Investigating the role of HMGN2 in the self-renewal and neuronal differentiation of ECCs using the CRISPR-Cas9 knockout system

By

Abdulmajeed Abdulghani A. Sindi
MSc.

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Institute of Cancer Sciences
College of Medical, Veterinary and Life Science
University of Glasgow
G12 8QQ

February 2017

Abstract

High mobility group N (HMGN) proteins are highly dynamic nuclear proteins that bind to nucleosomes, modulate chromatin structure and regulate transcription, replication and DNA repair. *Hmgn2* mRNA is highly expressed in zones of neurogenesis in the embryonic and adult brain, but its role in these areas is unclear. In order to investigate the role of HMGN2 in stem cell self-renewal and neuronal differentiation, we used the paired nickase CRISPR-Cas9 system to specifically mutate or knockout the *Hmgn2* gene in P19 embryonal carcinoma cells. Loss of HMGN2 did not affect the proliferation rate or cell cycle distribution in undifferentiated P19 cells. However, an increased rate of spontaneous neuronal differentiation was observed in HMGN2 knockout cells under conditions that normally promote self-renewal. Consistent with this, gene expression analysis revealed that several pro-neural and neuronal genes were upregulated in the HMGN2 knockout cells. Loss of HMGN2 was associated with changes in the profiles of several histone modifications at specific genomic loci. Highly efficient neuronal differentiation of P19 cells was demonstrated using an adherent cell induction protocol, and it was shown that the production of HMGN2 protein increased during the differentiation process. HMGN2 knockout cells showed greater induction of several pro-neural and neuronal-specific genes during the early stages of neuronal differentiation. These results suggest that HMGN2 plays an important role in maintaining self-renewal capacity and regulating neural induction in embryonal carcinoma cells.

Table of Contents

Table of Contents

List of Tables

List of Figures

Acknowledgement

*Firstly, I am grateful to The **Almighty God** for establishing me to complete this project.*

*I am extremely grateful to my principle supervisor **Dr. Katherine L. West** for her support and guidance. The thesis would not have been possible without her help, support, patience and unsurpassed knowledge about all aspects of my research. Her timely advice, meticulous scrutiny, scholarly advice and scientific approach have assisted me to accomplish the project. I would like to extend my sincere thank to my co-supervisor **Dr. Adam West** who provides me invaluable contribution and experimental advices to the work presented in this thesis. He also provided me an access to all his lab facilities*.

*I also thank my advisors Dr. **Helen Wheaden** and Dr. **Adam West** for providing a critical review of my work and advice on further experiment and support during the mini-viva.*

*Immeasurable thank extend to all members in Katherine West group, Dr. **Ohoud Rehbini**, **Sylvia Garza-Manero**, **Banna**r and **John Kennedy** (pervious lab technician), as well as members in Adam West group, Dr. **Sabarinadh Chilaka**, Dr. **Narendra Kumar**, Dr. **Ileana Guerrini**, Dr. **Aliya Al-Hosni**, **Alejandro Rivera**, **Valentine Jeantet**, **Laura Richmond** and **Mauro Berta** for their supports, encouragements, suggestions, motivations, valuable and intensive discussions, suggestions and most importantly being good and special friends. **They** helped me a lot in most of the experimental issues. I also thank Dr. **Torsten Stein** (University of Glasgow, Institute of Cancer Sciences) and Dr. **Andrew Hamilton** (University of Glasgow, School of Medicine, Dentistry and Nursing) and their previous and current members, especially Dr. **Claire Cairnery**, Dr. **Ayman Ibrahim** and **Sarah El-Messeiry**. A thank also for Tim Harvey and Tom Gilbey (FACS Facility, Institute of Cancer Sciences) who helped me with cell sorting.*

*I would like to express my gratitude toward my parents, **Abdulghan**i and **Sana**, and my sisters and brothers (**Asma**, **Esra**, **Alla**, **Mohammed**, **Ahmad**, **Aknan** and **Amjad**) for the encouragement, love and endless support during my life and this project. I owe a deep sense of gratitude to my beloved wife, **Bayan**, for her constant encouragement, support and love, as well as for great patience at all times. She has assisted me a lot in completing this project. A grateful thank to my daughter, **Raneem** who served as my inspiration to pursue this undertaking. Thanks to all **my friends** here in Glasgow or Saudi Arabia for their support. I have to admit that the achievement of completing this project is only attributable to **Allah** and then the support from all people mentioned here.*

*Lastly, I would like to acknowledge **Albaha University** (Faculty of applied medical sciences) and Saudi Arabian Cultural Bureau in London for supporting me financially to complete this project.*

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institutions.

Abdulmajeed Sindi

Definitions/Abbreviations

Abbreviation	Definition
3'	Three prime end
5'	Five prime end
5caC	5-carboxlcytosine
5fC	5-forylcytocine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
A	Adenine
aa	Amino acid
ADMP	Anti-dorsalising morphogenetic factor
APC	Adenomatous polyposis coli
APC gene	Gene encode adenomatous polyposis coli protein. This gene is classified as a tumour suppressor gene
AraC	Cytosine ß-D-arabinofuranoside hydrochloride
BER	Base excision repair
bHLH	Basic helix-loop-helix
BL2-A or Bl2-W	Pulse area or width
BMP	Bone morphogenic protein
Bp	Base pair
BrdU	bromo-deoxyuridine
BRG1	Brhama—related gene1
BRM	Brahma
BSA	Bovine serum albumin
C	Cytocine
C	control
C-terminus	It is the end of protein or polypeptide, terminated by a free carboxyl group (-COOH)
CaMK-II	Calcium and Calmodulin-dependent Serine/Threonine Protein Kinase Type II
Cas	CRISPR-associated
Cas9H840A	Mutant *Streptococcus pyogenes* (Histidine at position 840 changed Alanine)
CBF1	Centromere binding factor
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation-sequencing
CHUD	Chromatin unfolding domain
CHUD/ RD	Chromatin unfolding domain/ C-regulatory domain
CK 1	Casein kinase 1
CNS	Central nervous system
CP	Nucleosome core particle
CRD	Chromatin regulatory domain
CRISPR	Clustered regularly-interspaced short palindromic repeats
crRNAs	CRISPR RNAs
Ct	Cycle threshold
DAI	Days after induction
DAPI	4',6-diamidino-2-phenylindole
DAPT	N-[N-(3,5-Difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester
dCas9	Inactivated S. pyogenes dCas9 (D10A, H840A)
dCas9 ^{VP64}	Inactivated S. pyogenes dCas9 (D10A, H840A) fused to VP64 transactivator domain
Dcx	Doublecortin
del	Deletion
Dg	Dentate gyrus
dH_2O	Distilled water
DHSs	DNase I hypersitistive sites
Dir	Direct
DMEM/F12	Dulbecco's Modified Eagle Medium/Ham's F-12
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNA-PK	DNA-dependent protein kinase
DNMTs	DNA methyltransferases
dNTPs	Deoxynucleotide
DSBs	Double strand breaks
DTT	Dithiothreitol
E. coli	Escherichia coli
EBs	Embryoid bodies
ECCs	Embryonic carcinoma cells
EDTA	Ethylenediaminetetraacetic acid

EdU	5-ethynl-2'-dexyuridine
EGL	External germinal layer
EGTA	ethylene glycol-bis(ß-aminoethyl ether)-N,N,N',N'-tetraacetic acid is an aminopolycarboxylic acid, a chelating agent
EpiSCs	Epiblast stem cells
ES/ ESCs	Embryonic stem cells
Esrrb	*Estrogen-related receptor b*
EX	Exon
Ezh2	Zeste homologue 2
F	Forward
FACS	Fluorescence-activated cell sorting
FGF	Fibroblast growth factor
FGF8	Fibroblast Growth Factor 8 Recombinant Human Protein
FRAP	Fluorescence recovery after photobleaching
Fs	Frameshift
Fs	Frame-shift
FSC-A	Forward scatter-area
g	Force or relative centrifugal force (RCF)
G	Guanine
G1 and G2	Cell cycle gap phases
Gapdh or G3pdh	Glyceraldehyde 3-phosphate dehydrogenase
GC	Guanine-cytosine
GDF6	Growth and differentiation fctor-6
GFAP	Glial fibrillary acidic protein
GFP/ EGFP	Green fluorescent protein
Glut2	Glucose transporter gene
Glyt1	glycine transporter1
gp130	Glycoprotein 130
Gpi1	Glucose phosphate isomerase1
gRNA	Guide ribonucleic acid
GSK3	Glycogen synthesis kinase-3
h	Human
H3	Histone 3
H3K14ac	Histone 3 lysine 14 acetylation
H3K27ac	Histone 3 lysine 27 acetylation
H3K27me3	Histone 3 lysine 27 trimethylation
H3K36me2	Histone H3 Lysine 36 Dimethylation
H3K4me1	Histone 3 lysine 4 monomethylation
H3K4me3	Histone 3 lysine 4 trimethylation
H3K9ac	Histone 3 lysine 9 acetylation
HATs	Histone acetyltransferases
HD	Homeodomain
HDACi	Histone deacetylase inhibitor
HDACs	Histone deacetylases
HDMs	Histone demethylases
HEPES	N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid
Herp	*Hes-related protein*
HiFi	High-Fidelity
HMG	High mobility group
HMGN/ Hmgn	High mobility group-nucleosome binding domain
HMGN1	High mobility group-nucleosome 1
Hmgn2/ N2	High mobility group-nucleosome 2
HMGN2ΔNBD	HMGN2 with mutation deleting part of NBD
HMGN2ΔNBD_CHUD	HMGN2 with mutation delete NBD and CHUD or part of them.
Hmgn3	High mobility group-nucleosome 3
Hmgn4	High mobility group-nucleosome 4
HMTs	Histone methyltransferases
HN/LN	High or Low *Nanog* cells
HP1	Heterochromatin protein 1
HR	Homologous recombination
HRP	Horseradish peroxidase
ICM	Inner cell mass
Id	Inhibitory of DNA-binding
IDLVs	Integrase-defective lentiviral vectors
IF	Immunofluorescence
IGF	Insulin-like growth factor
IL-6	Interleukin-6
Ins	Insertion
iPSCs	Induced pluripotent stem cells
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
JAK-STAT3	Janus kinase-signal transducer and activator of

	transcription 3
Kb	Kilobase
Klf4	Krüppel-factor4
KO	knockout
l	Litre
LacZ	Lac operon. The gen product of LacZ is ß-galactosidase which cleaves lactose, a disaccharide, into glucose and galactose.
LIF	Leukaemia inhibitor factor
LIFR	LIF receptor
LRP	Lipoprotein receptor-related pro
Lys or K	Lysine
M	Marker or ladder
M	Molar
m	Mouse
M	mitosis
MAP2	Microtubule-associated protein 2
MAPK	Mitogen-activated protein kinase
Mash1	Known as Ascl1 (Achaete-scute homolog 1)
MBD	Methyl-CpG binding domain
MEFs	Mouse embryonic fibroblasts
Mg	Magnesium
$Mg2^+$	Magnesium ion
mins	Minutes
ml	Millilitre
MLL2	Mixed-lineage leukaemia 2
mM	Milli-molar
mm	Millimetre
MmHmgn2ExIV	Mouse Hmgn2 exon IV
MmN2ATG	Mouse Hmgn2 ATG
MNase	Micrococcal nuclease
MRN	Mre11/Rad50/Nbs1
Msat	M-satellite
MSK1	Mitogen and stress-activated kinase 1
MW	Molecular weight
N-terminus	The start of a protein or polypeptide terminated by an amino acid with a free amine group (-NH2)
NB or NBA	Neurobasal media
NBCS	New born calf serum
NBD	Nucleosome binding domain
NC	Negative control
NeuroD	Also called Beta2, is a basic helix loop helix transcription factor
NF-160	Neurofilaments medium
NFH	Neurofilament heavy
NFL	Neurofilament light
NFM	Neurofilament medium
ng	Nanogram
Ngn1	Neurogenin1
NHEJ	Non-homologous end joining
NLS	Nuclear localisation signal
nM	Nano-molar
NPCs	Neural progenitor/precursor cells
NRSE	neuron-restrictive silencer element
NSCs	Neural stem cells
NuRD	Nucleosome remodelling and deacetylase
Oct4	Octamer-binding transcription factor 4 also known Pou5F1 gene
oligos	Oligonucleotide
P	Proline
P	Phosphorylation
p.MN	pMultiNick
PAM	Protospacer adjacent motif
Pax6	Paired box -6 protein or gene
PBS	Phosphate-buffered saline
PCAF	P300/CBP-associated factor
pCas9D10A	Mutant *Streptococcus pyogenes* (Aspartate at position 10 changed Alanine)or/and plasmid with only Cas9D10A
PcG	Polycomb group
PCR	Polymerase chain reaction
Pen/strep	Penicillin streptomycin
pGuide001-4_mU6	Plasmid has individual guide followed by number of gRNA and specific promoter (mU6)
PI	Propidium iodide
PI	PAM-interacting
PITX2	
Pmol	picomole
pMultiCas9D10A	Plasmid contains Cas9 and multiguide cassette
pMultiNick_001-4	Plasmid multinickase/ multiguide followed by

	number
Pr.	Primer
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
pUC	The origin of replication
PVDF	Polyvinylidene difluoride
QPCR or Q-PCR	Quantitative polymerase chain reaction
qRT-PCR	Quantitative real time-polymerase chain reaction
R or REV	Reverse
RA	Retinoic acid
Raldh2	Retinal dehydrogenease-2
RARs	Retinoic acid receptors
Rdh10	Retinol dehydrogenease-10
RE1	Repressor element 1
REC	Alpha helical recognition
Rest	RE1 Silencing Transcription Factor
RFLP	Restriction fragment length polymorphism
RFP	
RG	Radial glial
RNA	Ribonucleic acid
RNAi	RNA interference
RNAi	RNA interference
RNase A	Ribionuclease A
Rplp0	Ribosomal Protein Lateral Stalk Subunit P0
RPM	Revolutions per minute
RRSARLSA	R-arginie, S-serine, A-alanine, L-leucine
RS	Restriction site
RSK2	Ribosomal S6 kinase 2
RXRs	Retinoid X receptors
SD/STDVP	Standard deviation
SDS	Sodium dodecyl sulfate
Sec or S	Second
sgRNA	Single guide RNA
SNPs	Single nucleotide

	polymorphisms
SpCas9	*Streptococcus pyogenes Cas9*
SQRT	The square root
SSC-A	Side scatter-area
SVZ	Subventricular zone
SW12/SNF2	switching/source nonfermenting
T	Thymine
T2A	self-cleaving 2A peptide
TALENs	Transcription activator-like effector nucleases
TCF/LEF	T cell factor/lymphoid enhancer factor
TET	Ten-eleven translocation
TfcPl1	Transcription factor CP2-like protein 1
TFs	Transcription factors
TGFβ	Transformation growth factor beta
tracrRNA	Trans activating CRISPR RNA
Trkβ	*Tropomyosin-related kinase β*
trxG	Trithorax group
TSS	Transcription start site
U	Unit
UTRs	Untranslated regions
v	Voltage
VZ	Ventricle zone
WT	Wild-type
Wt-tr	Electroporated wild type
X-gal	5-bromo-4-chloro-3-indolyl-ß-D-galactopyranoside
ZFNs	Zing finger nucleases
Zfp521	Zing finger protein521
Zfp521	Zing finger protein 521
ß-ME	2-mercaptoethanol
Δ	Delta
μg	Microgram
μl	Microliter
μM	Micro-molar

Chapter 1 Introduction

1.1 Embryonic stem/carcinoma cells

Embryonic stem cells (ESCs) are derived from the inner cell mass (ICM) of the blastocyst, which is the first stage of cellular differentiation during embryo development (Hattori et al., 2007). The cells in the ICM are the cells that create all tissues of the body but do not give rise to the extraembryonic endoderm and trophoblast. The isolation and *in vitro* culture of these cells from mouse and human facilitate the study of the cellular and molecular biology of early embryogenesis. ESCs have two important features, which are self-renewal and pluripotency. Self-renewal means that they can undergo unlimited symmetrical cell divisions without differentiation. Secondly, these cells have also the potential to differentiate into all tissues including three germ layers (endoderm, mesoderm and ectoderm), and this feature is attributed to pluripotency (reviewed by Vastenhouw et al., 2012). The unlimited capacity of self-renewal and the ability to differentiate into three germ layers are the main properties for considering these cells as an infinite resource for cell therapy and replacement.

Embryonic carcinoma cells (ECCs) are also pluripotent stem cells, and isolated from teratocarcinomas produced by grafting early mouse embryos into adult mice tissue (reviewed by Smith, 2001; Friel et al., 2005). ECCs were isolated and cultured *in vitro* for the first time in 1967 by Finch and Ephrussi. One example of ECCs is the P19 cell line, which was derived from a teratocarcinoma formed after transplanting a 7.5-day embryo into the testis of adult mice (Mcburney, 1993). Therefore, these cells can be propagated in culture while retaining the capacity to differentiate into any of the three primary germ layers. The ECCs are similar to ESCs because both can differentiate into extra-embryonic endoderm-like cells or neuroectoderm derivatives, e.g., neurons (reviewed by Castro-Obregon and Covarrubias, 1996). When mouse ECCs or ESCs are sited in a blastocyst, both can play a role in embryonic development and contribute to normal tissues, resulting in a chimeric embryo (Brinster, 1974; Papaioannou et al., 1975). However, ECCs form chimeras less efficiently than ESCs, and this may refer to the accumulation of the genetic and epigenetic changes acquired in ECCs (Draper et al., 2002). As in ESCs, *Oct4*, *Nanog*, and *Sox2* are also expressed

in ECCs (Niwa et al., 2000; Andrews et al., 1996). In addition to these transcription factors, there is a wide range of markers that are identical for ESCs and ECCs; these can be used to characterise ESCs and ECCs and to monitor their differentiation. They include cell surface markers, such as SSEA1 for mouse (m)ESCs and (m)ECCs, and SSEA3/4 after differentiation, and intracellular markers, for example, alkaline phosphatase (Solter, 2006; Ramirez et al., 2011).

Although ECCs and ESCs are similar in morphology, growth behaviour and marker expression, there are some differences between them. ECCs are cancer-derived and mostly aneuploid, while ESCs maintain a diploid karyotype. Nevertheless, both ESCs and ECCs can acquire karyotyping changes during prolonged culture (Andrew et al., 2005). In addition, P19 cells were originally reported to have a normal mice euploid karyotype (Angello et al., 1997). *In vitro*, ECCs can differentiate into neuroectoderm or striated muscle when aggregated in presence of retinoic acid (RA) or DMSO, respectively, and into mesoderm and endoderm when treated with RA in monolayer (Mummery et al., 1987).

As ESCs and ECCs share the same features of self-renewal and pluripotency, several studies have identified the molecular mechanism responsible for these features (Hattori et al., 2004; Pan and Pie, 2003). The self-renewal and differentiation of ESCs can be regulated by a combination of extrinsic and intrinsic factors such as signalling pathways and transcription factors (TFs), respectively. Three canonical TFs, Oct4, Nanog and Sox2, are required for specifying the stemness of ESCs, and are the core pluripotency circuitry (reviewed in Huang et al., 2015). Chromatin structure and epigenetic mechanisms play key roles in regulating these processes. These different components are discussed in more detail below.

1.1.1 Extrinsic factors regulating the self-renewal and pluripotency of ESCs

The self-renewal of ESCs, which is the capability to produce two identical stem cell daughters, requires repression of differentiation during proliferation. Several studies propose that the self-renewal ability is dependent on a combination of factors, including the maintenance of proliferation, the inhibition of apoptosis and blocking differentiation (Figure 1.1) (Zhang and Li,

2005; Ying et al., 2008). As self-renewal and differentiation can be regulated by intrinsic factors, several signalling pathways have also been described to be associated with pluripotency and self-renewal of ESCs, such as the LIF, BMP and Wnt pathways (Figure 1.1).

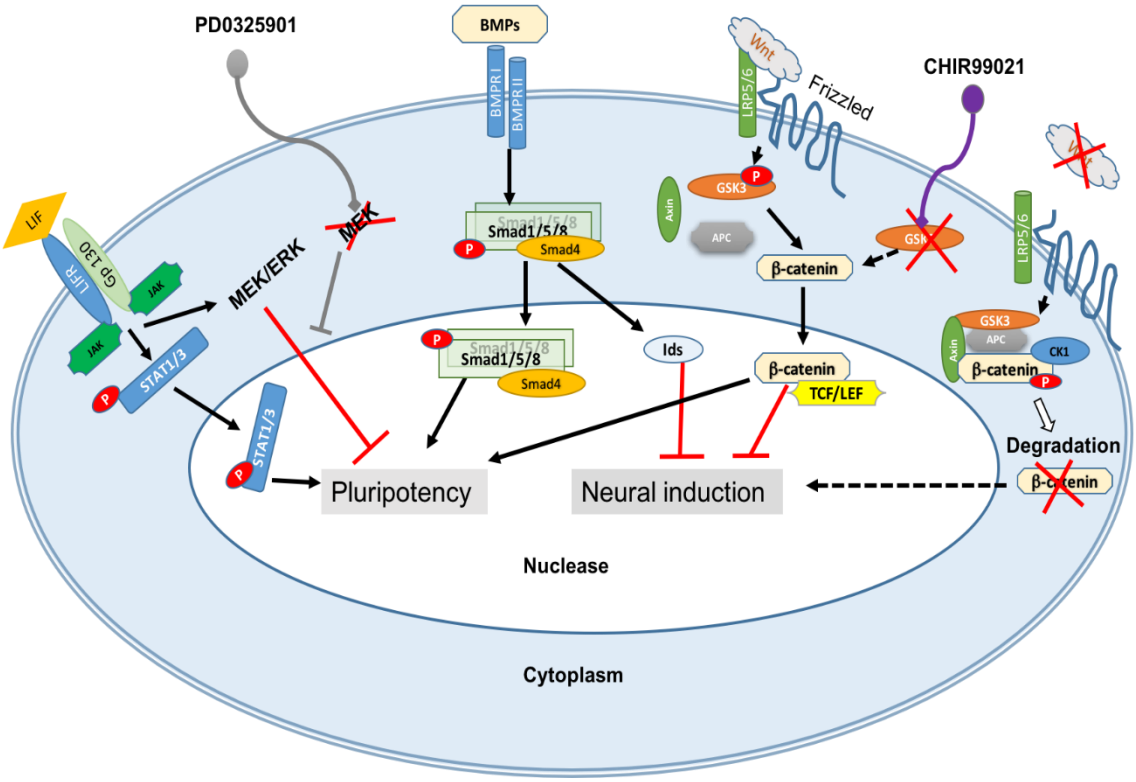


Figure 1.1: Signalling pathways (LIF, BMP and Wnt) involved in pluripotency of mESC. Three major signalling pathways maintain the self-renewal and pluripotency, which are LIF, BMP and Wnt. LIF can promote the pluripotency of mESCs by activation of JAK-STAT pathway. LIF binds to its receptor (LIFR) that forms heterodimer complex with gp 130. This leads to phosphorylation of JAK, LIFR and gp 130 (not shown), as well as phosphorylates STAT1/3. Phosphorylated STAT1/3, which is a transcription factor, translocates into the nucleus to regulate pluripotent genes. Activation of LIFR and gp 130 also results in the activation of another pathway (MAPK) and its downstream MEK/ERK proteins, which inhibit the pluripotency state by inducing the non-neural differentiation. The inhibition of MEK by an inhibitor (PD032591) can promote the self-renewal in presence of LIF. BMP signalling is initiated by binding of BMP4 as an example to its receptor (BMPR I and II, type I and II serine-threonine) and phosphorylates Smad1/5/8. A complex of two phosphorylated Smad1/5/8 with common Smad4 enters into the nucleus and cooperates with other TFs to regulate target gene expression that are important for pluripotency. Smad1 can also stimulate Id genes that inhibit neural induction. Thirdly, Wnt signalling can promote the self-renewal by using GSK3 inhibitor (CHIR99021) in presence of LIF and MEK inhibitor. GSK3, in absence of Wnt, forms a destruction complex with adenomatous polyposis coli (APC), casein kinase 1 (CK1), and Axin to phosphorylate β-catenin. Phosphorylated β-catenin is degraded which results in promoting neural differentiation. However, the activation of Wnt signalling by binding to its receptor Frizzled with co-receptors such as lipoprotein receptor-related protein (LRP5/6) leads to phosphorylation of GSK3, preventing the formation of destruction complex and accumulation of unphosphorylated β-catenin that enters the nucleus. In the nucleus, β-catenin interacts with T-cell factor/lymphoid enhancer factor (TCF/LEF) and binds to specific motif and regulates the transcription of target genes for mESCs self-renewal. P refers to phosphorylation. {Original figure modified from Zhang and Li, 2005; Ying et al., 2008}.

1.1.1.1 Leukaemia Inhibitor Factor (LIF)

Although the mESCs can be expanded and cultured *in vitro*, medium containing all necessary metabolites and nutrients is not sufficient for maintaining the undifferentiated ESCs. It was discovered that co-culture of mESCs with feeder cells could support the maintenance of these mESCs, as a self-renewal factor is provided by the feeder cells. Later, it was found that the mESCs can be maintained by leukaemia inhibitor factor (LIF, a member of interleukin-6, IL-6, family of cytokine) in feeder-free medium (William et al. 1988; Smith et al., 1988). LIF withdrawal from mESCs culture leads to rapid differentiation into a mixed population of mesoderm and endoderm cells (Niwa et al., 1998).

LIF functions by activating the Janus kinase-signal transducer and activator of transcription 3 (JAK-STAT3) signalling pathway (Figure 1.1) (Niwa et al 1998; Boeuf et al., 1997). The activation of JAK-STAT3 signalling pathway is initiated by binding of LIF into cell surface LIF receptor (LIFR). This binding induces the recruitment of another common receptor subunit of the IL-6 family called glycoprotein 130 (gp130) to form a heterodimer complex which mediates downstream signal transduction (reviewed in Onishi and Zandstra, 2015). This complex resulted in a rapid activation of JAKs that phosphorylate LIFR and gp130. Three Jak kinases (Jak1, Jak2 and Tyk2) are activated by LIF in ESCs (Ernst et al., 1996). Consequently, STATs which are TF are recruited to the receptors and phosphorylated by Jaks on their tyrosine residues. The dimerized STAT1/3 is then translocated into the nucleus for direct regulation of the transcription of several genes (reviewed in Matsuda et al. 1999; Smith, 2001; Onishi and Zandstra, 2015).

Matsuda et al. (1999), who used a constitutive expression of a conditionally active form of STAT3 (STAT3-estrogen receptor) that can be phosphorylated by 4-hydroxytamoxifen, showed that activation of STAT3 independently from LIF is sufficient for maintaining the self-renewal of mESCs. Several studies have focused on identifying potential candidates for pluripotency factors as the downstream targets of STAT3. Expression of the c-Myc transcription factor is directly regulated by STAT3, and sustained expression of *c-Myc* promotes self-renewal and maintenance of pluripotency independent of LIF (Cartwright et al., 2005). Additional TFs have been identified as downstream targets of STAT3 such

as Krüppel-factor4 (Klf4) and Transcription factor CP2-like protein 1 (TfcPl1) (Hall et al., 2009; Bourillot et al., 2009; Ye et al., 2013; Martello et al., 2013). Klf4, which is a zinc finger protein and one of the four important TFs for reprogramming somatic to induced pluripotent stem cells, is stimulated by STAT3 (Bourillot et al., 2009; Takahashi and Yamanaka, 2006). Depletion of Klf2, Klf4 and Klf5 results in ESCs differentiation (Jiang et al., 2008). However, constitutive expression of TfcPl1, but not Klf4, is sufficient to maintain undifferentiated Stat3 null ESCs (Hall et al., 2009; Martello et al., 2013). According to Bourillot et al. (2009), some of the downstream targets of STAT3 are also regulated by Nanog which clearly highlights the mechanisms underlying functional redundancy between STAT3 and Nanog in mESCs although the absence of direct regulation between Nanog and STAT3.

In addition to JAK-STAT pathway, the LIF and gp130 complex also stimulates the MAPK pathway (mitogen-activated protein kinase) which has a cascade of trans-phosphorylation including RAS/Raf/MEK/ERK) (Figure 1.1). However, this pathway generates a "pro-differentiation" signal rather than supporting self-renewal (Burdon et al., 1999a; Burdon et al., 1999b). Therefore, the self-renewal of undifferentiated ESCs is enhanced by MEK inhibitor, which can inhibit this specific pathway but leave STAT3 signalling pathway active (Burdon et al., 1999a). Unlike mESCs, human (h)ESCs do not rely on JAK-STAT signalling for maintaining their self-renewal (Humphrey et al., 2004; Onishi and Zandstra, 2015).

1.1.1.2 Bone morphogenic protein (BMP)

Another main signalling pathway important in self-renewal of ESCs is bone morphogenic protein (BMP) (Friel et al., 2005). This signalling pathway is provided by serum since LIF alone in serum-free media cannot inhibit the neural differentiation (Ying et al., 2003a). Although ESCs can be cultured in serum replacement medium with LIF, high cell density is required as stem cell colonies never form from single cells in the low-density culture (Ogawa et al., 2004). However, ESCs can be propagated and stem cell colonies can be formed from a single cell in serum-free medium if BMP4, or BMP2 or GDF6 (growth and differentiation fctor-6) are supplemented with LIF (Ying et al., 2003a). The BMP signalling pathway is well-known to inhibit neural differentiation (Ying et al.,

2003b). The BMP signalling pathway is mediated by the interaction of BMPs with heterodimeric type I and II serine-threonine receptors (Figure 1.1). Upon binding, cytoplasmic R-Smads (Smad1, Smad5 and Smad8) are phosphorylated. Two phosphorylated R-Smads form a heterotrimeric complex with a common Smad4 (co-Smad). The Smad heterotrimeric complex translocates into the nucleus and cooperates with other TFs to regulate target gene expression (reviewed by Zhang and Li et al., 2005).

Smad1 can stimulate Id (inhibitory of DNA-binding) gene expression, and Id proteins inhibit the functions of pro-neural basic helix-loop-helix (bHLH) transcription factors such as Mash1 (Figure 1.1) (Ying et al., 2003a). The binding sites of SMAD in Id gene promoters were subsequently identified in mouse ESCs (Fei et al., 2010). Genome-wide occupancy studies identified that SMAD1 is also bound at several common targets along with the core pluripotent factors (Oct4, Sox2 and Nanog) (Chen et al., 2008). Although BMP belongs to transformation growth factor beta (TGFβ) superfamily that includes TGFβ/Activin/Nodal and BMP/GDF, the function of BMP in sustaining the self-renewal of ESCs is restricted to BMP/GDF because regulation of Id2 expression by Smad1, 5, or 8 is only achieved through BMPs but not TGFβ (Ying et al., 2003a).

1.1.1.3 Wnt pathway

Generally, Wnt signalling is stimulated by binding of Wnt protein to Frizzled family receptors and co-receptors such as lipoprotein receptor-related protein (LRP5/6) (Figure 1.1). The activation of Wnt signalling promotes nuclear accumulation of unphosphorylated β-catenin in the cytoplasm that then translocates into the nucleus. β-catenin in the nucleus is associated with a TF known as T-cell factor/lymphoid enhancer factor (TCF/LEF) that binds to a specific motif and regulates the transcription of target genes. In the absence of Wnt ligand, GSK3 negatively regulates this signalling pathway by forming a destruction complex with adenomatous polyposis coli (APC), casein kinase 1 (CK1), and Axin, which degrades the phosphorylated β-catenin by targeting it for ubiquitination. Upon Wnt ligand stimulation, the assembly of this complex is inhibited.

In ESCs, inhibition of glycogen synthesis kinase-3 (GSK3) was shown to enhance the self-renewal of mouse and human ESCs (Sato et al., 2004). In serum-free medium, using the MEK inhibitor improves ESC propagation for several passages after LIF withdrawal but with poor clonogenicity because of compromised growth and viability (Ying et al., 2008; Huang et al., 2015). However, cell viability is enhanced by using a GSK3 inhibitor (CHIR99021), which results in increased metabolic and biosynthetic ability (Ying et al., 2008). In addition, the promotion of self-renewal by the GSK3 inhibitor is attributed to the stabilisation of β-catenin, the essential factor of the canonical Wnt signalling pathway (Ying et al., 2008; Ye et al., 2012; Wray et al., 2011). Ye et al. (2012) indicate that overexpression of β-catenin recapitulates the effect of GSK3 inhibitor in maintaining the undifferentiated ESCs. Furthermore, it was found that GSK-3 inhibition is important for ESCs by reducing the repressive effect of TCF3; this is the predominant Tcf/Lef in ESCs and it has been reported to behave primarily as a transcriptional repressor in the pluripotency gene network (Wary et al., 2011).

1.1.2 Intrinsic factors for regulating the self-renewal of ESCs

1.1.2.1 Core transcription factors (Oct4, Sox2 and Nanog)

Analysis of the molecular network that is responsible for regulation of the mouse ESCs pluripotency identified Oct4, Sox2 and Nanog as the main TFs involved in this mechanism. Firstly, Octamer-binding transcription factor 4 (Oct4), which is a member of POU family, is expressed in the embryo during the pre-implantation period (reviewed by Chambers and Smith, 2004; Huang et al., 2015). Its expression is detected in the ICM but not in the trophectoderm, and it is maintained in the epiblast before becoming restricted to the migratory primordial germ cells (Kehler et al., 2004). *Oct4*$^{-/-}$ embryos can survive to the morula stage but are developmentally compromised at the blastocyst stage because the *Oct4*$^{-/-}$ cells in the ICM tend to differentiate into trophectodermal cells (Nichols et al., 1998). In addition, Niwa et al. (2000) showed that precise expression of Oct4 in ESCs is required for maintaining self-renewal, since twofold higher expression induces differentiation into primitive endoderm and mesoderm, whereas a reduction to half causes differentiation into trophectoderm. This TF is bound to DNA, specifically to the octamer motif (ATGCAAAT), to regulate other several genes involved in pluripotency. Oct4 is

also known to activate or repress other target genes by cooperating with various transcriptional co-factors such as Sry-related factor (Sox2) (Avilion et al., 2003).

Sox2, a TF containing a high-mobility group (HMG) DNA-binding domain, interacts with Oct4 and plays a role in ESCs pluripotency. Sox2 is expressed in a wide variety of cells including the ICM, epiblast, extra-embryonic ectoderm and neural tissues (reviewed in Hung et al., 2015). Knockdown or inducible deletion of *Sox2* in ESCs results in the induction of differentiation into trophectoderm-like cells, a similar finding to Oct4 (Ivanova et al., 2006; Masui et al., 2007). This indicates the importance of Sox2 and Oct4 in the maintenance of ESCs. Several studies revealed that Oct4 and Sox2 bind to the enhancers of many other pluripotency-associated genes such as *Fgf4*, *Lefty1* and *Nanog,* and activate their expression (Kuroda et al., 2005; Nakatake et al., 2006). The *cis*-regulatory element bound by the Sox2-Oct4 complex consists of neighbouring Sox (CATTGTA) and Oct (ATGCAAAT) elements (Loh et al., 2006). Moreover, Oct4 and Sox2 expression are also regulated by a positive-feedback loop that means Oct4/Sox2 bind to their own enhancers and regulate their own expression (Tomioka et al., 2002; Loh et al., 2006).

The third transcription factor that coordinates with Oct4 and Sox2 in sustaining the pluripotency of ESCs is Nanog, which is a homeodomain-containing protein. The expression of Nanog is restricted to morula, ICM, early germ cells and in the proximal epiblast in the region of the presumptive primitive streak (reviewed in Chambers and Smith, 2004; Friel et al., 2005). Therefore, its expression appears slightly later than Oct4. It was shown that Nanog can be regulated by Oct4/Sox2 binding to its enhancer (Kuroda et al., 2005). *Nanog*$^{-/-}$ embryos appeared initially to give pluripotent cells but the ICM spontaneously differentiates into the extra-embryonic endoderm (Mitsui et al., 2003). The overexpression of Nanog supports the self-renewal of undifferentiated ESCs in absence of extrinsic LIF in serum-supplemented medium, though Nanog and STAT3 do not regulate each other. Although the function of Nanog in maintaining of ESCs status is independent of LIF, both maximise the self-renewal efficiency (Mitsui et al., 2003; Chambers et al., 2007). Similarly, in serum-free media, ESCs with Nanog overexpression can sustain their ability of self-renewal without BMP signalling, probably because of direct regulation of Nanog on Id gene expression (Ying et al., 2003a). Expression

of Nanog fluctuates in mouse ESCs, which leads to population heterogeneity (Chambers et al., 2007; Kalmar et al., 2009).

Several genes such as *Esrrb*, *Foxd3* and *REST* have been identified by genome – wide mapping to have Nanog binding sites, and Oct4 and Nanog overlap substantially in their targets (Chen et al., 2008; Ivanova et al., 2006; Loh et al., 2006). *Estrogen-related receptor b* (*Esrrb*) was shown to be directly bound by Nanog, and Nanog enhances the binding of RNA PolIII at *Esrrb* promoter which leads to increased expression (Festuccia et al., 2012). Finally, it can be argued that Nanog is not strictly required for establishing the pluripotency of ESCs, since although $Nanog^{-/-}$ ESCs are prone to differentiate, ESCs can self-renew indefinitely in the permanent absence of Nanog (Chambers et al., 2007). In addition, Nanog is not essential for reprogramming somatic cells into induced pluripotent stem cells (iPSCs) (Takahashi and Yamanaka, 2006; Schwarz et al., 2014).

1.2 Neural induction and neurogenesis

As mentioned previously, the cells in the ICM can be specified into the three germ layers, including the ectoderm. The first sign of the developing nervous system is the neural plate. Neural plate, which is originated from the ectoderm of gastrulating vertebrate embryo, contain a single layer of neuroepithelial cells. These cells then undergo multiple and rapid symmetrical divisions to expand the neural plate. The neural plate then invaginates to form the neural tube that eventually develops into all parts of central nervous system (CNS) (reviewed by Martynoga et al., 2012). The brain is formed from the anterior end of neural plate while the spinal cord is formed from the posterior end. The neural plate and tube are composed of a single layer of neuroepithelial cells that are considered to be multipotent neural stem cells (NSCs) as they able to produce all cells in CNS including the neurons, oligodendrocytes, and astrocytes (reviewed in Ozair et al., 2013). The neuroepithelial cells can migrate up and down the apical–basal axis during the cell cycle.

Some of the neuroepithelial cells at the ventricle zone (VZ) undergo asymmetric division, generating an identical stem cell and a more differentiated daughter cell, either progenitor or a neuron. At the same time, some of the progenitors

take on the identity of radial glial (RG) cells. RG cells are bipolar-shaped cells that are thought to be the primary progenitor cells for most of the neurons (either directly or indirectly) and linage-restricted intermediate progenitors that drive astrocyte and oligodendrocyte formation (reviewed in Gotz and Huttner, 2005). Unlike the early neuroepithelial cells, most radial glial cells are restricted to the generation of a single cell type, either astrocytes, oligodendrocytes or neurons (Grove et al., 1993; Noctor et al., 2001). The production of neurons from RG cells occurs first during development, followed by glial cell production. Another type of progenitor called basal/subventricular zone (SVZ) progenitors originate from the mitosis of neuroepithelial and RG cells at the apical surface. These basal progenitors appear at time of neuron production, and divide symmetrically to produce two neurons (Noctor et al., 2004; Miyata et al., 2004). The neurons produced from the basal/subventricular progenitors were seen to be more abundant than those derived from apical/ventricular progenitors in the telencephalon (Haubensak et al., 2004). The production of neurons from the asymmetric and symmetric division of NSCs and their derivative progenitor cells is known as neurogenesis.

1.2.1 The default model of neural induction, and signalling pathways that regulate neural induction and neurogenesis

Neural induction in vertebrates is triggered by signals from a specialised cluster of cells (the organizer) that promotes neural development in the ectoderm (reviewed in Stern, 2007). The neural induction process is thought to be the 'default model', which means that ectodermal cells acquire a neural fate following the elimination of the instructive signalling molecules (BMPs) that drive epidermal cell formation. It was found that depletion of *Bmp-2*, *-4*, *-7* and ADMP (anti-dorsalising morphogenetic factor) or the organizer (where the ADMP is expressed) resulted in a greatly enlarged brain in Xenopus embryos (Reversade et al., 2005a and b). Further evidence supports the default model. For example, it was found that the organizer secretes three BMP antagonists (Noggin, Chordin and Follistatin) and removal of all three antagonists in *Xenopus* and *zebrafish* resulted in complete loss of neural tissue (Smith and Harland, 1992; Sasai et al., 1994; Hemmatibrivanlou et al., 1994; Khokha et al., 2005). In agreement with this model, inhibition of BMP4 is known to be sufficient to directly induce neural fate in mammalian embryo and pluripotent mESCs and hESCs (Finley et al., 1999;

Ozair et al., 2013). In addition to BMP signalling, fibroblast growth factor (FGF) and insulin-like growth factor (IGF) family members, which bind tyrosine kinase receptors and signal via the MAPK cascade, also play role in the neural induction (Wills et al., 2010). FGF/MAPK is also known to inhibit BMP signalling by promoting the phosphorylation and degradation of SMAD1 (Pera et al., 2003). However, some studies have suggested neural induction requires FGF signalling independently of its ability to BMP inhibition (Aubin et al., 2004; Linker and Stern, 2004; Stern, 2007).

Although the inhibition of BMP signalling is required to induce anterior forebrain-like tissue, Wnt/β-catenin, FGF and RA signalling are also required for the induction of posterior regions of the nervous system (Ozair et al., 2013). However, because Wnt/β-catenin can inhibit neural plate formation at the onset of gastrulation, this signalling pathway needs to be inhibited in the anterior neural tissue in the default model (Min et al., 2011; Pere et al., 2014). Wnt signalling stabilises SMAD1 in a β-catenin-dependent manner and prolongs the duration of BMP signalling (Pere et al., 2014). Therefore, the dual inhibition of BMP and Wnt is required for anterior forebrain. RA plays an important role in patterning and morphogenesis of a developing embryo in the posterior region of nervous system, and reduced level of RA signalling in vitamin-A-deficiency quail resulted in a paucity of neurons in the spinal cord with a loss of proneural gene expressions such as *Neurogenin1/2* (*Ngn1/2*) (Maden et al., 1996, Diez et al., 2003). Expression of NeuroD, which is a bHLH transcription factor, was increased as a result of increasing RA level in chick spinal cord explants, indicating induction of the neural differentiation with increased RA (Diez et al., 2003). Not only in the hindbrain, some studies indicate that RA plays role in patterning of the embryonic forebrain and cell survival in the telencephalon (Schneider et al., 2001; Halilagic et al., 2003).

RA, which is vitamin A-derived, is small lipophilic molecule that performs as a ligand for nuclear RA receptors (RARs). Binding RA to RARs alters the conformation of RARs and converts them from transcriptional repressors to activators. RARs act in heterodimeric combinations with retinoid X receptors (RXRs) and RAR/RXR is bound to DNA motifs known as RAREs. The concentration of RA is strictly regulated during development by different enzymes. These

include Rdh10 (retinol dehydrogenease-10) and Raldh2 (retinal dehydrogenease-2) that both mediate the biosynthesis of RA from retinal, while RA is degraded by the hydroxylase Cy26A1 (Rhinn and Dolle, 2012; Pere et al., 2014). Several studies showed the importance of RA in the regulation of neurogenesis and neuronal differentiation, although the exact role of RA in neurogenesis is not completely understood. RA was shown to induce the differentiation of ESCs and ECCs such as P19 cells into neuronal and glial cells *in vitro* (Jones-Villeneuve et al., 1982; Janesick et al., 2015). Moreover, mutation of the RARα-ligand binding domain in a P19-derived cell line (RAC65) inhibits neuronal differentiation, which also indicates the importance of RARα in neuronal differentiation (Kruyt et al., 1992; Pratt et al., 1990). One direct target of RAR is *HoxA1,* which was shown to be essential for differentiating of ESCs into neurons (Kolm and Sive, 1995; Langston et al., 1997). Several other genes can also be targeted by RAR. The level of RA is also critical for the number and timing of primary neurons production during embryo development (Franco et al., 1999; Blumberg et al., 1997; Sharp and Goldstone, 1997). RA-deficient mice were shown to have fewer mature and immature neurons ($NeuN^+$ and Dcx^+, respectively) of the hippocampal dentate gyrus (Jacobs et al., 2006). In addition, NSCs in the proliferative VZ of cortex also need RA to differentiate into intermediate progenitor cells and post-mitotic neurons (Siegenthaler et al., 2009). In addition to the role of RA in differentiation, it was found the neurons derived from ESCs after RA exposure have significant neurite growth (Theus et al., 2006)

Another major signalling pathway involved in neurogenesis is the Notch pathway, which involves cell-cell signalling between adjacent cells. The Notch signalling pathway involves 25 genes and has many functions in neural differentiation, including switching a neural cell's fate, sustaining the potency of the NSCs and determining the fate of ESCs (Marei et al., 2012). The transition from neuroepithelial to RG cell identity is promoted by Notch signalling and its downstream TFs (Hes1 and Hes5) (Gaiano et al., 2000). In addition, the FGF signalling pathway is also necessary for this transition because overexpression of FGF10 results in the appearance of RG markers, while mutant *Fgf10* mouse embryos had prolonged neuroepithelial expansion and delayed neurogenesis (Sahara and O'Leary, 2009). Similarly, the activation of Notch in mESC derivatives after withdrawal of LIF promotes neural differentiation, whereas

inhibition of Notch blocks formation of neural progenitors. The ability of Notch ligands to promote neural progenitor formation required FGF receptor-mediated signalling (Lowell et al., 2006).

The Notch ligands are members of Delta and Jagged family. The ligand-receptor binding results in the cleavage of Notch by γ-secretase and the intracellular domain of the Notch receptor, known as NICD, is then released from the cell membrane and transports into the nucleus. NICD binds to the CBF1 (centromere binding factor) repressor complex converting it to an activator complex. The activator complex NCID/CBF1 promotes the expression of target genes such as *Hes* and *Herp* (*Hes-related protein*), which are bHLH transcription factors and inhibit proneural genes such as *Mash1* and *Ngn* (reviewed by Yoon and Gaiano, 2005). Therefore, Notch signalling can inhibit neuronal differentiation. Several studies concluded that disruption of any Notch pathway member such as ligand, receptor and modulator leads to more neuronal differentiation markers and reduction in progenitor markers (Hitoshi et al., 2002; Xue et al., 1999; Yun et al., 2002; Zhong et al., 2000; Petersen et al., 2004; Handler et al., 2000).

1.2.2 ESCs and ECCs as models for neuronal differentiation *in vitro*

Pluripotent stem cells and multipotent NSCs can both contribute to the production of neurons *in vitro*. Mouse or human ESCs and their equivalent embryonal carcinoma stem cells (ECCs) have been used as models for studying the biological mechanisms involved in the neuronal fate specification and many neural developmental disorders, and regenerative medicine applications. mESCs that require LIF, BMP and Wnt for sustaining the naïve (ground) state of pluripotency are considered to be at an earlier stage of development than hESCs (Ying et al., 2008; Brons et al., 2007; Tesar et al., 2007; Ozair et al., 2013). However, adding of FGFs or Wnt inhibitor to mESCs transitions them into epiblast stem cells (EpiSCs) that are similar to hESCs in their requirement for FGF and Activin/Nodal signalling in the maintenance of pluripotency (Guo et al., 2009; ten-Berge et al., 2011). mECCs can be considered to be mEpiSCs (Bain et al., 1994; Ozair et al., 2013). Different protocols for deriving neurons from mESCs/ECCs *in vitro* have been used and mostly rely on adding neural inducers such as RA or withdrawing differentiation inhibitors such as LIF and BMP. This

includes: culturing ESC/ECCs as embryoid bodies (EBs) in serum- and RA-containing medium, co-culturing of ESCs with cell lines that have neural-inducing activity, depriving the ESCs of cell-cell interactions and signals by low-density culture in serum-free medium, or more recently culturing cells in serum-free and defined monolayer culture conditions (Smith et al., 1988; Ying et al., 2003b; Wataya et al., 2008; Ozair et al., 2013; Chuang et al., 2015).

EB formation has been widely used because it mimics the environment that produces neuroectoderm *in vivo,* in which it provides appropriate cell-cell interactions and signalling (Cai and Grabel, 2007). These EB aggregates, that resemble the anterior pre-streak stage embryo, contain the central epiblast-like cells surrounded by hypoblast-like cells (Maye et al., 2004). Although the formation of EBs does not require RA, the production of neural lineage cells from EBs can be enhanced by the addition of RA (Guan et al., 2001). Therefore, ESC embryoid bodies are cultured in suspension for 4 days in absence of RA, and then for another 4 days but in presence of RA (4-/4+), before plating them on defined substrates for 7 days (Bain et al., 1995). Although this protocol can yield up to approximately 40% of neuron-like cells, this protocol seems to produce a heterogeneous culture of cells such as mesoderm derivatives (Cai and Grabel, 2007). Another drawback from this protocol is the requirement of up to 2 weeks culturing. However, Ying and colleagues (2003b) developed a protocol to direct the neural lineage from feeder-independent ESCs under serum-free conditions in adherent monolayer culture. This protocol, which used cell line 46C that expresses green fluorescent protein (GFP) under the control of the promoter for an early neuroectodermal marker gene, *Sox1,* can yield up to 75% Sox1-positive cells by day 5. In this protocol, the ESCs were plated on gelatine-coated dishes in absence of LIF and serum-free defined medium (N2B27). The NSCs (Sox1-GFP$^+$ cells) that are generated by monolayer differentiation form rosettes-structure in which cells elongate and align radially similar to RG cells in the neural tube. In addition, these Sox1-GFP$^+$ cells can be directed to all cell types in the neural fate.

1.3 Epigenetic factors regulate the self-renewal and neural differentiation of ESCs

Although the self-renewal and pluripotency of ESCs and neural differentiation are regulated by extrinsic and intrinsic factors such as pluripotency- or neural-related genes, chromatin structure and epigenetic factors are also be involved in the regulation of these processes. Precise information for determining cellular function, identity, and morphology is provided by the regulation of gene expression via epigenetic factors. Epigenetic mechanisms have also been shown to be vital during all the developmental stages of human life because they are responsible for forming multi-cellular organisms (reviewed by Olynik and Rastegar, 2012). The term epigenetics, which was first coined by Waddington in 1942, can be defined as the alterations that occur above or on the DNA; these alterations modify the chromatin organisation and thereby regulate gene expression (Waddington, 2012).

The basic repeat unit of chromatin is called a nucleosome, and is formed of DNA and two molecules of each core histone, H2A, H2B, H3 and H4 (reviewed by Nemeth and Langst, 2004; Vastenhouw et al., 2012). Each histone is composed of a central globular domain that lies within the nucleosome, and the C- and N-terminal tail domains that are subject to different post-translation modifications. The 10 nm nucleosomal array fibre is folded into a higher-order 30 nm structure; based on the compaction it can be either euchromatin or heterochromatin (open or densely compacted structure, respectively) (reviewed by Nemeth and Langst, 2004; Atkinson and Armstrong, 2008). The histone tails and DNA are subjected to several chemical changes such as methylation and acetylation that subsequently alter the chromatin structure and/or gene expression (reviewed in Olynik and Rastegar, 2012). These changes on the DNA or histone tails are reversible and are known as epigenetic marks. Chromatin structure and epigenetic factors play roles in different cellular function and during human development. Epigenetic mechanisms include ATP-dependent chromatin remodelling, DNA methylation and histone modification.

1.3.1 ATP-dependent chromatin remodelling

Gene expression can be affected by modifying the chromatin architecture or DNA accessibility at a specific gene or its regulatory regions. Nucleosome repositioning, or the exchange of histone variants, can be accomplished by destabilising the interaction between DNA and histone proteins by a protein complex containing an ATPase subunit (Vignali et al., 2000). This is called ATP-dependent chromatin remodelling. There are several of these complexes, for example, the switching/source nonfermenting (SW12/SNF2) complex, and its human analogue known as BAF that contains two distinct ATPase subunits BRM (Brahma) and BRG1(Brhama—related gene1) (Wang et al., 2004). Chromatin remodelling is important for both gene activation and repression, and some chromatin remodelling complexes are cell-type specific and developmental-stage specific (reviewed in Ho and Crabtree, 2010).

1.3.2 DNA methylation

DNA methylation refers to the addition of a methyl group at position 5 of the pyrimidine ring of a cytosine residue within a CpG dinucleotide, forming 5-methylcytosine (5mC). The addition of methylation group into DNA is mediated by DNA methyltransferases (DNMTs). *De novo* DNA methylation in early development is catalysed by DNMT3a and DNTMT3b, while DNA methylation during later development and adulthood is maintained by DNMT1 that can methylate the hemi-methylated CpG site on one DNA strand (Sikorska et al., 2008). Transcriptional repression can be caused by DNA methylation when CpG dinucleotides in a promoter region are highly methylated, suggesting the density of DNA methylation is critical (Deb-Rinker et al., 2005; Marei et al., 2012). DNA methylation can repress gene expression either by preventing the binding of crucial TFs, or by recruiting repressive factors via methyl-CpG binding domain (MBD) proteins such as MBD1 and MeCP2 (Khavari et al., 2010; Ariff et al., 2012).

In contrast, DNA demethylation can be achieved by suppressing DNMT1, which leads to loss of DNA methylation during successive rounds of replication, or by enzyme that deletes the methyl group without a need of replication process. Recently, ten-eleven translocation (TET) proteins which include TET1, TET2 and TET3 were discovered as DNA demethylases (Tahiliani et al., 2009; Ito et al.,

2010). These proteins can catalyse the hydroxylation of 5mC to 5-hydroxymethylcytosine (5hmC), and the oxidation of 5hmC to 5-forylcytocine and 5-carboxlcytosine (5fC and 5caC, respectively) (reviewed in Li et al., 2015). 5fC and 5caC can be further processed by thymine-DNA glycosylase followed by base excision repair, and this is known as active DNA demethylation. However, 5hmC cannot be recognised by DNMT1 during DNA replication (reviewed in Li et al., 2015). Thus, TET protein-mediated 5mC oxidation can regulate global or locus-specific 5mC and/or 5hmC levels dynamically by facilitating both active DNA demethylation and passive DNA methylation dilution, thereby modulating gene expression.

1.3.3 Histone acetylation and methylation marks

Many of the lysine residues in the N-terminal core histone tails can be modified by processes such as acetylation or methylation. Histone acetylation tends to be associated with transcriptional activation and is catalysed by histone acetyltransferases (HATs) such as P300/CBP, while histone deacetylation is catalysed by histone deacetylases (HDACs) (reviewed by Olynik and Rastegar, 2012). Acetylation of H3 and H4 tails reduces the compaction of chromatin *in vitro* (Garcia-Ramirez et al., 1995; Tse et al., 1998) and acetylated chromatin is more accessible to interacting proteins *in vivo*, as showed by the increase in sensitivity to nucleases (Hebbes et al., 1994; Krajewski and Becker, 1998). The change in chromatin structure due to histone acetylation is probably caused by neutralising the positive charge on the lysine residues, that in turn weakens the interaction between the histone tails and the negatively charged linker DNA (Zupkovitz et al., 2006). In addition, histone acetylation can function as binding site for proteins which can directly or indirectly regulate transcription. For example, acetylated lysine can recruit co-activator with homeodomains, such as CBP/p300, that then promote transcription (reviewed by Josling et al., 2012; Chan and Thangue, 2001).

On the other hand, histone methyltransferases (HMTs) and histone demethylases (HDMs) are responsible for adding and eliminating the histone methylation marks, respectively (reviewed by Hu et al., 2012). Unlike histone acetylation, lysine residues can be mono, di, or trimethylated. In addition, histone methylation marks can be either permissive or repressive marks. For instance,

histone H3 trimethylation at lysine 4 (H3K4me3, promoter mark) and H3K36me3 (gene body mark) are considered as activation marks for transcription; in contrast H3K9me or H3K27me are repressive marks and are usually associated with constitutive and facultative heterochromatin, respectively (Golebiewska et al., 2009).

The histone methylation marks H3K4me3 and H3K27me3 are mediated by trithorax group (trxG) and polycomb group (PcG) proteins, respectively. PcG and trxG function in distinct multiprotein complexes. PcG proteins are mainly found in two complexes known as polycomb repressive complex 1 and 2 (PRC1 and PRC2). Gene silencing by PcG is mediated initially by the recruitment of PRC2, and the SET domain in the zeste homologue 2 (Ezh2) subunit of PRC2 then tri-methylates H3K27, which serves as a docking site for PRC1. The RING1B and BMI1 subunits of PRC1 catalyse ubiquitination of H2AK119. These changes occurring at the promoter are thought to silence gene expression by impeding the elongation of RNA polymerase or by inhibiting the transcriptional activator mark, H3K4me3 (Ku et al., 2008; Wang et al., 2009; Schmitges et al., 2011). On the other hand, H3K4me3 can be mediated by trxG proteins such as Ash1 and TRX that also can antagonise the recruitment of H3K27me3, indicating crosstalk between permissive and suppressive marks. Although TrxG and PcG antagonise the recruitment of H3K27me3 and H3K4me3 marks, respectively, at specific sites, several thousands of the H3K4me3-enriched promoters in ESCs also carry H3K27me3, a situation referred to a bivalent domain (Bernstein et al., 2006). Recently, it was identified that another histone methyltransferase called mixed-lineage leukaemia 2 (MLL2) is responsible for depositing H3K4me3 at bivalent promoters, and loss of MLL2 eliminates H3K4me3 specifically on bivalent promoters (Hu et al., 2013; Denissov et al., 2014).

1.3.4 The role of epigenetics in ESCs self-renewal and neural differentiation

1.3.4.1 Chromatin structure and nucleosome remodelling

Analysis of the differences between pluripotent stem cells and differentiated somatic cells revealed that the chromatin structure is distinct in stem and progenitor cells. The chromatin in stem/progenitor cells is typically more open

than in differentiated cells (Spangrude et al., 1988). Indeed, it was shown that ESCs have fewer DAPI dense foci than differentiated cells, indicating a lower abundance of constitutive heterochromatin (Gaspar-Maia et al., 2011). However, the differentiation of ESCs requires a progressive loss of cellular potential that indicates the augmentation of epigenetic restriction (Gifford et al., 2013). The global chromatin compaction was observed to be less accessible upon the induction of differentiation by RA, since the chromatin of induced ESCs is less sensitive to nucleases such as micrococcal nuclease (MNase) (Schaniel et al., 2009). In the same study, the foci of H3K9me3, which is a histone mark required for recruitment of heterochromatin protein 1 (HP1) and establishment and maintenance of heterochromatin, were compared in mESCs and RA-treated mESCs nuclei. They found that the number of H3K9me3 foci is dramatically increased upon RA treatment, suggesting that differentiated cells have considerably higher heterochromatin (Schaniel et al., 2009).

The role of ATP-dependent chromatin remodelling is crucial during embryo development and neural differentiation. It was shown that members of a SWI/SNF sub-complex, specifically Smarcc1/Baf155, are necessary to repress Nanog expression during mESC differentiation (Schaniel et al., 2009). Moreover, a switch from BAF45a and BAF53a subunits to the homologous BAF45b, BAF45c and BAF53b subunits during the transition from the proliferative neural progenitor cells (NPCs) to postmitotic neurons, respectively, is important for proper neuronal differentiation (Lessard et al., 2007). It was also reported by Wu et al. (2009) that the chromatin remodelling enzyme, Brm, in a combination with certain histone modifications, regulates the *Ngn1* gene which encodes to a basic-helix-loop-helix protein involved in the neurogenesis. The H3K27me3 repressive mark is lost while H3K9/14ac and H3K4me3 are gained upon RA induction. As consequence, the recruitment of Brm drives the expression of *Ngn1*.

1.3.4.2 Role of DNA methylation

DNA methylation is essential during embryo development. Although the DNA of the totipotent zygote is not globally methylated apart from the imprinted genes, during subsequent cell cleavage the DNA is selectively and gradually remethylated by DNMT3 (reviewed by Bloushtain-Qimron et al., 2009; Isagawa et

al., 2011). In ESCs, the level of DNA methylation is consistent with the status of gene transcription. For example, in ESCs the CpG islands at promoters of pluripotency-associated genes such as *Oct4* and *Nanog* genes are hypomethylated, while they become hypermethylated in NSCs or more committed cells (Deb-Rinker et al., 2005; Wang et al., 2009; Cho et al., 2011). Conversely, the promoters of tissue-specific genes such as dimethylarginine dimethylaminohydrolase, which is an enzyme involving in switching the fate of NSCs from neurogenesis towards astrogliogenesis, are methylated in ESCs (Fouse et al., 2008; Isagawa et al., 2011; Bäckdahl et al., 2009; Cortese et al., 2011). Nevertheless, DNA methylation is not required for maintaining the self-renewal of mESCs, as mESCs that lack of all three DNA methyltransferases (DNMT1/3a/3b) were able to grow in the undifferentiated state despite the absence of CpG methylation (Tsumura et al., 2006). The triple DNMTs knockout ESCs exhibited defects in differentiation however. Similarly, Dnmt1 and Dnmt3b knockout mice are embryonic lethal, while *Dnmt3a*$^{-/-}$ mice display stunted growth and neonatal mortality (Li et al., 1992; Okano et al., 1999). The results of ESCs and mouse knockout studies highlight the importance of DNA methylation for embryonic development and cellular differentiation. Consistent with these result, the level of DNMT enzymes was found to increase upon the induction of NSC differentiation (Singh et al., 2009).

1.3.4.3 Role of histone acetylation

Histone acetylation is related to transcriptionally active domains and its levels correlate with gene expression. The promoters of pluripotent-associated genes such as *Oct4* and *Nanog* are enriched by H3 and H4 acetylation marks, but the acylation marks are lost during induction and this is consistent with their expression (Hattori et al., 2004; Kimura et al., 2004; Markowetz et al., 2010). Not only is histone acetylation important in pluripotent cells (ESCs and NSCs), HDACs were also shown to be important for repressing the alternative lineage-determination genes when the cells commit to a specific lineage. For instance, the transcription repressor factor known as repressor RE1-silencing transcription factor (REST) binds to repressor element 1 (RE1) elements at the neuronal-specific genes in non-neuronal cells and ESCs, and recruits two HDAC-containing corepressor complexes in order to suppress these genes (Ballas et al., 2001; Bruce et al., 2004; Hsieh et al., 2004). The global level of histone acetylation

marks such as H3K9ac was shown to decrease during ESCs differentiation (Golebiewska et al., 2009; Saraiva et al., 2010; Cho et al., 2011). According to Lee et al (2003), inhibiting global histone deacetylation by treatment with trichostatin A, which is a reversible HDAC inhibitor that targets most class I and II HDACs, prevents ESCs differentiation. Moreover, inhibition of histone deacetylases by valproic acid in adult NPCs induced differentiation into neuronal lineage but inhibited astrocyte or oligodendrocyte differentiation (Hsieh et al., 2004). The induction of neurogenic potential as result of HDACi treatment was mediated by up-regulation of the neurogenic basic helix-loop-helix transcription factor NeuroD. HDAC1 and 2 were seen to be expressed distinctly in neurons and glial cells in which HDAC1 is expressed in NS/NP cells and glia while HDAC2 is initiated in neural progenitors and is up-regulated in post-mitotic neuroblasts and neurons, but not in fully differentiated glia (MacDonald and Roskams, 2008). Overall, histone acetylation and deacetylation are crucial for gene regulation during neural differentiation.

1.3.4.4 Role of histone methylation and bivalent domains

Histone methylation marks at the promoter regions of pluripotency-related genes are correlated with transcription status. For example, the *Oct4* gene promoter in ESCs is marked by the permissive methylation mark (H3K4me3) but not by repressive marks (H3k27me3 and H3K9me2/3), which were found to be highly enriched at same region in differentiated cells (Deb-Rinker et al., 2005; Wang et al., 2009; Cho et al., 2011). This is consistent with the findings from genome-wide analysis that indicates differentiated cells have large regions enriched by H3K27me3 and H3K9me3 compared to ESCs, while the gene body mark H3K36me3 is comparable between differentiated cells and ESCs (Hawkins et al., 2010). However, some developmental regulatory genes such as *Pax6*, which is key a regulatory factor of eye and brain development, are silenced in ESCs by PRC2 and H3K27me3 (Lee et al., 2006). These genes are selectively activated upon the differentiation, and this requires demethylation of H3K27me3 by Jmjd3, and a decrease in PRC2 binding at promoter (Burgold et al., 2008; Wang et al., 2009). The interplay between histone methylation marks, H3K9me3 and H3K4me3, during NSCs differentiation can also be seen during the shift from neuronal differentiation to astrocyte differentiation. FGF2 signalling can result in the removal of H3K9me3 and the addition of H3K4me3 around the STAT-

binding sites on the *GFAP* (glial fibrillary acidic protein) and *S100-B* genes that are expressed in astrocytes, which, in turn, allows STAT3 to access to these sites (Song and Ghosh, 2004).

Many developmental-regulatory genes in ESCs have been shown to be regulated by bivalent domains, in which the active H3K4me3 and repressive H3K27me3 marks are found together at gene promoters (Bernstein et al., 2006). Because both marks are present at the same time, these genes are poised in ESCs but are ready for rapid activation or silencing during differentiation (Bernstein et al., 2006). For instance, commitment of NSCs into neurons requires the further inhibition of poised astrocyte-developmental genes that are marked by bivalent domains. Therefore, the H3K4me3 of bivalent domains at the promoters of the astrocyte lineage genes is lost, while H3K27me3 is maintained. The silencing of these promoters is enhanced by H3K9me2 and H3K9me3 (Golebiewska et al., 2009). However, when the NSCs differentiate into astrocytes, the promoters of these genes increase the H3K4me3 mark but lose the H3K27me3 mark. The removal of histone H3K27me3 from these bivalent domains was shown to be mediated by the H3K27me3-specific demethylase, Jmjd3 (Burgold et al., 2008).

Methylation of H3K27 and H3K4 can be catalysed by PcG and trxG complexes, respectively, and these multiprotein complexes are important for ESCs self-renewal and differentiation. Using the depletion system (RNAi) and chromatin immunoprecipitation followed by sequencing (ChIP-Seq) revealed that MLL2 specifically deposits H3K4me3 at bivalent domains, whereas another complex, Set1/2, catalyses H3K4me3 at the transcription start sites of active gene in ESCs (Clouaire et al., 2012; Hu et al., 2013). Wdr5 and Ash21 are subunits found in all Set1A/B and MLL complexes. Consistent with their expression being down-regulated upon ESCs differentiation, depletion of WDR5 or ASH2L result in down-regulation of pluripotent genes and loss of ESCs self-renewal and pluripotency (Ang et al., 2011; Wan et al., 2013). WDR5 knockout reduces the global level of H3K4me3 (Ang et al., 2011). However, knockdown of another core subunit of all Set1A/B and MLL complexes, DPY-30, has little effect on self-renewal but results in improper ESC neuronal differentiation (Jiang et al., 2011). Interestingly, depletion of DPY-30 alters the H3K4me3 specifically at bivalent domains in ESCs (Jiang et al., 2011), proposing a selective developmentally related function of

this subunit. These studies emphasise the vital role of H3K4me3 and related enzymes in maintaining and differentiating pluripotent ESCs.

Similarly, PcG complexes can affect the self-renewal or differentiation of ESCs. Mutation of the individual subunits Eed, Suz12 or Ezh2 of PRC2 results in normal ESC self-renewal but defective differentiation capacity (Pasini et al., 2004; Pasini et al., 2007; Wang et al., 2002; Shen et al., 2008; Leeb et al., 2010). Similarly, depletion of PRC1 components such as RING1B and BMI1 also impairs proper differentiation (Leeb and Wultz, 2007; Van der Lugt et al., 1994; Aloia et al., 2013). Simultaneous depletion of RING1B and EED in ESCs provokes a stronger inclination toward differentiation, though self-renewal can still be sustained under careful culture conditions, and prolonged differentiation results in cell death (Leeb et al., 2010). Taken together, the epigenetic factors that regulate chromatin structure and gene expression play vital role in maintaining or inducing the neural differentiation.

1.4 Chromatin architectural proteins: the HMG superfamily

As mentioned earlier, chromatin structure plays a vital role in regulating the accessibility of genomic DNA. The dynamic nature of the chromatin, which is caused by the interplay between histone and non-histone architectural proteins and DNA, is important in regulating many cellular processes. Chromatin architectural proteins which bind to chromatin and change its configuration and accessibility include linker histones eg histone H1 and the high mobility group (HMG) protein superfamily (reviewed in Postnikov and Bustin, 2016). Linker histones bind to nucleosomes close to entry-exit site of linker DNA and promote chromatin compaction (reviewed by Woodcock et al., 2006). However, the HMG proteins, which were discovered in mammalian cells more than 40 years ago, promote chromatin decondensation. Furthermore, HMG proteins have been shown to be involved in various biological DNA-dependent processes, such as the regulation of DNA replication, DNA repair, and gene expression (Bustin, 1999; Ueda et al., 2009). Although all proteins in the HMG superfamily can modulate chromatin structure, the HMG superfamily is subdivided into three families based on their functional motif. These include HMG-AT hook (HMGA family), HMG-box (HMGB family) and HMG-nucleosome binding (HMGN family) (Bustin, 2001). The

first two families bind in the minor groove of DNA with no preferential DNA sequence, while HMGN proteins are able to bind directly to the core nucleosome particle (reviewed by Reeves, 2010). This section will focus on the HMGN family.

1.4.1 HMGN proteins

The HMGN family has 6 members: HMGN1 (96 amino acids, aa), HMGN2 (90 aa), HMGN3a (99 aa), HMGN3b (77 aa), HMGN4 (92 aa), and HMGN5 (406 aa) (Bustin, 2001; West et al., 2001; Furusawa and Cherukuri, 2010). All of them have a similar structure, which is composed of a bipartite nuclear localisation signal (NLS), a nucleosome binding domain (NBD) and lastly a negatively charged C-terminal domain known as the chromatin unfolding domain (CHUD) or regulatory domain RD) (Figure 1.2). However, the shorter HMGN3b splice form lacks the CHUD, while HMGN5 has very long CHUD (Figure 1.2). The major functional motif of these proteins is the NBD, which is highly conserved and positively charged and has 30 amino acid residues (Cripps et al., 1992). This domain confers a distinct property to this group. The NBD binds to the core nucleosome without any DNA sequence preference, and each nucleosome can be bound by two HMGN molecules (Postnikov et al., 1995; Bustin, 2001b; West et al., 2003). The NBD contains a unique highly conserved peptide motif (RRSARLSA), which is thought to be the core region for binding to nucleosome (Figure 1.2). A mutation in this conserved motif can abolish the binding of HMGN to the nucleosome, while a point mutation or deletion in other domains only decreases the binding affinity (Ueda et al., 2008). These proteins are found in vertebrates but not in lower eukaryotes such as yeast or *Drosophila*.

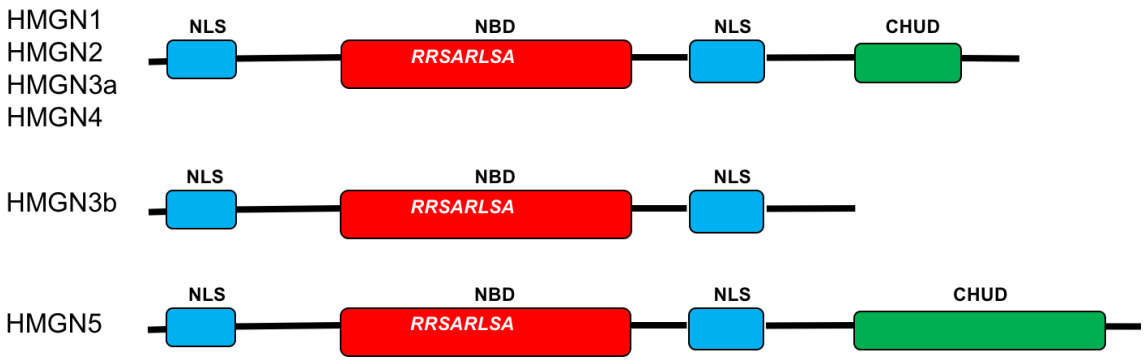


Figure 1.2: Schematic diagram of HMGN protein variants.
HMGN proteins have three functional domains, which are the bipartite nuclear localization signal (NLS, blue), a nucleosome binding domain (NBD, red), and a C-terminal chromatin unfolding domain (CHUD, green). Schematic diagrams of HMGN variants. HMGN1, 2, 3a, and 4 have very similar structures but different numbers of amino acids. However, HMGN3b and HMGN5 are different. The former lacks a C-terminal domain, but HMGN5 has very long C-terminal domain (an acidic tail) that contains negatively charged amino acid repeats (Bustin, 2001; Furusawa and Cherukuri, 2010; West et al., 2004a).

1.4.2 Interaction of HMGNs with nucleosome/chromatin

Because the function of HMGNs relies on their binding to the chromatin, the interaction of HMGNs with nucleosomes has been investigated *in vitro* and *in vivo* by several studies. The interaction of HMGN with nucleosome via the NBD was investigated using different synthesised peptides. Only peptides that contained a complete, intact NBD were able to bind nucleosomes with properties to the full-length HMGN proteins (Crippa et al., 1992; Postnikov et al., 1994). However, *in vivo* and *in vitro* analysis using mutagenesis, quantitative fluorescence recovery after photobleaching (FRAP), fluorescence imaging and mobility shift assay indicate that although the RRSARLSA motif is the main determinant of specific interaction with nucleosomes, mutations at several regions influence the affinity binding of HMGN with nucleosomes (Ueda et al., 2008). Each nucleosome is bound by two HMGN molecules, as shown by mobility shift assays under physiological conditions (Crippa et al, 1992). In addition, HMGN1 and 2 proteins were found to bind *in vivo* and *in vitro* to nucleosomes as homodimer complexes, which means either two molecules of only HMGN1 or HMGN2 can bind to the same nucleosome (Postnikov et al., 1995; Kato et al., 2011). During mitosis, the two conserved serine residues in the core NBD of HMGNs are phosphorylated, leading to the release of HMGNs from nucleosomes (Prymakowska-Bosak et al., 2001).

Several studies also investigated in detail the exact interaction between HMGN and nucleosomes using different methods. These studies concluded that the conserved motif of the NBD binds to nucleosomes at the acidic patch in the H2A-H2B dimer, while the C-terminal end of the NBD contacts the DNA close to the entry-exit site of linker DNA. The long C-terminal tail of HMGN is extended near the linker DNA and the dyad axis, and is in close proximity to the N-terminal of histone H3 (Shick et al., 1985; Alfonso et al., 1994; Trieschmann et al., 1998; Kato et al., 2011). These studies also provide insight into the mechanism by which HMGNs could modulate chromatin compaction and promote transcriptional activation, as will be discussed in the next section. The interaction of HMGNs with chromatin is highly dynamic as shown by FRAP analysis; they bind to nucleosomes for a short time and are constantly moving from one nucleosome to another (Catez et al., 2004; Ueda et al., 2009). The average binding time lasts for several seconds and is longer than the period of movement; therefore, most of the time, the HMG molecules are associated with the chromatin. One percent of nucleosomes in a mammalian cell can be bound by HMGNs at any one time, but because they move continuously and rapidly, HMGN proteins can bind to an entire genome in less than 1 minute (Catez et al., 2006). Although most of HMGN members seem to bind to all nucleosomes in the genome, HMGN5, which has a long acidic C-terminal domain, is preferably associated with euchromatin rather than heterochromatic regions (Rochman et al., 2009).

