D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. *Science* 250(4986), pp. 1429–1432. Available at: http://www.sciencemag.org/content/250/4986/1429.abstract.

Ghanem, N., Jarinova, O., Amores, A., Long, Q., Hatch, G., Park, B.K., Rubenstein, J.L.R. and Ekker, M. 2003. Regulatory roles of conserved intergenic domains in vertebrate Dlx bigene clusters. *Genome research* 13(4), pp. 533–43. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=430168&tool=pmcentrez&rendertype=abstract.

Gong, S., Zheng, C., Doughty, M.L., Losos, K., Didkovsky, N., Schambra, U.B., Nowak, N.J., Joyner, A., Leblanc, G., Hatten, M.E. and Heintz, N. 2003. A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* 425(6961), pp. 917–25. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14586460.

Goodrich, L. V. 1997. Altered Neural Cell Fates and Medulloblastoma in Mouse patched Mutants. *Science* 277(5329), pp. 1109–1113. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.277.5329.1109.

Greengard, P., Allen, P.B. and Nairn, A.C. 1999. Beyond the dopamine receptor: The DARPP-32/protein phosphatase-1 cascade. *Neuron* 23(3), pp. 435–447.

Grundemann, D. and Schomig, E. 1996. Protection of DNA during preparative agarose gel electrophoresis against damage induced by ultraviolet light. *BioTechniques* 21, pp. 898–903.

Guillemot, F. 2007. Spatial and temporal specification of neural fates by transcription factor codes. *Development (Cambridge, England)* 134(21), pp. 3771–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17898002.

Gulacsi, A. and Anderson, S.A. 2006. Shh maintains Nkx2.1 in the MGE by a Gli3-independent mechanism. *Cerebral cortex (New York, N.Y. : 1991)* 16 Suppl 1(suppl_1), pp. i89–95. Available at: http://cercor.oxfordjournals.org/cgi/content/long/16/suppl_1/i89.

Gutekunst, C. a, Levey, a I., Heilman, C.J., Whaley, W.L., Yi, H., Nash, N.R., Rees, H.D., Madden, J.J. and Hersch, S.M. 1995. Identification and localization of huntingtin in brain and human lymphoblastoid cell lines with anti-fusion protein antibodies. *Proceedings of the National Academy of Sciences of the United States of America* 92(19), pp. 8710–8714.

Han, D.W., Tapia, N., Hermann, A., Hemmer, K., Höing, S., Araúzo-Bravo, M.J., Zaehres, H., Wu, G., Frank, S., Moritz, S., Greber, B., Yang, J.H., Lee, H.T., Schwamborn, J.C., Storch, A. and Schöler, H.R. 2012. Direct reprogramming of fibroblasts into neural stem cells by defined factors. *Cell stem cell* 10(4), pp. 465–72. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22445517.

Hanashima, C., Shen, L., Li, S.C. and Lai, E. 2002. Brain Factor-1 Controls the Proliferation and Differentiation of Neocortical Progenitor Cells through Independent Mechanisms. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 22(15), pp. 6526–6536.

Haskell, G.T. and LaMantia, A.-S. 2005. Retinoic acid signaling identifies a distinct precursor population in the developing and adult forebrain. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 25(33), pp. 7636–47. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16107650.

Heinrich, C., Blum, R., Gascón, S., Masserdotti, G., Tripathi, P., Sánchez, R., Tiedt, S., Schroeder, T., Götz, M. and Berninger, B. 2010. Directing astroglia from the cerebral cortex into subtype specific functional neurons. *PLoS biology* 8(5), p. e1000373. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2872647&tool=pmcentrez&rendertype=abstract.

Helms, A.W., Battiste, J., Henke, R.M., Nakada, Y., Simplicio, N., Guillemot, F. and Johnson, J.E. 2005. Sequential roles for Mash1 and Ngn2 in the generation of dorsal spinal cord interneurons. *Development (Cambridge, England)* 132(12), pp. 2709–19. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1351036&tool=pmcentrez&rendertype=abstract.

Hemmings, H.C., Greengard, P., Tung, H.Y. and Cohen, P. 1984. DARPP-32, a dopamine-regulated neuronal phosphoprotein, is a potent inhibitor of protein phosphatase-1. *Nature* 310(5977), pp. 503–505.

Henke, R.M., Meredith, D.M., Borromeo, M.D., Savage, T.K. and Johnson, J.E. 2009. Ascl1 and Neurog2 form novel complexes and regulate Delta-like3 (Dll3) expression in the neural tube. *Developmental biology* 328(2), pp. 529–40. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2698949&tool=pmcentrez&rendertype=abstract.

Henrique, D., Hirsinger, E., Adam, J., Le Roux, I., Pourquié, O., Ish-Horowicz, D. and Lewis, J. 1997. Maintenance of neuroepithelial progenitor cells by Delta-Notch signalling in the embryonic chick retina. *Current biology : CB* 7(9), pp. 661–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9285721.

Herrup, K. and Yang, Y. 2007. Cell cycle regulation in the postmitotic neuron: oxymoron or new biology? *Nature reviews. Neuroscience* 8(5), pp. 368–78. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17453017.

Hodges, A., Strand, A.D., Aragaki, A.K., Kuhn, A., Sengstag, T., Hughes, G., Elliston, L.A., Hartog, C., Goldstein, D.R., Thu, D., Hollingsworth, Z.R., Collin, F., Synek, B., Holmans, P.A., Young, A.B., Wexler, N.S., Delorenzi, M., Kooperberg, C., Augood, S.J., Faull, R.L.M., Olson, J.M., Jones, L. and Luthi-Carter, R. 2006. Regional and cellular gene expression changes in human Huntington's disease brain. *Human molecular genetics* 15(6), pp. 965–77. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16467349.

Horton, S., Meredith, A., Richardson, J.A. and Johnson, J.E. 1999. Correct coordination of neuronal differentiation events in ventral forebrain requires the bHLH factor MASH1. *Molecular and cellular neuroscience* 14, pp. 355–69. Available at: http://www.sciencedirect.com/science/article/pii/S1044743199907911.

Houart, C., Caneparo, L., Heisenberg, C., Barth, K., Take-Uchi, M. and Wilson, S. 2002. Establishment of the telencephalon during gastrulation by local antagonism of Wnt signaling. *Neuron* 35(2), pp. 255–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12160744.

Hsieh-Li, H.M., Witte, D.P., Szucsik, J.C., Weinstein, M., Li, H. and Potter, S.S. 1995. Gsh-2, a murine homeobox gene expressed in the developing brain. *Mechanisms of development* 50(2-3), pp. 177–86. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7619729.

Hu, W. 2011. Parkinson's disease is a TH17 dominant autoimmune disorder against accumulated alpha-synuclein. *Nature Precedings*. Available at: http://precedings.nature.com/documents/6176/version/1.

Huang, P., He, Z., Ji, S., Sun, H., Xiang, D., Liu, C., Hu, Y., Wang, X. and Hui, L. 2011. Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* 475(7356), pp. 386–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21562492.

Huelsken, J. and Behrens, J. 2002. The Wnt signalling pathway. *Journal of Cell Science* 115(21), pp. 3977–3978. Available at: http://jcs.biologists.org/cgi/doi/10.1242/jcs.00089.

Jeon, I., Choi, C., Lee, N., Im, W., Kim, M., Oh, S.-H., Park, I.-H., Kim, H.S. and Song, J. 2014. In Vivo Roles of a Patient-Derived Induced Pluripotent Stem Cell Line (HD72-iPSC) in the YAC128 Model of Huntington's Disease. *International journal of stem cells* 7(1), pp. 43–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4049731&tool=pmcentrez&rendertype=abstract.

Jeon, I., Lee, N., Li, J.-Y., Park, I.-H., Park, K.S., Moon, J., Shim, S.H., Choi, C., Chang, D.-J., Kwon, J., Oh, S.-H., Shin, D.A., Kim, H.S., Do, J.T., Lee, D.R., Kim, M., Kang, K.-S., Daley, G.Q., Brundin, P. and Song, J. 2012. Neuronal properties, in vivo effects, and pathology of a Huntington's disease patient-derived induced pluripotent stem cells. *Stem cells (Dayton, Ohio)* 30(9), pp. 2054–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22628015.

Jessell, T.M. 2000. Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nature reviews. Genetics* 1(1), pp. 20–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11262869.

Johnson, R., Zuccato, C., Belyaev, N.D., Guest, D.J., Cattaneo, E. and Buckley, N.J. 2008. A microRNA-based gene dysregulation pathway in Huntington's disease. *Neurobiology of disease* 29(3), pp. 438–45. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18082412.

Johri, A. and Beal, M.F. 2012. Antioxidants in huntington's disease. *Biochim Biophys Acta* 29(6), pp. 997–1003.

Jones, E.G., Coulter, J.D., Burton, H. and Porter, R. 1977. Cells of origin and terminal distribution of corticostriatal fibers arising in the sensory-motor cortex of monkeys. *The Journal of comparative neurology* 173(1), pp. 53–80.

Jones, K.S. and Connor, B. 2012. Intrinsic regulation of adult subventricular zone neural progenitor cells and the effect of brain injury. *American journal of stem cells* 1(1), pp. 48–58. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3643385/.

Joseph, R. 2011. *Neuroscience, Neuropsychology, Neuropsychiatry, Brain &Mind: Introduction, Primer & Overview.* University Press.

Jung, A.B. and Bennett, J.P. 1996. Development of striatal dopaminergic function. I. Pre- and postnatal development of mRNAs and binding sites for striatal D1 (D1a) and D2 (D2a) receptors. *Developmental Brain Research* 94(2), pp. 109–120.

Kallur, T., Darsalia, V., Lindvall, O. and Kokaia, Z. 2006. Human fetal cortical and striatal neural stem cells generate region-specific neurons in vitro and differentiate extensively to neurons after intrastriatal transplantation in. *Journal of neuroscience Research* 84(4), pp. 1630–1644. Available at: http://onlinelibrary.wiley.com/doi/10.1002/jnr.21066/full.

Kelly, C.M., Dunnett, S.B. and Rosser, A.E. 2009. Medium spiny neurons for transplantation in Huntington's disease. *Biochemical Society transactions* 37(Pt 1), pp. 323–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19143656.

Kiecker, C. and Niehrs, C. 2001. A morphogen gradient of Wnt/beta-catenin signalling regulates anteroposterior neural patterning in Xenopus. *Development (Cambridge, England)* 128(21), pp. 4189–201. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11684656.

Kim, D., Kim, C.-H., Moon, J.-I., Chung, Y.-G., Chang, M.-Y., Han, B.-S., Ko, S., Yang, E., Cha, K.Y., Lanza, R. and Kim, K.-S. 2009. Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. *Cell stem cell* 4(6), pp. 472–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2705327&tool=pmcentrez&rendertype=abstract.

Kim, J., Su, S.C., Wang, H., Cheng, A.W., Cassady, J.P., Lodato, M.A., Lengner, C.J., Chung, C.-Y., Dawlaty, M.M., Tsai, L.-H. and Jaenisch, R. 2011. Functional integration of dopaminergic neurons directly converted from mouse fibroblasts. *Cell Stem Cell* 9(5), pp. 413–419. Available at: http://www.sciencedirect.com/science/article/pii/S1934590911004425.

Kim, J.H., Lee, S.-R., Li, L.-H., Park, H.-J., Park, J.-H., Lee, K.Y., Kim, M.-K., Shin, B.A. and Choi, S.-Y. 2011. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PloS one* 6(4), p. e18556. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3084703&tool=pmcentrez&rendertype=abstract.

Kim, S.H., Thomas, C. a, André, V.M., Cummings, D.M., Cepeda, C., Levine, M.S. and Ehrlich, M.E. 2011. Forebrain striatal-specific expression of mutant huntingtin protein in vivo induces cell-autonomous age-dependent alterations in sensitivity to excitotoxicity and mitochondrial function. *ASN Neuro* 3(3), p. e00060.

Kiyama, H., Seto-Ohshima, a and Emson, P.C. 1990. Calbindin D28K as a marker for the degeneration of the striatonigral pathway in Huntington's disease. *Brain research* 525(2), pp. 209–214.

Kohtz, J.D., Baker, D.P., Corte, G. and Fishell, G. 1998. Regionalization within the mammalian telencephalon is mediated by changes in responsiveness to Sonic Hedgehog. *Development (Cambridge, England)* 125(24), pp. 5079–89. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9811591.

Kooy, D. Van Der and Fishell, G. 1987. Neuronal birthdate underlies the development of striatal compartments. *Brain Research* 401, pp. 155–161.

Kozak, M. 1987. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *Journal of Molecular Biology* 196(4), pp. 947–950. Available at: http://www.sciencedirect.com/science/article/pii/0022283687904189.

Krackow, S., Vannoni, E., Codita, a, Mohammed, a H., Cirulli, F., Branchi, I., Alleva, E., Reichelt, a, Willuweit, a, Voikar, V., Colacicco, G., Wolfer, D.P., Buschmann, J.-U.F., Safi, K. and Lipp, H.-P. 2010. Consistent behavioral phenotype differences between inbred mouse strains in the IntelliCage. *Genes, brain, and behavior* 9(7), pp. 722–31. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20528956.

De la Pompa, J.L., Wakeham, a, Correia, K.M., Samper, E., Brown, S., Aguilera, R.J., Nakano, T., Honjo, T., Mak, T.W., Rossant, J. and Conlon, R. a 1997. Conservation of the Notch signalling pathway in mammalian neurogenesis. *Development (Cambridge, England)* 124(6), pp. 1139–48. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9102301.

Landles, C. and Bates, G.P. 2004. Huntingtin and the molecular pathogenesis of Huntington's disease. Fourth in molecular medicine review series. *EMBO reports* 5(10), pp. 958–63. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1299150&tool=pmcentrez&rendertype=abstract.

Lau, S., Rylander Ottosson, D., Jakobsson, J. and Parmar, M. 2014. Direct Neural Conversion from Human Fibroblasts Using Self-Regulating and Nonintegrating Viral Vectors. *Cell reports* 9(5), pp. 1673–1680. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25482564.

Lee, G., Papapetrou, E.P., Kim, H., Chambers, S.M., Tomishima, M.J., Fasano, C. a, Ganat, Y.M., Menon, J., Shimizu, F., Viale, A., Tabar, V., Sadelain, M. and Studer, L. 2009. Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. *Nature* 461(7262), pp. 402–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2784695&tool=pmcentrez&rendertype=abstract.

Letinic, K., Zoncu, R. and Rakic, P. 2002. Origin of GABAergic neurons in the human neocortex. *Nature* 417(6889), pp. 645–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12050665.

Lewis, J. 1996. Neurogenic genes and vertebrate neurogenesis. *Current opinion in neurobiology* 6(1), pp. 3–10. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8794055.

Li, S., Cheng, A.L., Li, H. and Li, X. 1999. Cellular Defects and Altered Gene Expression in PC12 Cells Stably Expressing Mutant Huntingtin. 19(13), pp. 5159–5172.

Liao, W., Tsai, H. and Wang, H. 2008. Modular patterning of structure and function of the striatum by retinoid receptor signaling. *Proceedings of the …* 105(18), pp. 6765–70. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2373312&tool=pmcentrez&rendertype=abstract.

Liao, W.-L., Wang, H.-F., Tsai, H.-C., Chambon, P., Wagner, M., Kakizuka, A. and Liu, F.-C. 2005. Retinoid signaling competence and RARbeta-mediated gene regulation in the developing mammalian telencephalon. *Developmental dynamics : an official publication of the American Association of Anatomists* 232(4), pp. 887–900. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15736225.

Lin, L., Yuan, J., Sander, B. and Golas, M.M. 2015. In Vitro Differentiation of Human Neural Progenitor Cells Into Striatal GABAergic Neurons. *Stem cells translational medicine* 4(7), pp. 775–88. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25972145.

Lindsell, C.E., Boulter, J., DiSibio, G., Gossler, A. and Weinmaster, G. 1996. Expression patterns of Jagged, Delta1, Notch1, Notch2, and Notch3 genes identity ligand-receptor pairs that may function in neural develpment. *Molecular and cellular* *neuroscience* 8, pp. 14–27. Available at: http://www.sciencedirect.com/science/article/pii/S1044743196900408.

Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S., Downes, M., Yu, R., Stewart, R., Ren, B., Thomson, J.A., Evans, R.M. and Ecker, J.R. 2011. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* 471, pp. 68–73.

Lobo, M.K., Karsten, S.L., Gray, M., Geschwind, D.H. and Yang, X.W. 2006. FACS-array profiling of striatal projection neuron subtypes in juvenile and adult mouse brains. *Nature neuroscience* 9(3), pp. 443–52. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16491081.

Long, J.E., Cobos, I., Potter, G.B. and Rubenstein, J.L.R. 2009a. Dlx1&2 and Mash1 transcription factors control MGE and CGE patterning and differentiation through parallel and overlapping pathways. *Cerebral cortex (New York, N.Y. : 1991)* 19 Suppl 1(July), pp. i96–106. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2693539&tool=pmcentrez&rendertype=abstract.

Long, J.E., Swan, C., Liang, W.S., Cobos, I., Potter, G.B. and Rubenstein, J.L.R. 2009b. Dlx1&2 and Mash1 transcription factors control striatal patterning and differentiation through parallel and overlapping pathways. *The Journal of comparative neurology* 512(4), pp. 556–72. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2761428&tool=pmcentrez&rendertype=abstract.

Lujan, E., Chanda, S., Ahlenius, H., Südhof, T.C. and Wernig, M. 2012. Direct conversion of mouse fibroblasts to self-renewing, tripotent neural precursor cells. *Proceedings of the National Academy of Sciences of the United States of America* 109(7), pp. 2527–32. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3289376&tool=pmcentrez&rendertype=abstract.

Lumsden, a, Sprawson, N. and Graham, a 1991. Segmental origin and migration of neural crest cells in the hindbrain region of the chick embryo. *Development (Cambridge, England)* 113(4), pp. 1281–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1811942.

Lumsden, A. and Keynes, R. 1989. Segmental patterns of neuronal development in the chick hindbrain. *Nature* 337, pp. 424–28. Available at: http://europepmc.org/abstract/MED/2644541.

Lunkes, A. and Mandel, J. 1998. A cellular model that recapitulates major pathogenic steps of Huntington ' s disease. 7(9), pp. 1355–1361.

Ma, L., Hu, B., Liu, Y., Vermilyea, S.C., Liu, H., Gao, L., Sun, Y., Zhang, X. and Zhang, S.C. 2012. Human embryonic stem cell-derived GABA neurons correct locomotion deficits in Quinolinin Acid-lesioned mice. *Cell Stem Cell* 10(4), pp. 455–464.

Malone, S., Rietbrock, A. and Scherbaum, F. 1998. The GIANT Analysis System (Graphical Interactive Aftershock Network Toolbox). *Seismological Research Letters* 69(1), pp. 40–45.

Marchand, R. and Lajoie, L. 1986. Histogenesis of the striopallidal system in the rat. Neurogenesis of its neurons. *Neuroscience* 17(3), pp. 573–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3703249.

Marchetto, M.C.N., Carromeu, C., Acab, A., Yu, D., Yeo, G.W., Mu, Y., Chen, G., Gage, F.H. and Muotri, A.R. 2010. A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell* 143(4), pp. 527–39. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3003590&tool=pmcentrez&rendertype=abstract.

Maretto, S., Cordenonsi, M., Dupont, S., Braghetta, P., Broccoli, V., Hassan, a B., Volpin, D., Bressan, G.M. and Piccolo, S. 2003. Mapping Wnt/beta-catenin signaling during mouse development and in colorectal tumors. *Proceedings of the National Academy of Sciences of the United States of America* 100(6), pp. 3299–304. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=152286&tool=pmcentrez&rendertype=abstract.

Marin, O., Anderson, S. a and Rubenstein, J.L. 2000. Origin and molecular specification of striatal interneurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 20(16), pp. 6063–76. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10934256.

Marro, S., Pang, Z., Yang, N., Tsai, M., Qu, K., Chang, H.Y., Sudhof, T.C. and Werning, M. 2011. Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. *Cell stem cell* 9(4), pp. 374–382. Available at: http://www.sciencedirect.com/science/article/pii/S1934590911004334.

Martin, G.R. 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the United States of America* 78, pp. 7634–7638.

Martín-Ibáñez, R., Crespo, E., Esgleas, M., Urban, N., Wang, B., Waclaw, R., Georgopoulos, K., Martínez, S., Campbell, K., Vicario-Abejón, C., Alberch, J., Chan, S., Kastner, P., Rubenstein, J.L. and Canals, J.M. 2012. Function in Developing Striatal Matrix Neurons. *Stem Cells and Development* 21(12), pp. 2239–2251.

Martindale, D., Hackam, A., Wieczorek, A., Ellerby, L., Wellington, C., McCutcheon, K., Singaraja, R., Esfarjani, P.K., Devon, R., Kim, S.U., Bredese, D.E., Tufaro, F. and Hayden, M.R. 1998. Lenght of huntingtin and its polyglutamine tract influences localization and frequency of intracellular aggregates. *Nature Genetics* 18, pp. 150–154.

Martinez, Y., Béna, F., Gimelli, S., Tirefort, D., Dubois-Dauphin, M., Krause, K.H. and Preynat-Seauve, O. 2012. Cellular diversity within embryonic stem cells: Pluripotent clonal sublines show distinct differentiation potential. *Journal of Cellular and Molecular Medicine* 16(3), pp. 456–467.

Martynoga, B., Morrison, H., Price, D.J. and Mason, J.O. 2005. Foxg1 is required for specification of ventral telencephalon and region-specific regulation of dorsal telencephalic precursor proliferation and apoptosis. *Developmental biology* 283(1), pp. 113–27. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15893304.

Maurisse, R., De Semir, D., Emamekhoo, H., Bedayat, B., Abdolmohammadi, A., Parsi, H. and Gruenert, D.C. 2010. Comparative transfection of DNA into primary and transformed mammalian cells from different lineages. *BMC biotechnology* 10(1), p. 9. Available at: http://www.biomedcentral.com/1472-6750/10/9.

McGuinness, T., Porteus, M., Smiga, S., Bulfone, A., Kingsley, C., Qiu, M., Liu, J.K., Long, J.E., Xu, D. and Rubenstein, J.L.R. 1996. Sequence, Organization, and Transcription of the Dlx-1 and Dlx-2 Locus. *Genomics* 35(3), pp. 473–485. Available at: http://www.sciencedirect.com/science/article/pii/S0888754396903870.

Méndez-Gómez, H.R. and Vicario-Abejón, C. 2012. The homeobox gene Gsx2 regulates the self-renewal and differentiation of neural stem cells and the cell fate of postnatal progenitors. *PloS one* 7(1), p. e29799. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3252334&tool=pmcentrez&rendertype=abstract.

Moore, K.L., Persuad, T.V.N. and Torchia, M.G. 2011. *The Developing Human: Clinically Oriented Embryology*. 9th ed. philadelphia: Elsevier Health Sciences.

Morizane, A., Doi, D., Kikuchi, T., Nishimura, K. and Takahashi, J. 2011. Small-molecule inhibitors of bone morphogenic protein and activin/nodal signals promote highly efficient neural induction from human pluripotent stem cells. *Journal of neuroscience research* 89(2), pp. 117–26. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21162120.

Moyer, J.T., Wolf, J. a and Finkel, L.H. 2007. Effects of dopaminergic modulation on the integrative properties of the ventral striatal medium spiny neuron. *Journal of neurophysiology* 98(6), pp. 3731–3748.

Nat, R. and Hovatta, O. 2004. In vitro neural differentiation of human embryonic stem cells. *Journal of cellular and molecular medicine* 8(4), pp. 570–571.

Nawara, M., Szczaluba, K., Poirier, K., Chrzanowska, K., Pilch, J., Bal, J., Chelly, J. and Mazurczak, T. 2006. The ARX mutations: a frequent cause of X-linked mental retardation. *American journal of medical genetics. Part A* 140, pp. 727–732.

Niclis, J., Trounson, A., Dottori, M., Ellisdon, A., Bottomley, S., Verlinsky, Y. and Cram, D. 2009. Human embryonic stem cell models of Huntington disease. *Reproductive BioMedicine Online* 19(1), pp. 106–113. Available at: http://www.sciencedirect.com/science/article/pii/S1472648310600533.

Nicola, S.M., Surmeier, D.J. and Malenka, R.C. 2000. Dopaminergic modulation of neural excitability in the striatum and nucleus accumbens. *Annual review of nuroscience* 23, pp. 185–215. Available at: http://www.annualreviews.org/doi/abs/10.1146/annurev.neuro.23.1.185.

Nicoleau, C., Varela, C., Bonnefond, C., Maury, Y., Bugi, A., Aubry, L., Viegas, P., Bourgois-Rocha, F., Peschanski, M. and Perrier, A.L. 2013. Embryonic stem cells neural differentiation qualifies the role of Wnt/β-Catenin signals in human telencephalic specification and regionalization. *Stem cells (Dayton, Ohio)* 31(9), pp. 1763–74. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23818270.

Nishi, a, Bibb, J. a, Snyder, G.L., Higashi, H., Nairn, a C. and Greengard, P. 2000. Amplification of dopaminergic signaling by a positive feedback loop. *Proceedings of the National Academy of Sciences of the United States of America* 97(23), pp. 12840–12845.

Nishi, a, Snyder, G.L. and Greengard, P. 1997. Bidirectional regulation of DARPP-32 phosphorylation by dopamine. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 17(21), pp. 8147–8155.

O'Malley, J., Woltjen, K. and Kaji, K. 2009. New strategies to generate induced pluripotent stem cells. *Current opinion in biotechnology* 20(5), pp. 516–21. Available at: http://www.sciencedirect.com/science/article/pii/S0958166909001074.

Ohkubo, Y., Chiang, C. and Rubenstein, J.L.R. 2002. Coordinate regulation and synergistic actions of BMP4, SHH and FGF8 in the rostral prosencephalon regulate morphogenesis of the telencephalic and optic vesicles. *Neuroscience* 111(1), pp. 1–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11955708.

Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T. and Yamanaka, S. 2008. Generation of mouse induced pluripotent stem cells without viral vectors. *Science (New York, N.Y.)* 322(5903), pp. 949–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18845712.

Olsson, M., Björklund, a and Campbell, K. 1998. Early specification of striatal projection neurons and interneuronal subtypes in the lateral and medial ganglionic eminence. *Neuroscience* 84(3), pp. 867–76. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9579790.

Ouimet, C.C. and Greengard, P. 1990. Distribution of DARPP-32 in the basal ganglia: an electron microscopic study. *Journal of Neurocytology* 19(1), pp. 39–52.

Ouimet, C.C., Miller, P.E., Hemmings Jr., H.C., JR., Walaas, S.I. and Greengard, P. 1984. Darpp-32, a dopamine-and adenosine 3': 5'-monophosphate-regulated phosphoprotein enriched in dopamine-innervated brain regions. III. Immunocytochemical localization. *The journal of neuroscience : the official journal of the Society for Neuroscience* 4, pp. 111–24. Available at: http://www.jneurosci.org/content/4/1/111.short.

Paek, H., Antoine, M.W., Diaz, F. and Hébert, J.M. 2012. Increased β-catenin activity in the anterior neural plate induces ectopic mid-hindbrain characteristics. *Developmental dynamics : an official publication of the American Association of Anatomists* 241(2), pp. 242–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3266450&tool=pmcentrez&rendertype=abstract.

Pan, Z.Z. 2012. Transcriptional control of Gad2. *Transcription* 3, pp. 68–72.

Pang, Z.P., Yang, N., Vierbuchen, T., Ostermeier, A., Fuentes, D.R., Yang, T.Q., Ami, C., Vittorio, S., Marro, S., Sudhof, T.C. and Werning, M. 2012. Induction of human neuronal cells by defined transcription. *Nature* 476(7359), pp. 220–223.

Panganiban, G. and Rubenstein, J.L.R. 2002. Developmental functions of the Distal-less/Dlx homeobox genes. *Development (Cambridge, England)* 129(19), pp. 4371–86. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12223397.

Park, B.K., Sperber, S.M., Choudhury, A., Ghanem, N., Hatch, G.T., Sharpe, P.T., Thomas, B.L. and Ekker, M. 2004. Intergenic enhancers with distinct activities regulate Dlx gene expression in the mesenchyme of the branchial arches. *Developmental biology* 268(2), pp. 532–45. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15063187.

Park, I.-H., Arora, N., Huo, H., Maherali, N., Ahfeldt, T., Shimamura, A., Lensch, M.W., Cowan, C., Hochedlinger, K. and Daley, G.Q. 2008. Disease-specific induced pluripotent stem cells. *Cell* 134(5), pp. 877–86. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2633781&tool=pmcentrez&rendertype=abstract.

Parras, C.M., Galli, R., Britz, O., Soares, S., Galichet, C., Battiste, J., Johnson, J.E., Nakafuku, M., Vescovi, A. and Guillemot, F. 2004. Mash1 specifies neurons and oligodendrocytes in the postnatal brain. *The EMBO journal* 23(22), pp. 4495–505. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=526464&tool=pmcentrez&rendertype=abstract.

Pauly, M.-C., Döbrössy, M.D., Nikkhah, G., Winkler, C. and Piroth, T. 2013. Organization of the human fetal subpallium. *Frontiers in neuroanatomy* 7(January), p. 54. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3893616&tool=pmcentrez&rendertype=abstract.

Pei, Z., Wang, B., Chen, G., Nagao, M., Nakafuku, M. and Campbell, K. 2011. Homeobox genes Gsx1 and Gsx2 differentially regulate telencephalic progenitor maturation. *Proceedings of the National Academy of Sciences of the United States of America* 108(4), pp. 1675–80. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3029701&tool=pmcentrez&rendertype=abstract.

Penrod, R.D., Kourrich, S., Kearney, E., Thomas, M.J. and Lanier, L.M. 2011. An embryonic culture system for the investigation of striatal medium spiny neuron dendritic spine development and plasticity. *Journal of neuroscience methods* 200(1), pp. 1–13. Available at: http://www.sciencedirect.com/science/article/pii/S0165027011003347.

Petryniak, M. a, Potter, G.B., Rowitch, D.H. and Rubenstein, J.L.R. 2007. Dlx1 and Dlx2 control neuronal versus oligodendroglial cell fate acquisition in the developing forebrain. *Neuron* 55(3), pp. 417–33. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2039927&tool=pmcentrez&rendertype=abstract.

Pfisterer, U., Kirkeby, A., Torper, O., Wood, J., Nelander, J., Dufour, A., Björklund, A., Lindvall, O., Jakobsson, J. and Parmar, M. 2011. Direct conversion of human fibroblasts to dopaminergic neurons. *Proceedings of the National Academy of Sciences of the United States of America* 108(25), pp. 10343–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3121829&tool=pmcentrez&rendertype=abstract.

Pickel, V.M. and Heras, A. 1996. Ultrastructural localization of calbindin-d and GABA in the matrix compartmnet of teh rat caudate-putamen nuclei. *Neuroscience* 71, pp. 167–178.

Pinal, C.S. and Tobin, A.J. 1998. Uniqueness and redundancy in GABA production. *Perspectives on developmental neurobiology* 5(2-3), pp. 109–18. Available at: http://europepmc.org/abstract/med/9777629.