1.4.3 The molecular mechanisms of HMGNs

HMGN proteins function by modulating the structure and activity of chromatin. In particular, they promote an unfolding of higher-order chromatin structure (Furusawa and Cherukuri, 2010; Rochman et al., 2011; Barkess et al., 2012). HMGNs are responsible the formation of more accessible chromatin that in turn enhances the gene transcription, DNA replication and repair. Since the function of HMGNs relies on their interaction with nucleosome, the nature of this interaction is critical. The nucleosomal binding site of HMGNs overlaps with the linker histone binding site (Shick et al., 1985; Alfonso et al., 1994; Trieschmann et al., 1998; Kato et al., 2011). In addition, although linker histone, H1, increases the chromatin compaction, HMGNs act to reduce chromatin folding (West et al., 2003). The contrasting functions and overlapping site of interaction suggest that the binding of H1 to the nucleosome may be counteracted by the

binding of HMGNs (Catez et al., 2004). Indeed, HMGN1 antagonised the chromatin compaction and transcriptional inhibition induced by H1 in SV40 minichromosomes (Ding et al., 1997). In addition, by using FRAP analysis to monitoring the binding of GFP-H1 to chromatin in living cells, it was shown that microinjection of HMGN1 or HMGN2 reduces the residence time of H1; this indicates competition between the two proteins *in vivo* (Catez et al., 2002; Catez et al., 2004). However, it was observed that HMGNs also caused decondensation of nucleosomal arrays lacking H1 (Vestner et al 1998; Trieschmann et al., 1995), which suggests another mechanism may also be involved.

HMGN1/2 were found to increase nucleosome stability as shown by thermal denaturation, circular dichroism and nuclease digestion (Crippa et al., 1992; Bustin, 2001; Shimahara et al., 2013). It is possible that increased nucleosome stability because of HMGN binding would inhibit nucleosome remodelling. Although HMGN1 was not shown to have an effect on ATP-dependent chromatin remodelling by the SWI/SNF complex (Hill et al., 2005), this was contradicted by another study conducted by Rattner and colleagues (2009). They observed that HMGN1 and HMGN2 repress the ATP-dependent nucleosome remodelling via ACF which has an ISWI ATPase subunit, as well as by BRG1, which is the ATPase subunit of some human SWI/SNF complexes. This repression is mediated by reducing the interaction of ACF with nucleosomes. However, the inhibition of ATP-dependent nucleosome remodelling was not observed with mutant HMGN proteins that cannot bind to nucleosomes (Rattner et al., 2009).

In addition to the ability of HMGNs to compete with H1 and inhibit chromatin remodelling, HMGN proteins can modulate the level of histone modifications (Lim et al., 2005; Rochman et al., 2011; Barkess et al., 2012). Previously, it was observed that HMGN1 lacking the CHUD was unable to induce chromatin unfolding (Trieschman et al., 1995). By examining the interaction of HMGNs with nucleosomes, the C-terminal of HMGN proteins was found to be located near to the N-terminal tail of histone H3. Therefore, the histone H3 post-translational modifications could be directly affected by this interaction (Trieschmann et al., 1998; Kato et al., 2011). Lim and colleagues (2004) first demonstrated that HMGN1 can modulate histone modifications; specifically, the loss of HMGN1

increased the rate of anisomycin-induced phosphorylation of H3S10 and altered the induction of a subset of immediate-early genes in mouse embryonic fibroblasts (MEFs). In addition, the acetylation level of histone 3 at lysine 14 was lower in HMGN1$^{-/-}$ compared to HMGN1$^{+/+}$ MEFs, and HMGN1 was shown to increase the level of H3K14ac by enhancing the action of histone acetyltransferases (HAT) *in vivo* and PCAF *in vitro* (Lim et al., 2005). The increase in histone acetylation level may lead to the enhanced expression of certain genes, as it well-known that histone acetylation at regulatory regions is correlated with gene expression and unfolded chromatin. Therefore, HMGN proteins can enhance the open structure of chromatin, either by competing with linker histones, inhibiting ATP-dependent chromatin remodelling, or modulating the level of post-translational modifications.

1.4.4 The effect of HMGN proteins on gene transcription

The effect of HMGN proteins on gene expression was initially suggested by the observation that DNase I hypersensitive sites appeared to be associated with HMGN1/2 proteins (Weisbrod and Weintraub, 1979). Moreover, data indicated that nucleosomes from actively transcribed genes are preferentially bound by HMGN1/2, and that most of the acetylated chromatin is bound by HMGN2 (Malik et al., 1984; Postnikov et al., 1991). Several studies subsequently investigated the effect of HMGN1/2 on the transcription using nucleosomes assembled using cell-free extract of *Xenopus* eggs or *Drosophila* embryos, or SV40 minichromosomes (Crippa et al, 1993; Ding et al., 1994; Paranjape et al., 1995; Ding et al., 1997; Weigmann et al., 1997). These studies concluded that incorporation of HMGN1/2 into *in vitro*-assembled or *in vivo*-preassembled chromatin enhances the transcription, but only in the chromatin context and not from naked DNA templates. In addition, some studies suggested that the enhancement of transcription by HMGN1/2 is because they enhance the elongation of RNA polymerase II or III enzymes (Ding et al., 1994; Ding et al., 1997), while some found that HMGN2 enhances the rate of transcriptional initiation (Paranjape et al., 1995).

1.4.5 HMGN proteins bind at genomic regulatory sites

Recent genome-wide mapping studies have linked HMGN proteins to DNase I hypersensitive sites (DHSs, a hallmark of chromatin regulatory sites) and transcriptionally active regions in different tissues (Cuddapah et al., 2011; Deng et al., 2015; Zhang et al., 2016). ChIP-seq for HMGN1 in human $CD4^+$ T cells indicates that the HMGN1 is enriched at regulatory sites including DHSs, active promoters, functional enhancers, and transcription factor binding sites (Cuddapah et al., 2011). Similarly, HMGN1 binding sites overlap with DHS sites in *in-vitro* derived-NPCs at gene promoters (higher enrichment at promoters containing CpG islands). Loss of HMGN1 causes a reduction in the total number of DHS sites in ESCs and NPCs, and most of retained-DHS signals are at the promoters containing CpG islands (Deng et al., 2013). However, HMGN1 and HMGN2 were found to perform synergistically to maintain the DHS landscape in MEFs, and knockout of both proteins was required to significantly affect the distribution and intensity of DHS sites (Deng et al., 2015). From all these different studies, it appears that HMGN proteins are highly enriched at DHS sites; disregulation of HMGNs affects the DHSs landscape, particularly at the regulatory sites, and this in turn, affects the maintenance of the transcription profile.

Most of the HMGN1 binding sites that are preferentially enriched at CpG island promoters also overlap with H3K4me3 enrichments in *in vitro* derived-NPCs. In addition, the level of HMGN1 binding at these promoters is correlated with gene expression status (Deng et al., 2013). Using MNase-Seq (micrococcal nuclease digests followed by deep sequencing of DNA) it was shown that the loss of HMGN proteins stabilises the nucleosome positioning at transcription start sites in NPCs and MEFs (Deng et al., 2013, Deng et al., 2015). However, this is contradictory with the previous observations that binding of HMGNs to nucleosomes increases nucleosome stability (Crippa et al., 1992; Bustin, 2001). Taken together, the data suggest that HMGN proteins are crucial for maintaining DHSs at promoters and enhancers, and for appropriate nucleosome positioning at TSSs.

1.4.6 HMGN as a general modulator for the cellular transcription

Various studies have shown that HMGN proteins affect the gene expression and cellular phenotype (Hock et al., 2007; Rochman et al., 2011). It is arguable whether the HMGN proteins are general or specific modulators for cellular transcription. By comparing the transcriptional profile in four tissues derived from HMGN1, 3 or 5 functional knockout mice with wild-type littermates, it was found that HMGN variants are not specific regulators but they modulate the fidelity of the cellular transcriptional profile in a tissue-specific and HMGN variant-specific manner (Kugler et al., 2013). Additionally, most of the differential expressed genes in HMGN1$^{-/-}$, HMGN2$^{-/-}$ or HMGN1$^{-/-}$N2$^{-/-}$ mice in the same tissue or from four different tissues do not overlap with each other, indicating tissue and variant specificity (Deng et al., 2015). The differential effects on modulating the DHSs profiling seen among HMGN1$^{-/-}$, HMGN2$^{-/-}$ or HMGN1$^{-/-}$N2$^{-/-}$ may be in agreement with this (Deng et al., 2015). Although HMGN proteins are thought to be general transcriptional modulators on chromatin templates, they can also be specific transcription regulators of specific genes. For example, HMGN3 stimulates glycine transporter (*Glyt1*) and glucose transporter gene (*Glut2*) expressions, while HMGN1 enhances heat shock protein 70 (*Hsp70*) and estrogen-regulated genes (*Tff1* and *Fos*) (West et al., 2004; Belova et al., 2008; Zhu and Hansen, 2007; Ueda et al., 2009; Barkess et al., 2012). However, most of the transcription profiles in these studies are limited. Regardless whether HMGN proteins affect the cellular transcription profile or specific genes, the induction or inhibition of cellular transcription, or more specifically the pluripotency and developmentally regulated genes, may affect the pluripotency and differentiation potential of ES and EC cells.

1.4.7 Redundancy between HMGN proteins

The redundancy between the HMGN1 and 2 strongly indicates that these proteins share the same cellular functions. In fact, over 40% of amino acids in the two proteins are conserved, and the two protein share chemical and physical properties such as net charge, conformation and interaction with nucleosomes (Bustin et al., 1990). In addition, both can enhance the transcriptional activity only in the chromatin context (Trieschmann et al., 1995; Crippa et al., 1993). Furthermore, a recent study (Deng et al., 2015) found that 76% of HMGN2

binding sites in NPCs overlap with HMGN1 binding sites, and loss of HMGN1 results in an increase of HMGN2 binding at HMGN1-unique sites with a reduction at HMGN2-unique sites, indicating a redundancy between HMGN members. In addition, the combined loss of both proteins remodels the DHSs landscape of MEF cells, as discussed before (Deng et al., 2015). In the same study, more severe phenotypical effects were seen in HMGN1$^{-/-}$N2$^{-/-}$ mice than in mice with one variant knockout (see next section for details). All this evidence supports a shared role for these proteins in cellular function.

In contrast, some evidence indicates that HMGN proteins can function in a variant and tissue-specific manner. At the molecular level, although the net charge is identical between these proteins, the density of negative charges in the C-terminal domain differs. In addition, two molecules of either HMGN1 or 2 are bound to nucleosomes under cooperative conditions to form a homodimer complex (Postnikov et al., 1995). Moreover, some histone modifications are modulated by one variant of HMGN but not another (Lim et al. 2004; Ueda et al., 2006). The genetically altered mice for HMGN variants indicates some variant-specific phenotypes and have cellular transcriptional alterations that suggest no functional redundancy (Rochman et al., 2011; Kugler et al., 2013; Deng et al., 2015). For instance, HMGN1 knockout mice have behavioural abnormalities, increased tumourigenicity and impaired DNA repair, whereas HMGN3 knockout mice are mildly diabetic (Birger et al., 2003; Birger et al., 2005; Ueda et al., 2009). Therefore, although there is considerable redundancy between HMGN1 and 2, they also appear to have some distinct functions.

1.4.8 HMGN proteins and development

Embryonic development is controlled by the precise regulation of tissue-specific gene expression. HMGN proteins that regulate chromatin structure and cellular transcription are probably involved in the regulation of this process. The expression of HMGN proteins are developmentally regulated during mouse and *Xenopus* embryogenesis and cellular differentiation (Begum et al., 1990; Crippa et al., 1991; Lehtonen et al., 1998; Shakoori et al., 1993; Lehtonen and Lehtonen, 2001; Furusawa et al., 2006). HMGN1 and HMGN2 are ubiquitously expressed in embryonic tissues of vertebrates, particularly in undifferentiated cells but less so in adult cells. However, HMGN3 and HMGN5 proteins show

distinct developmental and tissue-specific expression (in pancreatic cells, eye and brain for HMGN3, and in placenta for HMGN5) (Ito and Bustin, 2002; West et al., 2001; Hock et al., 2007; Shirakawa et al., 2009). Examining the expression of HMGN1 during preimplantation showed that HMGN1 is present in oocytes and throughout all preimplantation stages including blastocyst stage (Deng et al., 2013). Furthermore, the expression of HMGN1 and HMGN2 should decrease during erythropoiesis, chondrogenesis, and myogenesis in order to ensure proper differentiation (Pash et al., 1993; Furusawa and Cherukuri, 2010). For example, the inhibition of myoblast differentiation via aberrant induction of *Hmgn1* expression is associated with the down-regulation of myogenic determination factors (Pash et al., 1993). The diversity in *Hmgn* expression among tissues may indicate that a distinct subset of genes is regulated by individual HMGN.

Cellular differentiation relies on changes in the gene expression, which are regulated within a temporal and spatial pattern (Sikorska et al., 2008; Furusawa and Cherukuri, 2010). Studies using knockout mice, transient depletion, and over-expression of HMGN proteins have provided evidence of their role in cellular differentiation during embryonic development (Körner et al., 2003, Mohamed et al., 2001 and Pash et al., 1993). Transient depletion of HMGNs from one- or two-cell mouse embryos slowed the progression of preimplantation development (Mohamed et al., 2001), and misregulated HMGN expression during *Xenopus* embryogenesis resulted in malformed tadpole embryos (Körner et al., 2003). This data suggests that HMGN proteins may perform important roles in ESCs and during development.

Hmgn1 knockout mice have defects in corneal maturation, increased tumourigenicity and impaired ability to repair damaged DNA. They also have behavioural abnormalities that are correlated with altered expression of *MeCP2*, a methylCpG binding protein that is highly expressed in neurons (Briger et al., 2003; Briger et al., 2005; Ueda et al., 2009; Abuhatzira et al., 2011). More recently, mice with knockouts of both HMGN1 and HMGN2 have been studied (Deng et al., 2015). In general, the double knockout mice have more severe phenotypes that the individual knockouts, including increased food intake but decreased body mass; changes in the levels of albumin, alpha amylase activity, triglycerides, cholesterol, and insulin; changes in the distribution of leukocyte

subpopulations; a reduced corneal thickness but also an enlarged eye axial length (Deng et al., 2015). Although the knockout mice do not seem to have a significant developmental impairment apart from the eye development, the expression of *Hmgns* during the developmental or adult mice brain is correlated with the neurogenesis.

1.4.9 HMGNs and neurogenesis

Based on the analysis of the mRNA expression from Allen Brain Atlas, *Hmgn1* and *2* appear to be highly expressed in zones of neurogenesis in the developing and adult mouse brain (Figure 1.3). This includes the VZ of the embryonic brain at day 11.5 and 13.5 and the VZ and SVZ at E15.5 (Katherine west, unpublished data). Similarly, using immunohistochemistry HMGN1, HMGN2 and HMGN3 were found highly expressed in the VZ and SVZ of mouse forebrain at embryonic day E18.5 (Nagao et al., 2014). In addition, Hmgn1 and 2 mRNA are also expressed in the external germinal layer (EGL) of the developing cerebellum from E15.5 through to P4 and P14, which is the period of EGL progenitor cell proliferation. Also, they are expressed in the dentate gyrus of the hippocampus and SVZ of the lateral ventricles in adult brain (P14 and P28) (Katherine west, unpublished data). Although the high expression of *Hmgn1* and *Hmgn2* in these tissues suggests they play an important role, the function of these proteins in neurogenesis during development is mostly unknown.

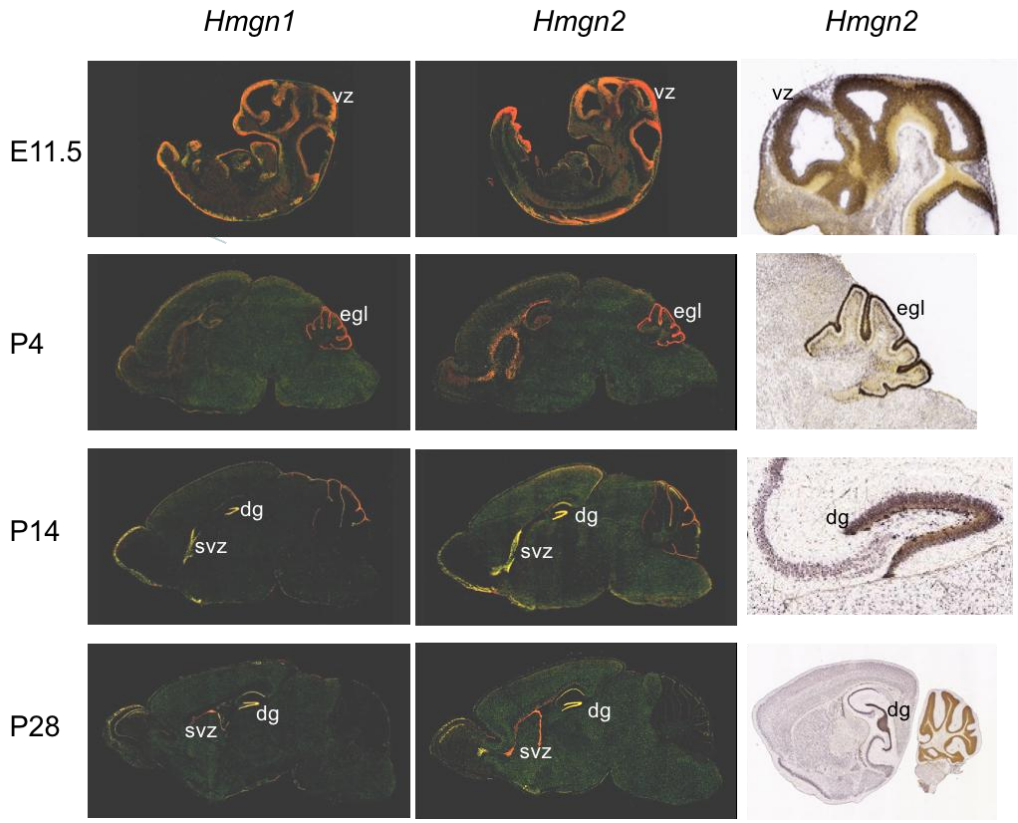


Figure 1.3: The expression of Hmgn1 and Hmgn2 in regions of neurogenesis in developing and adult mice.
In situ hybridisation (ISH) data on sagittal mouse brain sections from the Allen Brain Institute. Left and middle columns: expression masked ISH data for *Hmgn1* (left) and *Hmgn2* (middle). Right column – original *Hmgn2* ISH data, zoomed in to focus on the head (E11.5), developing cerebellum (P4) or dentate gyrus (P14). The right-hand P28 image shows a different sagittal section to illustrate the dentate gyrus more extensively. vz: ventricular zone, egl: an external germinal layer of the cerebellum, svz: subventricular zone; dg: dentate gyrus. The figures were obtained from Katherine West.

Although HMGN1 and HMGN2 can regulate several types of cellular differentiation, data about whether HMGNs influence neural differentiation is contradictory and incomplete. A study by Deng et al, (2013) revealed a reduction of nestin-positive cells in the SVZ region of *Hmgn1*$^{-/-}$ mice; however, NPCs and neurons derived from mESCs lacking HMGN1 did not have altered expression of *nestin* or *tropomyosin-related kinase* β (*Trk*β), which serve as differentiation markers for NPCs and adult neurons, respectively. Another study by Nagao et al, (2014) demonstrated that HMGN1, 2 and 3 regulate the cell fate of NPCs *in vivo* and *in vitro*. Specifically, higher expression of HMGNs stimulated the astrocytic fate, whereas lower expression promoted the neuronal lineage.

Preliminary data from our lab showed that knockdown of HMGN1/2 in P19 embryonal carcinoma cells down-regulated the expression of *Oct4*, *Nanog*, and

Sox2, which suggested that HMGNs may facilitate the maintenance of the pluripotency in stem cells (Mohan, 2012). In addition, some genes were differentially expressed when the cells were differentiated into neurons. Further research is required to investigate the role of HMGN1 and HMGN2 in ESCs and neural development. In the previous experiment, the HMGN1 or HMGN2 was knocked down by using siRNA, which reduces protein expression for a short period (Mohan, 2012). However, the study of the role of HMGNs during development is likely to require longer term silencing, and an inducible approach is desirable.

1.5 Gene editing

Genome editing or engineering can be referred to the process of manipulating the genomic DNA sequence or its outcome in a specific-targeted manner (Hsu et al., 2014; Perez-Pinera et al., 2012). The ability to make precise DNA modifications, which include either inserting, deleting or replacing DNA sequences, has advanced many aspects in life science research, including studies of genome structure, regulation and function. Targeted gene editing has been developed by the advent of a series of programmable targeted DNA-binding nucleases which include meganucleases, zing finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs). In addition to these synthetic nucleases, the most recent generation of genome editing technologies is based on RNA-guided endonucleases and this is known as clustered regularly-interspaced short palindromic repeats (CRISPR)-associated protein (Cas9). Both DNA-binding and RNA-guided nucleases are used to introduce double-strand DNA breaks (DSBs) which are a crucial step for performing targeted genome editing. DSBs-mediated by nucleases can stimulate the natural DNA repair mechanisms in cells to repair the deleterious damage by either homologous recombination (HR) or non-homologous end joining (NHEJ) (Rudin et al., 1989; Bibikova et al., 2001; Bibikova et al., 2002).

HR and NHEJ pathways are induced by DSBs and can be powerful mechanisms for genetic engineering of eukaryotic genomes (Figure 1.4). DNA repair by HR can be used for inserting specific point mutations or longer sequences into the host genome through the exchange of the targeting region with an exogenous template carrying homology sequences on both sides (Perez-Pinera et al., 2012;

Truong et al., 2014). The first HR model for repair of a DSB was identified based on the observations of transformation of linear plasmids that carried sequences homologous to yeast DNA (Orr-Weaver and Szostak, 1983; San Filippo et al., 2008). In contrast, repair by NHEJ often leads to the formation of different indel (insertion/deletion) mutations because this pathway rejoins the break via direct ligation of the both DNA ends without any needs for sequence homology (Dudas and Chovanec, 2004; Jasin and Rothstein, 2013). This error-prone repair pathway starts by recruiting the Ku70/Ku80 end-binding protein complex that activates other proteins necessary for the ligation of two ends of the damaged DNA (Perez-Pinera et al., 2012). Insertions or deletions of various lengths can be obtained with this method, which can be used to disrupt the open reading frame of a coding sequence, or the binding site of a transcription factor at a promoter.

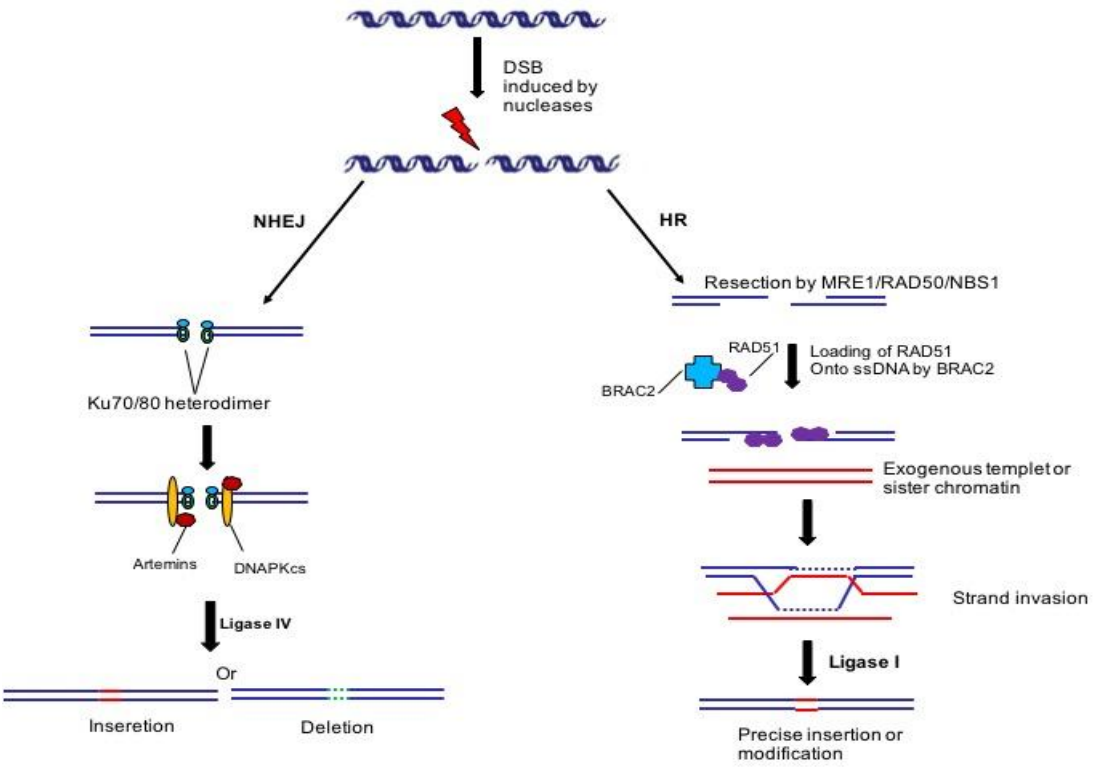


Figure 1.4: Nuclease-induced DSBs are repaired by either non-homologous end joining (NHEJ) or homologous recombination(HR).
After the nucleases induce a DSB, two major repair pathways can amend the DNA lesion: NHEJ and HR. The pathways for repairing DSBs have components involved in both signalling and repairing DNA damage. NHEJ, which is error-prone and occurs in G0/G1, is based on DNA-dependent protein kinase (DNA-PK) that is recruited by Ku70/Ku80 heterodimer. DNA-PK, in turn, recruits and activates another protein (Artemis) that is responsible for processing DNA ends before ligation via DNA ligase IV/XRCC4 complexes. On the other hand, HR, using an identical sequence such as an exogenous template or normally the sister chromatid to repair DSBs, is an error-free repair pathway and is usually used during the S and G2 phases of the cell cycle. In HR, single-stranded 3` overhangs are created at the region of DSBs by the MRN complex (Mre11/Rad50/Nbs1). HR is depended on the function of BRAC1 protein that can recruit other DNA repair proteins. BRAC2 is also involved in this pathway and important for RAD51 (nucleoprotein filament protein) localisation to the 3` overhang. RAD51 then catalyses the search for homologous target sequences, invades the sister chromatid at the site of homology, and then initiates DNA synthesis, using the homologous template. The final gaps at the end of the newly synthesized sequence are then ligated (modified from Lord et al., 2006; San Filippo et al., 2008; Postel-Vinay et al., 2012; Truong et al., 2014).

1.5.1 Meganucleases

Since the first reports of using the I-SceI meganuclease in mammalian cells, the use of nucleases to stimulate HR has been applied in different situations. Meganucleases are a family of naturally occurring restriction enzymes that are characterised by having extended DNA recognition sequences (14-40 bp) (Sander and Joung, 2014). Theses endonucleases can be classified into 5 families depending on sequence and structure motifs. The most studied family, LAGLIDADG, is a group of proteins that are capable of recognising and cleaving

the exon-exon junction sequence wherein their intron resides (Silva et al., 2011). The engineering of meganucleases is challenging as the DNA recognition and cleavage function of these endonucleases are in the same scaffold (Smith et al., 2006; Silva et al., 2011). Additionally, there is a lack of clear correspondence between the meganuclease enzyme residues and their target DNA sequence specificity, which means that they have not been widely implemented as a genome engineering tool (Hsu et al., 2014).

1.5.2 Zinc finger nucleases (ZFNs)

Although meganucleases cannot easily be used in gene editing because of their limitation in target sequences, the DNA-binding domains in ZFNs and TALENs are separated from the cleavage domains, which allows more flexibility in the engineering their DNA-binding specificity. The ZF domain, which is formed from a Cys_2-His_2 zinc finger and was firstly described in *Xenopus laveis*, is the most common DNA binding motif in human genome (Miller et al., 1985; FitzPatrick and McKay, 2015). The ZF motif is composed of 30 amino acids and has the specificity to recognise 3 nucleotides of DNA sequence (Silva et al., 2011). The ZF domain can also be engineered by different methods such as site-directed mutagenesis to target any nucleotide triplet (Pabo et al., 2001). Different combinations of customised multi-finger arrays can be assembled and fused to the cleavage domain of a type IIS restriction enzyme, *FokI*, in order to recognise long DNA sequences and induce targeted genome alterations. Although many groups developed different approaches to design and assemble multi-finger arrays (Meader et al., 2008), these approaches were not applicable for many laboratories due to the need to consider “context-dependent” ZF binding preferences because of the crosstalk between adjacent finger domains in the array (Sander and Joung, 2014). Therefore, the assembly of functional ZFNs with high specificity for a desired DNA target is still challenging because of the need for extensive ZF array screening (Hsu et al., 2014). Furthermore, the ZFNs were shown to have endonuclease activity at off-target sites (Petek et al., 2010), which is not desirable for precise gene editing.

1.5.3 Transcription activator-like effector nucleases (TALENs)

TALEs are natural DNA-binding proteins produced by plant pathogenic bacteria, *Xanthomonas,* to facilitate bacterial infection (FitzPatrick and McKay, 2015). The discovery of TALEs provides another choice for engineering DNA-binding proteins. The DNA-binding domain of TALEs is composed of repeat sequences that are 33-35 amino acids in length. Each repeat can recognise only one nucleotide, and the binding specificity is determined by the amino acids at positions 12 and 13 of each repeat (Boch et al., 2009; Perez-Pinera et al., 2012). Similar to ZFNs, the modular TALE repeats can be assembled in desired combinations for recognising specific DNA sequence, and also fused to the cleavage functional domain of *FokI* to introduce DSBs. Although the assembly of amino acid repeats is challenging due to the difficulty of cloning repeated sequences, several protocols for the rapid and effective assembly of custom TALE arrays have been described (Miller et al., 2011; Reyon et al., 2012). Since the nuclease *FokI* performs as a dimer, two customised TALEN arrays recognising two sites with a spacer sequence of 12-20 bp between them have been used successfully for altering genomes from yeast to human (Carrol, 2014). The design and assembly of TALENs is simpler and faster compared to ZFNs, but constructing the repetitive sequences of specific TALE array is still labour intensive and costly (Hsu et al., 2014; FitzPatrick and McKay, 2015). In addition, the highly repetitive sequences of TALENs may create a barrier for delivering them using lentiviral systems (Sander and Joung, 2014).

1.5.4 CRISPR/Cas9

The most recent and advanced system for gene editing is derived from a bacterial adaptive immune system and is based on an RNA-guided nuclease. It is known as clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein (Cas9). CRISPR refers to a cluster of repetitive short sequences (25-40 nucleotides) separated by variable sequences integrated from invader DNA sequences. This type of cluster was firstly observed in the *Escherichia coli* genome during the characterisation of the DNA segment containing the *iap* gene, which encodes an enzyme important for the isozyme conversion of alkaline phosphatase (Ishino et al.1987). The CRISPR sequences are found in prokaryotic genomes including bacteria and archaea. These clusters are

transcribed to RNAs that are able to direct proteins encoded by the CRISPR-associated (*Cas*) gene to cleave genome invaders (Hale et al., 2009). *Cas* genes are frequently found adjacent to the CRISPR loci (Hale et al., 2009). There are three different types of CRISPR systems which are CRISPR type I, II and III. Type I and III are characterised by having multiple Cas proteins at the CRISPR loci that are required together with RNAs to recognise and cleave the matching sequences in the invader sequences. In contrast, the type II CRISPR, which is comprehensively studied, has fewer Cas proteins and only the Cas9 enzyme within the gene cluster mediates DNA cleavage (Garneau et al., 2010).

1.5.4.1 The components of CRISPR-Cas9 type II

In the type II CRISPR system, the clustered arrays are naturally processed into two different RNAs: CRISPR RNAs (crRNAs, also called the spacer) that consist of the invading DNA sequence (called the protospacer) and sequence from the array, and trans activating CRISPR RNA (tracrRNA). The tracrRNA hybridises with crRNA through complementary sequences, and both form a complex with endonuclease Cas9 that is able to recognise and cleave the exogenous DNA (Deltcheva et al., 2011). A study by Jinek et al. (2012) demonstrated that these dual RNAs can be fused together in single guide RNA (sgRNA), and that either the programmable dual RNAs or the sgRNA can guide the Cas9 endonuclease *in vitro* to introduce DSB in the target DNA; this raised the possibility of using this system in gene targeting (Figure 1.5a). The sgRNA consists of 20 nucleotides at the 5' side that determine the DNA target site through Watson-Crick base-pairing, followed by a protospacer adjacent motif (PAM) sequence that is recognised by Cas9 (Jinek et al., 2012). The PAM sequence for *Streptococcus pyogenes* (SpCas9), which is the most commonly used system, is 5'-NGG-3', although another alternative sequence (NAG) has been reported to be less efficiently targeted by SpCas9 (Cho et al., 2013; Hsu et al.,2013; Jiang et al., 2013). SpCas9 cuts both DNA strands 3 bp upstream the PAM sequence (Figure 1.5a).

The Cas9 endonuclease protein, which is 160 kDa, cleaves the double strand via two conserved nuclease domains (HNH and RuvC). The HNH is a single domain and cleaves the complementary DNA strand, while RuvC is divided into three subdomains, containing RuvC I and RuvC II/III domains flanking a central HNH

domain (Figure 1.5b and c) (Cong et al., 2013; Hsu et al., 2014). The crystal structure of SpCas9 in a complex with sgRNA and targeted DNA showed how these domains, along with other domains such as the alpha helical recognition (REC) domain and the PAM-interacting (PI) domain, are positioned (Nishimasu et al., 2014; Jink et al., 2014). The SpCas9 structure indicated ways in which SpCas9 could be engineered for different purposes such as increasing sequence specificity and changing the specificity for the PAM sequence (Saymaker et al., 2016; Kleinstriver et al., 2016; Paquet et al., 2016). The SpCas9 gene is relatively small (around 4.1 kb) compared to TALENs which are around 6 kb, making delivery of the system more straightforward (Cho et al., 2013).

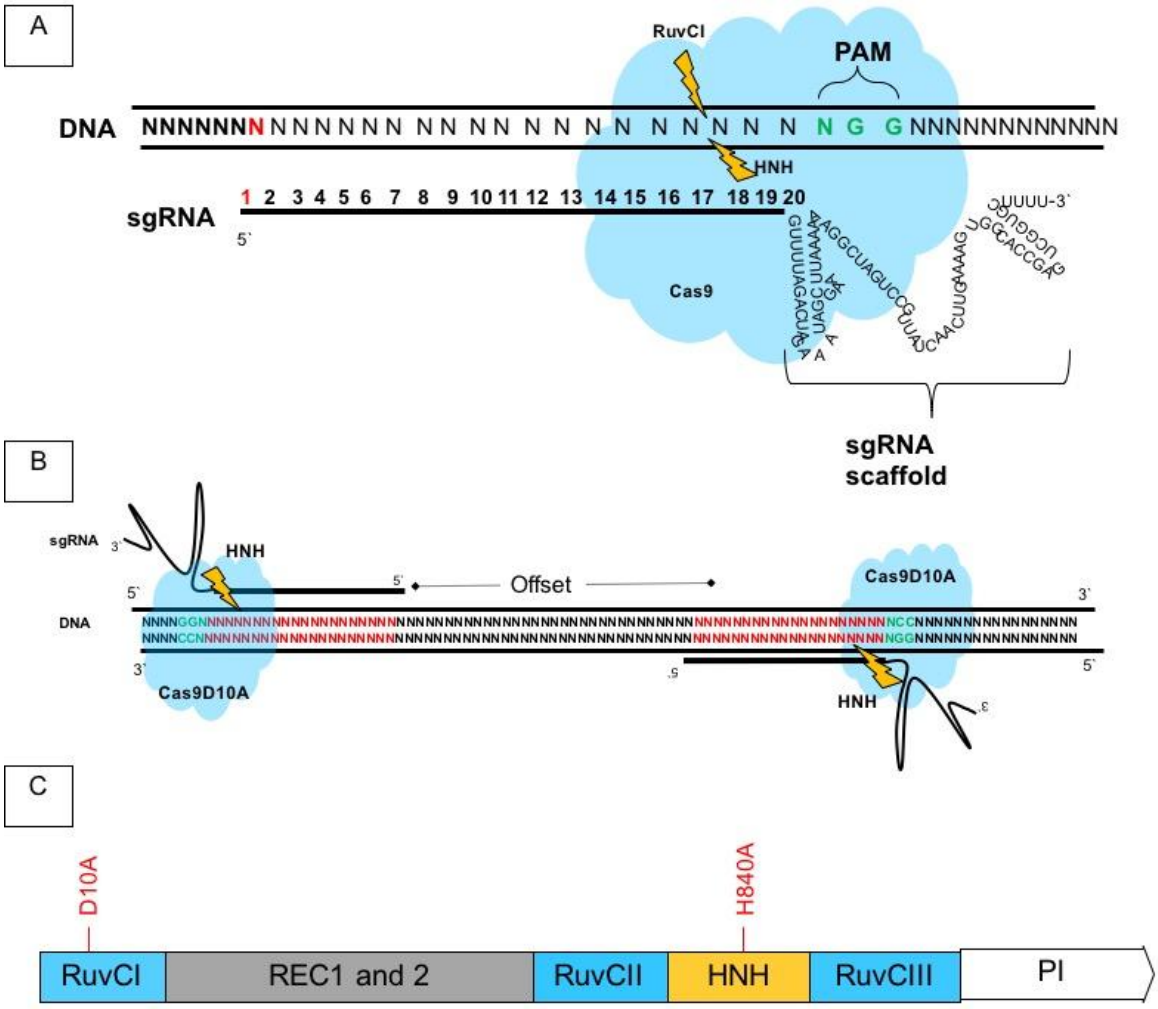


Figure 1.5: The CRISPR-Cas9 and paired nickase systems.
A-The CRISPR-Cas9 system requires at least two main elements to introduce DSBs. These are the synergetic gRNA and Cas9 protein. The sgRNA is composed of 20 nucleotides that are complementary to target site and sgRNA scaffold. The first nucleotide (number 1 in red) should be either A or G depending on the polIII promoter used to transcribe it. On the DNA, the 5'-NGG-3' (green sequences) for PAM recognition should be preceded by the target sequence. The two nuclease domains (HNH and RuvCI) of Cas9 cleave before the 3 nucleotides from the PAM sequence and generate blunt-end DSBs. **B-**One of the nuclease domains of Cas9 can be inactivated to nick only one strand of DNA, and by nicking both strands using a pair of gRNAs the DSBs repair pathway will be stimulated. A pair of Cas9D10A molecules will generate 5`-overhangs as shown in the figure. The same gRNAs can be used with Cas9H840A to form 3`-overhangs. **C-**Schematic illustration of the Streptococcus pyogenens Cas9 domain organisation. This includes the two nuclease domains which cleave the complementary strand (HNH) and non-complementary strand (RuvC). The RuvC domain is divided into 3 subdomains that are assembled together in the folded protein. There are two other domains are α-helical recognition (REC1 and 2), which form the REC lobe and facilitate the target binding, and PAM-interacting (PI) domain. The residues that can be mutated to convert the nuclease into a DNA nickase are indicated in red: substitution of aspartate to alanine at position 10 (D10A) in the RuvCI domain, and histidine to alanine at position 840 (H840A) in the HNH domain (Modified from Jinek et al., 2013; Nishimasu et al., 2014; Ran et al., 2013; Hsu et al., 2014; Sander and Joung et al., 2014).

1.5.4.2 CRISPR-Cas9 is a model for gene editing

Unlike ZFNs and TALENs that are technically challenging and time consuming to make custom nucleases, the simplicity of using CRISPR-Cas9 by only changing the

20 nucleotides of the gRNA to induce DSBs at any region of the DNA, and the ability to re-engineer SpCas9, have enabled this system to be rapidly and widely implemented as a gene editing tool. Indeed, different studies have shown that RNA-programmed SpCas9 can be efficiently used with no apparent toxicity in variety of cells and species types such as human cell lines, stem cells, bacteria, zebrafish and mouse (Cong et al., 2013; Mali et al., 2013; Cho et al., 2013; Jiang et al., 2013; Hawang et al., 2013; Wang et al., 2013; Shen et al., 2013). The efficiency of mediating cleavage at the target site varies from cell to cell and from locus to the locus in the same cell type; for example, it can reach up to 38% or 65% in human and bacteria, respectively (Mali et al., 2013; Jiang et al., 2013). Furthermore, CRISPR-Cas9 has been shown to stimulate HR as a result of DSBs at similar or greater efficiencies to TALENs (Mali et al., 2013). The use of either sgRNA (gRNA+48 bp scaffold) or crRNA that both contain the same 20 nucleotides matching the target site revealed a discrepancy in the efficiency of inducing the DSBs (Cong et al., 2013). This is probably because of the differences in the length or stability of the sgRNA. Indeed, Hsu et al., (2013) who compared the dual crRNA:tracrRNA and different lengths of sgRNA found that sgRNA with +85 scaffold mediates DNA cleavage five times higher than the corresponding crRNA:tracrRNA, because this length preserves the hairpin loop structure in the 3` of the tracrRNA. This reflects the importance of sgRNA design for optimal Cas9 activity.

The design of the CRISPR-Cas9 system this allows the use of multiple gRNAs with a single Cas9 to target different DNA sequences of interest simultaneously (Cong et al., 2013). For example, by using SpCas9 with multiple gRNAs, different indel mutations were obtained in three and five genes in rat cells and mouse embryonic stem cells, respectively (Li et al., 2013; Wang et al., 2013). Moreover, different groups have developed vector systems for multiple gene targeting, in which multiple guide sequences can be encoded into a single vector to enable simultaneous editing of several sites within the mammalian genome (Kabadi et al., 2014; Sakuma et al., 2014). Lastly, SpCas9 can be engineered to be catalytically inactive which offers the possibility of regulating gene expression by fusing it to activator or silencer factors (Mali et al., 2013b; Charpentier and Doudna, 2013). Therefore, the simplicity of CRISPR-Cas9 programming, the capacity for multiplexed target recognition, and the existence

of many natural type II CRISPR-Cas9 system variants have allowed notable developments using this technology to precisely and efficiently target and edit genomic loci of a wide range of cells and organisms. However, the major concern with this technology is that the guide RNA is a short sequence that may give rise to off-target effects.

1.5.4.3 The specificity of CRISPR-Cas9

As gene editing causes permanent alteration to the genome, targeting specificity is crucial, especially for clinical applications and gene therapy. Optimal gene editing is to just alter the intended gene or sequences without affecting other regions in the genome. Both ZFN and TALENs were found to have off-target activities and also the longer TALENs are more tolerant of mismatches (Guilinger et al., 2014; Gabriel et al., 2011). The specificity of CRISPR-Cas9 is dependent on the gRNA because the recognition of Cas9 to its target is based on the RNA-DNA interaction.

The specificity of CRISPR-Cas9 has been studied by using gRNAs bearing different mismatches to a specific site, or by predicting and testing the activity of SpCas9 at possible off-target sequences that are similar to the on-target site. The specificity of SpCas9 was thought to be dependent on the first 8-12 bases at the 3` end of the spacer (gRNA), which is named the “seed sequence”, because mismatches up to position 11 abolish genomic cleavage while mismatches in the 5` of spacer can be tolerated (Jinek et al., 2013; cho et al., 2013). However, a single mismatch at any position of the gRNA, even at the 3` end, can still direct cleavage by SpCas9 at some tested sites (Fu et al., 2013). Although not all single or double mismatches can lead to mutations, up to 5-13 mismatches can be tolerated by SpCas9 in some cases (Fu et al., 2013; Wang et al., 2015). Moreover, a mismatch at one position of the gRNA can have different effects on SpCas9 activity from one DNA target to another (Hsu et al., 2013; Fu et al 2014). Collectively, different studies concluded that multiple mismatches between sgRNA and target DNA can be tolerated depending on the quantity, position, base identity and distribution of the mismatches (Hsu et al., 2013; Fu et al., 2013; Mali et al., 2013b; Shen et al., 2014). The ability of SpCas9 to introduce indel mutations at off-target sites varies from site to another; at some off-target

sites the indel mutation frequency is 2 to 5%, while in the others the frequency is similar to that at the intended target (Hsu et al., 2013; Fu et al., 2014).

Most of the studies investigating the specificity of SpCas9 were conducted on a subset of the possible off-target sites. However, any 20 nucleotide sequence normally has hundreds to thousands of potential off-target sites with different mismatches (Fu et al., 2013; Sander and Joung, 2013). In addition, Cas9 was shown to be able to tolerate an off-target site with a one base bulge caused by skipping a single nucleotide when gRNA and DNA are paired together (Wang et al., 2015). Moreover, because SpCas9 can also target site with the alternative PAM sequence (NAG), off-target sites preceding NAG need to be considered. Indeed, SpCas9 was shown to be active but less efficient at off-target sites with this alternative PAM sequence (Hsu et al., 2013; Mali et al., 2013b). Whole-genome sequencing or any unbiased approach such as genome-wide labelling of DSBs may detect all or most off-target mutations. Whole exome sequencing was performed to identify off-target mutations induced by Cas9 in three human K562 cell line clones; no mutations were found in the three lines although there was a high false-positive rate associated with sequencing analysis (Cho et al., 2014). This indicates that it is possible to obtain no or low off-target mutations with careful selection of gRNAs. In addition to selecting a unique gRNA sequence, different methods and approaches have been identified to increase SpCas9 specificity, as described below.

1.5.4.4 Different approaches to improve the specificity of CRISPR-Cas9

Several studies have used different strategies to improve CRISPR-Cas9 specificity. Firstly, the concentration of Cas9 or sgRNA seems to be important for the catalytic activity of SpCas9 at either the on- or off-target sites. Mismatches seem to be better tolerated at high SpCas9 concentrations, because reducing its concentration led to a reduction in the off-target mutation rate (Hsu et al., 2013); however, the on-target mutation rate was also decreased (Fu et al., 2013; Wang et al., 2015). Secondly, the addition of two extra guanines at the 5` end of the sgRNA, or the use of truncated sgRNAs which are 2 to 3 nucleotides shorter at the 5` end, have been shown to improve SpCas9 specificity because they have better discrimination between on- and off-target sites (Cho et al., 2014; Fu et al., 2014). However, these strategies, especially the extra GG, also

reduced the efficiency of on-target genome editing. Another strategy that was also used to improve the specificity of SpCas9 is known as “paired nickases”, which means adjacent offset nicks are generated by using two gRNAs and mutant Cas9 at the target site.

1.5.4.5 Paired gRNA nickases

Specificity can be enhanced by using mutant Cas9 and a pair of offset gRNAs that are complementary to opposite strands of the target site, leading to the formation of double nicks at the target DNA (Figure 1.5b). Nicking both DNA strands can lead to site-specific DSBs, whereas individual nicks are mainly repaired by the high fidelity base excision repair pathway (BER); therefore, off-target mutations are minimised (Mali et al., 2013b; Dinov and Husbscher, 2013; Ran et al., 2013b). To achieve this with the Cas9 system, one of the catalytic domains of Cas9 (RuvC and HNH) that mediate DNA cleavage needs to be mutated. Two different single substitutions are known to deactivate the functionality of either HNH or RuvC. A substitution of aspartate to alanine (Cas9D10A) in the RuvC I domain, or histidine to alanine (Cas9H840A) in the HNH domain, converts the nuclease into a DNA nickase (Jinak et al., 2013; Cong et al., 2013). Cas9D10A nicks the complementary strand of the DNA, while Cas9H840 nicks the non-complementary strand. Mali et al (2013b) found that creating offset nicks to generate DSBs is highly efficient for inducing gene disruption mediated by either NHEJ or HR, especially with the 5` overhang generated by Cas9D10A. Another study compared the efficiency of wild type Cas9 and mutant Cas9D10A to introduce a mutation at targeted sites and found that the on-target efficiencies are comparable (Cho et al., 2014). In addition, Ran et al (2013b) demonstrated that using the Cas9D10A paired nickase can diminish off-target mutations by 50- to 1,500-fold in human cell lines without affecting the efficiency of DNA cleavage at on-target sites. Similarly, Wang et al (2015) compared the off-target activity of wild type Cas9 and Cas9D10A by using integrase-defective lentiviral vectors (IDLVs), that can detect such off-target cleavage with a frequency as low as 1%. They found no detectable off-target activity with Cas9D10A although the wild type CRISPR-Cas9 produced multiple off-target cleavages. Lastly, unlike the wild type Cas9 nuclease, the paired nickase was not shown to promote unwanted translocations associated with off-

target cleavage (Cho et al., 2014). Altogether, these studies clearly imply that the Cas9D10A nickase approach improves specificity over the wild type Cas9.

1.6 The main aims and objectives of this project

The main aim of this project is to study the role of HMGN2 in stem cells and during *in vitro* neuronal differentiation. A mouse embryonal carcinoma stem cell line (P19 cells) was selected for this purpose because it is a cost-efficient alternative to mESCs, is easy to grow and displays most of the stem cell properties of mESCs. P19 cells are also able to differentiate down the neural lineage with high efficiency.

Objectives:

1- Generate functional HMGN2 knock-out lines using the paired nickase CRISPR-Cas9D10A strategy.

2- Characterise the genotype and HMGN expression in these newly generated lines.

3- Characterise the expression of HMGN2 during neuronal differentiation of wild type P19 cells using the adherent culture method.

4- Investigate whether loss of HMGN2 affects the expression of genes involved in pluripotency and neuronal differentiation.

5- Investigate whether the loss of HMGN2 affects chromatin structure by studying key histone acetylation and methylation marks.

Chapter 2 Materials and Methods

2.1 Cell culture and monoclonal lines

2.1.1 Reagents and instruments

Reagents	Abbreviation	Manufacture or source	Catalogue number
Dulbecco's Modified Eagle Medium/Ham's F-12	Advanced DMEM/F12	Gibco	12634010
New born calf serum	NBCS	Gibco	16010159
GlutaMAX™ Supplement	GlutMAX	Gibco	35050061
N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid	HEPES	Gibco	15630056
Penicillin streptomycin	Pen/strep	sigma	P4333
Tryple express enzyme, no phenol red	Tryple express	Gibco	12604013
Phosphate-buffered saline	PBS	Gibco	10010056
Trypan blue	--	Gibco	15250061
Recovery cell culture Freezing medium	--	Gibco	12648010
Coolcell	--	Biocision	BCS-170
Nalgene Cryo 1°C freezing container	--	Thermo Scientific	5100-0001
Haemocytometer	--	Marienfeld-Superior	0680010
Centrifuge	--	BOECO	C-28A
BioMat^2 Class II Microbiological safety cabinet	--	Medical Air Technology	--
Vortex GenIE2	--	Scientific industries	SI-0236
Laboratory fridge	--	BioCold	BIO130FRSS
Mini spin plus	--	Eppendorf	F-45-12-11
Inverted microscope	--	Olympus	Ck2
Galaxy S CO2 incubator	--	New Brunswick	170-200

2.1.2 Cell culture

Mouse embryonic carcinoma cell, P19, was obtained from Dr. Andrew Hamilton (Institute of Cancer Sciences, University of Glasgow) at passage 14. The

undifferentiated P19 cells was maintained in advanced DMEM/F12 containing 10% NBCS and 1X GlutaMAX. P19 adherent cells were incubated in standard condition which is 37^{o}C and 5% CO_2 for 2 days. The cells were passaged every two days as following:

1- The media for culturing P19 cells after two days was discarded.

2- The adherent cells were washed by PBS and then enzymatically dissociated by Tryple Express for 5 minutes in 37^{o}C and 5% CO_2.

3- The reaction was deactivated by adding advanced DMEM/F12 containing 10% NCBS with other supplements and detached cells were transferred into falcon tube.

4- The cells were centrifuged for 5 minutes at 155g (1000 rpm) to discard the deactivated Tryple with media.

5- The cell pellet was resuspended in 10 ml fresh complete media and the cells were seeded into new untreated tissue culture plastic-ware at 1:10 ratio.

6- The cells were then incubated at standard condition for another two days' till reaching to 80-90% cell confluence.

2.1.3 Cell cryopreservation

For freezing the cells, the adherent cells were disassociated from culture vessels as usual. The cells were collected and counted using the haemocytometer and 1-2X10^{6} of cells were used for freezing 1 stock (cryogenic vial). The total amount of required cells was transferred into new falcon tube and pellet by centrifuging the mixture at 155g for 5 minutes. After discarding the supernatant, the cell pellet was resuspended (drop-wise) in the total volume of recovery freezing media. 1 ml aliquots were added into each cryogenic vial. The vials were kept inside the cryovial containers (CoolCell) and directly stored in -80^{o}C for 1-2 days. The cryogenic vials then moved to the liquid nitrogen for long term storage.

For thawing a frozen stock of cells, the complete media was warmed up at 37^{o}C and the cells in cryogenic vial were taken from liquid nitrogen on ice. The frozen stock was rapidly thawed in 37^{o}C (beads bath) and drop-wised into 8-10 ml of complete media. The diluted recovery freezing medium in complete media was discarded after centrifuging the mixture at 155g for 5 minutes. The cell pellet was resuspended in fresh complete media and plated into culture vessels (Nunc) for two days.

2.1.4 FACS sorting and monoclonal lines

The transfected P19 cells with different vectors including the multiguide or a combination of pGuide and pMultiCas9D10A were plated for two days in same wild-type P19 culture's condition (section 2.1.2). After two days, the transfection efficiencies were measured by fluorescence-activated cell sorting (FACS) using Attune Acoustic Focusing Cytometer (Applied Bio system) or BD FACSAria Fusion instrument (BD Biosciences) if the cell sorting is required. At this stage, the GFP-positive cells in some of the transfections were sorted with a BD FACSAria Fusion instrument. Sorting of transfected cells was carried out in collaboration with Tim Harvey (FACS Facility, Institute of Cancer Sciences, University of Glasgow). The sorted populations were either automatically plated into 96-well plates at ratio 1 cell per well or sorted in tube and incubated for another 2 days before performing the monoclonal lines. The monoclonal lines for the GFP sorting of the transfection of HMGN2-ExI-CRISPR was automatically performed 2 days' post-transfection. The sorted and unsorted populations transfected by CRISPR-Cas9D10A-ExIV were cloned using the limiting dilution to establish monoclonal lines 4 days' post-transfection. The limiting dilution was carried out as following:

1- After detaching the cells by Tryple as mentioned in 2.1.2, the cells in suspension were counted using Trypan blue.

2- Calculate the volume of the media and number of cells needed based on the number of the wells and seeding ratio 1 or 2 cell/s per well. Seeding densities 1 or 2 cell per well were previously tested to determine the efficiency of generating single colonies. The 2 cells per well yielded in around the double the number of single colonies obtained from plating 1

cell per well at day 6, 8 and 12 after plating. For example, if 1 cell per well is required in 96 wells, the calculation will be as following:

Volume of media is: 1 cell per well X 96 wells X 200 μl per well= 19.2 ml (for pipetting issue, 21 ml instead)

Number of cells required in 21 ml for 1 cell per well is: 21 ml / 200 μl per well= 105 cells in 21 ml.

3- To make 21 ml containing 105 cells, based on the cell count 10^5 cells was added and mixed into final volume of 1 ml fresh media to form 10^5 cells in 1 ml suspension.

4- From the 10^5 cells suspension 100 μl was transferred by new tips into 900 μl fresh media to make 10^4.

5- In same way, the 10^3 cells in 1000 μl was made up and this mixture is assumed to have 1 cell per 1 μl.

6- Therefore, because in previous example 105 cells is desired in 21 ml, 105 μl of 10^3 mixture was added to 20.895 ml media.

7- Finally, 200 μl of the media containing cells were added into each well of the 96-well plate after mixing consistently.

The monoclonal lines from either sorted or unsorted populations were cultured in advanced DMEM/F12 media containing 20% NBCS, 1X GlutaMAX, 2% HEPES and 1X pen/strep to maintain the PH at the normal level and prevent the contamination. The cultures were kept till observing single colonies. The monoclonal cells were dissociated from culture vessels by Tryple Express as normal and the monoclonal lines were expanded and sub-cultured as wild-type P19 cells.

2.2 Neural induction

2.2.1 Reagents

Reagents	Abbreviation	Manufacture	Catalogue number
Dulbecco's Modified Eagle Medium/Ham's F-12	Advanced DMEM/F12	Gibco	12634010
Laminin from mouse engelbreth-holm-swarm tumour	Laminin	Sigma	L2020
N2-supplemented	N2	Gibco	17502048
B27- supplement	B27	Gibco	17504044
Retinoic acid	RA	Sigma	R2625-100
N-[N-(3,5-Difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester	DAPT	Sigma	D5942
Neurobasal media	NB	Gibco	21103049
Cytosine ß-D-arabinofuranoside hydrochloride	AraC	Sigma	C6645-25
Fibroblast Growth Factor 8 Recombinant Human Protein	FGF8	Gibco	PHG0184
GlutaMAX™ Supplement	GlutMAX	Gibco	35050061

2.2.2 Method

The neural induction using the monolayer adherent culture were carried out as mentioned previously by Nakayama et al (2014) with some modifications.

1- Briefly, cultured dishes were coated with 5 µg/ml of laminin diluted in PBS for at least 4 hours. Although 2.5 µg/ml of laminin was used by Nakayama, different concentrations up to 10 µg/ml were tested but 5 µg/ml was the optimum concentration and selected for all subsequent experiments. For immunofluorescence, the laminin was added into 24-well plate containing sterilised cover slips (cover slips were washed with dH_2O-ethanol-2XdH_2O and then autoclaved and dried before use).

2- Before plating, the laminin was discarded and the dishes were washed with PBS three times.

3- Prepare the neural induction media [N2-supplemented advanced DMEM/F12 (1ml 100X N2:100 ml media) containing 500 nM retinoic acid (RA), 10 ng/ml FGF8, 2 mM GlutaMAX and 10 μM DAPT (γ-secretase inhibitor)]. RA and DAPT are dissolved in DMSO while the FGF8 in distilled water.

4- The adherent wild-type P19 cells, that were cultured as normal, were collected by dissociating cells by Tryple and the healthy cells were counted.

5- After counting the healthy cells and preparation of the induction media $1X10^5$ cells per ml were prepared and 500 μl, 2 and 10 ml were added into 24 well, 6 well plates and 100mm dishes, respectively. In separate experiment, different seeding densities were tested but no differences have not been noticed.

6- The cells were adherently cultured for 4 days in neural induction media.

7- After 4 days, the neural induction media was replaced by B27-supplemented neurobasal medium (2 ml of 50X B27:100 ml media) containing 2 mM GlutaMAX and 8 μM cytosine arabinoside hydrochloride (AraC) and the cells cultured for two days.

8- After 2 days of incubating cells in media containing AraC, the AraC was withdrawal and the only B27-supplemented neurobasal medium with 2 mM GlutaMAX was gently changed every 2 days till the day of harvesting.

9- The cells incubated at 37^oC and 5% CO_2 during the whole process. The N2-DMEM/F12 and B27-NB media were filter sterilise before adding other reagents and can be stored at 4^oC for up to 4 weeks.

2.3 Plasmid constructs and creation of Golden Gate for 2 or 4 multiplex CRISPR guide expression

2.3.1 Plasmid's constructs

All plasmids backbones were obtained from Dr. Adam west lab. The Cas9 nuclease that is from *Streptococcus pyogenes* (SpCas9) require a guide RNA that directs SpCas9 to a genomic target locus. The expression of wild-type SpCas9-T2A-GFP, mutant pMultiCas9D10A-T2A-GFP and pCas9D10-T2A-GFP are driven by CAG promoter (Figure 2.1). The difference between the last two is pMultiCas9D10A-T2A-GFP has multiguide cassette. The wild-type SpCas9-T2A-GFP was used alone or in combination with individual pGuide vector to check the efficiency of specific gRNA to direct Cas9 for introducing mutation at that specific locus. The pMultiCas9D10A-T2A-GFP was used for transfection the CRISPR against exon IV of *Hmgn2* whereas pCas9D10-T2A-GF with two combination pGuide vectors for CRISPR against exon I.

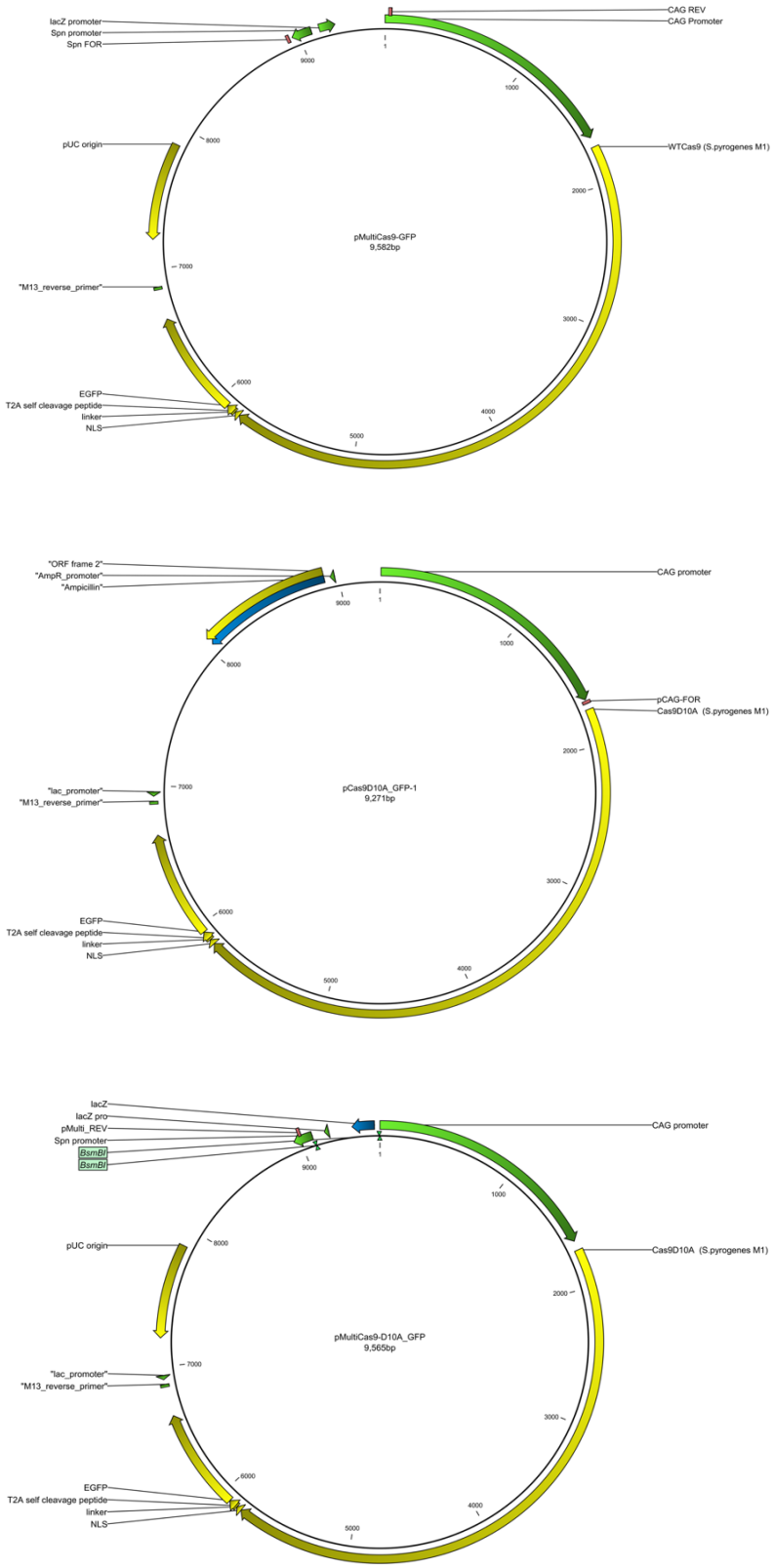
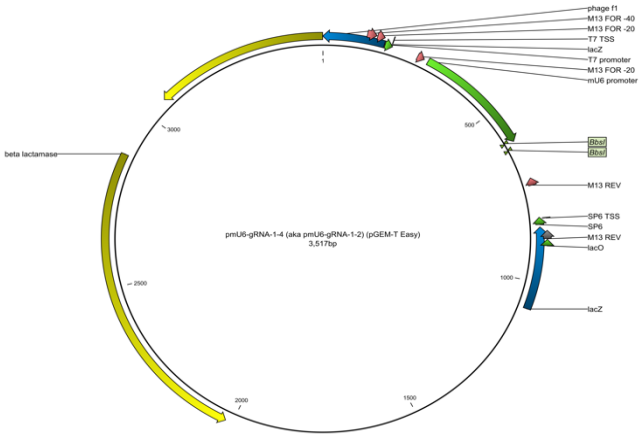
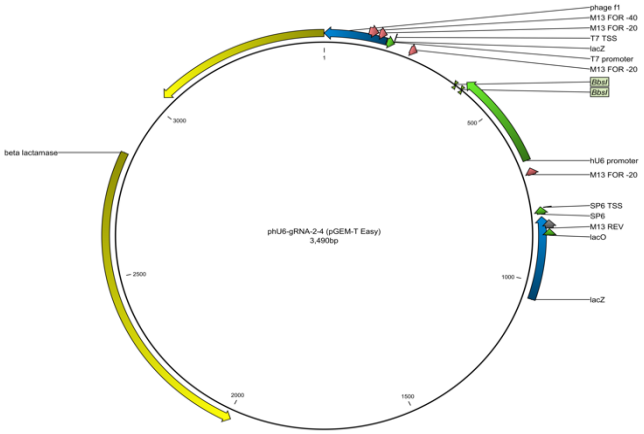
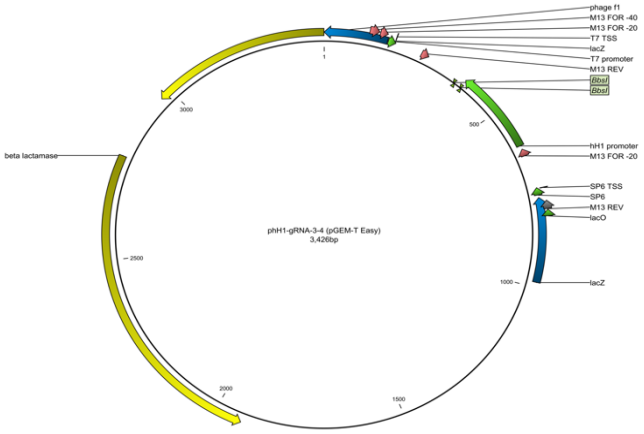
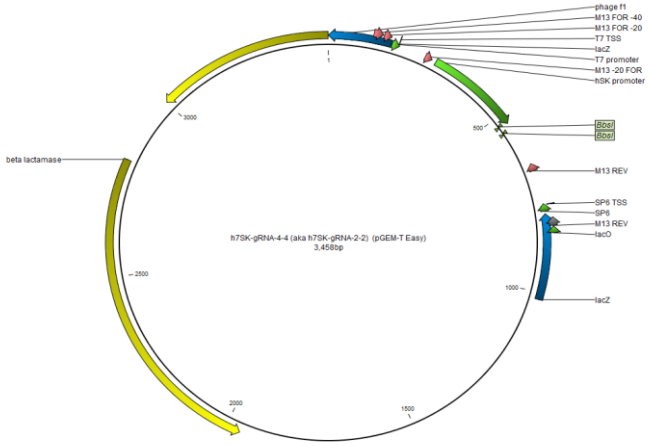
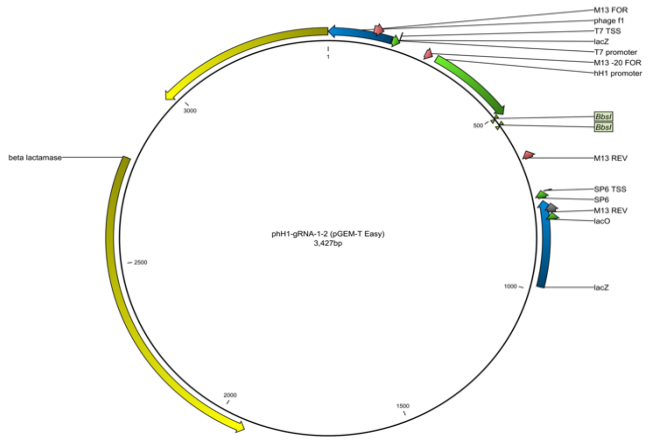
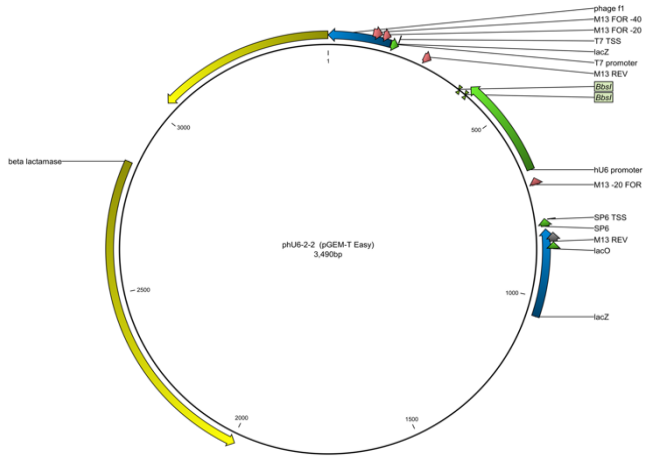


Figure 2.1: wild-type and mutant SpCas9 plasmids.
The name and size of each vector are given inside each plasmid. CAG promoter drives the expression of wild-type or mutant Cas9 and GFP that is linked by T2A linker and self cleavage peptide. The sequence of primer on backbone and used for colony PCR or sequencing after assembling the gRNAs cassettes in pMutliCas9D10A_GFP are annotated by red bracket and named as (p.Multi_REV). NLS is nuclear localisation signal, pUC origin is the origin of replication initiation.

Four independent promoters including mU6, hU6, hH1 and h7SK to minimise sequence repetition drive the expression of sgRNAs. These promoters are in 6 different vectors as following:

Parental expression vector name	Vector backbone
mU6_gRNA_1/4 (aka mU6_gRNA_1/2)	
hU6_gRNA_2/4	
hH1_gRNA_3/4	

h7SK_gRNA_4/4 (aka h7SK_gRNA_2/2)	
hH1_gRNA_1/2	
hU6_gRNA_2/2	

The first four are for the 4 multiplex gRNA assemblies while the last two and aka mU6_gRNA_1/2 and aka h7SK_gRNA_2/2 are essential for 2 multiplex gRNA assemblies.

2.3.2 Design CRISPR-Cas9 nickase for targeting the exon IV or I of mouse *Hmgn2*

Two nickases were designed for exon IV while one nickase for exon I (near ATG) of *Hmgn2*. The suitable gRNAs for this purpose was selected using the website (http://crispr.mit.edu/). The first two nickases are called as MmHmgn2ExIV_A68/B88 and MmHmgn2ExIV_A87/B91 while the third for exon I is named MmN2ATG_A69/B77. The numbers are according to the score of each guide in the nickase (Figure 3.2 and 4.1). Each nickase was selected based on different criteria such as the overlap with exon and whether there is a restriction enzyme site in a useful place. The protospacer oligonucleotides (gRNA) are typically 20 bases long and should start with G or A to avoid adding an extra non-matching base at the 5' end. The protospacer that will be cloned next hH1 promoter must have an A at 5’ end while the other promoters require G. However, the protospacers can be as short as 17 bases if there is a G or A base at the 5' end of the genomic target sequence. All the sequences of oligonucleotides for protospacers are given in Table 2.1. Once the paired guides chosen, primers for a genomic PCR for RFLP analysis of Cas9-mediated mutations are designed at http://primer3.ut.ee/. The primers were ordered from MWG Eurofins. Different primer sets were designed and selected for CRISPR-Cas9D10A exon I or IV. All the primer sets and their sequences are given in the Table 2.2.

Oligo name	Sequence	Protospacer (gRNA)
MmHmgn2ExIV_ A87_mU6T	TTGTTTGGCTGGCTTACCTTCGC	TTGTTTGGCTGGCTTACCTTCGC AACCGACCGAATGGAAGCGCAAA
MmHmgn2ExIV_ A87_mU6B	AAACGCGAAGGTAAGCCAGCCAA	
MmHmgn2ExIV_ B91_hU6T	CACCGGTGCTAATTCTAGACT	CACCGGTGCTAATTCTAGACT CCACGATTAAGATCTGACAAA
MmHmgn2ExIV_ B91_hU6B	AAACAGTCTAGAATTAGCACC	
MmHmgn2ExIV_ B88_hH1T	TCCCACCTAAAAAGGCCCCTGCGA	TCCCACCTAAAAAGGCCCCTGCGA TGGATTTTTCCGGGGACGCTCAAA
MmHmgn2ExIV_ B88_hH1B	AAACTCGCAGGGGCCTTTTTAGGT	
MmHmgn2ExIV_ A68_h7SKT	CCTCGGAGCAGGTTTCTTATTT	CCTCGGAGCAGGTTTCTTATTT CCTCGTCCAAAGAATAAACAAA
MmHmgn2ExIV_ A68_h7SKB	AAACAAATAAGAAACCTGCTCC	
MmHN2ATG_A6 9_ h7SKT	CCTCGTACCTTTCTTTTGGGCA	CCTCGTACCTTTCTTTTGGGCA CATGGAAAGAAAACCCGTCAAA
MmN2ATG_A69_ h7SKB	AAACTGCCCAAAAGAAAGGTAC	
MmN2ATG_B77_ hH1T	TCCCACCGCCAGCCTCGGGCT	TCCCACCGCCAGCCTCGGGCT TGGCGGTCGGAGCCCGACAAA
MmN2ATG_B77_ hH1B	AAACAGCCCGAGGCTGGCGGT	

Table 2.1: The oligonucleotide sequences for cloning new protospacers into sgRNA expression vectors.
The blue bases are referred to the overhang that is complementary to the specific gRNA vector. The red sequences are the complementary to genomic target sequence. The first base of the genomic target sequence (highlight letter) is critical and was selected based on the first base in each promoter. Notably, only human H1 promoter starts with an A while the other promoters start with G. The oligonucleotides start with MmHmgn2ExIV for CRISPR against exon IV of *Hmgn2* while the ones starting with MmN2ATG are oligonucleotides for CRISPR against exon I (or ATG). The letter with number coming after the MmHmgn2ExIV_ or MmN2ATG_ such as (A87 in first oligo) refers to the name and score specific for each oligo. Each oligo creates different overhangs that are complementary to the overhangs generated on the digested parental expression vectors.