Poitras, L., Ghanem, N., Hatch, G. and Ekker, M. 2007. The proneural determinant MASH1 regulates forebrain Dlx1/2 expression through the I12b intergenic enhancer. *Development (Cambridge, England)* 134(9), pp. 1755–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17409112.

Porteus, M.H., Bulfone, A., Liu, J.K., Puelles, L., Lo, L.C. and Rubenstein, J.L. 1994. DLX-2, MASH-1, and MAP-2 expression and bromodeoxyuridine incorporation define molecularly distinct cell populations in the embryonic mouse forebrain. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 14(11 Pt 1), pp. 6370–83. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7965042.

Qiu, M., Bulfone, a, Martinez, S., Meneses, J.J., Shimamura, K., Pedersen, R. a and Rubenstein, J.L. 1995. Null mutation of Dlx-2 results in abnormal morphogenesis of proximal first and second branchial arch derivatives and abnormal differentiation in the forebrain. *Genes & Development* 9(20), pp. 2523–2538. Available at: http://www.genesdev.org/cgi/doi/10.1101/gad.9.20.2523.

Quinn, J.C., Molinek, M., Martynoga, B.S., Zaki, P. a, Faedo, A., Bulfone, A., Hevner, R.F., West, J.D. and Price, D.J. 2007. Pax6 controls cerebral cortical cell number by regulating exit from the cell cycle and specifies cortical cell identity by a cell autonomous mechanism. *Developmental biology* 302(1), pp. 50–65. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2384163&tool=pmcentrez&rendertype=abstract.

Radcliffe, P. and Mitrophanous, K. 2004. Multiple gene products from a single vector: 'self-cleaving' 2A peptides. *Gene Therapy* 11(23), pp. 1673–74. Available at: http://www.nature.com/doifinder/10.1038/sj.gt.3302361.

Rallu, M., Machold, R., Gaiano, N., Corbin, J.G., McMahon, A.P. and Fishell, G. 2002. Dorsoventral patterning is established in the telencephalon of mutants lacking both Gli3 and Hedgehog signaling. *Development (Cambridge, England)* 129(21), pp. 4963–74. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12397105.

Ranen, N.G., Stine, O.C., Abbott, M.H., Sherr, M., Codori, a M., Franz, M.L., Chao, N.I., Chung, a S., Pleasant, N. and Callahan, C. 1995. Anticipation and instability of IT-15 (CAG)n repeats in parent-offspring pairs with Huntington disease. *American journal of human genetics* 57(3), pp. 593–602.

Regad, T., Roth, M., Bredenkamp, N., Illing, N. and Papalopulu, N. 2007. The neural progenitor-specifying activity of FoxG1 is antagonistically regulated by CKI and FGF. *Nature cell biology* 9(5), pp. 531–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17435750.

Rigamonti, D., Bauer, J.H., De-Fraja, C., Conti, L., Sipione, S., Sciorati, C., Clementi, E., Hackam, A., Hayden, M.R., Li, Y., Cooper, J.K., Ross, C.A., Govoni, S., Vincenz, C. and Cattaneo, E. 2000. Wild-Type Huntingtin Protects from Apoptosis Upstream of Caspase-3. *The Journal of Neuroscince* 20(10), pp. 3705–3713. Available at: http://www.jneurosci.org/content/20/10/3705.long.

Ring, K., Tong, L., Balestra, M., Javier, R., Zwilling, Y.A., Li, G., Walker, D., Zhang1, W.R., Kreitzer, A.C., Huang, Y. and Huang, and Y. 2012. Direct reprogramming of mouse and human fibroblasts into multipotent neural stem cells with a single factor. *Cell stem cell* 11(1), pp. 100–109. Available at: http://www.sciencedirect.com/science/article/pii/S1934590912002895.

Robey, E. 1997. Notch in vertebrates. *Current Opinion in Genetics & Development* 7(4), pp. 551–557. Available at: http://www.sciencedirect.com/science/article/pii/S0959437X97800858.

Rosecrans, A., JR, P. and JA., J. 2009. Conditioned Place Preference. In: Buccafusco, J. ed. *Methods of Behavior Analysis in Neuroscience*. 2nd ed. CRC Press.

Ross, C. a. and Tabrizi, S.J. 2011. Huntington's disease: From molecular pathogenesis to clinical treatment. *The Lancet Neurology* 10(1), pp. 83–98. Available at: http://dx.doi.org/10.1016/S1474-4422(10)70245-3.

Ross, S.E., Greenberg, M.E. and Stiles, C.D. 2003. Basic helix-loop-helix factors in cortical development. *Neuron* 39(1), pp. 13–25. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12848929.

Rouchka, E.C., Phatak, A.W. and Singh, A. V 2008. Effect of single nucleotide polymorphisms on Affymetrix ® match-mismatch probe pairs. *Bioinformation* 2(9), pp. 405–11. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2533060/.

Royce, G.J. 1982. Laminar origin of cortical neurons which project upon the caudate nucleus: a horseradish peroxidase investigation in the cat. *The Journal of comparative neurology* 205(1), pp. 8–29.

Rubenstein, J.L. and Puelles, L. 1994. Homeobox gene expression during development of the vertebrate brain. *Current topics in developmental biology* 29, pp. 1–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7828435.

Rubenstein, J.L.R., Shimamura, K., Martinez, S. and Puelles, L. 1998. Regionalization of the prosencephalic neural plate. *Annual review of nuroscience* 21, pp. 445–77.

Ryan, M.D., King, a M. and Thomas, G.P. 1991. Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *The Journal of general virology* 72, pp. 2727–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1658199.

Saga, Y., Kobayashi, M., Ohta, H., Murai, N., Nakai, N., Oshima, M. and Taketo, M.M. 1999. Impaired extrapyramidal function caused by the targeted disruption of retinoid X receptor RXRgamma1 isoform. *Genes to cells : devoted to molecular & cellular mechanisms* 4(4), pp. 219–28. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10336693.

Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual.*

Sari, Y. 2011. Huntington's Disease: From Mutant Huntingtin Protein to Neurotrophic Factor Therapy. *International Journal of Biomedical Science* 7(2), pp. 89–100. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=21841917.

Schambra, U.B., Duncan, G.E., Breese, G.R., Fornaretto, M.G., Caron, M.G. and Fremeau, R.T. 1994. Ontogeny of D1A and D2 dopamine receptor subtypes in rat brain using in situ hybridization and receptor binding. *Neuroscience* 62(1), pp. 65–85.

Schneider, R. a, Hu, D., Rubenstein, J.L., Maden, M. and Helms, J. a 2001. Local retinoid signaling coordinates forebrain and facial morphogenesis by maintaining FGF8 and SHH. *Development (Cambridge, England)* 128(14), pp. 2755–67. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11526081.

Schuurmans, C. and Guillemot, F. 2002. Molecular mechanisms underlying cell fate specification in the developing telencephalon. *Current opinion in neurobiology* 12(1), pp. 26–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11861161.

Shimamura, K. and Rubenstein, J.L. 1997. Inductive interactions direct early regionalization of the mouse forebrain. *Development (Cambridge, England)* 124(14), pp. 2709–18. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9226442.

Sillivan, S.E. and Konradi, C. 2011. Expression and function of dopamine receptors in the developing medial frontal cortex and striatum of the rat. *Neuroscience* 199, pp. 501–514. Available at: http://dx.doi.org/10.1016/j.neuroscience.2011.10.004.

Simeone, A., Acampora, D., Panneset, M., Esposito, M.D., Stornaiuolo, A., Gulisanott, M., Mallamacit, A., Kastury, K., Druck, T., Huebner, K.A.Y. and Boncinelli, E. 1994. Cloning and characterization of two members of the vertebrate Dlx gene family. *Developmental Biology* 91, pp. 2250–2254.

Sipione, S. and Cattaneo, E. 2001. Modeling Huntington's disease in cells, flies, and mice. *Molecular neurobiology* 23(1), pp. 21–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11642542.

Smith, J.R., Vallier, L., Lupo, G., Alexander, M., Harris, W. a and Pedersen, R. a 2008. Inhibition of Activin/Nodal signaling promotes specification of human embryonic stem cells into neuroectoderm. *Developmental biology* 313(1), pp. 107–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18022151.

Soldati, C., Bithell, A., Johnston, C., Wong, K.-Y., Stanton, L.W. and Buckley, N.J. 2013. Dysregulation of REST-regulated coding and non-coding RNAs in a cellular model of Huntington's disease. *Journal of neurochemistry* 124(3), pp. 418–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23145961.

Soldner, F. and Jaenisch, R. 2012. iPSC disease modeling. *Science* 338, pp. 1155–1156.

Sommer, C. a, Stadtfeld, M., Murphy, G.J., Hochedlinger, K., Kotton, D.N. and Mostoslavsky, G. 2009. Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem cells (Dayton, Ohio)* 27(3), pp. 543–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19096035.

Son, E.Y., Ichida, J.K., Wainger, B.J., Toma, J.S., Rafuse, V.F., Woolf, C.J. and Eggan, K. 2011. Conversion of mouse and human fibroblasts into functional spinal motor neurons. *Cell stem cell* 9(3), pp. 205–18. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3188987&tool=pmcentrez&rendertype=abstract.

Song, J., Lee, S.-T., Kang, W., Park, J.-E., Chu, K., Lee, S., Hwang, T., Chung, H. and Kim, M. 2007. Human embryonic stem cell-derived neural precursor transplants attenuate apomorphine-induced rotational behavior in rats with unilateral quinolinic acid lesions. *Neuroscience letters* 423(1), pp. 58–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17669593.

Staahl, B.T., Tang, J., Wu, W., Sun, A., Gitler, A.D., Yoo, A.S. and Crabtree, G.R. 2013. Kinetic analysis of npBAF to nBAF switching reveals exchange of SS18 with CREST and integration with neural developmental pathways. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33(25), pp. 10348–

61. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3685834&tool=pmcentrez&rendertype=abstract.

Stamataki, D., Ulloa, F., Tsoni, S. V, Mynett, A. and Briscoe, J. 2005. A gradient of Gli activity mediates graded Sonic Hedgehog signaling in the neural tube. *Genes & development* 19(5), pp. 626–41. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=551582&tool=pmcentrez&rendertype=abstract.

Stock, D.W., Ellies, D.L., Zhao, Z., Ekker, M., Ruddle, F.H. and Weiss, K.M. 1996. The evolution of the vertebrate Dlx gene family. *Proceedings of the National Academy of Sciences of the United States of America* 93(20), pp. 10858–63. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=38247&tool=pmcentrez&rendertype=abstract.

Stoof, J. and Kebabian, J. 1981. Opposing roles for D-1 and D-2 dopamine receptors in efflux of cyclic AMP from rat neostriatum. *Nature* 294(26), pp. 366–368. Available at: http://www.nature.com/nature/journal/v294/n5839/abs/294366a0.html.

Storm, E.E., Garel, S., Borello, U., Hebert, J.M., Martinez, S., McConnell, S.K., Martin, G.R. and Rubenstein, J.L.R. 2006. Dose-dependent functions of Fgf8 in regulating telencephalic patterning centers. *Development (Cambridge, England)* 133(9), pp. 1831–44. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16613831.

Stoykova, a, Treichel, D., Hallonet, M. and Gruss, P. 2000. Pax6 modulates the dorsoventral patterning of the mammalian telencephalon. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 20(21), pp. 8042–50. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11050125.

Stoykova, A. and Gruss, P. 1994. Roles of Pax-Genes Expression Patterns in Developing and Adult Brain as Suggested by. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 14(3), pp. 1395–1412.

Stühmer, T., Anderson, S.A., Ekker, M. and Rubenstein, J.L.R. 2002. Ectopic expression of the Dlx genes induces glutamic acid decarboxylase and Dlx expression. *Development (Cambridge, England)* 129, pp. 245–252.

Stühmer, T., Puelles, L., Ekker, M. and Rubenstein, J.L.R. 2002. Expression from a Dlx gene enhancer marks adult mouse cortical GABAergic neurons. *Cerebral cortex (New York, N.Y. : 1991)* 12(1), pp. 75–85. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11734534.

Sturrock, R. 1980. A developmental study of the mouse neostriatum. *Journal of anatomy* 130(2), pp. 243–261. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1233130/.

Suh, Y., Obernier, K., Hölzl-Wenig, G., Mandl, C., Herrmann, A., Wörner, K., Eckstein, V. and Ciccolini, F. 2009. Interaction between DLX2 and EGFR regulates proliferation and neurogenesis of SVZ precursors. *Molecular and cellular neurosciences* 42(4), pp. 308–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19683576.

Surmacz, B., Fox, H., Gutteridge, A., Fish, P., Lubitz, S. and Whiting, P. 2012. Directing differentiation of human embryonic stem cells toward anterior neural ectoderm using small molecules. *Stem cells (Dayton, Ohio)* 30(9), pp. 1875–84. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22761025.

Sussel, L., Marin, O., Kimura, S. and Rubenstein, J.L. 1999. Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: evidence for a transformation of the pallidum into the striatum. *Development (Cambridge, England)* 126(15), pp. 3359–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10393115.

Svenningsson, P., Nishi, A., Fisone, G., Girault, J.-A., Nairn, A.C. and Greengard, P. 2004. DARPP-32: an integrator of neurotransmission. *Annual review of pharmacology and toxicology* 44, pp. 269–296.

Szabo, E., Rampalli, S., Risueño, R.M., Schnerch, A., Mitchell, R., Fiebig-Comyn, A., Levadoux-Martin, M. and Bhatia, M. 2010. Direct conversion of human fibroblasts to multilineage blood progenitors. *Nature* 468(7323), pp. 521–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21057492.

Szucsik, J.C., Witte, D.P., Li, H., Pixley, S.K., Small, K.M. and Potter, S.S. 1997. Altered forebrain and hindbrain development in mice mutant for the Gsh-2 homeobox gene. *Developmental biology* 191(2), pp. 230–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9398437.

Szymczak, A.L., Workman, C.J., Wang, Y., Vignali, K.M., Dilioglou, S., Vanin, E.F. and Vignali, D. a a 2004. Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nature biotechnology* 22(5), pp. 589–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15064769.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131(5), pp. 861–72. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18035408.

Takahashi, K. and Yamanaka, S. 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* 126(4), pp. 663–676.

Tamamaki, N., Fujimori, K.E. and Takauji, R. 1997. Origin and route of tangentially migrating neurons in the developing neocortical intermediate zone. *The Journal of neuroscience* 17(21), pp. 8313–23. Available at: http://www.jneurosci.org/content/17/21/8313.short.

Tamamaki, N., Sugimoto, Y., Tanaka, K. and Takauji, R. 1999. Cell migration from the ganglionic eminence to the neocortex investigated by labeling nuclei with UV irradiation via a fiber-optic cable. *Neuroscience research* 35(3), pp. 241–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10605947.

Tamura, S., Morikawa, Y., Iwanishi, H., Hisaoka, T. and Senba, E. 2004. Foxp1 gene expression in projection neurons of the mouse striatum. *Neuroscience* 124(2), pp. 261–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14980377.

Tang, B., Becanovic, K., Desplats, P. a., Spencer, B., Hill, A.M., Connolly, C., Masliah, E., Leavitt, B.R. and Thomas, E. a. 2012. Forkhead box protein p1 is a transcriptional repressor of immune signaling in the CNS: Implications for transcriptional dysregulation in huntington disease. *Human Molecular Genetics* 21(14), pp. 3097–3111.

Tao, W. and Lai, E. 1992. Telencephalon-restricted expression of BF-1, a new member of the HNF-3/fork head gene family, in the developing rat brain. *Neuron* 8(5), pp. 957–66. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1350202.

Tapscott, S.J., Davis, R.L., Thayer, M.J., Cheng, P.F., Weintraub, H. and Lassar, A.B. 1988. MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. *Science (New York, N.Y.)* 242(4877), pp. 405–11. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3175662.

Tauber, E., Miller-Fleming, L., Mason, R.P., Kwan, W., Clapp, J., Butler, N.J., Outeiro, T.F., Muchowski, P.J. and Giorgini, F. 2011. Functional gene expression profiling in yeast implicates translational dysfunction in mutant huntingtin toxicity. *The Journal of biological chemistry* 286(1), pp. 410–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3012999&tool=pmcentrez&rendertype=abstract.

Telezhkin, V., Brazier, S.P., Cayzac, S.H., Wilkinson, W.J., Riccardi, D. and Kemp, P.J. 2010. Mechanism of inhibition by hydrogen sulfide of native and recombinant BKCa channels. *Respiratory physiology & neurobiology* 172(3), pp. 169–78. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20576528.

Thier, M., Wörsdörfer, P., Lakes, Y.B., Gorris, R., Herms, S., Opitz, T., Seiferling, D., Quandel, T., Hoffmann, P., Nöthen, M.M., Brüstle, O. and Edenhofer, F. 2012. Direct conversion of fibroblasts into stably expandable neural stem cells. *Cell stem cell* 10(4), pp. 473–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22445518.

Thomas, E. a., Coppola, G., Tang, B., Kuhn, A., Kim, S., Geschwind, D.H., Brown, T.B., Luthi-Carter, R. and Ehrlich, M.E. 2011. In vivo cell-autonomous transcriptional abnormalities revealed in mice expressing mutant huntingtin in striatal but not cortical neurons. *Human Molecular Genetics* 20(6), pp. 1049–1060.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S. and Jones, J. effery M. 1998. Embryonic stem cell lines derived from human blastocysts. *Science* 282, pp. 1145–1147. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.282.5391.1145.

Toresson, H. and Campbell, K. 2001. A role for Gsh1 in the developing striatum and olfactory bulb of Gsh2 mutant mice. *Development (Cambridge, England)* 128(23), pp. 4769–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11731457.

Toresson, H., Potter, S.S. and Campbell, K. 2000. Genetic control of dorsal-ventral identity in the telencephalon: opposing roles for Pax6 and Gsh2. *Development (Cambridge, England)* 127, pp. 4361–4371.

Trushina, E., Dyer, R.B., Ii, J.D.B., Eide, L., Tran, D.D., Vrieze, B.T., Mcpherson, P.S., Bhaskar, S., Houten, B. Van, Zeitlin, S., Aebersold, R., Hayden, M., Joseph, E., Seeberg, E., Dragatsis, I., Doyle, K., Chacko, C., Mcmurray, C.T., Ure, D., Legendre-guillemin, V., Mandavilli, B.S., Mcniven, M., Parisi, J.E. and Bender, A. 2004. Mutant Huntingtin Impairs Axonal Trafficking in Mammalian Neurons In Vivo and In Vitro Mutant Huntingtin Impairs Axonal Trafficking in Mammalian Neurons In Vivo and In Vitro. *Molecular and Cellular Biology* 24(18), pp. 8195–8209.

Valjent, E., Bertran-Gonzalez, J., Hervé, D., Fisone, G. and Girault, J.-A. 2009. Looking BAC at striatal signaling: cell-specific analysis in new transgenic mice. *Trends in neurosciences* 32(10), pp. 538–47. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19765834.

Victor, M.B., Richner, M., Hermanstyne, T.O., Ransdell, J.L., Sobieski, C., Deng, P.-Y., Klyachko, V.A., Nerbonne, J.M. and Yoo, A.S. 2014. Generation of Human Striatal

Neurons by MicroRNA-Dependent Direct Conversion of Fibroblasts. *Neuron* 84(2), pp. 311–323. Available at: http://www.cell.com/article/S0896627314009143/fulltext .

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Südhof, T.C. and Wernig, M. 2010. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* 463(7284), pp. 1035–41. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2829121&tool=pmcentrez&rendertype=abstract.

Waclaw, R.R., Allen, Z.J., Bell, S.M., Erdélyi, F., Szabó, G., Potter, S.S. and Campbell, K. 2006. The zinc finger transcription factor Sp8 regulates the generation and diversity of olfactory bulb interneurons. *Neuron* 49(4), pp. 503–16. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16476661.

Waclaw, R.R., Wang, B. and Campbell, K. 2004. The homeobox gene Gsh2 is required for retinoid production in the embryonic mouse telencephalon. *Development (Cambridge, England)* 131(16), pp. 4013–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15269172.

Waclaw, R.R., Wang, B., Pei, Z., Ehrman, L. a and Campbell, K. 2009. Distinct temporal requirements for the homeobox gene Gsx2 in specifying striatal and olfactory bulb neuronal fates. *Neuron* 63(4), pp. 451–65. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2772064&tool=pmcentrez&rendertype=abstract.

Walker, F.O. 2007. Huntington's disease. *Lancet* 369(9557), pp. 218–28. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17240289.

Wang, B., Long, J.E., Flandin, P., Pla, R., Waclaw, R.R., Campbell, K. and Rubenstein, J.L.R. 2013. Loss of Gsx1 and Gsx2 function rescues distinct phenotypes in Dlx1/2 mutants. *The Journal of comparative neurology* 521(7), pp. 1561–84. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3615175&tool=pmcentrez&rendertype=abstract.

Wang, B., Waclaw, R.R., Allen, Z.J., Guillemot, F. and Campbell, K. 2009. Ascl1 is a required downstream effector of Gsx gene function in the embryonic mouse telencephalon. *Neural development* 4(February), p. 5. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2644683&tool=pmcentrez&rendertype=abstract.

Wang, Y. and Steinbeisser, H. 2009. Molecular basis of morphogenesis during vertebrate gastrulation. *Cellular and molecular life sciences : CMLS* 66(14), pp. 2263–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19347571.

Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.-H., Li, H., Lau, F., Ebina, W., Mandal, P.K., Smith, Z.D., Meissner, A., Daley, G.Q., Brack, A.S., Collins, J.J., Cowan, C., Schlaeger, T.M. and Rossi, D.J. 2010. Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell stem cell* 7(5), pp. 618–30. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3656821&tool=pmcentrez&rendertype=abstract.

Warren, N., Caric, D., Pratt, T., Clausen, J. a, Asavaritikrai, P., Mason, J.O., Hill, R.E. and Price, D.J. 1999. The transcription factor, Pax6, is required for cell proliferation and differentiation in the developing cerebral cortex. *Cerebral cortex (New York, N.Y.: 1991)* 9(6), pp. 627–35. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10498281.

Watanabe, K., Kamiya, D., Nishiyama, A., Katayama, T., Nozaki, S., Kawasaki, H., Watanabe, Y., Mizuseki, K. and Sasai, Y. 2005. Directed differentiation of telencephalic precursors from embryonic stem cells. *Nature neuroscience* 8(3), pp. 288–96. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15696161.

Watanabe, K., Ueno, M., Kamiya, D., Nishiyama, A., Matsumura, M., Wataya, T., Takahashi, J.B., Nishikawa, S., Nishikawa, S., Muguruma, K. and Sasai, Y. 2007. A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nature biotechnology* 25(6), pp. 681–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17529971.

Wichterle, H., Garcia-Verdugo, J.M., Herrera, D.G. and Alvarez-Buylla, A. 1999. Young neurons from medial ganglionic eminence disperse in adult and embryonic brain. *Nature neuroscience* 2(5), pp. 461–66. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10321251.

Wictorin, K. 1992. Anatomy and connectivity of intrastriatal striatal transplants. *Progress in Neurobiology* 38(6), pp. 611–639. Available at: http://www.sciencedirect.com/science/article/pii/030100829290044F.

Wilson, C.J. 1987. Morphology and synaptic connections of crossed corticostriatal neurons in the rat. *The Journal of comparative neurology* 263(4), pp. 567–580.

Wilson, S. and Houart, C. 2004. Early steps in the development of the forebrain. *Developmental cell* 6(2), pp. 167–181. Available at: http://www.sciencedirect.com/science/article/pii/S1534580704000279.

Wilson, S. and Rubenstein, J. 2000. Induction and dorsoventral patterning of the telencephalon. *Neuron* 28, pp. 641–651. Available at: http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.elsevier-1c12714d-7083-36a5-9e0d-582445783fff.

Woltjen, K., Michael, I.P., Mohseni, P., Desai, R., Mileikovsky, M., Hämäläinen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M., Kaji, K., Sung, H.-K. and Nagy, A. 2009. piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* 458(7239), pp. 766–70. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3758996&tool=pmcentrez&rendertype=abstract.

Wu, H. and Sun, Y.E. 2006. Epigenetic regulation of stem cell differentiation. *Pediatric research* 59(4 Pt 2), p. 21R–5R. Available at: http://dx.doi.org/10.1203/01.pdr.0000203565.76028.2a.

Wu, J.I., Lessard, J. and Crabtree, G.R. 2009. Understanding the words of chromatin regulation. *Cell* 136(2), pp. 200–206.

Xuan, S., Baptista, C. a, Balas, G., Tao, W., Soares, V.C. and Lai, E. 1995. Winged helix transcription factor BF-1 is essential for the development of the cerebral hemispheres. *Neuron* 14(6), pp. 1141–52. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7605629.

Yang, Y., Jiao, J., Gao, R., Yao, H., Sun, X.-F. and Gao, S. 2013. Direct conversion of adipocyte progenitors into functional neurons. *Cellular reprogramming* 15(6), pp. 484–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3848438&tool=pmcentrez&rendertype=abstract.

Yoo, A.S., Staahl, B.T., Chen, L. and Crabtree, G.R. 2009. MicroRNA-mediated switching of chromatin-remodelling complexes in neural development. *Nature* 460(7255), pp. 642–646.

Yoo, A.S., Sun, A.X., Li, L., Shcheglovitov, A., Portmann, T., Li, Y., Lee-messer, C., Dolmetsch, R.E., Tsien, R.W. and Crabtree, G.R. 2011. MicroRNA-mediated conversion of human fibroblasts to neurons. *Nature* 476, pp. 228–231.

Yoshihara, S., Omichi, K., Yanazawa, M., Kitamura, K. and Yoshihara, Y. 2005. Arx homeobox gene is essential for development of mouse olfactory system. *Development (Cambridge, England)* 132(4), pp. 751–762.

Yun, K., Fischman, S., Johnson, J., Hrabe de Angelis, M., Weinmaster, G. and Rubenstein, J.L.R. 2002. Modulation of the notch signaling by Mash1 and Dlx1/2 regulates sequential specification and differentiation of progenitor cell types in the subcortical telencephalon. *Development (Cambridge, England)* 129(21), pp. 5029–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12397111.

Yun, K., Garel, S., Fischman, S. and Rubenstein, J.L.R. 2003. Patterning of the lateral ganglionic eminence by the Gsh1 and Gsh2 homeobox genes regulates striatal and olfactory bulb histogenesis and the growth of axons through the basal ganglia. *The Journal of comparative neurology* 461(2), pp. 151–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12724834.

Yun, K., Potter, S. and Rubenstein, J.L. 2001. Gsh2 and Pax6 play complementary roles in dorsoventral patterning of the mammalian telencephalon. *Development (Cambridge, England)* 128, pp. 193–205.

Zeron, M.M., Hansson, O., Chen, N., Wellington, C.L., Leavitt, B.R., Brundin, P., Hayden, M.R. and Raymond, L. a. 2002. Increased sensitivity to N-methyl-D-aspartate receptor-mediated excitotoxicity in a mouse model of Huntington's disease. *Neuron* 33(6), pp. 849–860.

Zerucha, T., Stühmer, T., Hatch, G., Park, B.K., Long, Q., Yu, G., Gambarotta, A., Schultz, J.R., Rubenstein, J.L. and Ekker, M. 2000. A highly conserved enhancer in the Dlx5/Dlx6 intergenic region is the site of cross-regulatory interactions between Dlx genes in the embryonic forebrain. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 20, pp. 709–721.

Zhang, N., An, M.C., Montoro, D. and Ellerby, L.M. 2010. Characterization of Human Huntington ' s Disease Cell Model from Induced Pluripotent Stem Cells. *pLos Currents* 2, pp. 1–11. Available at: http://currents.plos.org/hd/article/characterization-of-human-huntingtons-disease-cell-model-from-induced-pluripotent-stem-cells-2/.

Zhang, Y., Pak, C., Han, Y., Ahlenius, H., Zhang, Z., Chanda, S., Marro, S., Patzke, C., Acuna, C., Covy, J., Xu, W., Yang, N., Danko, T., Chen, L., Wernig, M. and Sudhof, T.C. 2013. Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* 78(5), pp. 785–798. Available at: http://www.sciencedirect.com/science/article/pii/S0896627313004492.

Zheng, Z. and Diamond, M.I. 2012. Huntington disease and the huntingtin protein. *Progress in molecular biology and translational science* 107, pp. 189–214. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22482451.

Zuccato, C., Belyaev, N., Conforti, P., Ooi, L., Tartari, M., Papadimou, E., MacDonald, M., Fossale, E., Zeitlin, S., Buckley, N. and Cattaneo, E. 2007. Widespread disruption of repressor element-1 silencing transcription factor/neuron-restrictive silencer factor occupancy at its target genes in Huntington's disease. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27(26), pp. 6972–83. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17596446.

Zuccato, C., Ciammola, A., Rigamonti, D., Leavitt, B.R., Goffredo, D., Conti, L., MacDonald, M.E., Friedlander, R.M., Silani, V., Hayden, M.R., Timmusk, T., Sipione, S. and Cattaneo, E. 2001. Loss of huntingtin-mediated BDNF gene transcription in Huntington's disease. *Science (New York, N.Y.)* 293(5529), pp. 493–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11408619.

Zuccato, C., Tartari, M., Crotti, A., Goffredo, D., Valenza, M., Conti, L., Cataudella, T., Leavitt, B.R., Hayden, M.R., Timmusk, T., Rigamonti, D. and Cattaneo, E. 2003. Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. *Nature genetics* 35(1), pp. 76–83.

Zuccato, C., Valenza, M. and Cattaneo, E. 2010. Molecular Mechanisms and Potential Therapeutical Targets in Huntington ' s Disease. *Physiol Rev* 90(3), pp. 905–981.