Primer name or set	Amplification locus	Sequence	Note
Primer set 1 forward or mN2-nickase-PCR1-F	Exon IV of *Hmgn2*	AGCTGACTAGTGAGGAGACC	The reverse primer is common
Primer set 2 forward or mN2-nickase-PCR2-F		TCTCTGTGATCCCGTTGCAT	
Primer set 1 and 2 reverse or mN2-nickase-PCR1/2-R		TGCAGTCCTCCAGAGTTGTT	
mN2ATG_PCR1F	Exon I of *Hmgn2*	CGCGCCCCTCATTTATTACC	The combination of different forward and reverse primers were tested but the only pair worked is mN2ATG_PCR4F and mN2ATG_PCR1R
mN2ATG_PCR2F		GTTTGCAAAAGCCAGGTCTG	
mN2ATG_PCR3F		GAGTTTAAACCCCGCCCAC	
mN2ATG_PCR4F		GGCTTGGAGAAGTGGGTAGA	
mN2ATG_PCR1R		TGTGCAAACAAACCTCTCCG	
mN2ATG_PCR2R		ACAGGGAGCGGGATAAACAG	
mN2ATG_PCR3R		CTTTCGGGTTTGAATCGCCC	
mN2ATG_PCR4R		GCCCAACTAAAGTCCGAGAC	

Table 2.2: Forward and reverse primers used in amplification of exon I or IV.

2.3.3 Protocol for multiplex CRISPR guide expression vectors assembly

The detailed protocol for assembly of custom lentiviral vectors expressing up to four sgRNAs of choice and active Cas9, dCas9 or dCas9^{VP64} was previously described by Kabadi et al (2014). Here, we did not use lentiviral vectors and some modifications were applied. For creation multiplex CRISPR guide expression vector assembly, two main steps are involved. The process is initiated with cloning of short annealed oligos (guide protospacer matching genomic target) into a compatible RNA polymerase III promoter cassette and then Golden Gate assembly cloning such that promoter-gRNA cassettes into tandem inverted repeat units. Both steps involve the use of Type II(S) restriction enzymes which cleave outside their recognition site to create unique overhang. One of these restriction enzyme is *BsmB*I that is active in the same buffer as T4 DNA ligase. The creation of different overhangs and the active restriction enzyme in the same ligation buffer facilitate the assembly of two or four expression cassettes into final pMultiguideCas9D10A-T2A-GFP vector in one step.

Step 1: cloning of short annealed oligos into a compatible RNA polymerase III promoter cassette.

This step was performed for creating individual pGuide vector of either CRISPR-Cas9D10A targeting exon I and IV.

1- Two complementary single stranded DNA oligos for each desired genomic target, that were mentioned previously (Table 2.1), were annealed as following: 8 µl of each sense and antisense oligos (both 10 µM) was mixed with 2 µl of 10X T4 DNA ligase buffer (NEB), followed by melting and reannealing in a thermal cycler with the program:

Temperature	Time	Note
96°C	300 sec	with a −0.3◦C/s rate
85°C	20 sec	
75°C	20 sec	
65°C	20 sec	
55°C	20 sec	
45°C	20 sec	
35°C	20 sec	
25°C	20 sec	

2- After annealing of complementary oligos, the overhang of each protospacer was phosphorylated because it requires for ligation later on. The phosphorylation was accomplished by adding 1 µl of 25 mM ATP and 1 µl T4 Polynucleotide Kinase (NEB), followed by incubation at 37°C for 60 minutes. The enzyme then was inactivated by heat at 65°C for 20 minutes.

3- 5 µg of each gRNA backbone plasmid was digested by 1 µl of *Bbs*I (NEB, R0539S) in 3 µl 10X NEB buffer 2.1. The digestion occurs at 37°C for 2 hours.

4- The overhangs of gRNA backbone plasmids were dephosphorylated by adding 0.5 µl of alkaline phosphatase calf intestine (NEB, M0290S) for further 1 hour at 37°C.

5- The linearization of plasmids was checked by running 2 µl of each reaction, plus undigested plasmids as control, in 1% agarose gel electrophoresis for approximately 1 hour.

6- The linearized plasmids were purified using Qiagen PCR clean up kit (QIAquick PCR purification kit, 28104).

7- Subsequently, each protospacer was ligated into the purified desired vector (Table 2.3) by Qiagen PCR clean up kit as the below reaction. The reaction was incubated at 16°C for 60 minutes on a thermal cycler:

Reagent	Volume
Annealed oligonucleotides	1
T4 DNA ligase (NEB, M0202S)	0.5 µl
10X NEB DNA ligase buffer (fresh aliquots)	1 µl
50 ng of purified vector	X µl
dh_2O	Up to 10 µl

Cloning of guide vectors			
Insert oligo	**Parental expression vector name**	**Individual pGuide vector**	**GG assembly**
MmHmgn2Ex4_A87_mU6	mU6_gRNA_1/4 (aka mU6_gRNA_1/2)	pGuide001-4_mU6 or pGuide001-2_ mU6	Vector contains 4 multiplex gRNA is combined of green and called pMultiNick_003-4, while 2 multiplexes is either in blue or dark red known as pMultiNick_001-4 or pMultiNick_002-4, respectively
MmHmgn2Ex4_B91_hU6	hU6_gRNA_2/4	pGuide002-4_hU6	
MmHmgn2Ex4_B91_hU6	hU6_gRNA_2/2	pGuide002-2_hU6	
MmHmgn2Ex4_B88_hH1	hH1_gRNA_3/4	pGuide003-4_hH1	
MmHmgn2Ex4_B88_hH1	hH1_gRNA_1/2	pGuide001-2_hH1	
MmHmgn2Ex4_A68_h7SK	h7SK_gRNA_4/4 (aka h7SK_gRNA_2/2)	pGuide004-4_h7SK or pGuide002-2_h7SK	
MmN2ATG_A69_ h7SK	h7SK_gRNA_4/4 (aka h7SK_gRNA_2/2)	pGuide004-4_h7SK\ExI	Not assembled in multinickase vector
MmN2ATG_B77_ hH1	hH1_gRNA_1/2	pGuide001-2_hH1\ExI	

Table 2.3: Inserted oligo into desired vector and the assembly of different pGuide vector into final vector for Golden Gate assembly.

8- 2.5 µl of each ligation was transformed into fresh subcloning efficiency DH5α competent cells (Life Technology) as following:

a. 25 μl aliquots of DH5α competent cells in 1.5 ml microcentrifuge tube were thawed on ice.

b. Add 2.5 μl of each DNA from each reaction gained after step 5 and 1 μl of pUC19 control to the cells and mix gently.

c. Incubate the cells with DNA on ice for 30 minutes and then heat shock at 42°C water bath for 20 seconds without shaking.

d. Put the tubes again on ice for 2 minutes before adding 470 μl of pre-warmed SOC media [10 ml of LB broth (20 g/l LB broth and 5 g/l NaCl), 0.2 ml of 2 M glucose and 500 μl of 2 M $Mgcl_2$].

e. The tubes were incubated at 37°C for 1 hour at 225 rpm.

f. After the incubation, 10-100 μl of each transformation was plated onto LB agar (20 g/l of LB broth, 5 g/l of NaCl and 10 g/l of the agar) plates containing 100 μg/ml ampicillin and incubated overnight at 37°C. Typically, >90% of the colonies contain the desired ligation product.

9- Different colonies were selected and the DNA plasmid was expanded and purified using either Midi- or Mini-preparation (Qiagen, 27104 and 12943, respectively). In the same time, colony PCR using a vector primer (T7 or SP6) and the bottom strand oligo for the guide protospacer, which has been cloned into the plasmid, was performed to screen transformations. Colony PCR was performed using *Taq* DNA polymerase (BioLabs, M0257S) in 25 μl reaction.

10- In addition, the insertion of protospacer into its expression promoter cassette was validated by sequencing prior to proceeding to step 2. The sequencing was commercially performed using the T7 primer (TAATACGACTCACTATAGGG) for plasmids 1-4, 4-4 and 1-2 but SP6 (ATTTAGGTGACACTATAGAA) for the rest plasmids.

Step 2: Assembly of either the two or four promoter-gRNA cassettes into a destination vector (pMultiCas9D10A-T2A-GFP) using Golden Gate cloning.

This stage only done for CRISPR against exon IV as multiguide plasmids (all-in-one-vector) containing 2 or 4 gRNAs are required. However, because of the issue of multiguide plasmid seen in CRISPR targeting exon IV, it was decided to use the individual gRNA vectors (pGuide) along with pCas9D10A-T2A-GFP for targeting exon I. Therefore, this step was only done for assembly gRNAs targeting exon IV but not exon I.

1- To assemble the two or four different promoter-gRNA constructs into the desired destination vector, 130 ng of each sgRNA expression plasmid (pGuide vectors) and 125 ng of pMultiCas9-T2A-GFP destination vector are combined with 1µl of T4 DNA ligase (NEB), 1 µl of *BsmB*I (NEB) and 2 µl of 10× T4 ligase buffer (NEB) in a 20 µl reaction volume.

2- The reaction was then incubated for the Golden Gate assembly as mentioned bellow:

Temperature	Time	Cycles
37°C	10 minutes	5
16°C	15 minutes	
37°C	30 minutes	-
80°C	5 minutes	-

3- 2 µl of the reaction was transformed into SURE2 competent cells (Agilent, 200152), which have been engineered to allow the cloning of certain DNA segments that are "unclonable" in conventional E. coli strains, following the manufacturer's instructions.

a) 14-ml BD Falcon polypropylene round-bottom tubes for each reaction in step 1 and 2 were pre-chilled on ice. In addition, one tube is for the pUC18 control.

b) The cells were thawed on ice. When thawed, gently mix and aliquot 100 µl of cells into each of the two pre-chilled tubes.

c) 2 µl of the ß-ME provided with this kit was added to each aliquot of cells.

d) Swirl the tubes gently. Incubate the cells on ice for 10 minutes and swirl gently every 2 minutes.

e) 2 µl of the experimental DNA was added to one aliquot of cells. Dilute the pUC18 control DNA 1:10 with sterile dH2O, then add 1 µl of the diluted pUC18 DNA to the other aliquot of cells.

f) Swirl the tubes gently, then incubate the tubes on ice for 30 minutes.

g) Heat-pulse the tubes in a 42°C water bath for 30 seconds. The duration of the heat pulse is critical.

h) The tubes were incubated on ice for 2 minutes.

i) 0.9 ml of preheated (42°C) SOC media was added into each tube and the tubes were incubated at 37°C for 1 hour with shaking at 225 rpm.

4- The transformations are plated (~ 150 µl) onto LB agar plates containing 50 µg/ml spectinomycin (Sigma, S4014) and incubated overnight at 37°C. Colonies were screened by lacZ-based blue-white screening using IPTG and X-gal. Typically, >90% of the colonies contain the desired ligation product.

5- Golden Gate assembled pMultiguide plasmids were digested either with *Xho*I or *Xba*I-*Spe*I (CutSmart) to ensure the ligation was achieved.

6- Colony PCR and sequencing were carried out to validate the ligation and the sequences using common primer on the vector backbone, called as pMulti-REV (ACTGGGTTCGTGCCTTCATC). The cassette expressions of pMultiNick_001-4, pMultiNick_002-4 and pMultiNick_003-4 are given in Figure 2.2

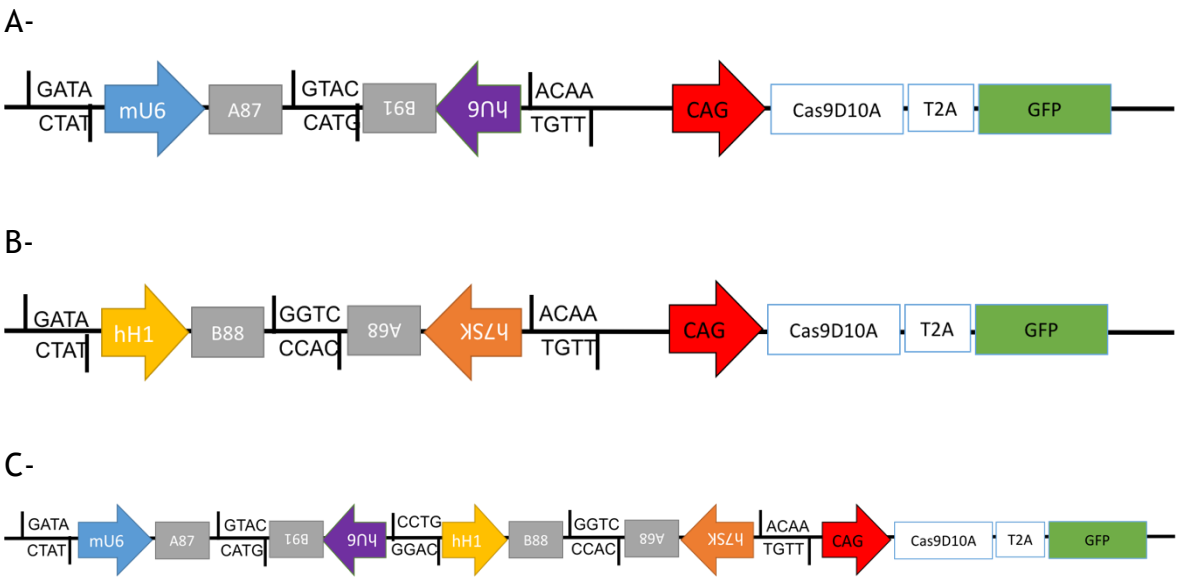


Figure 2.2: The expression cassettes encoded by different pMultiguide vectors.
The expression cassettes for all pMultiguide plasmids (all-in-one-vectors) obtained after Golden Gate cloning. pMultiNick_001-4 (**A**) and pMultiNick_002-4 (**B**) have 2 multiguide RNAs while pMultiNick_003-4 (**C**) contains 4 gRNAs. *BsmB*I digestion for each individual gRNA vector creates different 4 bp overhangs which specifies their position and orientation when ligated together.

2.4 Neon transfection

The Neon electroporation system (Life Technologies) is based on using the pipette tip as an electroporation chamber to efficiently transfect a wide variety of mammalian cells. The principle of using Neon tip as chamber is that the Neon tip maximises the gap size between the two electrodes while minimising the surface area of each electrode. This will result a more uniform electric field, minimal pH change, less ion formation and negligible heat generation.

P19 cell is slightly hard to transfect. Different transfection methods including Amaxa nucleofection and lipofectamine was previously tested but no more than 15-50% transfection efficiency was obtained for pmaxGFP plasmid. Firstly, the Neon system was used to determine the transfection efficiency for plasmid DNA (pmaxGFP) in ECCs (P19) using two commercial programs. The programs include one compatible with P19 cell line while the other for mouse ESCs (Table 2.4). 89 and 97% transfection efficiency were gained with P19 and mESCs programs, respectively. Therefore, the mESCs program was used in all subsequent transfections. In addition, different concentrations of plasmid were optimised.

Program	Pulse voltage (V)	Pulse width (ms)	Pulse number	Cell density (cells/ml)	Transfection efficiency	Viability	Tip type
mESC	1,400	10	3	$1X10^7$	88%	96%	10 µl
P19	1,275	30	1	$5X17^6$	90%	80%	10 µl

Table 2.4: : The provided Neon optimised program for Neon transfection.
The mESC and P19 programs that have been optimised by Life Technologies and tested for the transfections. The same conditions were implanted but 100 µl tip was used instead of 10 µl.

The transfection efficiency was also assessed for SpCas9-T2A-GFP and pMultiCas9D10A-T2A-GFP plasmids. In addition, each individual gRNA was tested by co-transfection of each gRNA with wild-type SpCas9-T2A-GFP. However, the transfection for knocking out HMGN2 protein in P19 cells by Neon was performed using different pMultiguides (all-in-one) that have mutant Cas9D10A on the backbone vectors and either two gRNAs (pMultiNick001-4 and pMultiNick002-4) and 4 multiplex gRNAs (pMultiNick003-4) for targeting exon IV of *Hmgn2* genes at different regions. On the other hand, the strategy has been modified for targeting exon I in which transfecting two individual pGuide vectors with SpCas9D10A-T2A-GFP to avoid the viability issues that result from transfecting one multiguide vector.

The transfection was carried out according to the manufacturer's protocol using 100 µl Neon tips (Life Technologies). Briefly, on the days of the electroporation, 1-1.5$X10^6$ of trypsinised and neutralised cells were centrifuged at 155g for 5 minutes at room temperature. Then, they were washed by PBS and resuspended in R buffer, that supplied by kit, with variable concentration of plasmids (10, 15 or 25 µg, see Table 2.5). The mixture of plasmid DNA, cells and electroporation buffer was mixed and then aspirated using a 100 µl Neon tip. The already optimised program for mouse embryonic stem cells was applied (see Table 2.4). After the electroporation, the cells were then immediately plated onto 6 well plate containing 2 ml of pre-warmed media and then incubated under standard growth conditions before assessing the transfection efficiency by a fluorescent microscope (24 hours) and by FACS 48 hours' post-transfection. Other different ways, including leaving the mixture (cells, buffer and DNA) on ice for 5 minutes' prior the electroporation and/or 10 minutes at room temperature post the electroporation that may affect the viability of cells were examined.

Transfection purpose	Plasmid	Concentration
Control	SpCas9-T2A-GFP	15 μg
Checking efficiency for gRNA	pGuide001-4_mU6/A87 + SpCas9-T2A-GFP	10 μg 15 μg
Checking efficiency for gRNA	pGuide002-4_hU6/B91 + SpCas9-T2A-GFP	10 μg 15 μg
Checking efficiency for gRNA	pGuide003-4_hH1/B88 + SpCas9-T2A-GFP	10 μg 15 μg
Checking efficiency for gRNA	pGuide004-4_h7SK/A68+ SpCas9-T2A-GFP	10 μg 15 μg
Control, empty vector	pMultiCas9D10A-T2A-GFP	15 μg
CRISPR targeting exon IV of *Hmgn2*	pMultinick001_4	15 μg
CRISPR targeting exon IV of *Hmgn2*	pMultinick002_4	15 μg
CRISPR targeting exon IV of *Hmgn2*	pMultinick003_4	15 μg
Control for CRISPR targeting exon I	SpCas9D10A-T2A-GFP	10 μg
Checking efficiency for gRNA	pGuide001-2hH1\ExI+ SpCas9D10A-T2A-GFP	5 μg 10 μg
Checking efficiency for gRNA	pGuide004-4_h7SK\ExI+ SpCas9D10A-T2A-GFP	5 μg 10 μg
CRISPR targeting exon I of *Hmgn2*	SpCas9D10A-T2A-GFP	10 μg
	p.Guide001-2_hH1\ExI	5 μg
	p.Guide004-4_h7SK\ExI	5 μg

Table 2.5: Different Neon transfections and the plasmid concentrations.

2.5 PCR-RFLP assay

The DNA was extracted from different samples using DNeasy Blood and Tissue Kit (Qiagen, 69504) as manufacture protocol. 200-300 ng DNA was used per each reaction. Different DNA enzymes were used to amplify the CRISPR-targeted region. This includes *Taq* DNA polymerase (New England BioLabs, M0257S), Herculase II Fusion DNA polymerase (Agilent, 600677), *Phusion* High-Fidelity DNA polymerase (New England BioLabs, M0530) and *KAPA* HiFi PCR Kit (KAPABiosystems, KR0368). Most the procedures were followed as manufacture and the final reactions and conditions for amplifying exon I and IV of *Hmgn2* are listed in Table 2.6. Different optimisations including several annealing temperatures and different concentrations of PCR enhancing agents (DMSO and betaine) were examined. The primer sets used for amplifying the exons IV and I are listed in Table 2.2.

CRISPR	Reaction components	Per 50 µl reaction	Final concentration	Condition
Exon I	5X KAPA GC buffer	10 µl	1X	Initial denaturation- 5 mins- 95°C (1 cycle) Denaturation- 20 seconds- 98°C Annealing – 15 seconds- 59°C Extension- 15 seconds- 72°C (30 cycles) Final extension- 5 mnis- 72°C (1cycle)
	10 mM dNTP mix	1.5 µl	0.3 mM each	
	10 µM mN2ATG_PCR4F	1.5 µl	0.3 µM	
	10 µM mN2ATG_PCR1R	1.5 µl	0.3 µM	
	1 U/µl KAPA HiFi DNA pol.	1 µl	0.5 U	
	Templet DNA	Variable	200 ng	
	Betaine	-	1 M	
	PCR-grade water	Up to 50 µl	-	
Exon IV	10X ThermoPol reaction buffer	5 µl	1X	Initial denaturation- 2 mins- 95°C (1 cycle) Denaturation- 30 seconds- 95°C Annealing – 30 seconds- 58.1°C Extension- 30 seconds- 72°C (30 cycles) Final extension- 5 mnis- 72°C (1cycle)
	10 mM dNTPs	1 µl	200 µM	
	10 µM forward primer	1 µl	0.2 µM	
	10 µM reverse primer	1 µl	0.2 µM	
	Taq DNA Polymerase	0.25 µl	1.25 U	
	Templet	Variable	300 ng	
	Nuclease-free water	To 50 µl	-	

Table 2.6: PCR reaction components and conditions for amplifing exon I and IV of *Hmgn2* where CRISPR targets.

The desirable PCR products (10 µl) were checked in 1-2.2% agarose gel electrophoresis and if desired band was obtained the rest of PCR products were purified by QIAquick PCR purification kit (28104) before subjecting them to digestion. For the RFLP assay, 10 µl of cleaned up PCR products were incubated with 1 µl of restriction enzymes and 1.5 µl of 1X NEB buffer that is compatible with the enzymes (Table 2.7) in 15 µl total volume per reaction. The incubation degree and period vary based on the restriction enzymes used (Table 2.7). 10 and 15 µl of undigested and digested PCR products were respectively loaded in 1-2.2% agarose gel that was run for 45-90 minutes at 100v. DNA ladders were also loaded in each gel. The bands were visualised by staining the gel with SYBER safe DNA gel stain (Life Technology, S33102) and the images were taken by Flouro Image Analyser machine (Fujifilm, FLA-5000).

Restriction enzyme	Suitable buffer	Temp (°C)	Time (hour)
*Xba*I (R0145S)	CutSmart	37	1
*Bsal*I (R0555S)	CutSmart	55	1
*BfuA*I (R0701S)	3.1	50	1
*Ban*II (R0119S)	CutSmart	37	2
*Nla*III (R0125S)	CutSmart	37	2

Table 2.7: The restriction enzymes used in RFLP assay for CRISPR targeting either exon I or IV.

2.6 Cloning and sequencing

All the amplicons (PCR products) amplified from either exon I or IV from different lines and wild-type P19 cells were send to GATC Biotech for sequencing. Initially, the PCR products were cleaned up by QIAquick PCR purification kit exactly as manufacturer's instructions. The amplified and purified PCR products were subjected to TA cloning before sending each pure clone to sequencing. Because some of the PCR products, especially the PCR products from CRISPR targeting exon I of *Hmgn2*, were amplified using high fidelity DNA polymerase, the addition of 3' A overhangs to PCR products was performed as following:

	Final Concentration	Volume(µl)
Purified PCR product	0.15 to 1.5 pmol	3
dATP (1 mM)	0.2 mM	2
PCR ThermoPol Buffer with Mg (10x)	1x (1.5 mM $MgCl_2$)	1
Taq DNA Polymerase (5 U/µl)	1U	0.2
dH_2O	-	up to 10 µl

The reaction was incubated at 70°C for 30 minutes and 2 µl of each reaction was subjected to TA cloning. The PCR products with additional 3'A was ligated by T4 DNA ligase into pGEM-T Easy Vector (Promega, A1360) as exactly described by manufacture. The ligated vectors were transformed into DH5α competent cells (Life Technology) as mentioned previously. Several colonies were selected and mini-preped using Qiagen kit. The validation of ligation was confirmed using *EcoR*I restriction enzyme before sending samples to sequencing. T7 primer was used for sequencing.

2.7 Western blotting

2.7.1 Protein extraction and quantification

The undifferentiated, differentiated and transfected P19 cells/lines were harvested as normal and then washed twice with PBS. The cells were either resuspended in 300-500 µl whole cell lysis buffer (CLB, 45 mM of Tris-HCl pH 7.5, 1 mM EDTA, 1% SDS, 10% glycerol, 2 mM DTT, 100 µM sodium orthovanadate, 10 mM sodium butyrate and 400 µl of mini- tablet protease inhibitor cocktail dissolved in 1 ml dH_2O) or RIPA buffer (Life Technology, 89900). Alternatively, the lysis buffer was added into adherent cells and scraped off with plastic scrapper. The lysate cells were then homogenised by p200 pipette and sonicated for 15-30 seconds or treated with Benzonase Nuclease (Sigma, E1014) to digest the DNA before storing them at -80^{o}C. The frozen lysates were thawed on water and kept on ice during all process.

The total protein extracts were quantified either by Coomassie blue staining or by Pierce™ BCA Protein Assay kit (Life Technologies, 23225) according to the manufacturer's instructions or based on β-Actin. For Coomassie blue staining, whole cell lysates were mixed with 1X NuPAGE LDS sample buffer (Life Technologies) and incubated at 70^{o}C for 10 minutes. After loading the samples into 4-12% or 12% acrylamide gel (NuPAGE Novex Bis-Tris gel from Life Technologies), the proteins were separated by electrophoresis in MOPS buffer for 40 minutes. The gel was then stained with simply blue safe stain (Life Technologies, LC6060) for 30 minutes. Equal protein concentration for different sample was then loaded on a subsequent gel for immunoblotting. For quantification based on β-Actin, the western immunoblotting assay was performed with antibody against β-Actin. Then, the samples were quantified based on the β-Actin band intensity of each sample using Aida program.

2.7.2 Western immunoblotting assay

Once the protein samples were quantified, the immunoblotting was done using the NuPAGE electrophoresis system (Invitrogen) as manufacture instruction with some modifications. The reducing and non-reducing conditions were both used but the reducing condition mostly used with high MW proteins. After denaturing

the sample by heating at 70^{o}C for 10 minutes and loading samples in wither gradient 4-12% or non-gradient 12% acrylamide gel (NuPAGE Novex Bis-Tris gel from Life Technologies), the gel was run at 200V constant for 35 minutes in MOPS SDS running buffer (Life Technologies). Although NuPAGE electrophoresis system describes the most common western transferring method using the XCell II Blot Module, the proteins were transferred to PVDF membranes using Trans-Blot^{R} TurboTM transfer system (Bio-Rad). PVDF membranes (Bio-Rad, 21703) were subsequently blocked with blocking buffer [1X PBS buffer with 0.05% Tween-20 and 5% non-fat dairy milk (Marvel)] for 1 hour at room temperature on an orbital shaker. The membranes were then incubated overnight at 4^{o}C on shaker with primary antibodies diluted in the same blocking buffer as mentioned previously at the various dilutions (Table 2.8). After washing the membranes with 1X PBS buffer with 0.05% Tween-20 three times for 5 minutes each, they were incubated for 60-120 minutes with the appropriate horseradish peroxidase HRP-conjugated secondary antibodies (Table 2.8) in blocking buffer. Finally, the reaction of bound antibodies was detected with enhanced chemiluminescence reagent called SuperSignal™ West Dura Extended Duration Substrate (Pierce) ECL reagent using a CCD camera imaging system (LAS3000, Fuji). Novex sharp prestained ladder (Invitrogen, LC5800) was used.

The chromatin samples from different lines were also subjected for western blotting assay. However, before proceeding to immunoblotting assay 200 µl of sonicated chromatins was treated by Benzonase Nuclease (Sigma, E1014) for 15 minutes at room temperature as following reaction:

	Volume (µl)	Final concentration
Chromatin samples	200	-
10X TBS (200 mM Tris pH7.5 and 1.5 M NaCl)	20	1X
10% Triton X-100	20	1%
100 mM $Mgcl_2$	2.4	1X
Benzonase (250 U/µl)	1	1 U/µl

The reaction was reversed cross-linked by adding 2 µl of 5 M NaCl at 70^{o}C for 1 hour. The samples (20 µg, 2-6 µl) were then used for western blot assay.

Antibody	Species raised	Supplier	Dilution	Catalogue number
Nestin	Mouse	Abcam	1 µg/ml	2Q178
MAP2	Chicken	Abcam	1/2000	Ab5392
βIII-tubulin	Rabbit	Abcam	1 µg/ml	Ab18207
NF-160	Rabbit	Novus	1/5000	NB300-133
GFAP	Rabbit	Abcam	1/2500	Ab7269
HMGN1	Rabbit	-	1/1000	-
HMGN2	Rabbit	-	1/4000	-
Oct4	Rabbit	Abcam	1 µg/ml	Ab19857
β-Actin	Goat	Santa Cruz	1/1000	Sc-1615
Nanog	Rabbit	Abcam	1/1000	Ab80892
H3K9ac	Rabbit	EMD Millipore	1/1000	07-352
H3K27ac	Rabbit	EMD Millipore	1/1000	07-360
H3K14ac	Rabbit	EMD Millipore	1/2000	07-353
H3K4me3	Rabbit	EMD Millipore	1/1000	07-473
H3K27me3	Rabbit	EMD Millipore	1/1000	07-449
H3K36me2	Rabbit	EMD Millipore	1/1000	07-369
H3K4me1	Rabbit	EMD Millipore	1/1000	07-436
H3 pan ct	Rabbit	EMD Millipore	1/20000	07-690
Stabilized Peroxidase conjugated Goat Anti-Rabbit	Goat	Life technology	1/1000	32460
Polyclonal Rabbit anti-Goat immunoglobulins/HRP	Rabbit	Dao	1/1000	P0160
Stabilized Goat anti-Mouse HRP-Conjugated	Goat	PIERCE	1/1000	1858413

Table 2.8: The primary and secondary antibodies used for immunoblotting.

2.8 Total mRNA extraction and quantitative real-time PCR (qRT-PCR)

2.8.1 Total RNA extraction

The RNA was isolated from undifferentiated and differentiated P19 cells using RNeasy® Mini kit (Qiagen) and an on-column DNase digestion with the RNase-free DNase set (Qiagen, 79254) according to the manufacturer's instructions. Briefly, after washing the cell pellets with PBS, 200-500 µl of RLT lysis buffer was added and then homogenised by pipetting with P200 several times. One volume of 70% ethanol was added to homogenised lysate and mixed by pipetting. The mixture,

around 700 µl, was then transferred into RNeasy spin column and centrifuged at 10,000 rpm for 15 seconds. 350 µl Buffer RW1 was added to RNeasy column and centrifuged at the same speed mentioned before. DNase digestion was conducted at this point by adding 80 µl of DNase I (10 µl DNase stock and 70 µl RDD buffer) directly to column and incubated for 15 minutes at room temperature. Again, 350 µl of RW1 buffer was added to RNeasy column and centrifuged at the same speed and time. 500 µl of RPE buffer was added twice to the RNeasy spin column and centrifuged at 10,000 rpm for 15 seconds. The RNA was finally eluted into 1.5 ml RNase-free microcentrifuge tubes by adding 30–50 µl RNase-free water directly to the spin column membrane. RNA concentration was determined by spectrophotometer Nanodrop. The quantified RNA was either used immediately for cDNA synthesis or stored at -80°C.

2.8.2 cDNA synthesis

After extracting the total RNA, the first strand cDNA was synthesised using SuperScript™ III kit from Invitrogen. In addition, oligo $(dT)_{20}$ primer was used for cDNA synthesis to reverse transcribed polyadenylated transcripts only. Three main steps, including denaturation and annealing, reverse transcription and finally RNA digestion, were performed consequently as following:

2.8.2.1 Denaturation and annealing

The following reagents were combined and incubated at 65°C in thermo-cycler for 5 minutes and then cooled on ice rapidly for 5 minutes.

Reagents	Final volume (µl)
300 ng of extracted RNA in 10 µl of dh_2O	10
10 mM dNTPs	0.5
100 µM Oligo $(dT)_{20}$	0.5
Total	11

2.8.2.2 Reverse transcription

The cDNA synthesis master mix was set up as following:

Reagents	Final volume (µl)
5X RT buffer	3
DTT (0.1M)	1
RNaseOUT™	0.5
Super Script™ III	0.5
Total	5

Then, 5 µl of master mix was added to each mixture (RNA, dNTPs and primer) of pervious step (2.8.2.1). After adding the master mix to reaction, the mixture was incubated at 25ºC for 10 minutes following 50ºC for 50 minutes. Finally, the reaction was heat inactivated at 85ºC for 5 minutes.

2.8.2.3 RNA digestion

The pervious reaction was firstly cooled on ice for 30 seconds. Subsequently, the RNA was digested by adding 1 µl of RNase H to the reactions and incubated for 20 minutes at 37ºC. The cDNA then was used directly for real-time amplification.

2.8.3 The quantitative real-time PCR (qRT-PCR)

The cDNA synthesis was then followed by qRT-PCR to profile the gene expression of different genes. The cDNA used in qRT-PCR amplification was firstly diluted 1:5 ratios, and 5 µl of diluted cDNA was used in each PCR reaction. The PCR reaction was performed using transparent 96-well PCR plates (Abgene) in 25 µl total volume for each reaction, typically in triplicate for each sample. Each reaction contained all the reagents as shown in Table 2.9.

SYBER-green assay	Final volume (µl)
cDNA	5
SYBR PCR mix	12.5
Forward primer (2.5-7.5 µM)	3
Reverse primer (2.5-7.5 µM)	3
dh_2O	1.5

Table 2.9: The reaction set up for qRT-PCR.

qRT-PCR was performed using FastStart Universal SYBR Green Master (04913914001, Roche) in the MxPro 3000P thermo cycler and detection system (Stratagene). The SYBR-green assay is principally based on the detection of double stranded PCR products that were amplified by oligonucleotide primers for each specific gene. The thermal cycle profile for qRT-PCR is given in Table 2.10. A step of dissociation curve (melting curve) was included in each reaction to assess whether the detectable fluorescent dye is for single and specific products.

Segment	Cycles number		
1	1	10 minutes	95ºC
2	40	15 seconds	95ºC
		1 minute	60ºC
3	1	1 minute	95ºC
		30 seconds	55ºC
		30 seconds	95ºC

Table 2.10: The thermal cycle profile showing different cycle for qRT-PCR.

The PCR primers used for the gene expression profile and their sequences are shown in Table 2.11. Most of primer sets were designed by pervious PhD students (Mohan, thesis 2012 and Rehbini, PhD thesis 2016) using the Primer 3 software and taking all criteria for primer design in consideration. These primers are mainly designed in the exon-exon junction with melting temperature between 55 to 60^{o}C. In addition, the concentration and efficiency of each primer sets were optimised by comparing the dissociation curves and the standard amplification curves for detecting the primer dimer structure and checking PCR efficiency, respectively, in + and - RT.

Gene	Primer	Sequence	Primer efficiency
Hmgn1	Forward Reverse	AGA GAC GGA AAA CCA GAG TCC AG CGT GAT GGA TGC TTA GTC GGA	1.98
Hmgn2	Forward Reverse	AAA AGG CCC CTG CGA AGA A TGC CTG GTC TGT TTT GGC A	1.96
Oct4	Forward Reverse	CGT TCT CTT TGG AAA GGT GTT CA GGT TCT CAT TGT TGT CGG CTT C	1.98
Sox2	Forward Reverse	GGA ACA GCA TGG CGA GCG G CGT TCA TGT GCG CGT AGC TG	1.96
Nanog	Forward Reverse	ACC TGA GCT ATA AGC AGG TTA AG TCA GAC CAT TGC TAG TCT TC	2.0
Nestin	Forward Reverse	AAA GTT CCA GCT GGC TGT CAC TTC CAG ACT AAG GGA CAT	1.95
Ngn1	Forward Reverse	TTC ATG CAT TAT GGA TCC CTC CAG TCC AGT GCC TGA ATA	--
Mash1	Forward Reverse	CCA ACT GGT TCT GAG GAC CTG CTG CCA TCC TGC TTC CAA A	1.93
NeuroD	Forward Reverse	TCT TTG AAA GCC CCC TAA CTG AAGTTG CCA TTG ATG CTG AGG	--
Map2	Forward Reverse	TCT GCC TCT AGC AGC CGA AG CAC TGT GGC TGT TTG TTC TG	1.96
Nf160	Forward Reverse	CTC AGC AGC TAC CAG GAC AC CGA TCT CGA TGT CCA GGG CC	1.94
Zfp521	Forward Reverse	GAG CGA AGA GGA GTT TTT GG AGT TCC AAG GTG GAG GTC AC	1.95
Pax6	Forward Reverse	GGC GGA GTT ATG ATA CCT ACA C GAA CTT GGA CGG GAA CTG AC	--
Rest	Forward Reverse	GTG CGA ACT CAC ACA GGA GAA CG AGT CCG CAT GTG TCG CGT TAG A	1.94
β-Actin	Forward Reverse	GTG AAA AGA TGA CCC AGA TC GTG TGG GTG ACC CCG TCT CC	1.97
Gpi1	Forward Reverse	TCC GTG TCC CTT CTC ACC AT TGG CAG TTC CAG ACC AGC TT	(Bouma et al. 2004)
Rplp0	Forward Reverse	GCT TTC TGG AGG GTG TCC G ACG CGC TTG TAC CCA TTG AT	--
Gapdh	Forward Reverse	GAT GCC CCC ATG TTT GTG AT GGT CAT GAG CCC TTC CAC AAT	2.0
α-tubulin	Forward Reverse	ACC CAC GGT CAT CGA TGA AGT T TCC TTG CCA ATG GTG TAG TGG C	1.98

Table 2.11: Different primers used for gene expression profile.

2.8.3.1 qRT-PCR calculation and statistic analysis

Changes in gene expression during neural induction (single biological replicates)

The fold change in gene expression in differentiating neuronal cells compared to undifferentiated cells was calculated using the comparative delta-delta Ct method. All starting Ct values used in the calculations are the average of technical triplicate qPCR reactions. Following the recommendation by different studies of using more than single gene for normalisation of qRT-PCR data (Vandesompele et al., 2002; Hellemans et al., 2004; Derveaux et al., 2010), three housekeeping genes were used as controls: *RPLP0*, *GP*I1 and *GAPDH* (*ribosomal protein large P0*, *Glucose-6-Phosphate Isomerase 1* and *Glyceraldehyde 3-phosphate dehydrogenase*, respectively). This is thought to be more robust as it reduces reliance on a single housekeeping gene, which could change unexpectedly in response to the experimental conditions.

$$Ct_{(housekeeping)} = (Ct_{(RPLPO)}+Ct_{(GPI1)+}Ct_{(GAPDH)})/3$$

$$\Delta Ct = Ct_{(gene)} - Ct_{(housekeeping)}$$

$$\Delta\Delta Ct = \Delta Ct_{(experimental\ condition)} - \Delta Ct_{(control\ condition)}$$

$$\text{Fold change} = 2^{-\Delta\Delta Ct}$$

The control condition is undifferentiated parental cells in all cases, and the experimental conditions are the different time points after neuronal differentiation in parental and knockout cells.

Error bars represent the standard deviation (sd) of the technical triplicate PCR reactions, and were calculated as follows:

$$error_{(\Delta Ct)} = \sqrt{(2^{sd(gene)} + 2^{sd(housekeeping)})}$$

$$error_{(\Delta\Delta Ct)} = \sqrt{(2^{error(\Delta Ctexp)} + 2^{error(\Delta Ctcontrol)})}$$

$$\text{Up error}_{(fold\ change)} = 2^{-(\Delta\Delta Ct - error(\Delta\Delta Ct))} - \text{fold change}$$

$$\text{Down error}_{(fold\ change)} = \text{fold change} - 2^{-(\Delta\Delta Ct + error(\Delta\Delta Ct))}$$

Relative gene expression in HMGN2 knockout cells compared with parental cells (multiple biological replicates)

All starting Ct values used in the calculations are the average of technical triplicate qPCR reactions.

For each RNA sample:

$Ct_{(housekeeping)} = (Ct_{(RPLPO)}+Ct_{(GPI1)+}Ct_{(GAPDH)})/3$

$\Delta Ct_{(gene/HK)} = Ct_{(gene)} - Ct_{(housekeeping)}$

gene expression relative to housekeeping genes: $expression_{(gene/HK)} = 2^{-\Delta Ct\ (gene/HK)}$

The average gene expression (relative to housekeeping genes) and its standard deviation were then calculated from all biological replicates from a particular cell line:

Av exp = (rep 1 $expression_{(gene/HK)}$ + rep 2 $expression_{(gene/HK)}$ + rep 3 $expression_{(gene/HK)}$)/3

Average expression in each cell line was normalised to that in parental cells to give relative expression as follows:

Relative expression = $Av\ exp_{(KO\ cells)}$ / $Av\ exp_{(wt\ cells)}$

Student's two-tailed T-tests were performed using the values for $expression_{(gene/HK)}$ from each replicate in order to determine whether expression in HMGN2 knockout cell lines was significantly different to that in parental cells. P values <0.05 were considered to be significant.

2.9 Immunofluorescence (IF)

Immunofluorescence staining was conducted to examine different pluripotent and different neural linage markers. Standard cell culture was applied to grow cells and specifically $0.5\text{-}1X10^5$ P19 cells were seeded into 24 well plate containing sterilised coated cover slips. If the cells were induced into neural linage the cover-slips were coated with 5 µg/ml laminin (Sigma, L2020). Undifferentiated P19 and neural induced cells at different time points were fixed by adding 250 µl of 4% paraformaldehyde in PBS for 20 minutes at room temperature, followed by incubating the cells with 100 µl chilled methanol at -20ºC for 1 minutes. The cells were then washed by 1X PBS. The permeabilisation step was then performed using 0.1% tritonX-100/PBS for 10 minutes at room temperature. The permeabilised cells were washed once with PBS and 20mM glycine for 10 minutes. Subsequently, the cells were blocked by 100 µl IF blocking buffer (0.3% tritonX-100 and 2.5% horse serum in PBS) for an hour at room temperature. The primary antibodies (Table 2.12) diluted in IF blocking

buffer were added to fixed cells and incubated for 1 hour at room temperature. After washing the excess antibody three times with 0.3% tritonX-100/PBS for 10 minutes each, the reaction was detected by conjugated secondary antibodies (Alexa Flour 488, Table 2.12) diluted in IF blocking buffer and incubated for 1 hour in a dark place at room temperature. Then, the cells were washed three times with PBS for 10 minutes each and once with dh_2o. The cells were finally mounted with ProLong Gold Antifade Mountant with DAPI (Life Technology, P36931) to counter-stain the nuclei. The slides were left in dark place for 24 hours before storing them at 4°C. The slides were left at room temperature for at least 30 minutes before they were examined and pictured by fluorescent microscope (Olympus 1X51) containing camera. The images were taken and coloured using Cell P program and the scale bars were set using Image J.

Antibody	Species raised	Supplier	Concentration	Catalogue number
Nestin	Mouse	Abcam	1/300	2Q178
MAP2	Chicken	Abcam	1/5000	Ab5392
βIII-tubulin	Rabbit	Abcam	1 µg/ml	Ab18207
NF-160	Rabbit	Novus	1/1000	NB300-133
GFAP	Rabbit	Abcam	1/1000	Ab7269
HMGN1	Rabbit	-	1/1000	-
HMGN2	Rabbit	-	1/1000	-
Oct4	Rabbit	Abcam	1/500	Ab19857
Alexa Fluor Goat anti-mouse-488	Goat	Life Technology	1/1000	A-21121
Alexa Fluor Donkey anti-rabbit-594	Donkey	Life Technology	1/1000	A-21207
Alexa Fluor Donkey anti-rabbit-488	Donkey	Life Technology	1/1000	A-21206
Alexa Fluor Goat anti-mouse-594	Goat	Life Technology	1/1000	A-11020
Alexa Fluor Goat anti-chicken-594	Goat	Life Technology	1/500	A-11042
Alexa Fluor Donkey anti-goat-488	Donkey	Life Technology	1/1000	A-11055

Table 2.12: The primary and secondary antibodies used in immunofluorescent assay.

2.10 Proliferation assay (EdU)

The proliferation assay was carried out using the Click-iT® plus EdU Alexa Fluor® 488 imaging kit (Life Technology, C10637) as manufacturer's instructions. EdU

(5-ethynl-2'-dexyuridine) which is a nucleoside analog of thymidine is incorporated into DNA during active DNA synthesis and detected by a click reaction. The click reaction is based on a copper catalysed covalent reaction between a picolyl azide on Alexa flour and an alkyne that is on EdU. Briefly, 0.5-1X10^5 of either undifferentiated and differentiated cells at different time points were grown on coverslips for 24 hours before treatment. The cells were incubated with 10 μM EdU for additional 24 hours. The cells were then fixed by 4% paraformaldehyde in PBS for 15 minutes, followed by 0.5% tritonX-100 as a permeabilisation step for 20 minutes after washing the cells with 3% BSA in PBS twice. The reaction was detected by incubating the cells with Click-iT plus reaction cocktail (see the instruction for preparation) to each well for 30 minutes at room temperature in a dark place. The reaction cocktail was removed and cells were washed once with 1 ml of 3% BSA in PBS. The cells were washed with PBS and then either mounted with Prolong Gold Mountant containing DAPI or proceeded for antibody protein labelling following the pervious procedures (section 2.9, from the blocking stage).

2.11 Cell cycle analysis

For cell cycle analysis by quantification the DNA content, a fluorescent nucleic acid dye (propidium iodide, PI) was used to identify the cell proportion that are in three interphase stages (G1, S phase, G2) of the cell cycle. 2X10^6 of cells were harvested in normal manner (section 2.1.2). The cells were washed with PBS and resuspended in 250 μl PBS. 750 μl of absolute ethanol was added by drop-wise with vortex to ensure the fixation of all cells and minimise the cell-clump formation. The cells were kept with ethanol for 30 minutes on ice. The ethanol was removed by centrifugation at higher speed for 5 minutes. The cells pellets were washed twice with PBS (higher speed used). Subsequently, the cell pellets were treated with Ribionuclease A (50 μl of 100 μg/ml RNase A) in order to remove the RNA and to ensure that only the DNA is stained. 800 μl of 40 μg/ml PI solution per 2 million cells were added directly to cells with RNase A. The cells were mixed well and incubated for 30 minutes at room temperature and protected from the light. 20,000 cells were analysed by FACS in the PI/RNase A solution at a low flow rate (100 events/seconds).

2.12 Chromatin immunoprecipitation (ChIP)-QPCR

The cells from different lines were cultured and expanded as normal (2.1.2) in 15 mm dishes. One plate from different lines with 70-80% confluence was trypsinised and counted. The number of plates required for $7.5X10^7$ was calculated based on the number of cell count per plate. The medium of each plate was changed by 20 ml advanced DMEM/F12 without serum since the serum can reduce the formaldehyde fixation efficiency. The chromatin was cross-linked by adding 2.2 ml of 10% formaldehyde (1% final concentration) drop-wise to the 20 ml and cells. The 10% formaldehyde was prepared from 16% stock diluted in formaldehyde solution (50 mM HEPES pH 8.0, 0.1 M NaCl, 1 mM EDTA and 0.5 mM EGTA). The dishes were rocked on orbital shaker for 10 minutes at room temperature. The cross-linking reactions were quenched by adding either 1.5 or 1 ml of 1 or 1.5 M glycine, respectively. The cells were washed by 20 ml pre-chilled PBS and resuspended in 5 ml pre-chilled PBS before scraping them off and collecting them into 50 ml falcon tubes. Subsequent work was performed on ice and pre-chilled buffers and solutions were used unless otherwise specified. The cells were pelleted by centrifuging at 1200 rpm at 4^oC for 5 minutes. The pellet was lysed in RIPA buffer (10 mM Tis-HCl pH 7.6, 1 mM EDTA pH 7.6, 0.1% SDS, 0.1% sodium deoxycholate, 1% triton X-100, plus 10 mM sodium butyrate) containing 1X protease inhibitor (cOmplete, Mini, EDTA-free Cocktail Tablets from Roche Applied Science, 11836170001, 1 tablet per 10 ml). 1 ml of RIPA buffer with inhibitors was added per $7.5X10^7$ cells and left on ice for more than 10 minutes. The lysates were then stored at -80^oC till sonication.

At the day of sonication, the lysates were thawed on wet ice and the chromatin was sheared by sonicator (SONICATOR 3000) for 4.5-8.5 minutes (total process time) depending on the cell number used. The sonication process was carried out at amplitude 3, 10 seconds on, 30 seconds off and power voltage 10%. Then, 25 μl aliquots of each sonicated chromatin were reverse cross-linked by adding 4 μl of 5 M NaCl, 4 μl RNase A, 8 μl of 20 mg/ml Proteinase K and 75 μl ddH_2O. The reactions were incubated at 56^oC for 1 hours and DNA was purified by QIAquick PCR clean up kit (Qiagen). The samples were quantified by Nanodrop and 300 ng was run in 1% agarose gel electrophoresis for 1 hour. Lysates were

sonicated to result in the most of DNA fragments with 100-1000 bp length. Lysates were divided into 250 μl aliquots and stored at -80°C.

For immunoprecipitation, 250 μl aliquot of each sample were thawed on ice and additional 10 mM sodium butyrate (deactylase inhibitor), sodium orthovanadate (phosphatase inhibitor) and 1X protease inhibitor were added to aliquots. The lysates were centrifuged at 12,750 rpm at 4°C for 5 minutes, and the supernatants containing the sonicated chromatin were transferred to a new tube. In parallel, 50 μl magnetic Protein A (Invitrogen, 10002D) were blocked and conjugated to an antibody by washing and resuspending twice in PBS, 0.5% BSA, 0.5% Tween-20. Then, the antibodies (see Table 2.13) were added and bound to the beads by rotating for 1 hour at room temperature. Blocked antibody-conjugated beads were then placed on a magnet and the supernatant was removed. Subsequently, around 5 μg (equivalent to $1X10^6$ cells) of sonicated lysates per each IP were added to the beads followed by incubation for 3 hours at 4°C with rotation (half the amount of sonicated lysates was included for the input that left on ice till the elution step). After the incubation, the beads were washed subsequently twice with RIPA buffer (10 mM Tis-HCl pH 8.0, 1 mM EDTA pH 8.0, 140 mM NaCl, 0.1% sodium deoxycholate, 0.1% SDS, 1% triton X-100 and 1X protease inhibitor), twice with RIPA-500 (10 mM Tis-HCl pH 8.0, 1 mM EDTA pH 8.0, 500 mM NaCl, 0.1% sodium deoxycholate, 0.1% SDS, 1% triton X-100), once with RIPA-LiCl (10 mM Tis-HCl pH 8.0, 1 mM EDTA pH 8.0, 250 mM LiCl, 0.5% sodium deoxycholate, 0.5% NP40, 1% triton X-100), and twice with TE pH 8.0 (10 mM Tis-HCl pH 8.0 and 1 mM EDTA). The beads of each sample and the input were then incubated with 70 μl of elution buffer (0.5% SDS, 300 mM NaCl, 5mM EDTA, 10 mM Tris-HCl pH 8.0) and 2.5 μl of Proteinase K for 1 hour at 55°C to revert formaldehyde cross-linking. However, 1 μl of RNase A was also added to the input only. Then, the supernatants were transferred to new tubes, and DNA was finally purified by using Qiagen Minelute column (PB buffer). 50 μl Qiagen elution buffer (10 mM Tris pH 8.0) was used.

Antibody	Species raised	Supplier	µl per IP	Catalogue number
H3K9ac	Rabbit	EMD Millipore	3	07-352
H3K27ac	Rabbit	EMD Millipore	3	07-360
H3K14ac	Rabbit	EMD Millipore	3	07-353
H3K4me3	Rabbit	EMD Millipore	2.5	07-473
H3K27me3	Rabbit	EMD Millipore	3	07-449
H3K4me1	Rabbit	EMD Millipore	2	07-436
H3 pan ct	Rabbit	EMD Millipore	2	07-690
IgG	Rabbit	Sigma	2	I5006
HMGN2	Rabbit	--	2.5	--

Table 2.13: The antibodies and their concentrations used for immunoprecipitation of ChIP assay.

qPCR for ChIP

The ChIP DNA from each IP (50 µl) was diluted as 1 in 5 by adding 200 µl of 10 mM Tris-HCl, pH 8.0 while the input DNA was quantified and 0.1 ng/µl was used. qPCR was performed using FastStart Universal SYBR Green Master in the MxPro 3000P thermo cycler and detection system (Stratagene). A master mix for each primer set was as shown in Table 2.14.

Master mix	Final volume (µl)
2X SYBR green PCR master mix (contain fluorescent dye, polymerase, dent's, buffer)	12.5
Forward & reverse primers	5
dh_2O	2.5
Total	20

Table 2.14: The components of the master mix prepared for each primer set for ChIP-qPCR.

Then, 5 µl of each diluted DNA was added into well of transparent 96-well PCR plates (Abgene) and 20 µl of master mix were then added to form 25 µl total volume for each reaction. The electronic pipettor was used for maximising the accuracy. Each sample was typically performed as triplicates. After adding the diluted DNA and master mix, the plate was sealed and centrifuged for 5 minutes at 3000 rpm and 4^{o}C. The thermal cycle profile for qRT-PCR are given in Table 2.15. A step of dissociation curve (melting curve) was included in each reaction to assess whether the detectable fluorescent dye is for single and specific products.

Segment	Cycles number		
1	1	10 minutes	95ºC
2	40	30 seconds	95ºC
		1 minute	60ºC
3	1	1 minute	95ºC
		30 seconds	60ºC
		30 seconds	95ºC

Table 2.15: The thermal cycle profile showing different cycle for q-PCR.

The PCR primers used for ChIP-qPCR and their sequences are shown in Table 2.16. The primer sets were previously designed by Katherine West (Barkess et al., 2012) and PhD student (Rehbini, PhD thesis 2016) was previously tested.

All starting Ct values used in the calculations are the average of technical triplicate qPCR reactions.

For each qPCR primer set, the Ct value obtained from the ChIP for HMGN2 or a histone modification was normalised to the Ct value obtained from 0.5 ng of input DNA, using the ΔCt calculation. Normalisation to input controls for varying representation of genomic regions in the starting chromatin sample, and for differences in PCR primer amplification efficiency.

$\Delta Ct = Ct_{(ChIP)} - Ct_{(input)}$

Fold change = $2^{-\Delta Ct}$

The fold change represents the relative level of HMGN2 binding, or the relative amount of the histone modication, at that primer set. The scale is arbitrary, as the 0.5 ng of input used in the qPCR reaction was chosen arbitrarily.

Error bars are derived from the standard deviation (sd) of the technical triplicate PCR reactions, and were calculated as follows:

$error_{(\Delta Ct)} = \sqrt{(2^{sd(ChIP)} + 2^{sd(input)})}$

$Up\ error_{(fold\ change)} = 2^{-(\Delta Ct - error(\Delta Ct))} - fold\ change$

$Down\ error_{(fold\ change)} = fold\ change - 2^{-(\Delta Ct + error(\Delta Ct))}$

In some figures, as indicated in the legend, data were further normalised to the average fold change for H3 across all primer sets tested. This control for varying concentrations of “ChIP-able” nucleosomal DNA in each chromatin

sample, which can vary depending on the efficiency of crosslinking and sonication.

Example:

average fold change for H3 = (H3 fold change$_{(primer\ a)}$ + H3 fold change $_{(primer\ b)}$ + H3 fold change$_{(primer\ c)}$ / 3

H3K4me3/H3 at primer set a = H3K4me3 fold change$_{(primer\ a)}$ / average fold change for H3

Gene	Position	Primer	Sequence	Final concentration
Nanog	-1081	Forward Reverse	GGAAGAACCACTCCTACCAATACTCA CGTAACATCTCCCATGTGAAGACTC	300 nM 900 nM
Nanog	-219	Forward Reverse	TCTTTAGATCAGAGGATGCCCCCTAAGC AAGCCTCCTACCCTACCCACCCCCTAT	300 nM 300 nM
Nanog	929	Forward Reverse	TCAGCCCAGTACTCAGGCTTGT AGCCTAGCAGCCTCTTGGTTCT	300 nM 300 nM
Nanog	1740	Forward Reverse	TAACTGGACCCTCTGACTGGCT CCCACCATCTTTTCTGCTAGTACAAG	300 nM 300 nM
Glyt1	-5394	Forward Reverse	CTGGGCATGCAGGGCGTACT TGCCAGGAGCAATGGCGGAA	300 nM 900 nM
Glyt1	312	Forward Reverse	CCCCAGGATGTGTATGGATGT CCACAGATACCCGCAAGCA	300 nM 900 nM
Oct4-1	-1781	Forward Reverse	GTGAGCATGACAGAGTGGAGGAA TCTCTGGCCCTCTCCATGAAT	300 nM 900 nM
Oct4-2	-399	Forward Reverse	GTGGGTAAGCAAGAACTGAGGA TGGAGAGCCTAAAACATCCATT	300 nM 900 nM
Oct4-3	410	Forward Reverse	CAATGCCGTGAAGTTGGAGA TCACTTACCTCCTCGGGAGTTG	300 nM 900 nM
Oct4-4	2070	Forward Reverse	TCAAGGCTTCTCACCTCCAGA ACAGAGAAAGCAAGGCAGCAG	900 nM 300 nM
Oct4-5	3342	Forward Reverse	GCTGCCTCACTCACTCTGTTTG TGGCTGAACACCTTTCCTGAA	300 nM 300 nM
β-Actin	-	Forward Reverse	CGCCATGGATGACGATATCG CGAAGCCGGCTTTGCACATG	300 nM 300 nM
M-satellite /Msat	-	Forward Reverse	GACGACTTGAAAAATGACGAAATC CATATTCCAGGTCCTTCAGTGTGC	300 nM 300 nM

Table 2.16: Primer sets used in ChIP-qPCR assay.

Chapter 3 Generating HMGN2 functional knocked-out lines using CRISPR-Cas9 system that targets exon IV

3.1 Introduction

The ability to disrupt specific genes at the genomic or RNA level sequences is a powerful approach for understanding their functions. RNA interference (RNAi) has been widely used to investigate gene function; however, it has some drawbacks such that it provides a temporary and incomplete inhibition of gene expression, and can cause unpredictable off-target effects (Shalem et al., 2014). Another strategy is to target the genome of living cells and make permanent gene modifications using knock-out or knock-in technology (Hsu et al., 2014). This is mainly achieved by homologous recombination that involves integrating an exogenous DNA sequence into the genomic DNA of cells. Because homologous recombination events occur infrequently, this may present a difficulty for specific genomic sites or for large-scale uses of gene targeting (Capecchi, 1989; Hsu et al., 2014).

Genome editing of cell lines in a precise and targeted manner has been greatly facilitated by using the CRISPR-Cas9 system, which is based on RNA-guided nucleases (Sander and Joung, 2013). CRISPR-Cas9 is an effective tool for multiple gene editing purposes. For instance, when Cas9 is directed by guide RNA (gRNA) to the coding region of a gene, it may create frameshift mutations that result in premature stop codons and truncated proteins (Ran et al., 2013a). Other applications include deletions generated by the use of multiple gRNAs. For example, CRISPR-Cas9-mediated deletion of several exons of the dystrophin gene was shown to restore the correct reading frame and generate functional dystrophin protein in myoblasts from patients with Duchenne muscular dystrophy (Ousterout et al., 2015). Moreover, because the targeting specificity of Cas9 depends only on gRNA sequence, several studies have developed methods for large-scale, systematic, genetic screening in human and mouse cells using lentiviral vectors that are able to deliver pooled gRNAs (Shalem et al., 2014; Wang et al., 2016).

However, one major concern of using CRISPR-Cas9 in gene targeting is the off-target activity, which could cause genome instability and disruption of other genes (Wang et al., 2015). Using paired gRNAs and Cas9D10A or Cas9H840A mutants to mediate double stranded DNA nicks has been shown to significantly reduce the off-target mutagenesis rate (Shen et al., 2014; Ran et al., 2013b). This is because the chance of both gRNAs binding close to each other, elsewhere in the genome is rare (Shen et al., 2014). Furthermore, according to Fu et al (2014), the Cas9D10A nickase can be combined with truncated 17-18 base gRNAs, which further reduces off-target cleavage without compromising the efficiency of on-target mutation.

The CRISPR-Cas9 or –Cas9D10A system can simultaneously target multiple distinct genomic loci by co-expressing a single Cas9 or Cas9D10A protein with two or more sgRNAs. However, co-transfecting multiple gRNA plasmids could reduce transfection efficiency, or cause variable levels of expression because of differences in copy number. To avoid this problem, several groups have developed delivery systems that combine all essential components for CRISPR-Cas9 in single vector (Kabadi et al., 2014; Sakuama et al., 2014; Chiang et al., 2016). By using an all-in-one vector, the efficiencies of transfection and gene mutagenesis are maximised, particularly during multiplex genome editing (Sakuma et al., 2014).

In this chapter, we have used a single vector system developed by Kabadi et al (2014), which is able to express Cas9 or its variants and up to four gRNAs from independent RNA polymerase III promoters. The use of different promoters is to minimise the recombination caused by sequence repetition, which leads to plasmid instability, and to maximise the expression efficiency. The expression cassette for each gRNA is incorporated into final vector by a convenient Golden Gate assembly method that is well established for modular assembly of the DNA-binding repeats of TALENs (Kabadi et al., 2014; Sakuma et al., 2014). Golden Gate assembly is facilitated by the use of Type II(S) restriction enzymes, such as *BsmB*I, which cleave a set number of bases to one side of the recognition site. This allows the design of different overhangs to specify the position and orientation of fragments during a mixed ligation reaction. This type of enzyme is active in the same ligation buffer as T4 DNA ligase.

3.2 Objectives

The main aim of this chapter is to use the CRISPR-Cas9 nickase technology to generate several lines of P19 cells in which the nucleosome binding domain (NBD) of HMGN2 is functionally disrupted, and characterise the mutations in these lines.

1. Design CRISPR-Cas9D10A nickases for targeting exon IV of *Hmgn2* gene locus.

2. Creation of several vectors includes individual gRNA or multiple gRNA arrays (multiguides) vectors to carry out genome editing at exon IV of *Hmgn2* gene locus. This was done in collaboration with the laboratory of Adam West.

3. Transfection of either individuals or multigudes vectors in P19 cells to ensure the efficiency of different gRNAs for introducing mutation at that locus.

4. Generation and screening of different monoclonal lines from the poly-population of P19 cells transfected with the most efficient paired gRNAs vector.

5. Characterise the putative knock-out lines of P19 cells that generated in objective 3.

    a. Determine the exact mutations in all candidate lines by RFLP assay and sequencing.

    b. Characterise the remaining proteins in the candidate lines, if remains, and ensure that the mutant form of HMGN2 protein loss functionality in lines that have a disrupted nucleosome binding domain (NBD).

3.3 Design of CRISPR-Cas9D10A nickase for exon IV of *Hmgn2* gene

The *Hmgn2* gene, which is localised to chromosome 1p36.11, is composed of 6 exons interrupted by 5 introns and all the intron/exon junctions follow the "GT-AT" rule (Landsman et al., 1989; Popescu et al., 1990). Exons I and VI contain the uninterrupted 5' and 3' untranslated regions and the coding sequences for the first 5 and the last 11 amino acids, respectively (Landsman et al., 1989). The four middle exons, II, III, IV and V, encode 15, 10, 17 and 32 amino acids, respectively. The NBD, which is composed of 30 amino acids, is mainly encoded by exons III and IV (Figure 3.1). Targeting of the *Hmgn2* gene is complicated by the fact that human and mouse genomes contain more than 30 retropseudogenes for Hmgn2, which are similar to the cDNA sequence of *Hmgn2* (Popescu et al., 1990; Johnson et al., 1992 and 1993; Strichman-Almashanu et al., 2003). The intronless human *Hmgn4* is an expressed retropseudogene that has a very similar sequence to the Hmgn2 transcript: the genes are 86% identical at the DNA level across the open reading frame (Birger et al., 2001). It seems to have originated from an accidental insertion of an Hmgn2 retropseudogene next to an active promoter (Birger et al., 2001).

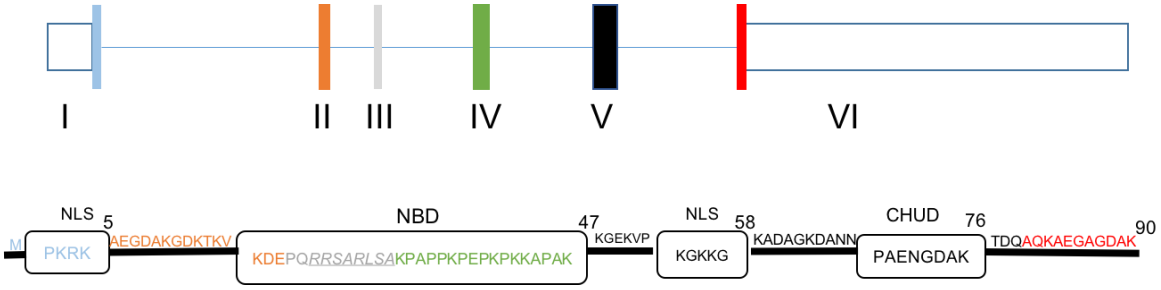


Figure 3.1:Schematic diagrams of *Hmgn2* gene and its protein sequence.
The first diagram represents the *Hmgn2* genes that has 6 exons (I-VI) interrupted by 5 introns. The first exon (I) and last exon (VI) have 5' and 3' untranslated regions (unfilled boxes). Each exon and its translated amino acids from that specific exon are given in same colour. The HMGN2 protein which is composed of 90 amino acids have 3 different domains (boxes) including: two nuclear localisation signal (NLS), nucleosome binding domain (NBD) and chromatin unfolding domain (CHUD). The amino acid sequence is given inside the boxes for the domains or on the lines. The numbers on the top correspond to the amino acid position of the last amino acid in each domain. Within the NBD, there is conserved motif (italic and underlined) that confers the binding to nucleosomes (Adapted from Landsman et al., 1989; Gonzalez-Romero et al., 2014).

We used the paired nickase and truncated gRNAs in order to ensure the maximum specificity for targeting the essential nucleosome binding domain of *Hmgn2*. Removal of this domain should prevent HMGN2 from binding to

nucleosomes. For targeting the NBD of the mouse *Hmgn2* gene, two different nickases were selected: Cas9_mN2_Ex4_A68/B88 (18 and 20 nucleotides, respectively) and Cas9_mN2_Ex4_A87/B91 (both are 17 nucleotides) (Figure 3.2). According to Fu et al (2014), either one or both truncated gRNAs in each nickase can direct cleavage and mutagenesis at the target site. Both nickases target exon IV of the *Hmgn2* gene which encodes part of the NBD. The reason for selecting two different nickases at the same region was because it was previously shown that not all gRNAs are equally active to produce null alleles (Cho et al., 2014; Fu et al., 2014). Each gRNA was selected based on several criteria including: recognition of both exon and intron sequence to avoid targeting pseudogenes, the presence of an indicative restriction enzyme site, and avoiding similar sequences found in the other *Hmgn* variants, especially *Hmgn4*.

In addition to specific considerations for targeting *Hmgn2*, different factors can affect the cooperative nicking that leads to the DSBs. Firstly, the distance between the two gRNAs (offset) plays a role in mutation efficiency. Ran et al, (2013b) showed that gRNA pairs with offsets between -4 and 20 bp lead to the highest NHEJ rates (up to 40%), but that offsets can be up to 100 bp and still have modest indel mutation rates. In addition to the offset window, the type of overhang also affects the efficiency of introducing mutations through NHEJ repair. The double nickase using Cas9D10A generates 5` overhangs, and has been reported to result in more NHEJ events than the 3`overhang generated by Cas9H840A (Mali et al., 2013b). However, Shen et al (2014) could not see differences between 5` or 3` overhangs in producing indel mutations. In this study, Cas9D10A is used to generated 5' overhangs, and the offsets of the two sets of gRNAs, Cas9_mN2_Ex4_A68/B88 and Cas9_mN2_Ex4_A87/B91, are 16 bp and 5 bp, respectively (Figure 3.2).

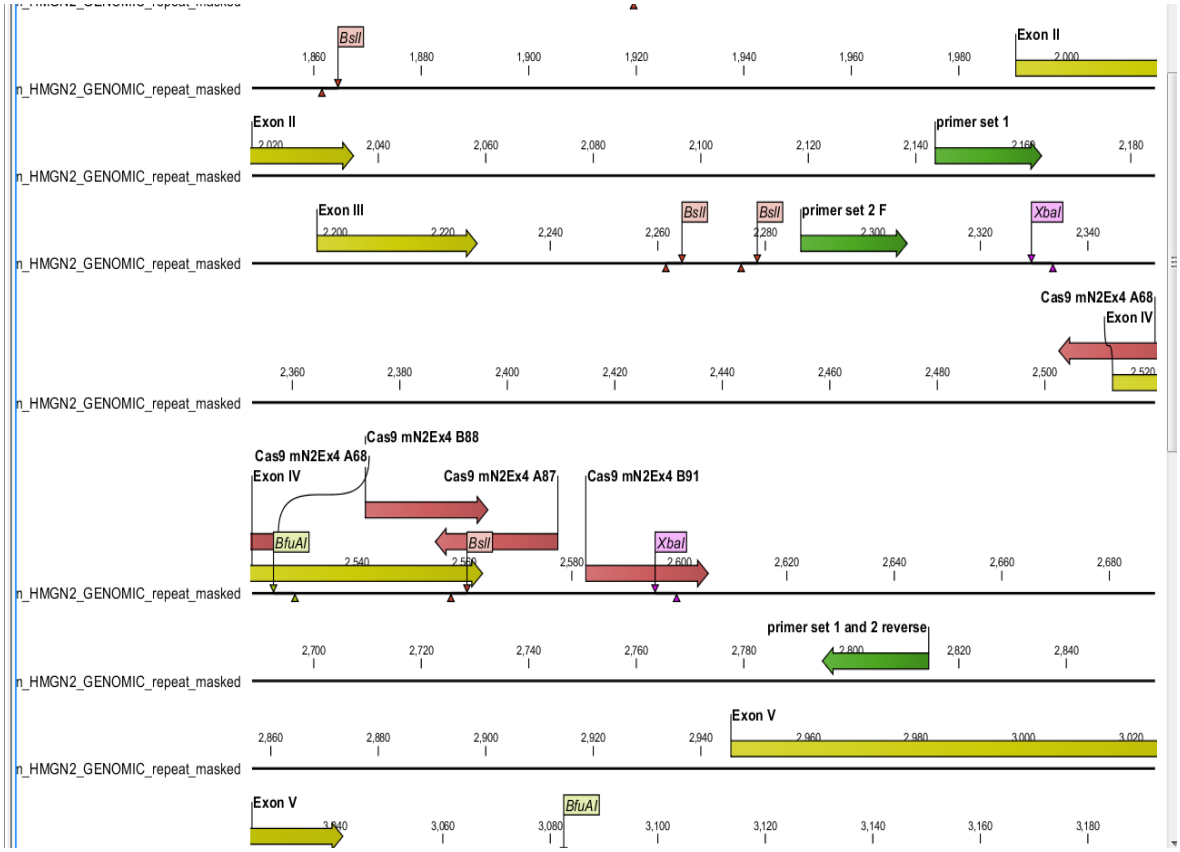


Figure 3.2: Screenshot of CLC file showing different gRNAs targeting exon IV of mouse *Hmgn2* gene and the two primer sets.
Two different nickases were selected for targeting exon IV of mouse *Hmgn2* gene. Each nickase contains two gRNAs (red arrows). The first nickase is composed of Cas9_mN2Ex4_A68 and Cas9_mN2Ex4_B88 while the another nickase includes Cas9_mN2Ex4_A87 and Cas9_mN2Ex4_B91. Each gRNA sequence contains an indicative restriction enzyme site. Two primer sets were also designed (set 1 and set 2, green arrows). Both sets have same reverse primer sequence but different forward primers. The yellow arrows refer to the exons of *Hmgn2* genes.

3.4 Creation of single vector containing multiguides for disrupting *NBD* of *Hmgn2* gene by paired gRNAs nickase

This involves two steps which are: i) construct individual gRNA vector by ligating the annealed oligonucleotides into expression cassette and ii) assemble different gRNAs in single vector by Golden Gate cloning.

3.4.1 Construct the expression cassette of each individual gRNA vectors

Each gRNA was created by annealing two complementary single stranded DNA oligos (MmHmgn2Ex4_A68, _B88, _A87 and _B91), then cloning into appropriate vectors with compatible RNA polymerase III promoters. Individual gRNA vectors

after insertion each spacer into its RNA polymerase III promoter are named as pGuide001-2_hH1/B88, pGuide02-2_hU6/B91, pGuide001-4_mU6/A87, pGuide002-4_hU6/B91, pGuide003-4_hH1/B88 and pGuide004-4_h7SK/A68 (Table 3.1). pGuide001-2_hH1/B88 and pGuide002-2_hU6/B91 are important for creating a 2-gRNA array, whereas vectors pGuide001-4_mU6/A87, pGuide002-4_hU6/B91, pGuide003-4_hH1/B88 and pGuide004-4_h7SK/A68 are necessary for either 2 or 4 guides multiplex.

Parental plasmid or backbone	Inserted Oligo (spacer)	Individual gRNA vector (pGuide vector)	Short name (either before or after insertion)
mU6_1-4	MmHmgn2Ex4_A87	pGuide001-4_mU6/A87	1-4
hU6_2-4	MmHmgn2Ex4_B91	pGuide002-4_hU6/B91	2-4
hH1_3-4	MmHmgn2Ex4_B88	pGuide003-4_hH1/B88	3-4
h7SK_4-4	MmHmgn2Ex4_A68	pGuide004-4_h7SK/A68	4-4
hU6_2-2	MmHmgn2Ex4_B91	pGuide002-2_hU6/ B91	2-2
hH1_1-2	MmHmgn2Ex4_B88	pGuide001-2_hH1/B88	1-2

Table 3.1: Individual gRNA vectors after insertion of each oligo into its parental plasmid.

In order to achieve this, all parental vectors (hH1_1-2, hU6_2-2, mU6_1-4, hU6_2-4, hH1_3-4 and h7SK_4-4) were first digested with the Type II (S) restriction enzyme, *Bbs*I. As shown in Figure 3.3, the six vectors were efficiently digested by *Bbs*I.