Appendix

2.1 Table shows the dysregulated genes between Ctip2$^{-/-}$ homozygous, Ctip2$^{-/+}$ heterozygous and wild-type:

Probe set ID	p-value	Regulation	FCAbsolute	Fold change	Log Fold change	Gene symbol
1422860_at	8.11E-05	up	5.52421	5.52421	2.465768	Nts
1425699_a_at	9.05E-05	up	3.433145	3.433145	1.7795308	Abhd14a
1448293_at	9.13E-05	up	2.6417077	2.6417077	1.4014708	Ebf1
1417605_s_at	9.20E-05	up	5.183407	5.183407	2.3739007	Camk1
1424092_at	9.41E-05	up	10.151953	10.151953	3.3436854	Epb4.1
1435311_s_at	9.49E-05	up	2.457243	2.457243	1.2970405	
1455469_at	9.60E-05	up	5.0760446	5.0760446	2.3437047	Slc6a7
1438228_at	1.03E-04	up	2.405829	2.405829	1.2665341	4930452B06Rik
1456512_at	1.04E-04	up	3.8267944	3.8267944	1.9361364	Pdzrn4
1421679_a_at	1.04E-04	up	18.726233	18.726233	4.226989	Cdkn1a

1460617_s_at	1.08E-04	up	1.6370419	1.6370419	0.7110913	Rab6b
1456005_a_at	1.09E-04	up	7.3031363	7.3031363	2.8685162	Bcl2l11
1416361_a_at	1.10E-04	up	2.9430172	2.9430172	1.5572959	Dync1i1
1428213_at	1.13E-04	up	1.5718087	1.5718087	0.65242565	Nsmce4a
1455064_at	1.20E-04	up	1.879241	1.879241	0.91015005	Rab36
1440209_at	1.27E-04	up	5.260762	5.260762	2.3952718	Mar-01
1454656_at	1.28E-04	up	1.7252355	1.7252355	0.7867933	Spata13
1427122_at	1.30E-04	up	4.2831583	4.2831583	2.098675	Copg2as2
1416561_at	1.31E-04	up	2.265817	2.265817	1.1800313	Gad1
1430526_a_at	1.38E-04	up	1.7694229	1.7694229	0.8232789	Smarca2
1442379_at	1.39E-04	up	3.3677416	3.3677416	1.7517815	EG574403
1423495_at	1.43E-04	up	2.2837672	2.2837672	1.1914155	Decr2
1435486_at	1.48E-04	up	5.3985577	5.3985577	2.432574	Pak3

1421990_at	1.65E-04	up	1.8222594	1.8222594	0.8657284	Syt1
1428136_at	1.73E-04	up	2.47847	2.47847	1.3094498	Sfrp1
1438454_at	1.75E-04	up	3.063769	3.063769	1.6153076	B430203M17Rik
1426929_at	1.76E-04	up	2.3975563	2.3975563	1.2615647	Brunol4
1441598_at	1.85E-04	up	1.6369637	1.6369637	0.7110223	Tmeff2
1459861_s_at	1.85E-04	up	1.7437768	1.7437768	0.8022154	Kdm2b
1456137_at	1.89E-04	up	3.8626266	3.8626266	1.9495822	Nrxn3
1416840_at	2.01E-04	up	2.1153219	2.1153219	1.0808772	Mid1ip1
1435404_at	2.03E-04	up	2.8292687	2.8292687	1.5004292	Disp2
1454900_s_at	2.10E-04	up	1.3680533	1.3680533	0.45212445	Mycbp2
1458842_at	2.22E-04	up	3.2285945	3.2285945	1.6909063	Odz1
1421118_a_at	2.31E-04	up	2.002576	2.002576	1.001857	Gpr56
1442873_at	2.32E-04	up	25.567024	25.567024	4.6762123	

1455298_at	2.34E-04	up	3.936704	3.936704	1.9769882	
1454672_at	2.40E-04	up	6.487682	6.487682	2.6977031	
1450051_at	2.43E-04	up	1.5632838	1.5632838	0.6445797	Atrx
1417924_at	2.45E-04	up	6.8353887	6.8353887	2.7730234	Pak3
1447139_at	2.51E-04	up	2.7812095	2.7812095	1.4757124	Bcl7c
1430668_a_at	2.52E-04	up	1.5340186	1.5340186	0.617316	Ankra2
1448933_at	2.60E-04	up	2.1550417	2.1550417	1.1077157	Pcdhb17
1457446_at	2.60E-04	up	2.7305598	2.7305598	1.4491967	Opcml
1457072_at	2.62E-04	up	2.5279438	2.5279438	1.3379644	Bcl11a
1438957_x_at	2.78E-04	up	3.1672223	3.1672223	1.6632181	Cds2
1451458_at	2.83E-04	up	2.4060595	2.4060595	1.2666723	Tmem2
1416896_at	2.84E-04	up	4.676809	4.676809	2.2255244	Rps6ka1
1428157_at	2.85E-04	up	2.0293076	2.0293076	1.0209875	Gng2

1458499_at	1.25E-04	down	2.3589375	-2.3589375	-1.2381371	Pde10a
1425553_s_at	1.26E-04	down	1.9368467	-1.9368467	-0.9537098	Hip1r
1431228_s_at	1.31E-04	down	1.367688	-1.367688	-0.4517391	4930526I15Rik
1435227_at	1.51E-04	down	2.678508	-2.678508	-1.4214296	Bcl11b
1455800_x_at	1.72E-04	down	1.5760868	-1.5760868	-0.6563469	Samm50
1426454_at	1.76E-04	down	6.7189307	-6.7189307	-2.7482316	Arhgdib
1423803_s_at	1.97E-04	down	1.5766994	-1.5766994	-0.6569076	Gltscr2
1419389_at	2.02E-04	down	3.8294756	-3.8294756	-1.9371468	Pde10a
1416185_a_at	2.14E-04	down	1.4288716	-1.4288716	-0.5148763	Adh5
1450339_a_at	2.17E-04	down	3.5837295	-3.5837295	-1.8414618	Bcl11b
1416512_at	2.21E-04	down	1.47332	-1.47332	-0.5590708	Nubp2
1418277_at	2.37E-04	down	1.5374179	-1.5374179	-0.6205093	rp9
1416146_at	2.43E-04	down	1.3863947	-1.3863947	-0.4713381	Hspa4

1425050_at	2.49E-04	down	1.4895205	-1.4895205	-0.574848	Isoc1
1435553_at	2.56E-04	down	2.325122	-2.325122	-1.2173065	Pdzd2
1428214_at	2.60E-04	down	1.9322833	-1.9322833	-0.9503066	Tomm7
1448978_at	2.61E-04	down	3.3793082	-3.3793082	-1.7567279	Ngef
1428507_at	2.67E-04	down	1.4174502	-1.4174502	-0.50329804	Hdhd2
1423802_at	2.71E-04	down	1.6305734	-1.6305734	-0.70537937	Camkv
1416350_at	2.86E-04	down	1.7930807	-1.7930807	-0.8424404	Klf16
1460506_s_at	2.87E-04	down	1.4123412	-1.4123412	-0.4980887	Ndufc2

2.2 Table of the target genes that found to play fundamental roles in the development of forebrain and striatal medium spiny neuron fate:

Gene Name	Process
achaete-scute complex homolog 1 (Drosophila)	Neurogenesis.
adhesion molecule with Ig like domain 1	
basic leucine zipper and W2 domains 2	
calcium/calmodulin-dependent protein kinase I	
cell adhesion molecule with homology to L1CAM	
dedicator of cytokinesis 7	
disabled homolog 1 (Drosophila)	
Doublecortin	
doublecortin-like kinase 1	
drebrin 1	
ELAV (embryonic lethal, abnormal vision, Drosophila)-like 3 (Hu antigen C)	
enabled homolog (Drosophila)	
ephrin B2	
ephrin B3	
G-protein signalling modulator 1 (AGS3-like, C. elegans)	
Gene Name	
growth arrest specific 7	
hairy and enhancer of split 5 (Drosophila)	
kinesin family member 2A	
L1 cell adhesion molecule	
LIM homeobox protein 6	
myelin transcription factor 1-like	
myocyte enhancer factor 2C	
myocyte enhancer factor 2D	
neurogenic differentiation 1; neurogenic differentiation 5	
neurogenic differentiation 2	
neuron navigator 1	
neuronal guanine nucleotide exchange factor	
neuropilin 1	
orthopedia homolog (Drosophila)	
PHD finger protein 10	
platelet-activating factor acetylhydrolase, isoform 1b, subunit 1	
roundabout homolog 1 (Drosophila)	
RUN and FYVE domain containing 3	
slit homolog 2 (Drosophila)	
spastin	
SRY-box containing gene 11	
achaete-scute complex homolog 1 (Drosophila)	Activator transcription factor.
activating transcription factor 2	
activating transcription factor 5	
activating transcription factor 7 interacting protein	
AF4/FMR2 family, member 3	
ataxin 7-like 3	
bromodomain and WD repeat domain containing 1	
BTB and CNC homology 2	
calmodulin binding transcription activator 1	
cAMP responsive element binding protein 3-like 1	
Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 4	
chromodomain helicase DNA binding protein 8	
churchill domain containing 1	
circadian locomoter output cycles kaput	
cone-rod homeobox containing gene	
CREB binding protein	

cysteine-serine-rich nuclear protein 3	
D site albumin promoter binding protein	
dachshund 1 (Drosophila)	
DEAH (Asp-Glu-Ala-His) box polypeptide 9	
delta/notch-like EGF-related receptor	
DNA methyltransferase (cytosine-5) 1	
E2F transcription factor 1	
E74-like factor 2	
early B-cell factor 1	
endothelial PAS domain protein 1	
enhancer of yellow 2 homolog (Drosophila)	
ets variant gene 1	
eyes absent 2 homolog (Drosophila)	
family with sequence similarity 120, member B	
forkhead box D3	
forkhead box F2	
forkhead box H1	
forkhead box K1	
Friend leukemia integration 1	
heterogeneous nuclear ribonucleoprotein D-like	
IKAROS family zinc finger 2	
interferon activated gene 204	
interferon regulatory factor 3	
interleukin enhancer binding factor 2	
Kruppel-like factor 1 (erythroid)	
Kruppel-like factor 11	
Kruppel-like factor 17	
Kruppel-like factor 3 (basic)	
Kruppel-like factor 5	
Kruppel-like factor 6	
Kruppel-like factor 7 (ubiquitous)	
LIM domain containing preferred translocation partner in lipoma	
LIM homeobox protein 3	
mediator complex subunit 1	
mediator complex subunit 13	
mediator complex subunit 13-like	
mediator complex subunit 15	
mediator of RNA polymerase II transcription, subunit 11 homolog (S. cerevisiae)	
mediator of RNA polymerase II transcription, subunit 12 homolog (yeast)-like	
mediator of RNA polymerase II transcription, subunit 25 homolog (yeast)	
mediator of RNA polymerase II transcription, subunit 28 homolog (yeast)	
mediator of RNA polymerase II transcription, subunit 8 homolog (yeast)	
mediator of RNA polymerase II transcription, subunit 9 homolog (yeast)	
Meis homeobox 1	
membrane-bound transcription factor peptidase, site 2	
microphthalmia-associated transcription factor	
MYB binding protein (P160) 1a	
myeloblastosis oncogene-like 1	
myocyte enhancer factor 2C	
myocyte enhancer factor 2D	
MYST histone acetyltransferase monocytic leukemia 4	
N-acetyltransferase 14	
nascent polypeptide-associated complex alpha polypeptide	
neurogenic differentiation 1; neurogenic differentiation 5	
nuclear factor I/A	

nuclear factor I/B	
nuclear factor of activated T-cells 5	
nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2	
nuclear protein in the AT region	
nuclear receptor subfamily 2, group E, member 1	
nuclear receptor subfamily 5, group A, member 1	
nuclear respiratory factor 1	
nuclear transcription factor-Y alpha	
one cut domain, family member 2	
paraspeckle protein 1	
peroxisome proliferator activated receptor gamma	
POU domain, class 2, transcription factor 1	
POU domain, class 3, transcription factor 2	
pre B-cell leukemia transcription factor 1	
pre B-cell leukemia transcription factor 2	
RAR-related orphan receptor alpha	
ras responsive element binding protein 1	
recombination signal binding protein for immunoglobulin kappa J region	
regulatory factor X, 4 (influences HLA class II expression)	
RNA binding motif protein 39	
SET domain containing 1B	
Sp7 transcription factor 7	
Spi-B transcription factor (Spi-1/PU.1 related)	
splicing factor proline/glutamine rich	
SRY-box containing gene 19; SRY-box containing gene 4	
SRY-box containing gene 2	
SRY-box containing gene 6	
steroid receptor RNA activator 1	
sterol regulatory element binding transcription factor 1	
SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1	
SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2	
thyrotroph embryonic factor	
trans-acting transcription factor 1	
trans-acting transcription factor 3	
transcription factor 4	
transducin (beta)-like 1 X-linked	
transformation related protein 53 binding protein 1	
transformation/transcription domain-associated protein	
ubiquitin specific peptidase 49	
v-maf musculoaponeurotic fibrosarcoma oncogene family, protein B (avian)	
YY1 transcription factor	
zinc finger homeobox 3	
zinc finger protein 143	
zinc finger protein 326	
zinc finger protein 410	
zinc finger protein 521	
zinc finger protein, autosomal	
ZXD family zinc finger C	

AE binding protein 2	Transcriptional repressor.
B-cell CLL/lymphoma 11A (zinc finger protein)	
B-cell leukemia/lymphoma 11B	
BTB and CNC homology 2	
C-terminal binding protein 2	
CUG triplet repeat, RNA binding protein 2	
DNA methyltransferase (cytosine-5) 1	

DNA methyltransferase 3A	
E4F transcription factor 1	
Ets2 repressor factor	
GATA-like 1	
Kruppel-like factor 11	
Kruppel-like factor 12	
Kruppel-like factor 17	
Kruppel-like factor 3 (basic); similar to BKLF	
Kv channel interacting protein 3, calsenilin	
MAF1 homolog (S. cerevisiae)	
MYB binding protein (P160) 1a	
MYST histone acetyltransferase monocytic leukemia 4	
Max dimerization protein 3	
Max interacting protein 1	
NF-kappaB repressing factor	
PHD finger protein 21A	
SAP30-like	
SET domain containing (lysine methyltransferase) 8; predicted gene 8590	
Sin3A associated protein	
TGFB-induced factor homeobox 1	
TSC22 domain family, member 1	
WT1-interacting protein	
YY1 transcription factor	
acidic (leucine-rich) nuclear phosphoprotein 32 family, member A	
activating transcription factor 7 interacting protein	
additional sex combs like 1 (Drosophila)	
additional sex combs like 3 (Drosophila)	
amine oxidase, flavin containing 1	
arginine glutamic acid dipeptide (RE) repeats	
chromobox homolog 4 (Drosophila Pc class)	
chromobox homolog 6; neuronal pentraxin receptor; Cbx6-Nptxr readthrough transcripts	
chromodomain helicase DNA binding protein 8	
cofactor of BRCA1	
cold shock domain protein A	
core-binding factor, runt domain, alpha subunit 2, translocated to, 3 (human)	
cryptochrome 1 (photolyase-like)	
dachshund 1 (Drosophila)	
forkhead box D3	
forkhead box K1	
forkhead box P1	
forkhead box P2	
forkhead box P4	
hairy and enhancer of split 5 (Drosophila)	
hairy/enhancer-of-split related with YRPW motif 1	
heterogeneous nuclear ribonucleoprotein A/B	
heterogeneous nuclear ribonucleoprotein D-like	
histone deacetylase 11	
histone deacetylase 8	
homeobox, msh-like 1	
hypothetical protein LOC100046298; akirin 2	
interferon activated gene 204	
lysine (K)-specific demethylase 2B	
mediator complex subunit 13	
mediator complex subunit 13-like	
mediator of RNA polymerase II transcription, subunit 12 homolog (yeast)-like	
membrane-bound transcription factor peptidase, site 2; similar to zinc finger, X-linked, duplicated B; Yy2 transcription factor	

mesoderm induction early response 1, family member 2	
methyl CpG binding protein 2	
myelin basic protein expression factor 2, repressor	
nuclear receptor co-repressor 1	
nuclear receptor co-repressor 2	
nuclear receptor coactivator 5	
nuclear receptor interacting protein 1	
nuclear receptor subfamily 1, group D, member 2; predicted gene 5827	
nuclear receptor subfamily 2, group E, member 1	
nuclear receptor subfamily 5, group A, member 1	
paired box gene 4	
polycomb group ring finger 3	
predicted gene 13886; TAR DNA binding protein	
predicted gene 14457; predicted gene 14503; zinc finger protein 161; predicted gene 14509	
prohibitin 2	
proliferation-associated 2G4; predicted gene 5297	
ras responsive element binding protein 1	
recombination signal binding protein for immunoglobulin kappa J region	
retinoblastoma 1	
retinoblastoma binding protein 7; predicted gene 6382	
ring finger protein 2	
scaffold attachment factor B2	
similar to MBD2 (methyl-CpG-binding protein)-interacting zinc finger protein; similar to MBD2-interacting zinc finger; histone H4 transcription factor	
similar to NFkB interacting protein 1; protein phosphatase 1, regulatory (inhibitor) subunit 13 like	
similar to c-Maf long form; avian musculoaponeurotic fibrosarcoma (v-maf) AS42 oncogene homolog	
similar to transcriptional regulator protein; SAP30 binding protein	
sin3 associated polypeptide	
special AT-rich sequence binding protein 1	
splicing factor proline/glutamine rich (polypyrimidine tract binding protein associated); similar to PTB-associated splicing factor	
suppression of tumorigenicity 18	
suppressor of variegation 4-20 homolog 2 (Drosophila)	
trans-acting transcription factor 1	
transcription elongation regulator 1 (CA150)	
transcriptional regulator, SIN3B (yeast)	
transducin-like enhancer of split 1, homolog of Drosophila E(spl)	
transducin-like enhancer of split 3, homolog of Drosophila E(spl)	
tripartite motif-containing 24	
tripartite motif-containing 27	
v-maf musculoaponeurotic fibrosarcoma oncogene family, protein B (avian)	
zinc finger protein 148	
zinc finger protein 318	
zinc finger protein 503	
zinc finger protein 521	
zinc finger protein 639	
zinc finger, MYND domain containing 11	
zinc fingers and homeoboxes 3	

B-cell leukemia/lymphoma 11B	Neuron development.
Eph receptor A4	
Eph receptor A7	
Eph receptor B2	
FIG4 homolog (S. cerevisiae)	
Kruppel-like factor 7 (ubiquitous)	
L1 cell adhesion molecule	
LIM homeobox protein 2	
LIM homeobox protein 3	
LIM homeobox protein 6	
MYC binding protein 2	
POU domain, class 4, transcription factor 3	
RAS protein-specific guanine nucleotide-releasing factor 1	
RAS-related C3 botulinum substrate 1	
RAS-related C3 botulinum substrate 3	
SLIT and NTRK-like family, member 4	
Unc-51 like kinase 2 (C. elegans)	
achaete-scute complex homolog 1 (Drosophila)	
activated leukocyte cell adhesion molecule	
adenosine A2a receptor	
adhesion molecule with Ig like domain 1	
amyloid beta (A4) precursor protein-binding, family B, member 2	
ankyrin 3, epithelial	
autophagy-related 7 (yeast); similar to AGP7	
biregional cell adhesion molecule-related/down-regulated by oncogenes (Cdon) binding protein	
cadherin 23 (otocadherin)	
cadherin 4	
calcium channel, voltage-dependent, P/Q type, alpha 1A subunit	
cell adhesion molecule with homology to L1CAM	
chemokine (C-X-C motif) ligand 12	
crumbs homolog 1 (Drosophila)	
deleted in colorectal carcinoma	
dihydropyrimidinase-like 5	
distal-less homeobox 5	
dopamine receptor D1A	
doublecortin	
doublecortin-like kinase 1	
dystonin; hypothetical protein LOC100047109	
enabled homolog (Drosophila)	
ephrin B3	
ets variant gene 1	
glutamate receptor ionotropic, NMDA3A	
growth arrest specific 1	
growth arrest specific 7	
guanine nucleotide binding protein, alpha q polypeptide	
guanine nucleotide binding protein, alpha transducing 1	
huntingtin	
kelch-like 1 (Drosophila)	
kinesin family member 5C	
leukemia inhibitory factor	
leukocyte specific transcript 1	
methyl CpG binding protein 2	
microtubule-associated protein 1B	
microtubule-associated protein 1S	
microtubule-associated protein 2	
myocardial infarction associated transcript (non-protein coding)	
netrin 3	
neurofilament, light polypeptide	

neurogenic differentiation 2	
neuron-glia-CAM-related cell adhesion molecule	
neuropilin 1	
neurotrophin 3	
neurturin	
nuclear receptor subfamily 2, group E, member 1	
numb-like	
one cut domain, family member 2	
paired box gene 6	
phosphatase and tensin homolog	
plexin A3	
predicted gene 8566; superoxide dismutase 1, soluble; similar to Superoxide dismutase	
protein kinase C, iota	
protein kinase, cGMP-dependent, type I	
protein tyrosine phosphatase, receptor type Z, polypeptide 1	
reelin	
reticulon 4 receptor-like 1	
ribosomal protein L24; predicted gene 9385; predicted gene 7380	
roundabout homolog 1 (Drosophila)	
runt related transcription factor 1	
sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A	
similar to Ena-VASP-like; Ena-vasodilator stimulated phosphoprotein	
similar to PBX3a; pre B-cell leukemia transcription factor 3	
similar to RIKEN cDNA 2610109H07 gene; RIKEN cDNA 2610109H07 gene	
similar to clusterin; clusterin	
slit homolog 2 (Drosophila)	
sodium channel and clathrin linker 1	
superoxide dismutase 2, mitochondrial	
taurine upregulated gene 1	
thymus cell antigen 1, theta	
topoisomerase (DNA) II beta	
tyrosine hydroxylase	
v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)	
ventral anterior homeobox containing gene 2	
wingless-related MMTV integration site 3A	

ATP/GTP binding protein 1	Neuron differentiation.
B-cell leukemia/lymphoma 11B	
BCL2-associated athanogene 1	
Eph receptor A4	
Eph receptor A7	
Eph receptor B2	
FIG4 homolog (S. cerevisiae)	
Kruppel-like factor 7 (ubiquitous)	
L1 cell adhesion molecule	
LIM domain binding 1	
LIM homeobox protein 2	
LIM homeobox protein 3	
LIM homeobox protein 6	
MYC binding protein 2	
POU domain, class 3, transcription factor 2	
POU domain, class 3, transcription factor 4	
POU domain, class 4, transcription factor 3	
RAR-related orphan receptor alpha	
RAS protein-specific guanine nucleotide-releasing factor 1	

RAS-related C3 botulinum substrate 1	
RAS-related C3 botulinum substrate 3	
SLIT and NTRK-like family, member 4	
SRY-box containing gene 2	
SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1	
Unc-51 like kinase 2 (C. elegans)	
achaete-scute complex homolog 1 (Drosophila)	
activated leukocyte cell adhesion molecule	
activator of basal transcription 1	
adenosine A2a receptor	
adhesion molecule with Ig like domain 1	
amyloid beta (A4) precursor protein-binding, family B, member 2	
ankyrin 3, epithelial	
autophagy-related 7 (yeast); similar to AGP7	
biregional cell adhesion molecule-related/down-regulated by oncogenes (Cdon) binding protein	
cadherin 23 (otocadherin)	
cadherin 4	
calcium channel, voltage-dependent, P/Q type, alpha 1A subunit	
cell adhesion molecule with homology to L1CAM	
cell division cycle 42 homolog (S. cerevisiae); predicted gene 7407	
ceroid-lipofuscinosis, neuronal 8	
chemokine (C-X-C motif) ligand 12	
crumbs homolog 1 (Drosophila)	
deleted in colorectal carcinoma	
delta-like 1 (Drosophila)	
dihydropyrimidinase-like 5	
distal-less homeobox 1	
distal-less homeobox 2	
distal-less homeobox 5	
dopamine receptor D1A	
doublecortin	
doublecortin-like kinase 1	
dystonin; hypothetical protein LOC100047109	
enabled homolog (Drosophila)	
ephrin B3	
ets variant gene 1	
glutamate receptor ionotropic, NMDA3A	
growth arrest specific 1	
growth arrest specific 7	
guanine nucleotide binding protein, alpha q polypeptide	
guanine nucleotide binding protein, alpha transducing 1	
hairy and enhancer of split 5 (Drosophila)	
huntingtin	
inhibitor of DNA binding 4	
kelch-like 1 (Drosophila)	
kinase non-catalytic C-lobe domain (KIND) containing 1	
kinesin family member 5C	
leukemia inhibitory factor	
leukocyte specific transcript 1	
methyl CpG binding protein 2	
microtubule-associated protein 1B	
microtubule-associated protein 1S	
microtubule-associated protein 2	
myocardial infarction associated transcript (non-protein coding)	
myotrophin	
netrin 3	
neurofilament, light polypeptide	
neurogenic differentiation 2	

neuron-glia-CAM-related cell adhesion molecule	
neuropilin 1	
neurotrophin 3	
neurotrophin 5	
neurturin	
nuclear receptor subfamily 2, group E, member 1	
numb-like	
one cut domain, family member 2	
orthopedia homolog (Drosophila)	
paired box gene 6	
phosphatase and tensin homolog	
plexin A3	
predicted gene 8566; superoxide dismutase 1, soluble; similar to Superoxide dismutase	
proprotein convertase subtilisin/kexin type 9	
protein kinase C, iota	
protein kinase, cGMP-dependent, type I	
protein tyrosine phosphatase, receptor type Z, polypeptide 1	
recombination signal binding protein for immunoglobulin kappa J region	
reelin	
reticulon 4 receptor-like 1	
ribosomal protein L24; predicted gene 9385; predicted gene 7380	
roundabout homolog 1 (Drosophila)	
runt related transcription factor 1	
sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A	
similar to Ena-VASP-like; Ena-vasodilator stimulated phosphoprotein	
similar to PBX3a; pre B-cell leukemia transcription factor 3	
similar to RIKEN cDNA 2610109H07 gene; RIKEN cDNA 2610109H07 gene	
similar to Stat3B; signal transducer and activator of transcription 3	
similar to clusterin; clusterin	
slit homolog 2 (Drosophila)	
sodium channel and clathrin linker 1	
superoxide dismutase 2, mitochondrial	
taurine upregulated gene 1	
thymus cell antigen 1, theta	
topoisomerase (DNA) II beta	
transforming growth factor, beta receptor I	
tubby-like protein 3	
tyrosine hydroxylase	
v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)	
ventral anterior homeobox containing gene 2	
wingless-related MMTV integration site 3A	

LIM homeobox protein 6	Generation of neuron in forebrain.
POU domain, class 3, transcription factor 4	
SRY-box containing gene 2	
asp (abnormal spindle)-like, microcephaly associated (Drosophila)	
autophagy-related 7 (yeast); similar to AGP7	
distal-less homeobox 1	
distal-less homeobox 2	
guanine nucleotide binding protein, alpha q polypeptide	
orthopedia homolog (Drosophila)	
paired box gene 6	
plexin A3	
wingless-related MMTV integration site 3A	

LIM homeobox protein 6	Forebrain in neuron differentiation.
POU domain, class 3, transcription factor 4	
SRY-box containing gene 2	
autophagy-related 7 (yeast); similar to AGP7	
distal-less homeobox 1	
distal-less homeobox 2	
guanine nucleotide binding protein, alpha q polypeptide	
orthopedia homolog (Drosophila)	
paired box gene 6	
plexin A3	
E2F transcription factor 1	Forebrain development.
LIM homeobox protein 2	
LIM homeobox protein 3	
LIM homeobox protein 6	
N-deacetylase/N-sulfotransferase (heparan glucosaminyl) 1	
POU domain, class 3, transcription factor 2	
POU domain, class 3, transcription factor 4	
RAS-related C3 botulinum substrate 1	
SRY-box containing gene 2	
SRY-box containing gene 3	
achaete-scute complex homolog 1 (Drosophila)	
aldehyde dehydrogenase family 1, subfamily A3	
alpha thalassemia/mental retardation syndrome X-linked homolog (human)	
amyloid beta (A4) precursor-like protein 1	
apoptotic peptidase activating factor 1	
asp (abnormal spindle)-like, microcephaly associated (Drosophila)	
autophagy-related 7 (yeast); similar to AGP7	
bone morphogenetic protein receptor, type 1A	
centrosomal protein 120	
chordin	
deleted in liver cancer 1	
disabled homolog 1 (Drosophila)	
distal-less homeobox 1	
distal-less homeobox 2	
dopamine receptor D1A	
doublecortin-like kinase 1	
fibroblast growth factor receptor substrate 2	
forkhead box P2	
guanine nucleotide binding protein, alpha q polypeptide	
homeobox, msh-like 1	
huntingtin	
inhibitor of DNA binding 4	
neurofibromatosis 1	
nuclear factor I/B	
nuclear receptor co-repressor 1	
nuclear receptor co-repressor 2	
nuclear receptor subfamily 2, group E, member 1	
numb-like	
orthopedia homolog (Drosophila)	
paired box gene 6	
platelet-activating factor acetylhydrolase, isoform 1b, subunit 1	
plexin A3	
protein kinase, cGMP-dependent, type I	
recombination signal binding protein for immunoglobulin kappa J region	
reelin	
regulatory factor X, 4 (influences HLA class II expression)	
similar to RIKEN cDNA 2610109H07 gene; RIKEN cDNA 2610109H07 gene	
sine oculis-related homeobox 3 homolog (Drosophila)	

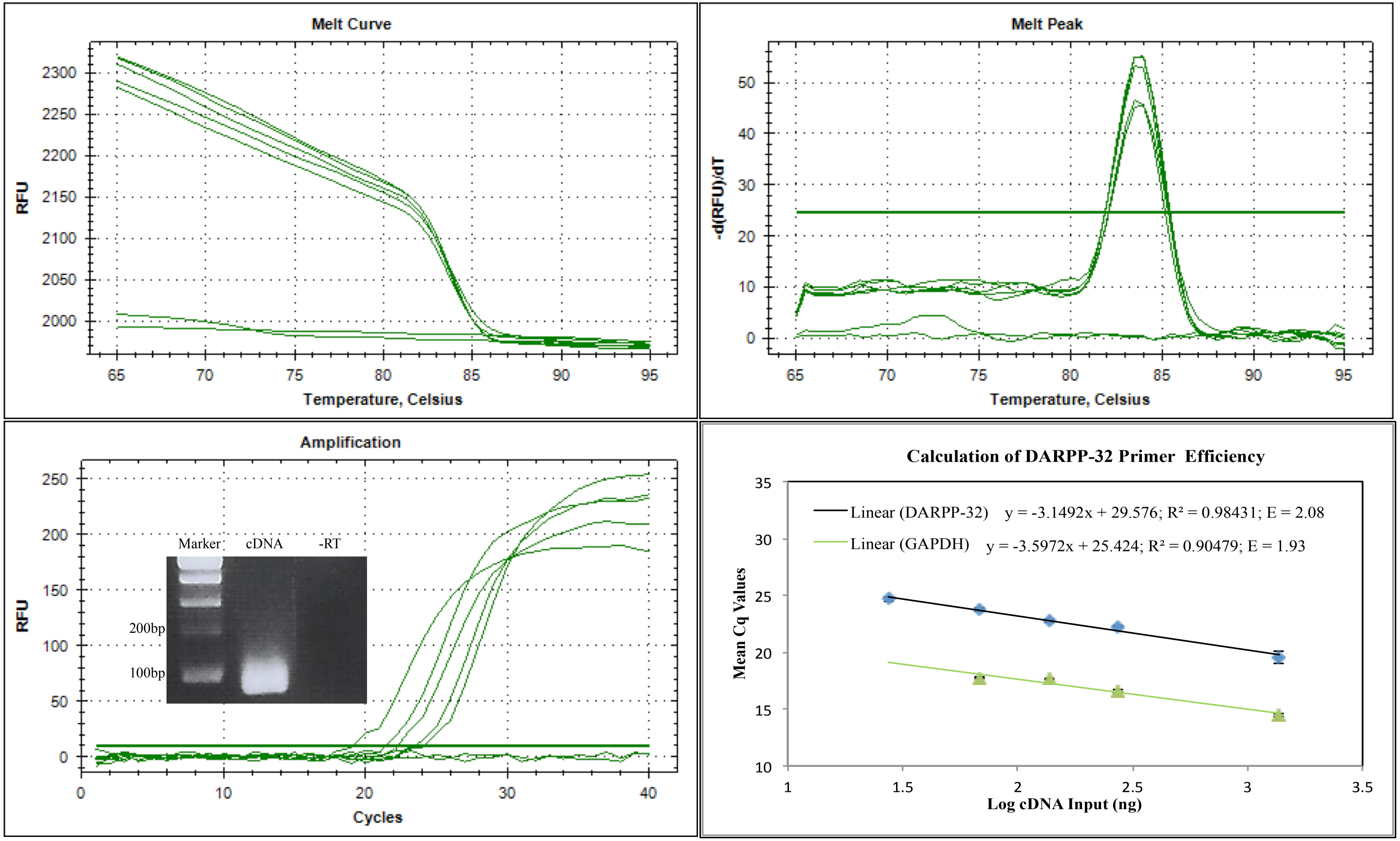
topoisomerase (DNA) II beta	
wingless-related MMTV integration site 3A	
LIM homeobox protein 2	Telencephalon development.
LIM homeobox protein 6	
POU domain, class 3, transcription factor 2	
RAS-related C3 botulinum substrate 1	
SRY-box containing gene 2	
achaete-scute complex homolog 1 (Drosophila)	
aldehyde dehydrogenase family 1, subfamily A3	
autophagy-related 7 (yeast); similar to AGP7	
centrosomal protein 120	
disabled homolog 1 (Drosophila)	
distal-less homeobox 1	
distal-less homeobox 2	
dopamine receptor D1A	
forkhead box P2	
huntingtin	
inhibitor of DNA binding 4	
neurofibromatosis 1	
nuclear receptor co-repressor 2	
nuclear receptor subfamily 2, group E, member 1	
paired box gene 6	
platelet-activating factor acetylhydrolase, isoform 1b, subunit 1	
plexin A3	
reelin	
sine oculis-related homeobox 3 homolog (Drosophila)	
wingless-related MMTV integration site 3A	
Eph receptor B2	Regulation of neurogenesis.
Meis homeobox 1	
POU domain, class 3, transcription factor 2	
SRY-box containing gene 2	
TGFB-induced factor homeobox 1	
Unc-51 like kinase 2 (C. elegans)	
X-ray repair complementing defective repair in Chinese hamster cells 6	
achaete-scute complex homolog 1 (Drosophila)	
adhesion molecule with Ig like domain 1	
asp (abnormal spindle)-like, microcephaly associated (Drosophila)	
bone morphogenetic protein receptor, type 1A	
cadherin 4	
calcium channel, voltage-dependent, P/Q type, alpha 1A subunit	
cone-rod homeobox containing gene	
delta-like 1 (Drosophila)	
distal-less homeobox 1	
distal-less homeobox 2	
hairy and enhancer of split 5 (Drosophila)	
homeo box D3	
hypothetical protein LOC100044170; X-ray repair complementing defective repair in Chinese hamster cells 4	
inhibitor of DNA binding 4	
kinase non-catalytic C-lobe domain (KIND) containing 1	
leucine rich repeat containing 4C	
microtubule-associated protein tau	
neurofibromatosis 1	
neurofilament, light polypeptide	
neurofilament, medium polypeptide	
neuroligin 1	
neuropilin 1	
neurotrophin 3	
nuclear receptor subfamily 2, group E, member 1	

orthopedia homolog (Drosophila)	
paired box gene 6	
plexin A3	
pre B-cell leukemia transcription factor 1; region containing RIKEN cDNA 2310056B04 gene; pre B-cell leukemia transcription factor 1	
steroidogenic acute regulatory protein	
tetratricopeptide repeat domain 3	
thymus cell antigen 1, theta	
tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide	
vascular endothelial growth factor C	
Eph receptor B2	Positive regulation of neurogenesis.
SRY-box containing gene 2	
X-ray repair complementing defective repair in Chinese hamster cells 6	
achaete-scute complex homolog 1 (Drosophila)	
adhesion molecule with Ig like domain 1	
cadherin 4	
hypothetical protein LOC100044170; X-ray repair complementing defective repair in Chinese hamster cells 4	
microtubule-associated protein tau	
neurofilament, light polypeptide	
neurotrophin 3	
orthopedia homolog (Drosophila)	
paired box gene 6	
steroidogenic acute regulatory protein	
vascular endothelial growth factor C	
Rac GTPase-activating protein 1; predicted gene 1859	Neuroblast proliferation.
achaete-scute complex homolog 1 (Drosophila)	
asp (abnormal spindle)-like, microcephaly associated (Drosophila)	
fibroblast growth factor receptor substrate 2	
inhibitor of DNA binding 4	
numb-like	
platelet-activating factor acetylhydrolase, isoform 1b, subunit 1	
wingless-related MMTV integration site 3A	
ADP-ribosylation factor-like 4A	Fate cell differentiation.
C-terminal binding protein 2	
SH2B adaptor protein 2	
chibby homolog 1 (Drosophila)	
cyclin D1	
glutathione peroxidase 1	
glycogen synthase kinase 3 beta	
integrin alpha 6	
mediator complex subunit 1	
methyltransferase like 8	
nuclear receptor co-repressor 2	
nudix (nucleoside diphosphate linked moiety X)-type motif 7	
peroxisome proliferator activated receptor gamma	
predicted gene 14506; BCL2/adenovirus E1B interacting protein 3; predicted gene 6532; similar to E1B 19K/Bcl-2-binding protein homolog	
regulator of G-protein signaling 2	
runt-related transcription factor 1; translocated to, 1 (cyclin D-related)	
selenium binding protein 1; hypothetical protein LOC100044204	
solute carrier family 2 (facilitated glucose transporter), member 4	
stromal cell derived factor 4	
transducin (beta)-like 1 X-linked	
Eph receptor B2	Regulation of neuron differentiation.
Meis homeobox 1	
POU domain, class 3, transcription factor 2	

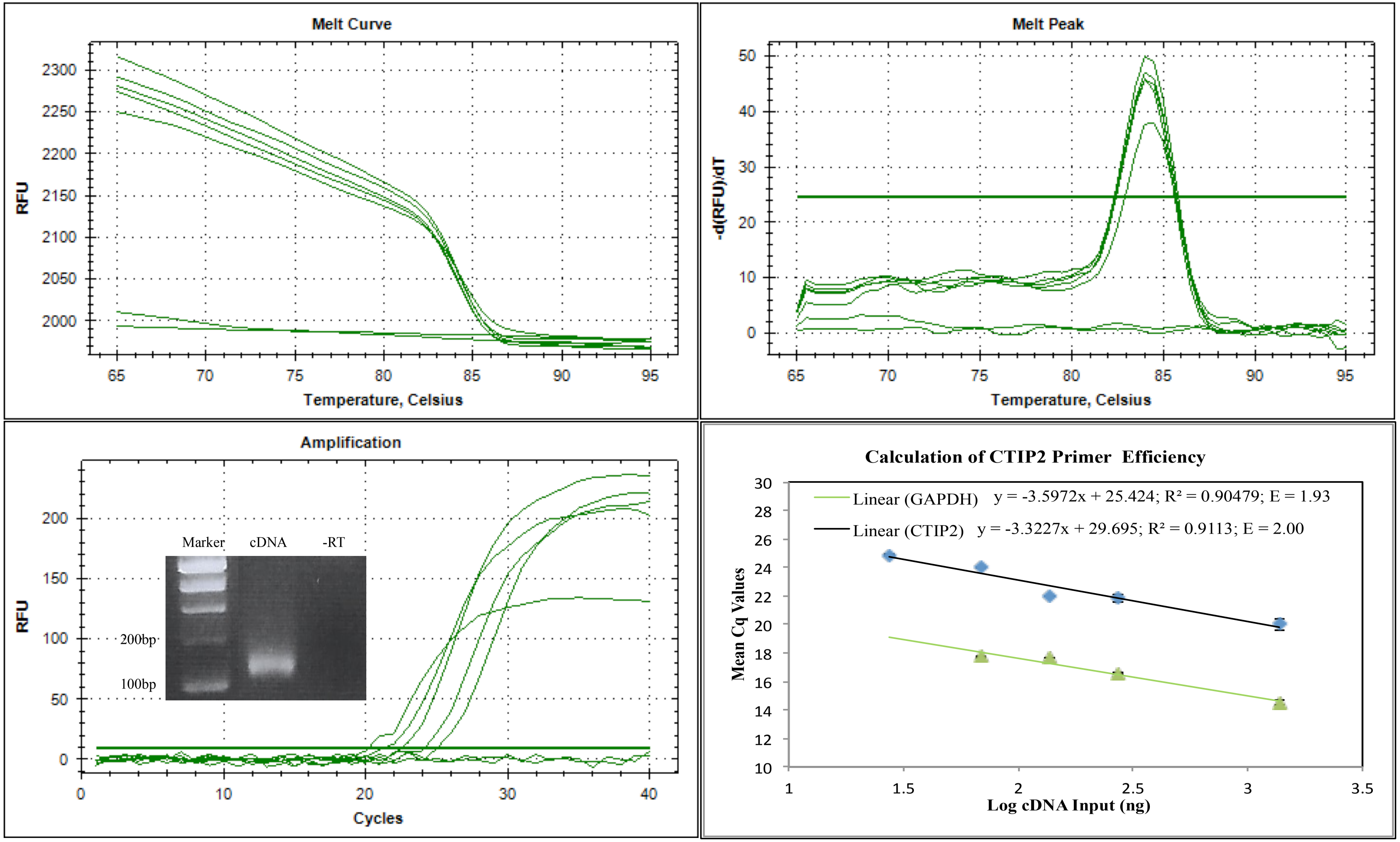
SRY-box containing gene 2	
TGFB-induced factor homeobox 1	
Unc-51 like kinase 2 (C. elegans)	
achaete-scute complex homolog 1 (Drosophila)	
adhesion molecule with Ig like domain 1	
asp (abnormal spindle)-like, microcephaly associated (Drosophila)	
cadherin 4	
calcium channel, voltage-dependent, P/Q type, alpha 1A subunit	
cone-rod homeobox containing gene	
delta-like 1 (Drosophila)	
hairy and enhancer of split 5 (Drosophila)	
homeo box D3	
inhibitor of DNA binding 4	
kinase non-catalytic C-lobe domain (KIND) containing 1	
leucine rich repeat containing 4C	
microtubule-associated protein tau	
neurofilament, light polypeptide	
neurofilament, medium polypeptide	
neuroligin 1	
neuropilin 1	
nuclear receptor subfamily 2, group E, member 1	
paired box gene 6	
plexin A3	
pre B-cell leukemia transcription factor 1; region containing RIKEN cDNA 2310056B04 gene; pre B-cell leukemia transcription factor 1	
tetratricopeptide repeat domain 3	
thymus cell antigen 1, theta	
tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide	
LIM homeobox protein 6	Forebrain neuron fate commitment.
distal-less homeobox 1	
distal-less homeobox 2	
paired box gene 6	
Eph receptor A7	
achaete-scute complex homolog 1 (Drosophila)	
ataxia telangiectasia mutated homolog (human)	Regulation of neuron apoptosis.
leukocyte specific transcript 1	
lymphotoxin A	
neurofibromatosis 1	
nuclear receptor subfamily 3, group C, member 1	
proprotein convertase subtilisin/kexin type 9	
POU domain, class 4, transcription factor 3	
apoptotic peptidase activating factor 1	
ataxia telangiectasia mutated homolog (human)	Neuron apoptosis.
caspase 3	
huntingtin	
predicted gene 14506; BCL2/adenovirus E1B interacting protein 3; predicted gene 6532; similar to E1B 19K/Bcl-2-binding protein homolog	
tumor necrosis factor receptor superfamily, member 21	