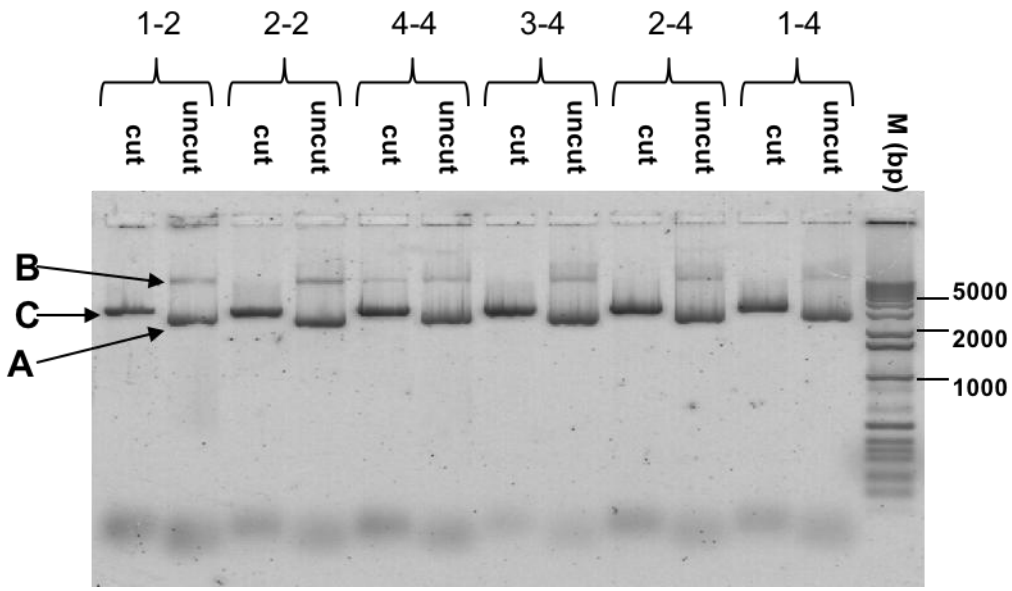


Figure 3.3: *Bbs*I digestion of each pGuide vectors for cloning.
Each vector (1-2, 2-2, 4-4, 3-4, 2-4 and 1-4) was digested for 2 hours with *Bbs*I on the thermal cycler. The approximate size for all vectors is 3,500 bp and when they are digested by *Bbs*I only 20 bp are removed. Band A is the supercoiled plasmid while B and C are non-specific band (probably nicked circle DNA) and linearized plasmid, respectively. 1-2, 2-2, 1-4, 2-4, 3-4 and 4-4 are abbreviations for the vectors (hH1_1-2, hU6_2-2, mU6_1-4, hU6_2-4, hH1_3-4 and h7SK_4-4, respectively). M is for marker. The marker is 1 Kb Plus DNA ladder from Life Technology. 1% agarose gel was used for electrophoresis.

Annealed oligos were then ligated into the digested backbone vector and transformed into competent cells. Colonies were screened by colony-PCR (Figure 3.4). All tested colonies showed the expected PCR products (Figure 3.4), although there is also a non–specific band, as the PCR reaction conditions were not fully optimised. One clone from each ligation was selected for further work, and the sequence of the gRNA in each vector was validated by commercial sequencing. Most of the constructs contained the correct sequence of the inserted gRNA and its corresponding promoter (Data not shown). However, the sequencing of pGuide001-4_mU6/A87 (1-4) vector revealed a single base pair deletion at an *EcoR*I that is upstream the mU6 promoter (data not shown). Because the deletion does not interfere with the expression cassette, this plasmid was taken forward into the next experiments.

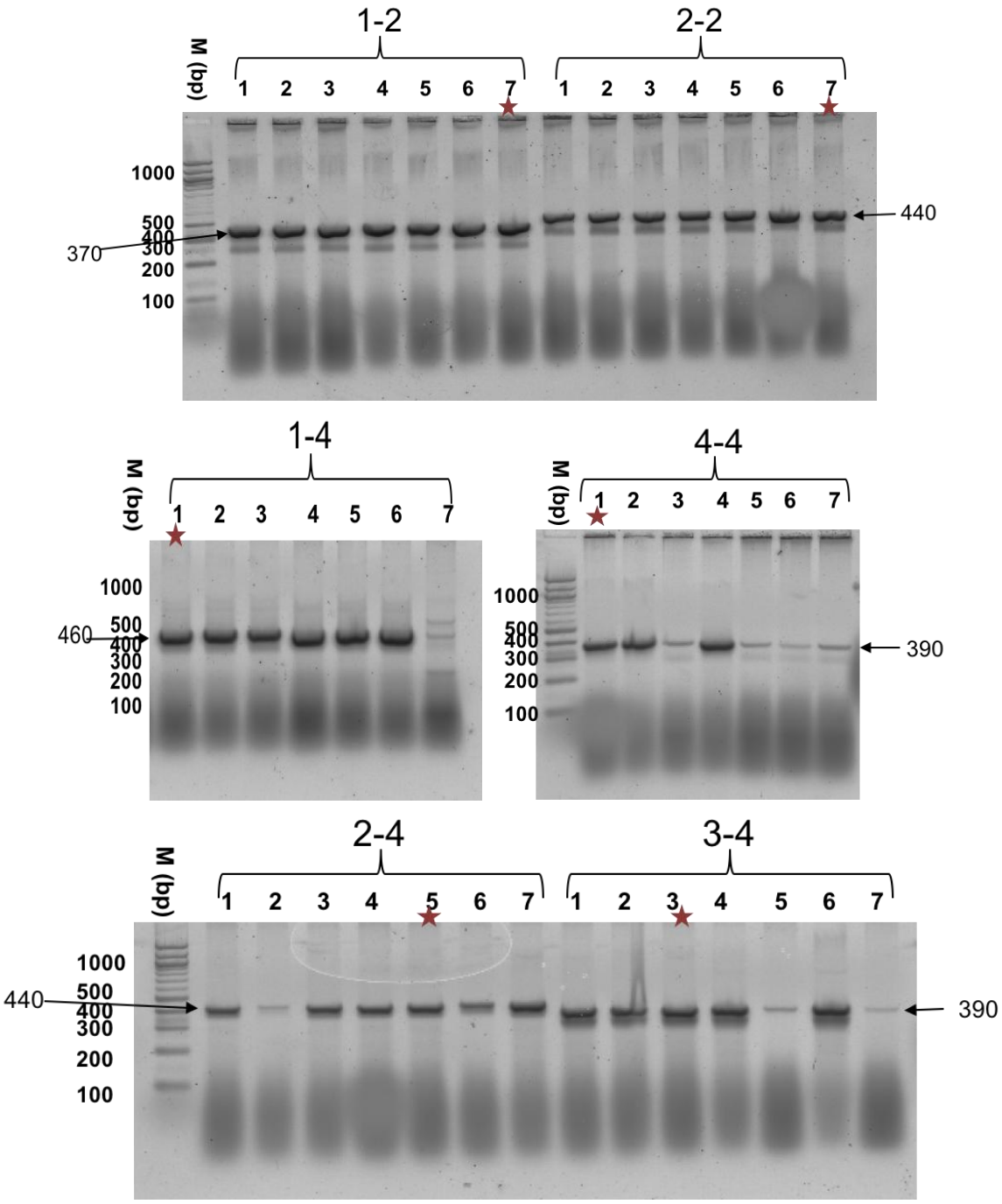


Figure 3.4: Colony PCR for individual gRNA vectors after inserting the annealed oligos. After cloning of each short annealed oligos into compatible RNA polymerase III promoter cassette, each ligation was transformed into DH5α E. coli (Life Technology). 7 colonies were selected per each transformation and these colonies of each gRNA vector were labelled from 1 to 7 (given on the top of each lane). Colony PCR was performed to screen the transformants using vector primer and the bottom strand oligo for the guide RNA. The expected band size of each colony PCR reaction is indicated by arrow for each gRNA vector. The red stars represent the plasmids (colonies) that were selected for further experiments. 1-2, 2-2, 1-4, 2-4, 3-4 and 4-4 are abbreviations for pGuide001-2_hH1/B88, pGuide02-2_hU6/B91, pGuide001-4_mU6/A87, pGuide002-4_hU6/B91, pGuide003-4_hH1/B88 and pGuide004-4_h7SK/A68, respectively. M is for the marker which is 100 bp DNA ladder (New England BioLabs). 1% agarose gel was used for electrophoresis.

3.4.2 Determining the efficiency of each gRNA in targeting exon IV of *Hmgn2* gene

3.4.2.1 Determine the transfection efficiency of the Cas9 wild type vector with individual gRNA vectors

To determine the ability of each gRNA to direct the cleavage by Cas9 and generate a mutation at exon IV of the *Hmgn2* gene in P19 cells, four vectors containing the four different gRNAs were tested: pGuide001-4_mU6/A87, pGuide002-4_hU6/B91, pGuide003-4_hH1/B88 and pGuide004-4_h7SK/A68. pGuide001-2_hH1/B88 and pGuide002-2_hU6/B91 were not examined because they have the same gRNAs as pGuide003-4_hH1/B88 and pGuide002-4_hU6/B91, respectively. The Neon transfection conditions and procedures were previously optimised (see the method section, chapter 2). The four tested vectors were individually transfected into P19 cells alone or along with wild type Cas9 linked to a reporter gene encoding green fluorescent protein (GFP). The transfection efficiencies ranged from 35 to 65% (Figure 3.5 and Table 3.2). The graphs in Figure 3.5 are examples for different transfection efficiencies. However, these transfection efficiencies only reflect the effectiveness of delivering the Cas9, not the gRNA vectors. From the scatter plot, there is no obvious difference in the cell morphology, distribution, or viability between the control (wild type and Cas9) and other vectors (Figure 3.5).

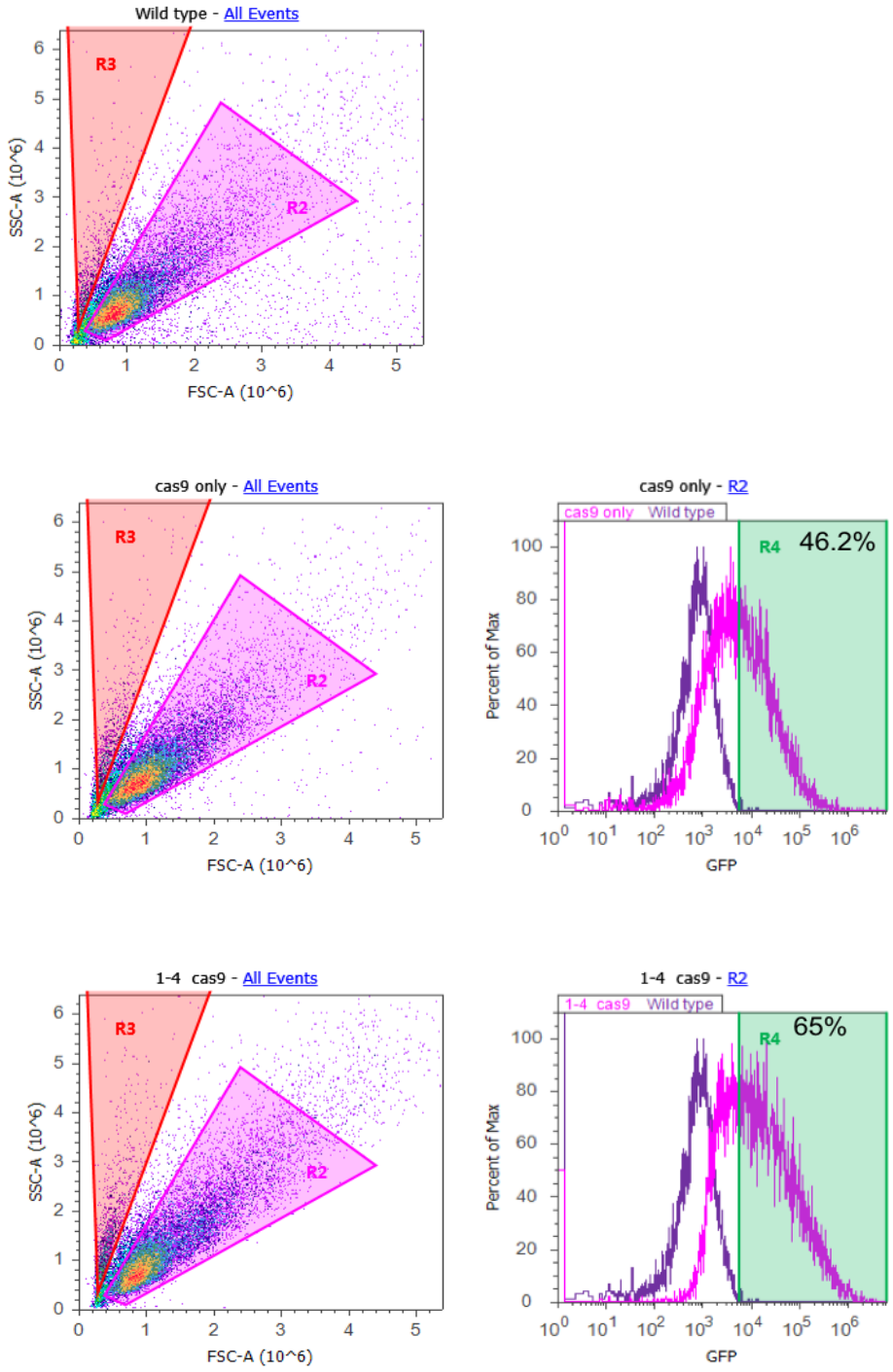


Figure 3.5: The transfection efficiencies for individual gRNA vectors with wild type Cas9 in P19 cells two days after Neon transfection.
The left and right panels are dot blots and histograms, respectively. The dot blot in the top row is for the un-transfected P19 wild type cells. The P19 cell were transfected with Cas9 only (control, middle row) or co-transfected with Cas9 and pGuide001-4_mU6/A87 or pGuide002-4_hU6/B91 or pGuide003-4_hH1/B88 or pGuide004-4_h7SK/A68. However, the co-transfection with Cas9 and pGuide001-4_mU6/A87 is only given here as an example (bottom row). The highest effective transfection was obtained when cell co-transfected with two plasmids in which 65% of cells were green positive (gate R4). The green negative population was not gated and overlays on the wild type peak (dark purple). The transfection efficiencies were measured 2 days post the transfection (given inside histograms). The dot blot showed the events read by cytometer machine (Attune Acoustic Focusing Cytometer) and the healthy population was gated in R2. The dead cells are gated in R3 as the cells in this gate were red when treated with propidium iodide in previous experiment (PI, data not shown). The only cells in R2 gate were blotted in the histograms. Transfections were carried out using 15 μg of the Cas9 vector and 10 μg of the guide vector.

Transfection	GFP(%)
Cas9 only	46.2
pGuide 001-4 and Cas9	65
pGuide 002-4 and Cas9	35
pGuide 003-4 and Cas9	48
pGuide 004-4 and Cas9	61

Table 3.2: All the transfection efficiencies for gRNAs with Cas9 using Neon system.

3.4.2.2 Determine the efficiency of each individual gRNA to introduce a mutation at exon IV by RFLP

In order to optimise the restriction fragment length polymorphism (RFLP) assay, DNA from wild type P19 was amplified using primer sets 1 and 2 (Figure 3.2) to ensure the effectiveness of primers and restriction enzymes used in. Both primer sets were efficient, giving products of 671 bp for set 1 and 528 bp for set 2 (Figure 3.6, lane 2 and 8, respectively). These products were also tested for digestion by *Xba*I, *Bsl*I or *BfuA*I. *Bsl*I should report mutations introduced by the B91 guide, and *Xba*I should report mutations introduced by guides B88 and A87. Both enzymes efficiently cut the PCR products from primer set 1 or 2 (Figure 3.6). However, *BfuA*I, which should report mutations introduced by pGuide004-4_h7SK/A68, was unable to cut the PCR products from set 1 and 2. Closer inspection of the properties of this enzymes revealed that it needs two recognition sites per DNA fragment in order to cut. As PCR products 1 and 2 only contain one *BfuA*I *site,* the enzyme is unable to cut. No other restriction enzyme sites were suitable for RFLP analysis of mutations introduced by the A68 guide.

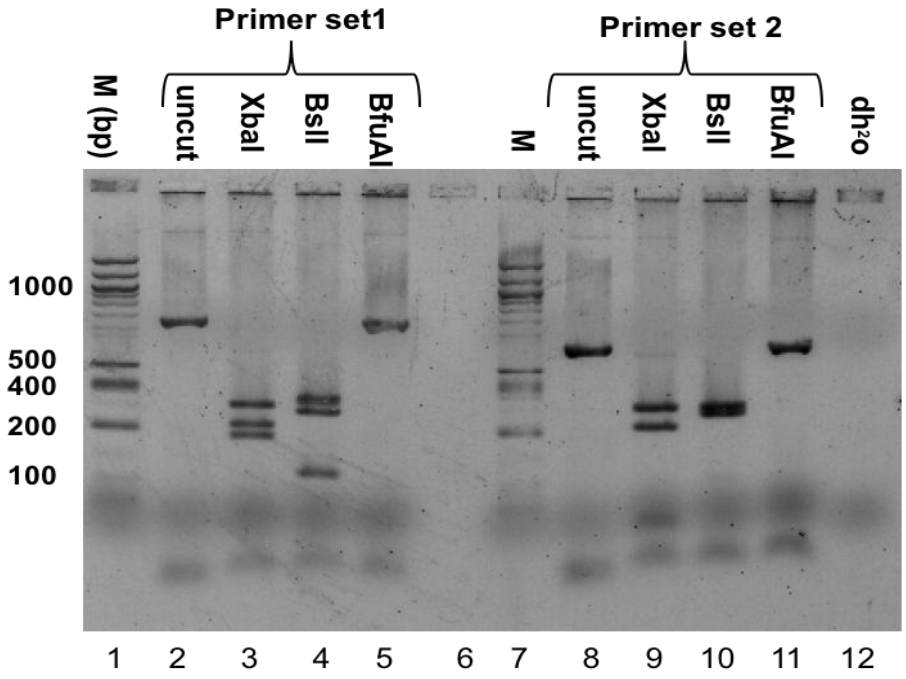
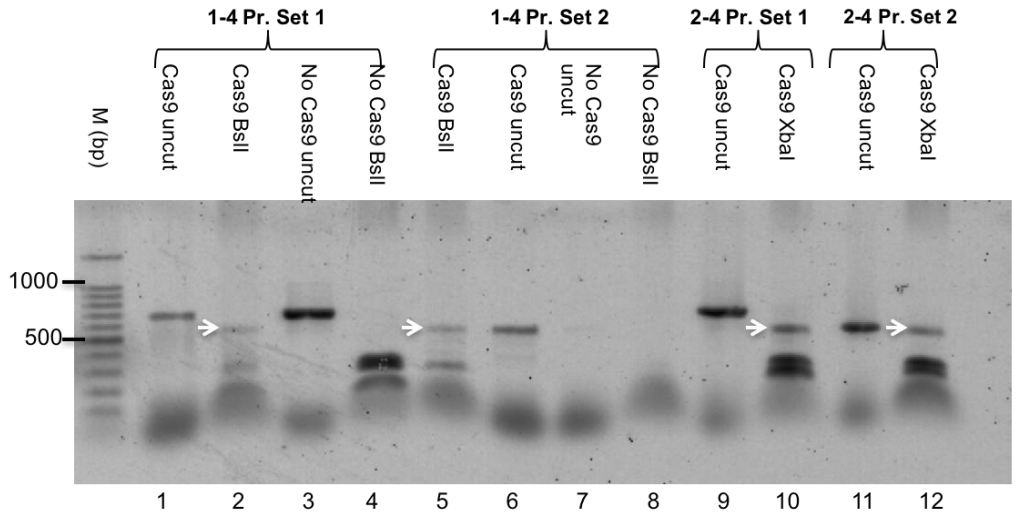
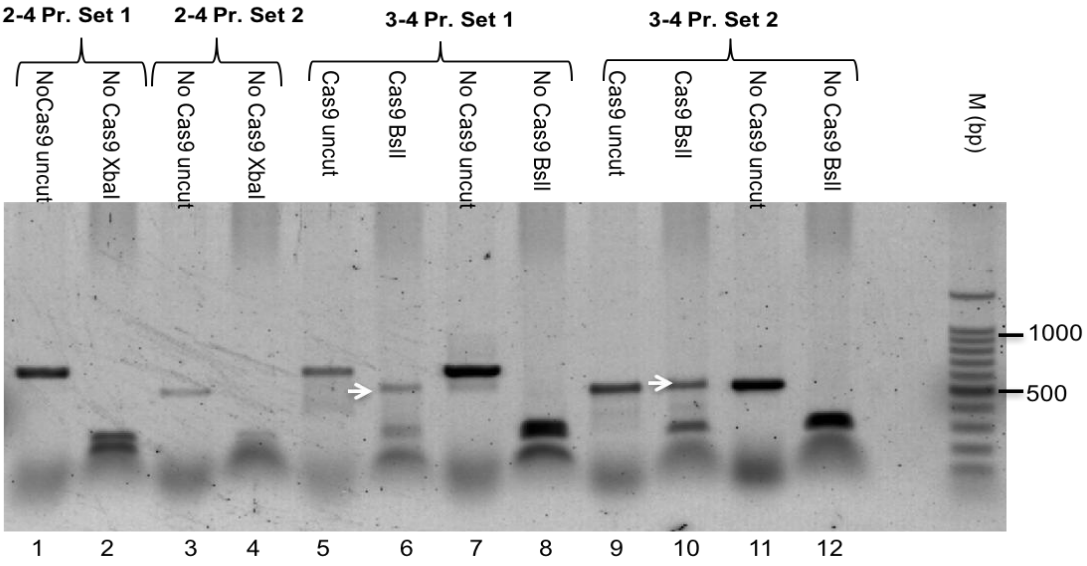


Figure 3.6: Determine the effectiveness of RFLP-assay using wild type P19 DNA.
The P19 wild type DNA was extracted and the region around exon IV of *Hmgn2* gene was amplified by two different primer sets (set 1 and set 2). The optimum annealing temperature for set 1 and 2 that was previously determined using the gradient PCR (data not shown), is 58.1^{o}C. The undigested products obtained from primer set 1 and 2 are expected to be 671 and 528 bp, lane 2 and 8, respectively. These products were subjected to digestion by three restriction enzymes (*Xba*I, *Bsl*I and *BfuA*I) that interfere with cleavage site of Cas9 for 1 hour. *Xba*I enzyme cut the PCR product from primer set 1 into 190, 215 and 266 bp, whereas the product of primer set 2 into 47, 215 and 266 bp. *Bsl*I digests the PCR products of primer set 1 and 2, and yields products of (18, 117, 255 and 281 bp) and (255 and 273 bp), respectively. The products from *BfuA*I digestion are 381 and 290 bp for primer set 1, and 238 and 290 bp for primer set 2. Negative control using dh_2o with dNTPs and primer set 1 only was used in PCR reaction (lane 12). The digestion temperatures for *Xba*I, *Bsl*I and *BfuA*I enzymes are 37, 55 and 50 ^{o}C, respectively. M refers to marker which is 100 bp DNA ladder from New England BioLabs. 1% agarose gel was used for electrophoresis.

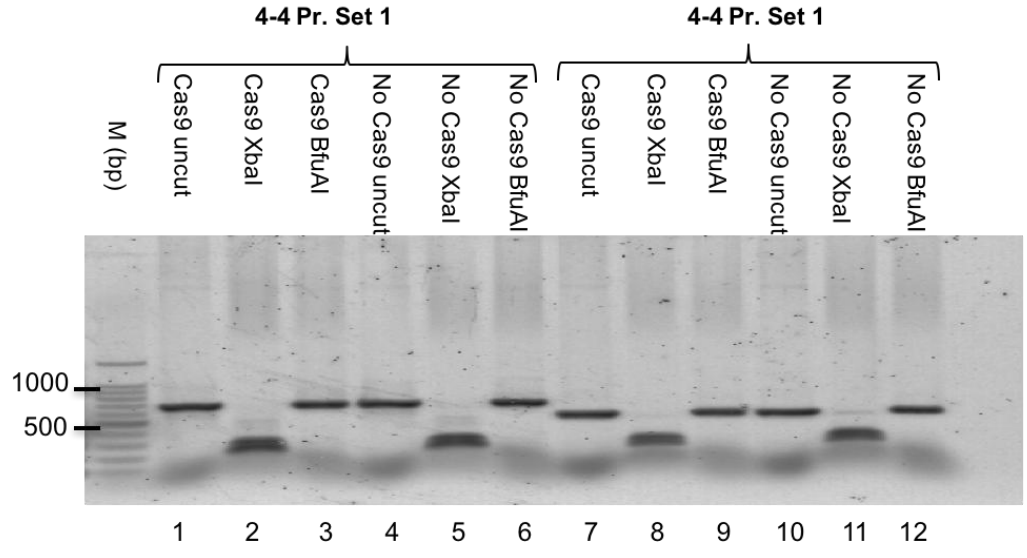
The ability of each gRNA to generate mutations at the *Hmgn2* locus was tested by performing RFLP assays on DNA extracted from the earlier transfections. Firstly, in Figure 3.7a, lanes 1-8 contain DNA from transfections carried out with plasmid pGuide001-4_mU6/A87 (abbreviated as 1-4), either in the presence or absence of Cas9. In lanes 1-4, PCR primer set 1 was used, whereas in lanes 5-8, PCR primer set 2 was used. The uncut PCR product of the control transfection (only gRNA) is 671 bp (lane 3), and the *Bsl*I digestion of this product showed the expected pattern of bands (18, 117, 255 and 281 bp, lane 4). In the presence of Cas9 and pGuide001-4_mU6/A87, the *Bsl*I digestion product revealed clearly an extra band (lane 2, white arrow) that is expected to be around 536 bp. This indicates a proportion of cells has a deletion that disrupts the recognition site of *Bsl*I. A similar result of *Bsl*I digestion was obtained with primer set 2 and can be seen in Figure 3.7a, lane 5 (white arrow).


1-4 Pr. Set 1
1-4 Pr. Set 2
2-4 Pr. Set 1
2-4 Pr. Set 2
M (bp)
Cas9 uncut
Cas9 BsII
No Cas9 uncut
No Cas9 BsII
Cas9 BsII
Cas9 uncut
No Cas9 uncut
No Cas9 BsII
Cas9 uncut
Cas9 XbaI
Cas9 uncut
Cas9 XbaI
1000
500
1 2 3 4 5 6 7 8 9 10 11 12

A-


2-4 Pr. Set 1
2-4 Pr. Set 2
3-4 Pr. Set 1
3-4 Pr. Set 2
NoCas9 uncut
No Cas9 XbaI
No Cas9 uncut
No Cas9 XbaI
Cas9 uncut
Cas9 BsII
No Cas9 uncut
No Cas9 BsII
Cas9 uncut
Cas9 BsII
No Cas9 uncut
No Cas9 BsII
M (bp)
1000
500
1 2 3 4 5 6 7 8 9 10 11 12

B-


4-4 Pr. Set 1
4-4 Pr. Set 1
M (bp)
Cas9 uncut
Cas9 XbaI
Cas9 BfuAI
No Cas9 uncut
No Cas9 XbaI
No Cas9 BfuAI
Cas9 uncut
Cas9 XbaI
Cas9 BfuAI
No Cas9 uncut
No Cas9 XbaI
No Cas9 BfuAI
1000
500
1 2 3 4 5 6 7 8 9 10 11 12

C-

D-

Allele	Undigested PCR product (bp)	*Bsl*I		*Xba*I		*Bfu*AI	
		WT	Mutant	WT	Mutant	WT	Mutant
Primer Set1	671	18, 117, 255 & 281	18, 117& ~536	190, 215 & 266	190 & ~481	290 & 381	~671
Primer Set 2	528	255 & 273	~528	47, 215 & 266	47 & ~481	238 & 290	~528

Figure 3.7: RFLP-assay for PCR products for different transfections of individual vectors. The DNA extracts of previous transfection in 3.4.2.1 were amplified using either primer set 1 and 2. The PCR product obtained from primer set 1 and 2 are 671 and 528 bp, respectively. Four individual gRNA vectors were transfected into P19 cells either alone or with wild type Cas9 vector using Neon transfection. These vectors include pGuide001-4_mU6/A87, pGuide002-4_hU6/B91, pGuide003-4_hH1/B88 and pGuide004-4_h7SK/A68 (1-4, 2-4, 3-4 and 4-4, respectively). The transfection of individual gRNA alone or with Cas9 were digested with the restriction enzyme that is specific for that gRNA. *BslI* was used for 1-4 and 3-4 while *XbaI* was used to digest the products of transfection with 2-4. Digestions with either *BfuAI* or *XbaI* (the latest was used because it cuts nearby the cleavage) were applied for transfection with 4-4. **A** contains all samples of 1-4 either amplified by primer (Pr) set 1 (lanes 1 to 4) or set 2 (lanes 5-8) and samples of transfection 2-4 with Cas9 (lanes 9 and 10 for primer set 1 while lanes 11 and 12 for primer set 2). **B** includes the remaining samples of 2-4 alone transfection (lanes 1 and 2 for primer set 1 whereas lanes 3 and 4 for primer set 2) and all samples of 3-4 transfection (primer set 1 in 5-8 lanes and primer set 2 in 9-12 lanes). Gel **C** have all samples of 4-4 transfection (lanes 1-6 for primer set 1 and 7-12 for set 2). In each primer set, there are four samples from two transfections {two samples from transfection of gRNA alone (named *no Cas9 uncut* and *no Cas9 enzyme* used for digestion) and two samples from transfection of gRNA with Cas9 (named *Cas9 uncut* and *Cas9 enzyme* used for digestion)}. Table in **D** shows the different PCR products from primer set 1 and 2 and the expected sizes after digestion of wild type or mutant alleles by different restriction enzymes. The new bands with variable sizes (given in table D) were gained as a result of disrupting the recognition site of restriction enzymes when only DNA was co-transfection with gRNA vectors and Cas9 together (white headed arrows). (Cas9 uncut) or (no Cas9 uncut) refer to undigested PCR products by relative enzyme. The digested PCR products from cells transfected by gRNA alone or with Cas9 are indicated as (no Cas9 or Cas9 followed by name of enzyme used, respectively). M refers to marker which is 100 bp DNA ladder from New England BioLabs. 1% agarose gel was used for electrophoresis.

In the same figure (Figure 3.7a), lanes 9-12 are for DNA from cells transfected by pGuide002-4hU6/B91 (abbreviated as 2-4) and Cas9. When DNA was either amplified by primer set 1 (lane 9) or set 2 (lanes 11), products of 671 and 528 bp were respectively obtained. *Xba*I digestion of the set 1 product generates bands of 190, 215 and 266 bp (Figure 3.7b, lane 2), whereas mutation by the gRNA should generate a new band of around 481 bp. This can be seen in Figure 3.7a, lane 10 (white arrow). *Xba*I digestion of the set 2 product yields bands of 47, 215 and 266 (Figure 3.7b, lane 4), and gRNA-directed mutation should result in a new band of around 481 bp. This can be observed Figure 3.7a, lane 12 (white arrow). This means that Cas9 was efficiently targeted by gRNA expressed from the pGuide002-4_hU6/B91 vector, and mutations were introduced that disrupt *Xba*I digestion.

Mutagenesis was also observed with the pGuide003-4_hH1/B88 vector (abbreviated as 3-4 in Figure 3.7b). Transfection of pGuide003-4_hH1/B88 along with Cas9 disrupts the recognition site of *Bsl*I in a proportion of cells: In Figure 3.7b, compare lanes 6 and 8 for primer set 1, and compare lanes 10 and 13 for primer set 2. Collectively, these data indicate that guides in pGuide001-4, 2-4 and 3-4 are directing cleavage by Cas9 and resulting in mutations of the *Hmgn2* gene at the desired location.

In summary, RFLP assays showed that three out four gRNAs were able to mediate indel mutations in the endogenous *Hmgn2* target gene. The efficiency of mutagenesis varied between the guides, with estimated efficiencies of 40%, 20% and 50% for pGuide001-4_mU6/A87, pGuide002-4hU6/B91 and pGuide003-4_hH1/B88, respectively. This correlates fairly well with the transfection efficiencies, which were 65%, 35% and 48%, respectively. This suggests that optimising transfection efficiency is vital for efficient mutagenesis.

As explained above, we were unable to perform the *Bfu*AI RFLP assay for mutagenesis directed by pGuide004-4_h7SK/A68, because *Bfu*AI requires two copies of its recognition sequence for cleavage to occur. To overcome this, it might be possible to design another primer set that amplifies a product carrying two *Bfu*AI recognition sites. Alternatively, the surveyor assay could be employed, which uses a mismatch-specific nuclease (surveyor nuclease) to recognise and cleave mismatches due to the presence of single nucleotide

polymorphisms (SNPs) or small insertions or deletions. However, as this guide acts in a nickase pair with pGuide003-4_hH1/B88, it was decided to proceed with investigating the efficiency of the nickase, rather than spend time optimising other assays for the pGuide004-4_h7SK/A68 guide on its own.

3.4.3 Generating of a single vector (multiguides) with multiple gRNAs

Our aim is to use paired nickase CRISPR-Cas9D10A to mutate the *Hmgn2* gene. To achieve this, an ideal approach is to utilise a single vector expressing an array of different gRNAs and the mutant Cas9D10A. As mentioned previously, Kabadi et al (2014) developed a Golden Gate cloning method to facilitate rapid and efficient cloning of multiple gRNA expression cassettes into a single vector expressing the desired Cas9 protein. Another advantage of this vector is that GFP is coupled to Cas9D10A via a self cleaving T2A linker peptide (Duda et al., 2014; Kabadi et al., 2014; Chiang et al., 2016). This means that transfected cells express GFP, so can be tracked by FACS. This section details the assembly of gRNA expression cassettes into an all-in-one vector that expresses Cas9D10A.

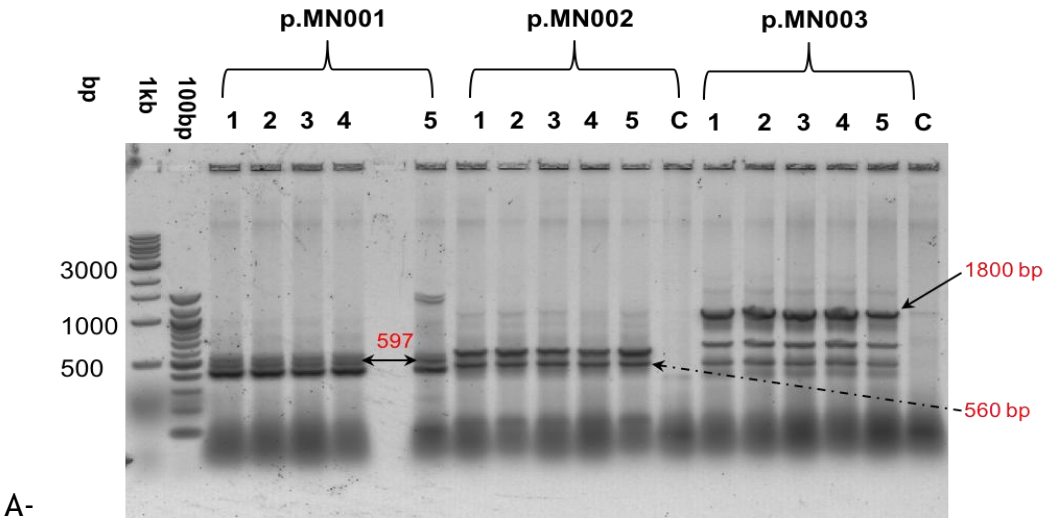
Three multiguide vectors were generated and named as pMultiNick_001-4, pMultiNick_002-4 and pMultiNick_003-4 (Table 3.3). The last vector expresses the gRNAs for the both nickases. The use of more than one nickase that targets two distal regions has mainly been applied to the generation of large deletions. For example, Sakuama et al (2014) used 3 nickases across the *APC* gene locus to obtain a large deletion of ~2.5 kb. The distance between the four gRNAs in this HMGN2 strategy is much smaller, and the rationale is that using these two nickases together may increase the chance of gene disruption by deleting the intervening small fragment. Indeed, it was shown that making adjacent DSBs resulted in high-efficiency deletion of the intervening small fragment (Mali et al., 2013a).

Multiguide vector	pGuide vectors
pMltiCas9D10A	No guide/ empty
pMultiNick_001	pGuide001-4_mU6/A87
	pGuide002-2_hU6/B91
pMultiNick_002	pGuide001-2_hH1/B88
	pGuide002-4_h7SK/A68
pMultiNick_003	pGuide001-4_mU6/A87
	pGuide002-4_hU6/B91
	pGuide003-4_hH1/B88
	pGuide004-4_h7SK/A68

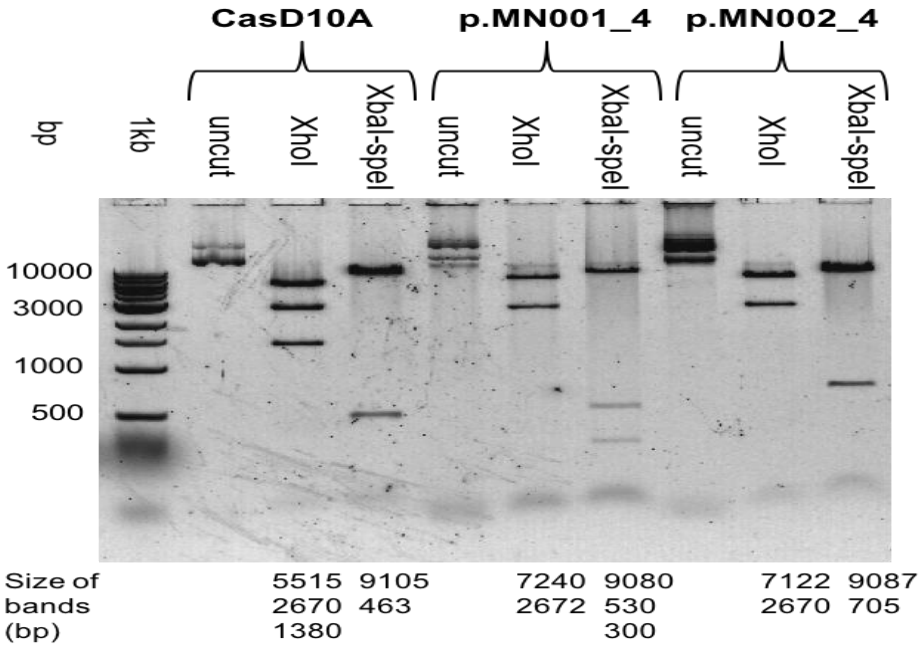
Table 3.3: The different multi-guides vectors assembled from pGuide vectors by Golden Gate.

The assembly of these different gRNAs from different pGuide vectors was performed by golden gate as described in chapter 2. After the transformation of each newly formed multiguide vector, colonies were screened by blue-white selection and colony-PCR. As demonstrated in Figure 3.8a, the majority of colonies for each plasmid generated the expected bands. Some additional bands were also observed because the optimal annealing temperatures for the PCR reactions had not been determined.

To confirm the identities of these plasmids, purified DNA was digested with *Xho*I alone or *Xba*I and *Spe*I double digestions (Figure 3.8b). The digestion of pMultiCas9D10A, pMultiNick_001-4 and pMultiNick_002-4 gave the expected products, confirming that these plasmids have been formed correctly. However, pMultiNick_003-4 showed some extra unpredictable bands when it was digested by *Xba*I and *Spe*I. Other colonies from the same transfection were then screened by restriction digest, and one (pMultiNick_003-5) was shown to have the correct digestion pattern (not shown). This was taken forward for further experiments. All three multiguide vectors were validated by sequencing to ensure the correct arrangements and sequences.



A-



B-

Figure 3.8: Colony PCR for multiguide vectors after Golden Gate assembly, and the multiguide plasmids digestion.
The promoter-gRNA cassettes were assembled as tandem inverted repeat units into Cas9D10A expression vectors. Three multiguide vectors were formed by the Golden Gate assembly cloning, including pMultiNick_001, _002 and _003. Each ligation was transformed into SURE chemically competent cells (Agilent). **A.** Because colonies were screened by lacZ-based blue/white screening, 5 white colonies were selected per each transformation. Colony PCR for different nickase vectors was performed to screen the assembly of gRNA cassettes by using universal primer complementary to the backbone sequence and one strand of the inserted oligo. The majority of colonies for each plasmid were given the expected bands (expected sizes were annotated next to the bands) with non-specific bands. Colonies from each pMultiguide vector were labelled from 1 to 5 and control (C, blue colony). **B.** Each selected pMultiguide (3.5-5μg) was digested either by *Xho*I or *Xba*I-*Spe*I to confirm the identities of these plasmids. One recognition site of *XhoI* is lost where the insertion occurs. Double digestion with *Xba*I-*Spe*I cut on the both side proximate the insertion. The expected bands sizes are indicated under each digestion. The marker is 1 Kb Plus DNA ladder from Life Technology. 1% agarose gel was used for electrophoresis.

3.4.3.1 Identifying the transfection efficiency of newly assembled multiguide vectors

P19 cells were transfected using Neon system with three different multiguide vectors (pMultiNick_001-4, pMultiNick_002-4 and pMultiNick_003-5). As shown in Figure 3.9, it appears that transfection with any of the multiguide vectors, even the empty vector (pMultiCas9D10A), resulted in a substantial reduction of cell numbers after two days of transfection. This was not observed when cells were exposed to the electric currents in the absence of any plasmid DNA.

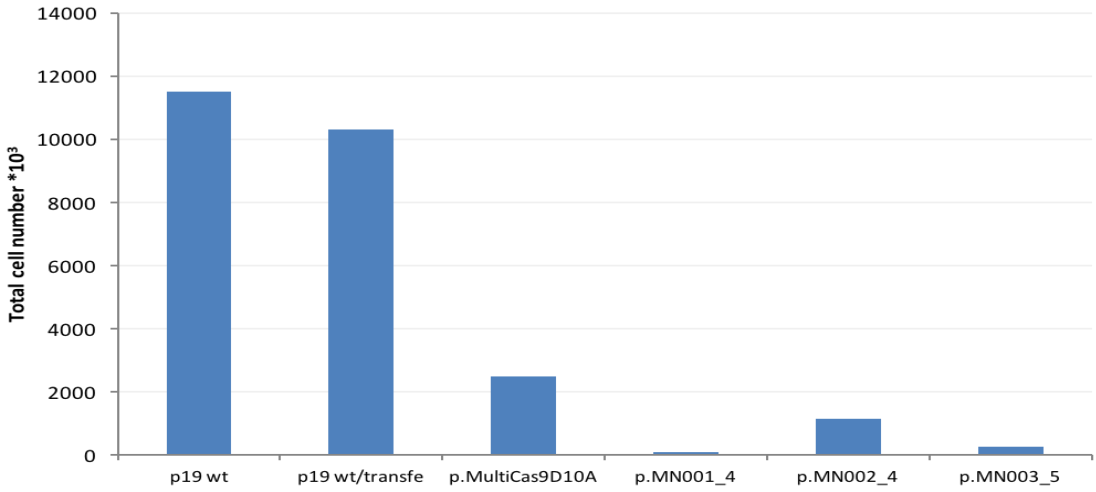
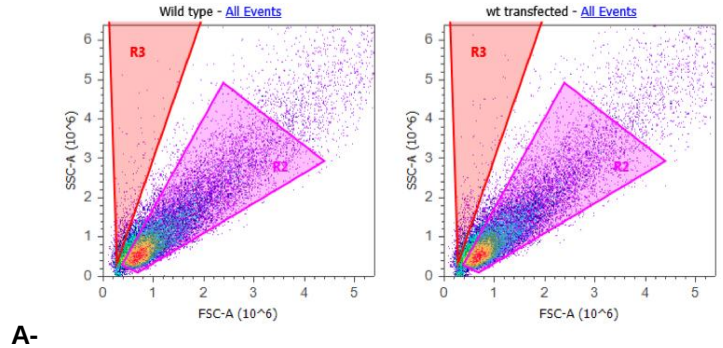
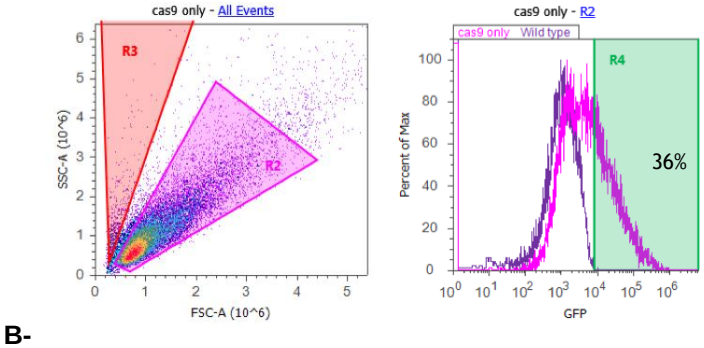


Figure 3.9: The cell counts after 2 days of transfecting different pMultiguide plasmids into P19 cell.
The total live cells were counted by haemocytometer after two days of transfection with different pMultiguide vectors using Neon machine. The cell number after two days were almost similar for electroporated or non-electroporated wild type. However, the cell viability was substantially reduced after 48 hours when the cells are transfected with any of pMultiguide vectors, even when the cells transfected with control empty plasmid that does not have gRNA (pMultiCas9D10A). p.MN stands for pMultiNick.

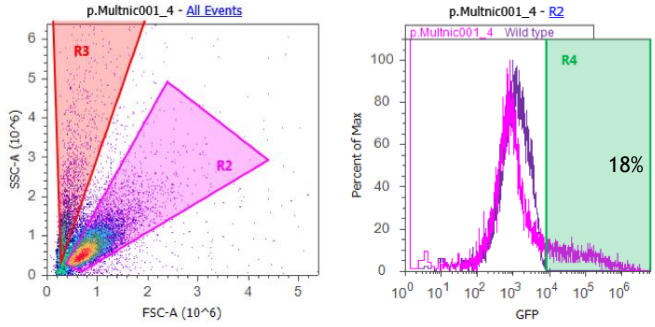
FACS analysis was used to assay for cell viability and transfection efficiency (Figure 3.10 and Table 3.4). In this assay, dead cells are removed by washing prior to the FACS analysis, so the reduced viability seen in Figure 8 is not apparent. The percentage of cells expressing GFP is 18%-30%, which is lower than the 35%-65% seen with individual guide plasmids (compare Table 3.4 with Table 3.2).



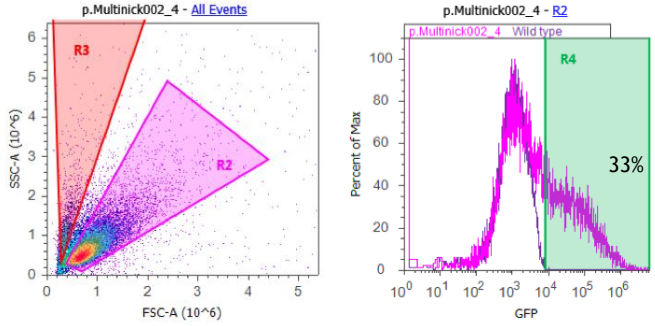
A-



B-



C-



D-

Figure 3.10: Histograms and dot blots showing the efficiency of transfecting pMultiguide vectors in P19 cells after 2 day.

A. The dot blots for wild type P19 cell that are either electroporated or not. The electroporation does not affect the property of healthy cells (R2). The P19 cell were transfected with vector including pMultiCas9D10A (**B**), pMultinick001_4 (**C**) and pMultinick002_4 (**D**). Transfections were carried out using 15 μg of each pMultiguide vectors. The dot blot showed the events read by cytometer machine (Attune Acoustic Focusing Cytometer) and the healthy population was gated in gate R2. The dead cells are gated in R3 as the cells in this gate were red when treated with propidium iodide in previous experiment (PI, data not shown). The only cells in R2 gate were blotted in the histograms. The peak on the histograms were shifted slightly that indicates there are two populations by BL-1 laser that detects the green fluorescence. These populations are either green negative or positive. The green negative population was not gated and overlays on the wild type peak (dark purple). The green positive cells were gated in gate R4 on the histograms and the percentages of GFP-positive cells are given on histograms of each transfection.

Transfection	GFP (%)	Healthy cells (R2, %)
Un-transfected wild-type	0.07	77
Electroporated wild type	0.03	77
pMultiCas9D10A	36	88
pMultiNick 001_4	18	67
pMultiNick 002_4	33	74
pMultiNick 003_5	30	77

Table 3.4: The GFP and the healthy cells percentages as calculated by FACS for all transfections with pMultiguide plasmids.

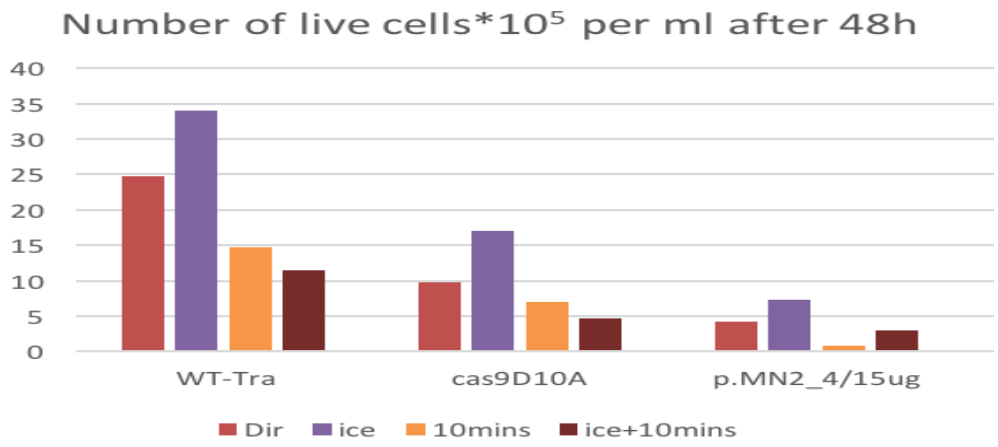
3.4.3.2 Optimising the transfection for the most powerful multiguide vector

Factors that might contribute to poor cell viability and low transfection efficiency are plasmid size, plasmid quality and poorly optimised transfection conditions. The multiguide plasmids are larger than the individual guide plasmids used earlier (9-10 kb vs. 3.5 kb), and it is known that large plasmids tend to have lower transfection efficiencies (Ohse et al., 1995).

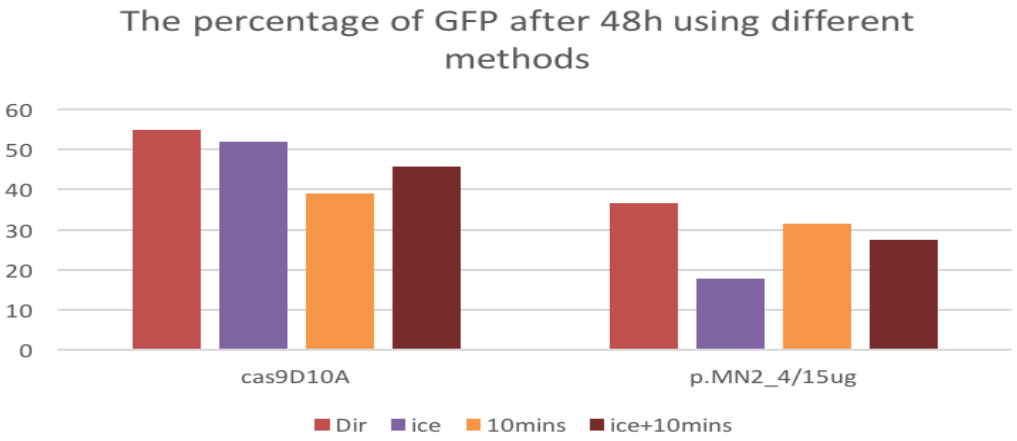
In an attempt to optimise transfection efficiencies, three alterations to the protocol were tested. These were i) increasing the plasmid concentration ii) pre-chilling the mixture (cells, DNA and buffer) on ice for 5 minutes before electroporation and iii) resting the mixture for 10 minutes at room temperature after the electroporation but before transferring the cells to the culture medium (Andreason and Evans, 1989; Potter, 2003). The last two are thought to reduce the stress caused by electroporation, thus possibly enhancing cell viability and transfection efficiency.

Pre-chilling the cells on ice did seem to improve cell viability (Figure 3.11a). However, it also reduced the percentage of transfected cells (Figure 3.11b), so no overall benefit was demonstrated. Post-electroporation incubation, or pre-chilling plus post-incubation, did not increase cell survival or transfection efficiency (Figure 3.11). Furthermore, increasing the amount of plasmid DNA from 15 µg to 25 µg did not increase transfection efficiency (data not shown).

It is notable that the transfection efficiency for the empty vector pMultiCas9D10A was 55% in the experiment shown in Figure 3.11b, compared to 36% in Figure 3.10. This day-to-day variation implies that other variables such as sample handling or cell health are important factors contributing to transfection efficiency.



A-



B-

Figure 3.11: The cell counts (A) and GFP percentages (B) after two days of transfecting P19 cells with two multiguide plasmids using different conditions.
The Neon transfection was carried out as previous in terms of the parameters. However, two different concentrations (15 and 25 μg) of p.MN2_4 (pMultinik002_4) were applied. In addition to the normal technique of Neon transfection in which the cells are transfected directly after adding the transfection buffer (Dir), three conditions were also used with either 15 or 25 μg of DNA (data for 25 μg transfections not shown). These are: leavening the mixture (cells, DNA and buffer) on ice for 5 minutes before the transfection (referred as ice), leaving the mixture at room temperature for 10 minutes after the electroporation (named as 10mins in the chart), and a combined between the latest two conditions (called ice+10mins).

3.4.3.3 Determine the proficiency of the paired nickase (multiguide vector) to introduce the DSB at the desired locus of *Hmgn2* gene

DNA was extracted from the transfections described in 3.4.3.1 in order to check the ability of each multiguide plasmid to generate mutations in exon IV of the *Hmgn2* gene. Because the majority of cells transfected by either pMultiNick_001-4 or pMultiNick_003-4 were dead, the RFLP assay could not be carried out. However, in the cells transfected with pMultiNick_002-4, the RFLP assays showed around 15-20% of alleles were resistant *Bsl*I digestion (Figure 3.12). The *Bfu*AI RFLP assay is not informative, as explained earlier, so the full extent of mutation resulting from this nickase treatment cannot be assayed here.

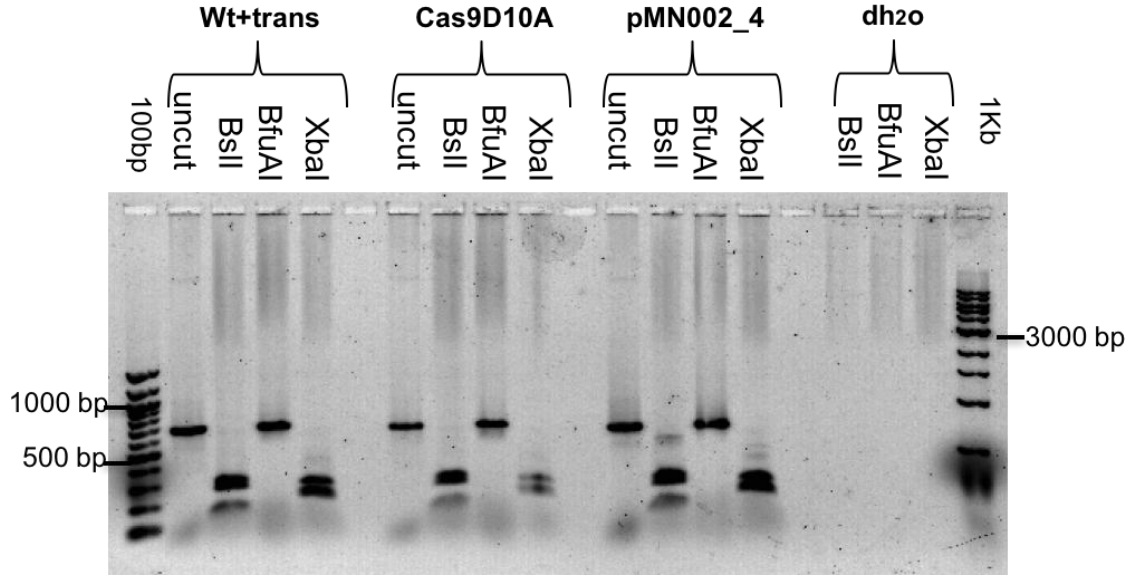


Figure 3.12: The RFLP-assay for polyclonal populations transfected by different pMultiguide plasmids.
The DNA samples were extracted from electroporated wild type (Wt-trans), transfections of pMultiCas9D10A (Cas9D10A) and pMultiNick_002_4 (pMN002_4). The transfections were described in 3.4.3.1. The samples were amplified by primer set 1 and then subjected for RFLP assay. For wild-type cells, digestion of PCR products amplified by primer set 1 (671 bp) are as following: (18,117, 255 and 281 bp) for *Bsl*I, (381 and 290 bp) and (190, 215 and 266 bp) for *Bfu*AI and *Xba*I, respectively. The *Xba*I digestion was used because it cuts nearby the cleavage. Proportion of the DNA extracted from the transfected cells with pMN002_4 was resistant to the cut by *Bsl*I as shown by presence of extra band. This extra band is around 600 bp and is smaller than the size of the uncut product. The markers are 100 bp DNA ladder (New England BioLabs) and 1 Kb Plus DNA ladder (Life Technology). 1% agarose gel was used.

3.5 Generating and screening monoclonal lines with disrupted NBD of *Hmgn2* gene

The pMultiNick_002-4 vector was used again in two independent transfections so that clonal lines derived from single cells (monoclonal lines) could be generated. The transfection efficiencies were around 35%, which is similar to those in the previous transfections (data not shown). In one transfection, the GFP positive cells were enriched using FACS sorting after 48 hours by FACS (Byrne et al., 2014). In total, 160 lines were recovered and screened for the partial or complete loss of HMGN2 protein by western blotting: 80 lines from the unsorted population and 80 lines from the sorted population.

The antibody for HMGN2 recognises an epitope at the C-terminus of the protein, so any insertions or deletions in exon IV that lead to a frameshift can be distinguished from the wild-type form of the protein. Some of the western blots are shown in Figure 3.13 and some clones were shown to be interesting for HMGN2 disruption. Of the lines shown here, S27 is the most promising candidate for a homologous parietal or complete HMGN2 knockout, as there appears to be no band for the HMGN2 protein. Lines C15, S45, 3, 13 and 16 have two forms of HMGN2, with the smaller forms presumably generated by loss of one or more exons. In lines 13 and 16, the upper band appears to be a different size to wild type HMGN2, which may indicate a deletion or insertion in the second HMGN2 allele. Clones 19 and 59 (data not shown for 59) do not appear to have any full length HMGN2 protein but only a truncated form.

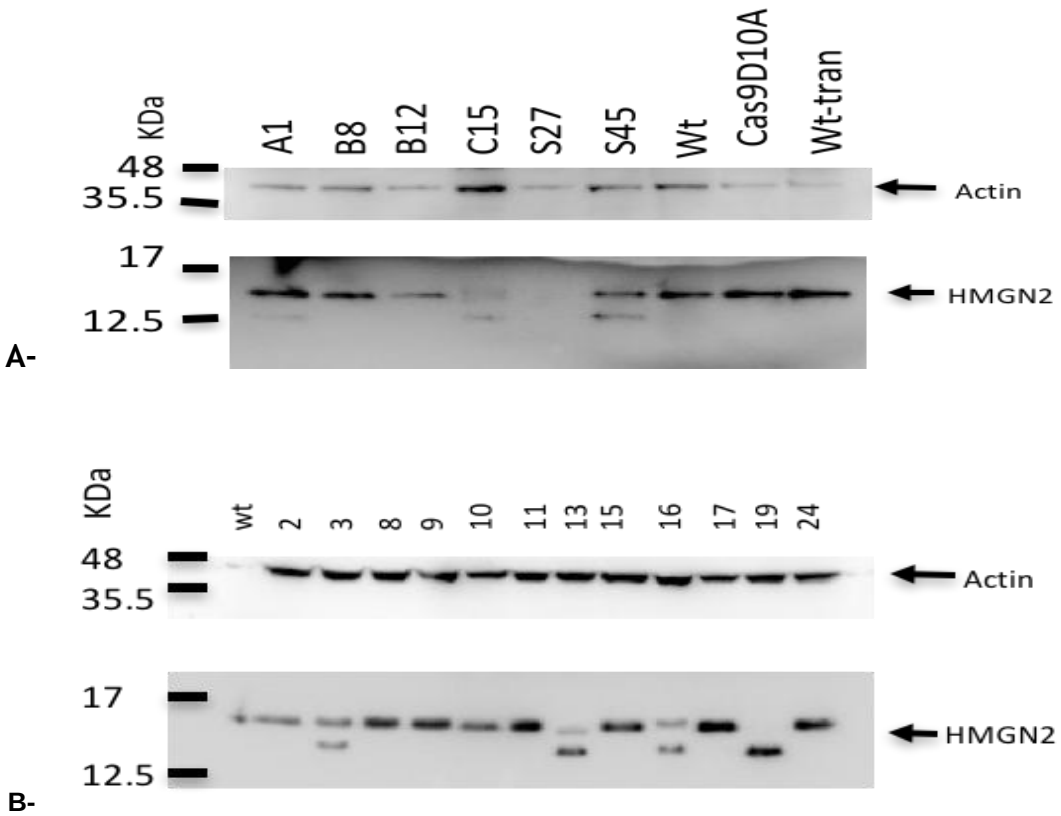


Figure 3.13: Some western blotting for different lines screened from two independent transfections.

Examples of western blots to screen different monoclonal lines are given. P19 cells were independently transfected twice using 15 μg of the multiguide plasmid (pMultinick002_4). The GFP cells from one transfection were enriched in one transfection using FACS sorting. The sorted and un-sorted cells were plated into 96 well plates to obtain monoclonal lines using limited dilution method. The recovered monoclonal cells were then screened for HMGN2 loss or disruption using antibody that recognises an epitope at the C-terminus of the protein. **A**. the different lines screened from unsorted poly-population while **B.** from sorted population. The same blot was cut and used against HMGN2 and β-Actin antibodies. The expected size for HMGN2 and β-Actin proteins are 17 and 42 kDa. The name of each line was given on the top. Wt stands for wild type and wt-trans for electroporated wild type. Both wt, wt-trans, plus Cas9D10A are negative controls. NuPAGE™ 12% Bis-Tris gels were used.

Interestingly, from 160 lines in total, only approximately 3% (5 lines) of lines appeared to have biallelic mutations leading to exon loss or frameshifts, while the majority (~20%, 31 lines) have monoallelic mutations. Although the transfection efficiency was similar in the sorted and unsorted populations, FACS sorting appeared to enhance the number of lines with biallelic mutations: four clones (13, 16, 19 and 59) were from the sorted population, whereas only one clone (S27) was obtained from the unsorted population (Figure 3.13).

3.6 Characterisation of monoclonal lines generated by CRISPR/Cas9 nickase targeting exon IV of *Hmgn2* gene

3.6.1 Determine the exact mutation in each line by using the RFLP assay and sequencing

Several of the monoclonal lines studied by western blotting were further screened using the *Bsl*I RFLP assay (Figure 3.14) and sequencing (Figure 3.15). Lines B8, B12, 22 and 41 all expressed full length HMGN2 protein as shown by western blotting, although differences in the intensity of the bands indicated that there might have been some mutations affecting protein production or stability (Figure 3.13). Genomic PCR using HMGN2 primer set 2 generated a single band that was the same size as the band from wt cells, suggesting there were no large scale insertions or deletions (Figure 3.14a). The *Bsl*I RFLP-assay showed the same pattern of bands for these lines as for the wt cells, so the *Bsl*I site was not disrupted in these cells (Figure 3.14b and c). The sequencing of the PCR products of these lines also indicates that these lines do only have the wild type allele (data not shown). It was concluded that these lines do not have mutations in the *Hmgn2* gene.

In contrast to the single PCR product from wt cells, the full length PCR products from clones A1, C15, S45, 13, 19, 28 and 48 had two bands, suggesting the presence of two alleles with different sizes (Figure 3.14a). Line 16 showed three bands, which could indicate the presence of more than two copies of *Hmgn2,* or that this line is not pure. The *Bsl*I RFLP assay on lines A1, C15, S45, 28, 48, and 16 generated four bands instead of the two that were expected (Figure 3.14b and c). These patterns of digestion are hard to interpret, and could arise from incomplete digestion, from a polyclonal mixture of cells with different mutated *Hmgn2* alleles, or from multiple copies of the *Hmgn2* allele in each line.

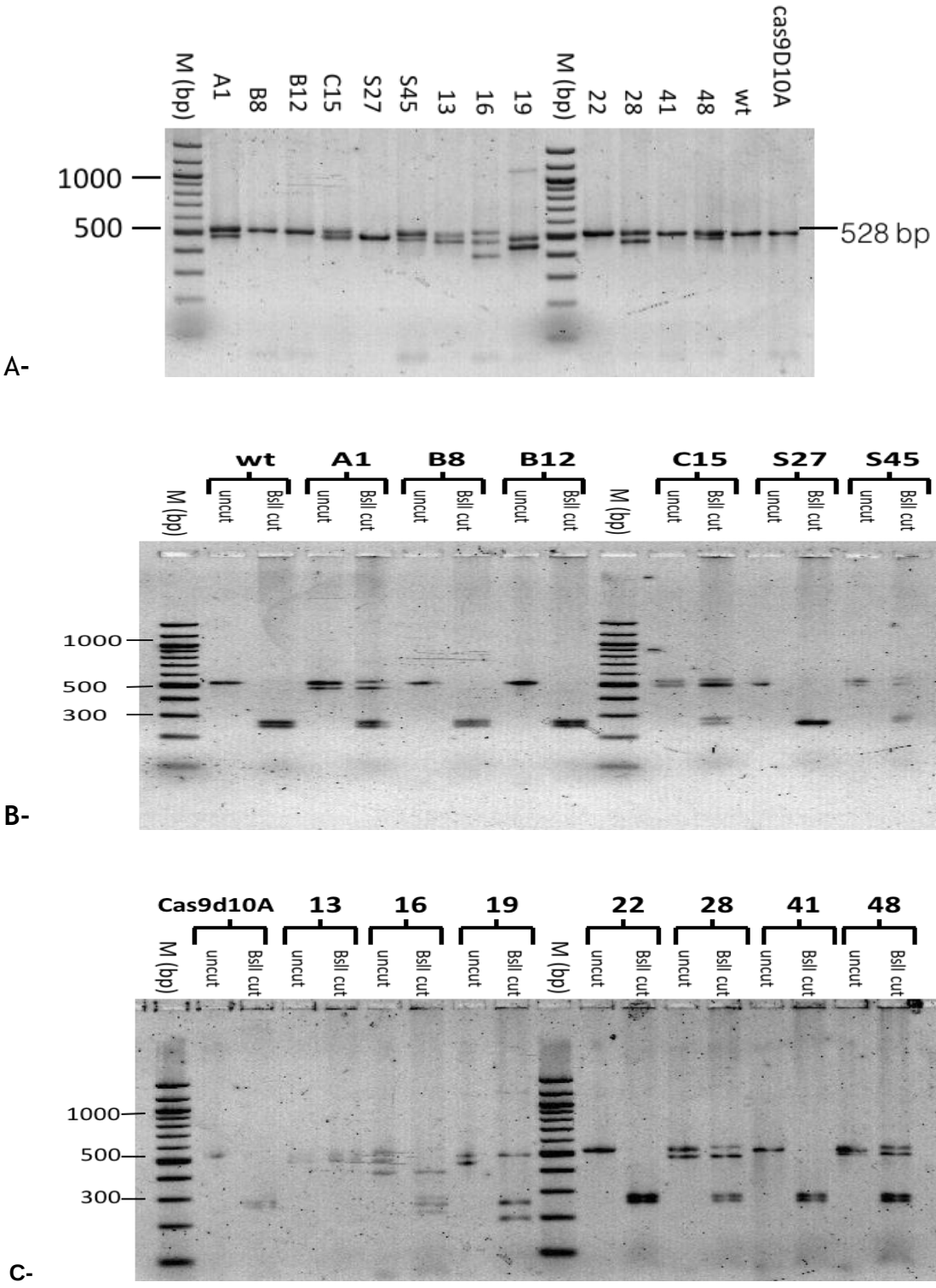


Figure 3.14: RFLP-assay for some of monoclonal lines generated after the transfection of pMultinick002_4 plasmid.

The DNA from each sample was amplified using primer set 2 and yields a product with 528 bp (**A**). The PCR products from lines (B8, B12, 22 and 41) are one band that has a similar size of wild type and Cas9D10A, whereas lines A1, C15, S27, S45, 13,16, 19, 28 and 48 showed either two or three different size bands. Then, all PCR products were used for the RFLP assay by *Bsl*I enzyme (**B** and **C**). The digestion pattern of the PCR product from B8, B12, 22 and 41 lines was similar as wild type and Cas9D10A (255 and 273 bp). However, A1, C15, S45, 28 and 48 lines, that showed two bands in the uncut, revealed 4 bands when the PCR products digested with *Bsl*I. In S27 line, only one band approximately 250 bp was obtained after the *Bsl*I digestion. Bands for uncut and cut of line 13 and 16 were not been able to visualise although the same amount was used in gel A. The digestion of PCR products (two bands~500 bp) of line 19 showed unique pattern. The upper band completely abolish the digestion by *Bsl*I while the lower band was digested into two bands (~200 and 250 bp). 1% agarose gel was used for electrophoresis.

Line S27, which does not show any HMGN2 protein by western blotting (Figure 3.13), gave a single, full length PCR product that was smaller than the wild type (Figure 3.14a). The *Bsl*I RFLP assay for S27 showed one strong band of around 250 bp, in contrast to the two bands of 255 and 273 bp in the wt sample (Figure 3.14b). In order to determine the exact mutations, the PCR product from the S27 line was cloned into a plasmid vector, and seven plasmid clones were sequenced. All seven clones had the same mutation - a deletion of 23 bp in exon IV that disrupts the original *Bsl*I recognition site but also creates a new *Bsl*I site (Figure 3.15 a-1). This sequencing data is consistent with the RFLP pattern that was observed. Taken together, these results are consistent with both *Hmgn2* alleles in the S27 line having the same deletion, possibly as a results of homologous recombination. This mutation will be referred to as $Hmgn2^{del23/del23}$.

The predicted protein from the $Hmgn2^{del23}$ allele would have a deletion of the last 8 amino acids of the NBD, and a frameshift resulting in a 13 amino-acid polypeptide instead of the HMGN2 C-terminal regulatory domain (Figure 3.15 a-2). This mutant protein is named as *p.K40_K48delfs*54* or HMGN2ΔNBD_CHUD. Therefore, this line is predicted to express a short form of HMGN2 with partial disruption of NBD and lack of the C-terminus. This explains the inability of anti-HMGN2 to detect any HMGN2 protein by western blotting (Figure 3.13), as this antibody is raised against a C-terminal epitope. Several other commercial antibodies raised against full length HMGN2 protein were tested, but failed to detect any protein, even the wild type protein (Data not shown).

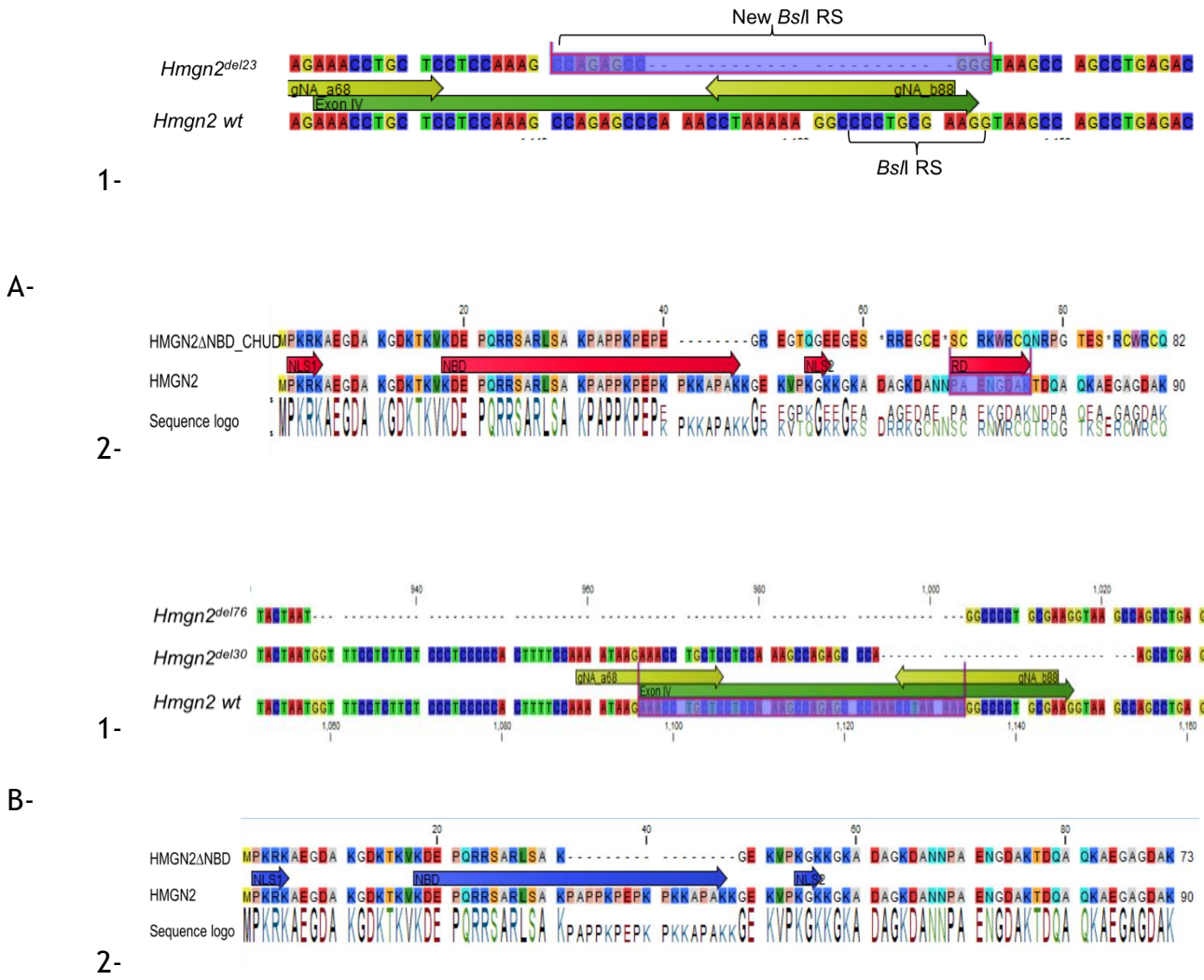


Figure 3.15: Screenshot from CLC showing the exact mutations (1) and predicted proteins (2) for S27 (A) and 19 (B) lines.
The green arrows represent the sequences of exon IV of *Hmgn2* while the yellow arrows are for gRNAs. 7 clones were sequenced from each cell line and aligned against the reference sequence. All clones for S27 revealed a deletion of 23 bp in the exonic sequence (A-1). The mutation leads to the formation of new recognition site for *Bsl*I (CCNNNNNNNGG) that is highlighted on the sequence (A-1). The old *Bsl*I recognition site (RE) is marked on the Hmgn2 wild type (wt) sequences. A-2 shows the predictable protein (amino acid sequence) of S27 mutant allele and the presence of three stop codons downstream the mutation (stars). Clones from 19 line displayed two different alleles: one with deletion of 76 bp (38 bp at the exon and 38 bp at the intron region), while the another with deletion of 30 bp (10 bp at the exon region). These two mutations affect either the splicing donor or acceptor (B-1). The outcome of these mutant alleles in 19 line is the skipping of whole exon (51 bp); therefore, short form of protein with 17 amino acids less than wild type (90 amino acid) will be produced (B-2). The red and blue arrows in A-2 and B-2, respectively, are the functional domains of HMGN2 which include NLS1/2 (nuclear localisation signal), NBD (nucleosome binding domain) and RD (C-regulatory domain).

Line 19, which has a single, shorter form of HMGN2 protein on the western blot (Figure 3.13), yielded two full-length PCR products (Figure 3.14a). One of these products was digested by *Bsl*I to give bands of ~200 and ~255 bp, whereas the other was resistant to digestion (Figure 3.14c). Cloning and sequencing of the

PCR products revealed two alleles with two different deletions at exon IV of *Hmgn2* (Figure 3.15 b-1). The genotype for this line is $Hmgn2^{del76/del30}$ in which the first allele has a deletion of 76 bp (38 bp in the exon IV), and the another allele has a deletion of 30 bp (23 bp exonic), which includes the *Bsl*I site. The sequences of these alleles are consistent with the results of the RFLP assay. Both mutations affect the splicing of exon IV of *Hmgn2* gene. The mutation in the $Hmgn2^{del76}$ allele causes the removal of the splice accepter sequence, while $Hmgn^{del30}$ allele has lost the splice donor sequence. The consequence of these mutations is likely to be the loss of exon IV from the mature mRNA transcript (Attanasio et al., 2003), and this was confirmed by qRT-PCR (data not shown). The loss of exon IV would result in a protein that is 17 amino acids shorter than the wt protein (named as *p.P32_K48del* or HMGN2ΔNBD) (Figure 3.15 b-2), which is consistent with the short HMGN2 form that was observed on the western blot.

Line 59 was also characterised, although the data is not shown here. One allele of line 59 expresses HMGN2 with a deletion of exon IV. As in line 19, this results in the shorter protein *p.P32_K48del*, which is missing the second half of the NBD. The second allele is predicted to express a variant of HMGN2 that has an intact NBD, but a different C-terminal domain due to a frameshift (*p.A46_Lys47insfs)*. As this line expresses two different HMGN2 protein forms, which could have differing effects of the cell, it was not used in most of the subsequent experiments.

3.6.2 Testing the ability of binding of the new mutant forms of *Hmgn2* to the chromatin

The 30 amino acid nucleosome binding demine (NBD) is highly conserved in HMGN variants (Crippa et al., 1992; Bustin, 2001). There is an internal motif of 8 amino acids (RRSARLSA) that is responsible for the specificity of HMGN binding to the nucleosome core particle (CP) compared to free DNA (Ueda et al., 2008). Moreover, mutation or phosphorylation of the two serine residues in this motif abolishes the binding of either HMGN1 or 2 to chromatin (Prymakowska-Bosak et al., 2000; Ueda et al., 2008). Although the RRSARLSA motif confers nucleosome binding specificity, other parts of the NBD were shown to be important for binding efficiency. Specifically, deletion of the last seven amino acids of the

NBD and the C-terminal domain reduced chromatin binding *in vivo* to 22% of wt protein, as measured by FRAP (fluorescence recovery after photobleaching) (Ueda et al., 2008). This mutation is similar to that in line S27 line, which has lost the last eight amino acids of the NBD and has an alternative 13 aa C-terminal region (HMGN2ΔNBD_CHUD, Figure 3.16). Line 19 has a larger deletion of the NBD, with 17 amino acids deleted, although the central RRSARLSA motif and the C-terminal domain are intact (HMGN2ΔNBD).

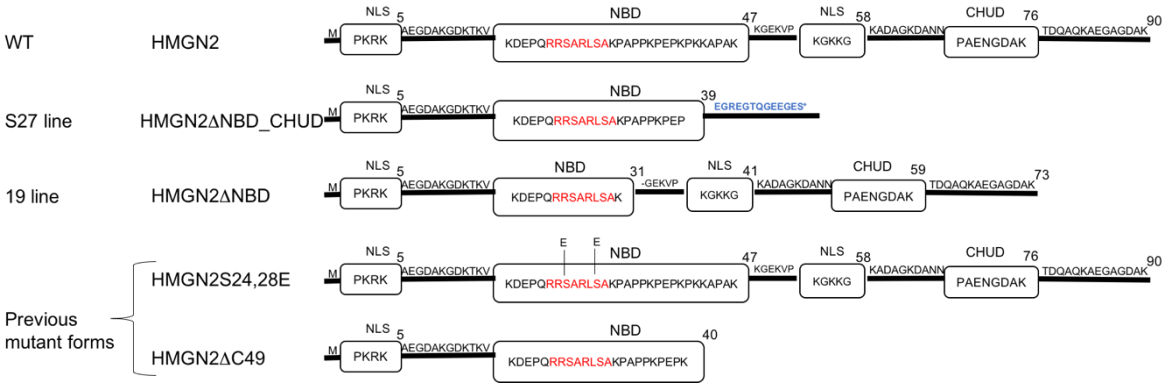


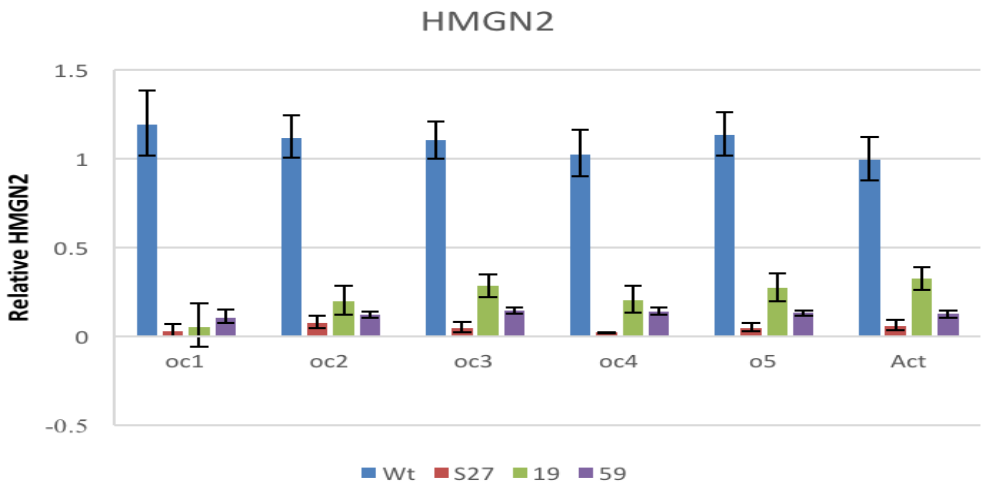
Figure 3.16: Outline of different mutant HMGN2 proteins.
Schematic representation of the structure of the HMGN2 protein and outline of the HMGN2 deletion mutants in S27 and 19 lines, plus in literature (HMGN2S24,28E and HMGN2ΔC49). The numbers at the end of each sequence correspond to amino acid position of the last residues of each domain. The known functional domains of the protein are indicated by black boxes (NBD, nucleosome binding domain; NLS, nuclear localization signal; CHUD, chromatin unfolding domain). The conserved HMGN motif that confers the binding to nucleosomes within the NBD is given in red. The blue amino acids of HMGN2ΔNBD_CHUD in S27 line are the alternative aa resulted from the frameshift mutation. The subsequent star after the blue aa represents the first predicted stop codon. The HMGN2 in HMGN2S24,28E has substitution of two serine (S) residues at position 24 and 28 to glutamic acid (E) (modified from Prymakowska-Bosak et al., 2001; Ueda et al., 2008).

In order to investigate whether protein HMGN2$^{p.P32_K48del}$ (HMGN2ΔNBD) in line 19 is able to bind to nucleosomes, chromatin immunoprecipitation (ChIP) assay followed by Q-PCR was performed. Line S27 was used as a negative control, as the HMGN2 antibody does not detect HMGN2 protein in this line. Line 59 was also included in this experiment, as it expresses the same HMGN2ΔNBD protein as line 19.

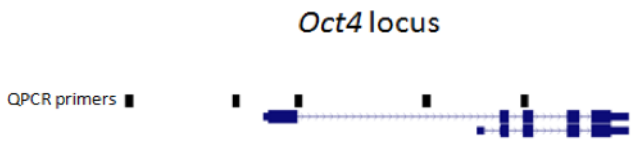
ChIP samples were interrogated using primer sets specific for several regions spanning the promoter and transcribed region of the *Oct4* gene, and also a primer set corresponding to the *β-actin* gene (Figure 3.17b) (Rehbini, PhD thesis 2016). The relative amount of HMGN2 at *Oct4* and *β-Actin* gene TSS in wild type cells was around 1 at all primer sets tested, which is consistent with previous

data from the lab (Rehbini 2016 and K West, personal communication) (Figure 3.17a). No HMGN2 binding was observed in the S27 line because the mutant protein is undetectable by the antibody. This demonstrates that the ChIP assay is specific for HMGN2.

Lines 19 and 59 ($HMGN2^{p.P32\_K48del/p.P32\_K48del}$ and $HMGN2^{p.P32\_K48del/p.A46\_Lys47insfs}$, respectively) have greatly reduced binding of HMGN2 at all the loci tested. Specifically, binding of $HMGN2^{p.P32\_K48d/p.P32\_K48del}$ in line 19 line was reduced to less than 30% of wild type binding. Binding of $HMGN2^{p.P32\_K48del/p.A46\_Lys47insfs}$, was reduced to 10-15% of wild type binding. This is consistent with there being less $HMGN2^{p.P32\_K48del}$ protein in line 59 than in line 19, because it is expressed from one rather than two alleles. This data shows that $HMGN2^{p.P32\_K48del}$ (HMGN2ΔNBD) retains some chromatin-binding ability, although it is considerably reduced compared to the full length protein.



A-



B-

Figure 3.17: The relative amount of HMGN2 binding across the *Oct4* gene in different lines (wild type, S27, 19 and 59).
(**A**) The binding of HMGN2 was determined using ChIP-QPCR in different locus of *Oct4* gene and near the transcription start site of *β-actin* (housekeeping gene) for all lines of P19 cells mentioned previously. The enrichment of HMGN2 in wild type cells was around 1 fold while it was merely detectable in other lines at all regions examined (compare the blue bar with other). *Oct4* locus were named from 1 to 5. Oc refers to *Oct4* whereas Act to *β-Actin* gene. The qPCR signal for HMGN2 was normalised to that observed for 0.5 ng input DNA, and that the scale is arbitrary. The error bars represent the average standard deviation from 3 technical PCR replicates. (**B**) The black bars identify the position of primer sets across the *oct4* gene, and they are ordered based on the transcription from oc1 to oc5 {oc1 (-1781), oc2 (-399), oc3 (+410), oc4 (+2070), oc5 (+3342)}.

3.7 Discussion

In order to generate functional knockouts of HMGN2, it was decided to target the essential nucleosome binding domain, which is encoded by exons III and IV (Ueda et al., 2008). A double nickase (CRISPR-Cas9D10A) strategy using truncated gRNAs was used, as this approach has been shown to minimise off-target mutagenesis (Shen et al., 2014; Fu et al., 2014).

Three multiguide plasmids were constructed, expressing either two or four gRNAs targeting exon IV of the *Hmgn2* gene. However, low transfection efficiency and high cell toxicity was observed with these multiguide plasmids, and only transfections with pMultiNick002_4 had enough viable cells for further culturing. Therefore, it might be necessary to co-transfect three vectors (two expressing gRNA, one expressing Cas9D10A) to avoid the toxicity from this large multiguide vector, although this may reduce the mutagenesis rate in the cells. Alternatively, as the SpCas9 gene is quite large (4.1 kb), the recently discovered orthogonal Cas9 from *Staphylococcus aureus* could be used, as it is 1 kb shorter (Ran et al., 2015). Other approaches such as delivering purified recombinant Cas9 protein that can induce site-specific mutations at frequencies of up to 79% in pluripotent stem cells can also be used (Kim et al., 2014).

In our hand, we only obtain up to 20% mutagenesis efficiency using the nickase system and all-in-one vector as estimated by RFLP assay in polyclonal population. However, 37-40% of indel mutagenesis were obtained from different studies who used all-in-one vector Cas9D10A (Ran et al., 2014; Chiang et al., 2016). The obvious reasons for this are the cell toxicity and low transfection efficiency that can affect the indel mutation efficiency. In addition to the low transfection efficiency and high cell toxicity, the detection of indel mutation at *Hmgn2* gene by RFLP seems to affect the accurate determination of indel efficiency. The RFLP sometimes underestimates the percentage because of the inability of detection the small mutations that do not affect the recognition sites of both enzymes. Alternative method for determining the indel efficiency can be surveyor assay that can detect any small differences between the mutant alleles and the wild type allele. Using this method may also be beneficial for the region that does not have an indicative recognition site or region with a recognition site but requires another site to work.

At least five lines (S27, 13, 16, 19 and 59) were found with biallelic mutations while the majority of indel mutations were found to be monoallelic. FACS sorting enhanced the screening strategy by enriching the number of lines with biallelic mutations. This is consistent with previous studies showing that FACS sorting can markedly enrich for clones that have mutations at the targeted gene (Duda et al., 2014; Chiang et al., 2016).

Two out of 160 screened lines are mostly interesting which are S27 and 19 because they have biallelic mutations that affect the products from both alleles and yielding truncated proteins. The protein in S27 line is lack of NLS2 and CHU domains and has a disrupted NBD (HMGN2ΔNBD_CHUD), while the protein in 19 has larger disruption of NBD but the N- and C-terminal are intact (HMGN2ΔNBD). S27 and 19 lines are named as S27ΔNBD_CHUD and 19ΔNBD, respectively. Losing of NLS2 in S27 would disrupt the localisation of mutant proteins but because the inability to detect this protein by any of antibodies it was hard if they are localised in nucleus or not. The localisation of the protein in 19 line is mainly in the nucleus (data not shown).

In terms of binding of mutant proteins in S27 and 19 lines, although we were unable to ChIP the truncated protein form in S27 because the antibodies against the C-terminal or whole protein failed to detect in western blot, the inability of detection this mutant protein in this line confirm the efficiency and specificity of antibody to detect the wild type HMGN2 in parental cells. In addition, we assumed the binding efficiency of mutant HMGN2 in S27 line to bind to the nucleosomes to be reduced because a pervious similar mutant peptide protein showed a reduction in the binding efficiency to the chromatin *in vivo* (Ueda et al., 2008). However, as the HMGN2 antibody can detect the truncated protein in 19 line, the ChIP experiment showed that the binding efficiency for the short form in 19 line has been significantly reduced around 80% comparing to wild type protein. Together, mutant HMGN2 proteins in 19ΔNBD lines (and probably S27 line) are unable to bind efficiently to the chromatin, or specifically to nucleosomes. Although these mutant lines targeted by CRISPR-Cas9D10Aa t exon IV of Hmgn2 gene can reduce the efficiency of mutant protein to bind the chromatin, a complete functional knockout was not achived by desgining CRISPR against exon IV. The complete functional knockout can be facilitated by

designing of CRISPR gRNA target exon III that is transcribed to the most HMGN motif that confers the binding to nucleosomes within the NBD.

In general, it is clearly that the efficiency of having biallelic functional knockouts was very low, approximately 3%, although the transfection efficiency and indel mutation efficiency were 35% and 20%, respectively. The reason of having low biallelic mutations can be due to the low transfection efficiency, as well as the probably of high efficient and accurate DNA repair in EC and ES cells. It was found that murine ESCs have a good DNA strand break repair capacity (Saretzki et al., 2004), but gene loss-of-function in ESCs and primary cells might be difficult comparing other cells. It was reported that the efficiency of mediating cleavage at the targeted site by CRISPR-Cas9 is varied from cell to cell and from locus to another locus in the same cell type (Mali et al., 201a3; Jiang et al., 2013). In some studies, the efficiency for knocking out a protein-coding gene has been reported to be 20%–60% in mouse ESCs and zygotes (Wang et al., 2013; Yang et al., 2013).

Finally, although the mutations in most lines with biallelic mutations are heterozygotes suggesting the double targeting by Cas9D10A, both mutations in S27 alleles are identical indicating the possibility of the homologous recombination in deriving the mutation in the another allele. There is a considerable variation in the type of mutations generated by CRISPR-Cas9D10A including small and large deletions, and to less extend small insertions, that result whole or partial exon loss, or the whole C-terminal loss due to the frameshift effect. In addition to the low efficiency of biallelic mutation, the variation in mutation types and their products complicates the chance of having cell line replicates. Only one line (19ΔNBD) was obtained and ensured that the mutant HMGN2 in this line has less binding to the chromatin, specifically to the nucleosomes

Chapter 4 Generating complete knocked-out lines of HMGN2 in P19 cells using CRISPR-Cas9 technology that target exon I

4.1 Introduction

Chapter 3 details the generation of monoclonal lines in which HMGN2 is mutated within its nucleosome binding domain. Nucleosome binding is required for HMGN2 to modulate chromatin structure (Bustin, 2001; Rochman et al., 2009), so it was considered that line 19ΔNBD contains a functional knockout of HMGN2. However, it cannot be ruled out that the remaining domains of the mutated HMGN2 could play other roles within the cell that are not dependent on nucleosome binding. In addition, it was decided that more than one knockout line was required in order to have a confidence that any phenotypes would be a result of the HMGN2 mutations, and not due to off-target mutagenesis or phenotypic drift under culture conditions.

A different strategy was chosen for the generation of more HMGN2 mutant lines. Targeting the beginning of the protein coding region such as the first start codon (ATG) is likely to disrupt the expression of most genes (Hruscha et al., 2013; Jiang et al., 2013b; Shalem et al., 2014), and is less likely to accidentally generate truncated protein artifacts with residual biological activity (Byrne et al., 2014). gRNAs targeting the 5` or 3` UTRs were shown to be highly ineffective (Doench et al., 2014). However, the incidence of an alternative start codon should be considered (Li et al., 2016). Thus, we selected two gRNAs closest to the ATG start codon of *Hmgn2* to maximise the disruption of the wild-type protein (Figure 4.1). The first exon has a 5` untranslated region (UTR) and a 15 base pair coding region. The first gRNA overlaps the first exon and extends into the first intron. To minimise the chances of targeting other sequences in the genome such as pseudogenes and *Hmgn4,* the other gRNA lies in the intronic sequence. The offset distance between these two gRNAs is 30 bp. The two gRNAs were called mN2ATG_A69_h7SK and mN2ATG_B77_hH1 and the Cas9-cleavage guided by these two gRNAs impedes the digestion of restriction

enzymes *Nla*III and *Ban*II, respectively. In addition, two primers that flank this region were also designed and named as mN2ATG_PCR1_forward and reverse.

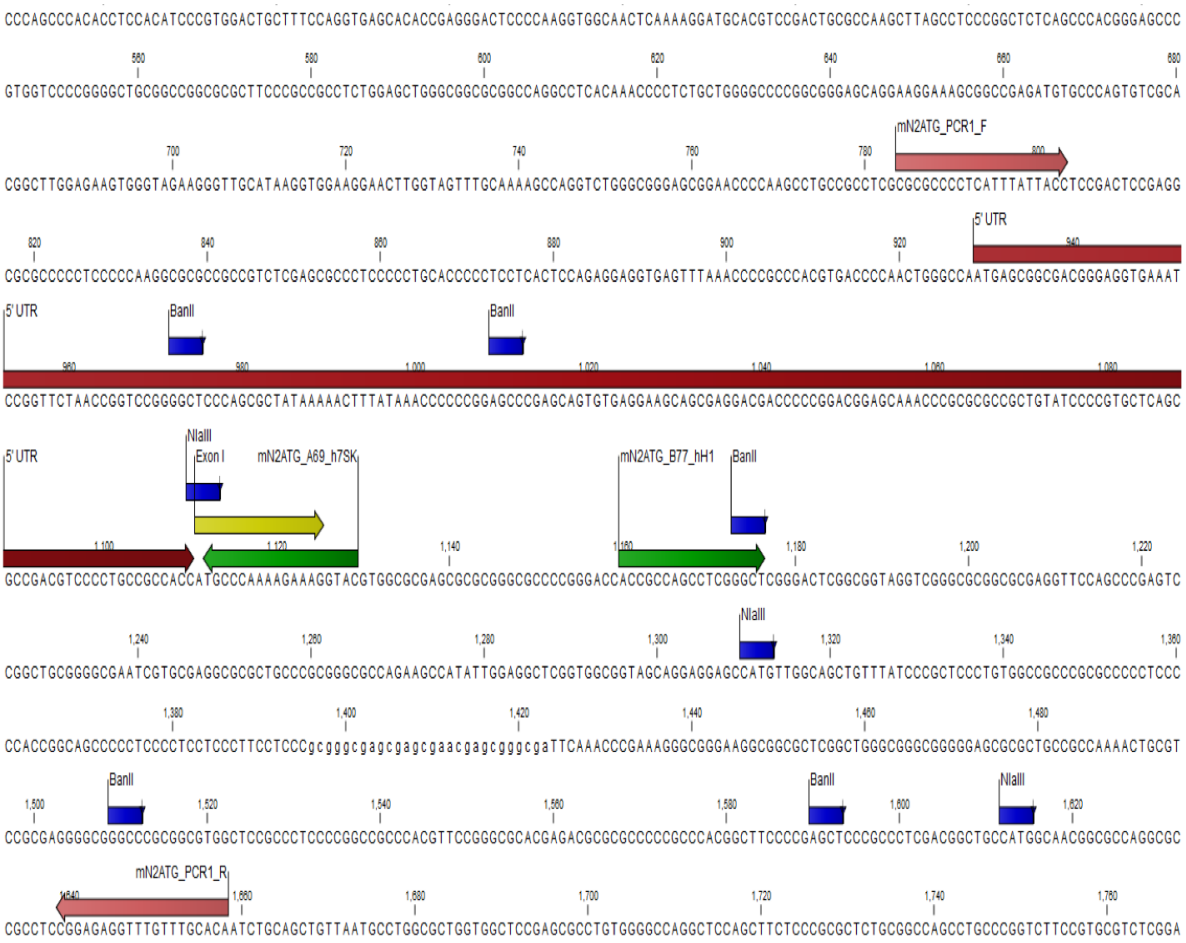


Figure 4.1: Screenshot of CLC program showing the two gRNAs used in the nickase for targeting exon I of *Hmgn2* gene.
The exon I of *Hmgn2* gene is only 15 bp as annotated by the yellow arrow. There is a 5` untranslated region that is just upstream the exon I (long red arrow). The first ATG in the exon I of *Hmgn2* is the first three sequences in the yellow arrow. There is PAM sequence just near the ATG and this is the CCA in which the A is part of ATG. The light green arrows represent the sequence for the gRNAs (mNATG_A69 and mN2ATG_B77). The two small red arrows are the initial primer set (mN2ATG_PCR1_F and R) to amplify this region, and the expected band by this set is 875 bp. *Nla*III and *Ban*II are restriction enzymes that have been used to track the mutation occur by mNATG_A69 and mN2ATG_B77, respectively. The blue annotation boxes represent the recognition sites for *Nla*III and *Ban*II.

4.2 Objectives

1. Use the CRISPR/Cas9 nickase technology to generate P19 cells with a complete knockout of HMGN2.

- Construct two vectors, each includes an individual gRNA to carry out genome editing at exon I of *Hmgn2* gene locus.

- Transfect vectors into P19 cells and test the efficiency of different gRNAs to introduce indel mutations at the desired locus.

- Generate monoclonal lines and screen for loss of HMGN2 protein using immunofluorescence.

2. Characterise the P19 cell lines generated in objective 1.

- Confirm the complete loss of HMNG2 protein by western blot and RNA expression.

- Identify the exact mutations in all candidate lines by sequencing.

4.3 Use CRISPR nickase technology to generate HMGN2-knockout cells

4.3.1 Construct gRNA expression vectors to target exon I of *Hmgn2* using a paired nickase

Construction of plasmids expressing gRNAs was carried out as described earlier (Materials and Methods, and Chapter 3). Complementary single stranded DNA oligos (mNATG_A69_h7SK top and bottom together, and mN2ATG_B77_hH1 top and bottom strand together) were annealed, then cloned into expression vectors h7SK_4-4 and hH1_1-2, respectively, after the parental vector was digested by *Bbs*I (similar as shown in chapter 3). These individual gRNA vectors were named p.Guide001-2_hH1\ExI and p.Guide004-4_h7SK\ExI in order to discriminate them from the exon IV CRISPR. Colony PCR was used to screen colonies. Most of the screened colonies had the protospacer inserted successfully into the backbone (Figure 4.2). One colony from each transfection was selected for further work, and the correct insertion of the gRNA sequences was confirmed by DNA sequencing.

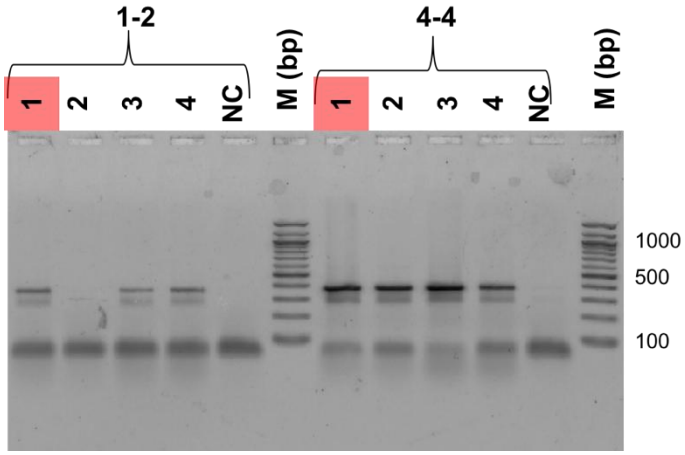


Figure 4.2: Colony PCR for the p.Guide vectors used in CRISPR-Cas9D10A targeting exon I. Vectors hH1_1-2 and h7SK_4-4 were firstly digested for 2 hours with *Bbs*I. After annealing of the complementary strand together for each gRNA (mNATG_A69 and mN2ATG_B77), the gRNAs were ligated into the compatible vectors containing different promoters. mNATG_A69 gRNA was ligated into h7SK_4-4 and the new plasmid was named as p.Guide004-4_h7SK\ExI while mN2ATG_B77 was ligated into hH1_1-2 vector and called p.Guide001-2_hH1\ExI (abbreviated as 4-4 and 1-2, respectively). Colony PCR for screening the transformation by using vector primer (T7) and bottom strand of the inserted oligo. The expected bands sizes for colony PCR are370 and 390 bp for p.Guide001-2_hH1\ExI and p.Guide004-4_h7SK\ExI, respectively. Colonies of each gRNA vector were labelled from 1 to 4. In addition, one blue colony is used as negative control (NC). The red highlighted numbers of colony represent the plasmids that were selected for further experiments. M stands for marker.

4.3.2 Transfection of P19 cells with gRNA plasmids

As detailed in chapter 3, transfection of the multiguide plasmid, MultiCas9D10A, appeared to be very toxic for P19 cells, possibly due to its large size. Therefore, it was decided to use individual gRNA plasmids instead of a multiguide plasmid for targeting the *Hmgn2* start codon. The disadvantages of this approach are that the three plasmids may be present in varying ratios in different cells, and that the gRNA expression vectors do not carry a reporter gene, so transfection efficiency is calculated from the pCas9-EGFP or pCas9D10A-EGFP plasmids.

To investigate the capabilities of the new gRNAs to introduce mutations at exon I of *Hmgn2*, transfections were carried out using the electroporation delivery system (Neon). The two gRNA expression plasmids (p.Guide001-2_hH1\ExI and p.Guide004-4_h7SK\ExI) were either electroporated individually with p.Cas9_EGFP or both together along with p.Cas9D10A_EGFP. Transfection efficiencies and the percentage of healthy cells are shown in Figure 4.3 and Table 4.1. Plasmids Cas9-EGFP or Cas9D10A-EGFP gave similar transfection efficiencies of around 60%. Co-transfection with one or two gRNA plasmids reduced the efficiency to 45- 55%. The percentage of healthy cells was 81% for untransfected cells, and 70-79% for transfected cells.

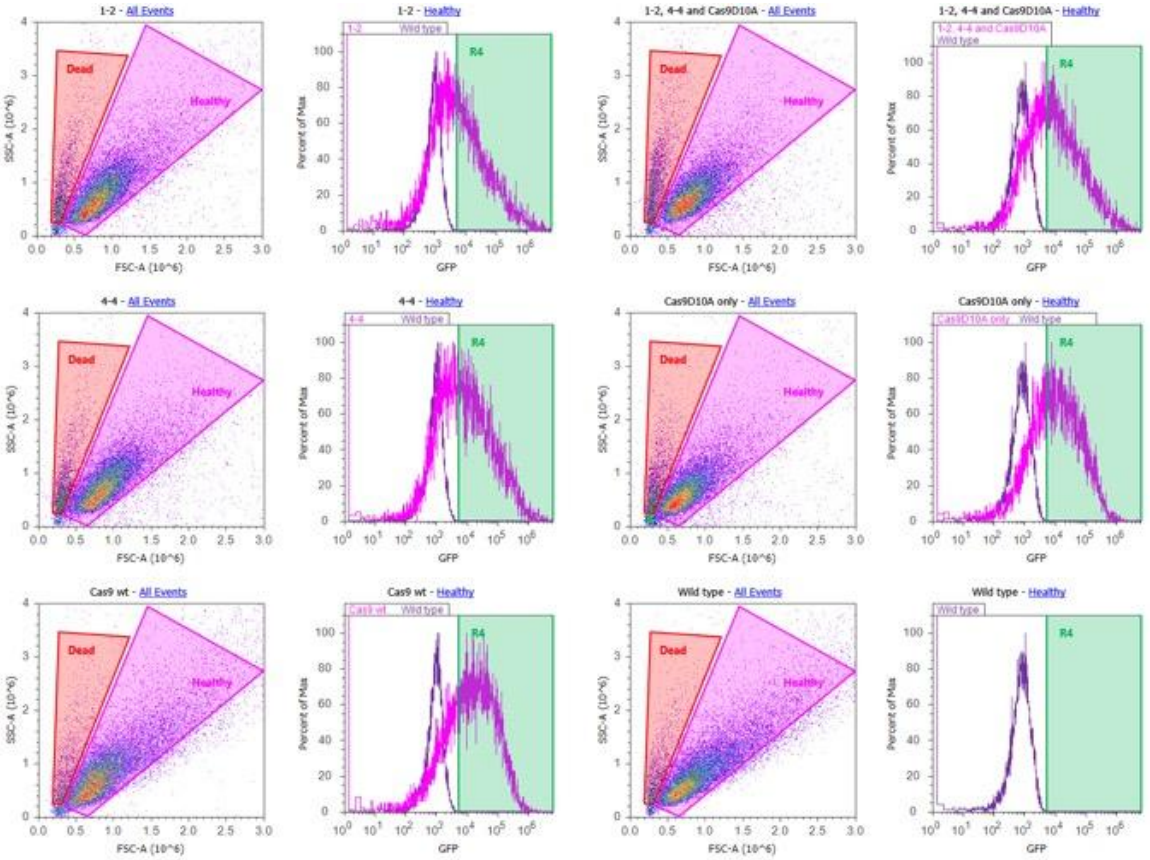


Figure 4.3: The transfection efficiency after two days of transfecting different plasmids combination targeting exon I of *Hmgn2*.
The dot blot and its histogram for each transfection were shown next to each other. The name of each transfection were given on the top of dot blots and histograms. 1-2 is a transfection of p.Guide001-2\ExI (5 μg) and wild type Cas9 (10 μg) while 4-4 for co-transfecting p.Guide004-4\ExI (5 μg) and wild type Cas9 (10 μg). However, 1-2, 4-4 and Cas9D10A refers to the transfection of three plasmids, including p.Guide001-2_hH1\ExI (5 μg), p.Guide004-4_h7SK\ExI (5 μg) and Cas9D10A (10 μg). The transfections of these plasmids into P19 cells were carried out by Neon transfection. The plasmid backbone of Cas9D10A differs from p.MultiCas9D10A that was used previously (Chapter 3). An empty vector containing Cas9 wild type or Cas9D10A were also used as control. The dead cells are gated as the cells in this gate were red when treated with propidium iodide in previous experiment (PI, data not shown). The X axis of the histograms is for channel BL-1 laser (GFP). The percentage of GFP was calculated from the healthy cells that fall in gate 4. Gate 4 was set just after the peak of the wild type. Table 4.1 shows the percentage of GFP positive cells and healthy cells for each transfection.

Transfection	Transfection efficiency (2days, %)	Healthy cell (%)
Cas9	60	79
p.Guide001-2_hH1\ExI and Cas9	45	78
p.Guide004-4_h7SK \ExI and Cas9	52	74
Cas9D10A	60	72
p.Guide001-2_hH1\ExI, p.Guide004-4_h7SK\ExI and Cas9D10A	55	70
Wild type	0.04	81

Table 4.1: The transfection efficiency and healthy cells as assessed by FACS two days post-transfection of different plasmids targeting exon I of *Hmgn2*.

4.3.3 Optimising the RFLP assay for exon I of the *Hmgn2* gene

It was intended to use RFLP to assay the ability of the gRNAs to introduce mutations at exon I. However, the first PCR attempt, using primers mN2ATG_PCR1_F and mN2ATG_PCR1_R (Figure 4.1) was unsuccessful (not shown). The genomic region around exon I is very rich in GC dinucleotides, so various conditions were altered, including the annealing temperature and concentrations of Mg^{2+}, primers and dNTPs (not shown), with no success. Three additional primer sets were designed (Figure 4.4a), and PCR reactions carried out with all sixteen combinations of forward and reverse primers. In addition, *Herculase II Fusion* and *KAPA* Hi-Fi polymerase enzymes were tested, as these are reported to efficiently amplify GC-rich templates. Most of the amplifications by different combination of primers lead to either non-specific bands or no products by either *Taq, Herculase II Fusion* or *KAPA* DNA enzyme (data not shown and Figure 4.5). Fortunately, one putative pair of primers show the specific band using *KAPA* DNA polymerase. This primer pair is mN2ATG_PCR4_forward and mN2ATG_PCR1_reverse and its product is 977 bp (Figure 4.5). Additional optimisation was required to reduce the formation of non-specific bands. Different annealing temperatures and PCR enhancing agents (DMSO and betaine) were tested. Betaine, which is an isostabilizing agent, equalizes the contribution of GC- and AT-base pairing to the stability of the DNA duplex, increases yield and specificity of PCR products and improves the PCR amplification of GC-rich

DNA sequence (Henke et al., 1997; Frackman et al., 1998). The most appropriate condition that reduced the appearance of non-specific bands was an annealing temperature of 59 ^{o}C with 1 M betaine (not shown).

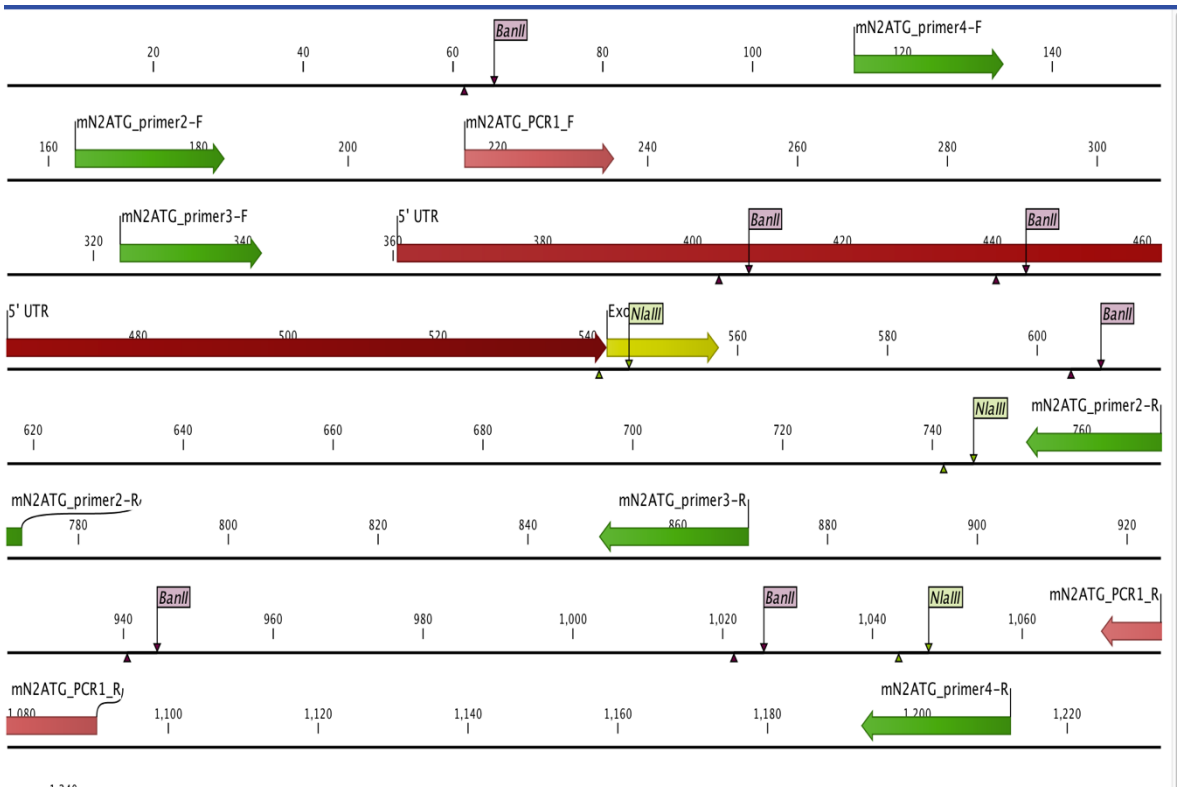


Figure 4.4: Different primer sets designed to amplify the sequence around exon I of *Hmgn2* gene.
A. Three primer sets in green were also designed and used in different combination to amplify the region of exon I of *Hmgn2* gene where the CRISP-Cas9 targets. These primer sets include mN2ATG_PCR2_F/R, mN2ATG_PCR3_F/R and mN2ATG_PCR4_F/R. The small red arrows are for previous primer set. The long dark red arrow is the untranslated region while the yellow is for exon I. F refers to the forward primers while R is the reverse.

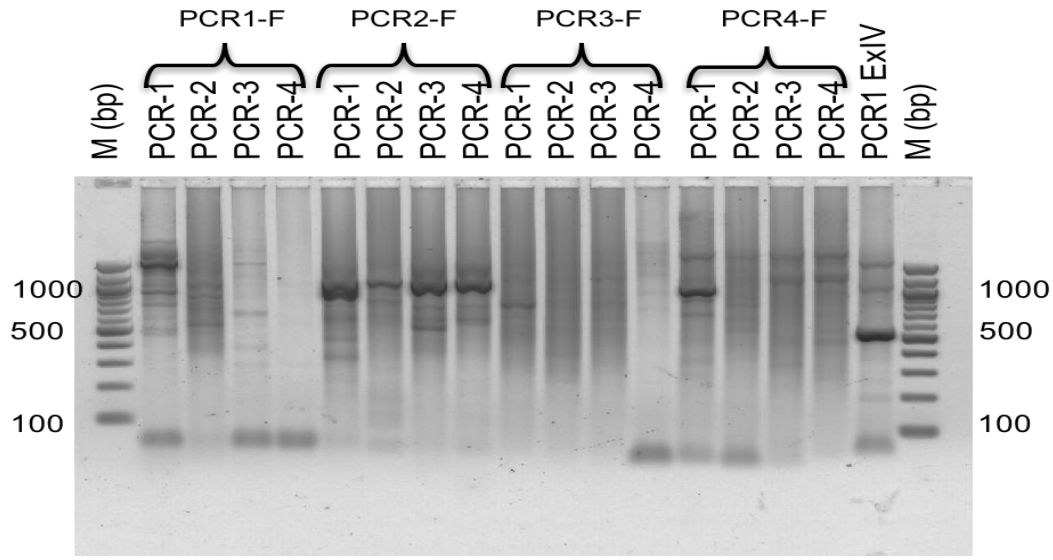


Figure 4.5: PCR amplification for exon I of *Hmgn2* using different combination of primers. The forward primers mN2ATG_PCR1 or 2 or 3 or 4_F were used with all the reverse primers (mN2ATG_PCR1 or 2 or 3 or 4_R) in numerical order. The PCR amplification was carried out using by *KAPA* HiFi DNA polymerase and wild-type P19 DNA. The annealing temperature is 57 ^{o}C and a 1% agarose gel was used. The predicted sizes for each PCR reaction are: PCR1F+1R (875 bp), PCR1F+2R (557 bp), PCR1F+3R (654 bp), PCR1F+4R (997 bp), PCR2F+1R (927 bp), PCR2F+2R (609 bp), PCR2-F3+R (706 bp), PCR2F+4R (1048 bp), PCR3F+1R (767 bp), PCR3F+2R (449 bp), PCR3F+3R (546 bp), PCR3F+4R (888 bp), PCR4F+1R (977 bp), PCR4F+2R (659 bp), PCR4F+3R (756 bp), and PCR4F+4R (1099 bp).

4.3.4 PCR-RFLP assay to investigate mutagenesis efficiency

The PCR-RFLP assay to detect gRNA-induced mutations was carried out using the optimised primer set (mN2ATG_PCR4_F and mN2ATG_PCR1_R) and conditions mentioned in previous section. This primer set gives a product of 977 bp. Indels induced by the two gRNAs, p.Guide001-2_hH1\ExI and p.Guide004-4_h7SK\ExI, are expected to interfere with recognition sites of *Ban*II and *Nla*III, respectively. *Ban*II digestion of PCR products from wild type cells is expected to produce bands of 37, 65, 81,146, 294 and 336 bp, while *Nla*III digestion produces bands of 43, 200, 302 and 432 bp (Figure 4.6).

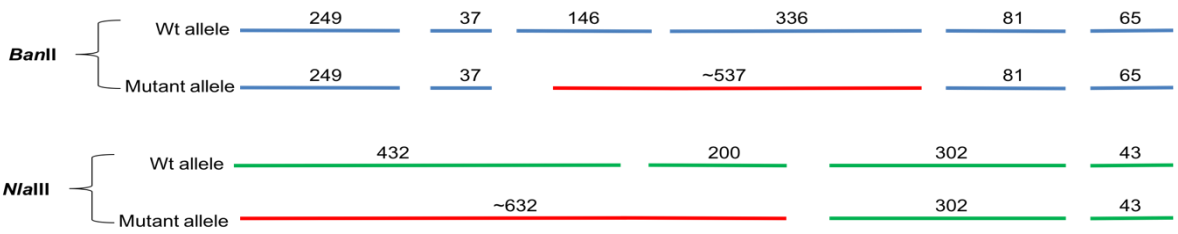


Figure 4.6: The digestion pattern for *Ban*II and *Nla*III. Schematic diagrams for the wild-type and mutant alleles from the *Ban*II and *Nla*III digestions for PCR products amplified using primer set (mN2ATG_PCR4_F and mN2ATG_PCR1_R). The sizes are given on the top of each band. The red bands form as a result of the abolishment of digestions.

All DNA samples obtained from different transfections (section 4.3.2) were subjected to PCR-RFLP. Interestingly, the full length PCR product from untransfected wt cells appeared to be a doublet rather than a single band of 977 bp, whereas the full length product from control cells transfected with Cas9 or Cas9D10A only was a single band (compare lanes 2 and 5 in Figure 4.7A and B). Moreover, digestion of the PCR product from wild type, but not Cas9 or Cas9D10A-transfected cells, by *Ban*II or *Nla*III gave additional bands at 350bp and 330 bp, respectively (indicated by stars in lanes 3 and 4, Figure 4.7A). These additional bands appeared at varying intensities in some of the other RFLP assays (e.g. lanes 9 and 10, Figure 4.7A, lanes 12 and 13, Figure 4.7B).

These additional bands could be due to incomplete digestion, an allelic variant with additional sequence in intron I, or some sort of PCR artefact. The band sizes are not totally consistent with incomplete digestion, as the additional *Ban*II band would be 417 bp, not 350 bp. Sequencing of several PCR clones from the wt DNA only revealed one sequence, and varying the concentration of DNA template or primers in the PCR reaction did not affect the production of these additional bands.

These additional bands make the PCR-RFLP digestion pattern more complex, but it is still possible to detect additional bands resulting from CRISPR-Cas mutagenesis. The full length PCR product for cells transfected with p.Guide001-2_hH1\ExI and wt Cas9 revealed at least two bands in the region of 977 bp, indicating the presence of different sized alleles in this mixed population of cells (lane 8, Figure 4.7A). *Ban*II digestion gave several additional bands at around 500 bp compared with the control (lane 9, Figure 4.7A), consistent with the loss of the *Ban*II site overlapping Guide 1-2 (B77, Figure 4.1). Specifically, the bands at 164 bp and 336 bp convert to around 500 bp because the mutant alleles abolish the digestion. It appears that around 50% of the alleles have been mutated, as judged by the reduction in intensity of the 164 bp band. There is also an additional band of about 220 bp present in the *Nla*III–treated sample, although it is not clear how mutagenesis could give rise to this band (lane 10, Figure 4.7A).

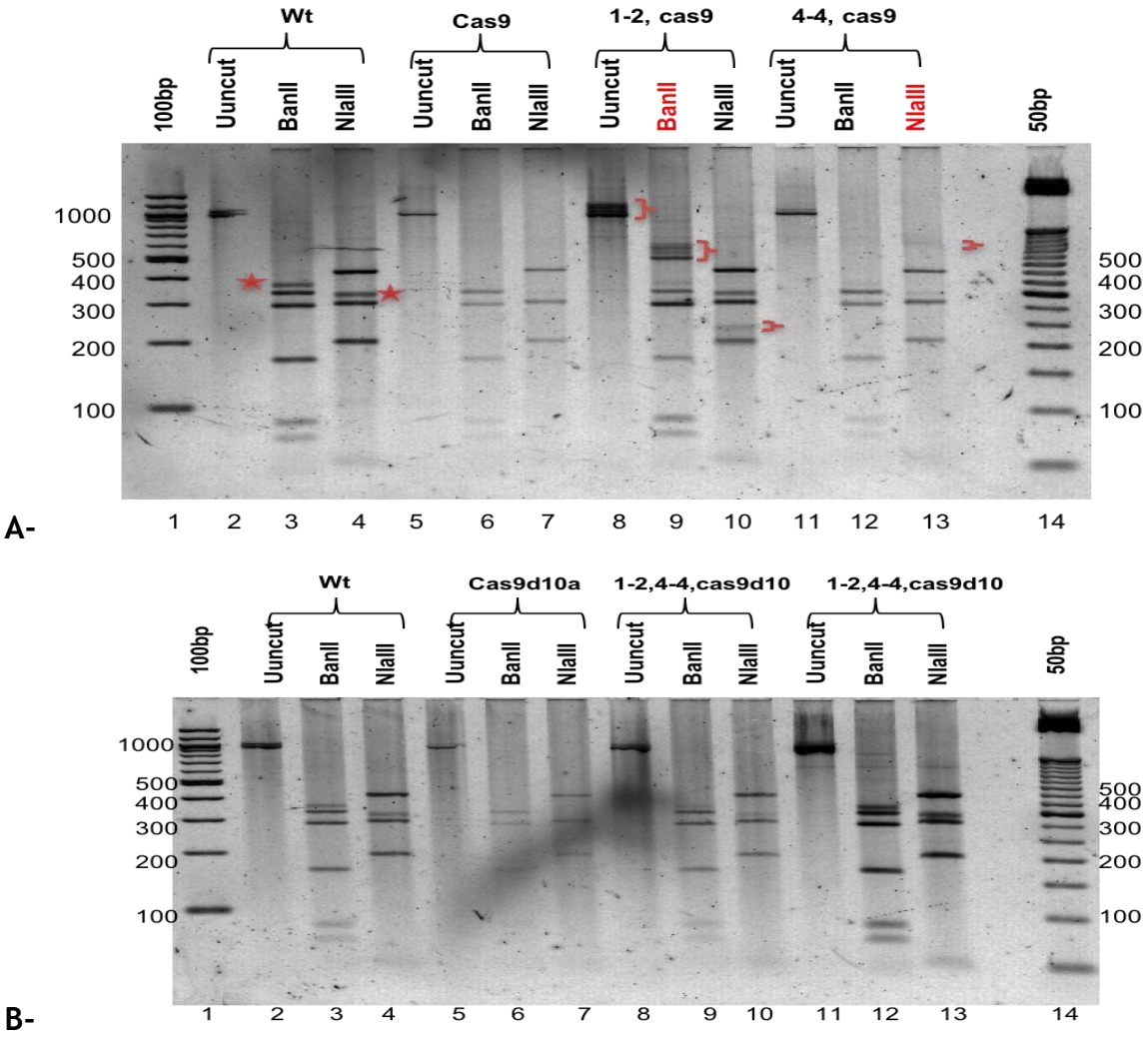


Figure 4.7: The RFLP assay for individual gRNA separately or together (nickase) for CRISPR-Cas9D10A targeting exon I.
RFLP analysis of cells transfected with (A) wt Cas9 or (B) mutant cas9D10A with or without gRNA plasmids. The primer set used here is mN2ATG_PCR4_forward and mN2ATG_PCR1_reverse, with an expected product size of 977 bp. Digestion of wt DNA by *Ban*II or *Nla*III should give products of (37, 65, 81, 164, 294 and 336 bp) and (43, 200, 302 and 432 bp), respectively. An unexpected extra band appears in the digestion of wild type DNA if it was either digested by *Ban*II (~350 bp, lane 3, red star) or *Nla*III (~330 bp, lane 4, red star). The uncut PCR amplification for DNA from cells transfected by p.Guide001-2_hH1\ExI and Cas9 wt generates several bands around the expected size (A, lane 8, red brackets). When the recognition site of *Ban*II is disrupted by gRNA in p.Guide001-2_hH1\ExI (1-2) plasmid, the two bands (164 and 336 bp) become around 500 bp (A, lane 9, red brackets). The gRNA expressed by p.Guide004-4_h7SK\ExI overlaps with the recognition site for *Nla*III. So, when this gRNA directs mutagenesis of the *Nla*III site, the digestion pattern becomes (43, 302 and around 632 bp) (A, lane 13). Lanes 11-13 of B contain the same samples as lanes 8-10, but with twice as much DNA loaded on the gel. A 2.2% agarose gel was used.

*Nla*III digestion of DNA from cells transfected with p.Guide004-4_h7SK\ExI and wt Cas9 revealed an additional faint band at around 632 bp (Figure 4.7A, lane 13). This band would be produced by the combination of the 200 bp and 432 bp bands, following the loss of the *Nla*III site overlapping the guide.

Finally, the DNA from cells transfected with both gRNAs and the Cas9D10A nickase was tested. The digestion pattern of *Ban*II and *Nla*III was the same as the wild type, with no additional bands visible (Figure 4.7B, lanes 8-13).

Overall, it was concluded that guide 1-2 is able to direct efficient cleavage by wt Cas9, whereas guide 4-4 is less effective. No mutagenesis was apparent when using the Cas9D10A nickase.

4.3.5 Immunofluorescent screening of transfected cells

The PCR-RFLP assay can only detect mutagenic events that directly affect the restriction enzyme recognition sites, so it is possible that other indels were undetected. As the ultimate aim is to create cell lines lacking HMGN2 expression, immunofluorescence was used as a highly sensitive method to screen the nickase-transfected cell population for individual cells lacking HMGN2 protein. Using this method, we estimated that 5-10% of cells that were negative for HMGN2 protein (Figure 4.8). This suggests that the nickase is functional, but the efficiency is too low for detection via the PCR-RFLP assay.

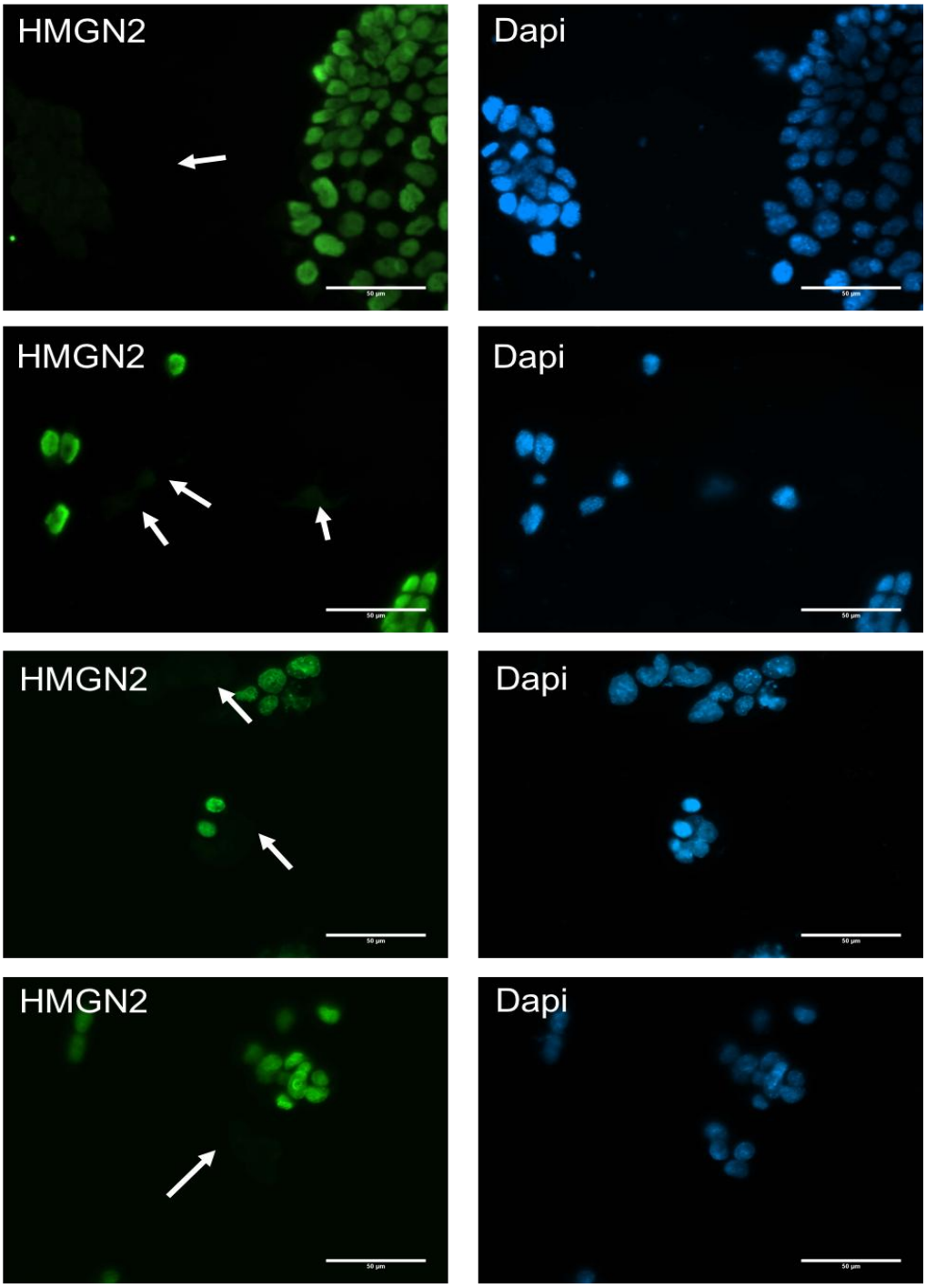


Figure 4.8: HMGN2-negative cells in polypopulation transfected with paired nickase targeting exon I of *Hmgn2* gene.
The mixed population of cells that were co-transfected by p.Guide001-2_hH1\ExI, p.Guide004-4h7SK\ExI and Cas9D10A, were stained with an antibody raised against a C-terminal peptide of HMGN2. HMGN2 is a nuclear protein (green signal) that co-localises with the Dapi staining (blue). The white arrows show the cells that do not have HMGN2 protein. The exposure time is 500 ms and 40X objective was used. Scale bar, 50 μM.

4.3.6 Screening for complete HMGN2 knocked-out lines

As the nickase approach was shown to be able to generate *Hmgn2*-knockout cells, albeit with low efficiency, the transfection was repeated, and FACS sorting was used to enrich for transfected cells. Individual cells were plated in the wells

of a 96 wells plates, to allow the growth of monoclonal lines. 86 lines were recovered and screened for a complete loss of HMGN2 protein by immunofluorescence. Immunofluorescence screening was used because it is less time consuming than western blotting for multiple samples, and provides a clear indication when there is complete loss of HMGN2 protein.

From the 86 lines, three lines were identified that seemed to have no HMGN2 protein: B8, B19 and B38 (Figure 4.9). Closer examination revealed that the B8 line has a small proportion of cells with a faint green signal, which is not visible in the other lines.

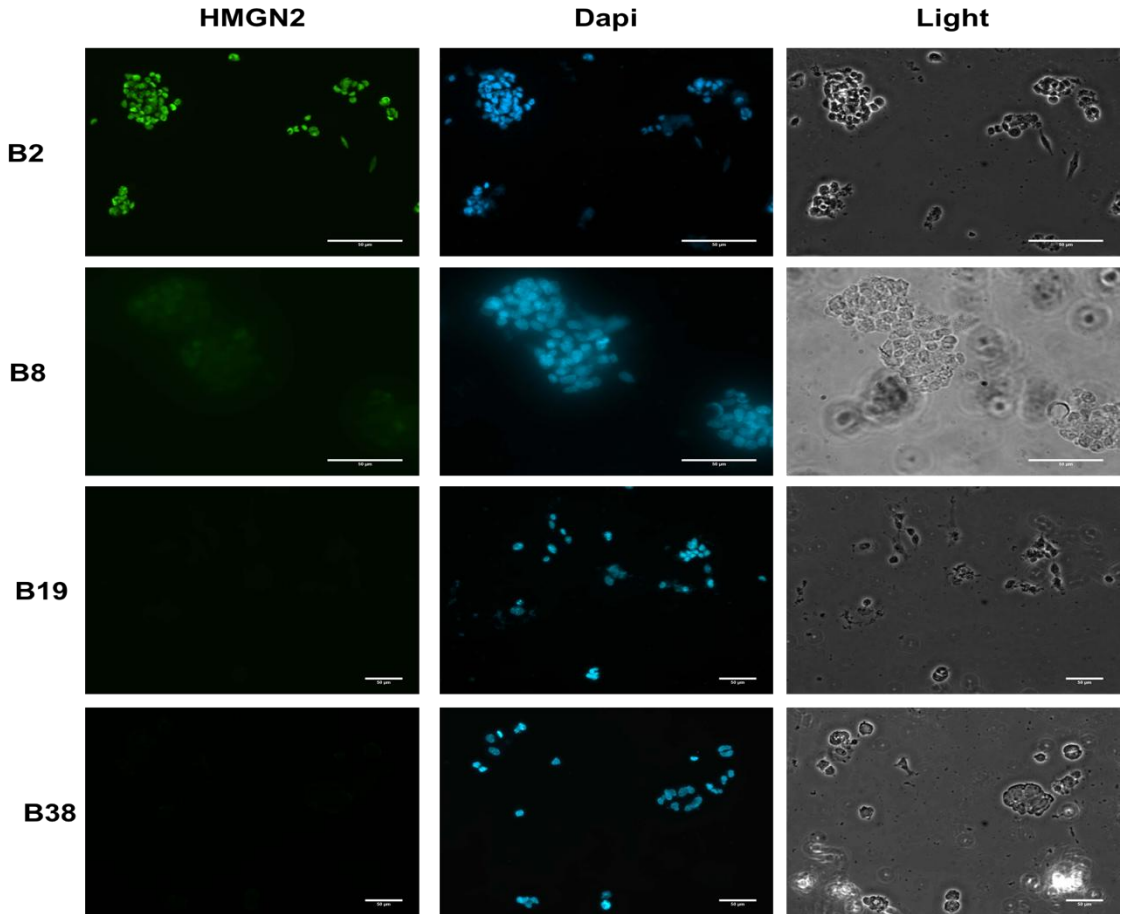


Figure 4.9: Some of screened monoclonal lines without HMGN2 protein from paired nickase targeting exon I of *Hmgn2* gene.
P19 cells were transfected by p.Guide001-2_hH1\ExI, p.Guide004-4_h7SK\ExI and Cas9D10A-GFP, were FACS sorted and plated in 96 well plates. The recovered lines were stained with anti-HMGN2. The line names are given on the right side. The exposure time is 1s, and 20X and 40X objectives were used. Scale bar, 50 μM.

4.4 Characterisation of monoclonal lines generated by CRISPR/Cas9 nickase targeting exon I of *Hmgn2* gene

4.4.1 Characterise mutant lines by western blotting

After identifying the potential lines to be used for further analysis, the loss of HMGN2 protein was also confirmed by western blot. In agreement with the immunofluorescence, B19 and B38 lines did not reveal any detectable bands corresponding to HMGN2 (Figure 4.10). B8 line showed a faint band at the size of the full length protein. This band is consistent with the presence of a few cells with faint signal that was observed in the immunofluorescence staining. Lines B1, B12 and B22 were also studied, in order to identify lines that could be used as negative controls. These lines did not show changes in HMGN2 protein levels.

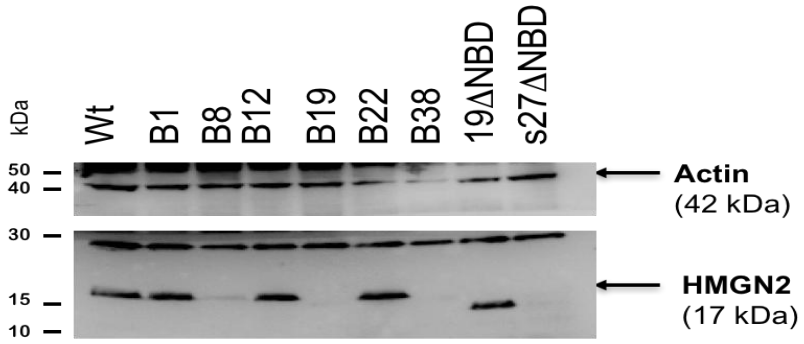


Figure 4.10: Western blotting conforming the complete loss of HMGN2 in the candidate lines from both CRISPRs.
B1, B8, B12, B19, B22 and B38 are different lines screened from CRISPR-cas9 that targets exon I of *Hmgn2* while 19ΔNBD and S27ΔNBD (used here as controls) are from CRISPR-cas9 that target exon IV. The same blot was cut and used against HMGN2 and β-Actin (housekeeping) antibodies. The name of each line was given on the top. Wt stands for wild type. The slower-migrating band in the HMGN2 blot is a non-specific band that is present in all samples. ΔNBD refers to lines with a deleted or disrupted nucleosome binding domain.

4.4.2 Characterise the mutant lines by RFLP

PCR-RFLP was used to assay genomic DNA purified from the monoclonal lines generated above. Lines B1 and B22, which have wt levels of HMGN2 protein, show the same digestion pattern as wt cells (Figure 4.11A, lanes 1-3, 4-6 and 16-18). This is the same pattern previously observed in Figure 4.7 (section 4.3.4). In contrast, line B12, which also shows wt levels of HMGN2 protein, has additional bands of approximately 600 and 700 bp following *Nla*III digestion (Figure 4.11A lane 12). In addition, the uncut PCR product from B12 cells shows two distinct bands (Figure 4.11A lane 10). It is possible that line B12 is not monoclonal, but

instead contains both wt and mutant cells. This was confirmed by sequencing of cloned PCR products that showed two mutant alleles and wild-type allele (data not shown) and immunofluorescent staining that showed positive and negative cells (data not shown). Thus, lines B1 and B22 appear to be suitable as negative controls, whereas B12 is not.

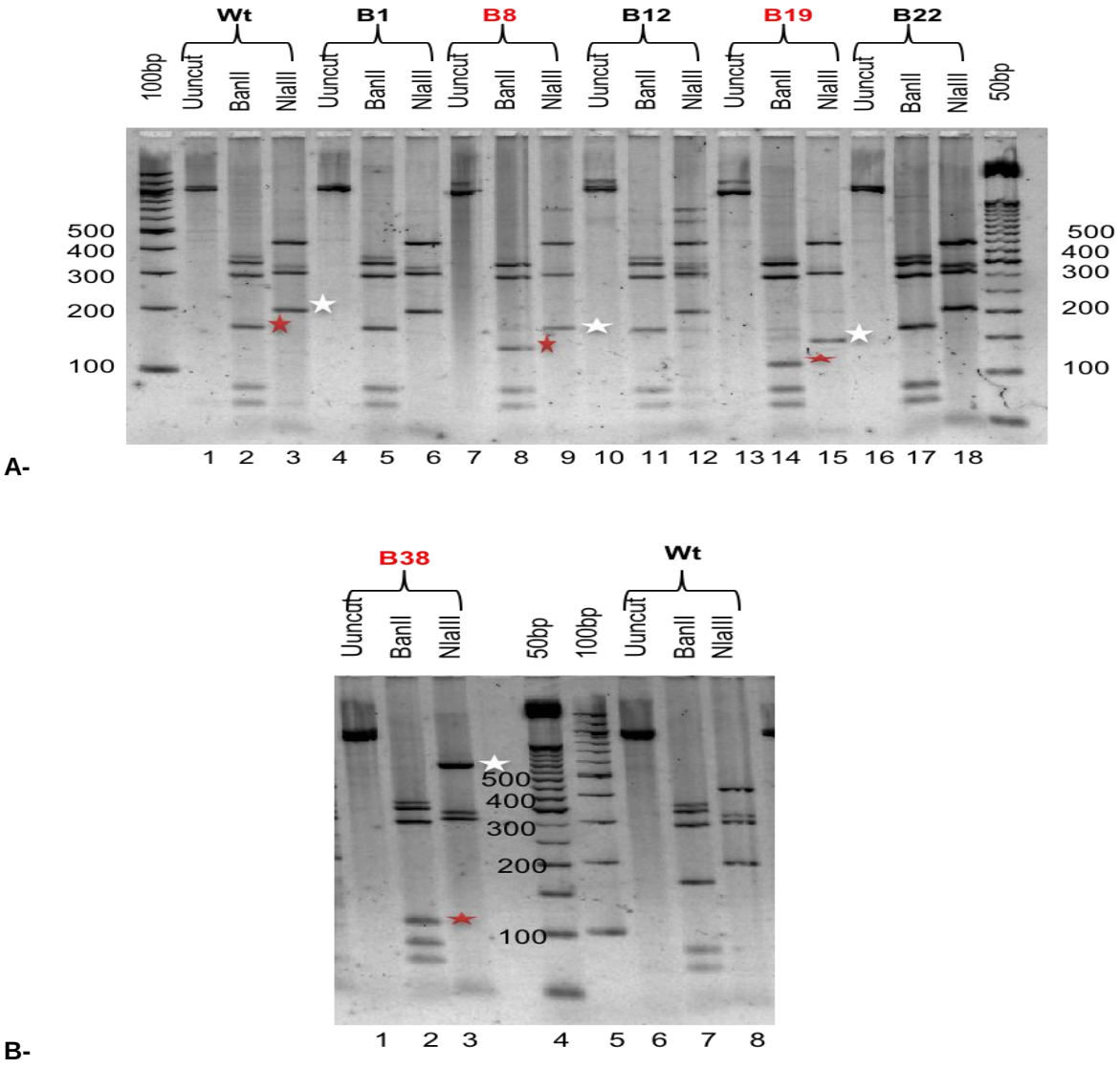


Figure 4.11: PCR-RFLP assay for monoclonal lines generated from CRISPR targeting exon I of *Hmgn2*.
The DNA from each sample was amplified using primer set (mN2ATG_PCR4_forward and mN2ATG_PCR1_reverse) that yields a product with 977 bp. When the wild type DNA was digested by *Ban*II or *Nla*III, the products should be digested into (37, 65, 81, 146, 294 and 336 bp) and (43, 200, 302 and 432 bp), respectively. As discussed previously, an unexpected extra faint band appears in the digestion of wild type DNA if it was either digested by *Ban*II (~350 bp, lane 2-A or 7-B) or *Nla*III (~330 bp, lane 3-A or 8-B). If the recognition site of *Ban*II is disrupted, the two bands (164 and 336 bp) are predicted to become around 500 bp. If the recognition site of *Nla*III is disrupted, the two bands (200 and 432 bp) are predicted to be converted to a single band of around 632 bp. In B8 and B19 lines, the mutations in both alleles do not affect the recognition sites for *Ban*II and *Nla*III but they change the sizes of bands 164 and 200 in *Ban*II and *Nla*III digestions, respectively (compare the red and white stars in A to corresponding stars in wild type in same gel). The mutation in B38 does not affect the *Ban*II recognition site but only gives a smaller band of around 120 bp instead of 146 bp. The *Nla*III site is disrupted in B38, with a larger band of ~600 bp. 2.2% agarose gel was used for electrophoresis.

Line B8, which has very low levels of HMGN2 protein expression, yielded two distinct uncut PCR products: a faint one at 977 bp and a second band that was a little shorter and stronger (Figure 4.11A, lane 7). Unexpectedly, the mutation in B8 did not alter the *Ban*II digestion by converting two bands (164 and 336 bp) into 537 bp; rather, the 164 bp band became shorter, to around 140 bp (Figure 4.11A, lane 8). Similarly, the 200 bp fragment from the *Nla*III digestion was truncated to around 170 bp (Figure 4.11A, lane 9). Thus, at least one allele of the B8 line appears to have a deletion that does not affect the *Ban*II or *Nla*III recognition sites. There is also a weak band at 700 bp of the *NlaIII* digestion which may be another mutant allele. Interpretation of the RFLP pattern is complicated by the fact that the two uncut PCR products have different intensities, possibly due to differences in amplification efficiency.

The other two lines that lost the HMGN2 protein completely were B19 and B38. The RFLP assay for the B19 line did not show disruption of the *Ban*II and *Nla*III sites. Instead, the 164 bp band from the *Ban*II digestion is truncated to around 120 bp, and the 200 bp band from the *Nla*III was truncated to around 150 bp, indicating a deletion between these two sites (Figure 4.11A, lanes 14-15).

The *Ban*II digestion for B38 line also revealed a truncation of the 164 bp band to around 120 bp (Figure 4.11B, lane 2), indicating deletion of some sequence. The *Nla*III site also appears to have been deleted in this line, as shown by the loss of the 200 bp and 432 bp bands, and the appearance of a new band at around 600 bp (Figure 4.11B, lane 3).

4.4.3 Determine the exact mutation in each monoclonal line by sequencing

The exact nature of the mutations was investigated by sequencing of multiple cloned PCR products. Firstly, the region around exon I of *Hmgn2* where the CRISPR-Cas9D10A targets was sequenced in B1 and B22 lines, and all the sequences from either B1 and B22 revealed only wild type allele. On the other hand, seven clones from the B8 line were sequenced, and all sequences corresponded to one allele with a deletion of 30 bp ($Hmgn2^{del30}$). This mutation deletes most of the coding region of exon I and the splicing donor site, but the

ATG is intact (Figure 4.12A). The *Nla*III and *Ban*II sites are not disrupted, confirming the results of the PCR-RFLP assay.

In the B19 line, four sequences corresponded to an allele with a deletion of 53 bp (*Hmgn2*^{del53}) (Figure 4.12B). Again, most of the coding region of exon I is deleted, but the *Nla*III and *Ban*II sites are intact. All the evidences (from protein and the RFLP and sequencing of DNA) endorse that the mutation in B8 and B19 lines are biallelic (homozygote) which means both alleles have the same mutation. However, because of having faint band in B8 western blotting and some cells with faint expression of HMGN2 as shown in the immunofluorescence, there is bias in the PCR amplification that affect the accuracy of interpreting the result of RFLP assay.

The sequencing (four sequences) of B38 DNA disclosed two different mutations which are a deletion of 59 bp in one allele and a deletion of 62 bp in the another (Figure 4.12C). The genotype for this line is Hmgn2$^{del59/del62}$. The start codon (ATG) and the associated *Nla*III site is deleted in both alleles, but the *Ban*II site is intact, in agreement with the PCR-RFLP assay. Therefore, the nature of mutations in different lines results in the loss of most exon I of *Hmgn2*.

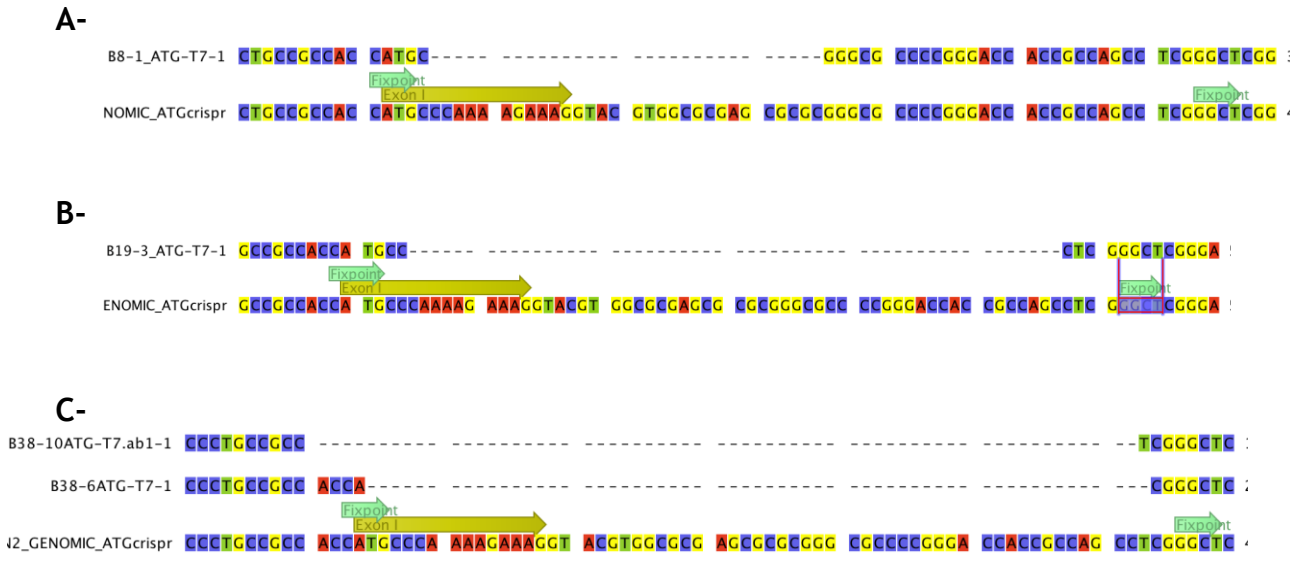


Figure 4.12: Screenshot from CLC program revealing the sequences of mutant alleles for knockout lines.
The yellow arrows represent the sequences of exon I of *Hmgn2* while the turquoise arrows are for *Nla*III and *Ban*II that interfere with gRNAs, mNATG_A69 and mN2ATG_B77, respectively. The turquois arrow that overlaps with exon is recognition site for *Nla*III whereas the one downstream is for *Ban*II. PCR products were cloned and sequenced and aligned against the reference sequence of mouse *Hmgn2*. (**A**) All seven clones for B8 had the same deletion of 30 bp. (**B**) All four clones for B19 had the same deletion 53 bp. (**C**) Clones from B38 line displayed two different alleles: three clones had a deletion of 59 bp, while another had a deletion of 62 bp.