3.1 Testing the primer by QPCR:

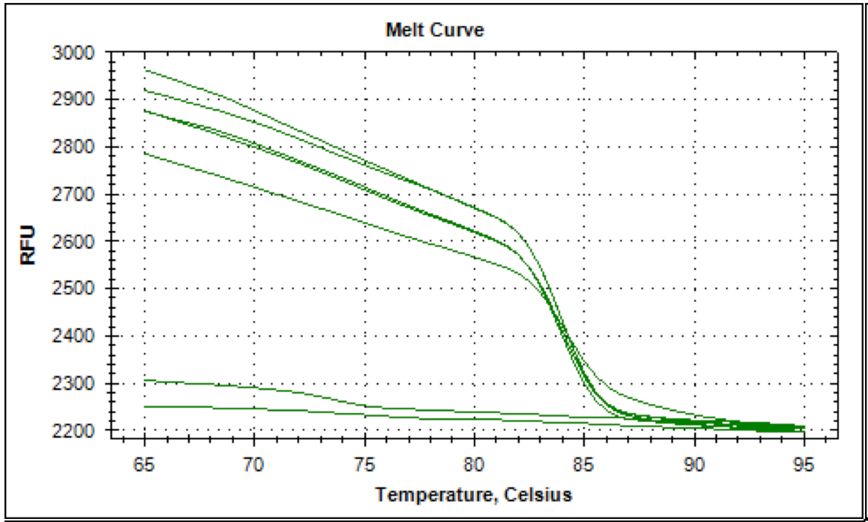
3.1.1 DARPP-32 primers

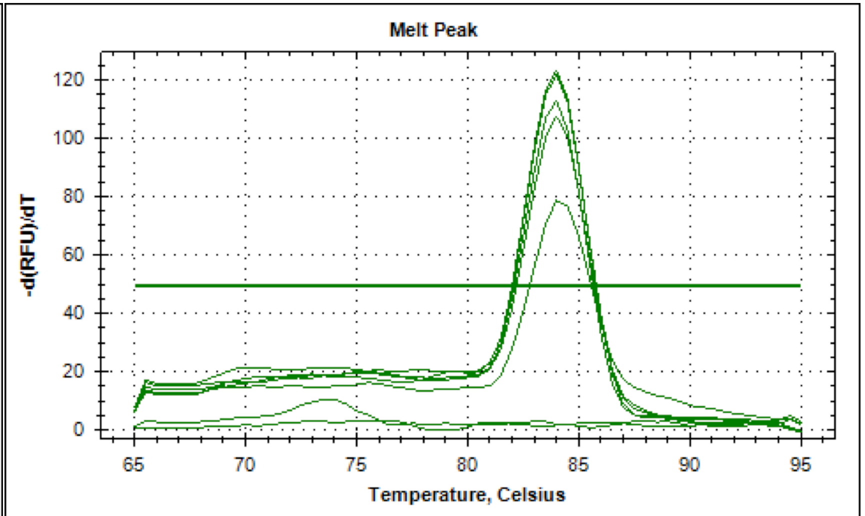


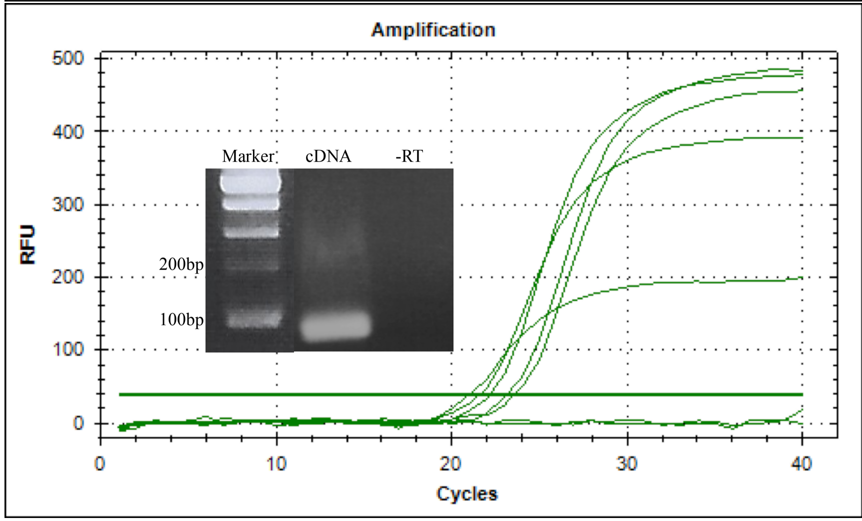
3.1.2 CTIP2 primers

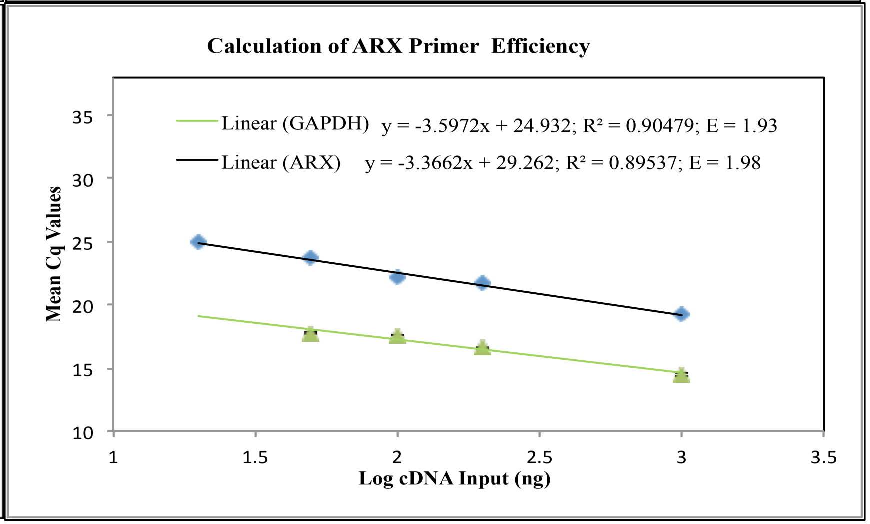


3.1.3 ARX primers

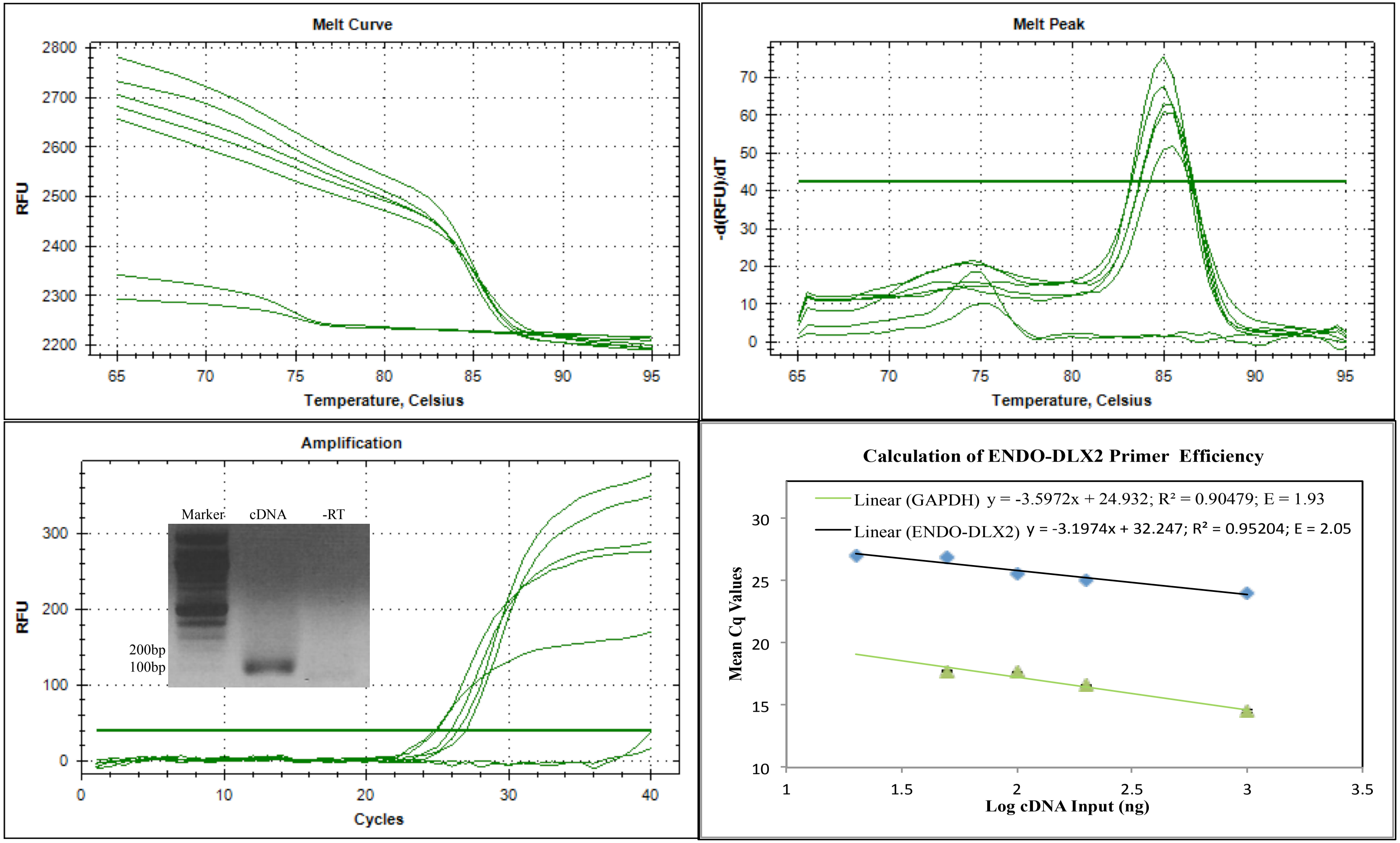




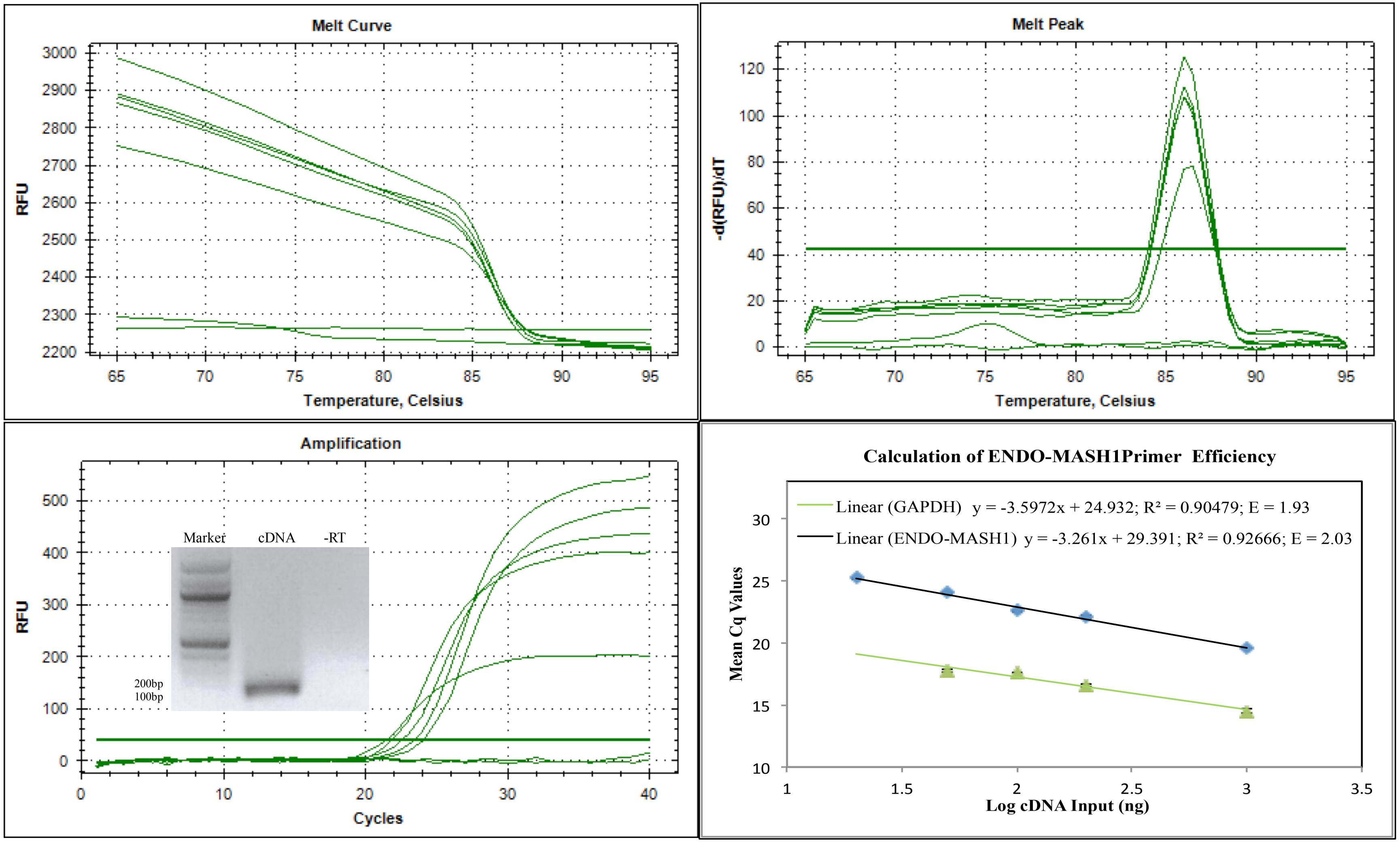




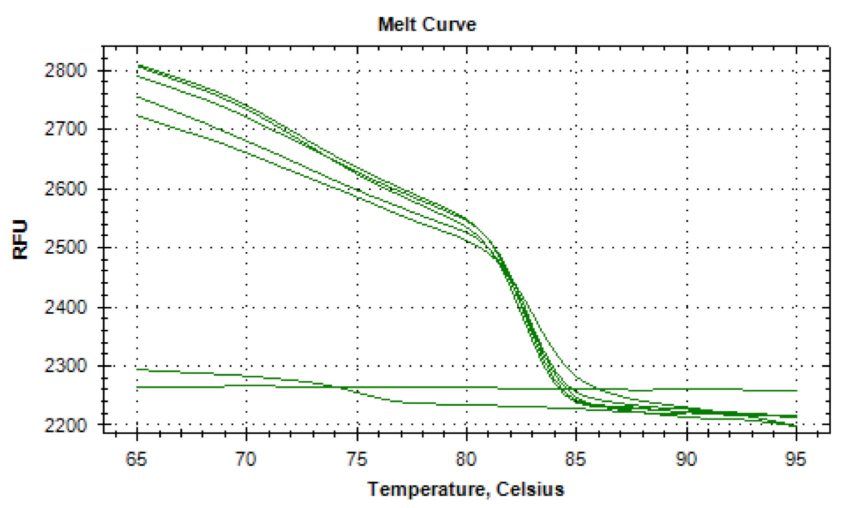
3.1.4 ENDO-DLX2 primers

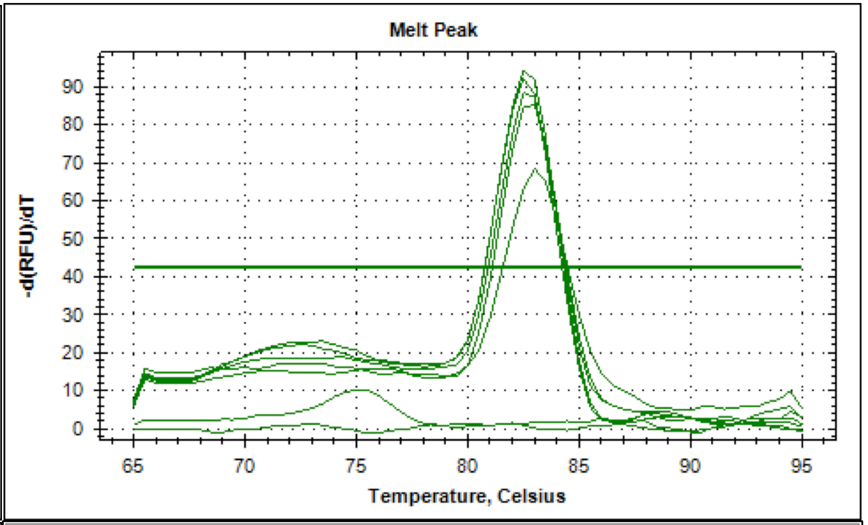


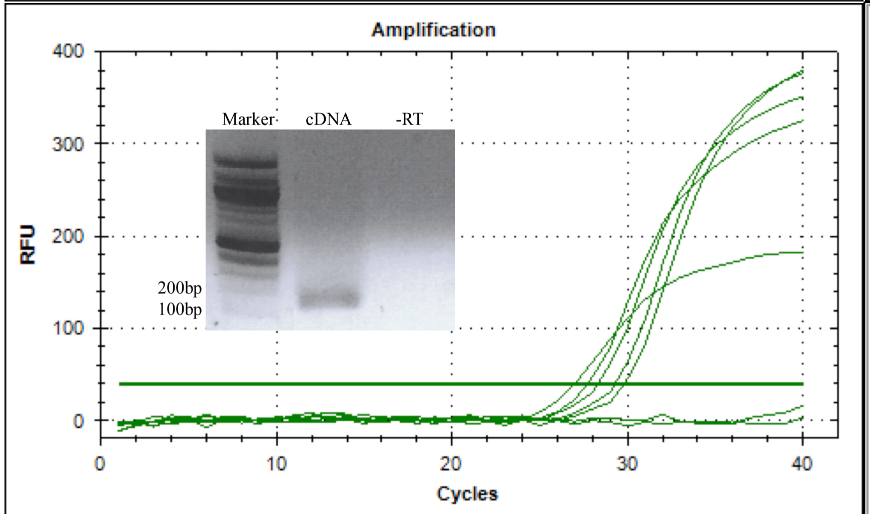
3.1.5 ENDO-MASH1 primers

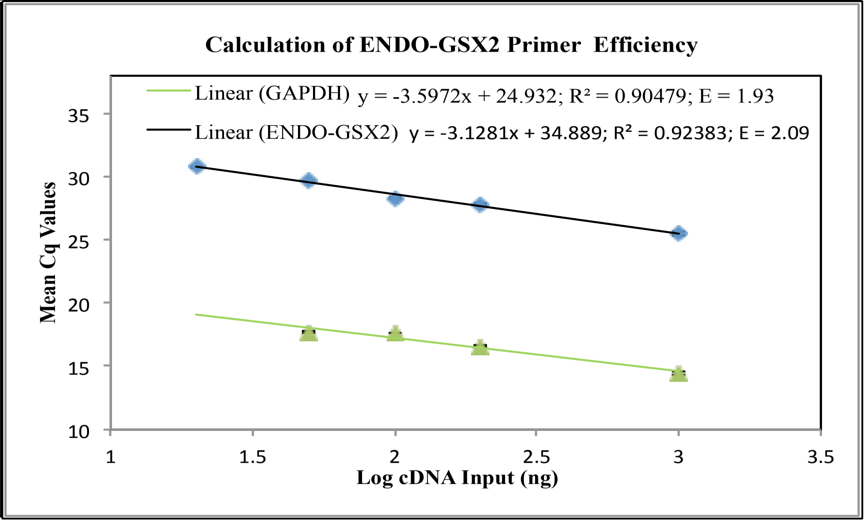


3.1.6 ENDO-GSX2 primers

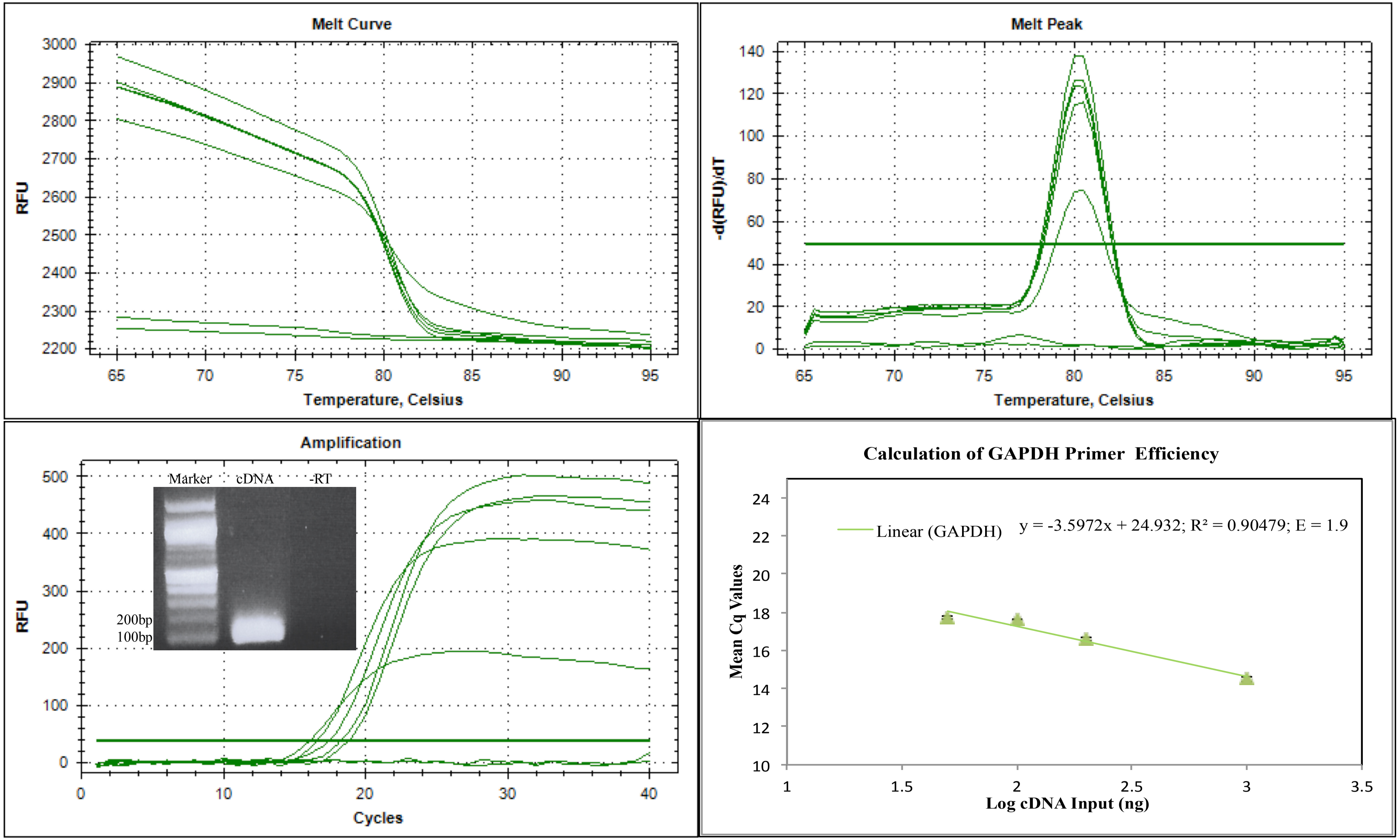




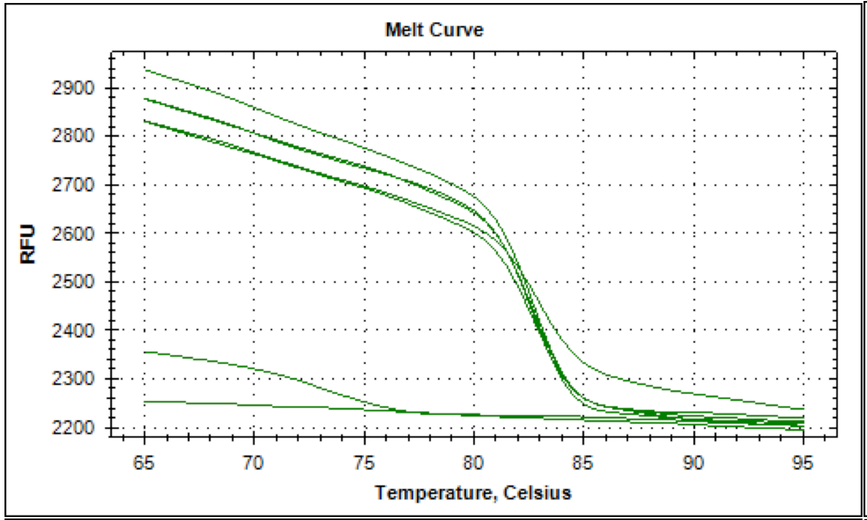


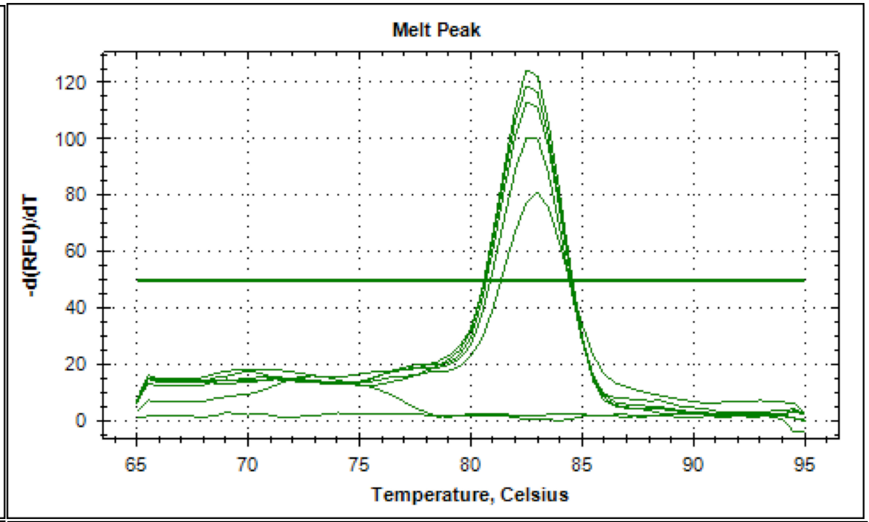


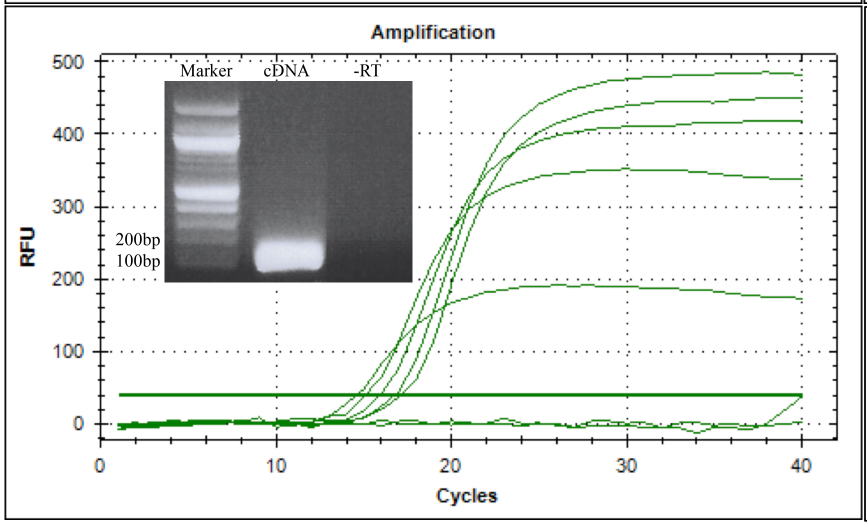
3.1.7 GAPDH primers

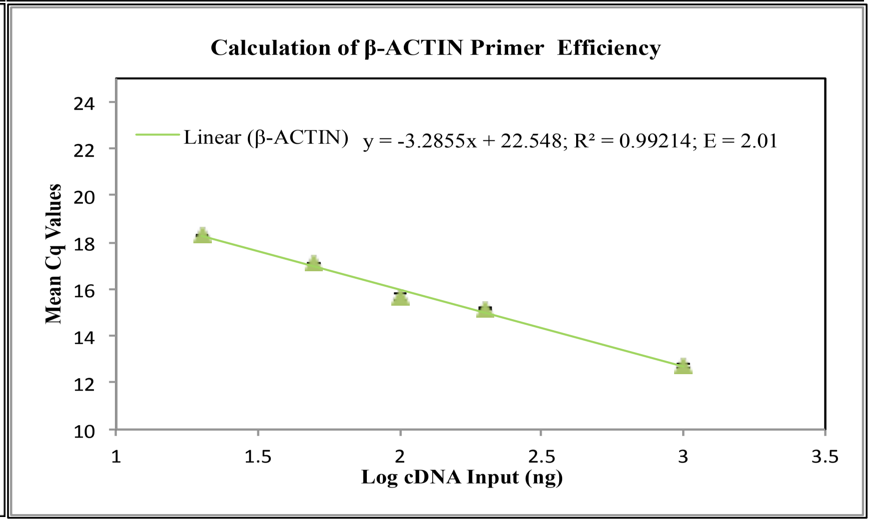


3.1.8 β-ACTIN primers

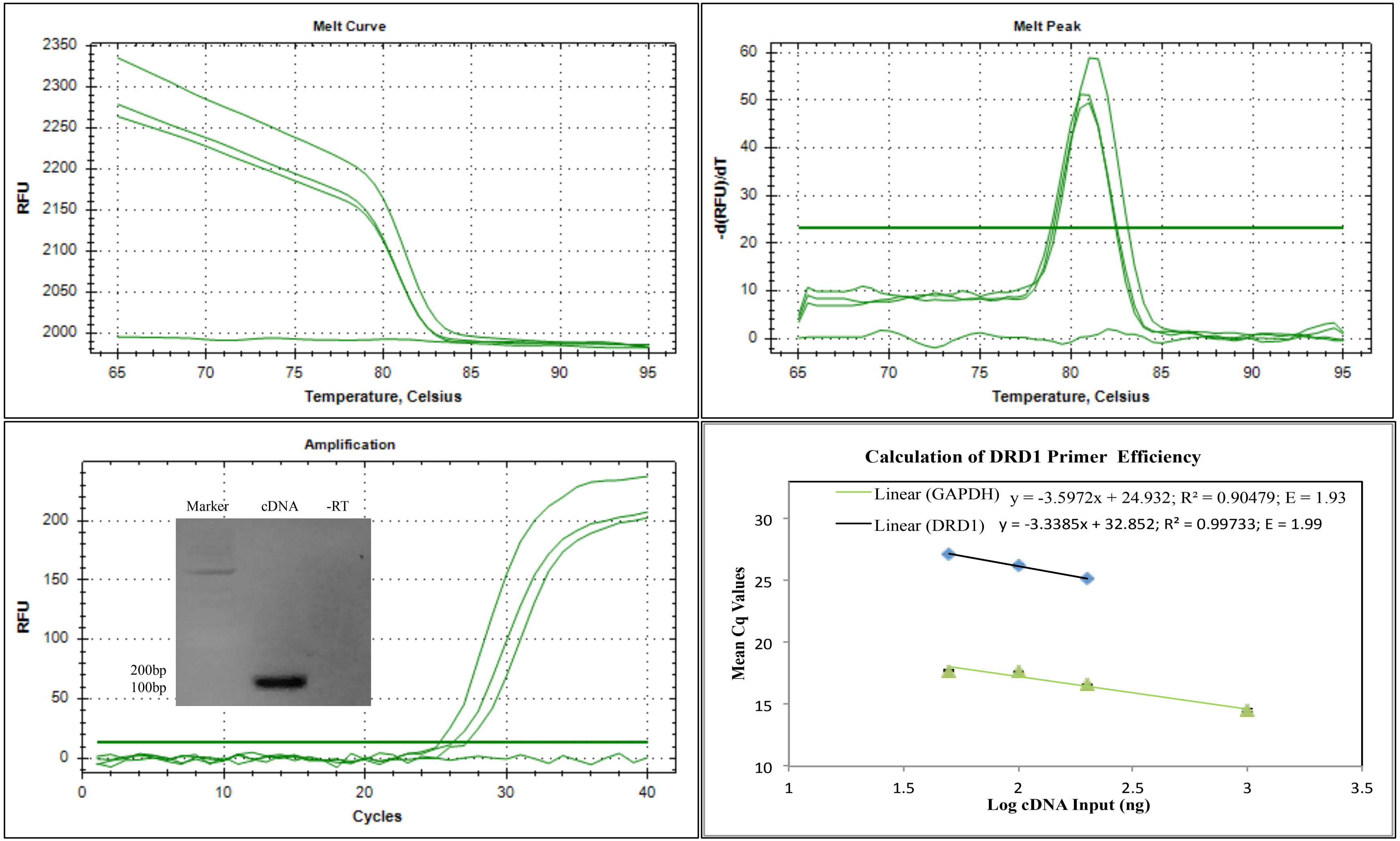




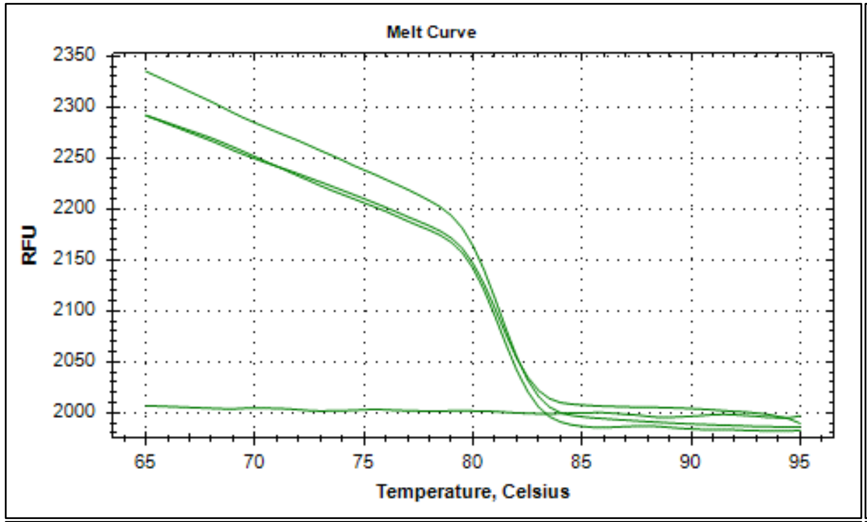


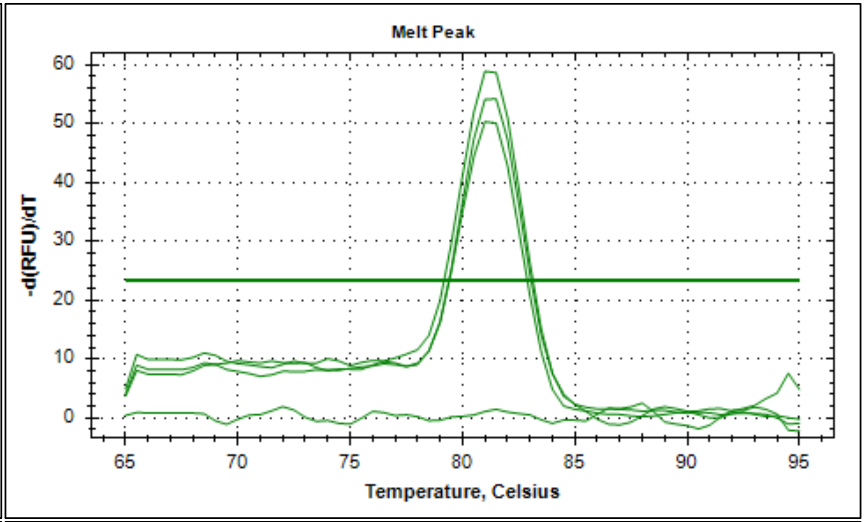


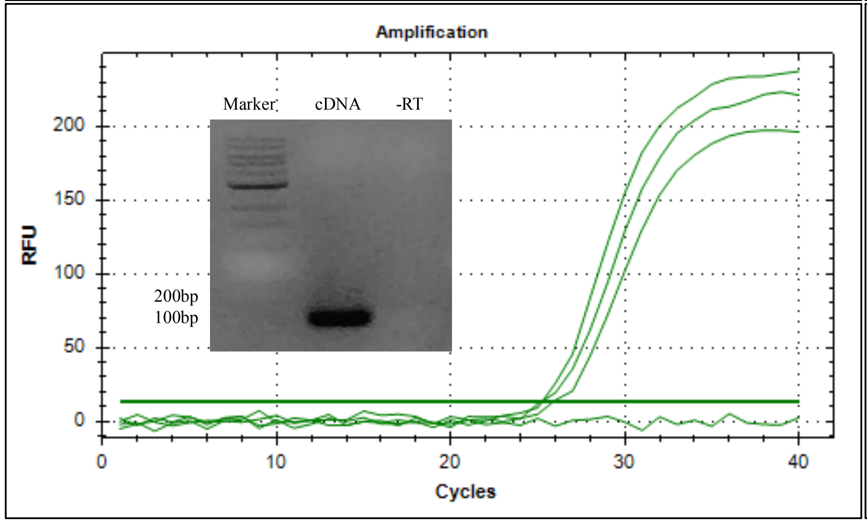
3.1.9 DRD1 primers

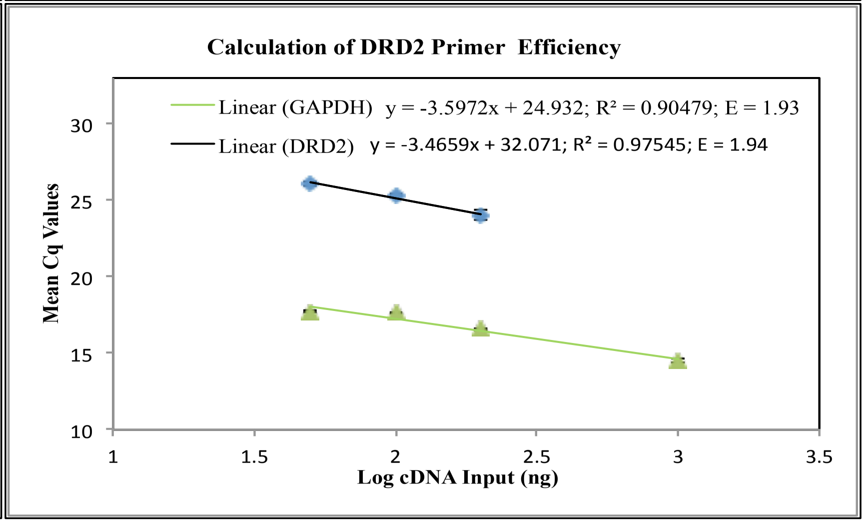


3.1.10 DRD2 primers

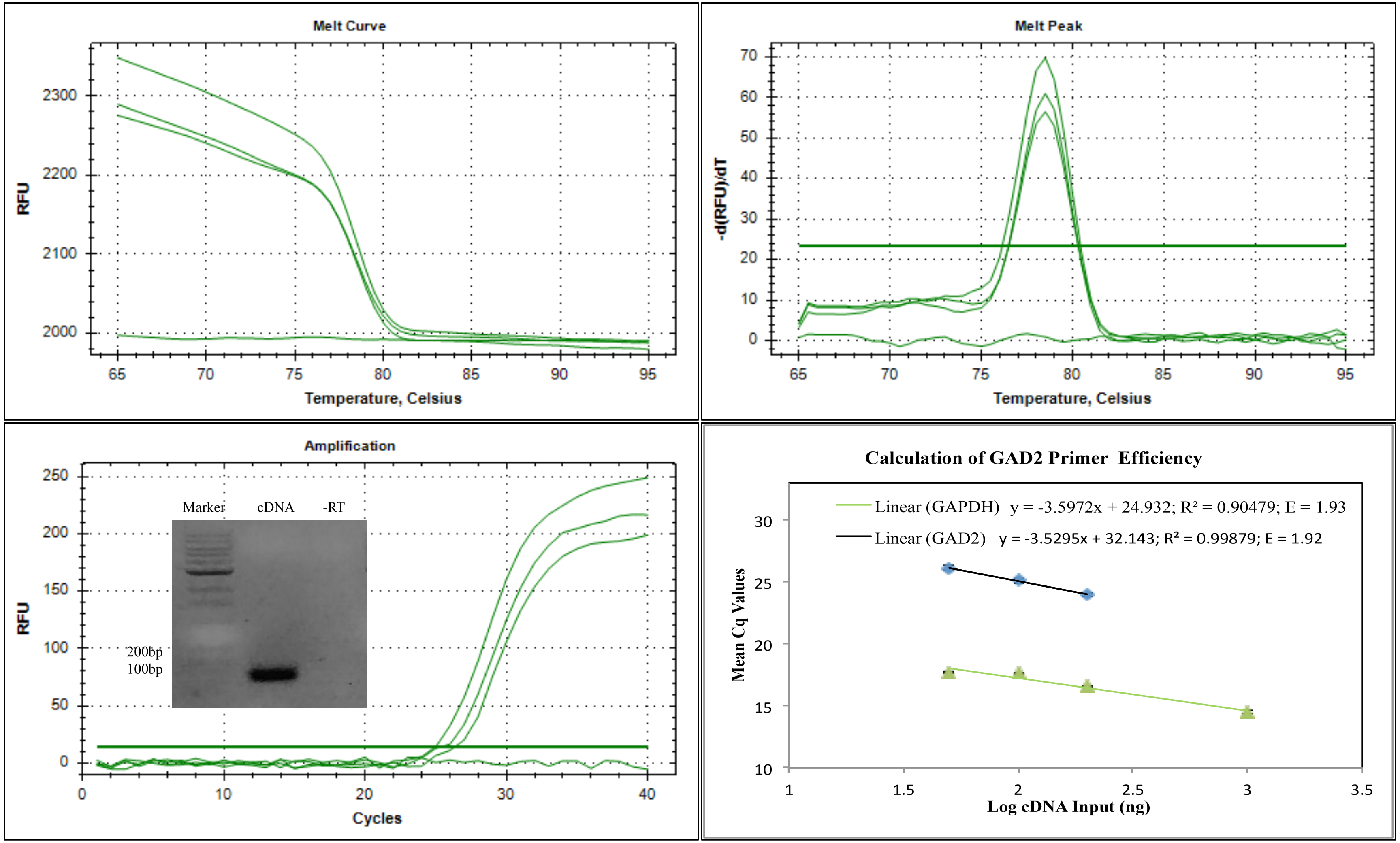




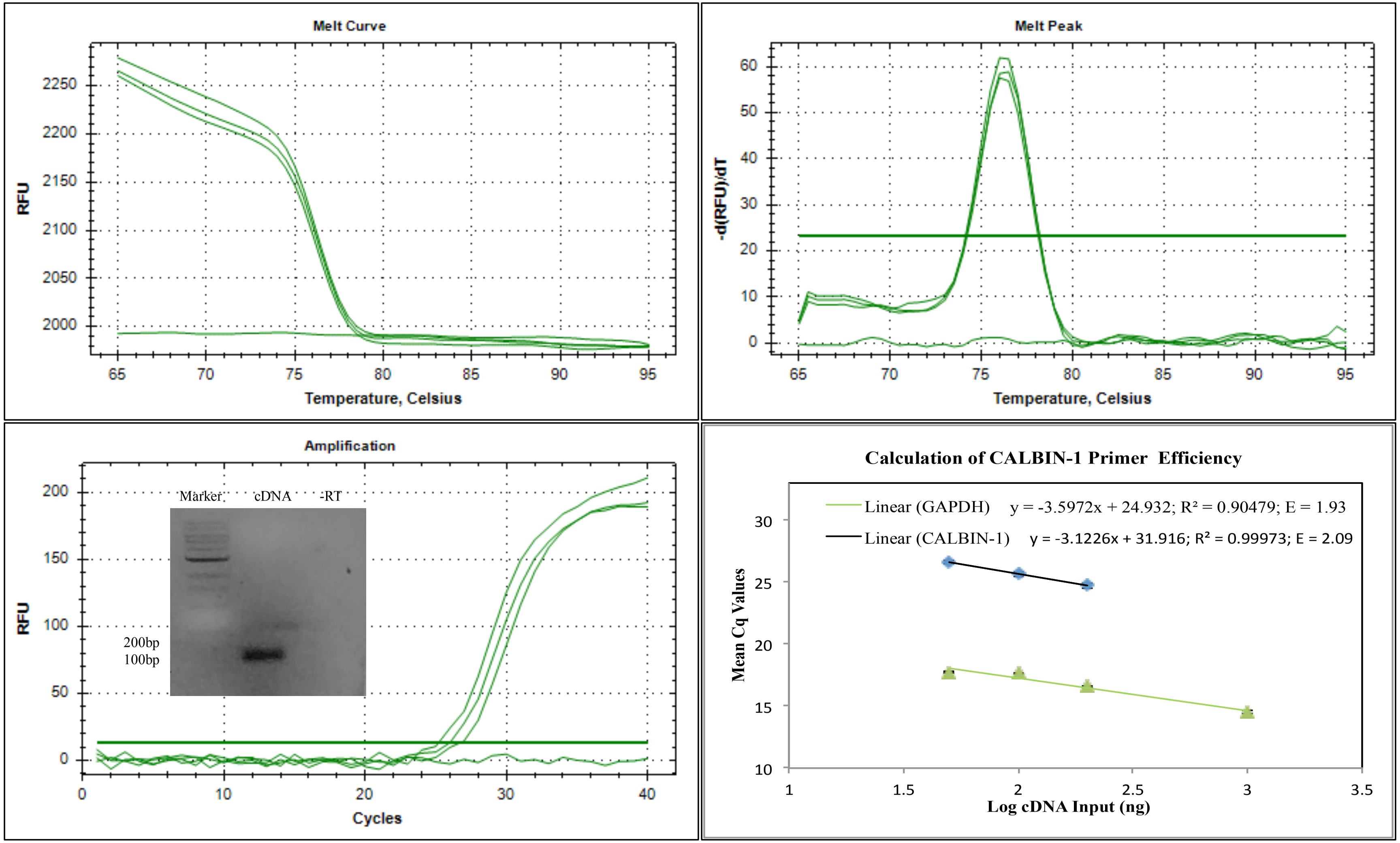




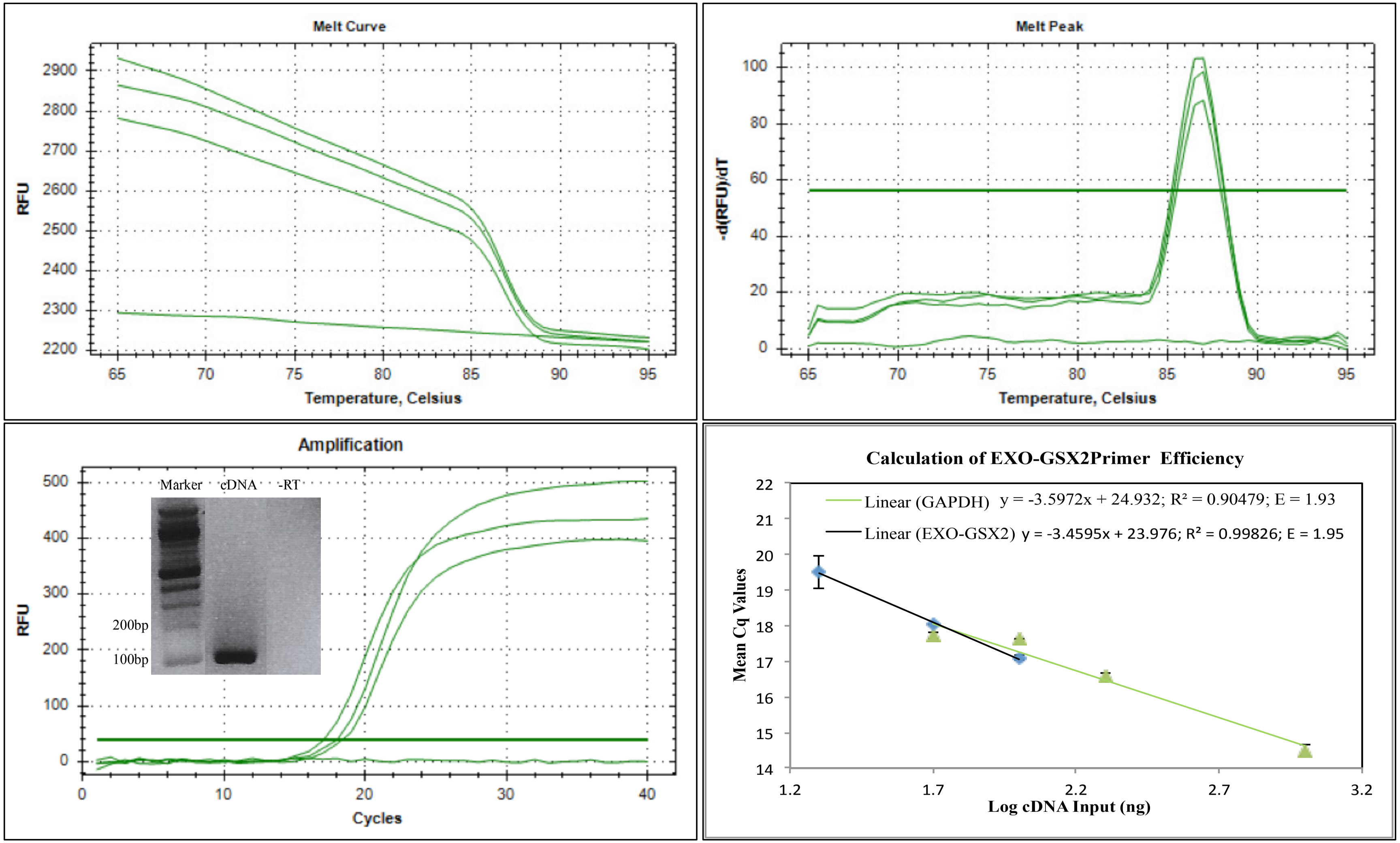
3.1.11 GAD2 primers



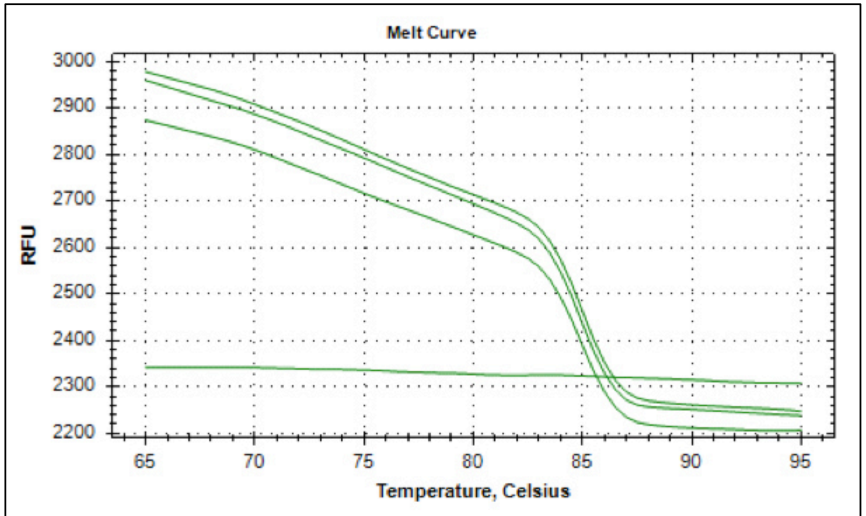
3.1.12 CALBIN-1 primers

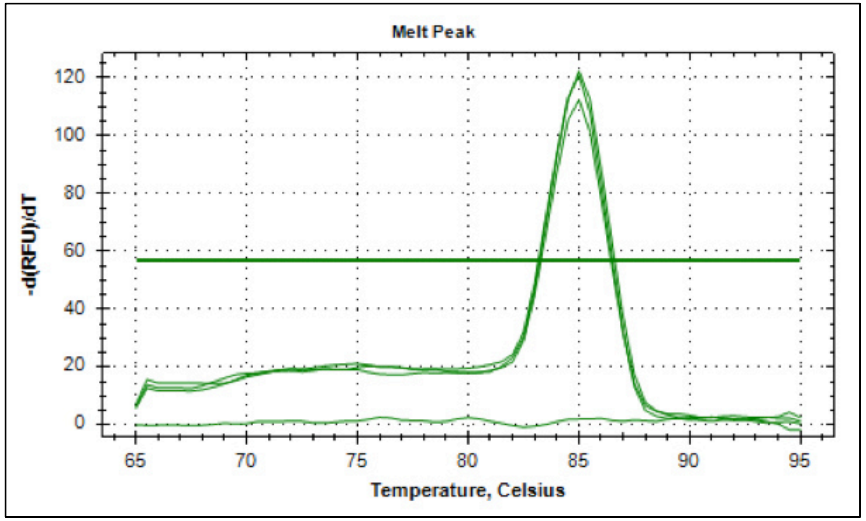


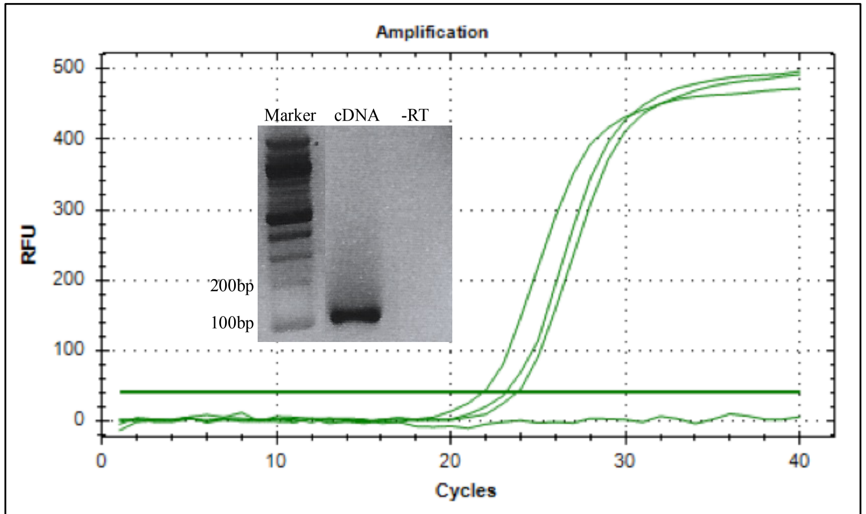
3.1.13 EXO-GSX2

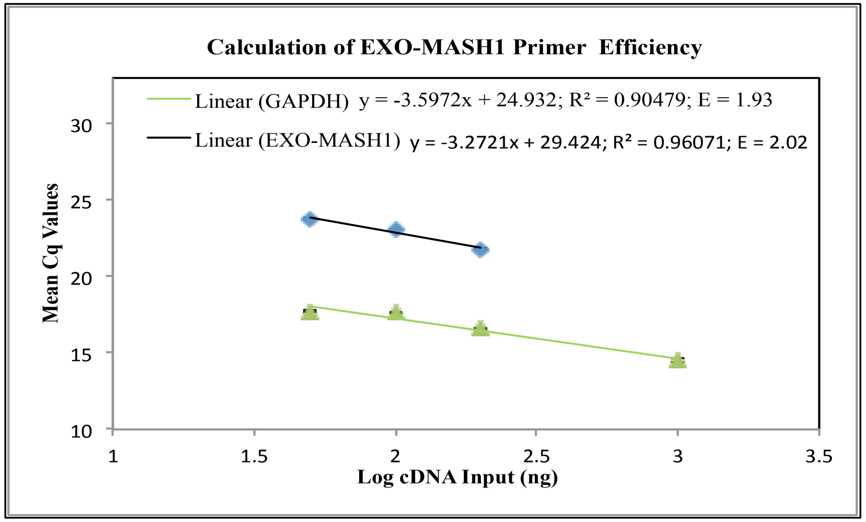