4.4.4 Characterise the *Hmgn2* mRNA expression in mutant lines

We also examined the mRNA expression of *Hmgn2* in the key lines obtained from CRISPR targeting either exon IV or I. The two negative control monoclonal lines (B1 and B22) for HMGN2 knockout, which were shown earlier to have a similar level of HMGN2 protein and a similar RFLP pattern as the parental cells, are also included. The *Hmgn2* mRNA expression in the negative control lines, B1 and B22, were up-regulated by 40% and 70% compared to the parental cells, although these increases were not statistically significant (Figure 4.13).

On the other hand, the B8ΔHMGN2, B19ΔHMGN2 and B38ΔHMGN2 lines had a 50-80% reduction of Hmgn2 mRNA (p-values of 0.016, 0.0001 and 7.7E-5, respectively) (Figure 4.13). It is not clear why the mRNA was reduced in these lines, as the mutations introduced by the CRISPR process do not overlap the qRT-PCR amplicon. The primer set amplifies exon V, and the forward primer crosses the exon IV-V boundary. It is likely that splicing of the *Hmgn2* mRNA was disrupted in these lines, as the mutations remove the exon I – intron I boundary, but it is not clear what splice products would be formed. It is possible that the new RNA product was less stable than the wild type, or that transcription was slower due to disrupted splicing patterns. It is worth noting that there is an ATG in the first intron, but it is in a different reading frame to the HMGN2 exons, so any peptides initiating here would be unrelated to HMGN2.

The 19ΔNBD and S27ΔNBD lines did not show any expression of Hmgn2 mRNA even though the mutant proteins were still translated in both lines. This is because the exon IV mutations disrupt the binding site for the forward primer. Therefore, the reduction in Hmgn2 qRT-PCR signal in these two lines does not reflect the expression of Hmgn2; however, another primer set designed against exon I and III showed a 50% reduction in expression of mutant Hmgn2 mRNA from these lines (not shown). This reduction could reflect a decrease in Hmgn2 mRNA transcription or stability as result of the mutations.

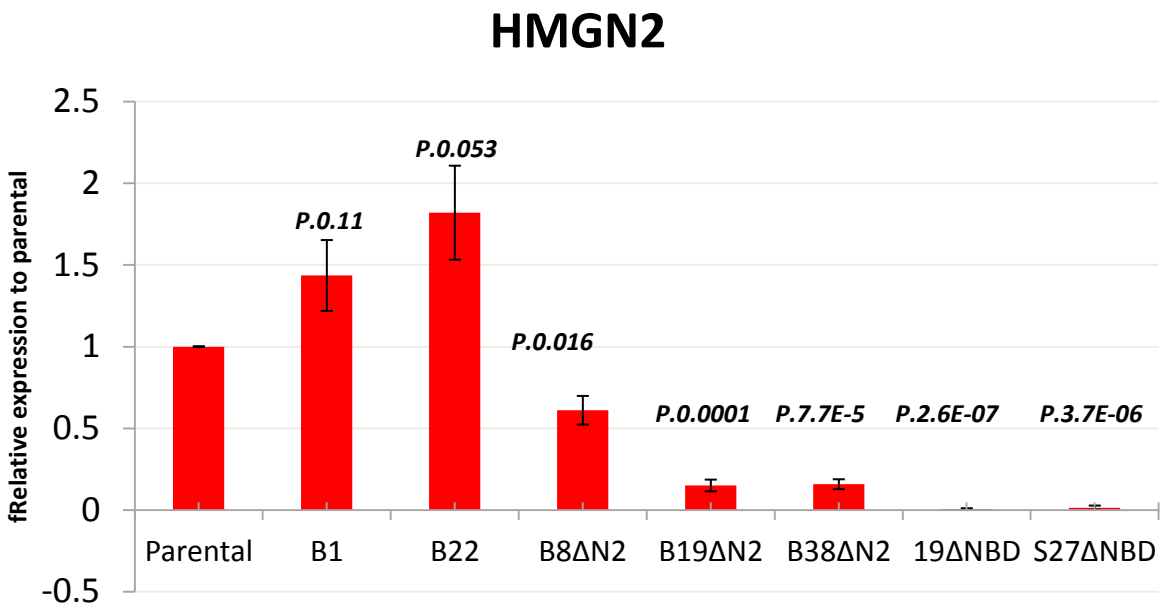


Figure 4.13: Hmgn2 mRNA expression in parental and different lines (negative control, null HMGN2 and disrupted NBD) derived from different CRISPR experiments.
"Parental" refers to the P19 wild type cells, while the B1 and B22 lines are negative control lines that went through the CRISPR transfection and clonal selection process. Lines with null HMGN2 (ΔN2) and lines with disrupted NBD (ΔNBD) were also included. B8ΔN2, B19ΔN2 and B38ΔN2 lines are derived from CRISPR against exon I while 19ΔNBD and S27ΔNBD are lines obtained from CRISPR against exon IV. The primer set was designed across an intron–exon junction and amplifies cDNA containing exon V. Relative gene expression in each cell line is the average from three biological replicates. Data was normalised to three housekeeping genes (*Gpi1*, *Rplp0* and *Gapdh*) and is plotted relative to that in parental cells (see methods for calculation). Error bars represent the standard deviation of the three biological replicate cultures for each line. The P values are shown on the top of each bar and were obtained from a paired t-test with a two-tailed test samples (equal variances), and the $P < 0.05$ is considered statistically significant.

4.5 Discussion

Many studies used the 5` exon or the ATG start codon as a targeting site for CRISPR-Cas9, as a frameshift here can cause a complete loss-of-function of the protein (Hruscha et al., 2013; Jiang et al., 2013b; Shalem et al., 2014). This strategy of targeting exon I near the first ATG was shown to generate a complete loss of HMGN2 protein in three lines, B8, B19 and B38. The first two lines were shown to have homozygous deletions in both alleles while the B38 has heterozygous deletions in the two alleles. All deletions introduce frameshifts in the *Hmgn2* coding region immediately before or after the ATG start codon, so no HMGN2-derived peptides can be produced.

Combining genotypic and phenotypic screens for functional knockout clones is the most effective way of screening (Chiang et al., 2016). Because our priority was to identify the biallelic functional knockout, we initially used phenotypic

screening either by western immunoblotting or immunofluorescence. Both western immunoblotting or immunofluorescence are equally effective in identifying the putative lines. The use of western immunoblotting is more efficient for screening the NBD knockout lines due to the ability of identifying truncated mutant proteins which cannot be detected by immunofluorescence. However, immunofluorescence can identify a mosaic cell culture more effectively. In most cases, the phenotypic screen was confirmed by genotypic screening of putative knockout clones using PCR-RFLP and sequencing to ensure both alleles were mutated.

Assessing autosomal genes by RFLP assay can sometimes be challenging, as there are two copies and CRISPR mediated mutation is sometimes heterozygous (Li et al., 2016). In some cases, it was difficult to interpret all findings in the RFLP assay because many additional factors including the presence of two bands (alleles) in the wild type PCR product with different intensities, unequal amplification for all alleles, and the presence of a small proportion of HMGN2-positive cells. An alternative approach for evaluating the cleavage activities of CRISPR–Cas9 at a target sequence involves enzymes that recognise and cleave at DNA mismatches – T7 endonuclease I and Surveyor Nuclease (Vouillot et al., 2015). The advantage of these enzymes is that they do not require the presence of a restriction enzyme site near the CRISPR target sequence. However, these enzymatic assays are also dependent on PCR amplification, and experience in this lab suggests they can be difficult to set up. The results can also be affected by the presence of heterozygous mutations or SNP sites (Li et al., 2016).

Co-transfection of multiple plasmids including Cas9D10A-GFP and two gRNA expression plasmids, can compromise the transfection and targeting efficiency in cells that are hard to transfect (Sakuma et al., 2014). However, the co-transfection of three plasmids in P19 cells avoided the cell toxicity that observed with the use of multiguide plasmids in the previous chapter. In addition, the transfection efficiency was higher with three individual plasmids. With transfection efficiencies of only 45-55%, FACS sorting was important to enrich for Cas9D10A-GFP - transfected cells. Addition of reporter genes to the gRNA expression plasmids would enable more accurate determination of transfection

efficiency and might improve FACS sorting efficiency, thus increasing the proportion of clones with mutations.

The frequency of introducing mutations at the target site are varied between gRNA sequences, and was not correlated with transfection efficiency. Specifically, the transfection efficiencies for wt Cas9 with p.Guide 1-2 or p.Guide 4-4 were 45% and 52% respectively, whereas the mutation frequencies estimated by RFLP were around 50% and <10% respectively. It was interesting that all of the HMGN2 knockout mutations introduced by the Cas9D10A nickase involved a deletion between the two gRNAs, and in most cases these deletions did not affect the restriction enzyme sites. This could explain why mutations were undetectable by RFLP in the original transfected population, yet 5-10% of cells were HMGN2-knockout as estimated by immunofluorescence. Use of a restriction enzyme that cuts in between the two nickase sites, such as *Sma*I in this case, could give a more sensitive RFLP assay for nickase mutagenesis. Previous studies have suggested that nickase-induced mutagenesis is as effective as wt Cas9, with indel efficiencies of 14-91% (Chiang et al., 2016, Cho et al., 2014).

Chapter 5 Analysis of HMGN2 mutant lines in the undifferentiated EC state

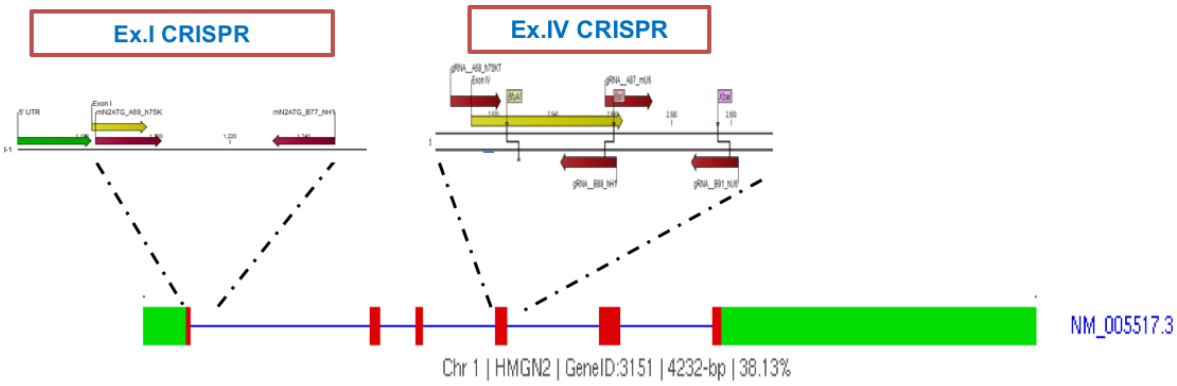
5.1 Introduction

To gain insight into the role of HMGN2 in undifferentiated P19 ECCs, we established different knockout lines using the CRISPR-Cas9 double nickase system as previously described in Chapter 3 and 4. Briefly, two CRISPR-Cas9D10A approaches were designed: one targeting exon I while the another is targeted against exon IV (Figure 5.1A). In total, 5 monoclonal lines were obtained from these approaches that were selected for further study. Line 19 has a deletion of exon IV in both alleles, and expresses protein HMGN2ΔNBD, which is lacking the second half of the NBD. Line S27 has a partial deletion of exon IV and a frameshift, resulting in protein HMGN2ΔNBD_CHUD that lacks both the last seven amino acids of the NBD, and the C-terminal chromatin unfolding domain (Figure 5.1). From here onwards, the lines are referred to as 19ΔNBD and S27ΔNBD for clarity.

The other lines (B8, B19 and B38) from CRISPR-Cas9D10A against the first exon of *Hmgn2* lead to complete removal of HMGN2 protein expression (Figure 5.1), although the B8 has faint expression in some cells in the culture as shown in western blot and immunofluorescent staining. These line were renamed as B8ΔHMGN2, B19ΔHMGN2 and B38ΔHMGN2.

This chapter described the analysis of the HMGN2 knockout and mutant cell lines in their undifferentiated state.

A-



B-

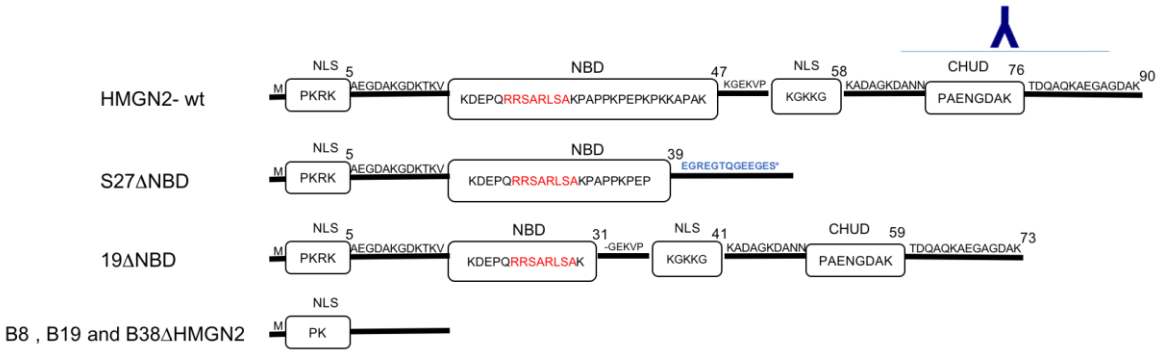


Figure 5.1: Schematic diagrams for *Hmgn2* gene and the loci for CRISP-Cas9D10A (A), and protein structure with amino acid sequences (B) in the wild type and mutant lines.
A- The *Hmgn2* gene that is located in chromosome 1 includes 6 exons (red boxes), and the first and last exon have 5' and 3' untranslated regions (green boxes). The lines between the boxes represent the introns. The two regions magnified at exon I and IV correspond to the regions where the CRISPR-Cas9D10A targets. **B**- Outlines of the HMGN2 deletion mutants in different lines obtained from CRISPR-Cas9D10A exon I or IV. The HMGN2 protein contains three domains which are two nuclear localisation signal (NLS), nuclear binding domain (NBD) and chromatin unfolding domain (CHUD). The amino acid sequence is given inside the boxes for the domains or on the lines. The numbers on the top correspond to the last amino acid in each domain or in protein. Within the NBD, there is conserved motif (red sequences) that confers the binding to nucleosomes. The blue amino acids sequences in S27ΔNBD line refer the frame-shift sequences and the asterisk refers to the first possible stop codon. S27ΔNBD and 19ΔNBD lines are obtained from CRISPR-Cas9D10A_ExIV while B8ΔHMGN2, B19ΔHMGN2 and B38ΔHMGN2 lines generated from CRISPR-Cas9D10A targeting exon I. B8ΔHMGN2 and B19ΔHMGN2 and B38ΔHMGN2 lost the splice donor of exon I and did not reveal HMGN2 protein in western blot (by antibody against the C-terminal, blue line and inverted Y) and their mRNAs were significantly reduced (primer set against exon IV and V).

5.2 Objectives

The main purpose of this chapter is to study the effect of knocking-out or mutating HMGN2 in undifferentiated ECCs (P19 cells):

1- Is HMGN1 upregulated in HMGN2 mutant or knockout cells to compensate for the loss of HMGN2?

2- Does the knockout or mutation of HMGN2 affect the proliferation of cell cycle distribution of P19 EC cells?

3- Does the knockout or mutant HMGN2 affect the expression of pluripotency genes in the undifferentiated P19 ECs?

4- Does the knockout or mutant HMGN2 affect certain histone modifications, either at a global level, or at specific genomic locations?

5.3 Does HMGN1 companste the loss of HMGN2 at mRNA and protein?

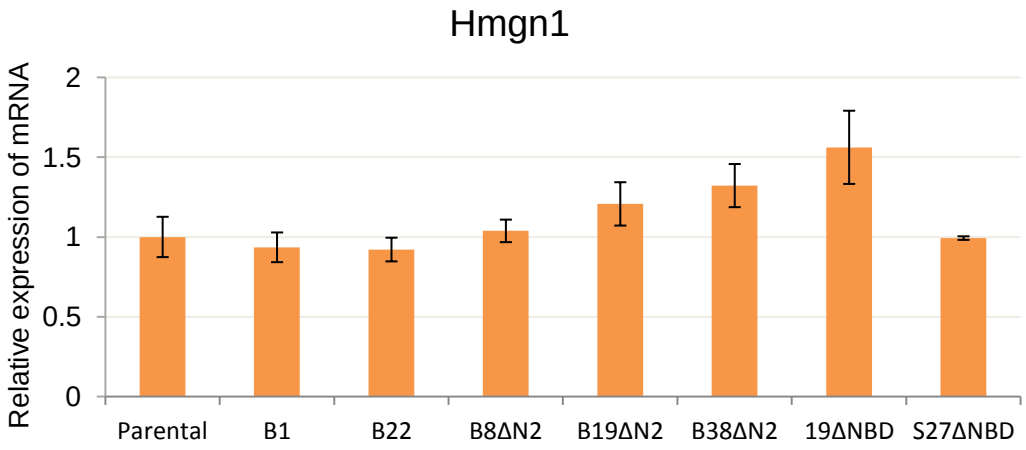
The HMGN1 and 2 are the most abundant members of the HMGN family in vertebrate cells, especially in stem cells (Birger et al., 2001). Previous studies have suggested there is likely to be some redundancy between these proteins, with each of them being able to functionally compensate for the loss of the other member to some extent (Deng et al., 2013; Deng et al., 2015). In order to investigate whether HMGN1 expression was altered in HMGN2 mutant cells, *Hmgn1* mRNA and protein levels were studied (Figure 5.3A). The level of *Hmgn1* mRNA was unchanged in the negative control lines B1 and B22, and in mutant lines B8ΔHMGN2 and S27ΔNBD. In lines B19ΔHMGN2, B38ΔHMGN2 and 19ΔNBD, there was a trend of up to 50% higher *Hmgn1* expression, but this was not statistically significant (Figure 5.2A).

However, the western blot for HMGN1 from whole cells extracts showed a possible increase in HMGN1 protein in all lines, but more obvious in B38ΔHMGN2, 19ΔNBD and S27ΔNBD lines that revealed more than double the amount of HMGN1 (Figure 5.2B). However, because this immunoblotting was done once and western assay is semi-quantitative, it might be required to have more replicates

before draw a conclusion, but from the RNA data which is more accurate there are no apparent differences in HMGN1 expression in all lines.

Similarly, Furusawa et al (2006) did not observe any change in the mRNA and protein of HMGN2 in *Hmgn1*$^{-/-}$ mice. Previous study using siRNA to knockdown either HMGN1 or HMGN2 in undifferentiated P19 cells suggested that there is no compensation between HMGN1 and 2 at the protein level when the another isoform was knocked-down (unpublished data from our lab).

A-



B-

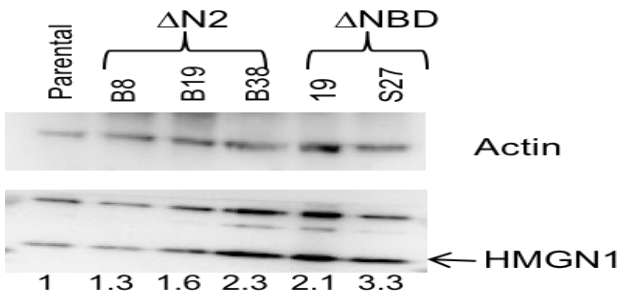


Figure 5.2: The mRNA and the protein level of Hmgn1 in the ΔHMGN2 and ΔNBD lines.
A- The *Hmgn1* expressions in parental cells (wild type), B1 and B22 lines (negative monoclonal selection), B8, B19 and B38 (ΔHMGN2), and 19 and S27 lines (ΔNBD, disrupted NBD). Relative gene expression in each cell line is the average from three biological replicates. Data was normalised to three housekeeping genes (*Gpi1*, *Rplp0* and *Gapdh*) and is plotted relative to that in parental cells (see methods for calculation). Error bars represent the standard deviation of the three biological replicate cultures for each line. Student T-test was done for all sample relative to parental cells but because none of them showed to be significant, they were not included in graph. p-values for B1, B22, B8, B19, B38, 19ΔNBD and S27ΔNBD are 0.64, 0.75, 0.27, 0.12, 0.09 and 0.94, respectively. **B-** The whole cell extracts from different lines were subjected to Western blot using antibody against the HMGN1 (11 kDa). The non-specific band in the same blot was used as internal control rather than β-Actin that showed unclear bands. Quantifying the bands (numbers at the bottom) was performed using the intensity of each band relatively to parental after normalising each sample to the non-specific band.

In addition to protein extracts, we also investigate if there is a compensation between most abundant HMGN protein from the chromatin extracts. The fixed chromatin samples from different lines were sonicated to shear chromatin into smaller fragments ranged between 100-3000 bp (Figure 5.3). Before western blotting of chromatin samples, crosslinks were reversed and nucleic acids removed by benzonase digestion, so that proteins were free to migrate according to their size and charge.

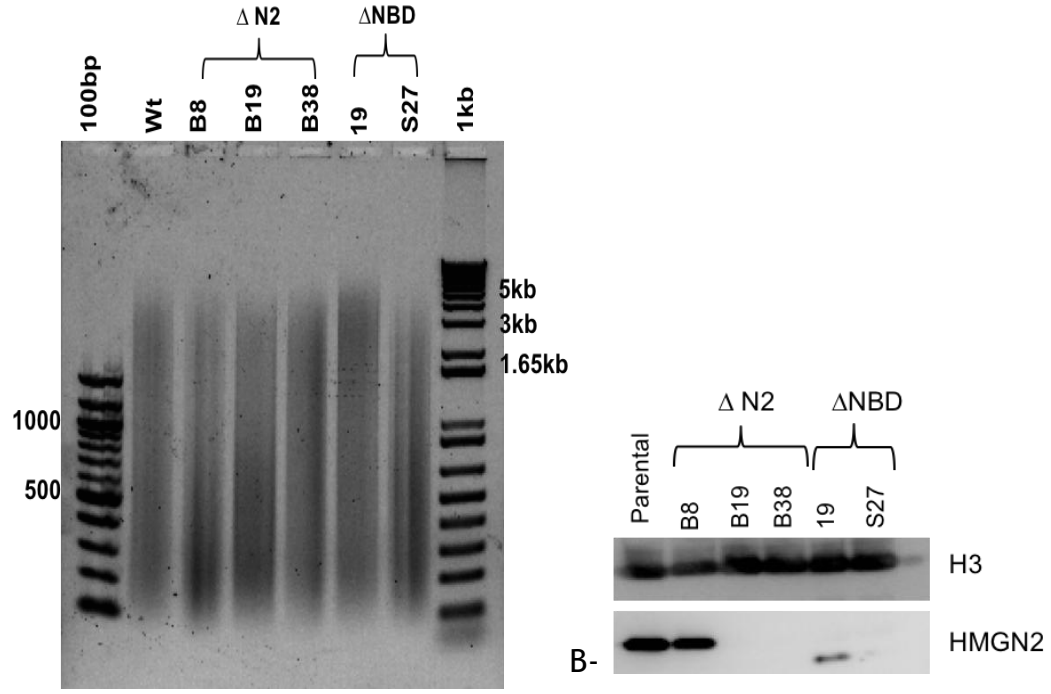


Figure 5.3: Shearing of chromatin lysates from different lines and verifying the loss of HMGN2.
A- Chromatin samples (from 0.9-1.9X10^8) of undifferentiated P19 cells and different monoclonal lines were sheared using a sonicator. The wild type, B19, B38, 19 and S27 were sonicated for 8 minutes while the chromatin from the B8 sample was sonicated for 2.5 minutes. The aliquots (25 µl) of DNA from each sample were reverse cross-linked by incubating with 5M NaCl (4 µl), 20 ng/ml proteinase K (8 µl) and Rnase A (4 µl) at 65 ^{o}C for 1 hour. The DNA was purified with PCR clean-up kit (Qiagen) and 300 ng was run in 1% agarose gel for an hour. **B**- Western blot for wild type cells (parental), HMGN2 knockout lines (B8ΔN2, B19ΔN2 and B38ΔN2), and mutant-NBD of HMGN2 lines (19ΔNBD and S27ΔNBD), showing the level of HMGN2 (17 kDa) in the chromatin extracts of all lines. This was performed on the chromatin extractions and H3 (15 kDa) was used as loading control. The cross-linked chromatin samples were treated with Benzonase (see martials and methods) to digest the nucleic acids and reverse cross-linked by NaCl before the immunoblotting.

Loss of HMGN2 in the chromatin samples was confirmed by western blotting. The HMGN2 protein was completely removed from B19ΔHMGN2 and B38ΔHMGN2, while the B8ΔHMGN2 line showed nearly wild-type levels of HMGN2 protein

(Figure 5.3B). Therefore, B8ΔHMGN2 was excluded from the chromatin assays only. The lines that express mutant HMGN2 were also included. 19ΔNBD that has a mutant form of HMGN2 with half of the NBD removed showed a substantial reduction in HMGN2 level and probably its binding efficiency to the chromatin (Figure 5.3B). In contrast, the HMGN2 mutant form in S27ΔNBD line was not detected. This is because the mutation causes a frameshift that disrupts the second half of the NBD and the C-terminal domain, and the antibody used to detect HMGN2 recognises the C-terminal domain.

In terms of HMGN1, the same samples isolated from fixed chromatin displayed around 30 to 90% higher of HMGN1 level in all lines either HMGN2$^{-/-}$ or ΔNBD comparing to HMGN2$^{+/+}$ (Figure 5.2C). Again, this is a semi-quantitative assay for the total level of protine and the requirement of performing ChIP followed by more quantitative assays such as q-PCR or sequencing are crucial to determine if the HMGN1 bound to chromatin is quantitatively affected.

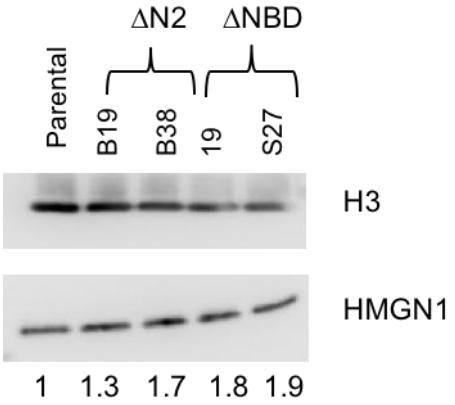


Figure 5.4: Western immunoblotting for HMGN1 on reverse cross-linked chromatin samples. The reverse cross-linked chromatin extractions were subjected to Western blot for HMGN1. Western blot for wild type cells (parental), HMGN2 knockout lines (B8ΔN2, B19ΔN2 and B38ΔN2), and mutant-NBD of HMGN2 lines (19ΔNBD and S27ΔNBD), showing the level of HMGN1 in chromatin extracts. The Histone H3 was used as loading control. The numbers at bottom of the blots represent the relative densities of each band to parental after normalizing to H3.

Several evidences indicate that a functional redundancy among HMGN proteins binding to the chromatin. For example, it was reported an enhancement of HMGN2 binding on the *Sox9* gene in limb bud cells obtained from *Hmgn1*$^{-/-}$ mice (Furusawa et al., 2006). In addition, a very recent study also concluded that loss of HMGN1 and 2 together in mouse embryonic fibroblasts (MEFs) modulates the DNase I hypersensitive (DHS) landscape, but the DHS are maintained when only

one variant is lost due to the compensatory binding by other HMGN isoforms (Deng et al., 2015). These observations and probably our result indicate the binding interplay between HMGN variants at the chromatin.

5.4 Knockout or mutation of HMGN2 does not affect the proliferation rate of P19 cells

It was found that the level of HMGN2 is higher in rapidly dividing cell culture (Hela cells) comparing to the slowly dividing cells (mouse liver cells) which may indicate a dependency on the differences in proliferation or dependency on different cell-types (Bustin et al., 1987). Lim et al, (2014) do not show difference in growth rate of HMGN1 knockout MEFs. However, it was shown that the removal or reduction of another isoform in this family (HMGN5), which is highly expressed in many cancer cells, can reduce the proliferation in some cancer cells (Chen et al., 2012; Weng et al., 2015). Although the effect of HMGN proteins on the proliferation is controversial, we compared the cell number between all lines because if there is a difference in the proliferation or cell death rate it will reflect on the cell number. Most of the kinetic studies of cell proliferation are based on the measurement of cell numbers in specific time intervals (Gollias et al., 2004).

The cell growth rate was assessed by counting the total number of viable cells two days after plating out 10^6 cells. All the lines had multiplied by approximately 10 fold in this time, with around 1 $X10^7$ cells present after 48 hours (Figure 5.5). Cell numbers in the HMGN2 knockout and mutant lines were not significantly different to those in the wt HMGN2 lines. Thus, it seems that loss or mutation of HMGN2 does not affect the net proliferation rate of mouse P19 EC cells.

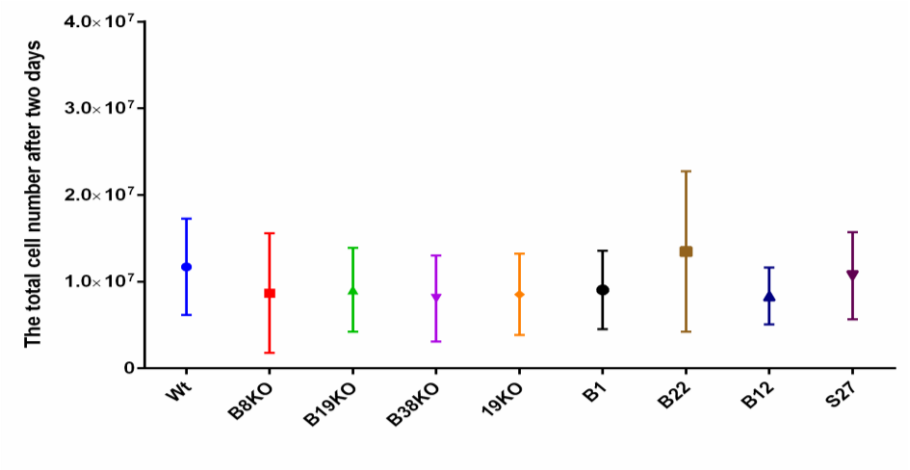


Figure 5.5: The cell growth rates of P19 monoclonal lines (ΔHMGN2, ΔNBD and negative control) and parental cells.
The cell growth rates were examined in wt (wt, B22 and B12), B19KO, B8KO, B38KO, 19ΔNBD and S27ΔNBD lines. 1 million cells from each line were seeded and live cells were counted after two days using Trypan blue. The mean of cell number and standard deviation from eight counts in three independent replicates are shown.

5.5 Knockout of HMGN2 does not change the cell cycle of P19 cells

Another approach to identify differences in cell division is to analyse the cell cycle distribution. It is well known that the cell division, cell death and cellular differentiation are linked with the entry and exit from the cell cycle. The cell cycle is a controlled process that refers to the replication and transmission of genetic material to daughter cells and it consists of four different stages including: two gap phases (G1 and G2), DNA synthesis phase (S-phase) and the cell division phase known as mitosis (M phase) (White and Dalton, 2005). HMGN variants have been implicated in regulation of DNA replication and cell division. HMGN2 mRNA and protein exist in the cell during the cell cycle with significant increase in the late S-phase in Hela cells but constant in Burkitt's lymphoma CA46 (Bustin et al., 1987; Morton et al., 1996). Although the HMGN proteins are mostly associated with chromatin, the mode of HMGN binding to the chromatin during the mitosis is cell cycle-dependent (Cherukuri et al., 2008). HMGN1 and 2 are displaced from the chromatin during mitosis, particularly in metaphase and early anaphase, as a result of negative charge caused by phosphorylating serines residues in NBD (Hock et al., 1998; Prymakowska-Bosak et al., 2001; Cherukuri et al., 2008). Therefore, we sought to investigate the effect of removal of HMGN2 on the cell cycle in all lines obtained from Chapter 3 and 4.

The cell cycle analysis was determined in $HMGN2^{+/+}$ and $HMGN2^{-/-}$ cells using flow cytometry and the Propidium Iodide (PI) to quantify the DNA content. The binding of this dye to DNA is directly proportional to the amount of DNA within the cells. PI is an intercalating dye that binds to DNA and double stranded RNA; therefore, it is always used in conjunction with Rnase A to remove RNA.

Since the analysis of cell cycle is based on the amount of DNA, it is crucial to eliminate cell doublets and debris from the analysis because they can interfere with the single cells in G2 phase. Firstly, the forward scatter (FSC) and side scatter (SSC) are plotted versus to each other and represented in 2-dimensional density plots to distinguish cell populations according to their granularity (SCC) and size (FSC) (da Silva et al., 2005). Figure 5.6A is an example of density blot with FSC versus SSC that was recorded for all events of the negative control B1 cell line ($HMGN2^{+/+}$). It shows one cell population, as did most of other cell lines (data not shown). Gate R1 (red) is drawn around the main bulk of the population, excluding cells at the top and bottom of the distribution.

In order to exclude doublets from the analysis, pulse area (BL2A) versus pulse width (BL2W) is plotted (Figure 5.4B). The BL2W signal can discriminate between a G1 doublet and a G2/M single cell because the diameter of the doublet is larger. In contrast, the BL2A signal is the same for both a G1 doublet and a G2/M single (Nunez, 2001). Gate R9 (dark blue) was drawn around the presumed single cell population, and gate R10 (green) was drawn around the presumed doublet population (Figure 5.4B).

Figure 5.4C is a FSC vs. SSC plot showing the bulk cells in gate R1 from graph A (red dots), compared to the presumed single cells in gate R9 from graph B (blue dots). It can be seen that the single cells in gate R9 overlap with bulk cells in gate R1, and indeed have a broader FSC vs. SSC distribution.

Figure 5.4D shows a plot of cell frequency vs. PI staining/DNA content. This type of graph enables the identification of the four distinct phases in a proliferating cell population: G0/G1, S, G2 and M-phase (Figure 5.6D, black line). The black line, B1, shows the total cell population, and the two sharp peaks represent the G1 and G2/M phases. Long tails on the right and left sides of these peaks represent debris and doublets that should be excluded from the analysis. The

red line, cells in gate R1, has smaller right hand tail. However, the blue line (cells in gate R9), has lost both the left and right hand tails. The green line (cells in gate R10), shows a large number of cells with higher DNA content than the main bulk of the population. This is consistent with these cells being doublets, and it is appropriate that they should be excluded from the cell cycle analysis. It was concluded that gate R9 does enrich for single cells, excluding debris and doublets, and so gate R9 was used for subsequent cell cycle analysis of the HMGN2 KO cell lines.

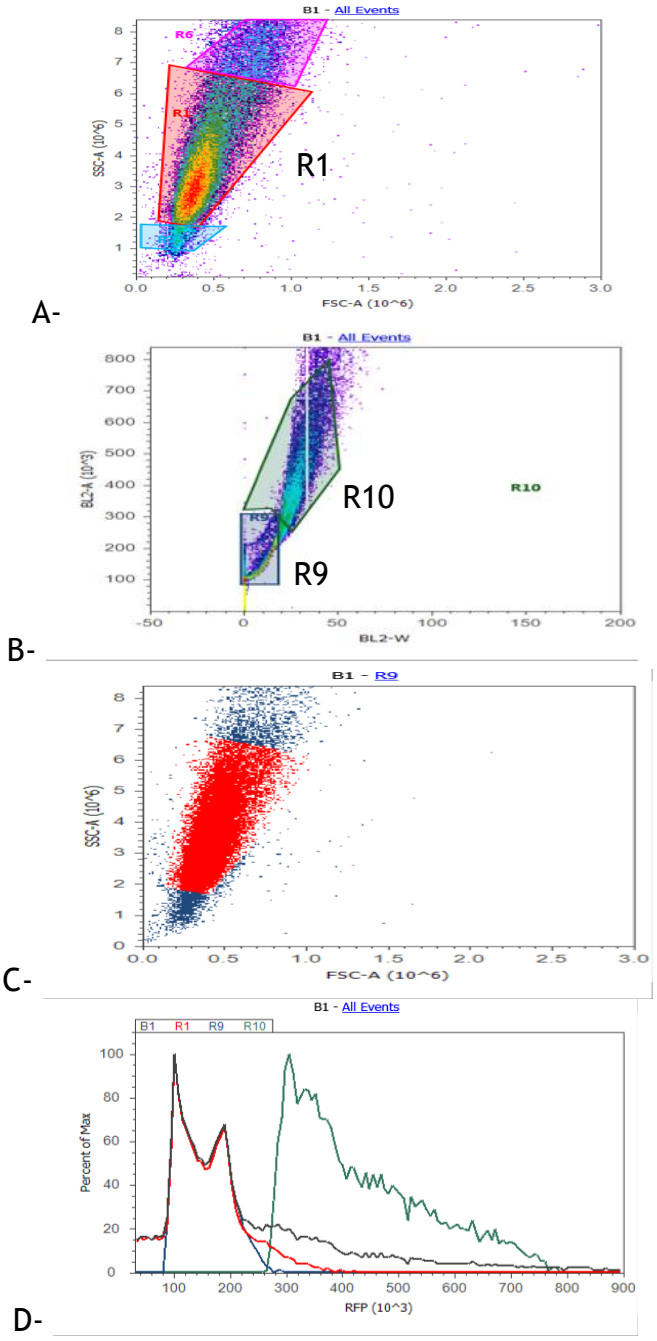
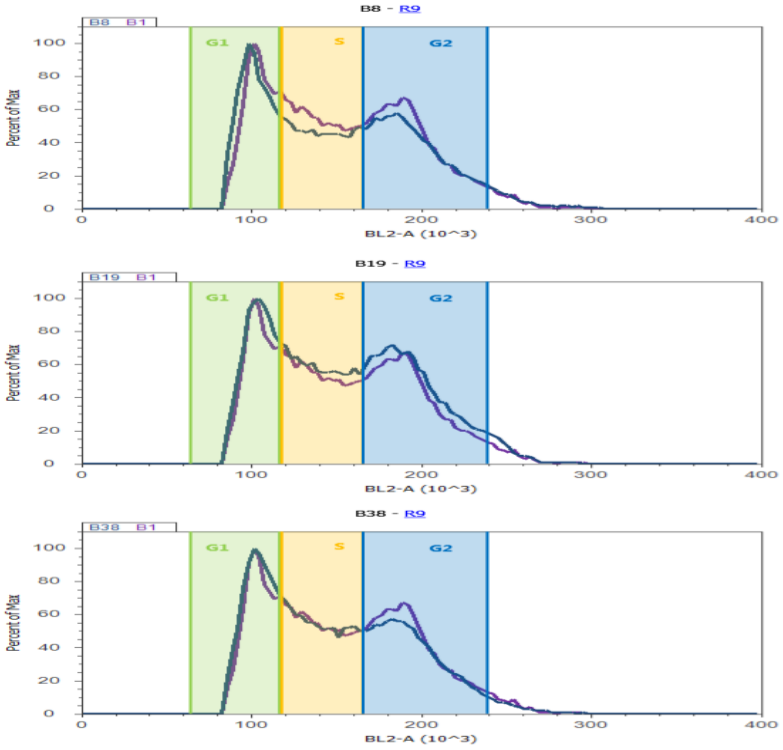


Figure 5.6: Setting up the gate for cell cycle analysis using the Attune® Acoustic Focusing Cytometer.
The B1 and B22 line were used as negative control for HMGN2 loss, and the dot blots and histograms for B1 lines were shown here as an example. The dot blot (**A**) is for forward scatter (FSC, size) versus side scatter (SSC, granularity) and showed most of the events for the B1 line with one population and few debris that are allocated in the bottom corner. The cells were gated into three gates including cells in R1 (red), R7 (light blue) and R6 (pink). Dot plot in **B** shows pulse width (BL2-W) versus area (BL2-A), and this is used to discriminate between single cells and aggregates. Single cells (G0/1 or G2/M) will have similar pulse width (transit time) values (gate R9) while the aggregates while have larger width values (gate R10). The cells gated in R1 (in **A**) and R9 (in **B**) were plotted together in new dot plot (**C**), and R1 is red dots while R9 is blue dots. (**D**) A one-parameter histogram of BL2-A (PI fluorescence, RFP) is plotted from different gates to compare between them, including all events (black), the single cells gated in (R9, dark blue), the doublets identified in plot **B** (gate R10, green) and cells gated in (R1, red). Y axis in the histogram represents the percentage of the maximum count whereas the X axis is for relative DNA content on the basis of staining with PI.

Cell cycle analysis was performed on the HMGN2 knockout lines B8ΔHMGN2, B19ΔHMGN2 and B38ΔHMGN2, and compared to the negative control B1 and B22 lines (Figure 5.7A). There are two ways of quantifying the percentage of cells in each cell cycle phase, including either by assuming and marking the fraction of each phase manually or by more complex deconvolution methods using curve fitting (Kaiser et al., 1982). Because the curve fitting method is not available in our flow cytometry software, phases were manually gated into three gates: G1, S and G2 (Figure 5.7A).

Comparing the pattern of cell cycle and the cells percentages in each phase do not reveal any substantial variation in all phases for B19ΔHMGN2 and B38ΔHMGN2 (Figure 5.7B). However, the number of cells in S-phase was significantly reduced in B8ΔHMGN2 compared to B1 and B22 (Figure 5.7B). Because this reduction only observed in 1 out 3 lines that have no HMGN2, it was concluded that this reduction is specific for this line and not associated with the removal of HMGN2.

A-



B-

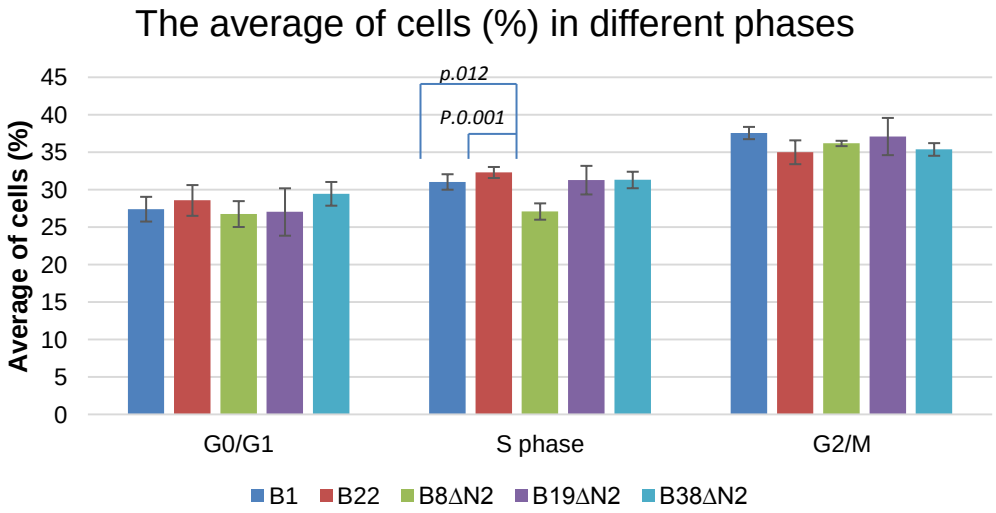


Figure 5.7: The cell cycle analysis for most monoclonal lines obtained from CRISPR targeting exon I.

A- The histogram plots for complete loss of HMGN2 lines [B8 (HMGN2 knockout in some cells), B19 and B38] compared with the negative control monoclonal line B1. BL2-A (PI fluorescence, RFP) is plotted from only single cells gate (R9, Figure 5.6B). Three further gates on the histogram, named as G1, S and G2, and refer to G0/1, S, and G2/M phase, respectively. Y axis in the histogram represents the percentage of the maximum count whereas the X axis is for relative DNA content on the basis of staining with PI. The percentages of cells in each phase were given by the Attune and plotted in (**B**) as an average of biological triplicates. Lines (B1 and B22 that are negative controls, and B8ΔHMGN2, B19ΔHMGN2 and B38ΔHMGN2) are included. The error bars represent the standard deviations of triplicate experiments. The P values of significant differences were given on the top of bar and were obtained from a paired t-test with a two-tailed test samples.

5.6 The effect of HMGN2 on the expression of pluripotent genes in ECCs

Since the HMGN can modulate the transcription, different studies have found that HMGN can positively or negatively regulate the expression of specific genes. For example, HMGN3a/b is bound and enhances the expression of glycine transporter 1 gene (*Glyt1*, its protein regulates the glycine concentration at synapses in the brain) whereas HMGN1 suppresses *Sox9* expression which is a master regulator of the chondrocyte linage (West et al., 2004; Furusawa et al., 2006; Barkess et al., 2012). Based on previous unpublished data from our lab (Mohan thesis, 2012), the expression of the main pluripotent genes (*Oct4*, *Nanog* and *Sox2*) were down-regulated by more than 80% in HMGN2 or HMGN1/2 knockdown undifferentiated P19 cells. This suggests that HMGN protein might regulate the expression of these genes. These genes are core transcription factors (TFs) that are responsible to sustain the pluripotent state of mouse ESCs (Hough et al., 2006).

We questioned whether the complete removal of HMGN2 can also decrease the expression of these genes. To investigate the differential expressions of pluripotent genes and other genes, total RNA was isolated by the RNeasy Mini Kit (Qiagen) from different samples and the RNA integrities were checked by agarose gel electrophoresis and spectrophotometer. Then, the expression of these pluripotent genes were evaluated by real time-quantitative PCR (RT-qPCR).

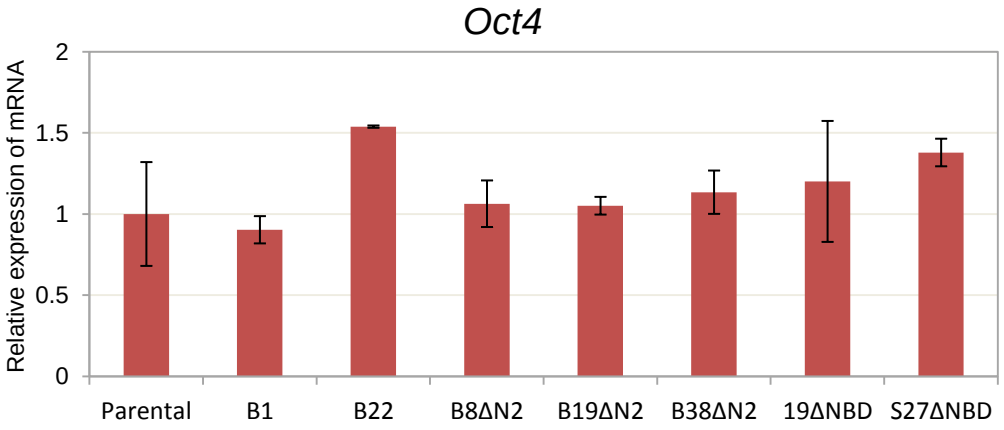
5.6.1 *Oct4* and *Sox2* expression are not regulated by HMGN2

The RT-qPCR results revealed that mutation or knockout of HMGN2 does not affect the mRNA expression of *Oct4* in any of the lines (Figure 5.6A). In contrast, levels of Oct4 protein appeared to increase by 2 - 5 fold in all chromatin extracts from mutant and knockout lines. This might indicate regulation at the post-transcriptional level, although more replicates of the western blotting on either whole cell extracts or chromatin are required to be confident about this result. Two of the cell lines had an upregulation of *Sox2* mRNA compared to the parental line – the negative control B22 and the HMGN2 knockout B38ΔHMGN2

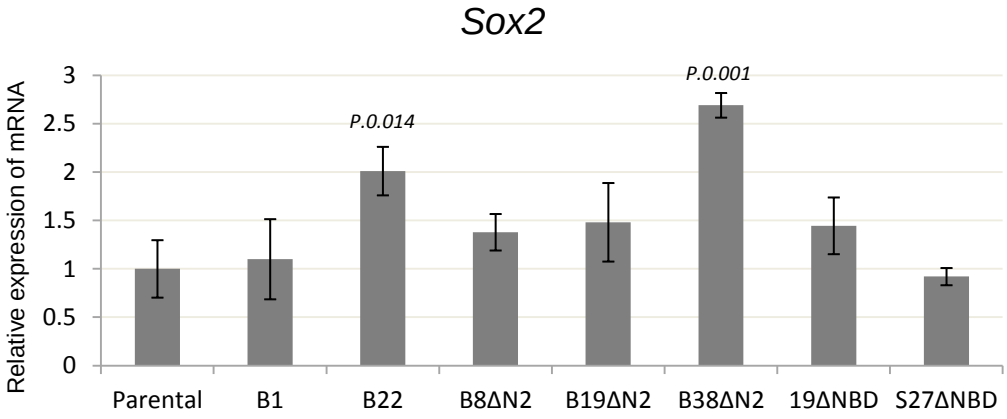
(Figure 5.8B). Due to the lack of consistency between cell lines, it was concluded that HMGN2 is unlikely to regulate *Sox2* expression.

In summary, the data from the knockout and mutant HMGN2 lines indicates that the expression of *Oct4* and *Sox2* are not regulated by HMGN2. This is not consistent with the data from the siRNA experiments, which showed that knockdown of HMGN1, HMGN2 or HMGN1/2 resulted in down-regulation of *Oct4* and *Sox2* (Mohan, 2012). One explanation for this discrepancy could be that the RNAi system had non-specific effects on these genes; however, this would not explain why the negative siRNA control did not have same effect. The other main difference between the two approaches is that the CRISPR-treated cells were clonally derived, and thus had to be able to survive for long period of culture. In contrast, the siRNA experiments were performed over a maximum of five days. It is possible that HMGN2$^{-/-}$ cells do have a reduction in *Oct4* expression initially, but that some cells adapt in order to survive during the monoclonal selection process, and this involves upregulation of *Oct4* through alternative mechanisms.

A-



B-



C-

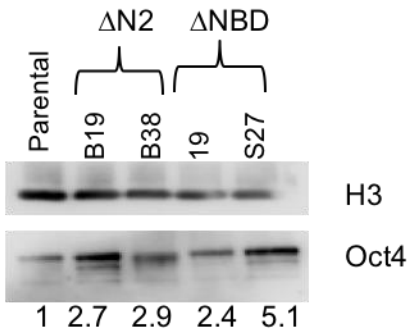


Figure 5.8: Expression of pluripotent genes, *Oct4* and *Sox2*, are not affected in null HMGN2 or disrupted NBD lines, while OCT4 protein increased in the HMGN2 knockout lines.
A- *Oct4* mRNA expression in parental cells (wild type), negative control lines (B1 and B22), in lines with no HMGN2 (B8, B19 and B38ΔHMGN2), and lines with disrupted NBD of HMGN2 (19 and S27ΔNBD). **B-** The level of *Sox2* mRNA was assayed in all lines mentioned in A. Relative gene expression in each cell line of *Oct4* or *Sox2* is the average from three biological replicates. Data was normalised to three housekeeping genes (*Gpi1*, *Rplp0* and *Gapdh*) and is plotted relative to that in parental cells (see methods for calculation). Error bars represent the standard deviation of the three biological replicate cultures for each line. The P-values were statistically analysed from 3 biological replicates compared to parental by Student's t-test. The T-test was performed for all samples but only the significant values were represented in the graph. **C-** Western blot for Oct4 (39 kDa) and H3 (15 kDa, loading control) were performed on the reverse cross-linked chromatin extractions. The densities that are given at the bottom of the plots were relative to parental after normalising to the internal control (H3).

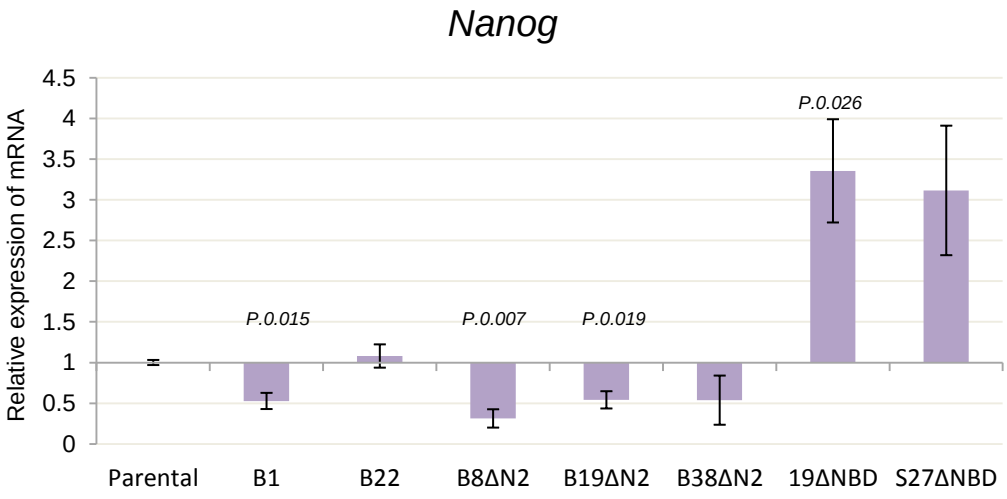
5.6.2 *Nanog* expression is differentially regulated in HMGN2 knockout and mutant cells

Expression of a third core transcription factor, *Nanog*, was significantly down-regulated in two of the HMGN2$^{-/-}$ lines, B8ΔHMGN2 and B19ΔHMGN2 (Figure 5.9). A reduction in *Nanog* expression was also observed in one of the negative control lines, B1, but not in the other negative control, B22. The level of *Nanog* protein from chromatin extracts was also reduced in lines B19ΔHMGN2 and B38ΔHMGN2 (Figure 5.9B). Western blotting of chromatin from the B1 line would assist to determine whether the reduction in *Nanog* protein was specific to the HMGN2 knockout cells.

In contrast, the two lines with the mutated NBD, 19ΔNBD and S27ΔNBD, revealed the opposite pattern. Specifically, the *Nanog* mRNA expression were up-regulated by 2.5 and 2 fold in the 19ΔNBD and S27ΔNBD lines, respectively, though only the increase in S27ΔNBD was statistically insignificant (Figure 5.9). Both 19ΔNBD and S27ΔNBD showed an increase in the level of chromatin-bound Nanog protein.

This data suggests that the knockout of HMGN2 has a different effect on *Nanog* expression than mutation of the NBD. The latter results in an increase in *Nanog* expression, whereas in cells lacking HMGN2, the *Nanog* expression is unchanged or reduced.

A-



B-

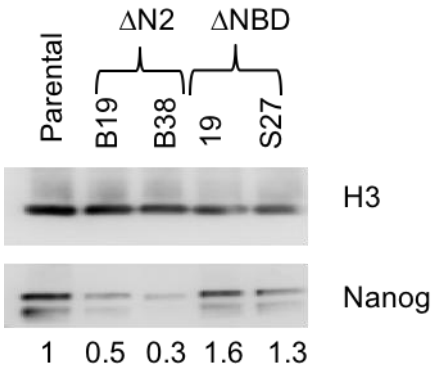


Figure 5.9: The *Nanog* mRNA expression and its protein level are down-regulated in HMGN2$^{-/-}$ lines while they are up-regulated in ΔNBD cell lines comparing to the parental cells.
A- *Nanog* mRNA expression was measured in parental (wild type), B1 and B22 lines (negative for loss HMGN2), and B8ΔN2, B19ΔN2, B38ΔN2, 19ΔNBD and S27ΔNBD lines. Relative gene expression in each cell line is the average from three biological replicates. Data was normalised to three housekeeping genes (*Gpi1*, *Rplp0* and *Gapdh*) and is plotted relative to that in parental cells (see methods for calculation). Error bars reflect the standard deviation from the fold changes of three biological replicates. The P-values were statistically analysed from 3 biological replicates compared to parental by Student's t-test. The T-test was performed for all samples but only the significant values were represented in the graph. P-values for B38ΔN2 and S27ΔNBD are 0.19 and 0.06, respectively. **B**- The protein level of Nanog was examined by western blot from chromatin extraction in parental, B19ΔN2, B38ΔN2, 19ΔNBD and S27ΔNBD lines. The expected band of Nanog is 35 kDa. The numbers at bottom of the blots represent the densities of each band relatively to parental after normalizing to H3 (loading control).

5.7 The consequences of HMGN2 loss or disruption on histone modification marks

The HMGN proteins have been shown to induce a less compact chromatin conformation and enhance the transcription from chromatin templates but not from naked DNA (Crippa et al., 1993; Ding et al., 1994: Trieschmann et al.,1995). Because the HMGN proteins affect transcription in the chromatin context, the interaction between the HMGN proteins and the nucleosomes is associated with their functions. Particularly, it was observed that the C-terminal

of HMGN proteins is localised near the N-terminal of histone H3 while the N-terminal interferes with the H2B (Kato et al., 2011). Various studies have revealed the influence of HMGN proteins on different histone modifications that are known to be critical for the chromatin structure and transcription. HMGN1, but not HMGN2, inhibits the phosphorylation of H3S10 and H3S28 mediated by RSK2 and MSK1(Ribosomal S6 kinase 2 and Mitogen and stress-activated kinase 1, respectively) by impeding their enzyme activity (Lim et al., 2004; Ueda et al., 2006). In addition, the binding of HMGN1, and to a lesser extent HMGN2, enhances the ability of P300/CBP-associated factor (PCAF) to acetylate H3K14, which is linked to transcriptional activation and gene expression (Herrera et al., 1999; Lim et al., 2005; Ueda et al., 2006). Similarly, HMGN3a/b also increase the H3K14ac at *Glyt1* gene and promotes its expression (Barkess et al., 2012). The modulation of histone modifications by HMGN proteins is dependent on the HMGN C-terminal domain, and the effect on histone modifications is clearly HMGN-variant specific (Ueda et al., 2006).

To gain insight into the role of HMGN2 on the histone modifications, we investigated the global changes in the level of various histone modifications in $HMGN^{+/+}$ and $HMGN^{-/-}$ lines and we also performed ChIP-qPCR assay at different sites of specific genes. The cross-linked chromatin samples from different lines were sonicated to shear the chromatin into smaller fragments ranged between 100-3000 bp, as shown previously in Figure 5.3. In addition, before western blotting of chromatin samples, crosslinks were reversed and nucleic acids removed by benzonase digestion, so that proteins were free to migrate according to their size and charge for checking the global level of histone modifications.

5.7.1 Knockout or mutation of HMGN2 modulates the global level of acetylation and methylation marks

Western blotting was performed in order to study global changes in histone modifications in HMGN2-mutant cells. As previous studies have indicated that acetylation of H3K14 is modulated by HMGN proteins, acetylation of H3K9, H3K14 and H3K27 was investigated. Firstly, the global level of H3K9ac increased in B19ΔHMGN2 and S27ΔNBD lines but did not change in the B38ΔHMGN2 and 19ΔNBD lines (Figure 5.10A). Because this inconsistency, it was hard to draw

conclusion from this. The other acetylation mark, H3K27ac, was slightly reduced in all lines apart from S27ΔNBD. The most constant change in acetylation across all the mutant lines was an increase in H3K14ac of up to 50% (Figure 5.10A).

Histone methylation can be associated with gene activation or repression, depending on the location of the modification. H3K4me1 and H3K4me3 are respectively associated with enhancers and active promoters, H3K27me3 is associated with gene silencing, and H3K36me2 is associated with transcription elongation. Cells with the HMGN2 knockout or mutation had increased global levels of H3K4me3, H3K27me3 and H3K36me2 (Figure 5.10B). Although the H3K4me1 increased in S27ΔNBD, its level in the other lines was close to the wild type level, suggesting it is not linked to HMGN2 status.

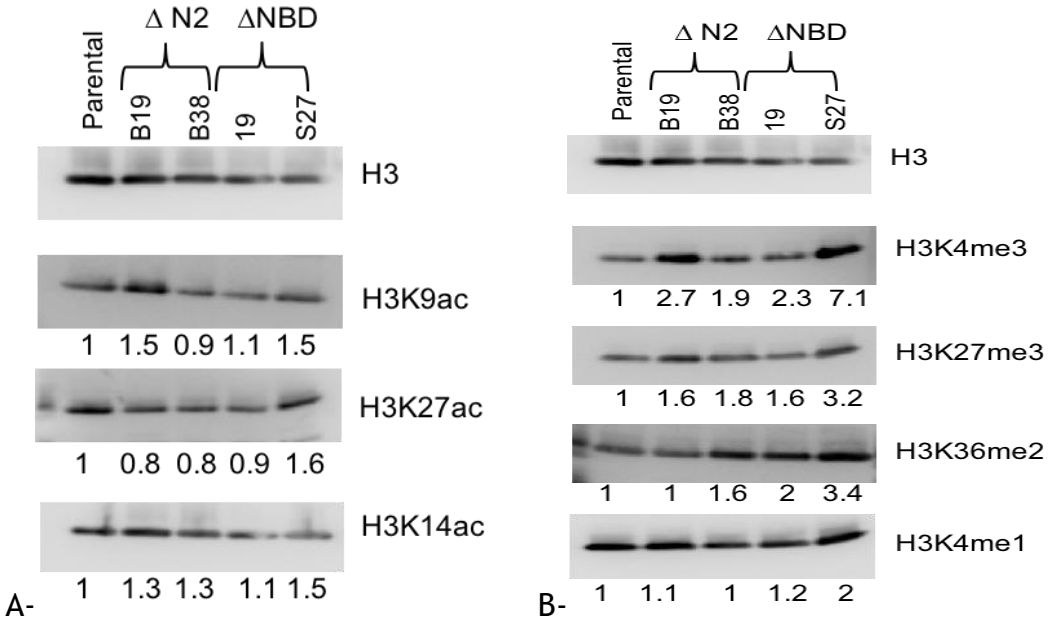


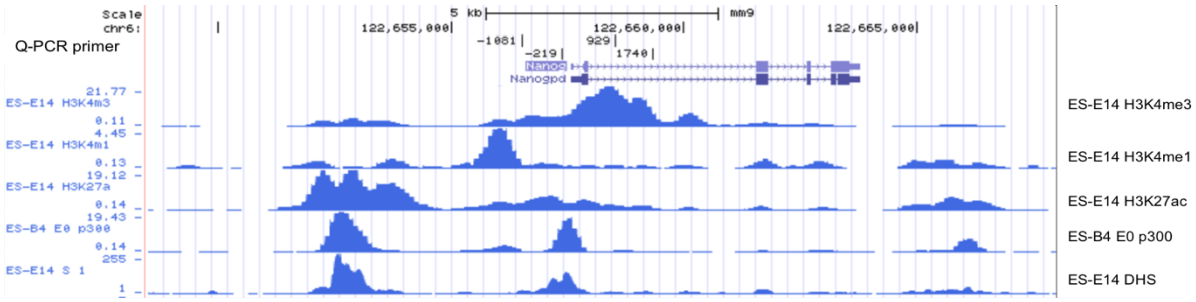
Figure 5.10: The global change of histone acetylation or methylation marks in undifferentiated P19 cells wild type and knockout lines.
A-The reverse cross-linked chromatin extractions from wild type (parental), null HMGN2 lines (B19 and B38), and mutant NBD line (19 and S27) were subjected to western blot for H3K9ac, H3K27ac and H3K14ac after digesting the DNA by benzonase. 10 mM of sodium butyrate was added to samples for inhibiting the histone deacetylase (HDAC). **B-** The western blots for histone methylation marks (H3K4me3, H3K27me3, H3K36me2 and H3K4me1). The immunoblots were performed on the chromatin extractions from previous samples and the band also was quantified similar as for acetylation. In **A** and **B**, the H3 was used as loading control and similar amount of each sample was loaded in different blots. The numbers at bottom of each blot represent the density of each band after normalizing to loading control and relative to parental cells. These immunoblots were performed in duplicate from same extractions and the another set revealed almost similar results.

5.7.2 Depletion and reduction of chromatin-bound HMGN2 in knockout and mutant cells

5.7.2.1 Determine the primer sets for ChIP-QPCR

In order to perform ChIP-QPCR assay for either HMGN2 or histone modification marks, appropriate primer sets were chosen. Firstly, the primer sets around the transcription start site (TSS) of *Nanog* gene were chosen because *Nanog* expression in the knockout lines appeared to be different than in wild type cells. Moreover, the change in *Nanog* expression was different in the complete HMGN2 knockout compared to the NBD-mutant lines. The four primer sets are distributed around the TSS of *Nanog* gene (-1081, -219, 929 and 1740) and reflect the distribution of H3K4me3 enrichment identified in mESCs from the UCSC genomic browser (Figure 5.11A, Rehbini's thesis, 2016). In addition to *Nanog* primers, two primer sets were designed at a putative enhancer (-5394) and the promoter (312) of the *Glyt1* gene. The promoter was mainly defined from the RNA expression whereas the putative enhancer was defined by enrichment of H3K4me1, H3K27ac, P300, and DNAse I hypersensitivity (Barkess et al., 2012). *Glyt1* expression was down-regulated in B19ΔHMGN2 and B38ΔHMGN2 (complete knockout lines) but unchanged in 19ΔNBD and S27ΔNBD (NBD-mutant lines, data not shown). Furthermore, another two primer sets were also used as negative and positive controls: a primer set at around 1 kb downstream of the *β-actin* TSS (positive control for permissive marks and negative for repressive marks) and a primer set at one of the major satellite repeats (M-sat, negative for permissive but positive control for repressive marks) (Barkess et al., 2012).

A-



B

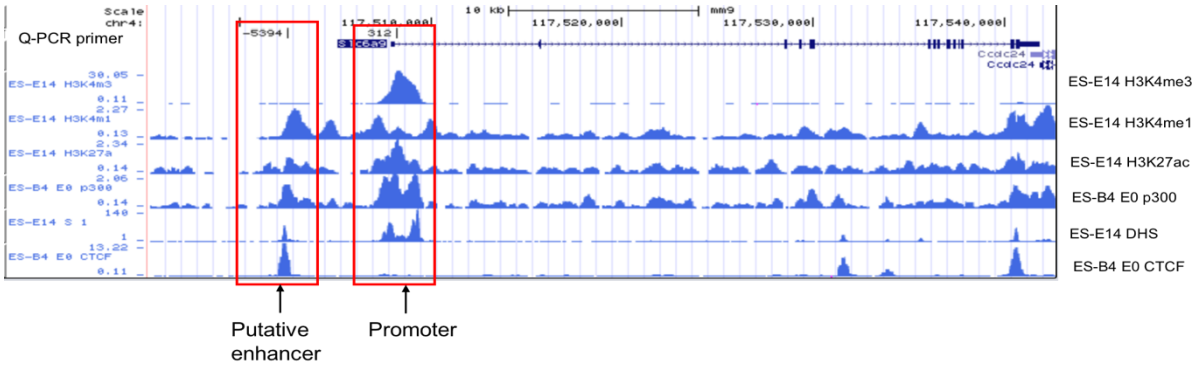


Figure 5.11: Screenshot from genome browser for identifying the primer sets at *Nanog* and *Glyt1 (Slc6a9)* genes.
Screenshot of online available data from the UCSC genome browser displaying the tracks of H3K4me3, H3K4me1, H3K27ac, P300 and DNase hypersensitive site (DHS) across the *Nanog* (**A**) and *Glyt1* (**B**) genes. The data was generated by the Ren laboratory (Ludwig Institute for Cancer research, USCD) as part of the ENCODE consortium (genome build NCBI37/mm9) (http://genome.ucsc.edu/) (Fujita et al., 2011; Rosenbloom et al., 2012). The data for H3K4me3, H3K4me1, H3K27ac and DHS are from mouse E14 cells, while the P300 and CTCF are from Bruce4 embryonic stem cells. Both diagrams illustrate the loci of primer sets used for ChIP-QPCR assay. **A**- The black boxes represent the position of primer sets across the *Nanog* gene, and their positions to TSS were indicated (-1081, -219, 929 and 1740). **B**- The identified primer sets for the promoter and putative proximal enhancer (red boxes) at *Glyt1* gene, and these two primer sets were named based on their position from TSS as -5394 and 312. Scale bar is given in A and B on top of the diagrams. The traces names are given in both sides.

5.7.2.2 Chromatin binding by HMGN2 is reduced in knockout and mutant P19 lines

The interaction of HMGN proteins with nucleosome is highly dynamic. However, it is questionable whether the binding of HMGNs with nucleosomes occurs randomly or more specifically at some sites of the genome. Genome wide mapping of HMGN1 and HMGN2 binding indicated that the binding of HMGN proteins at sites overlapping DHSs, that contain the regulatory elements such promoters and enhancers, is significantly stronger than sites not overlapping with DHSs (Cuddapah et al., 2011; Deng et al., 2013; Deng et al., 2015; Zhang et al., 2016). However, HMGN1, HMGN2 and HMGN3 were shown to have fairly even

distribution across a 50 kb region spanning the *Glyt1* gene in murine Hepa-1 cells (Barkess et al., 2012).

In this study, HMGN2 was found to be equally bound across the *Nanog* gene, at the promoter and putative enhancer of the *Glyt1* gene, and at the housekeeping gene, *β-actin,* in wild type ECCs (P19) with an average enrichment of 2 to 4-fold (Figure 5.12A). In particular, there was no increase in occupancy of HMGN2 between the primer sets that overlap with the peak of DHSs at *Nanog* (-219) and *Glyt1* (-5349 and 312), and the primers that do not overlay with DHS regions of *Nanog* gene (-1081, 929 and 1749). In contrast, the negative control for the permissive mark (M-sat) was not occupied by HMGN2 (0.2-fold) (Figure 5.12A).

To confirm the loss of HMGN2 enrichment in knockout lines, the knockout samples were also ChIPed using the antibodies against the C-terminal of HMGN2, H3 and IgG (background control). In Figure 5.12A, the complete knockout lines (B19ΔHMGN and B38ΔHMGN2) have shown a depletion of HMGN2 enrichment comparing to parental line at all loci of *Nanog*, *Glyt1* and *β-actin* genes. This confirms the loss of HMGN2 in these lines and the specificity of the antibody used in the ChIP assay. On the other hand, the binding of HMGN2 to chromatin in the 19ΔNBD line that has deletion in NBD was reduced by 75% at most locations (Figure 5.12A). The immunoprecipitation of histone, H3, showed almost same quantity of histone H3 at each specific site in all lines, except B19ΔHMGN that show higher H3 at the M-sat site compared to other lines at same region (Figure 5.12B). The background enrichment with rabbit IgG antibody was less than 0.5. In sum, the HMGN2 is occupied at all examined sites, except the M-sat, and the HMGN2 interaction with nucleosomes was lost in all HMGN2 knockout.

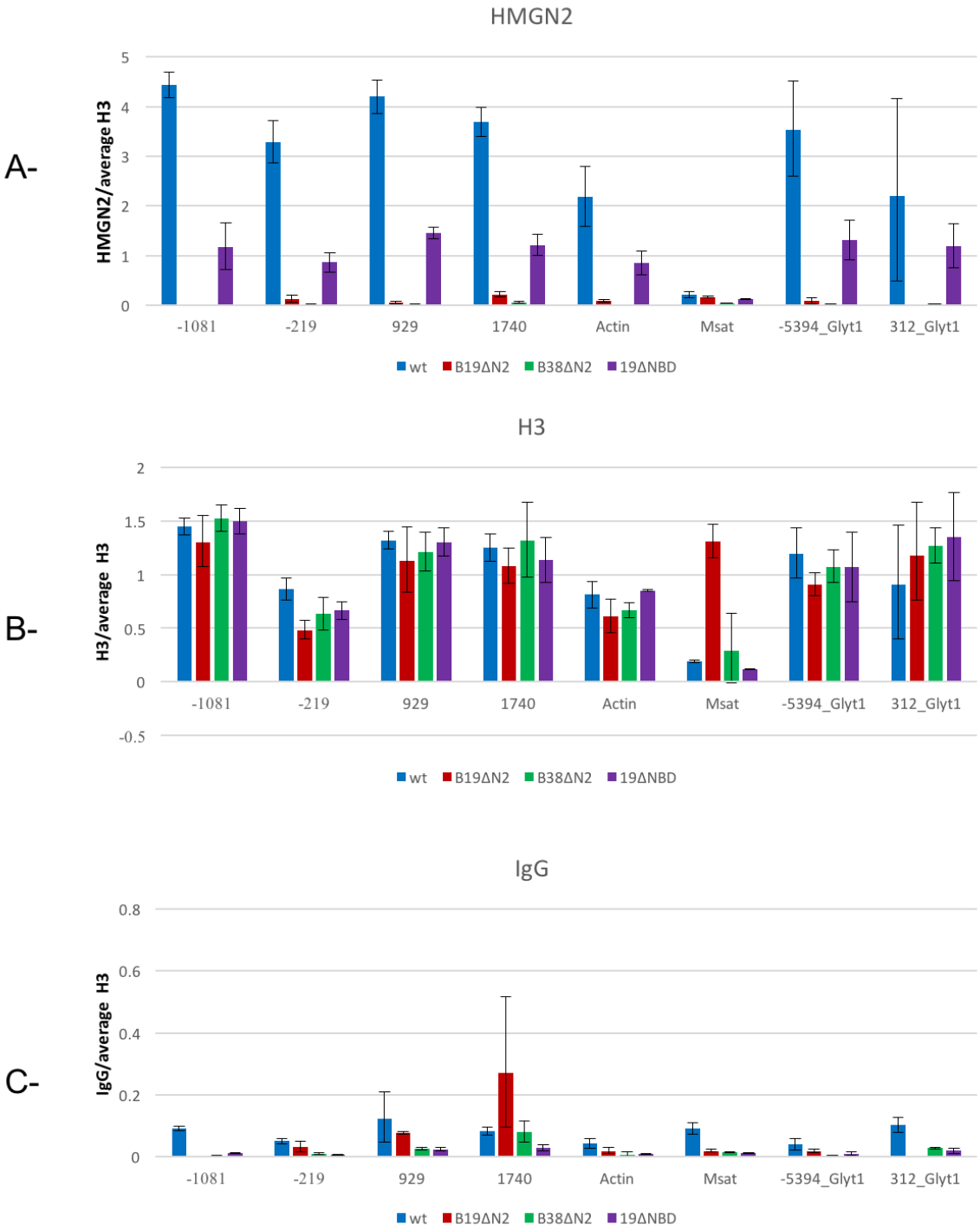


Figure 5.12: HMGN2 occupancy and level in wild type and knockout lines around the TSS of *Nanog*, at the TSS of *β-actin*, and promoter and enhancer of *Glyt1* gene.
A- ChIP-QPCR assay for HMGN2 in wild type cells, HMGN2 knockout lines (B19 and B38), and in line having mutant NBD of HMGN2 (19). Different primer sets were used around the TSS of *Nanog* gene (numbers indicate their positions relative to the TSS), at the TSS of *β-actin*, and putative enhancer (-5394) and promoter (312) of *Glyt1* gene. Negative control of enrichment was shown by using primer set for M-satellite (Msat). Results are reported as fold enrichment of immunoprecipitated DNA after normalising to input DNA and the H3 immunoprecipitation as described in the methods. The Error bars reflect the standard deviation from Q-PCR triplicates from one biological replicate. **B** and **C**- The ChIP-QPCR for the H3 enrichment and IgG (background control) at all region tested in different cell lines. The same analytical method used with HMGN2 were also applied here.

5.7.3 The effect of HMGN2 knockout or mutation on histone acetylation and methylation

5.7.3.1 The consequence of HMGN2 knockout or mutation on the acetylation marks at the *Nanog* and *Glyt1* genes

Since there were some differences in the global levels of histone acetylation marks in the HMGN2 knockout and mutant cells, we used CHIP to investigate the level of histone acetylation (H3K9ac, H3K14ac and H3K27ac) at all the loci studied previously. It is well known that acetylation marks are correlated with transcriptionally active genes, and ESCs, which have more permissive chromatin signature, contain high levels of histone H3 and H4 acetylation (reviewed in Mattout and Meshorer, 2010). Moreover, inhibition of histone deacetylation (HDAC) promotes the self renewal of mouse and human ESCs and prevents the differentiation (Ware et al., 2009).

H3K9ac: In wild type cells, H3K9ac is most highly enriched just after the TSS on both the *Nanog* and *Glyt1* genes (positions 929 and 312 respectively), as well as *β-actin* (Figure 5.13A). It is less or not enriched at positions -219 and 1740 of *Nanog*, and is not enriched at the putative *Glyt1* enhancer (-5394). In the knockout lines, the amount of H3K9ac mark was reduced by around 60% at *Nanog* position 929 in both B19ΔHMGN2 and B38ΔHMGN2. Reductions at other primer sets were less consistent, and there was no reduction at *β-actin*.

The level of H3K9ac enrichment in the 19ΔNBD line is very low at all primer sets, and one explanation could be that the immunoprecipitation did not work (Figure 5.13A). However, independent ChIP experiments by S. Svambaryte in the lab demonstrated a 75% reduction in H3K9ac at the *Glyt1* gene promoter in 19ΔNBD cells (data not shown). Further replicates are required in order to investigate this further.

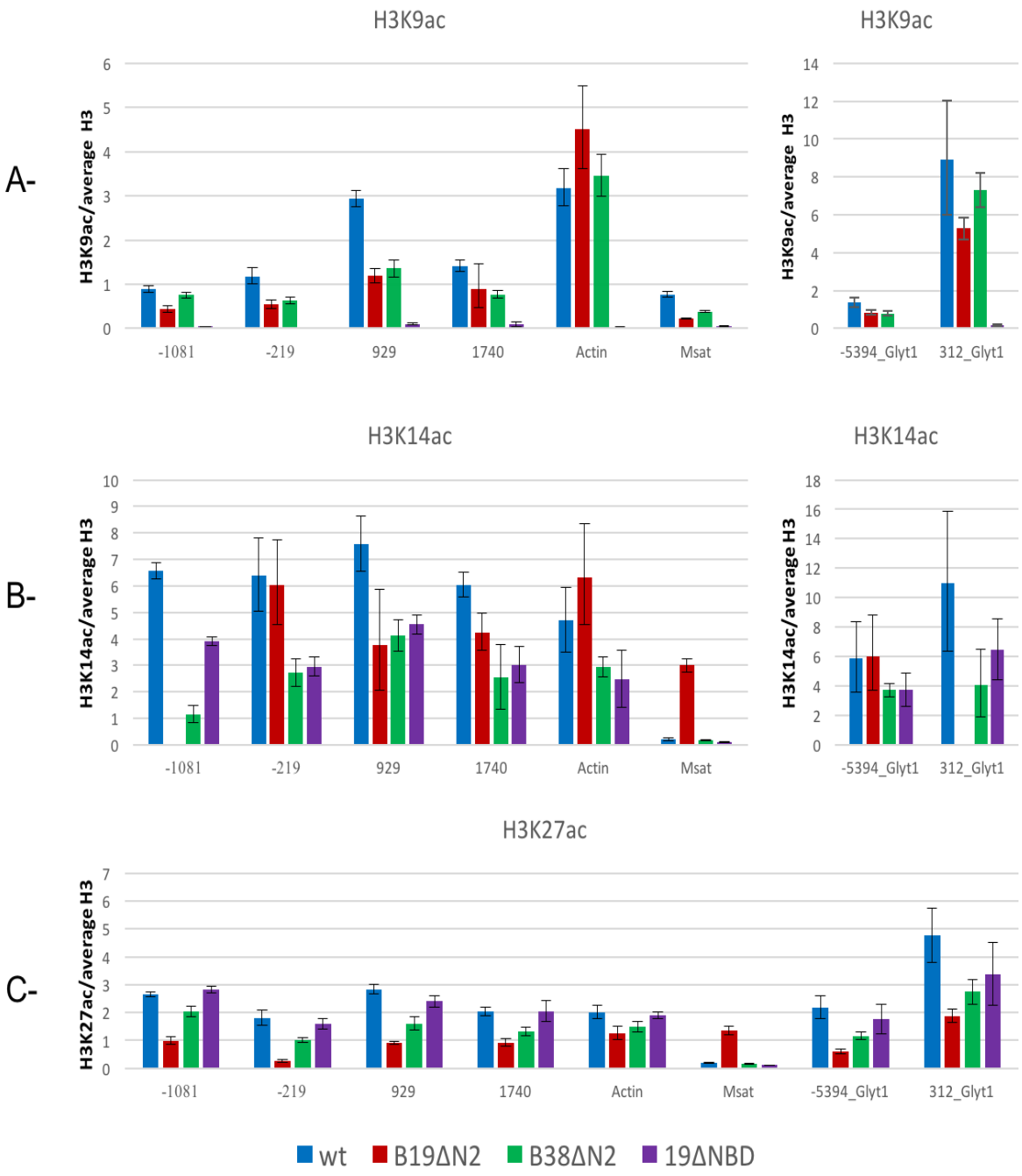


Figure 5.13: The amount of H3K acetylation at different lysine positions in the knockout lines and wild type.
ChIP–qPCR analysis of H3K9 (A), H3K27 (B) and H3K14 (C) acetylation binding around the TSS of *Nanog*, at *β-actin*, and promoter and putative enhancer of *Glyt1* gene. The blue, red, green and purple bars are for wild-type, B19ΔHMGN2, B38ΔHMNG2 and 19ΔNBD, respectively. Results are reported as fold enrichment of immunoprecipitated DNA after normalising to input DNA and the H3 immunoprecipitation as described in the methods. The Error bars reflect the standard deviation from Q-PCR triplicates from one biological replicate. Negative control of enrichment was shown by using primer set for M-satellite (Msat). The IgG and the H3 ChIP, and the position of primer sets across the *Nanog* and *Glyt1* genes were shown in Figure 5.12 and 5.11. The primer locus of *Nanog* are: -1081, -219, 929 and 1740. -5394_*Glyt1* and 312_*Glyt1* are the primer sets for the putative enhancer and promoter of *Glyt1* gene, respectively.

H3K14ac: H3K14ac shows a different profile to that of H3K9ac, with enrichment at all locations tested except for M-sat (Figure 5.13B). A similar profile was

previously observed across the *Glyt1* gene locus (Barkess et al., 2012). The level of H3K14ac was diminished by up to 50% at most of primer sets in lines B38ΔHMGN2 and 19ΔNBD. Data from the B19ΔHMGN2 line was less consistent, with some primer sets showing no signal, and a positive signal at M-sat that correlates with high levels of H3 in this line (Figure 5.12B).

Independent ChIP experiments by others in the lab also demonstrated a 50% reduction in H3K14ac at the *Nanog* gene in B38 cells (M. Sramek, data not shown) and at the *Glyt1* gene in 19ΔNBD cells (S. Svambaryte, data not shown).

H3K27ac: H3K27ac tends to be found at promoters and enhancers, and is often used to help defining active enhancers. In wild-type P19 cells, H3K27ac was most highly enriched (around 5-fold) at position 312 from *Glyt1* TSS (Figure 5.13C). Nevertheless, all the primer sets across the *Nanog* gene had a moderate enrichment of 2-3 fold, and this is identical to the available data from the genome browser (Figure 5.11). In the knockout B19ΔHMGN2 and B38ΔHMGN2 lines, H3K27ac level at *Nanog*, *Glyt1* and *β-actin* was also decreased by 30-60% (Figure 5.13C). However, H3K27ac was unchanged in the 19ΔNBD line comparing to its level in the parental line.

The ChIP data for H3K27ac is comparable with the changes in the global levels of H3K27ac observed by western blotting: 20% reductions in H3K27ac were observed in lines B19ΔHMGN2 and B38ΔHMGN2, whereas a smaller reduction of 10% was observed for 19ΔNBD (Figure 5.10). However, the changes in H3K9ac and H3K14ac at specific genomic locations do not correlate very well with the global changes observed by western blotting. This could reflect the fact that the western blotting is semi-quantitative, but it could also be because changes in modifications at specific loci such as promoters are not reflected elsewhere in the genome.

Taken together, the data indicates that knockout of HMGN2 has reduced the acetylation of H3K9, H3K14 and H3K27 at the *Nanog* gene locus. This is consistent with the reduction in *Nanog* expression in these knockout lines. In the mutant 19ΔNBD line, however, although H3K9ac and H3K14ac levels at the *Nanog* locus appeared to be reduced, *Nanog* mRNA expression was up-regulated. This suggests that the changes in H3 acetylation levels are not simply the

consequence of changes in transcription activity, but may be a direct result of the reduction in HMGN2 binding. This is consistent with other studies that found that HMGN proteins promote H3K14 acetylation (Herrera et al., 1999; Lim et al., 2005; Ueda et al., 2006; Barkess et al., 2012). Therefore, this data raises the possibility that HMGN2 may support and maintain several histone acetylation marks at individual genes; however further independent replicates are required to investigate this further.

5.7.3.2 The consequence of HMGN2 knockout or mutation on the histone methylation marks at the *Nanog* and *Glyt1* genes

Although the effect of HMGN proteins on the permissive histone acetylation and phosphorylation marks has been studied by other groups, the effect of HMGN2 on the permissive and suppressive histone methylation marks has not been elucidated. In addition, the observation of an increase in the global levels of H3K4me3 and H3K27me3 as shown by western blot led us to study these marks at the *Nanog* and *Glyt1* genes.

H3K4me3: H3K4me3 is a permissive mark that is typically highly enriched at active promoters. H3K4me3 is highly enriched at position +929 after the TSS of *Nanog* and at position +312 after the TSS of *Glyt1*, exactly as expected (Figure 5.14). In the same figure, the other regions of *Nanog* and the enhancer of *Glyt1* are either not or less enriched with H3K4me3, and this is identical to the peaks identified in genome browser of E14 cells (Figure 5.11). In addition, the H3K4me3 is also high at the TSS of positive control housekeeping gene (*β-actin*) but undetectable in M-satellite region. The enrichment of H3K4me3 is around 20-fold at the TSS of *Nanog* and *β-actin* while it is 40-fold at the TSS of the *Glyt1* gene.

The H3K4me3 mark enrichment at the TSS of *Nanog* was reduced by 50% in all three knockout and mutant lines. Enrichments at the TSS of *Glyt1* and *β-actin* were also reduced by 50% in lines B38ΔHMGN2 and 19ΔNBD but not in B19ΔHMGN2 (Figure 5.14). The reduction in H3K4me3 at the *Nanog* locus contrasts with the increase in global levels of H3K4me3 observed by western blotting (Figure 5.10).

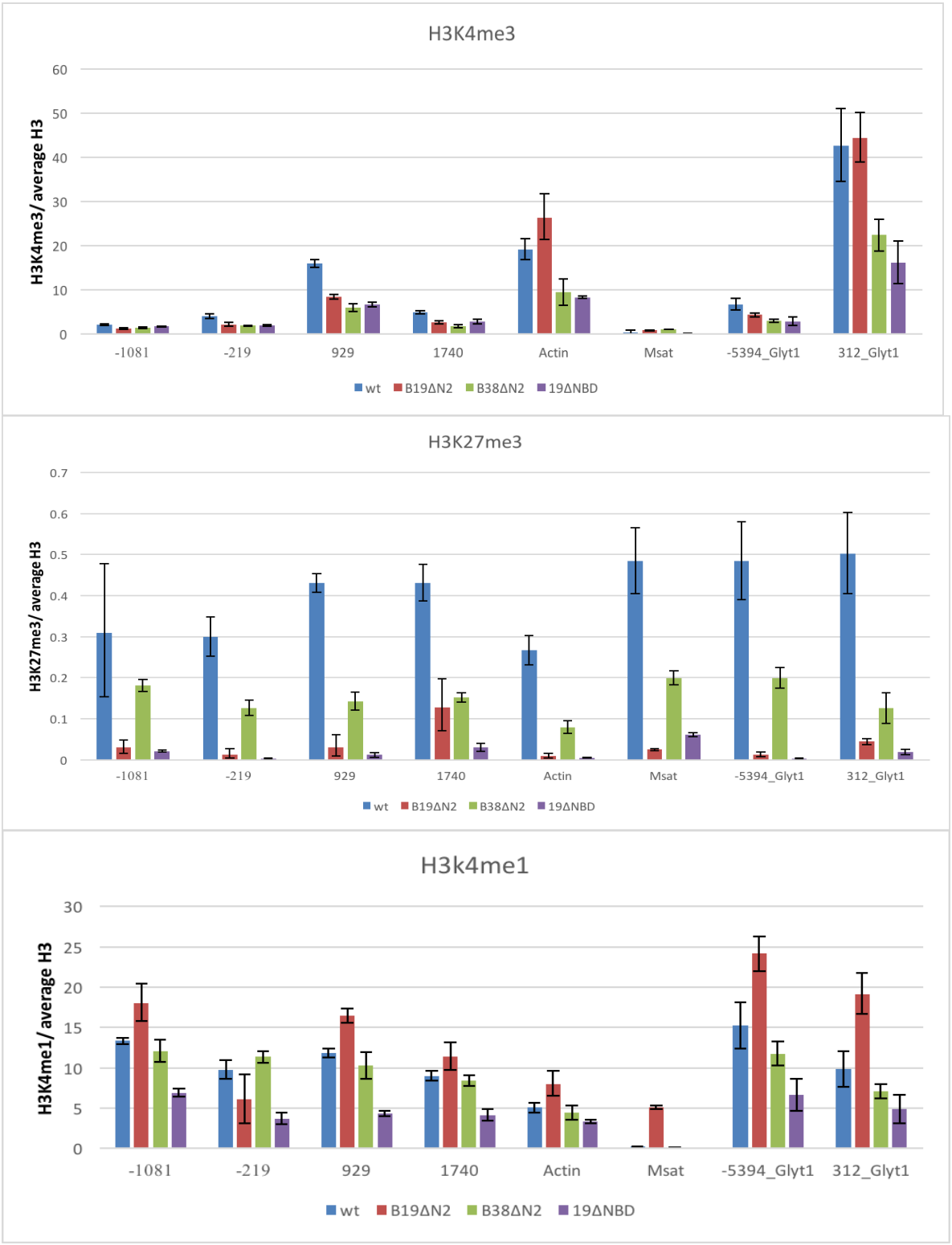


Figure 5.14: The level of histone methylation marks at different region of the genome. ChIP–qPCR analysis of H3K4me3 (top), H3K27me3 (middle) and H3K4me1 (bottom) binding around the TSS of *Nanog*, at *β-actin*, and promoter and putative enhancer of *Glyt1* gene. Results are reported as fold enrichment of immunoprecipitated DNA after normalising to input DNA and the H3 immunoprecipitation as described in the methods. The Error bars reflect the standard deviation from Q-PCR triplicates from one biological replicate. The IgG and the H3 ChIP, and the position of primer sets across the *Nanog* and *Glyt1* genes were shown in Figure 5.12 and 5.11. The primer locations at *Nanog* are: -1081, -219, 929 and 1740. -5394_*Glyt1* and 312_*Glyt1* are the primer sets for the putative enhancer and promoter of *Glyt1* gene, respectively.

H3K27me3: H3K27m3 is a repressive mark that is often found at silenced and bivalent genes. H3K27me3 was also examined at the same regions because the down-regulation of *Nanog* gene may be correlated with augmentation of this repressive mark. In fact, it was shown that the pluripotent genes gain tri-methylation of H3K27 during ESCs differentiation (Obier et al., 2015). The enrichment of H3K27me3 is in the region of 0.3 – 0.5 -fold in the wild type chromatin at all sites, including M-sat (Figure 5.14, middle panel). It is not clear how well the ChIP has worked in this experiment, as none of the primer sets shows a strong enrichment. However, as the enrichment of 0.3-0.5 is higher than that of IgG (less than 0.1), the H3K27me3 signal does appear to be higher than that of the background in the wild type cells. It is interesting that the H3K27me3 signal in all three HMGN2 knockout and mutant lines is reduced at all primer sets. Use of primer sets corresponding to regions known to be enriched for H3K27me3, such as bivalent gene promoters, could provide more robust data about the H3K27me3 levels in these cells.

H3K4me1: H3K4me1 is typically found both at promoters and at enhancers, and is often used to define the location of enhancers. Here, H3K4me1, was enriched between 10 and 15 fold at all primer sets over the *Nanog* locus in parental P19 cells (Figure 5.14, bottom panel). Similarly, the amount of H3K4me1 was 15-fold at the putative enhancer and 10-fold at promoter of *Glyt1* gene of the parental cells. However, this mark was only 5-fold enriched at the TSS of *β-actin,* and was undetectable at the M-satellite. The B38ΔHMGN2 line showed similar levels of H3K4me1 to wild type cells at all loci of *Nanog, Glyt1* and *β-actin*. Conversely, the other complete knockout line, B19ΔHMGN2, showed higher level of H3K4me1 comparing to parental line at most of the loci examined. As previously observed for some other ChIP experiments using the line, B19ΔHMGN2 shows higher signal with the major satellite primer set. Further replicates using a different batch of chromatin would be required to investigate this further. In contrast to the HMGN2 knockout lines, the NBD-mutant line (19ΔNBD) has a reduction in H3K4me1 signal of around 50% at all locations. Taken together, the results do not indicate a strong relationship between HMGN2 status and H3K4me1 enrichment at these genomic loci.

5.8 Discussion

5.8.1 The functional redundancy between HMGN variants

We studied the redundancy between the most abundant variants of HMGN proteins that are widely expressed in vertebrate cells. We observed that there is a possible increase in the HMGN1 protein which may indicate a sort of compensation in the removed or inefficient mutant HMGN2, though there was no a significant difference in the mRNA levels between all lines and parental cells. Although there was also an increase of HMGN1 in all chromatin extracted from HMGN2 knockout lines, the most redundant effect was seen in the lines that have mutant HMGN2 (less efficient to bind to chromatin). This is unexplainable because the compensation effect would be expected higher in the lines that do not have HMGN2. As this was examined using western blotting which is semi-quantitative, it would be beneficial to perform ChIP-QPCR on these sample to accurately identified if there is a compensation effect between these two variants.

A very recent study found that a combined loss of both proteins remodels the DHSs landscape of MEF cells (Deng et al., 2015). In the same study, more severe phenotypical effects were seen in mice with HMGN1$^{-/-}$N2$^{-/-}$ than mice with one variant knockout. In contrast, study of the transcription profile in disregulation of HMGN variants, which showed the changes in expression profile to be variant specific, suggesting the limited redundancy between these variants in the cellular transcription (Rochman et al., 2011). Moreover, the finding of that some histone modifications are modulated by one variant of HMGN, and the distinct domains in HMGN variants differentially modulate specific chromatin modifications (Lim et al., 2004; Ueda et al., 2006). Therefore, although there is redundancy between the HMGN1 and 2, it is still arguable that they have a distinct function.

5.8.2 The HMGN2 has no influence on the cell proliferation and cell cycle of P19 cells

The preliminary investigation of the role of HMGN2 during proliferation and cell division indicated that the removal or mutation of HMGN2 alone does not influence the cell growth rate or cell cycle in P19 EC cells. Previously, it was

shown that HMGN2 can increase the efficiency of replication *in vitro* (Vestner et al., 1998). In particular, using *in vitro* SV40 replication system and compared between minichromosomes containing normal or deficient of HMGN2, found that HMGN2 enhances the DNA replication efficiency by unfolding chromatin structure and increase the accessibility of the replication machinery (Vestner et al., 1998). So, it is questionable why this is not seen with the removal of HMGN2 *in vivo*. Since the replication efficiency was facilitated by HMGN2 through inducing more relaxed chromatin structure that is more accessible to the replication machinery, the another abundant variant, HMGN1, may also involve in enhancing the replication by compensating the removal of HMGN2. In fact, the *in vitro* replication system used in Vestner study relies on incorporation of HMGN2 into chromatin in absence of other HMGN variants. In agreement with our finding, Lim et al, (2004), who used flow cytometry and PI, did not observe any significant difference in the cell cycle progression between *Hmgn1*$^{-/-}$ and *Hmgn1*$^{+/+}$ MEFs.

5.8.3 HMGN2 regulates *Nanog* expression in EC P19 cells

In contrast to previous data from siRNA experiments, we found that loss or mutation of HMGN2 does not change the levels of *Oct4* and *Sox2* expression, but only regulates the expression of homeobox transcription factor, *Nanog*. Oct4, Nanog and Sox2 are part of the core network of TFs that are essential for maintaining pluripotency and preventing the differentiation of ES and EC cells. These TFs synergistically maintain ES cell identity, because the Oct4 and Sox2 can auto-regulate their expression through a feedback loop, and this complex is also recruited to the *Nanog* promoter to regulate its expression (Kuroda et al., 2005; Rodda et al., 2005). In addition, Nanog also can regulate genes related to pluripotency of P19 EC cells in which disregulation of *Nanog* (up or down) by siRNA resulted in analogous disregulation (up or down) of other pluripotent markers including *Oct4, Fgf4, Klf2, Rexl, Soxl, Yes*, and *Zfp143* but not *Sox2* (Choi et al., 2012).

However, it was reported that the reduction of *Nanog* expression in murine and human ES and EC cells leads to loss the pluripotency, and induces differentiation into extraembryonic endoderm and trophectoderm lineages; overexpression of *Nanog* in mES cells can promote self-renewal, sustain the level of Oct4 and resist

the cell differentiation (Mitsui et al., 2003 Chambers et al., 2003; Hatano et al., 2005; Hyslop et al., 2005). Therefore, disregulation of only *Nanog* expression, but not the other pluripotent TFs, by eliminating or disrupting the HMGN2 may affect the pluripotency and the differentiation state of the cells. By examining the morphological features between the parental and knockout lines, there were no apparent morphological differences between the lines, apart from the 19ΔNBD that grow as colonies rather than monolayer (as will be seen in Chapter 6). However, as it is going to be discussed in next chapter, the cell lines that showed a reduction in *Nanog* expression also showed an elevation level of neural-related genes which indicates a difference in the maintenance of the self-renewal ability. Although there was no apparent change in cells morphology, and the expressions of *Oct4* and *Sox2* were not changed in alongside *Nanog* disregulation, the maintenance of pluripotency in HMGN2$^{-/-}$ and mutant-HMGN2 lines is more likely to be affected as a result of *Nanog* reduction and neural gene induction. It is still necessary to investigate the extraembryonic endoderm genes (*Gata4* and *6*) as it was shown that the transient reduction of *Nanog* increased the expression of these genes in mES cells (Hough et al., 2006). Yet, it was reported by Choi et al (2012) that knockdown or overexpression of *Nanog* in P19 cells did not affect the expression of *Dab2*, *Gata4*, and *Gata6*, which are common markers for extraendodermal cell lineage after 48 hours.

The differential expression of *Nanog* varied, depending on the type of mutation in the *Hmgn2* gene. *Nanog* was down-regulated in the HMGN2 knockout lines (B8ΔHMGN2, B19ΔHMGN2 and non-significant in B38ΔHMGN2), while it was up-regulated in HMGN2 mutant lines, 19ΔNBD and S27ΔNBD, with the low binding efficiency to chromatin. This differential regulation between the complete knockout and mutant lines indicates that this gene is likely to be modulated by HMGN2, especially because HMGN2 is enriched around the TSS of *Nanog* gene as seen by chromatin immunoprecipitation. The reduction in *Nanog* expression is consistent with a role for HMGNs in transcriptional enhancement, but the increased level in the other lines is not, although gene expression profiling indicated that the disruption one or more HMGN proteins can result in both up- and down-regulation of different sets of genes (Rochman et al., 2011; Kugler et al., 2013). However, having diverse effects as a result of either complete removal or partial disruption of HMGN on the same gene has not been observed

before. It is possible that the two mutant HMGN2 forms in 19ΔNBD and S27ΔNBD, which share the N-terminal 28 amino acids but differ in C-terminal of the protein, have some sort of gain of function, or dominant positive activity, that is not simply a reduction in the activity of the full length protein. This novel activity could be associated with the mode of HMGN2 binding to nucleosomes. For example, the complete removal of HMGN2 could result in less accessible chromatin whereas the mutant-HMGN2 could form a complex that may be more accessible than that of wild type HMGN2 complex with nucleosomes. Alternatively, the mutant HMGNs could form novel interactions with transcription factors that directly or indirectly regulate *Nanog* expression. This can be possibly the situation especially because HMGN2 proteins are subjected for different posttranslational modification such as phosphorylation of serine or acetylation of lysine residues that can abolish, or maybe reduce, the binding affinity of these protein to chromatin (Prymakowska-Bosak et al., 2001; Bergel et al., 2000). Indeed, it was recently revealed that the lysine 2 on HMGN2, which is acetylated in normal breast tissue but not in breast tumour, can be deacetylated by HDAC6 to enhance Stat5a-mediate gene transcription (Medler et al., 2016).

Importantly, in addition to the differential regulation of *Nanog* expression in the HMGN2 mutant lines, we also noted differential expression between the parental and two monoclonal wild type lines. Various studies have shown that *Nanog* expression in ES or EC cells is heterogeneous between the cells in the same culture, and can be categorised as High or Low *Nanog* (HN and LN, respectively) cells (Chambers et al., 2007; Singh et al., 2007; Kalmar et al., 2009). This heterogeneity of *Nanog* expression is correlated with the probability of self-renewal and differentiation, in which the LN cells are more prone to differentiate. Thus, the significant reduction of *Nanog* in the B1 (negative control line) and slightly increase of *Nanog* in the other negative control, B22, may be comparable with the LN and HN populations, respectively. Moreover, it is also possible that deferential expression of *Nanog* in the HMGN2 knockout lines reflects the heterogeneity in the population. However, this seems unlikely due to the consistency of the data - *Nanog* expression in the three HMGN2 knockout lines was reduced, while it was increased in the two HMGN2 mutant lines with partial binding of HMGN2. Therefore, HMGN2 may control the *Nanog*

heterogeneity in EC P19 cells but more studies are still required to corroborate this observation. In general, it is clear that HMGN2 regulates the *Nanog* expression in EC P19 cells.

5.8.4 HMGN2 enhances the permissive histone marks at different site of the genome

We have confirmed the HMGN2 binding at different genomic loci of P19 cells. The comprehensive loss or reduction of HMGN2 binding in the complete knockout lines (B19ΔHMGN2 and B38ΔHMGN2) or in mutant-HMGN2 line (19ΔNBD), validates the ChIP assay by showing that enrichment is specific to HMGN2. HMGN2 enrichment was found equally at most of the examined sites, regardless of distance from the TSS, promoter or enhancer location, or overlap with DHS. Although the genome-wide distribution for HMGN1 and 2 identified distinct peaks at chromatin regulatory sites including DHSs, promoters, functional enhancers, and transcription factor binding sites in T cells and NPCs, these peaks were not detected in mES cells (Cuddapah et al., 2011; Deng et al., 2013). The authors concluded that this was due to the hyperdynamic nature of chromatin in mES cells (Cuddapah et al., 2011; Deng et al., 2013). Similar to mES cells, from the examined sites HMGN2 seems to bind equally to the gene body, promoter and enhancer regions in P19 cells. The only site that was not occupied by HMGN2 was M-satellite, which represents the heterochromatin regions. It might be valuable to perform ChIP-seq for HMGN2 in P19 cells to investigate how HMGN2 is distributed across the genome.

We next sought to identify the effect of HMGN2 on histone acetylation and methylation marks. Studies of the global level of different histone acetylation marks firstly suggests that H3K14ac may increase in the HMGN2$^{-/-}$ and mutant-NBD lines. However, this finding is in contrast with previous studies that reported that HMGN1 and 2 enhance the acetylation of H3K14 by PCAF (Herrera et al., 1999; Lim et al., 2005; Ueda et al., 2006). Further studies, using the site specific but more quantitative ChIP-QPCR, indicated that H3K14ac was consistently reduced at the *Nanog* gene in all mutant cell lines, with some reduction at the *Glyt1 and β-actin* genes. Secondly, the global level of H3K27ac was slightly reduced in both HMNG2$^{-/-}$ lines but to a lesser extent in the 19ΔNBD line. H3K27ac was also reduced at the *Nanog*, *Glyt1* and *β-actin* loci in complete

knockout lines but not in the 19ΔNBD line. Therefore, the reduction in H3K27ac enrichment are comparable either at global level or at three gene specific sites. In addition, the change in this mark is directly correlated with the expression of *Nanog* and *Glyt1*. Mainly, the H3K27ac enrichment was decreased in complete knockout lines that has low *Nanog* and *Glyt1* expressions, but it was not changed in 19ΔNBD line that has high *Nanog* and unchanged *Glyt1* level. Finally, although the level of H3K9ac, which can predict the pluripotency of ESCs and the reprogramming capacity (Hezroni et al., 2011), was varied between the lines, the enrichments of H3K9ac was diminished at the TSS of *Nanog*, but not at *β-actin*, in the HMGN2$^{-/-}$ lines. Taken together, this data suggests that HMGN2 enhances the histone acetylation marks of H3K9, H3K14 and H3K27 at certain genomic locations.

Global levels of several histone methylation marks were also altered when the HMGN2 knockout and mutant lines were compared using western blotting. Levels of H3K4me3 and H3K27me3 were increased in all mutant lines, while H3K4me1 was almost unchanged. In addition, H3K36me2 was increased in 3 out 4 lines. H3K4me3 and H3K27me3 are enriched at active and silenced promoters, respectively, but they can be found together at some promoters in ES cells and these are known as bivalent domains (Bernstein et al., 2006). The function of a bivalent domain is to maintain developmental genes in a poised state, which allows timely activation or repression of these gene upon differentiation, while keeping them at a very low level of expression in ES cells (Bernstein et al., 2006). Alteration in these two marks in P19 cells as a result of loss or disruption of HMGN2 could affect this transition state, as the majority of H3K27me3 and H3K4me3 in ES cells is found at bivalent domains (Voigt et al., 2013). Although the global level of H3K4me3 was increased in HMGN2$^{-/-}$ cells compared to wild-type, its enrichment was consistently reduced at the *Nanog* gene in HMGN2$^{-/-}$ and 19ΔNBD cells. The ChIP-QPCR for H3K27me3 suggests this mark is reduced in all the HMGN knockout and mutant cells; however, this conclusion is uncertain due to the lack of a positive control primer set. There are no consistent changes in H3K4me1 levels at the genomic locations tested.

Collectively, these preliminary results indicate that HMGN2 could directly or indirectly modulate several permissive marks (H3K9ac, H3K14ac, H3K27ac and

H3K4me3) and possibly one repressive mark (H3K27me3). Notably, the global reduction of H3K27ac and the increase of H3K27me3 in HMGN2$^{-/-}$ cells is interesting. These two modifications are on the same lysine reside, and it was reported that the active transcriptional mark H3K27ac can antagonise the transcriptional silencing by H3K27me3 that is mediated by polycomb repressive complex 2 (PRC2) (Tie et al., 2009). Thus, HMGN2 may directly or indirectly regulate the switch between the two marks. Although there are some differences in the changes in histone marks between the global levels and the ChIP-qPCR, this experiment was performed on one biological chromatin sample and more replicates are required. It is also important to note that the genomic locations that were studied might not be representative of other genes, and ChIP-seq is required in order to obtain a comprehensive picture of the changes in each histone mark.

Chapter 6 Analysis of HMGN2 mutant lines during the neural and neuronal-lineage commitment

6.1 Introduction

P19 cells, which are ECCs and derived from teratocaracinoma, can differentiate into all cell types of all 3 germ layers. Specifically, retinoic acid (RA) stimulates the neural induction and can produce neurons, astrocytes, microglias and oligodentrocytes (Bain et al., 1994). Thus, P19 cells are considered to be an excellent model system for analysing neural differentiation and activities. The most widely approach to induce the neural lineage of P19 cells relies on forming cell aggregates in suspension known as embryoid bodies (EBs) (Endo et al., 2009). There are some disadvantages for this method including a longer time for obtaining mature neurons, the presence of large three-dimensional clusters of cells, and a high level of heterogeneity in the cell population, such as large flattened non-neuronal proliferating cells that over time can comprise the majority of the population.

On the other hand, another method was developed by Nakayama et al. (2014) that does not require the formation of EB in a suspension culture. This method uses an adherent serum-free culture, as serum can be inhibitory for neural differentiation (Pachernik et al., 2005). In addition, it promotes a rapid and efficient induction of undifferentiated P19 cells into neurons within four days and into functional neurons by day 6. The P19 cells are induced by adding RA that stimulates the neural lineage in the presence of FGF8 and DAPT (N-[N-(3,5-Difluorophenacetyl-L-alanyl)]-(S)-phenylglycine t-butyl ester). FGF8 is an essential signalling molecule that regulates the normal development of the mid/hindbrain (Crossley et al., 1996). Furthermore, FGF8 overexpression induces the neural differentiation of P19 monolayer cells (Wang et al., 2006). DAPT is a γ-secretase inhibitor that inhibits Notch signalling by preventing its cleavage at the cell membrane. Notch is a transmembrane receptor mediating cell-cell communication and coordinating a signalling cascade. Notch signalling is important during neurogenesis in which it has a negative effect and suppresses neural differentiation. It blocks neural differentiation by repressing the expression of proneural genes in a cluster of cells that have neural potential via

the expression of basic helix-loop-helix (bHLH) transcriptional repressors (reviewed in Lai., 2004).

In addition to these factors that promote neural induction, cells should be plated on vessels coated with laminin as this promotes the formation of neurite extensions of P19 cells through the activation of ß1 integrin and CaMK-II (Easley et al., 2006; Nakayama et al., 2014). The further benefits of using a monolayer adherent culture are the homogenous induction of cells as most of cells are equally exposed to soluble factors and reagents, and the elimination of the proliferating non-neuronal cells from the culture during the neuronal maturation by adding cytosine arabinoside (AraC) that arrests the cell growth. The advantages of this induction method led us to apply this approach for characterising HMGN proteins during neuronal differentiation and investigating the role of HMGN2 during the differentiation.

The monolayer adherent method can be divided into four main steps including neural induction (first 2 days), axogenesis (from day 2 to 4), synaptogenesis (from day 4 to 6) and synapse maturation (from day 6 and onwards) (Nakayama et al., 2014). Figure 6.1 shows a schematic diagram for the process of neuronal differentiation.

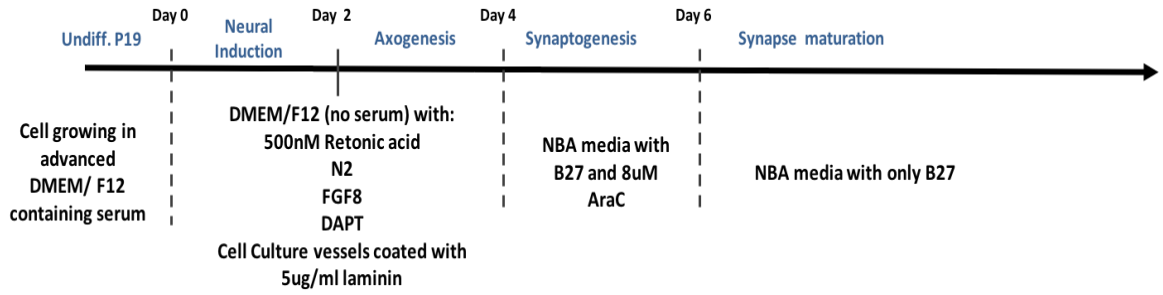


Figure 6.1: Time course of neural induction using monolayer adherent culture.
The undifferentiated P19 cells that grow in DMEM/F12 media containing serum are induced to neural lineage by withdrawing the serum and adding 500 nM retinoic acid, 10 ng/ml FGF8, 10 μM DAPT and N2 supplement. The cells were plated on dishes coated with 5 μg/ml laminin at day 0. The cells were grown in this condition for 4 days. Then, the media was changed into neural basal media containing B27 and 8 μM AraC for 2 days. The AraC was removed for the remaining of the process by changing the neural basal media with only B27 every 2 days. Undiff, DAPT, NBA and AraC stand for undifferentiated, N-[N-(3,5-Difluorophenacetyl-L-alanyl)]-(S)-phenylglycine t-butyl ester, neural basal media and cytosine arabinoside, respectively. N2 and B27 Supplements are recommended for growth and expression of NSCs and post-mitotic neurons in primary cultures, respectively. NBA media is used to grow neuronal cells from the hippocampus, cortex and other regions of the brain (adapted from Nakayama et al., 2014).

6.2 Objectives

The main aims of this chapter are to characterise the expression of HMGN1 and 2 during neural induction, and to identify the possible role of HMGN2 in regulating the expression of neural and neuronal-associated genes in undifferentiated P19 cells and during neuronal differentiation. Therefore, this can be divided into different objectives as following:

1- Characterise the neural induction of EC P19 cells using the adherent monolayer culture system.

2- Characterise HMGN1 and HMGN2 expression during the *in vitro* neuronal differentiation using the adherent monolayer culture system.

3- Identify whether the loss-of-function of HMGN2 leads to any differential expression of neural and neuronal-related genes at the undifferentiated state by comparing the expression between P19 wild-type cells and knockout lines produced in Chapter 3 and 4.

4- Study the role of HMGN2 protein during neuronal differentiation by inducing neuronal differentiation of the knockout lines, and comparing the neural-related markers between the wild-type and knockout lines.

6.3 Characterise the neural induction of EC P19 cells using adherent monolayer culture

6.3.1 Adherent monolayer culture efficiently produces NSCs and neurons from P19 ECCs

P19 wild-type cells were subjected to the neural induction using the monolayer adherent culture as described by Nakayama et al. (2014). The neural induction was firstly monitored by light microscopy (Figure 6.2). After inducing the P19 cells that have epithelial-like morphology down the neural lineage, the cells start to form an adherent cell aggregates at day 2. The cell aggregates are variable in size and do not have a defined structure. In addition, at day 2 some cells had neurite-like projections that extended outward and connected between the cell aggregates. The network of cellular processes was more apparent at day 6 after induction (Figure 6.2). At day 9, larger cell aggregates with multiple neuronal processes extending outwards were obvious. As the AraC was added at day 4 for 48 hours, many of proliferative cells died, as seen in the culture at day 6. These dead cells were washed out by changing the media on day 6. Therefore, the monolayer adherent neuronal differentiation system for P19 cells seems to be working efficiently, as an extensive network of projections was formed at day 6 and beyond.

In a separate experiment, the effect of seeding lower cell numbers, or using a cell strainer or both was investigated. However, these approaches did not reduce the formation of cell aggregates (data not shown). The formation of cells aggregates is likely to be a result of using the γ- secretase inhibitor (DAPT), since Nakayama and colleagues rule out the formation of these aggregates when they induced the neural induction without the use of DAPT. However, because Notch signalling is involved in maintaining the population of proliferating neuroepithelial progenitor (NPs), they observed that a large number of P19 cells continued to proliferate without having adherent cell aggregates, but only a small proportion of cells had neurite projections. Moreover, addition of γ-secretase inhibitor to the rosette cultures derived during adherent neuronal differentiation of mES cells resulted in changing the rosette structures into large rounded ganglion-like clusters with extensive neurite outgrowth (Abranches et

al., 2009). These structures are similar to the clusters seen here, especially at day 9.

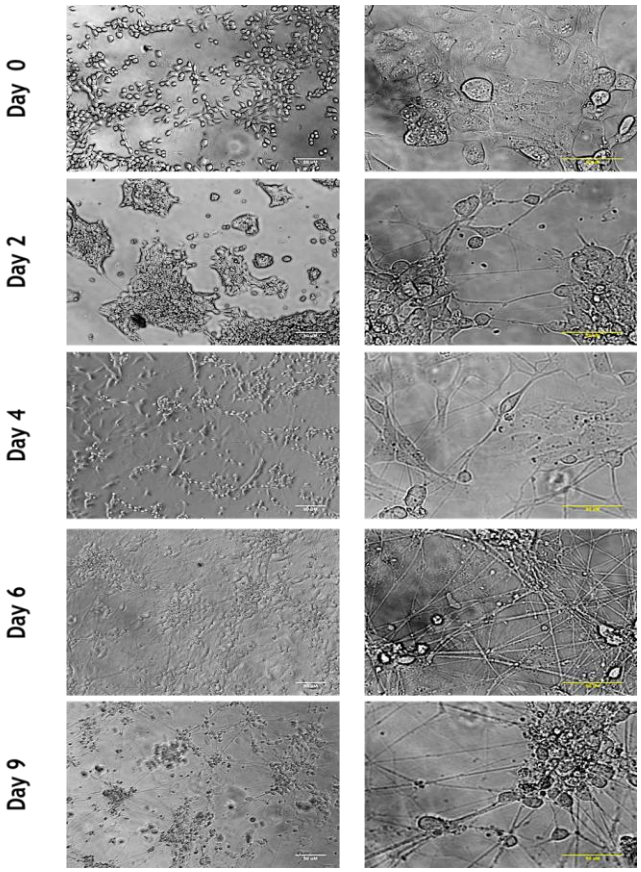


Figure 6.2: Light pictures of P19 wild-type cells during neural induction using monolayer adherent system.
Bright field images for P19 wild-type cells at day 0 (undifferentiated), 2, 4, 6 and 9 after induction. The P19 cells were induced by incubating the cells in DMEM/F12 media that contains 500 nM retinoic acid (RA),10 µM DAPT (gamma secretase inhibitor), 10 ng/ml FGF2 and N2 in absence of the serum. The cells were plated either in 24 well plate containing coverslips or 100 mm dishes that were previously coated with 5 µg/ml laminin. The cells were incubated in the DMEM/F12 media for 4 days and then changed with NBA media with B27 and 8 µM AraC for 2 days to kill the proliferative cells. After day 6 the media was changed every two days by NBA media with only B27. The undifferentiated P19 cells which have epithelial-like morphology start to form clusters by day 2 and the clusters are clearly seen by day 6 and 9. In addition, the cells projections were observed from day 2 and afterward. These projections that exit from the clusters form complex networks to connect the clusters together (day 6 and 9). The pictures were taken using the inverted microscope (Olympus IX5). Left panel pictures were taken under 20X while right panel under 40X. Scale bar, 50 µM.

The efficiency of neural induction and neuron production was also assessed by immunofluorescence (Figure 6.3) and immunoblotting (Figure 6.4) using different embryonic, neural stem cell (NSCs) and neuronal-specific markers. Oct4, an embryonic stem cell and pluripotency marker, was expressed in all cells before the induction. However, the nuclear signal of Oct4 staining was lost at days 3 and 10 after induction (Figure 6.3). The down-regulation of *Oct4* expression was confirmed by immunoblotting in Figure 6.4A, showing it is only detected in the

undifferentiated P19 cells and absent throughout the whole differentiation process until day 16. This suggests that most of the cells have lost their pluripotency state and have become more restricted. Consistently, another key regulatory protein for embryonic stem cells, Nanog, has also reduced during differentiation as shown by western blotting (Figure 6.4A). Therefore, this result confirmed that the P19 cells that were pluripotent before induction have lost their pluripotency potential after induction.

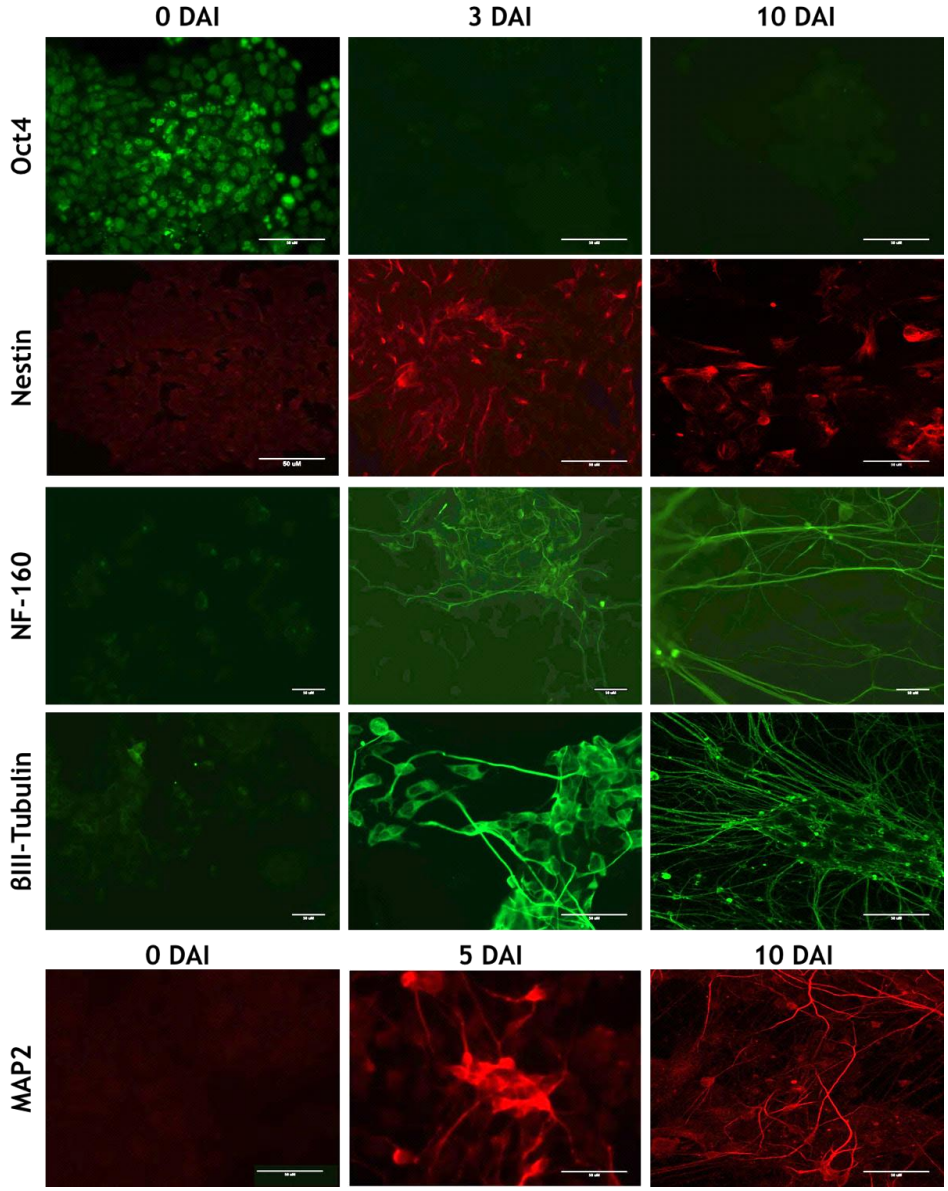


Figure 6.3: Efficient production of NSCs/NPCs and neurons from P19 ECCs using an adherent monolayer-serum free culture.
Immunofluorescence for Oct4, Nestin, Neurofilament 160 (NF-160) and β-III tubulin were performed at day 0, 3 and 10. MAP2 staining was performed in the undifferentiated (0 DAI) and 5 and 10 days after induction (DAI). The cells were plated on the coverslip in 24 wells plate till indicated days for staining. The pluripotent marker (Oct4) that is a transcription factor and found in the nucleus is seen positive only in cells before the induction but not after. However, the neural precursor (Nestin) and neuronal (NF-160, β-III tubulin and MAP2) markers were expressed after the induction although there is some positive faint signal in the undifferentiated P19 cells. The scale bar is 50 µM.

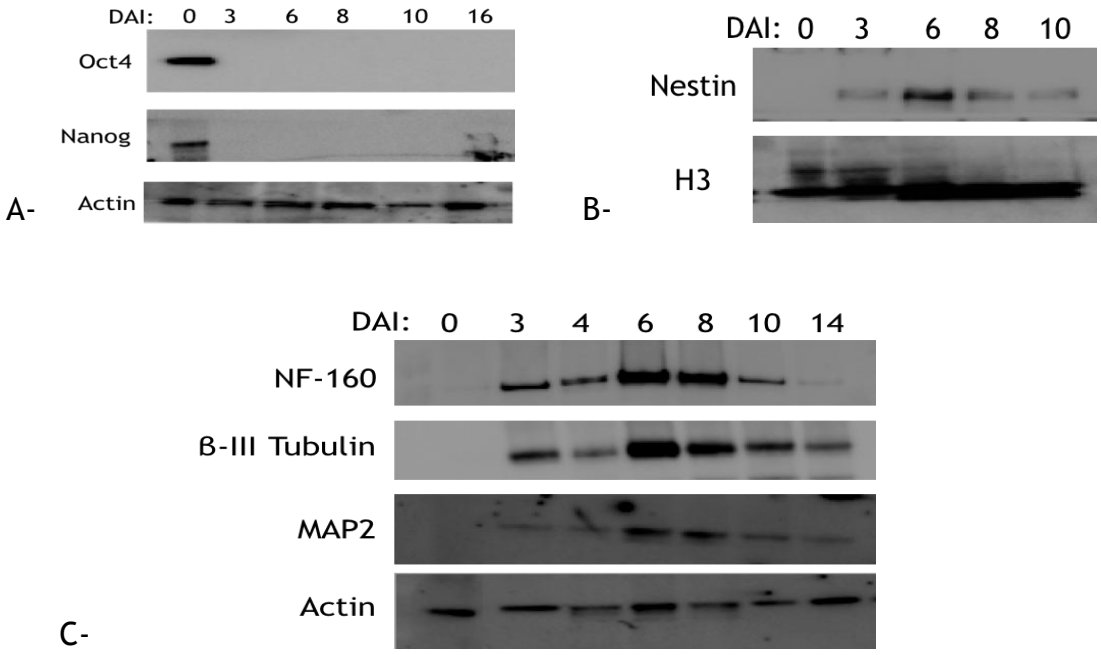


Figure 6.4: Immunoblotting for pluripotent, neural and neuronal-related proteins showing efficient production of NSCs and neurons from P19 ECCs using an adherent monolayer-serum free culture.
The protein level of pluripotent (Oct4 and Nanog), neural precursor (Nestin) and neuronal markers (MAP2, NF160 and β-III tubulin) were examined by western blot in A, B and C, respectively, for different time points during the neuronal differentiation. β-Actin and histon3 (H3) were used as internal controls. The data from western blotting is consistent with data from the immunofluorescence. In addition, the highest protein levels of all neural or neuron-specific proteins are at 6 and 8 days after induction. DAI is for days after induction. The band size for Oct4, Nanog, Nestin, MAP2, NF160 and β-III tubulin are 39, 35, 200, 70, 160 and 50 kDa, respectively.

On the other hand, Nestin, which is intermediate filament protein and marker of multipotent NSCs, was not detected in most undifferentiated P19 cells although a few cells had a faint signal (Figure 6.3). However, the Nestin signal is clearly apparent, with the expected pattern of mainly cytoplasmic intermediate filaments, at days 3 and 10 after induction. More Nestin positive cells were observed at day 3 compared to 10 days after induction. Similarly, the total level of Nestin protein, which was absent at day 0, significantly increased to reach a peak at day 6 and then was sharply reduced by day 10 (Figure 6.4B). Nestin is important for the survival and self-renewal of NSCs/NPCs during embryonic development, and is often reduced with the concomitant up-regulation of other tissue-specific intermediate filament proteins such as glial fibrillary acidic protein (GFAP) in astrocytes or neurofilament proteins in neurons (Park et al., 2010). The presence of some Nestin-positive cells at day 6 and 10 indicates that some cells are still NPCs/NSCs, and this may reflect heterogeneity within the system. Although Nestin is mainly considered as a NSC marker, Nestin can also be found in some more committed cells, including astrocytes (Clarke et al.,

1994). Immunofluorescence for GFAP during neural induction of P19 cells did not reveal $GFAP^+$ cells until day 10 (data not shown). This is consistent with the report by Nakayama et al. (2014), who showed that the addition of AraC restricts cells to the neuronal lineage, whereas omitting AraC leads to an increase in GFAP-positive cells.

To assess the production of neurons, different neuronal markers including ß-III tubulin (also known as Tuj1, a neuron-specific intermediate filament protein), NF-160 (neurofilament medium) and MAP2 (microtubule-associated protein) were studied (Figure 6.3 and Figure 6.4). These cytoskeleton proteins contribute to microtubule stability and functionality that are required for neurite formation and maintenance. Class III ß-tubulin is considered as one of earliest neuron-associated marker proteins because it is expressed either immediately before or during terminal mitosis in CNS (Menezes et al., 1994; Katsetos et al., 2003a). It is also distributed throughout the neuron and indiscriminately found in all neurons in the postnatal and mature brain (Menezes and Luskin., 1994). Its function is involved in the early neuritogenesis (Katsetos et al., 2003b). Map2 that is a member of MAP2/Tau family has three isoforms varying in the length of amino-terminal projection domain. The most predominant isoform of MAP2 is Map2c which is shorter than Map2a and Map2b (Kalcheva et al., 1995). Mammalian MAP2 is expressed mainly in neurons, but MAP2 immunoreactivity is also detected in some non-neuronal cells such as oligodendrocytes (Dehmelt and Halpain, 2004). Its expression is very weak in neuronal precursors and then becomes stronger after expression of ß-III tubulin (Menezes and Luskin., 1994). Directly after axonogenesis in developing cortical and hippocampal neuronal cultures, Tau gradually segregates into axons, whereas MAP2 segregates into the nascent dendrites which means that MAP2 is a marker for cell bodies and dendrites in mature neurons (Matus, 1990; Dehmelt and Halpain, 2004). Finally, neurofilaments are intermediate filaments and are composed of four subunits [neurofilament heavy (NFH), medium (NFM), light (NFL) and α-internexin] (Yuan et al., 2012). NFM or NF-160 is concurrently expressed with the beginning of neurite formation in the CNS (Nixon and Shea, 1992). Although the neurofilaments exist in perikarya and dendrites, they are particularly abundant in axons, where they are essential for the radial growth of axons during development, and the maintenance of axon caliber and the transmission of

electrical impulses along axons (Eyer and Peterson, 1994; Friede and Samorajski, 1970; Yum et al.,2009; Yuan et al., 2012).

At day 0, these neuronal-specific markers (MAP2, ß-III tubulin and NF-160) are almost negative in the majority of the undifferentiated cells, although there were few cells with a faint cytoplasmic signal (Figure 6.3). This is to be expected, especially for ß-III tubulin, because P19 cells sometimes display spontaneous differentiation, particularly at low cell densities (Paquin et al., 2002). However, there was no band for ß-III tubulin detected by immunoblotting at day 0 (Figure 6.4C). After the neural induction of P19 cells, robust staining was observed in the cell bodies and neurite-like cell processes at day 3 and 10 for ß-III tubulin and NF-160, and at days 5 and 10 for MAP2 (Figure 6.3). The intensity of staining for these markers is high in both cell bodies and processes at day 3, whereas by day 10 the staining in cell bodies is less intense than in the neuronal processes. The protein levels of MAP2, ß-III tubulin and NF-160 were also investigated by immunoblotting (Figure 6.4C). Though these proteins are detected from day 3 and onwards, the highest levels of these proteins were obtained at days 6 and 8, then they reduced at day 10 and day 14 (Figure 6.4C).

The expression of ß-III tubulin, NF-160 and MAP2 confirmed that immature and mature neurons were obtained from day 3 using this adherent differentiation method. According to Nakayama et al. (2014), this method of driving neurons from P19 cells results in mature and functional neurons within 6 days that were responsive to several neurotransmitters suggesting diverse subtypes of neurons. The reduction in neuronal-related proteins after day 8 could be due to the use of AraC, since it was previously reported that AraC is cytotoxic for cerebellar and sympathetic neurons in dose-dependent pattern (Dessi et al., 1995; Besirli et al., 2003). The neurotoxic effect of AraC is associated with the induction of reactive oxygen species that trigger DNA damage and neuronal apoptosis (Geller et al., 2001). Nevertheless, the AraC concentration used here is very low and is not expected to cause significant neuronal cytotoxicity.

Collectively, a significant increase in markers for NSCs and neurons, alongside the reduction in the levels of Oct4 and Nanog, indicate that this method drives the efficient production of NSCs and neurons, but not glia cells, from

undifferentiated P19 cells. Furthermore, the highest production of NSCs and neurons is fairly synchronous, mainly at day 6 after induction.

6.3.2 AraC does not kill all proliferative cells in adherent monolayer culture system

This neural induction method tends to produce a more homogenous cell population because mitotic cells are eliminated by the addition of AraC from day 4 to 6. AraC is incorporated into DNA of mitotic cells and inhibits DNA replication by inducing DNA cleavage mediated by topoisomerase I (Kufe et al., 1980; Cline et al., 1999). However, the presence of more than one cell type was noticed in the culture, and the cell confluence was still increasing with the time, and was especially apparent by day 16. Accordingly, it was important to investigate the presence of proliferative cells in the culture, which can be done by assaying for DNA synthesis. The traditional method for measuring DNA synthesis was based on the use of the radioactive nucleoside 3H-thymidine, but because it is time-consuming and hazardous it was replaced by antibody-based detection of the nucleoside analog bromo-deoxyuridine (BrdU), which is incorporated into newly synthesised DNA. This procedure requires a DNA denaturation step to expose the incorporated BrdU to the antibody, and this step can disrupt the integrity of the DNA and affect the cell morphology and antigen recognition sites (Wakayama et al., 2015). An alternative to BrdU, EdU (5-ethynyl-2′-deoxyuridine), is also a nucleoside analog of thymidine that is incorporated into DNA during active DNA synthesis (Life Technologies). Unlike BrdU that is detected by an antibody, the incorporated EdU is detected by the click chemical reaction that relies on a copper catalysing covalent reaction between the ethynyl group of the incorporated EdU in the double-stranded DNA and the Alexa Fluor 488 azide, forming the triazole bond (Life Technologies).

A proliferation assay was performed at different time points during the neuronal differentiation of P19 cells. Firstly, all the undifferentiated P19 cells were positive for EdU incorporation as expected (Figure 6.5). However, at day 3, some cells were clearly negative for EdU incorporation, which means they became no longer proliferative i.e. post-mitotic. The cells that were proliferative at day 3 are likely to be neural stem or progenitor cells (NSC/NPCs) because many cells were Nestin positive and Oct4/Nanog negative at that time.

At day 7 and 10 (after AraC treatment), most of the cells inside large aggregates are negative for EdU, and are therefore not proliferative (Figure 6.6, arrows). However, some cells outside the aggregates were positive for EdU. Some of these EdU positive cells were large, flattened cells with large nuclei and weak DAPI staining (Figure 6.5, day 7 arrow heads), whereas others had a more regular size (Figure 6.5, day 10). Preliminary data suggests that the large flattened cells are not Nestin-positive (not shown), but further characterisation of these EdU positive cells is required to determine whether they are from the neural lineage, or are a different type of cell. These results show that AraC does not eliminate all the proliferative cells in the culture, but by reducing the number of proliferative cells, it enriches the proportion of post-mitotic neuronal-like cells.

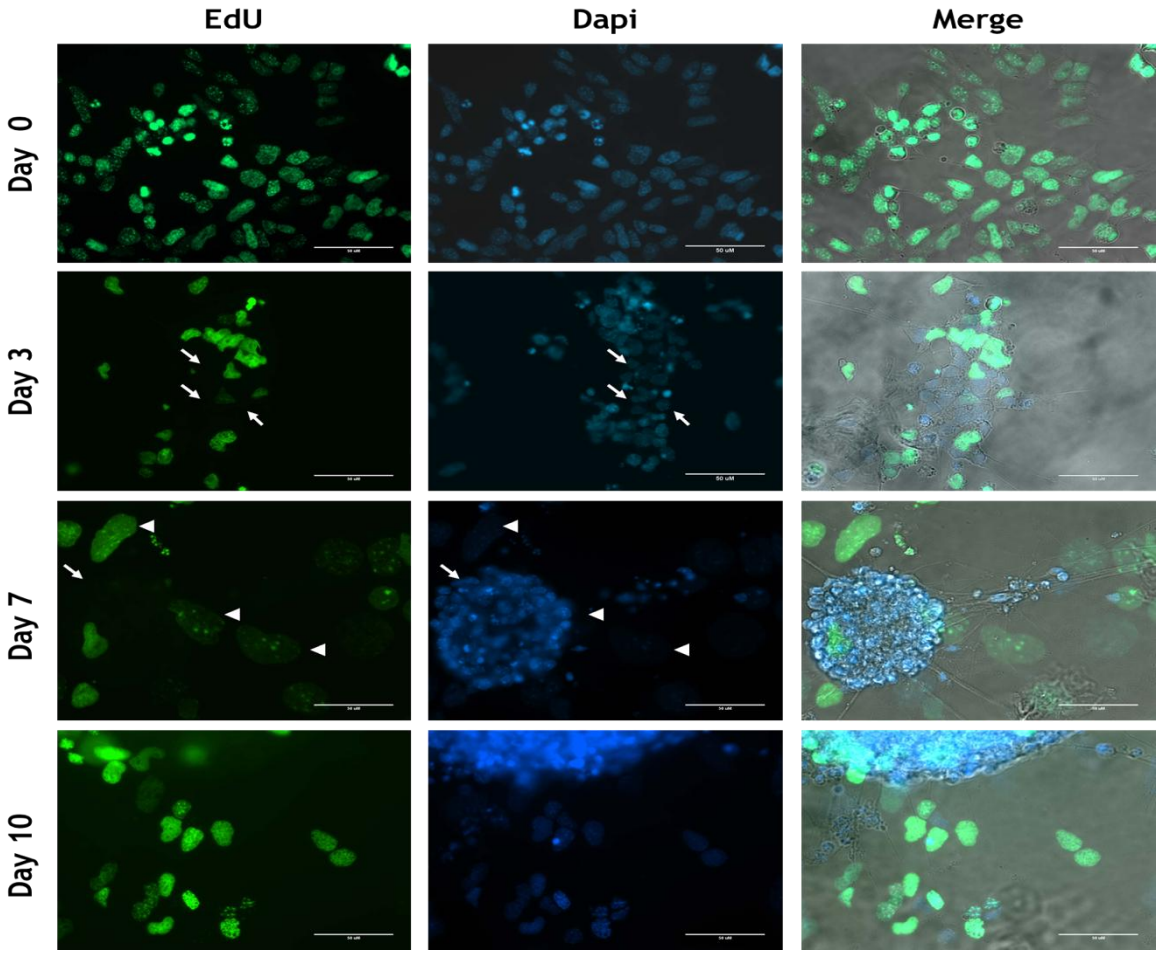


Figure 6.5: The proliferation assay, EdU, to detect the proliferative cells during the process of neural induction.
The cells were plated on the coverslip in 24 wells plate until indicated days for harvesting. 10 µM EdU was added to live cells for 24 hours and labelled using Click-iT Plus Alexa Fluor 448 Imaging Kit and Dapi nucleic acid stain (blue). Different time points were selected for the proliferation assay and include undifferentiated P19 cells (day 0) and 3, 7 and 10 days after induction. Coverslips were mounted with Prolong Gold Antifade Mountant and imaged using Olympus inverted microscope. Proliferating cells have green nuclei. The last column is a merge with the bright field image. Scale bar, 50 µM. Arrows are indicators for some negative cells while arrow heads are for EdU^{+} and faint DAPI.

6.4 Characterisation of HMGN1 and 2 expressions during neuronal differentiation *in vitro* using adherent monolayer culture

HMGN1 and HMGN2 have been characterised during the cellular differentiation processes such as erythropoiesis, myogenesis, osteoblast differentiation and kidney organogenesis (Crippa et al., 1991; Begum et al., 1990; Shakoori et al., 1993; Lehtonen and Lehtonen, 2001), but not during neural differentiation. HMGN3 is expressed in a tissue-specific manner, and is thought to be particularly highly expressed in glial cells (Ito and Bustin, 2002). Although HMGN4 is closely related to HMGN2 and is believed to have similar functions (Birger et al., 2001), HMGN5 has a longer acidic C-terminal domain that makes it dissimilar from all the other members, and has been extensively studied in cancer cells (Rochman et al., 2005; Chen et al., 2012; Gan et al., 2015). After extensive validation of neural induction and neuronal production by using an adherent serum-free culture, the most abundant HMGN family members, HMGN1 and HMGN2, were assessed during neuronal differentiation. From the previous section, although NSC/NPCs and neurons can be found 3 days after induction using serum-free monolayer adherent culture, the majority are produced at day 6.

Firstly, the level of HMGN1 protein, which seems to be less abundant in the undifferentiated P19 cells than HMGN2, is increased after neuronal induction till day 10, and then sharply reduced at day 14 and 16 (Figure 6.6). HMGN2, on the other hand, is gradually increased until day 8 but reduced at days 10, 14 and 16 as shown in western blot data from two biological replicates (Figure 6.6). In consistent with this, previously in our lab it was shown that the level of HMGN2 protein and RNA was increased, but HMGN1 was unchanged, up to day 6 using the EBs method (Mohan thesis, 2012).

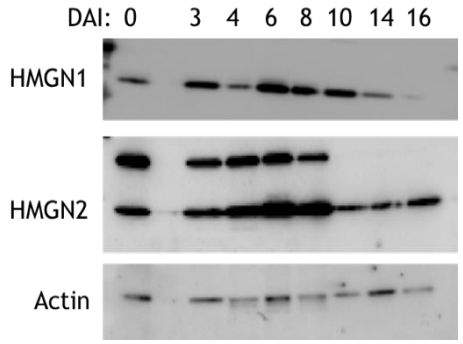


Figure 6.6: The level of HMGN1 and 2 proteins during neuronal differentiation.
The protein level of HMGN1 (11 kDa) and HMGN2 (17 kDa) in the undifferentiated wild-type P19 cells (0 DAI) and during neuronal differentiation (3, 4, 6, 8, 10, 14 and 16 DAI) were analysed using adherent monolayer culture. The protein used in these analyses are from the same biological samples used in characterising the cell-specific markers. HMGN2 showed a non-specific band at 25 kDa (upper band). β-Actin (42 kDa) was also used as internal control for western blot. DAI refers to days after induction. This data is reprehensive of two independent experiments.

Immunostaining for HMGN2 was performed to investigate whether HMGN2 is expressed in all cell types during the differentiation process. Co-staining for Nestin was also performed, as this should mark the NSC/NPC population (Figure 6.7). At day 0, all cells were equally stained by HMGN2 antibody but negative for Nestin, and the culture was very homogeneous (Figure 6.7). At day 3, the majority of cells express both HMGN2 and Nestin, but the intensity of HMGN2 staining is quite variable between cells, suggesting some heterogeneity in the culture.

At day 16, a reduction in HMGN2 expression was seen in many, but not all cells within the culture (Figure 6.7). This was often particularly apparent in the cell aggregates with extended neuronal processes, as observed at day 10 (Figure 6.8), and at day 16 by confocal microscopy (S. Manero, data not shown). This is consistent with the previous western blotting data, which showed a substantial reduction in HMGN2 protein at day 10. The high cell density and complexity of the culture at day 16 makes it hard to draw firm conclusion about the relationship between Nestin and HMGN2 expression.

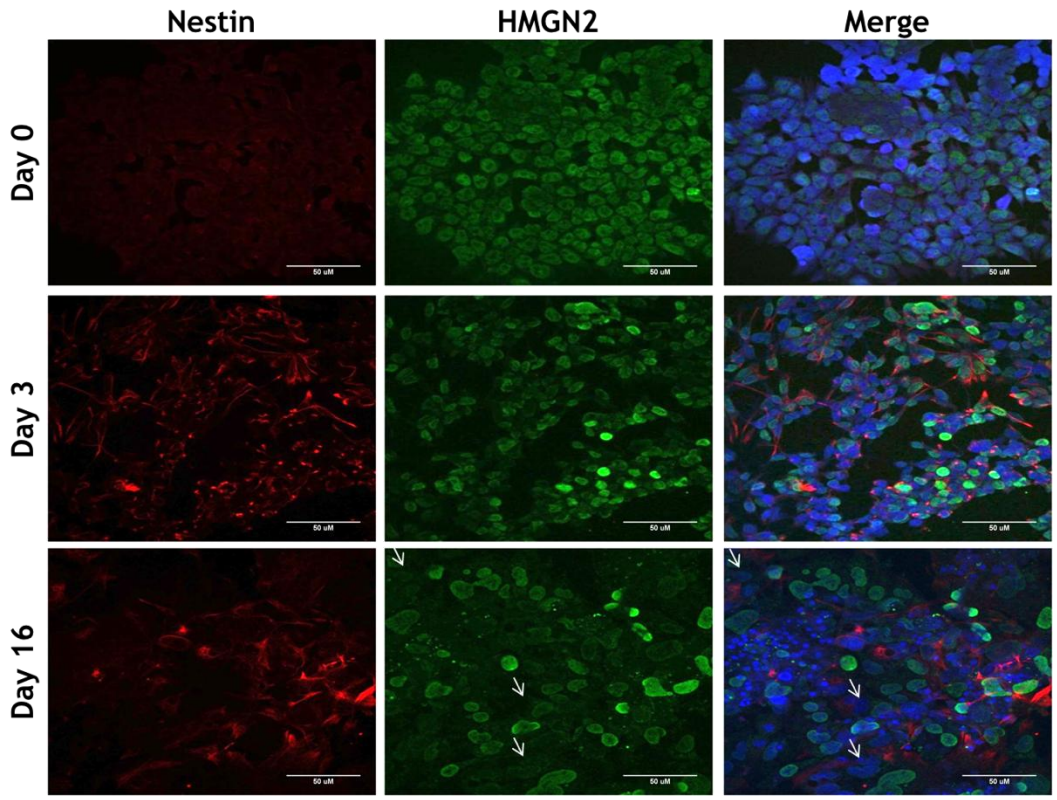


Figure 6.7: Double immunofluorescent staining for HMGN2 and Nestin at 0, 3 and 16 days during differentiation.
HMGN2, which has a nuclear localisation and the signal is comparable with Dapi (blue), is green signal whereas Nestin, which is type VI intermediate filament and cytoplasmic, is in red. Nestin is negative in the undifferentiated P19 cells but positive at day 3 and less at day 16. The majority of Nestin positive cells are HMGN2 positive either at day 3 or 16. At day 16, some cells are clearly either HMGN2 negative (arrows) or with faint HMGN2 signal. Scale bar, 50 µM. This differentiation and immunostaining were performed with collaboration with another Ph.D. student (Garza-Manero S.) but figures were taken by Garza-Manero S.

In order to investigate whether HMGN2 expression correlates with cell proliferation, dual staining for HMGN2 and EdU was performed on cells at day 10 (Figure 6.8). Many of the aggregates showed no proliferative cells, and were negative or very low signal for HMGN2 expression, whereas other, more isolated cells were both EdU and HMGN2 positive (Figure 6.8). The result suggests that the proliferative cells are highly HMGN2 positive but the post-mitotic neurons are either low or negative for HMGN2.

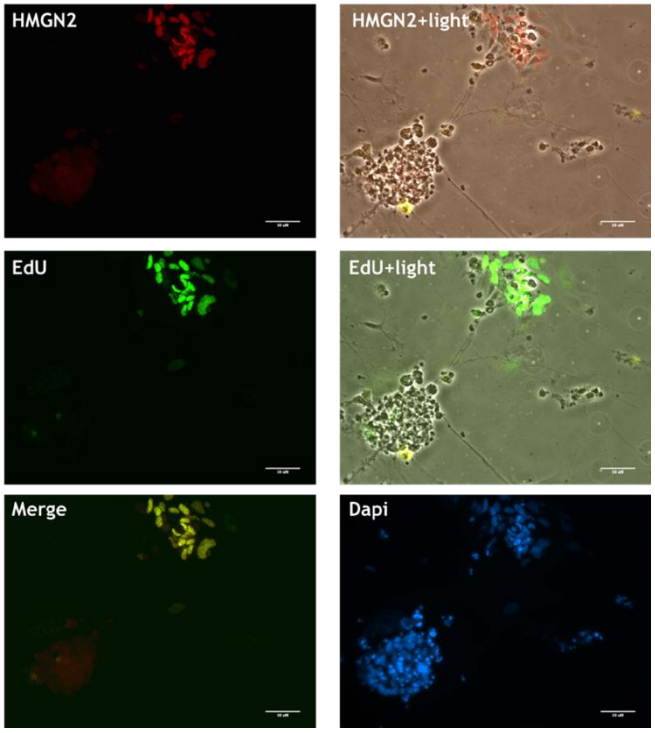


Figure 6.8: The HMGN2 staining and EdU labelling for detecting the proliferative cells performed at day 10 after neural induction.
The pre-incubated cells with the alkine-containing thymidine analog EdU (5-ethynyl-2'-deoxyuridine), for 24 hours were fixed with 3.7% formaldehyde and detected by a click reaction (a copper-catalysed azide-alkyne cycloaddition) using a fluorescent Alexa Flour 448 (green) that contains a picolyl azide moiety. In addition, the cells were also stained with HMGN2 antibody and secondary antibody (red) as exactly immunofluorescent protocol without fixation and permeabilization. The coverslips were mounted with Prolong Gold Antifade Mountant with Dapi (blue) and imaged using Olympus inverted microscope. The merged figures are either the red and green in the last figure of the left panel or red or green alone with the bright field images (the first two image on the right panel). This is representative of two figures. Scale bar, 50 µM.

In summary, HMGN2 is highly expressed during the early stages of *in vitro* neuronal differentiation when the culture is enriched in proliferative NSCs/NPCs, but it appears to be reduced in the post-mitotic neurons at later stages of the differentiation. Further investigations using dual labelling and confocal microscopy are required to confirm this result and to further characterise the proliferative, HMGN2 positive cells present at later stages in the culture. The increase of HMGN1 and 2 during the neural induction *in vitro* is comparable with high RNA expressions *in vivo* from the Allen Brain Atlas. In addition, as shown later in this chapter the RNA expression of *Hmgn1* and *2* during the another neural induction is increased to 2-3 folds at days 3 and 6 after induction of wild-type P19 cells (Figure 6.15). On the other hand, the reduction of HMGN1 and 2 proteins after day 10 and 8, respectively, can be as a result of using AraC that kill the proliferative cells. In addition, it also reflects the nature of their expression in post-mitotic neurons, as they were reported to be reduced at the

late stage of differentiation in different tissues, such as in erythroid cells (Kulkeaw et al., 2012).