3.1.14 EXO-MASH1

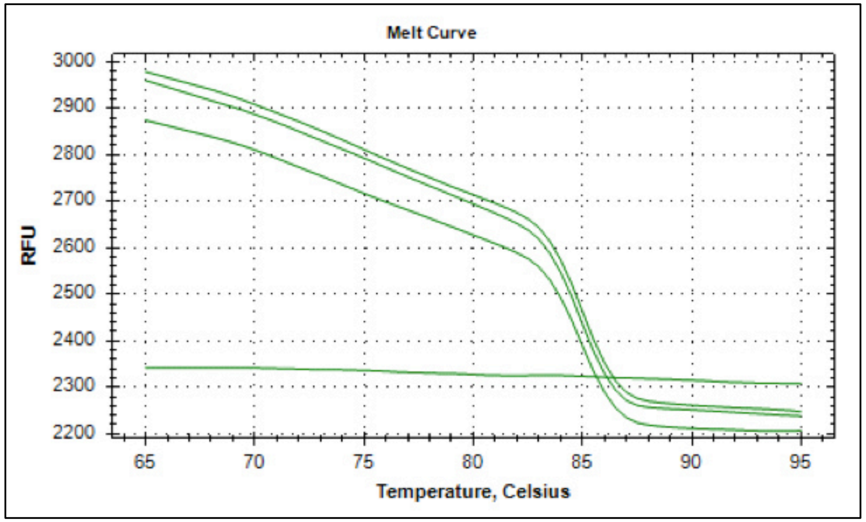


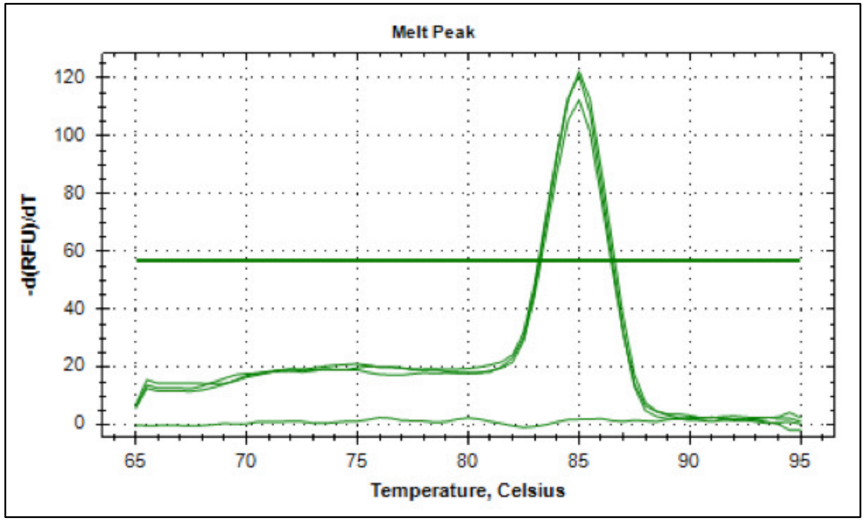


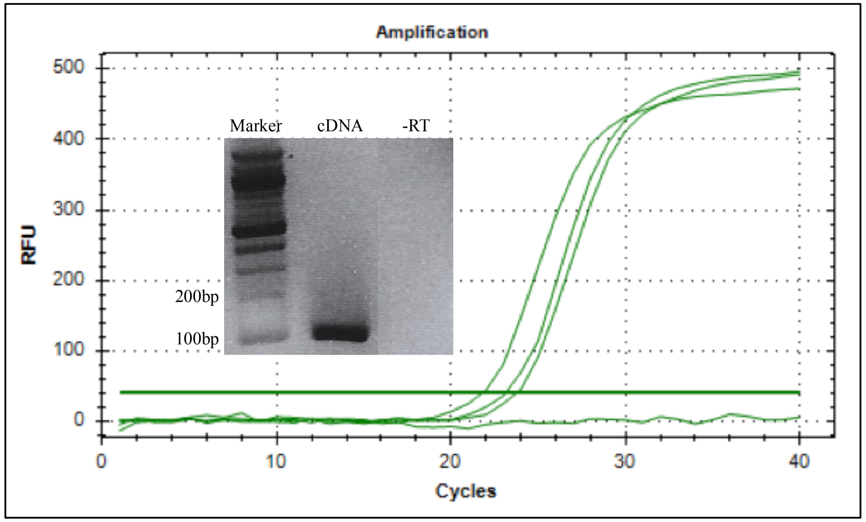


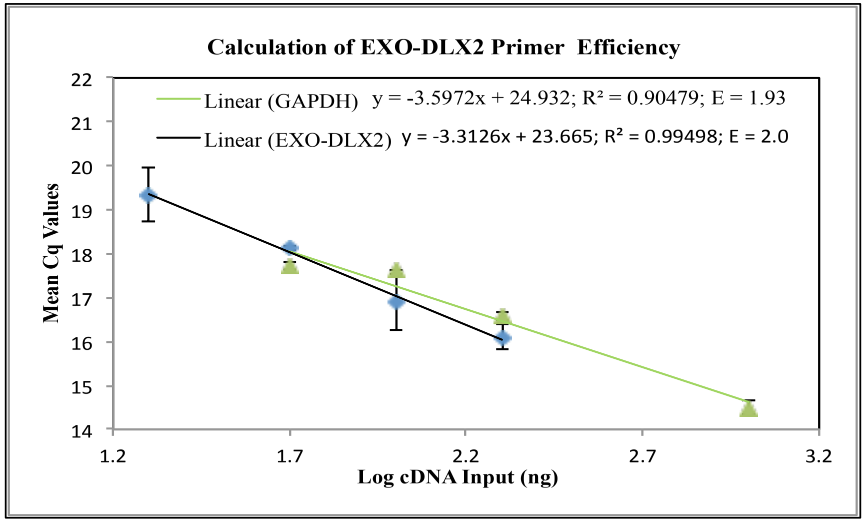


3.1.15 EXO-DLX2

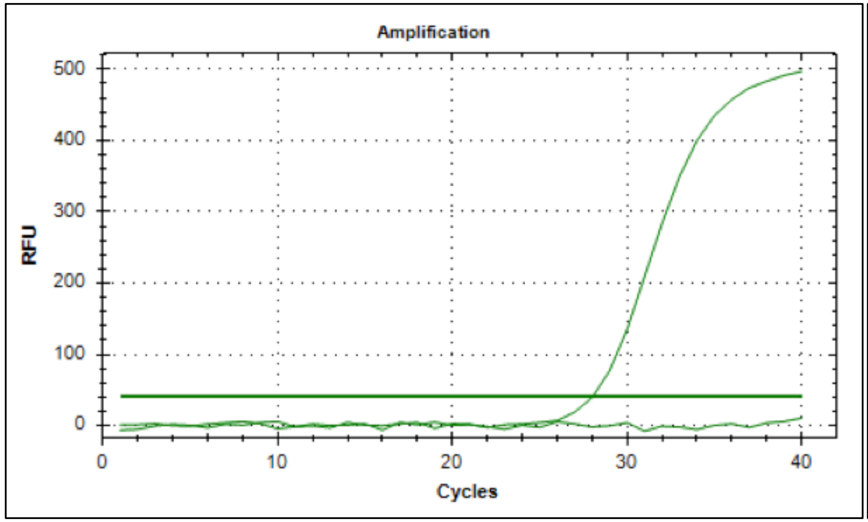


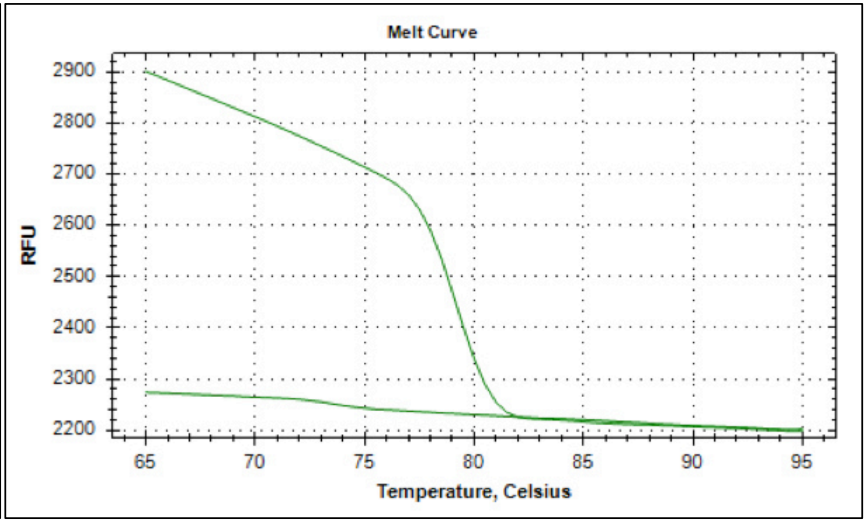


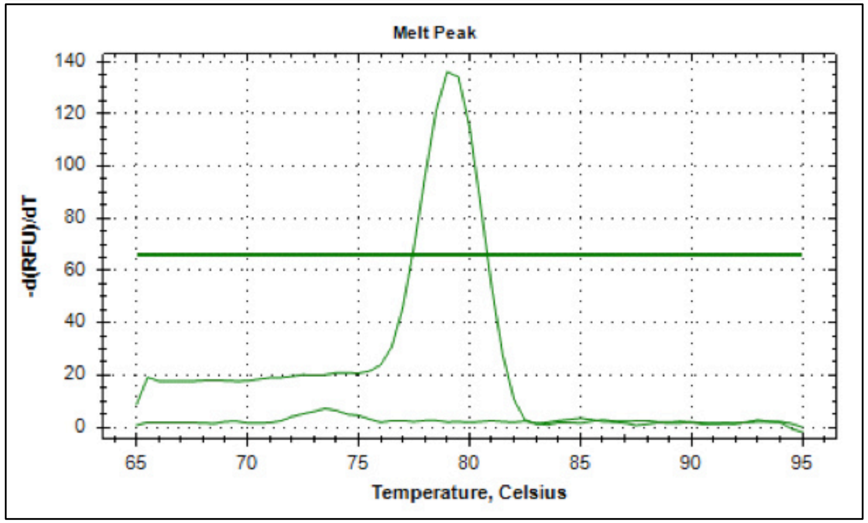




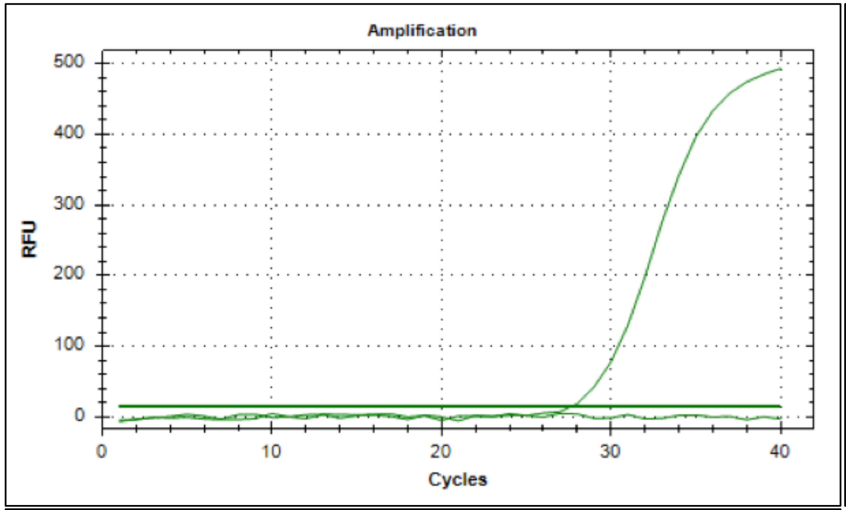
3.1.16 EBF1 (The efficiency of EBF1 primers has been tested before in Allen lab. Therefore, the ct value, melting curve and melt peak of EBF1 and –RT was shown below):

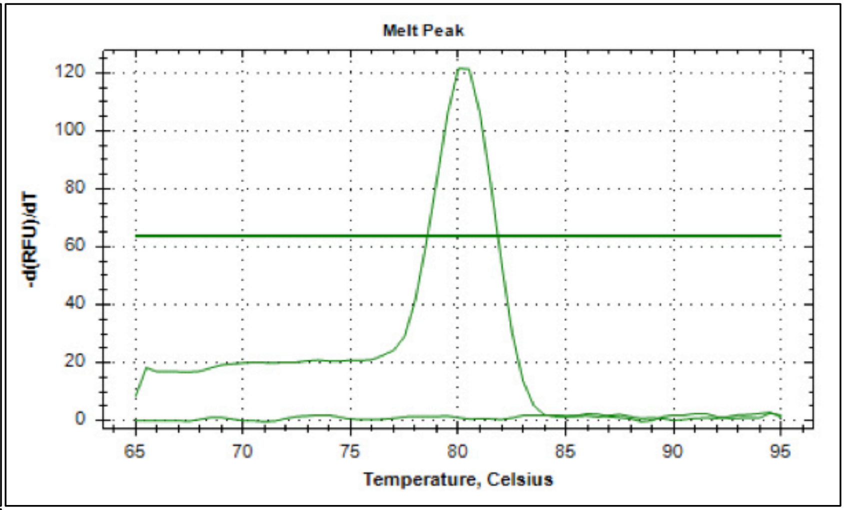


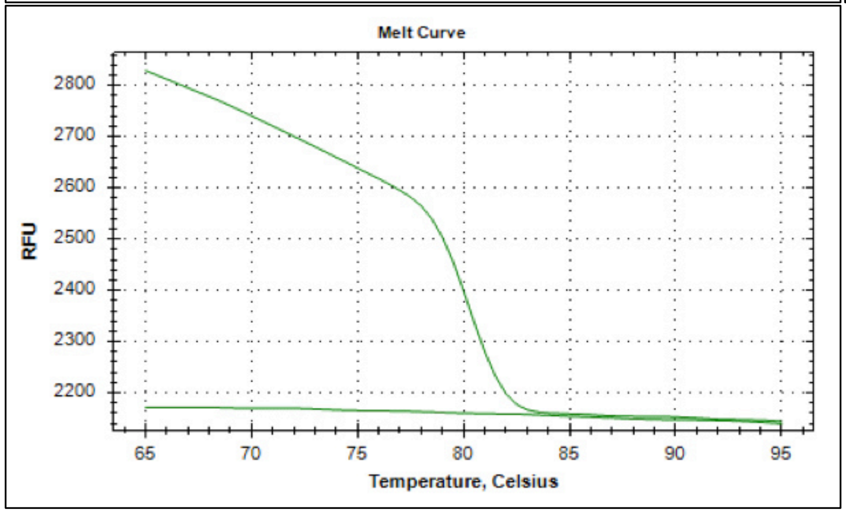




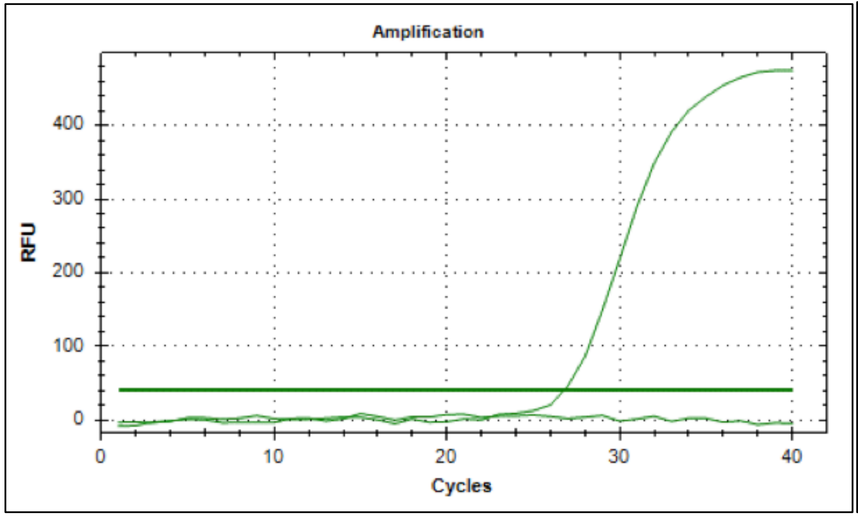
3.1.17 EMX2 (The efficiency of EMX2 primers has been tested before in Allen lab. Therefore, the ct value, melting curve and melt peak of EMX2 and –RT was shown below):

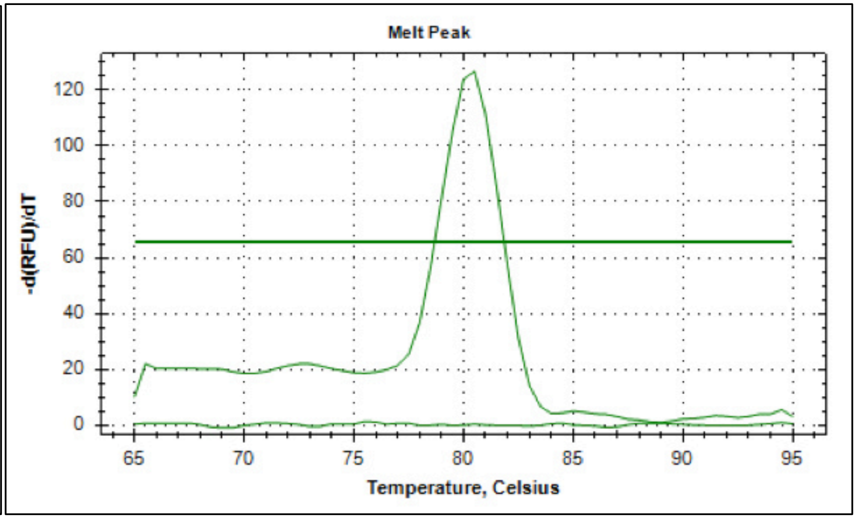


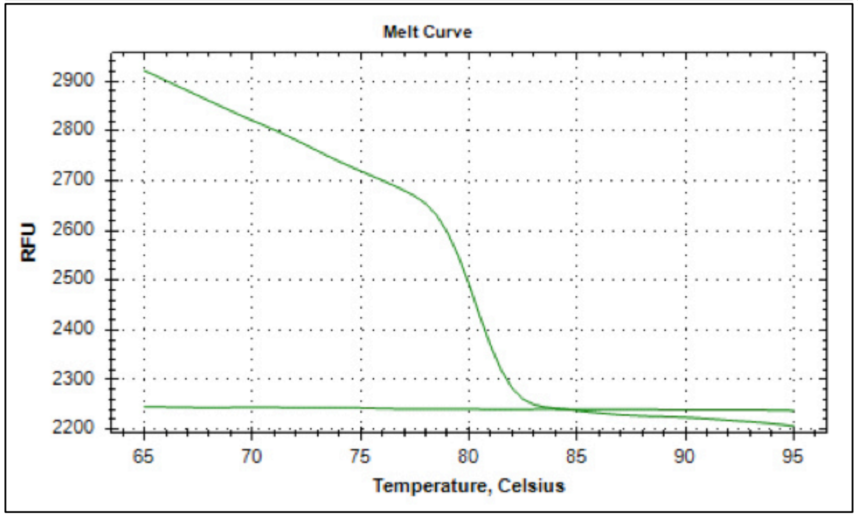




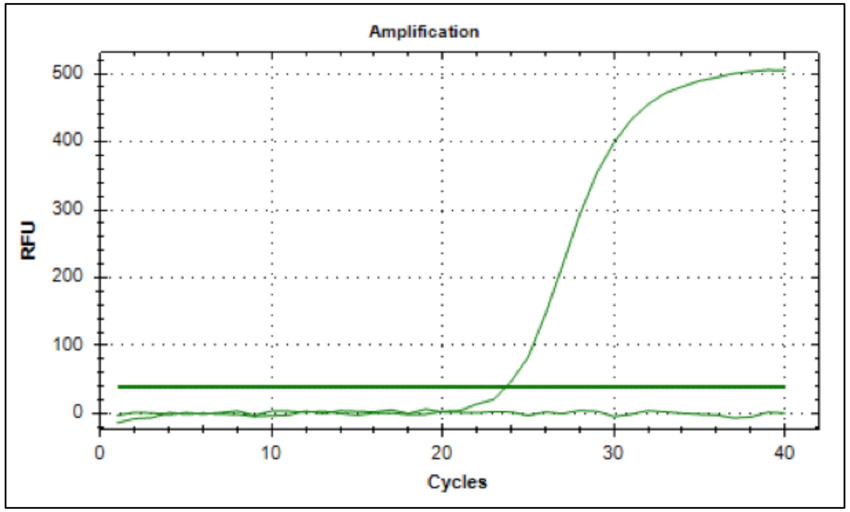
3.1.18 FOXP1 (The efficiency of FOXP1 primers has been tested before in Allen lab. Therefore, the ct value, melting curve and melt peak of FOXP1 and –RT was shown below):

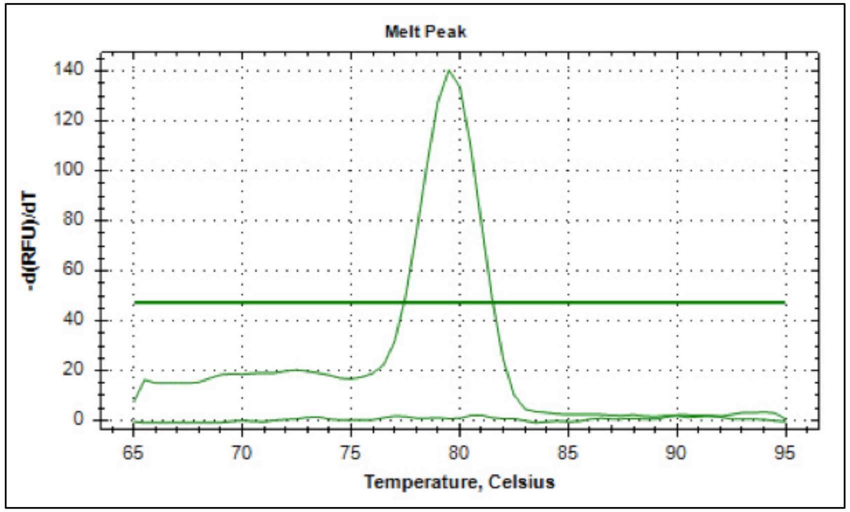


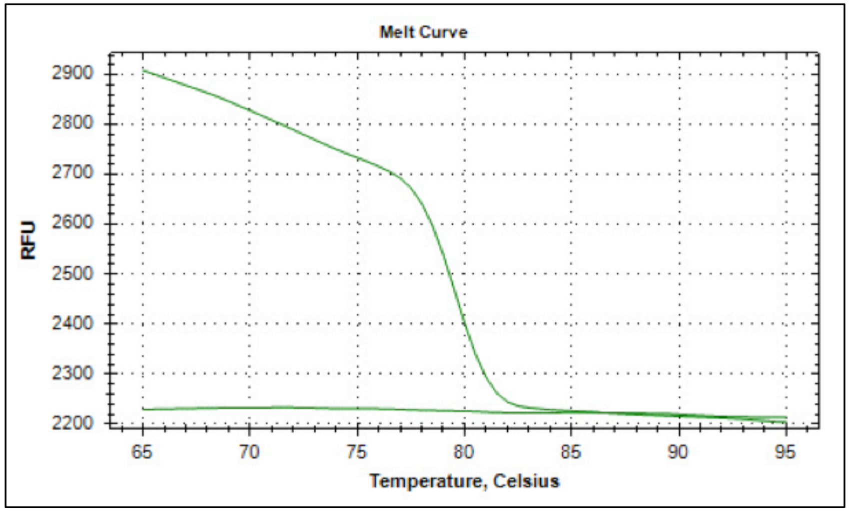




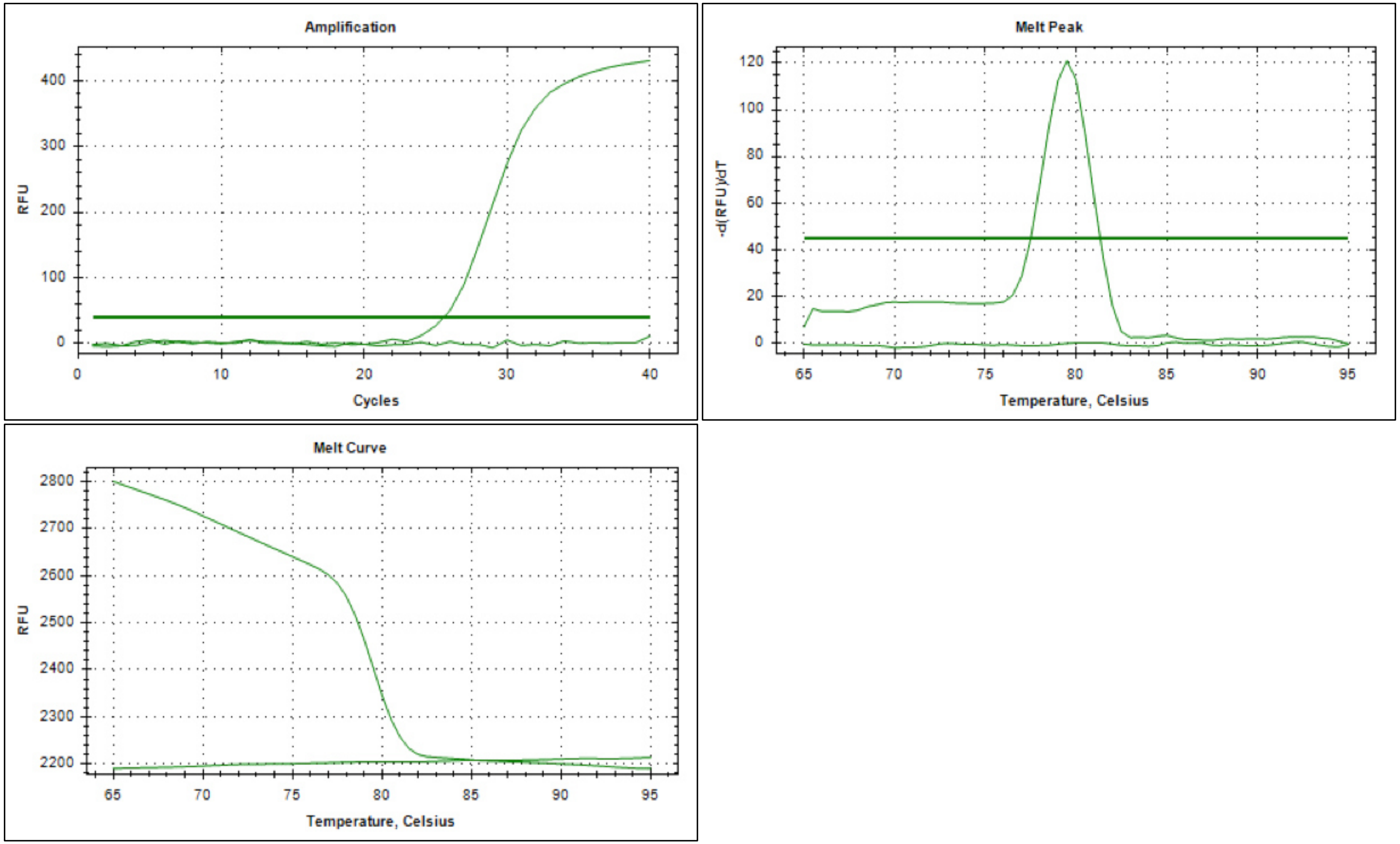
3.1.19 PAX6 (The efficiency of PAX6 primers has been tested before in Allen lab. Therefore, the ct value, melting curve and melt peak of PAX6 and –RT was shown below):







3.1.20 NKX2.1 (The efficiency of NKX2.1 primers has been tested before in Allen lab. Therefore, the ct value, melting curve and melt peak of NKX2.1 and –RT was shown below):



4.1 DNA Sequencing:

4.1.1 **P3X-2A-**DLX2 (3,595 bp)

```
CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAA
TAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGGCCGCTACAGGG
CGCTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGTTTCGGTGCGGGCCTCTTCGCTATTACGC
CAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTT
GTAAAACGACGGCCAGTGAGCGCGACGTAATACGACTCACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCC
ACGTGTCTTGTCCAGAGCTCGTCGACGAATTCAGCGCTCTCGAGACCGGTGCCGCCATGGGAGGATCCCAGTG
TACTAATTATGCTCTCTTGAAATTGGCTGGAGATGTTGAGAGCAACCCAGGTCCCAGATCTGAGGGCAGAGGA
AGTCTTCTAACATGCGGTGACGTGGAGGAGAATCCCGGCCCTTCTAGAGCCACGAAGCAAGCAGGAGATGTTG
AAGAAAACCCCGGTCCT
gct agc ATG act gga gtc ttt gac agt cta gtg gct gat atg cac tcg acc cag
atc gcc gcc tcc agc acg tac cac cag cac cag cag ccc ccg agc ggc ggc ggc
gcc ggc ccg ggt ggc aac agc agc agc agc agc agc ctc cac aag ccc cag gag
tcg ccc acc ctt ccg gtg tcc acc gcc acc gac agc agc tac tac acc aac cag
cag cac ccg gcg ggc ggc ggc ggc ggc ggg ggc tcg ccc tac gcg cac atg ggt
tcc tac cag tac caa gcc agc ggc ctc aac aac gtc cct tac tcc gcc aag agc
agc tat gac ctg ggc tac acc gcc gcc tac acc tcc tac gct ccc tat gga acc
agt tcg tcc cca gcc aac aac gag cct gag aag gag gac ctt gag cct gaa att
cgg ata gtg aac ggg aag cca aag aaa gtc cgg aaa ccc cgc acc atc tac tcc
agt ttc cag ctg gcg gct ctt cag cgg cgt ttc caa aag act caa tac ttg gcc
ttg ccg gag cga gcc gag ctg gcg gcc tct ctg ggc ctc acc cag act cag gtc
aaa atc tgg ttc cag aac cgc cgg tcc aag ttc aag aag atg tgg aaa agt ggt
gag atc ccc tcg gag cag cac cct ggg gcc agc gct tct cca cct tgt gct tcg
ccg cca gtc tca gcg ccg gcc tcc tgg gac ttt ggt gtg ccg cag cgg atg gcg
ggc ggc ggt ggt ccg ggc agt ggc ggc agc ggc gcc ggc agc tcg ggc tcc agc
ccg agc agc gcg gcc tcg gct ttt ctg ggc aac tac ccc tgg tac cac cag acc
tcg gga tcc gcc tca cac ctg cag gcc acg gcg ccg ctg ctg cac ccc act cag
acc ccg cag ccg cat cac cac cac cac cat cac ggc ggc ggg ggc gcc ccg gtg
agc gcg ggg acg att ttc gct agc
TAAGTCGACGGTACCTGGAGCACAAGACTGGCCTCATGGGCCTTCCGCTCACTGCCCGCTTTCCAGTCGGGAA
ACCTGTCGTGCCAGCTGCATTAACATGGTCATAGCTGTTTCCTTGCGTATTGGGCGCTCTCCGCTTCCTCGCT
CACTGACTCGCTGCGCTCGGTCGTTCGGGTAAAGCCTGGGGTGCCTAATGAGCAAAAGGCCAGCAAAAGGCCA
GGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCG
ACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTC
GTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGC
TTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGA
ACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGAC
TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCT
TGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTAC
CTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGC
AAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTC
AGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTT
AAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTA
ATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGA
TAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGAACCACGCTCACCGGC
TCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCC
TCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTG
TTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACG
ATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTC
AGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCAT
CCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAG
TTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGA
AAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTG
```

CACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGC
CGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGC
ATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTC
CGCGCACATTTCCCCGAAAAGTGCCAC

4.1.2 P3X-2A-MASH1 (3,319 bp)

```
CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAA
TAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGGCCGCTACAGGG
CGCTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGTTTCGGTGCGGGCCTCTTCGCTATTACGC
CAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTT
GTAAAACGACGGCCAGTGAGCGCGACGTAATACGACTCACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCC
ACGTGTCTTGTCCAGAGCTCGTCGACGAATTCAGCGCTCTCGAGACCGGTGCCGCCATGGGA
gga tcc ATG gaa agc tct gcc aag atg gag agc ggc ggc gcc ggc cag cag ccc
cag ccg cag ccc cag cag ccc ttc ctg ccg ccc gca gcc tgt ttc ttt gcc acg
gcc gca gcc gcg gcg gcc gca gcc gcc gca gcg gca gcg cag agc gcg cag cag
cag cag cag cag cag cag cag cag cag cag gcg ccg cag ctg aga ccg gcg gcc
gac ggc cag ccc tca ggg ggc ggt cac aag tca gcg ccc aag caa gtc aag cga
cag cgc tcg tct tcg ccc gaa ctg atg cgc tgc aaa cgc cgg ctc aac ttc agc
ggc ttt ggc tac agc ctg ccg cag cag cag ccg gcc gcc gtg gcg cgc cgc aac
gag cgc gag cgc aac cgc gtc aag ttg gtc aac ctg ggc ttt gcc acc ctt cgg
gag cac gtc ccc aac ggc gcg gcc aac aag aag atg agt aag gtg gag aca ctg
cgc tcg gcg gtc gag tac atc cgc gcg ctg cag cag ctg ctg gac gag cat gac
gcg gtg agc gcc gcc ttc cag gca ggc gtc ctg tcg ccc acc atc tcc ccc aac
tac tcc aac gac ttg aac tcc atg gcc ggc tcg ccg gtc tca tcc tac tcg tcg
gac gag ggc tct tac gac ccg ctc agc ccc gag gag cag gag ctt ctc gac ttc
acc aac tgg ttc gGA TCC
CAGTGTACTAATTATGCTCTCTTGAAATTGGCTGGAGATGTTGAGAGCAACCCAGGTCCCAGATCTGAGGGCA
GAGGAAGTCTTCTAACATGCGGTGACGTGGAGGAGAATCCCGGCCCTTCTAGAGCCACGAAGCAAGCAGGAGA
TGTTGAAGAAAACCCCGGTCCTGCTAGCTAAGTCGACGGTACCTGGAGCACAAGACTGGCCTCATGGGCCTTC
CGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAACATGGTCATAGCTGTTTCCTTG
CGTATTGGGCGCTCTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGGTAAAGCCTGGGGTGCC
TAATGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCC
GCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATA
CCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCC
GCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCG
TTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCG
TCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCG
AGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTG
GTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCAC
CGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCT
TTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTAT
CAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTA
AACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCC
ATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAA
TGATACCGCGAGAACCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCG
CAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGT
TCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTA
TGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGT
TAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCA
CTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCAT
TCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAG
CAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTG
AGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTG
GGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCAT
ACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGT
ATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC
```

4.1.3 **P3X-2A-**DLX2/MASH1 (4,309 bp)

```
CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAA
TAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGGCCGCTACAGGG
CGCTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGTTTCGGTGCGGGCCTCTTCGCTATTACGC
CAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTT
GTAAAACGACGGCCAGTGAGCGCGACGTAATACGACTCACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCC
ACGTGTCTTGTCCAGAGCTCGTCGACGAATTCAGCGCTCTCGAGACCGGTGCCGCCATGGGA
gga tcc ATG gaa agc tct gcc aag atg gag agc ggc ggc gcc ggc cag cag ccc
cag ccg cag ccc cag cag ccc ttc ctg ccg ccc gca gcc tgt ttc ttt gcc acg
gcc gca gcc gcg gcg gcc gca gcc gcc gca gcg gca gcg cag agc gcg cag cag
cag cag cag cag cag cag cag cag cag cag gcg ccg cag ctg aga ccg gcg gcc
gac ggc cag ccc tca ggg ggc ggt cac aag tca gcg ccc aag caa gtc aag cga
cag cgc tcg tct tcg ccc gaa ctg atg cgc tgc aaa cgc cgg ctc aac ttc agc
ggc ttt ggc tac agc ctg ccg cag cag cag ccg gcc gcc gtg gcg cgc cgc aac
gag cgc gag cgc aac cgc gtc aag ttg gtc aac ctg ggc ttt gcc acc ctt cgg
gag cac gtc ccc aac ggc gcg gcc aac aag aag atg agt aag gtg gag aca ctg
cgc tcg gcg gtc gag tac atc cgc gcg ctg cag cag ctg ctg gac gag cat gac
gcg gtg agc gcc gcc ttc cag gca ggc gtc ctg tcg ccc acc atc tcc ccc aac
tac tcc aac gac ttg aac tcc atg gcc ggc tcg ccg gtc tca tcc tac tcg tcg
gac gag ggc tct tac gac ccg ctc agc ccc gag gag cag gag ctt ctc gac ttc
acc aac tgg ttc gga tcc
CAGTGTACTAATTATGCTCTCTTGAAATTGGCTGGAGATGTTGAGAGCAACCCAGGTCCCAGATCTGAGGGCA
GAGGAAGTCTTCTAACATGCGGTGACGTGGAGGAGAATCCCGGCCCTTCTAGAGCCACGAAGCAAGCAGGAGA
TGTTGAAGAAAACCCCGGTCCT
gct agc ATG act gga gtc ttt gac agt cta gtg gct gat atg cac tcg acc cag
atc gcc gcc tcc agc acg tac cac cag cac cag cag ccc ccg agc ggc ggc ggc
gcc ggc ccg ggt ggc aac agc agc agc agc agc agc ctc cac aag ccc cag gag
tcg ccc acc ctt ccg gtg tcc acc gcc acc gac agc agc tac tac acc aac cag
cag cac ccg gcg ggc ggc ggc ggc ggc ggg ggc tcg ccc tac gcg cac atg ggt
tcc tac cag tac caa gcc agc ggc ctc aac aac gtc cct tac tcc gcc aag agc
agc tat gac ctg ggc tac acc gcc gcc tac acc tcc tac gct ccc tat gga acc
agt tcg tcc cca gcc aac aac gag cct gag aag gag gac ctt gag cct gaa att
cgg ata gtg aac ggg aag cca aag aaa gtc cgg aaa ccc cgc acc atc tac tcc
agt ttc cag ctg gcg gct ctt cag cgg cgt ttc caa aag act caa tac ttg gcc
ttg ccg gag cga gcc gag ctg gcg gcc tct ctg ggc ctc acc cag act cag gtc
aaa atc tgg ttc cag aac cgc cgg tcc aag ttc aag aag atg tgg aaa agt ggt
gag atc ccc tcg gag cag cac cct ggg gcc agc gct tct cca cct tgt gct tcg
ccg cca gtc tca gcg ccg gcc tcc tgg gac ttt ggt gtg ccg cag cgg atg gcg
ggc ggc ggt ggt ccg ggc agt ggc ggc agc ggc gcc ggc agc tcg ggc tcc agc
ccg agc agc gcg gcc tcg gct ttt ctg ggc aac tac ccc tgg tac cac cag acc
tcg gga tcc gcc tca cac ctg cag gcc acg gcg ccg ctg ctg cac ccc act cag
acc ccg cag ccg cat cac cac cac cac cat cac ggc ggc ggg ggc gcc ccg gtg
agc gcg ggg acg att ttc gct agc
TAAGTCGACGGTACCTGGAGCACAAGACTGGCCTCATGGGCCTTCCGCTCACTGCCCGCTTTCCAGTCGGGAA
ACCTGTCGTGCCAGCTGCATTAACATGGTCATAGCTGTTTCCTTGCGTATTGGGCGCTCTCCGCTTCCTCGCT
CACTGACTCGCTGCGCTCGGTCGTTCGGGTAAAGCCTGGGGTGCCTAATGAGCAAAAGGCCAGCAAAAGGCCA
GGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCG
ACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTC
GTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGC
TTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGA
ACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGAC
TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCT
TGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTAC
CTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGC
AAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTC
AGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTT
AAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTA
```

ATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGA
TAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGAACCACGCTCACCGGC
TCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCC
TCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTG
TTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACG
ATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTC
AGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCAT
CCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAG
TTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGA
AAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTG
CACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGC
CGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGC
ATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTC
CGCGCACATTTCCCCGAAAAGTGCCAC

4.1.4 **pCAGG** (6,407 bp)

```
TAGTTATTTCTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCC
ATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCA
TTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACT
ATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAA
TGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATC
TACGTATTAGTCATCGCTATTACCATGGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCC
CCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTTGTGCAGCGATGGGGGCGGGGGGGGGGG
GGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCA
ATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAG
CGCGCGGCGGGCGGGAGTCGCTGCGTTGCCTTCGCCCCGTGCCCCGCTCCGCGCCGCCTCGCGCCGCCCGCCC
CGGCTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGC
GCTTGGTTTAATGACGGCTCGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTAAAGGGCTCCGGGAGGGCCCTTT
GTGCGGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTGTGTGTGCGTGGGGAGCGCCGCGTGCGGCCCGCGCTG
CCCGGCGGCTGTGAGCGCTGCGGGCGCGGCGCGGGGCTTTGTGCGCTCCGCGTGTGCGCGAGGGGAGCGCGGC
CGGGGGCGGTGCCCCGCGGTGCGGGGGGGCTGCGAGGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGG
GGTGAGCAGGGGGTGTGGGCGCGGCGGTCGGGCTGTAACCCCCCCCTGCACCCCCCTCCCCGAGTTGCTGAGC
ACGGCCCGGCTTCGGGTGCGGGGCTCCGTGCGGGGCGTGGCGCGGGGCTCGCCGTGCCGGGCGGGGGGTGGCG
GCAGGTGGGGGTGCCGGGCGGGGCGGGGCCGCCTCGGGCCGGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCC
GGAGCGCCGGCGGCTGTCGAGGCGCGGCGAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGAGAGGGCGCA
GGGACTTCCTTTGTCCCAAATCTGGCGGAGCCGAAATCTGGGAGGCGCCGCCGCACCCCCTCTAGCGGGCGCG
GGCGAAGCGGTGCGGCGCCGGCAGGAAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCCGCCGTCCC
CTTCTCCATCTCCAGCCTCGGGGCTGCCGCAGGGGGACGGCTGCCTTCGGGGGGGACGGGGCAGGGCGGGGTT
CGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTAACCATGTTCATGCCTTCTTCTTTTTCCTACAG
CTCCTGGGCAACGTGCTGGTTGTTGTGCTGTCTCATCATTTTGGCAAAGAATTCTGCAGTCGACGGTACCGCG
GGCCCGGGATCCGCCCCTCTCCCTCCCCCCCCCCTAACGTTACTGGCCGAAGCCGCTTGGAATAAGGCCGGTG
TGCGTTTGTCTATATGTTATTTTCCACCATATTGCCGTCTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCT
GTCTTCTTGACGAGCATTCCTAGGGGTCTTTCCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGA
AGGAAGCAGTTCCTCTGGAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCTTTGCAGGCAGCGGAACCC
CCCACCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGCAAAGGCGGCACAACCC
CAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCTCCTCAAGCGTATTCAACAAGGGGC
TGAAGGATGCCCAGAAGGTACCCCATTGTATGGGATCTGATCTGGGGCCTCGGTGCACATGCTTTACATGTGT
TTAGTCGAGGTTAAAAAAACGTCTAGGCCCCCCGAACCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGA
TAATATGGCCACAACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTG
GACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGA
CCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGG
CGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGC
TACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGG
GCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAA
GCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAAC
TTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCG
GCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGA
GAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTAC
AAGTAAAGCGGCCGCGACTCTAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAA
CCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCT
TATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTT
GTGGTTTGTCCAAACTCATCAATGTATCTTAAGGCGTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGT
TAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGA
ATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAAC
GTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGG
GGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCC
GGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTC
ACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGG
GAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATA
ACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCCTGAGGCGGAAAGAACCAGCTGTGGAATGTGT
GTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTC
AGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCA
```

GCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCC
ATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTG
AGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGAT
TGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAA
CAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGA
CCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGT
TCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGG
CAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGC
ATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGAT
GGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCC
AGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCA
TGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACAT
AGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGT
ATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGG
GTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAA
AGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGT
TCTTCGCCCACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGAC
GGCAATAAAAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAGGGC
TGGCACTCTGTCGATACCCCACCGAGACCCCATTGGGGCCAATACGCCCGCGTTTCTTCCTTTTCCCCACCCC
ACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAACGTCGGGGCGGCAGGCCCTGCCATAGCCTCAGG
TTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTT
GATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCA
AAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGC
GGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATA
CCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACC
TCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAG
ACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGA
ACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGG
CGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTG
GTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGG
CGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACA
TGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT

4.1.5 **pCAGG-**DLX2 (7,622 bp)

TAGTTATTTCTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT
AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGC
CAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGC
AGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGG
CCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCT
ACGTATTAGTCATCGCTATTACCATGGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCC
CCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTTGTGC
AGCGATGGGGGCGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCG
GGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTT
CCTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGG
GAGTCGCTGCGTTGCCTTCGCCCCGTGCCCCGCTCCGCGCCGCCTCGCGCCGCCCGCCCCG
GCTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGC
TGTAATTAGCGCTTGGTTTAATGACGGCTCGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTA
AAGGGCTCCGGGAGGGCCCTTTGTGCGGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTGTG
TGTGCGTGGGGAGCGCCGCGTGCGGCCCGCGCTGCCCGGCGGCTGTGAGCGCTGCGGGCGC
GGCGCGGGGCTTTGTGCGCTCCGCGTGTGCGCGAGGGGAGCGCGGCCGGGGGCGGTGCCCC
GCGGTGCGGGGGGGCTGCGAGGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGT
GAGCAGGGGGTGTGGGCGCGGCGGTCGGGCTGTAACCCCCCCCTGCACCCCCCTCCCCGAG
TTGCTGAGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTGCGGGGCGTGGCGCGGGGCTCGC
CGTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGGCCGCCTCGGGCC
GGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCGGAGCGCCGGCGGCTGTCGAGGCGCGGC
GAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGAGAGGGCGCAGGGACTTCCTTTGTCC
CAAATCTGGCGGAGCCGAAATCTGGGAGGCGCCGCCGCACCCCCTCTAGCGGGCGCGGGCG
AAGCGGTGCGGCGCCGGCAGGAAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCC
GCCGTCCCCTTCTCCATCTCCAGCCTCGGGGCTGCCGCAGGGGGACGGCTGCCTTCGGGGG
GGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTA
ACCATGTTCATGCCTTCTTCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTGTTGTGCT
GTCTCATCATTTTGGCAAAGAATTCTGCAGTCGACGAATTCAGCGCTCTCGAGACCGGTGC
CGCCATGGGAGGATCCCAGTGTACTAATTATGCTCTCTTGAAATTGGCTGGAGATGTTGAG
AGCAACCCAGGTCCCAGATCTGAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGGAGG
AGAATCCCGGCCCTTCTAGAGCCACGAAGCAAGCAGGAGATGTTGAAGAAAACCCCGGTCC
T
gct agc ATG act gga gtc ttt gac agt cta gtg gct gat atg cac
tcg acc cag atc gcc gcc tcc agc acg tac cac cag cac cag cag
ccc ccg agc ggc ggc ggc gcc ggc ccg ggt ggc aac agc agc agc
agc agc agc ctc cac aag ccc cag gag tcg ccc acc ctt ccg gtg
tcc acc gcc acc gac agc agc tac tac acc aac cag cag cac ccg
gcg ggc ggc ggc ggc ggc ggg ggc tcg ccc tac gcg cac atg ggt
tcc tac cag tac caa gcc agc ggc ctc aac aac gtc cct tac tcc
gcc aag agc agc tat gac ctg ggc tac acc gcc gcc tac acc tcc
tac gct ccc tat gga acc agt tcg tcc cca gcc aac aac gag cct
gag aag gag gac ctt gag cct gaa att cgg ata gtg aac ggg aag
cca aag aaa gtc cgg aaa ccc cgc acc atc tac tcc agt ttc cag
ctg gcg gct ctt cag cgg cgt ttc caa aag act caa tac ttg gcc
ttg ccg gag cga gcc gag ctg gcg gcc tct ctg ggc ctc acc cag

act cag gtc aaa atc tgg ttc cag aac cgc cgg tcc aag ttc aag
aag atg tgg aaa agt ggt gag atc ccc tcg gag cag cac cct ggg
gcc agc gct tct cca cct tgt gct tcg ccg cca gtc tca gcg ccg
gcc tcc tgg gac ttt ggt gtg ccg cag cgg atg gcg ggc ggc ggt
ggt ccg ggc agt ggc ggc agc ggc gcc ggc agc tcg ggc tcc agc
ccg agc agc gcg gcc tcg gct ttt ctg ggc aac tac ccc tgg tac
cac cag acc tcg gga tcc gcc tca cac ctg cag gcc acg gcg ccg
ctg ctg cac ccc act cag acc ccg cag ccg cat cac cac cac cac
cat cac ggc ggc ggg ggc gcc ccg gtg agc gcg ggg acg att ttc
gct agc
TAAGTCGACGGTACCGCGGGCCCGGGATCCGCCCCTCTCCCTCCCCCCCCCCTAACGTTAC
TGGCCGAAGCCGCTTGGAATAAGGCCGGTGTGCGTTTGTCTATATGTTATTTTCCACCATA
TTGCCGTCTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTCTTGACGAGCATTC
CTAGGGGTCTTTCCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGC
AGTTCCTCTGGAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCTTTGCAGGCAGCGG
AACCCCCCACCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTG
CAAAGGCGGCACAACCCCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATG
GCTCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCCATTGTATG
GGATCTGATCTGGGGCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAAAC
GTCTAGGCCCCCCGAACCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAATATGG
CCACAACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCC
ACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGC
CCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACAT
GAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATC
TTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCC
TGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCA
CAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAAC
GGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCG
ACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTA
CCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTG
CTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCG
GCCGCGACTCTAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAA
AACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACT
TGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAA
AGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAAGGC
GTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCA
TTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGA
TAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAA
CGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAA
TCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCC
GATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAA
AGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCC
GCCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCG
GAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATA
ACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCCTGAGGCGGAAAGAACCAGC
TGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTAT
GCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCA

GGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTC
CGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAAT
TTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGA
GGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCG
TTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGC
TATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCT
GTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAA
CTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTG
TGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCA
GGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATG
CGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCA
TCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGA
GCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGC
GAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCC
GCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGC
GTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTG
CTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGT
TCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCA
CGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGG
ACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCTAG
GGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCA
ATAAAAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGT
CCCAGGGCTGGCACTCTGTCGATACCCCACCGAGACCCCATTGGGGCCAATACGCCCGCGT
TTCTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAAC
GTCGGGGCGGCAGGCCCTGCCATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAA
AACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAA
AATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGA
TCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGC
TACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGG
CTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCAC
TTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG
CTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAA
GGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACC
TACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGA
GAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCT
TCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAG
CGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGG
CCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATC
CCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT

4.1.6 **pCAGG-**MASH1 (7,346 bp)

```
TAGTTATTTCTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT
AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGC
CAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGC
AGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGG
CCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCT
ACGTATTAGTCATCGCTATTACCATGGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCC
CCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTTGTGC
AGCGATGGGGGCGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCG
GGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTT
CCTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGG
GAGTCGCTGCGTTGCCTTCGCCCCGTGCCCCGCTCCGCGCCGCCTCGCGCCGCCCGCCCCG
GCTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGC
TGTAATTAGCGCTTGGTTTAATGACGGCTCGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTA
AAGGGCTCCGGGAGGGCCCTTTGTGCGGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTGTG
TGTGCGTGGGGAGCGCCGCGTGCGGCCCGCGCTGCCCGGCGGCTGTGAGCGCTGCGGGCGC
GGCGCGGGGCTTTGTGCGCTCCGCGTGTGCGCGAGGGGAGCGCGGCCGGGGGCGGTGCCCC
GCGGTGCGGGGGGGCTGCGAGGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGT
GAGCAGGGGGTGTGGGCGCGGCGGTCGGGCTGTAACCCCCCCCTGCACCCCCCTCCCCGAG
TTGCTGAGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTGCGGGGCGTGGCGCGGGGCTCGC
CGTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGGCCGCCTCGGGCC
GGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCGGAGCGCCGGCGGCTGTCGAGGCGCGGC
GAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGAGAGGGCGCAGGGACTTCCTTTGTCC
CAAATCTGGCGGAGCCGAAATCTGGGAGGCGCCGCCGCACCCCCTCTAGCGGGCGCGGGCG
AAGCGGTGCGGCGCCGGCAGGAAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCC
GCCGTCCCCTTCTCCATCTCCAGCCTCGGGGCTGCCGCAGGGGGACGGCTGCCTTCGGGGG
GGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTA
ACCATGTTCATGCCTTCTTCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTGTTGTGCT
GTCTCATCATTTTGGCAAAGAATTCTGCAGTCGACGAATTCAGCGCTCTCGAGACCGGTGC
CGCCATGGGA
gga tcc ATG gaa agc tct gcc aag atg gag agc ggc ggc gcc ggc
cag cag ccc cag ccg cag ccc cag cag ccc ttc ctg ccg ccc gca
gcc tgt ttc ttt gcc acg gcc gca gcc gcg gcg gcc gca gcc gcc
gca gcg gca gcg cag agc gcg cag cag cag cag cag cag cag cag
cag cag cag cag gcg ccg cag ctg aga ccg gcg gcc gac ggc cag
ccc tca ggg ggc ggt cac aag tca gcg ccc aag caa gtc aag cga
cag cgc tcg tct tcg ccc gaa ctg atg cgc tgc aaa cgc cgg ctc
aac ttc agc ggc ttt ggc tac agc ctg ccg cag cag cag ccg gcc
gcc gtg gcg cgc cgc aac gag cgc gag cgc aac cgc gtc aag ttg
gtc aac ctg ggc ttt gcc acc ctt cgg gag cac gtc ccc aac ggc
gcg gcc aac aag aag atg agt aag gtg gag aca ctg cgc tcg gcg
gtc gag tac atc cgc gcg ctg cag cag ctg ctg gac gag cat gac
gcg gtg agc gcc gcc ttc cag gca ggc gtc ctg tcg ccc acc atc
tcc ccc aac tac tcc aac gac ttg aac tcc atg gcc ggc tcg ccg
gtc tca tcc tac tcg tcg gac gag ggc tct tac gac ccg ctc agc
ccc gag gag cag gag ctt ctc gac ttc acc aac tgg ttc gga tcc
```

CAGTGTACTAATTATGCTCTCTTGAAATTGGCTGGAGATGTTGAGAGCAACCCAGGTCCCA
GATCTGAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGGAGGAGAATCCCGGCCCTTC
TAGAGCCACGAAGCAAGCAGGAGATGTTGAAGAAAACCCCGGTCCTGCTAGCTAAGTCGAC
GGTACCGCGGGCCCGGGATCCGCCCCTCTCCCTCCCCCCCCCCTAACGTTACTGGCCGAAG
CCGCTTGGAATAAGGCCGGTGTGCGTTTGTCTATATGTTATTTTCCACCATATTGCCGTCT
TTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTCTTGACGAGCATTCCTAGGGGTC
TTTCCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTTCCTCT
GGAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCTTTGCAGGCAGCGGAACCCCCCA
CCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGCAAAGGCGG
CACAACCCCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCTCCTC
AAGCGTATTCAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCCATTGTATGGGATCTGAT
CTGGGGCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAAACGTCTAGGCC
CCCCGAACCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAATATGGCCACAACCA
TGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGG
CGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGC
AAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCG
TGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCA
CGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAG
GACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACC
GCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGA
GTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAG
GTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACC
AGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTC
GTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCGCGACT
CTAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCA
CACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTG
CAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTT
TTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAAGGCGTAAATTGT
AAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAAC
CAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGA
GTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGG
GCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTT
TTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAG
CTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGG
CGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTT
AATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTA
TTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATA
AATGCTTCAATAATATTGAAAAAGGAAGAGTCCTGAGGCGGAAAGAACCAGCTGTGGAATG
TGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCAT
GCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGT
ATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCC
CGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTAT
TTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTT
TTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATG
ATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCT
ATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCA
GGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGAC

GAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACG
TTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCT
GTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTG
CATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAG
CACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGG
GCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTC
GTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTG
GATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTAC
CCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGT
ATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAG
CGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTC
GATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCT
GGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCTAGGGGGAGGCT
AACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCAATAAAAAGA
CAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAGGGCT
GGCACTCTGTCGATACCCCACCGAGACCCCATTGGGGCCAATACGCCCGCGTTTCTTCCTT
TTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAACGTCGGGGCG
GCAGGCCCTGCCATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAAACTTCATT
TTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTA
ACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGA
GATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGG
TGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAG
AGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAAC
TCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTG
GCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCG
GTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAA
CTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGG
ACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGG
AAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTT
TTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTAC
GGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTC
TGTGGATAACCGTATTACCGCCATGCAT

4.1.7 pCAGG-DLX2/MASH1 (8,336 bp)

TAGTTATTTCTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT
AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGC
CAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGC
AGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGG
CCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCT
ACGTATTAGTCATCGCTATTACCATGGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCC
CCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTTGTGC
AGCGATGGGGGCGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCG
GGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTT
CCTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGG
GAGTCGCTGCGTTGCCTTCGCCCCGTGCCCCGCTCCGCGCCGCCTCGCGCCGCCCGCCCCG
GCTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGC
TGTAATTAGCGCTTGGTTTAATGACGGCTCGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTA
AAGGGCTCCGGGAGGGCCCTTTGTGCGGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTGTG
TGTGCGTGGGGAGCGCCGCGTGCGGCCCGCGCTGCCCGGCGGCTGTGAGCGCTGCGGGCGC
GGCGCGGGGCTTTGTGCGCTCCGCGTGTGCGCGAGGGGAGCGCGGCCGGGGGCGGTGCCCC
GCGGTGCGGGGGGGCTGCGAGGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGT
GAGCAGGGGGTGTGGGCGCGGCGGTCGGGCTGTAACCCCCCCCTGCACCCCCCTCCCCGAG
TTGCTGAGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTGCGGGGCGTGGCGCGGGGCTCGC
CGTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGGCCGCCTCGGGCC
GGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCGGAGCGCCGGCGGCTGTCGAGGCGCGGC
GAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGAGAGGGCGCAGGGACTTCCTTTGTCC
CAAATCTGGCGGAGCCGAAATCTGGGAGGCGCCGCCGCACCCCCTCTAGCGGGCGCGGGCG
AAGCGGTGCGGCGCCGGCAGGAAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCC
GCCGTCCCCTTCTCCATCTCCAGCCTCGGGGCTGCCGCAGGGGGACGGCTGCCTTCGGGGG
GGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTA
ACCATGTTCATGCCTTCTTCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTGTTGTGCT
GTCTCATCATTTTGGCAAAGAATTCTGCAGTCGACGAATTCAGCGCTCTCGAGACCGGTGC
CGCCATGGGA
gga tcc ATG gaa agc tct gcc aag atg gag agc ggc ggc gcc ggc
cag cag ccc cag ccg cag ccc cag cag ccc ttc ctg ccg ccc gca
gcc tgt ttc ttt gcc acg gcc gca gcc gcg gcg gcc gca gcc gcc
gca gcg gca gcg cag agc gcg cag cag cag cag cag cag cag cag
cag cag cag cag gcg ccg cag ctg aga ccg gcg gcc gac ggc cag
ccc tca ggg ggc ggt cac aag tca gcg ccc aag caa gtc aag cga
cag cgc tcg tct tcg ccc gaa ctg atg cgc tgc aaa cgc cgg ctc
aac ttc agc ggc ttt ggc tac agc ctg ccg cag cag cag ccg gcc
gcc gtg gcg cgc cgc aac gag cgc gag cgc aac cgc gtc aag ttg
gtc aac ctg ggc ttt gcc acc ctt cgg gag cac gtc ccc aac ggc
gcg gcc aac aag aag atg agt aag gtg gag aca ctg cgc tcg gcg
gtc gag tac atc cgc gcg ctg cag cag ctg ctg gac gag cat gac
gcg gtg agc gcc gcc ttc cag gca ggc gtc ctg tcg ccc acc atc
tcc ccc aac tac tcc aac gac ttg aac tcc atg gcc ggc tcg ccg
gtc tca tcc tac tcg tcg gac gag ggc tct tac gac ccg ctc agc
ccc gag gag cag gag ctt ctc gac ttc acc aac tgg ttc gga tcc

CAGTGTACTAATTATGCTCTCTTGAAATTGGCTGGAGATGTTGAGAGCAACCCAGGTCCCA
GATCTGAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGGAGGAGAATCCCGGCCCTTC
TAGAGCCACGAAGCAAGCAGGAGATGTTGAAGAAAACCCCGGTCCT
gct agc ATG act gga gtc ttt gac agt cta gtg gct gat atg cac
tcg acc cag atc gcc gcc tcc agc acg tac cac cag cac cag cag
ccc ccg agc ggc ggc ggc gcc ggc ccg ggt ggc aac agc agc agc
agc agc agc ctc cac aag ccc cag gag tcg ccc acc ctt ccg gtg
tcc acc gcc acc gac agc agc tac tac acc aac cag cag cac ccg
gcg ggc ggc ggc ggc ggc ggg ggc tcg ccc tac gcg cac atg ggt
tcc tac cag tac caa gcc agc ggc ctc aac aac gtc cct tac tcc
gcc aag agc agc tat gac ctg ggc tac acc gcc gcc tac acc tcc
tac gct ccc tat gga acc agt tcg tcc cca gcc aac aac gag cct
gag aag gag gac ctt gag cct gaa att cgg ata gtg aac ggg aag
cca aag aaa gtc cgg aaa ccc cgc acc atc tac tcc agt ttc cag
ctg gcg gct ctt cag cgg cgt ttc caa aag act caa tac ttg gcc
ttg ccg gag cga gcc gag ctg gcg gcc tct ctg ggc ctc acc cag
act cag gtc aaa atc tgg ttc cag aac cgc cgg tcc aag ttc aag
aag atg tgg aaa agt ggt gag atc ccc tcg gag cag cac cct ggg
gcc agc gct tct cca cct tgt gct tcg ccg cca gtc tca gcg ccg
gcc tcc tgg gac ttt ggt gtg ccg cag cgg atg gcg ggc ggc ggt
ggt ccg ggc agt ggc ggc agc ggc gcc ggc agc tcg ggc tcc agc
ccg agc agc gcg gcc tcg gct ttt ctg ggc aac tac ccc tgg tac
cac cag acc tcg gga tcc gcc tca cac ctg cag gcc acg gcg ccg
ctg ctg cac ccc act cag acc ccg cag ccg cat cac cac cac cac
cat cac ggc ggc ggg ggc gcc ccg gtg agc gcg ggg acg att ttc
gct agc
TAAGTCGACGGTACCGCGGGCCCGGGATCCGCCCCTCTCCCTCCCCCCCCCCTAACGTTAC
TGGCCGAAGCCGCTTGGAATAAGGCCGGTGTGCGTTTGTCTATATGTTATTTTCCACCATA
TTGCCGTCTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTCTTGACGAGCATTC
CTAGGGGTCTTTCCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGC
AGTTCCTCTGGAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCTTTGCAGGCAGCGG
AACCCCCCACCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTG
CAAAGGCGGCACAACCCCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATG
GCTCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCCATTGTATG
GGATCTGATCTGGGGCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAAAC
GTCTAGGCCCCCCGAACCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAATATGG
CCACAACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCC
ACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGC
CCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACAT
GAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATC
TTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCC
TGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCA
CAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAAC
GGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCG
ACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTA
CCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTG
CTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCG

GCCGCGACTCTAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAA
AACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACT
TGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAA
AGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAAGGC
GTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCA
TTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGA
TAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAA
CGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAA
TCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCC
GATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAA
AGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCC
GCCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCG
GAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATA
ACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCCTGAGGCGGAAAGAACCAGC
TGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTAT
GCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCA
GGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTC
CGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAAT
TTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGA
GGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCG
TTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGC
TATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCT
GTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAA
CTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTG
TGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCA
GGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATG
CGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCA
TCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGA
GCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGC
GAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCC
GCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGC
GTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTG
CTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGT
TCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCA
CGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGG
ACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCTAG
GGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCA
ATAAAAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGT
CCCAGGGCTGGCACTCTGTCGATACCCCACCGAGACCCCATTGGGGCCAATACGCCCGCGT
TTCTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAAC
GTCGGGGCGGCAGGCCCTGCCATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAA
AACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAA
AATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGA
TCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGC
TACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGG
CTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCAC
TTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG
CTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAA

```
GGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACC
TACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGA
GAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCT
TCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAG
CGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGG
CCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATC
CCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT
```

4.1.8 pCAGG-DLX2/GSX2 (8,540 bp)

TAGTTATTTCTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT
AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGC
CAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGC
AGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGG
CCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCT
ACGTATTAGTCATCGCTATTACCATGGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCC
CCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTTGTGC
AGCGATGGGGGCGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCG
GGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTT
CCTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGG
GAGTCGCTGCGTTGCCTTCGCCCCGTGCCCCGCTCCGCGCCGCCTCGCGCCGCCCGCCCCG
GCTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGC
TGTAATTAGCGCTTGGTTTAATGACGGCTCGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTA
AAGGGCTCCGGGAGGGCCCTTTGTGCGGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTGTG
TGTGCGTGGGGAGCGCCGCGTGCGGCCCGCGCTGCCCGGCGGCTGTGAGCGCTGCGGGCGC
GGCGCGGGGCTTTGTGCGCTCCGCGTGTGCGCGAGGGGAGCGCGGCCGGGGGCGGTGCCCC
GCGGTGCGGGGGGGCTGCGAGGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGT
GAGCAGGGGGTGTGGGCGCGGCGGTCGGGCTGTAACCCCCCCCTGCACCCCCCTCCCCGAG
TTGCTGAGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTGCGGGGCGTGGCGCGGGGCTCGC
CGTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGGCCGCCTCGGGCC
GGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCGGAGCGCCGGCGGCTGTCGAGGCGCGGC
GAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGAGAGGGCGCAGGGACTTCCTTTGTCC
CAAATCTGGCGGAGCCGAAATCTGGGAGGCGCCGCCGCACCCCCTCTAGCGGGCGCGGGCG
AAGCGGTGCGGCGCCGGCAGGAAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCC
GCCGTCCCCTTCTCCATCTCCAGCCTCGGGGCTGCCGCAGGGGGACGGCTGCCTTCGGGGG
GGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTA
ACCATGTTCATGCCTTCTTCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTGTTGTGCT
GTCTCATCATTTTGGCAAAGAATTCTGCAGTCGACGAATTCAGCGCTCTCGAGACCGGTGC
CGCCATGGGAGGATCCCAGTGTACTAATTATGCTCTCTTGAAATTGGCTGGAGATGTTGAG
AGCAACCCAGGTCCC
aga tct ATG tcg cgc tcc ttc tat gtc gac tcg ctc atc atc aag
gac acc tca cgg cct gcg ccc tcg ctg cct gaa ccg cac ccc ggg
ccg gat ttc ttc atc ccg ctt ggc atg ccg ccc cca ttg gtg atg
tcc gtg tcc ggc ccc ggc tgc ccg tcc cgc aag agc ggc gcg ttc
tgc gtg tgc cct ctc tgc gtc act tcg cac ctg cac tcc tct cgg
ggg tct gtg ggc gcc ggc agc ggg ggc gca ggg gcc ggg gtt acc
ggg gcc gga ggc agt ggg gtg gca ggg gcc gca ggg gca ctg cct
ctg ctt aag agc cag ttc tct tcg gct cct ggg gac gcg cag ttt
tgc ccg cgg gtg aac cat gcg cat cat cac cac cac ccg ccg cag
cac cac cat cac cat cat cag ccc cag cag cct ggc tcg gcc gcg
gcg gcg gca gca gca gca gcg gcg gcg gcg gcc gcg gcg gcc ttg
ggg cac ccg cag cac cac gca cct gtc tgc acc gcc acc acc tac
aac gtg gcg gac ccg cgg aga ttc cac tgc ctc acc atg gga ggc
tct gac gcc agc cag gta ccc aat ggc aag agg atg agg acg gcg
ttc act agc acg caa ctc ctg gag ctg gag aga gaa ttc tct tcc

aac atg tac ctg tct cga ctc cgg agg att gaa atc gcc act tac
ctg aac ctg tcg gag aag cag gtg aaa atc tgg ttt cag aac cgc
cga gtg aag cac aag aag gag ggg aag ggc acg cag agg aac agt
cac gcg ggc tgc aag tgc gtc ggg agc cag gtg cac tac gcg cgc
tcc gag gat gag gac tcc ctg tcg ccg gcc tca gcc aac gat gac
aag gag att tcc ccc tta aga tct
GAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGGAGGAGAATCCCGGCCCTTCTAGAG
CCACGAAGCAAGCAGGAGATGTTGAAGAAAACCCCGGTCCT
gct agc ATG act gga gtc ttt gac agt cta gtg gct gat atg cac
tcg acc cag atc gcc gcc tcc agc acg tac cac cag cac cag cag
ccc ccg agc ggc ggc ggc gcc ggc ccg ggt ggc aac agc agc agc
agc agc agc ctc cac aag ccc cag gag tcg ccc acc ctt ccg gtg
tcc acc gcc acc gac agc agc tac tac acc aac cag cag cac ccg
gcg ggc ggc ggc ggc ggc ggg ggc tcg ccc tac gcg cac atg ggt
tcc tac cag tac caa gcc agc ggc ctc aac aac gtc cct tac tcc
gcc aag agc agc tat gac ctg ggc tac acc gcc gcc tac acc tcc
tac gct ccc tat gga acc agt tcg tcc cca gcc aac aac gag cct
gag aag gag gac ctt gag cct gaa att cgg ata gtg aac ggg aag
cca aag aaa gtc cgg aaa ccc cgc acc atc tac tcc agt ttc cag
ctg gcg gct ctt cag cgg cgt ttc caa aag act caa tac ttg gcc
ttg ccg gag cga gcc gag ctg gcg gcc tct ctg ggc ctc acc cag
act cag gtc aaa atc tgg ttc cag aac cgc cgg tcc aag ttc aag
aag atg tgg aaa agt ggt gag atc ccc tcg gag cag cac cct ggg
gcc agc gct tct cca cct tgt gct tcg ccg cca gtc tca gcg ccg
gcc tcc tgg gac ttt ggt gtg ccg cag cgg atg gcg ggc ggc ggt
ggt ccg ggc agt ggc ggc agc ggc gcc ggc agc tcg ggc tcc agc
ccg agc agc gcg gcc tcg gct ttt ctg ggc aac tac ccc tgg tac
cac cag acc tcg gga tcc gcc tca cac ctg cag gcc acg gcg ccg
ctg ctg cac ccc act cag acc ccg cag ccg cat cac cac cac cac
cat cac ggc ggc ggg ggc gcc ccg gtg agc gcg ggg acg att ttc
gct agc
TAAGTCGACGGTACCGCGGGCCCGGGATCCGCCCCTCTCCCTCCCCCCCCCCTAACGTTAC
TGGCCGAAGCCGCTTGGAATAAGGCCGGTGTGCGTTTGTCTATATGTTATTTTCCACCATA
TTGCCGTCTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTCTTGACGAGCATTC
CTAGGGGTCTTTCCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGC
AGTTCCTCTGGAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCTTTGCAGGCAGCGG
AACCCCCCACCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTG
CAAAGGCGGCACAACCCCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATG
GCTCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCCATTGTATG
GGATCTGATCTGGGGCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAAAC
GTCTAGGCCCCCCGAACCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAATATGG
CCACAACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCC
ACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGC
CCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACAT
GAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATC
TTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCC
TGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCA

CAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAAC
GGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCG
ACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTA
CCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTG
CTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCG
GCCGCGACTCTAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAA
AACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACT
TGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAA
AGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAAGGC
GTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCA
TTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGA
TAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAA
CGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAA
TCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCC
GATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAA
AGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCC
GCCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCG
GAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATA
ACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCCTGAGGCGGAAAGAACCAGC
TGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTAT
GCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCA
GGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTC
CGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAAT
TTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGA
GGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCG
TTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGC
TATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCT
GTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAA
CTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTG
TGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCA
GGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATG
CGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCA
TCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGA
GCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGC
GAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCC
GCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGC
GTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTG
CTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGT
TCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCA
CGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGG
ACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCTAG
GGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCA
ATAAAAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGT
CCCAGGGCTGGCACTCTGTCGATACCCCACCGAGACCCCATTGGGGCCAATACGCCCGCGT
TTCTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAAC
GTCGGGGCGGCAGGCCCTGCCATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAA
AACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAA
AATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGA

TCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGC
TACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGG
CTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCAC
TTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTG
CTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAA
GGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACC
TACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGA
GAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCT
TCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAG
CGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGG
CCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATC
CCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT

4.1.9 **pCAGG-**DLX2/MASH1/GSX2 (9,254 bp)

TAGTTATTTCTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT
AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGC
CAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGC
AGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGG
CCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCT
ACGTATTAGTCATCGCTATTACCATGGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCC
CCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTTGTGC
AGCGATGGGGGCGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCG
GGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTT
CCTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGG
GAGTCGCTGCGTTGCCTTCGCCCCGTGCCCCGCTCCGCGCCGCCTCGCGCCGCCCGCCCCG
GCTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGC
TGTAATTAGCGCTTGGTTTAATGACGGCTCGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTA
AAGGGCTCCGGGAGGGCCCTTTGTGCGGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTGTG
TGTGCGTGGGGAGCGCCGCGTGCGGCCCGCGCTGCCCGGCGGCTGTGAGCGCTGCGGGCGC
GGCGCGGGGCTTTGTGCGCTCCGCGTGTGCGCGAGGGGAGCGCGGCCGGGGGCGGTGCCCC
GCGGTGCGGGGGGGCTGCGAGGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGT
GAGCAGGGGGTGTGGGCGCGGCGGTCGGGCTGTAACCCCCCCCTGCACCCCCCTCCCCGAG
TTGCTGAGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTGCGGGGCGTGGCGCGGGGCTCGC
CGTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGGCCGCCTCGGGCC
GGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCGGAGCGCCGGCGGCTGTCGAGGCGCGGC
GAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGAGAGGGCGCAGGGACTTCCTTTGTCC
CAAATCTGGCGGAGCCGAAATCTGGGAGGCGCCGCCGCACCCCCTCTAGCGGGCGCGGGCG
AAGCGGTGCGGCGCCGGCAGGAAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCC
GCCGTCCCCTTCTCCATCTCCAGCCTCGGGGCTGCCGCAGGGGGACGGCTGCCTTCGGGGG
GGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTA
ACCATGTTCATGCCTTCTTCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTGTTGTGCT
GTCTCATCATTTTGGCAAAGAATTCTGCAGTCGACGAATTCAGCGCTCTCGAGACCGGTGC
CGCCATGGGA
gga tcc ATG gaa agc tct gcc aag atg gag agc ggc ggc gcc ggc
cag cag ccc cag ccg cag ccc cag cag ccc ttc ctg ccg ccc gca
gcc tgt ttc ttt gcc acg gcc gca gcc gcg gcg gcc gca gcc gcc
gca gcg gca gcg cag agc gcg cag cag cag cag cag cag cag cag
cag cag cag cag gcg ccg cag ctg aga ccg gcg gcc gac ggc cag
ccc tca ggg ggc ggt cac aag tca gcg ccc aag caa gtc aag cga
cag cgc tcg tct tcg ccc gaa ctg atg cgc tgc aaa cgc cgg ctc
aac ttc agc ggc ttt ggc tac agc ctg ccg cag cag cag ccg gcc
gcc gtg gcg cgc cgc aac gag cgc gag cgc aac cgc gtc aag ttg
gtc aac ctg ggc ttt gcc acc ctt cgg gag cac gtc ccc aac ggc
gcg gcc aac aag aag atg agt aag gtg gag aca ctg cgc tcg gcg
gtc gag tac atc cgc gcg ctg cag cag ctg ctg gac gag cat gac
gcg gtg agc gcc gcc ttc cag gca ggc gtc ctg tcg ccc acc atc
tcc ccc aac tac tcc aac gac ttg aac tcc atg gcc ggc tcg ccg
gtc tca tcc tac tcg tcg gac gag ggc tct tac gac ccg ctc agc
ccc gag gag cag gag ctt ctc gac ttc acc aac tgg ttc gga tcc

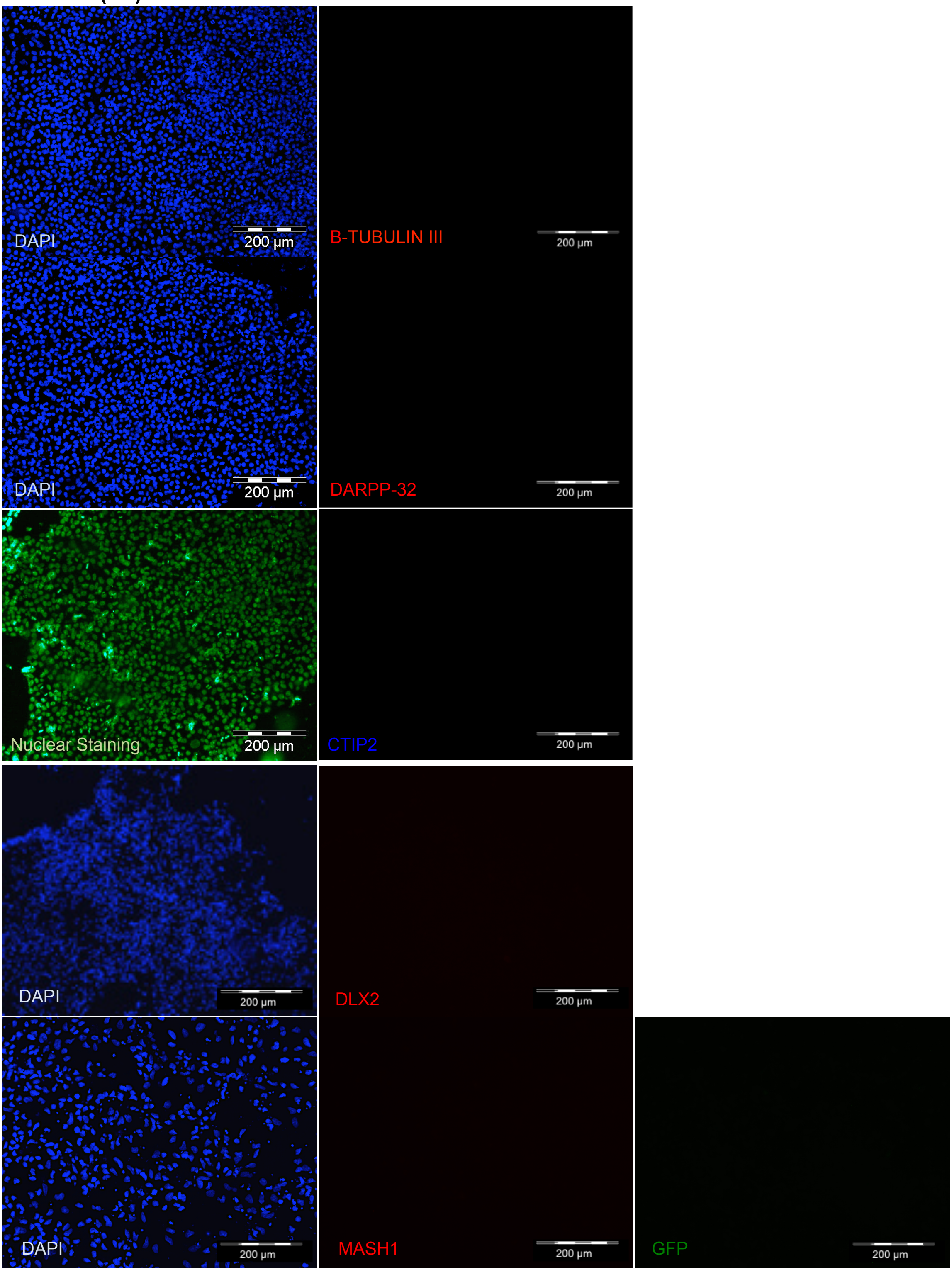
CAGTGTACTAATTATGCTCTCTTGAAATTGGCTGGAGATGTTGAGAGCAACCCAGGTCCC
aga tct ATG tcg cgc tcc ttc tat gtc gac tcg ctc atc atc aag
gac acc tca cgg cct gcg ccc tcg ctg cct gaa ccg cac ccc ggg
ccg gat ttc ttc atc ccg ctt ggc atg ccg ccc cca ttg gtg atg
tcc gtg tcc ggc ccc ggc tgc ccg tcc cgc aag agc ggc gcg ttc
tgc gtg tgc cct ctc tgc gtc act tcg cac ctg cac tcc tct cgg
ggg tct gtg ggc gcc ggc agc ggg ggc gca ggg gcc ggg gtt acc
ggg gcc gga ggc agt ggg gtg gca ggg gcc gca ggg gca ctg cct
ctg ctt aag agc cag ttc tct tcg gct cct ggg gac gcg cag ttt
tgc ccg cgg gtg aac cat gcg cat cat cac cac cac ccg ccg cag
cac cac cat cac cat cat cag ccc cag cag cct ggc tcg gcc gcg
gcg gcg gca gca gca gca gcg gcg gcg gcg gcc gcg gcg gcc ttg
ggg cac ccg cag cac cac gca cct gtc tgc acc gcc acc acc tac
aac gtg gcg gac ccg cgg aga ttc cac tgc ctc acc atg gga ggc
tct gac gcc agc cag gta ccc aat ggc aag agg atg agg acg gcg
ttc act agc acg caa ctc ctg gag ctg gag aga gaa ttc tct tcc
aac atg tac ctg tct cga ctc cgg agg att gaa atc gcc act tac
ctg aac ctg tcg gag aag cag gtg aaa atc tgg ttt cag aac cgc
cga gtg aag cac aag aag gag ggg aag ggc acg cag agg aac agt
cac gcg ggc tgc aag tgc gtc ggg agc cag gtg cac tac gcg cgc
tcc gag gat gag gac tcc ctg tcg ccg gcc tca gcc aac gat gac
aag gag att tcc ccc tta aga tct
GAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGGAGGAGAATCCCGGCCCTTCTAGAG
CCACGAAGCAAGCAGGAGATGTTGAAGAAAACCCCGGTCCT
gct agc atg act gga gtc ttt gac agt cta gtg gct gat atg cac tcg acc cag
atc gcc gcc tcc agc acg tac cac cag cac cag cag ccc ccg agc ggc ggc ggc
gcc ggc ccg ggt ggc aac agc agc agc agc agc agc ctc cac aag ccc cag gag
tcg ccc acc ctt ccg gtg tcc acc gcc acc gac agc agc tac tac acc aac cag
cag cac ccg gcg ggc ggc ggc ggc ggc ggg ggc tcg ccc tac gcg cac atg ggt
tcc tac cag tac caa gcc agc ggc ctc aac aac gtc cct tac tcc gcc aag agc
agc tat gac ctg ggc tac acc gcc gcc tac acc tcc tac gct ccc tat gga acc
agt tcg tcc cca gcc aac aac gag cct gag aag gag gac ctt gag cct gaa att
cgg ata gtg aac ggg aag cca aag aaa gtc cgg aaa ccc cgc acc atc tac tcc
agt ttc cag ctg gcg gct ctt cag cgg cgt ttc caa aag act caa tac ttg gcc
ttg ccg gag cga gcc gag ctg gcg gcc tct ctg ggc ctc acc cag act cag gtc
aaa atc tgg ttc cag aac cgc cgg tcc aag ttc aag aag atg tgg aaa agt ggt
gag atc ccc tcg gag cag cac cct ggg gcc agc gct tct cca cct tgt gct tcg
ccg cca gtc tca gcg ccg gcc tcc tgg gac ttt ggt gtg ccg cag cgg atg gcg
ggc ggc ggt ggt ccg ggc agt ggc ggc agc ggc gcc ggc agc tcg ggc tcc agc
ccg agc agc gcg gcc tcg gct ttt ctg ggc aac tac ccc tgg tac cac cag acc
tcg gga tcc gcc tca cac ctg cag gcc acg gcg ccg ctg ctg cac ccc act cag
acc ccg cag ccg cat cac cac cac cac cat cac ggc ggc ggg ggc gcc ccg gtg
agc gcg ggg acg att ttc gct agc
TAAGTCGACGGTACCGCGGGCCCGGGATCCGCCCCTCTCCCTCCCCCCCCCCTAACGTTACTGGCCGAAGCCG
CTTGGAATAAGGCCGGTGTGCGTTTGTCTATATGTTATTTTCCACCATATTGCCGTCTTTTGGCAATGTGAGG
GCCCGGAAACCTGGCCCTGTCTTCTTGACGAGCATTCCTAGGGGTCTTTCCCCTCTCGCCAAAGGAATGCAAG
GTCTGTTGAATGTCGTGAAGGAAGCAGTTCCTCTGGAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCT
TTGCAGGCAGCGGAACCCCCCACCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCT
GCAAAGGCGGCACAACCCCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCTCCTCAA
GCGTATTCAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCCATTGTATGGGATCTGATCTGGGGCCTCGGTG
CACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAAACGTCTAGGCCCCCCGAACCACGGGGACGTGGTTTTC
CTTTGAAAAACACGATGATAATATGGCCACAACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTG
CCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG

CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGT
GACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAG
TCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCG
CCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGG
CAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAG
AACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACC
AGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCT
GAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTC
GGCATGGACGAGCTGTACAAGTAAAGCGGCCGCGACTCTAGATCATAATCAGCCATACCACATTTGTAGAGGT
TTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTT
AACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTT
TTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAAGGCGTAAATTGTAAGCGTTAATA
TTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAAT
CCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTA
AAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCAC
CCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAG
AGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCG
CTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGT
CAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTA
TCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCCTGAGGCGGAAAGA
ACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAG
CATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGC
ATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTT
CCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGA
GCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATG
AGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTC
GGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCC
CGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTG
GCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTA
TTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTG
ATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGA
GCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCG
CCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATG
CCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGC
GGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGC
TTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCT
TCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCC
ACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCG
GGGATCTCATGCTGGAGTTCTTCGCCCACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGA
AGGAACCCGCGCTATGACGGCAATAAAAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGC
GGGGTTCGGTCCCAGGGCTGGCACTCTGTCGATACCCCACCGAGACCCCATTGGGGCCAATACGCCCGCGTTT
CTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAACGTCGGGGCGGCAGG
CCCTGCCATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATC
TAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAG
ACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAA
AAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGC
TTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTG
TAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCT
TACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACA
CAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGC
TTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCT
TCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTG
TGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTT
GCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCA
T

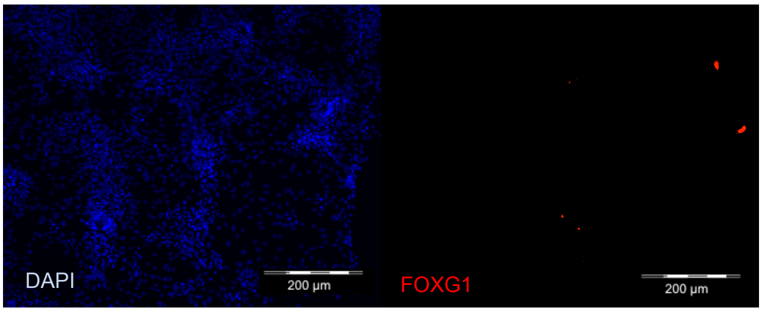
5.1 Negative and background control for immunocytochemistry

A) Negative control:

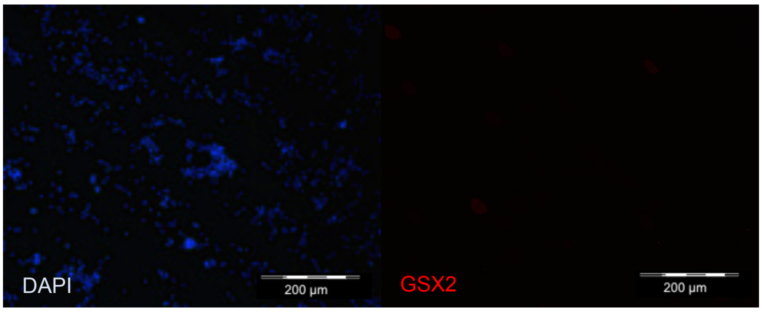
hESCs (H9):



hESC derived nrNPCs at PdD10:



Primary cells: Mlidbrain



B) Background control (secondary antibodies only):