6.5 The role of HMGN2 in the neural lineage commitment

6.5.1 Loss of HMGN2 does not affect the gross morphology of undifferentiated P19 cells

Before investigating whether the loss of HMGN2 affects neuronal differentiation it was important to compare the morphology between the HMGN2 knockout lines and controls. The parental P19 cells and the negative control line (B22) had typical morphological features of EC P19 cells. They grow as a monolayer of adherent epithelial-like cells with a high ratio of nucleus to the cytoplasm (Figure 6.9). In addition, very small protrusions are extended from cytoplasm in different directions. The complete HMGN2 knockout lines (B8ΔHMGN2, B19ΔHMGN2 and B38ΔHMGN2) and line that has a mutated HMGN2 (19ΔNBD) appear to have similar morphological features as the parental and B22 line (Figure 6.9). The only difference is seen in 19ΔNBD line in which the cells have a tendency to grow more as colonies than in the other lines (Figure 6.9). Therefore, there were no major morphological differences between the HMGN2 knockout lines and the wild-type.

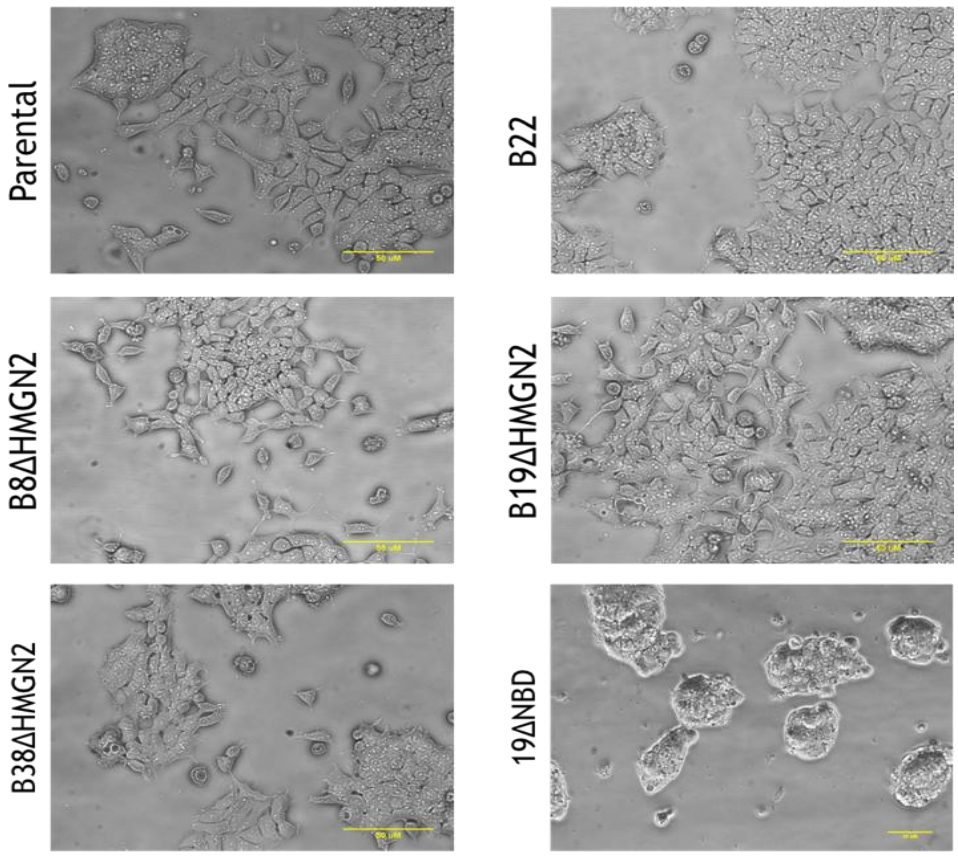


Figure 6.9: Bright field images for wild-type P19 cells and different HMGN2 knockout monoclonal lines in the undifferentiated state.
The B22 line is a negative control that has undergone the process of monoclonal derivation. The images were taken from live cells cultured on 100 mm dishes. ΔHMGN2 means the complete removal of HMGN2 protein while ΔNBD refers to the disrupted NBD of HMGN2. Scale bar, 50 µM.

6.5.2 Lack of complete HMGN2 induces the spontaneous neuronal differentiation of undifferentiated EC P19 cells

6.5.2.1 Loss of HMGN2 up-regulates Nestin expression in undifferentiated P19 cells

The unchanged morphology and *Oct4* and *Sox2* expression in HMGN2 knockout and mutant cells suggest that most of the cells are still pluripotent and similar to undifferentiated P19 wild-type. However, the significant reduction in *Nanog* expression in the HMGN2 knockout cells might be an indicator for a reduction in pluripotency. This contrasts with the upregulation of *Nanog* expression in the cells expressing mutant HMGN2 (19ΔNBD). According to Mitsui et al. (2003), Nanog is required to sustain the pluripotency in mouse ESCs, and Nanog-deficient ES cells lost pluripotency and differentiated into extraembryonic endoderm lineage.

In order to investigate whether markers of early differentiation down the neural lineage were misregulated in HMGN2 knockout or mutant cells, qRT-PCR was performed on the same RNA samples that were analysed previously. Expression of the neural stem cell marker, *Nestin*, was increased by almost 3.5-fold in two HMGN2 knockout lines, B8ΔHMGN2 and B38ΔHMGN2 (*P = 0.022* and *0.030*, respectively) (Figure 6.10). Although the third HMGN2 knockout line, B19ΔHMGN2, also revealed up-regulation of *Nestin* by 2.3-fold, the increase was not statistically significant (*P = 0.096*). *Nestin* expression in the control lines (B1 and B22), and the lines with mutated HMGN2 (19ΔNBD and S27ΔNBD) was not significantly increased.

This increase of *Nestin* expression and the reduction of *Nanog* expression suggests that the P19 cells with no HMGN2 may lose their self-renewal capability more than the wild-type and controls, and even more than the cells with mutant HMGN2. Although Nestin is used to identify the neuroepithelial stem cells (neural stem/progenitors of the developing nervous system), Nestin is also abundant in ES-derived progenitor cells that have potential to develop into neuroectodermal, endodermal and mesodermal lineage (Wiese et al., 2004). Therefore, the up-regulation of *Nestin* in HMGN2$^{-/-}$ P19 cells is not necessarily associated with the neural fate, but it may be a sign for progenitor cells from other lineages in the culture.

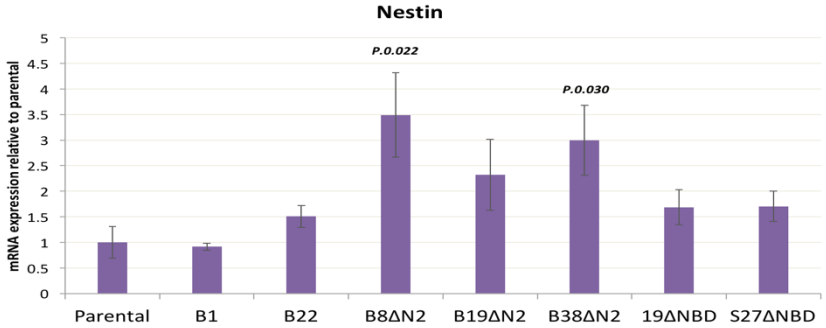


Figure 6.10: *Nestin* expression (neural precursor marker) is up-regulated in two complete HMGN2 knockout lines comparing to the parental cells and other negative control in the undifferentiated cells.
Nestin mRNA expression in parental (wild-type), B1 and B22 lines (negative controls), and B8ΔN2, B19ΔN2, B38ΔN2, 19ΔNBD and S27ΔNBD lines were determined. Relative gene expression in each cell line is the average from three biological replicates. Data was normalised to three housekeeping genes (*Gpi1*, *Rplp0* and *Gapdh*) and is plotted relative to that in parental cells (see methods for calculation). Error bars reflect the standard deviation from the fold changes of the three biological replicates. The P-values were statistically analysed from 3 biological replicates compared to parental by Student's t-test. The T-test was performed for all samples but only the significant values were represented in the graph.

6.5.2.2 Key proneural genes are up-regulated by loss of HMGN2 in P19 cells

In order to identify whether the high *Nestin* expression is related to the neural lineage or other lineages, the expression of other genes that are involved in neural-lineage determination was investigated. Several basic helix-loop-helix (bHLH) transcription factors have been shown to play role in regulation of neural stem cells. They are also implicated as mediators of neuronal or oligodendrocyte differentiation in the developing CNS. Although the *Olig1* and *Olig2* are the main factors for oligodendrocyte determination, *neurogenin1* and *2* (*Ngn1* and *Ngn2*), *Mash1*, and *NeuroD* promote neuronal differentiation (Sun et al., 2001; Ohtsuka and Kageyama, 2010).

Proneural BHLH factors can be divided into repressor or activator subtypes. The bHLH repressors such as *Hes* family proteins negatively regulate neuronal differentiation and maintain neural stem cell identity, whereas activator-type bHLH proteins that include *Ngns*, *Mash1*, *Math* and *NeuroD* promote neurogenesis (Ohtsuka and Kageyama, 2010). These activator and repressor bHLH factors act as heterodimers with the ubiquitously expressed E protein and preferentially bind to different DNA sequences. *Hes1* displays a higher binding affinity for the N-box (CACNAG) and C site (CACGCG) sequences and is also shown to repress the expression of neurogenic bHLH factors, specifically *Mash1* (Ross et al., 2003). On the other hand, the others type of bHLH factors bind with higher affinity to E-box (CANNTG) (Lee, 1997). The expression of *Hes1* and *5* is well known to be regulated by Notch signalling pathway, in which Notch activation results in the up-regulation of *Hes1* and *5* expressions (Ohtsuka and Kageyama, 2010). Here, we focused on the second type of bHLH factors because of their potential in promoting neurogenesis.

The activator bHLH factors (*Mash1*, *Ngn1* and *NeuroD*) can be detected early during embryogenesis (E8.5 and E9.5) (Lee, 1997). *Ngn1* and *Mash1* are expressed in neural precursor cells while *NeuroD* is found in post-mitotic precursors or differentiating neurons (reviewed in Ross et al., 2003; Kim et al., 2004). *Ngn1* and *2* are expressed in the cortical ventricular zone during development, which is where neuroepithelial precursor cells reside and neurogenesis occurs for a limited period of time (Sun et al., 2001). In many embryonic CNS regions, *Ngn1* and *Mash1* are expressed in complementary

domains that exhibit adjacent and non-overlapping boundaries, suggesting that they are involved in the development of distinct neuronal populations (Ma et al., 1997). For example, *Mash1* is expressed in the ventral telencephalon while *Ngn2* in dorsal telencephalon of forebrain and, in turn, they specify the GABAergic and glutamatergic neurons, respectively (Wilson and Rubenstein, 2000). As Ngn1 expression precedes NeuroD, it was found that Ngn1 is an up-stream regulator of NeuroD, which suggests Ngn1 is a determination factor while NeuroD is a differentiation factor (Seo et al., 2007). Thus, the ectopic expression of these genes in NSCs leads to down-regulation of *Nestin* and induction of neuronal markers, indicating to the determination of neuronal fate and acceleration the neurogenesis (Lee, 1997; Sun et al., 2001). Proneural bHLH factors, particularly *Ngn1*, also play a role in inhibiting astrocyte differentiation by sequestering the p300/Smad activator complex away from the *GFAP* promoter (Sun et al., 2001).

In our study, *Ngn1*, *Mash1* and *NeuroD* had very low expression in undifferentiated parental P19 cells, consistent with previous reports (Itoh et al., 1997). *Ngn1* expression was slightly increased in the negative lines B1 and B22 compared to wild-type parental cells. However, the complete removal of HMGN2 in the B8ΔHMGN2, B19ΔHMGN2 and B38ΔHMGN2 lines substantially up-regulated *Ngn1* expression by ~60, 15 and 35-fold, respectively. *Ngn1* expression did not change in HMGN2-mutant line 19ΔNBD.

Mash1 was also substantially up-regulated in HMGN2$^{-/-}$ cells, with increases of 64, 8 and 20-fold in the B8ΔHMGN2, B19ΔHMGN2 and B38ΔHMGN2 lines, respectively. *NeuroD* was also up-regulated in these lines, but the increase was smaller (2 – 4 fold), and there also appeared to be a 2-fold increase in the 19ΔNBD line.

In summary, the loss of HMGN2 in undifferentiated cells resulted in the up-regulation of key proneural genes, *Ngn1*, *Mash1* and *NeuroD*, which play roles in inducing the neuronal lineage specification. In contrast, mutation of the HMGN2 NBD does not affect *Ngn1* or *Mash1* expression.

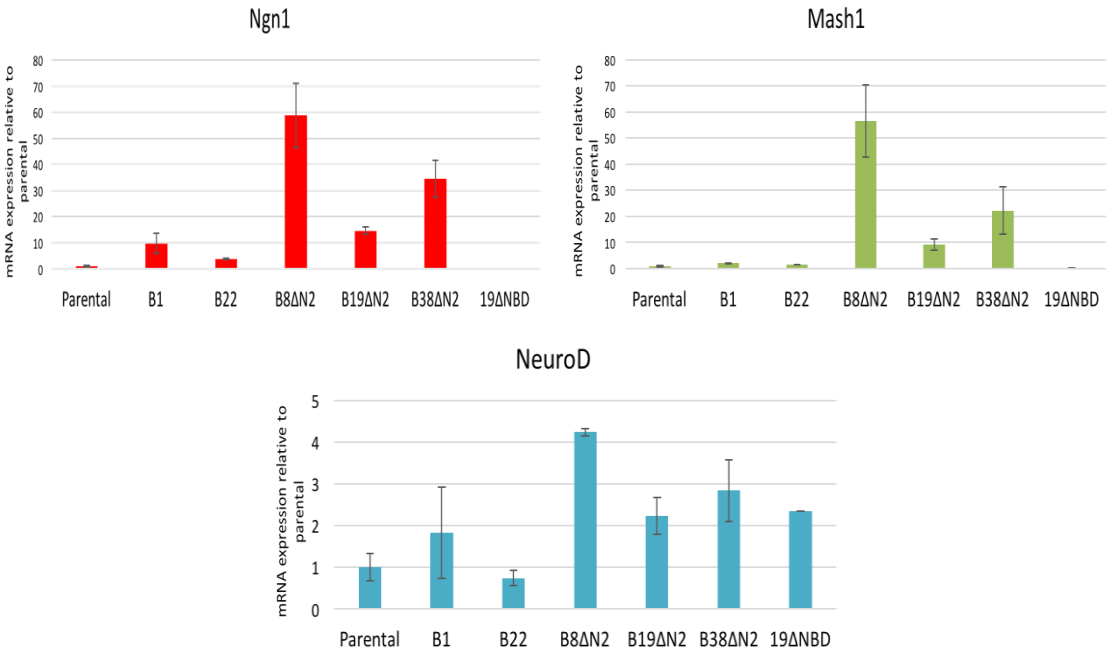


Figure 6.11: The expression of proneural bHLH transcription factors are increased in the HMGN2 knockout lines.
Ngn1, *Mash1* and *NeuroD* expression was compared from new RNA samples of parental cells (wild-type), negative control for HMGN2 loss (B1 and B22 lines), and in complete or partial HMGN2 knockout lines (B8ΔN2, B19ΔN2, B38ΔN2 and 19ΔNBD). Relative gene expression in each cell line is the average from two biological replicates, except 19ΔNBD which is from one biological replicate. Data was normalised to housekeeping gene (*Gpi1*) and is plotted relative to that in parental cells (see methods for calculation).

6.5.2.3 Neuronal marker genes are up-regulated in HMGN2$^{-/-}$ cells

Since several proneural genes and *Nestin* were up-regulated in all HMGN2$^{-/-}$ lines, which may mean more specification of undifferentiated P19 cells down the neural lineage, the expression of two neuronal-specific genes were investigated.

The expression of *Map2* was elevated by 7, 4 and 3-fold in B8ΔHMGN2, B19ΔHMGN2 and B38ΔHMGN2 cells, respectively, compared to the parental line (Figure 6.12). Similar to *Map2*, the *Nf160* expression is also up-regulated in HMGN2$^{-/-}$ lines (by 2.5 – 3.5 fold) but it is also slightly increased in the 19ΔNBD line and the B22 negative control (Figure 6.12).

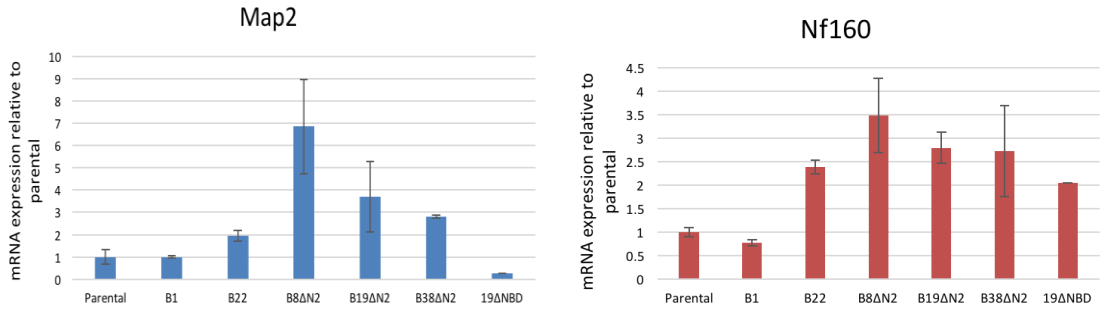


Figure 6.12: Up-regulation of neuronal-specific genes in the HMGN2$^{-/-}$ lines.
The expression of *Map2* and *Nf160* was compared between the lines: the parental cells (wild-type), negative control for HMGN2 loss (B1 and B22 lines), and in knockout or mutant HMGN2 lines (B8ΔN2, B19ΔN2, B38ΔN2 and 19ΔNBD). Relative gene expression in each cell line is the average from two biological replicates, except 19ΔNBD which is from one biological replicate. Data was normalised to housekeeping gene (*Gpi1*) and is plotted relative to that in parental cells (see methods for calculation).

There is clearly some variation in the levels of these genes in the parental and negative control lines, which could arise due to the process of clonal derivation. Furthermore, it is important to be aware of the risk of "false positive" results, where the testing of multiple genes in several cell lines increases the possibility that some of them appear to be up-regulated simply by chance. There are statistical methods to deal with multiple testing, such as the Bonferroni correction and the Benjamini–Hochberg procedure (McDonald, J.H. 2014. Handbook of Biological Statistics (3rd ed.). Sparky House Publishing, Baltimore, Maryland.). In this case, statistical significance for each line cannot be calculated as only two biological replicates were performed. Nevertheless, we consider that consistent upregulation of *Ngn1*, *Mash1* and *Map2* in three different knockout lines strongly suggests that that these changes in proneural and neuronal-specific gene expression are associated with the loss of HMGN2.

The qRT-PCR data cannot reveal whether the pro-neural and neuronal markers genes are up-regulated in all HMGN2$^{-/-}$ P19 cells, or whether a sub-population have started to differentiate into neurons. To address this question, immunostaining for ß-III tubulin and MAP2 was carried out in parental, B8ΔHMGN2, B19ΔHMGN2 and B38ΔHMGN2 cells. The result in Figure 6.13 demonstrates that although parental P19 cells have very few cells that are positive for ß-III tubulin, the HMGN2 knockout lines (B8ΔHMGN2, B19ΔHMGN2 and B38ΔHMGN2) have a substantially higher number of ß-III tubulin-positive cells with projections resembling neuritis. Correspondingly, the MAP2 staining showed

similar findings to ß-III tubulin, in which more neuron-like cells were found in the HMGN2 knockout lines (Figure 6.13).

Therefore, these data together strongly suggest that the upsurge in expression of proneural and neuronal-specific genes is a result of more cells become committed and losing their self-renewal potential, rather than an increase in expression in all cells. The expression of mesodermal (*Brachyury*) and endoderm (*Gata4*) marker genes did not show differential expression between the complete knockout and control (Garza-Manero S, personal communication). This data suggests that the level of spontaneous neuronal differentiation is increased in the HMGN2 knockout cells.

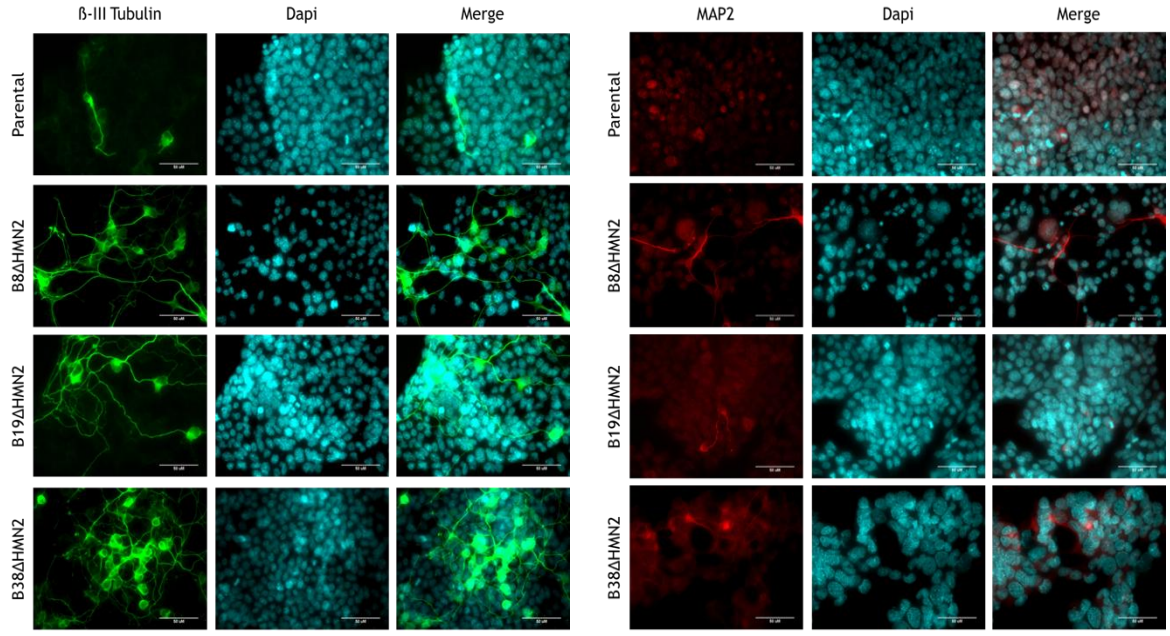


Figure 6.13: The ß-III tubulin and MAP2 staining in undifferentiated HMGN2 knockout lines and the parental P19 cells.
The undifferentiated cells were grown on coverslips for 2 days and stained with antibodies against the β-III tubulin (left panel) and MAP2 (right panel). It is clearly that the parental P19 cells have some positive cells for β-III tubulin and MAP2. However, the complete HMGN2 knockout lines showed more projection formed comparing the wild-type, seen in many fields. Coverslips were mounted with Prolong Gold Antifade Mountant with Dapi (nuclear stain, blue) and imaged using Olympus inverted microscope. Scale bar, 50 µM. (This immunofluorescence staining and the images were performed by Garza-Manero S. and Bailo M.).

6.5.3 Loss of HMGN2 alters induced neuronal differentiation in EC P19 cells

In order to investigate the role of HMGN2 in directed neuronal differentiation, the P19 lines were induced to differentiate using the monolayer adherent culture system described previously. All HMGN2$^{-/-}$ lines (B8ΔHMGN2, B19ΔHMGN2

and B38ΔHMGN2), the mutant HMGN2 line (19ΔNBD) and the parental P19 line were subjected for neural induction and the RNA and protein samples were collected at three-time points, which were 0, 3 and 6 DAI. As previously observed, cell aggregates with extended neurite-like projections were observed from day 2. An extensive network of aggregates and processes was clearly seen by day 6 in all lines (Figure 6.14). There were no apparent differences between the knockout lines and the parental P19 cells.

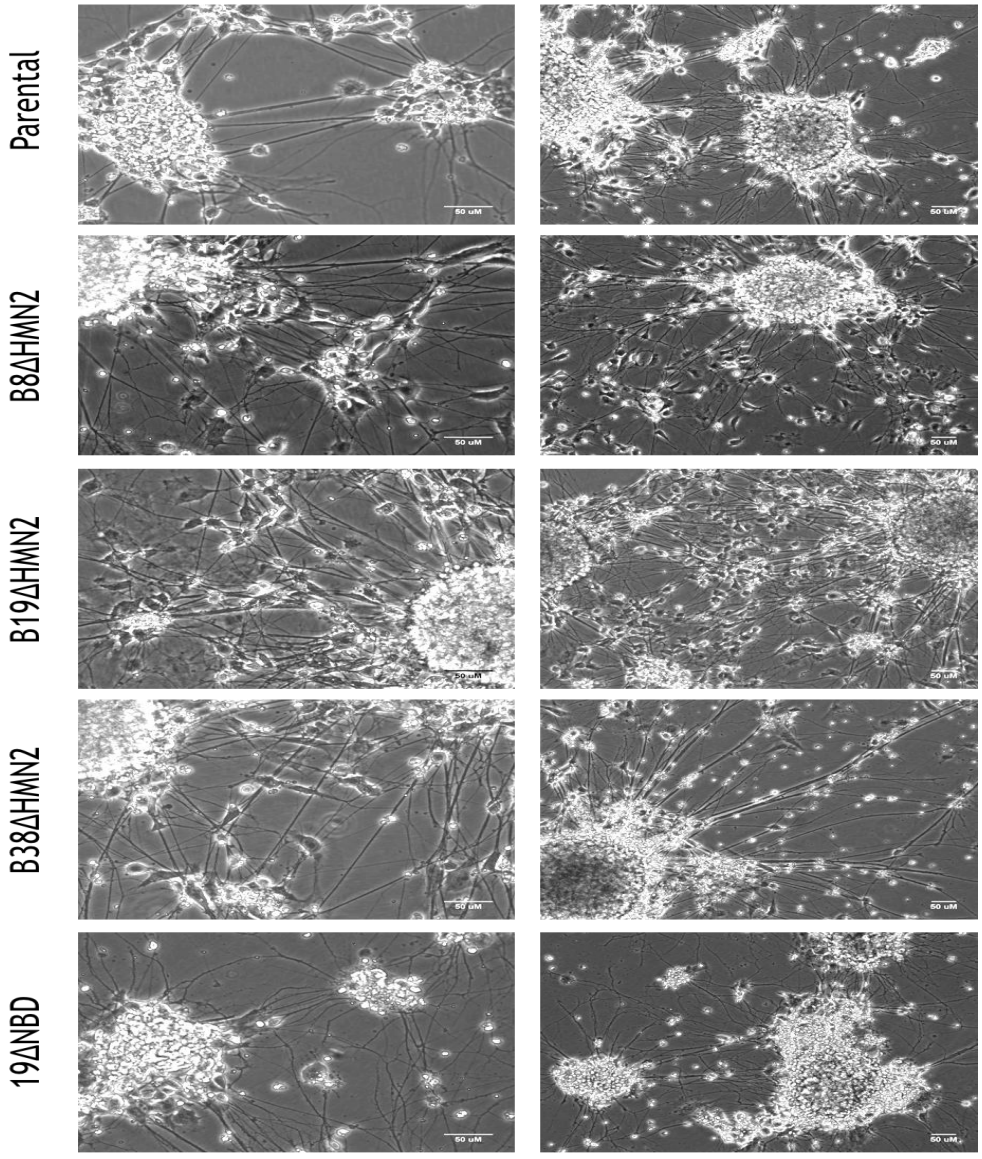


Figure 6.14: The light images for different lines showing efficient neural induction after 6 days.
Bright field images for live cells at day 6 after neural induction. The neural induction was accomplished using the monolayer adherent culture. The parental P19 cells and different knockout lines were induced by incubating the cells in DMEM/F12 media that contains 500 nM retinoic acid (RA),10 µM DAPT (gamma secretes inhibitor), 10 ng/ml FGF2 and N2 in absence of the serum. The cells were plated in only 100 mm dishes that were previously coated with 5 µg/ml laminin. The cells were incubated in the DMEM/F12 media for 4 days and then changed with NBA media with B27 and AraC for 2 days to kill the proliferative cells. Scale bar, 50 µM.

6.5.3.1 Identifying the *Hmgn1* and *2* expressions in all lines during neural induction

The *Hmgn1* expression and loss of *Hmgn2* in the knockout lines were assessed during neural induction. *Hmgn2* expression was up-regulated by 2- and 3-fold in parental P19 cells at day 3 and 6, respectively, compared to day 0 (Figure 6.15). As expected, *Hmgn2* mRNA expression in lines B19ΔHMGN2 and B38ΔHMGN2 was substantially less than that of the parental cells during the differentiation process, and *Hmgn2* mRNA was not detected in 19ΔNBD (Figure 4.13, chapter 4). It was noticeable that although *Hmgn2* mRNA in the B8ΔHMGN2 line was low in the undifferentiated cells, it had increased to parental levels by day 3 and day 6 (Figure 6.15).

HMGN2 protein expression is shown in the same figure (Figure 6.15). The level of HMGN2 protein was increased at day 6 in the parental cells (Figure 6.15), consistent with the mRNA data and with previous western blotting data (Figure 6.6). Although the undifferentiated B8ΔHMGN2 cells had a faint HMGN2 band in the previous analysis (Chapter 4), this time none of the knockout lines showed any HMGN2 expression in the undifferentiated state. However, a faint HMGN2 band was observed at day 3 in the B8ΔHMGN2 line, and by day 6, the level of HMGN2 protein in this line was almost as high as that in the parental cells. The reappearance of HMGN2 expression in B8ΔHMGN2 line may be because this is a mixed culture of proliferating and differentiating cells, so any contaminating $HMGN2^{+/+}$ cells may become more dominant by day 6.

The expression of *Hmgn1* mRNA in parental P19 cells was gradually up-regulated to around 1.8 fold at day 6 (Figure 6.15), which is consistent with the previous western blotting results (Figure 6.6). *Hmgn1* expression was slightly higher in lines B38ΔHMGN2 and 19ΔNBD at day 0, as previously observed (Figure 5.2). Similar to the parental cells, the expression of *Hmgn1* in the HMGN2 knockout lines was increased by around 1.6 fold by day 6. Therefore, the increase in *Hmgn1* expression during neuronal differentiation does not appear to be affected by the loss or mutation of HMGN2.

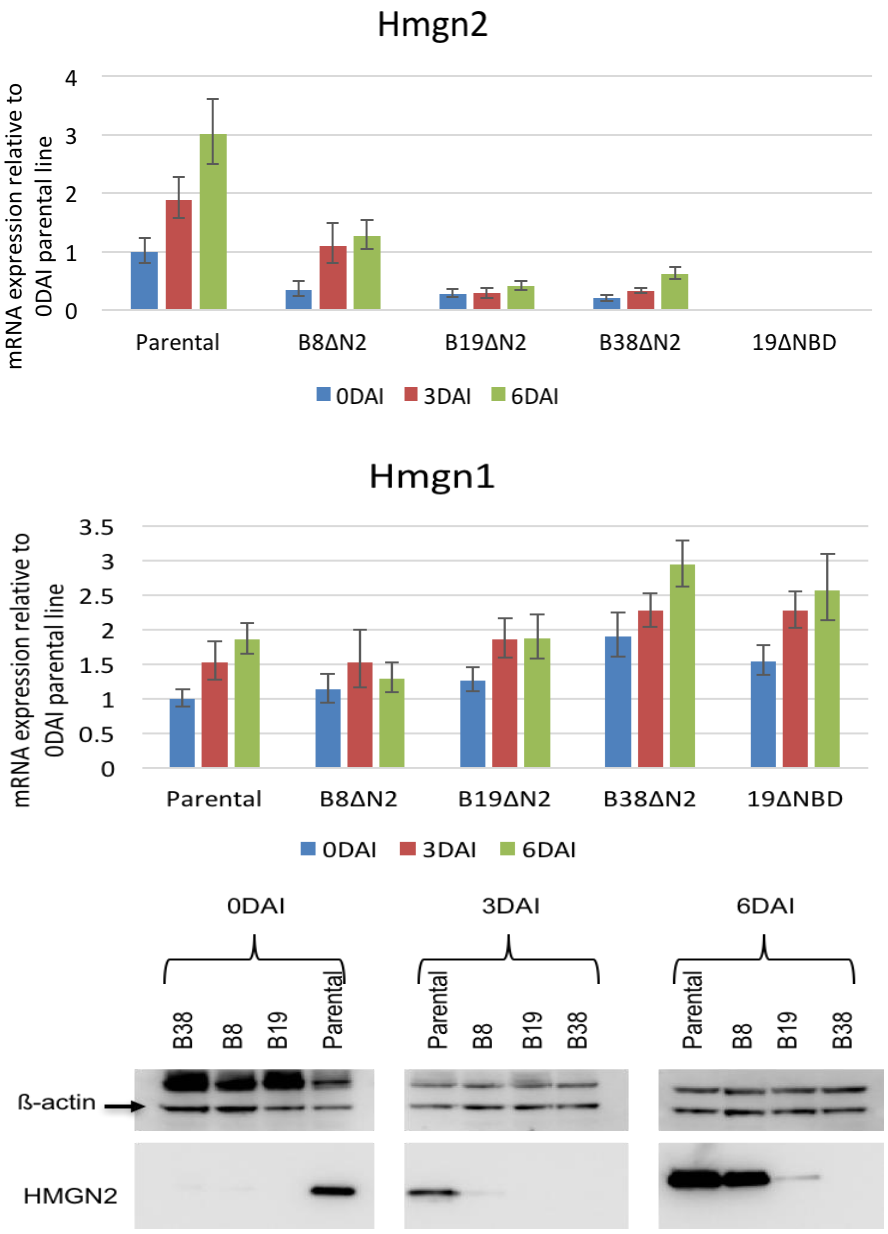


Figure 6.15: *Hmgn2* and *1* expressions and HMGN2 protein in all lines during the neuronal differentiation using the monolayer adherent culture.
The RNA expression of *Hmgn1* and *2* and the protein level of HMGN2 (17 kDa) in the parental P19 cells and different knockout lines (B8ΔHMGN2, B19ΔHMGN2, B38ΔHMGN2 and 19ΔNBD) at 0, 3 and 6 days after induction (DAI) using adherent monolayer culture. *Hmgn2* and *1* expressions are increased during differentiation at day 3 and 6 in the parental cells. Although the expression of *Hmgn2* in B8ΔHMGN2 was undetected in the undifferentiated cells (0DAI), it is expressed at day 3 and 6. The other complete knockout lines do not show any RNA expression in the undifferentiated and during neural induction. The *Hmgn2* is undetected in 19ΔNBD. However, the *Hmgn1* was increased in all lines. Gene expression in each sample was normalised to three housekeeping genes (*Gpi1*, *Rplp0* and *Gapdh*) and is plotted relative to that in parental cells at day 0 (see methods for calculation). Error bars reflect the standard deviation from RT-PCR triplicates from one biological replicate. DAI stands for days after induction while N2 refers to HMGN2. NBD is for nucleosome binding domain. The result in western blot confirming the RNA expression data, but showing unexpected faint band in B19ΔHMGN2. The protein sample of 19ΔNBD was not included. β-Actin (42 kDa) was used as internal control for western blot and showing non-specific band (upper bands).

6.5.3.2 The Oct4 and Nanog expressions are not differentially affected in the knockout lines during the neural induction

As mentioned before, the Oct4 was not differentially expressed in the knockout lines at day 0 (Figure 5.6). During neuronal differentiation, Oct4 expression was reduced by over 1000 fold in the parental cells. Similar reductions were observed in all the knockout and mutant lines.

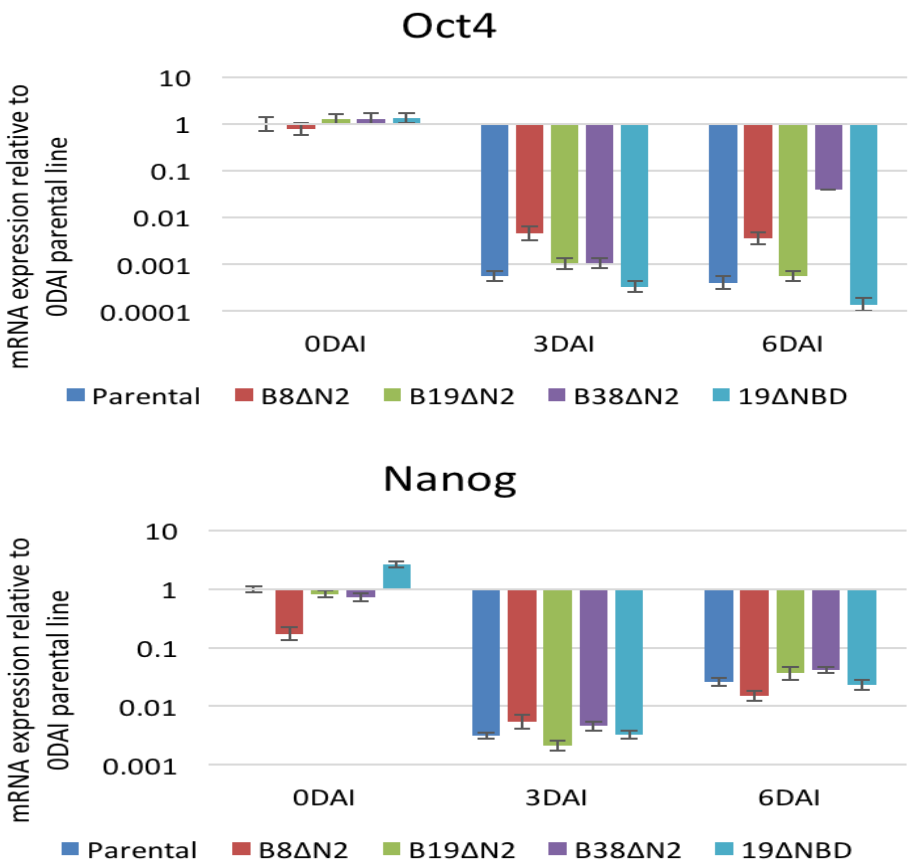


Figure 6.16: The down-regulation of pluripotent genes (*Oct4* and *Nanog*) during neural induction in all lines.
RNA expression of *Oct4* and *Nanog* was examined in the parental P19 cells and different knockout lines (B8ΔHMGN2, B19ΔHMGN2, B38ΔHMGN2 and 19ΔNBD) at 0, 3 and 6 days after induction (DAI) using adherent monolayer culture. Gene expression in each sample was normalised to three housekeeping genes (*Gpi1*, *Rplp0* and *Gapdh*) and is plotted relative to that in parental cells at day 0 (see methods for calculation). Error bars reflect the standard deviation from RT-PCR triplicates from one biological replicate.

The previous observations on *Nanog* expression in the undifferentiated cell lines were confirmed here: *Nanog* expression was reduced in B8ΔHMGN2, B19ΔHMGN2 and B38ΔHMGN2 cells, whereas it was up-regulated in 19ΔNBD cells at day 0. During neuronal differentiation, *Nanog* expression was profoundly reduced in all lines. There were no differences in the *Nanog* expression between the knockout

or mutant HMGN2 lines and parental P19 cells. Therefore, it was concluded that the expression of both pluripotency genes was reduced as expected during neural induction, with no differences between the knockout lines and parental cells.

6.5.3.3 The expressions of proneural bHLH genes are highly induced in the HMGN2$^{-/-}$ lines during neural induction

The basic bHLH transcription factors (*Ngn1*, *Mash1* and *NeuroD*) that drive the neuronal lineage were also studied during the neural induction of parental and mutated cell lines. Previously, results showed that these factors were highly up-regulated in the undifferentiated HMGN2$^{-/-}$ lines but not in the mutant HMGN2 line (19ΔNBD) (Figure 6.12). As shown earlier, *Ngn1*, *Mash1* and *NeuroD* are up-regulated in the undifferentiated HMGN2$^{-/-}$ cells (0DAI, Figure 6.17).

At day 3 after induction, the relative expression of *Ngn1* increased to over 1500 in the parental cells. In two of the HMGN2$^{-/-}$ lines (B8ΔHMGN2 and B19ΔHMGN2) expression of *Ngn1* was considerably higher than this, at around 4000 and 6000. However, in B38ΔHMGN2, *Ngn1* expression was similar to that of the parental cells (Figure 6.17). At day 6 after induction, the relative expression of *Ngn1* was reduced back down to 50 in the parental cells. Expression in all three HMGN2 knockout lines was reduced even further, to 15-25. The expression of *Ngn1* in the 19ΔNBD line showed a different pattern to that of the HMGN2 knockout lines, with lower expression at days 0 and 3, and similar levels to the parental cells at day 6.

Mash1 and *NeuroD* were also transiently induced by over 100 fold during the differentiation of the parental cells (Figure 6.17). This is in agreement with previous findings that found a transient high expression of the proneural bHLH factors during neural differentiation of RA-treated P19 cells (Itoh et al.,1997; Kim et al., 2004). All three HMGN2 knockout lines, B8ΔHMGN2, B19ΔHMGN2 and B38ΔHMGN2, had higher expression of *Mash1* and *NeuroD* at day 3, but there was variable expression between the lines at day 6 (Figure 6.17). Again, the HMGN2 mutant line 19ΔNBD had a different pattern, with similar levels of *Mash1* and *NeuroD* to wild-type at day 3, but higher levels at day 6 (Figure 6.17).

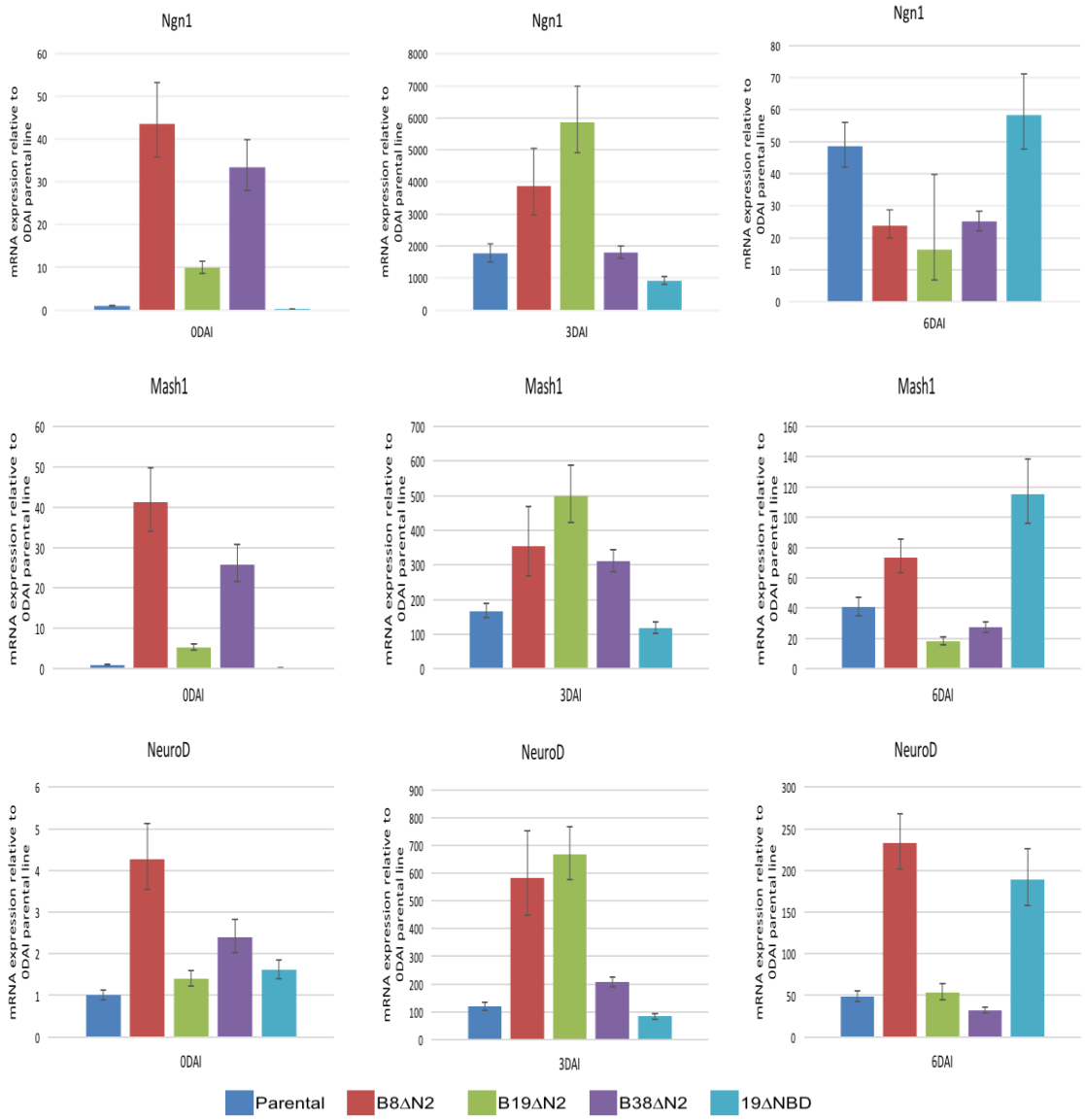


Figure 6.17: The up-regulation of proneural bHLH transcription factors [*Ngn1*, *Mash1* and *NeuroD*] expression at day 3 after neural induction in knockout lines.
Ngn1, *Mash1* and *NeuroD* expressions were compared between parental cells (wild-type) and complete or partial HMGN2 knockout lines (B8ΔN2, B19ΔN2, B38ΔN2 and 19ΔNBD) at days 0, 3 and 6 during induction. The data for each day is represented in different graphs due to the large differences in the fold change between different time points. Gene expression in each sample was normalised to three housekeeping genes (*Gpi1*, *Rplp0* and *Gapdh*) and is plotted relative to that in parental cells at day 0 (see methods for calculation). Error bars reflect the standard deviation from RT-PCR triplicates from one biological replicate. DAI stands for days after induction while N2 refers to HMGN2. NBD is for nucleosome binding domain.

6.5.3.4 HMGN2 knockout positively regulates the early neural TF, *Zfp521*, in the undifferentiated P19 cells and during neural induction

Zing finger protein 521 (Zfp521) is an essential transcription factor that directs ES cells into the neural lineage. Based on Kamiya et al. (2011), *Zfp521* transcripts are not expressed in the undifferentiated ES cells but are enriched (~60-fold higher) in an early neural lineage during ES cell differentiation.

Overexpression and knockdown experiments showed that *Zfp521* is necessary and sufficient to drive neuronal differentiation in culture (Kamiya et al., 2011). *Zfp521* is expressed in mouse embryos during gastrulation, specifically in the neuroectodermal region but not in the endoderm, mesoderm or non-neural ectoderm. Zfp521 is a nuclear protein with 30 Krüppel-like zinc fingers mediating multiple protein-protein interactions. It regulates transcription in diverse tissues and organs, including the expression of early neural genes. In addition, the protein promotes proliferation, delays differentiation and reduces apoptosis (Shen et al., 2011).

Here, the data shows that *Zfp521* was expressed at very low levels in undifferentiated P19 cells, but it was induced by around 5 fold during neural induction (Figure 6.18). Similar to the other neural-related genes, *Zfp521* expression was slightly up-regulated in the three HMGN2$^{-/-}$ lines, but was unchanged in 19ΔNBD cells, before neural induction (Figure 6.18). At 3DAI, *Zfp521* expression was around two fold higher in two of the HMGN2$^{-/-}$ lines (B8ΔHMGN2 and B19ΔHMGN2) compared to the parental cells (Figure 6.18). The other lines, B38ΔHMGN2 and 19ΔNBD, had similar *Zfp521* expression to the parental cells at day 3. At day 6, *Zfp521* expression was slightly reduced in most of the lines, and there was no consistent difference between knockout and parental cells (Figure 6.18). In conclusion, the expression of *Zfp521* was increased in all HMGN2$^{-/-}$ lines in undifferentiated P19 cells, corroborating the finding of more neural spontaneous differentiation in cells that lack HMGN2. *Zfp521* expression was also more highly induced in two HMGN2$^{-/-}$ lines when the cells were directed down the neural lineage.

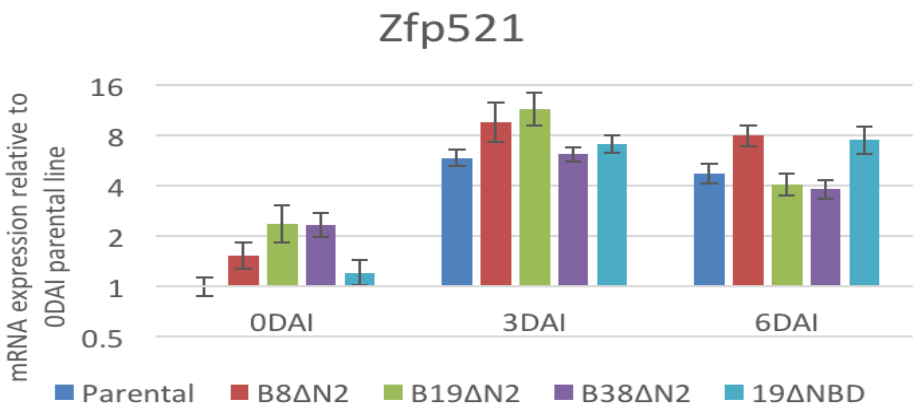


Figure 6.18: The complete HMGN2 knockout increased the expression of *Zfp521* in undifferentiated and at day 3 of neural induction.
RNA expression of *Zfp521* was assessed at day 0, 3 and 6. Gene expression in each sample was normalised to three housekeeping genes (*Gpi1*, *Rplp0* and *Gapdh*) and is plotted relative to that in parental cells at day 0 (see methods for calculation). Error bars reflect the standard deviation from RT-PCR triplicates from one biological replicate. DAI stands for days after induction while N2 refers to HMGN2. NBD is for nucleosome binding domain.

6.5.3.5 Loss of HMGN2 also regulates *Pax6* expression in the undifferentiated and during the neural induction

The *Pax6* gene encodes a homeobox and paired DNA-binding domain-containing protein that functions as a regulator of transcription (activator or repressor). This transcription factor is expressed in a spatiotemporal manner during the development of the central nervous system. Its expression starts at very early stage of CNS development, at day 8 of mouse embryogenesis in the presumptive forebrain and hindbrain neuroectoderm (Gajovic et al., 1997). It is also expressed in the forebrain and hindbrain and spinal cord of the neural tube at day E10.0 of the mouse embryos (reviewed in Osumi et al., 2008). In the neural tube, it is restricted to mitotically active cells of the ventral VZ, which comprises neuronal and glial progenitor cells (Walther and Gruss, 1991). In adults, *Pax6* is expressed in the hippocampal dentate gyrus and SVZ/rostral migratory stream, regions in which neuronal precursors exist during adult life (Nacher et al., 2005). The majority of cells in the forebrain SVZ of neonatal mice express either *Pax6* or *Olig2* (for oligodendrocyte fate determination), and rarely express both (Jang and Goldman, 2011). This reciprocal expression between *Pax6* and *Olig2* indicates the importance of these factors in determining the cell fate. Jang and Goldman (2011) found that overexpression of *Pax6* is sufficient to stimulate the neurogenic fate and inhibit a gliogenic fate in the neonatal SVZ. Nevertheless, a different study by Gomez-Lopez et al. (2011) concluded that

gliogenesis also requires *Pax6*, since defective differentiation into glial lineages was shown in *Pax6*-null neural stem cells. In general, Pax6 plays role in proliferation and neuronal differentiation of the neural stem/progenitor cells during embryogenesis and in adult (Ericson et al., 1997; Nacher et al., 2005; Osumi et al., 2008; Gomez-Lopez et al., 2011). *Pax6*-mutant rodents have imperfect neural tube patterning, changed progenitor proliferation, decreased neurogenesis, impaired migration of neuronal precursors, and other defects (Stoykova et al., 1996; Caric et al. 1997; Heins et al., 2002; Kroll and O'Leary, 2005; Quinn et al., 2007; Gomez-Lopez et al., 2011). Furthermore, *Pax6* plays an important role in eye, nose, and pancreas development. It was reported that *Pax6* is not expressed in mouse ES cells but is induced upon neural differentiation *in vitro* (Gajovic et al., 1997).

In P19 cells, *Pax6* expression is very low in the parental undifferentiated cells (Figure 6.20). Its expression is up-regulated 2 to 4-fold in undifferentiated HMGN2 Knockout lines (B8ΔHMGN2, B19ΔHMGN2 and B38ΔHMGN2) (Figure 6.19). 19ΔNBD, however, has a down-regulation of *Pax6* at day 0. Neural induction of P19 parental cells increases relative *Pax6* expression to 130 at day 3 (Figure 6.19). *Pax6* induction is higher in two of the HMGN2 knockout lines, and in the mutant HMGN2 line (relative expression of 215, 157 and 247 in lines B8ΔHMGN2, B19ΔHMGN2 and 19ΔNBD respectively, Figure 6.19). However, *Pax6* induction in the third HMGN2 knockout line is only 57, which is around half of the parental level at day 3. At day 6, *Pax6* expression is substantially reduced in all complete HMGN2 knockout lines, with ~4 times less expression compared to the parental cells (Figure 6.19).

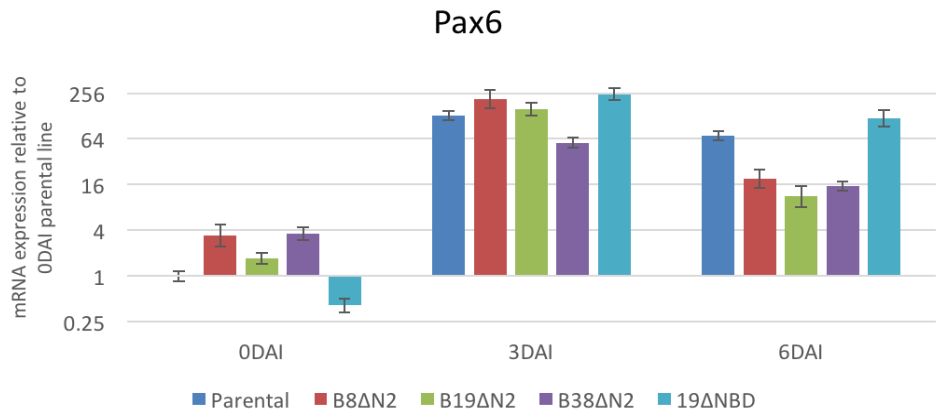


Figure 6.19: *Pax6* expression in different lines in the undifferentiated cells and during neuronal differentiation.
Pax6 mRNA expression is measured in parental (wild-type), B8ΔHMGN2, B19ΔHMGN2, B38ΔHMGN2 and 19ΔNBD lines at three-time points (0, 3 and 6 DIA). Gene expression in each sample was normalised to three housekeeping genes (*Gpi1*, *Rplp0* and *Gapdh*) and is plotted relative to that in parental cells at day 0 (see methods for calculation). Error bars reflect the standard deviation from RT-PCR triplicates from one biological replicate. DAI stands for days after induction while N2 refers to HMGN2. NBD is for nucleosome binding domain.

6.5.3.6 Differential expression of *Nestin* in the undifferentiated and during the neural differentiation

Nestin expression was also examined during the neural induction of parental and mutant P19 cell lines (Figure 6.20). As shown previously, Nestin was up-regulated in undifferentiated cells of the HMGN2 knockout lines, but not in the mutant HMGN2 line. During neural induction of the parental cells, *Nestin* expression was gradually increased up to 22 fold at day 6 (Figure 6.20). In contrast to the pro-neural transcription factors studied above, *Nestin* expression was not more highly induced in the HMGN2 knockout cells. In fact, *Nestin* expression appeared to be slightly lower in the knockout cells at day 3 and at day 6 (Figure 6.20). Unlike the complete knockout, the expression of *Nestin* in HMGN2 mutant, 19ΔNBD cells was comparable to wild-type at 0, 3 and 6 DAI.

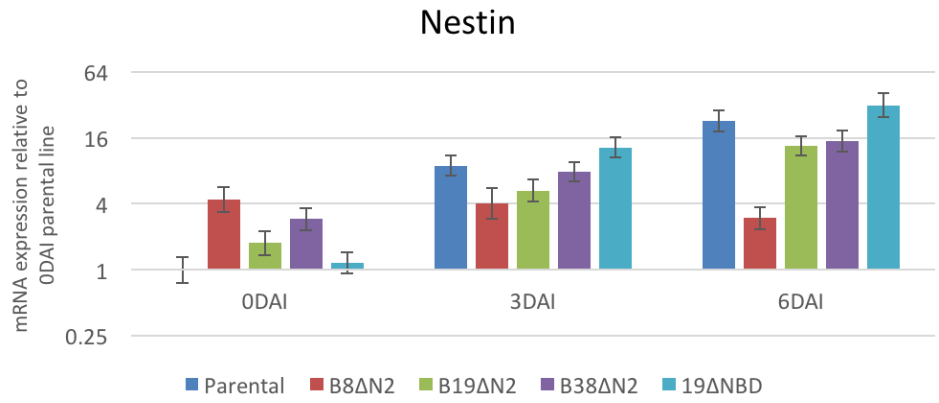


Figure 6.20: The expression of *Nestin* during the neural induction in all lines.
Nestin mRNA expression is measured in parental (wild-type), B8ΔHMGN2, B19ΔHMGN2, B38ΔHMGN2 and 19ΔNBD lines at three-time points (0, 3 and 6 DIA). Gene expression in each sample was normalised to three housekeeping genes (*Gpi1*, *Rplp0* and *Gapdh*) and is plotted relative to that in parental cells at day 0 (see methods for calculation). Error bars reflect the standard deviation from RT-PCR triplicates from one biological replicate. DAI stands for days after induction while N2 refers to HMGN2. NBD is for nucleosome binding domain.

6.5.3.7 The negative regulator of neurogenesis, *Rest*, is up-regulated in un-induced HMGN2$^{-/-}$ cells only

REST (RE1-silencing factor) is a negative regulator of neurogenesis that is involved in the repression of neural genes in the non-neuronal cells. This zinc finger protein binds to a 21 bp repressor element 1 (RE-1) that is also known as neuron-restrictive silencer element (NRSE) (Kraner et al., 1992; Mori et al., 1992). Transcriptional repression by REST is mediated by two corepressor complexes (mSin3 and CoREST) that recruit histone deacetylases (Ballas et al., 2005; Ooi and wood, 2007). *Rest* expression is high in non-neuronal cells, but low or undetectable in the *in vitro*-derived neuronal progenitors and neurons (Kraner et al., 1992; Mori et al., 1992; Ballas et al., 2005). However, REST also seems to have multiple functions that differ depending on the cell type and developmental stage. Indeed, it was reported that although the expression of REST is heterogeneous in the adult CNS, REST can be found in the hippocampus, midbrain and hindbrain of the adult brain (Greenway et al., 2007). REST protein has also been detected in NSCs derived from adult hippocampus, and in differentiated neurons derived from these NSCs (Kuwabara et al., 2004). Some evidence proposed that REST may be involved in the transcriptional activation of RE1-related genes during neurogenesis and adult NSCs (Kuwabara et al., 2004). It was shown that noncoding double-stranded RNA containing a RE1/NRSE sequence could convert REST from a repressor to an activator and thus induce neuronal

differentiation in adult hippocampal NSCs (Kuwabara et al., 2004). However, inhibition of REST function in murine NSCs, although this induces the ß-III tubulin expression, is not sufficient to induce differentiation into mature neurons (Greenway et al., 2007). Therefore, REST may have different mechanisms during neurogenesis depending on the developmental stage.

In our experiments, *Rest* expression was high in the undifferentiated parental cells, but was reduced to half at 3DAI and re-expressed to the initial level at 6DAI (Figure 6.21). This pattern of expression differs from *in vitro* differentiation of mouse ES cells induced by RA and formation of EB (Ballas et al., 2005). They mainly found that the REST protein, but not RNA, is posttranslationally down-regulated in cortical neural progenitor at day 3 after RA treatment, although *Rest* RNA expression is repressed in mature neurons.

In the undifferentiated HMGN2$^{-/-}$ lines, *Rest* expression was slightly up-regulated compared to parental cells, but unaffected in 19ΔNBD (Figure 6.21). During neuronal differentiation, the expression of *Rest* in the HMGN2 knockout lines was reduced to levels similar to the parental cells.

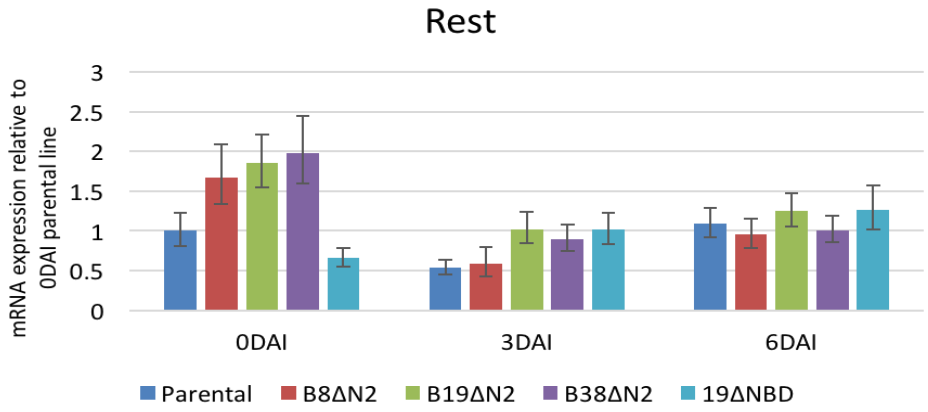


Figure 6.21: The expression of RE1 silencing transcription factor (*Rest*) during the neural induction in all lines.
Gene expression in each sample was normalised to three housekeeping genes (*Gpi1*, *Rplp0* and *Gapdh*) and is plotted relative to that in parental cells at day 0 (see methods for calculation). Error bars reflect the standard deviation from RT-PCR triplicates from one biological replicate. DAI stands for days after induction while N2 refers to HMGN2. NBD is for nucleosome binding domain.

6.5.3.8 The neuronal gene expressions in the undifferentiated P19 cells and after induction are enhanced by the removal of HMGN2

After observing higher induction of several early neural transcription factors at day 3 in most of HMGN2 knockout lines, we asked whether the neuronal-specific genes (*Map2* and *Nf160*) are also increased in the differentiated HMGN2$^{-/-}$ lines.

As previously shown, expression of *Map2* in undifferentiated cells was increased in the HMGN2 knockout cells (Figure 6.22 and Figure 6.12). The *Map2* expression was substantially increased during neural induction, with a relative expression of 32 at day 6 (Figure 6.22). Induction of *Map2* was higher in the knockout cells at day 3, with relative expression levels of 105, 100 and 42 in B8ΔHMGN2, B19ΔHMGN2 and B38ΔHMGN2 cells, respectively (Figure 6.22). In the HMGN2 mutant line 19ΔNBD, *Map2* was induced to a lower level than in the parental cells. Expression of *Map2* at day 6 was comparable in all the cell lines.

Nf160 expression was gradually increased in parental cells during neural induction, increasing from 10-fold at 3DAI to 20-fold at 6DAI. Nf160 was notably increased in B8ΔHMGN2 (22-fold), B19ΔHMGN2 (16-fold) and B38ΔHMGN2 (23-fold) at 3DAI compared to parental cells, whereas expression in 19ΔNBD cells was similar to that in parental cells (Figure 6.22). On the other hand, the expression of *Nf160* at day 6 did not have distinguishable pattern between different groups (Figure 6.22). Expression in B19ΔHMGN2 and B38ΔHMGN2 was similar to parental cells, whereas expression in B8ΔHMGN2 and 19ΔNBD was lower and higher than in parental cells, respectively (Figure 6.22).

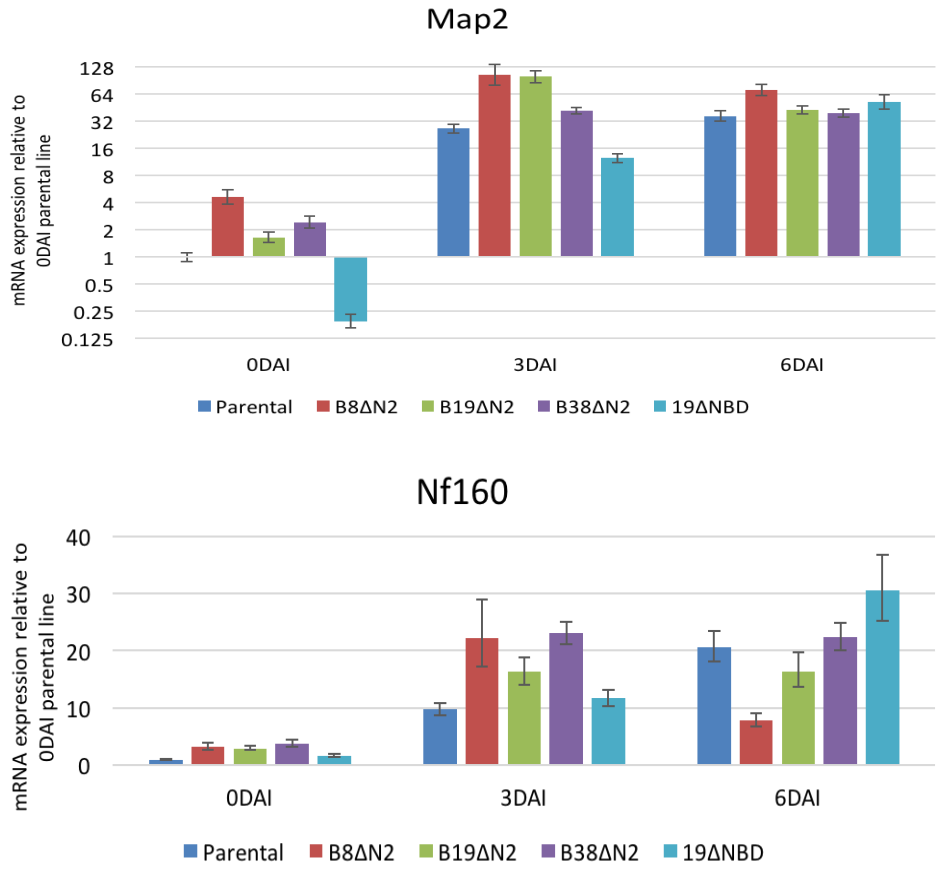


Figure 6.22: Up-regulation of neuronal-specific genes [*Map2* and *Nf160*] expression at day3 after neural induction in knockout lines.
Map2 and *Nf160* mRNA expressions are assessed in parental (wild-type), B8ΔHMGN2, B19ΔHMGN2, B38ΔHMGN2 and 19ΔNBD lines at three-time points (0, 3 and 6 DIA). Gene expression in each sample was normalised to three housekeeping genes (*Gpi1*, *Rplp0* and *Gapdh*) and is plotted relative to that in parental cells at day 0 (see methods for calculation). Error bars reflect the standard deviation from RT-PCR triplicates from one biological replicate. DAI stands for days after induction while N2 refers to HMGN2. NBD is for nucleosome binding domain.

In addition to RNA analysis during the neural induction, we also carried out western blotting at 0, 3 and 6 DAI for a neural stem cell marker (Nestin), and neuronal-specific markers (NF160 and ß-III tubulin) (Figure 6.23). As previously observed in wild type (Figure 6.4), Nestin and ß-III tubulin are detected before neural induction of the parental cells, and their levels are increased at 3DAI and 6DAI. NF160 is not detected before neural induction, but is highly expressed by day 6. The detection of ß-III tubulin but not NF160 in undifferentiated cells may be because ß-III tubulin is known to be expressed in immature neurons while NF160 is found in more mature neurons.

In terms of the HMGN2 knockout lines, firstly, Nestin level is almost equal in all HMGN2 knockout lines (B8ΔHMGN2, B19ΔHMGN2 and B38ΔHMGN2) at day 0, but it is substantially reduced in all knockout lines at days 3 and 6 (Figure 6.23).

Secondly, the neuronal marker (ß-III tubulin) was variable at 0DAI (Figure 6.23). Although ß-III tubulin protein level seems to be reduced in B8ΔHMGN2, B19ΔHMGN2 and B38ΔHMGN2 at day 3, higher level of this protein is in B8ΔHMGN2 and B19ΔHMGN2 knockout lines in comparison to parental cells at day 6. Finally, NF160 at day 6 are comparable in all the knockout lines and the parental cells. NF160 at day 3 is variable between the knockout lines in which B38ΔHMGN2 have almost similar level as the parental cells but less in B8ΔHMGN2 and B19ΔHMGN2 (Figure 6.23). From all these protein data, only the level of Nestin after the induction are comparable with the RNA expression data. Particularly, a reduction at day 3 and 6 are almost comparable with mRNA at the same time points. This was only performed once and additional replicates are also required.

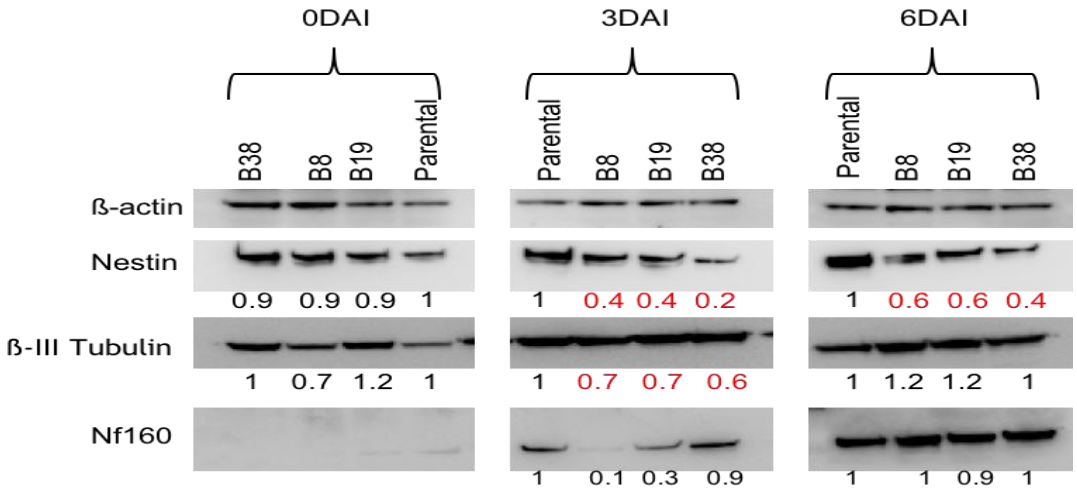


Figure 6.23: Nestin, ß-III tubulin and NF160 protein levels in undifferentiated cells and during neuronal differentiation.
The western blots for Nestin (200 kDa), ß-III tubulin (50 kDa) and NF160 (160 kDa) were performed using whole cell extracts of parental (wild-type), and complete HMGN2 knockout lines (B8ΔHMGN2, B19ΔHMGN2, B38ΔHMGN2) at 0, 3 and 6 days after induction (DAI). The protein sample of 19ΔNBD was not included. β-Actin (42 kDa) was used as internal control for western blot. The numbers at bottom of each blot represent the densities of each band after normalizing to β-Actin and then to parental cells of each time point.

6.6 Discussion

6.6.1 Characterisation of neural induction of EC P19 cells using adherent monolayer culture

The study of the molecular mechanisms during neural development can be accomplished by using ES or EC cells, which can be considered as powerful model systems for this purpose. EC cells such as P19 are used in neural

development studies because they have many features in common with pluripotent embryonic cells, and because they can produce of all cell type in the neural lineage (Mummery et al., 1987). Studies of directed EC differentiation into the neural lineage is typically based on differentiation via embryoid bodies in presence of non-specific morphogens (RA). Although this approach can yield populations with neuron-like cells, the efficiency of neural conversion is limited and lineage selection is usually required to ensure homogeneity of the differentiated population (Gajovic et al., 1997; Abranches et al., 2009). In addition, the undefined factors and endogenous cytokines present in EBs or serum may complicate efforts to investigate the role of specific proteins or signalling pathways in neural development (Velkey et al., 2013). Differentiation of ES or EC cells in monolayer cultures and defined media minimises these effects (Ying et al., 2003b; Nakayama et al., 2014) and, if combined with ectopic expression of a developmental gene, presents a unique opportunity to tease out the interactions of specific genes and growth factors in early development. Therefore, the neural induction of EC P19 cells was achieved using adherent culture with a serum-free defined medium as described earlier by Nakayama et al, (2014). This approach benefits from the rapid and selective enhancement of the neurons production.

Using this method, the neural induction of P19 cells was successfully achieved within 6 days after induction. The efficiency of neural induction was evaluated by the loss of pluripotent markers (Oct4 and Nanog) just immediately after induction, and the induction of neural and neuronal-specific markers (Nestin for NPCs and ß-II Tubulin, MAP2 and NF-160 for neurons). The peak of the neuronal markers (ß-II Tubulin, MAP2 and NF-160) was at day 6 and 8 after induction. The reciprocal change in marker gene expression indicates the loss of pluripotency and the determination of neural and neuronal lineage. Although cell projections were observed from day 2 and onwards, sophisticated neurite networks are clearly seen at day 6. Thus, the neuronal survival and formation of neurite networks after day 4 are clearly enhanced by changing the media to B27-supplemented Neurobasal media that was optimised for mature neuron survival (Brewer et al., 1993). Besides that, according to Nakayama and colleagues (2014) neurons at this time point can respond to several neurotransmitter signalling indicating the presence of functional neurons. However, there was a

clear reduction of neuronal specific proteins beyond day 8. This may be an indicator of synaptic pruning, which refers to the normal process of elimination of dysfunction synapses during brain maturation (Zhan et al., 2014), or neuronal death. The neuronal death is probably caused by the use of AraC or as result the lack of the astrocytes. Because the neuronal cytotoxicity by AraC is dose-dependant (Dessi et al., 1995), and the concentration of AraC was only 8 μM, the neuronal viability does not seem to be mainly affected by this treatment. In addition, glutamine was found to be toxic coincident with increased extracellular glutamate when neurons were cultured with reduced numbers of astrocytes (Newcomb et al., 1997). However, although this is possible, the obvious reason for the reduction in neuronal markers is the reduction in producing neurons as most of the cells in aggregates have became non-proliferative and most of the proliferative cells in culture have been killed by AraC.

Although the adherent neural induction of P19 cells is rapid and selective in producing neurons, it does not completely remove the heterogeneity and sophisticated cluster of cells are formed. During neural induction of the P19 cells, it has been noticed that the cells in the culture form adherent cell aggregates by day 2 and these aggregates become more condensed with time. It is technically challenging to analyse the cells inside the aggregates by immunofluorescence, as seen when characterising the HMGN2 expression. Several attempts including low cell seeding and/or using cell strainer were applied to overcome the aggregates formation, but they did not minimise this feature. This was also observed in the Nakayama study and thought to be a result of using the γ-secretase inhibitor, which blocks proteolytic cleavage of the Notch receptor. A similar finding of changing the rosette structures into large rounded ganglion-like clusters with extensive neurite outgrowth was also obtained when the rosette culture derived from mES cells in monolayer adherent culture were treated with DAPT for 2 days (Abranches et al., 2009). Notch signalling is known to be important in maintaining neural progenitors, and has also shown to be required for the maintenance of neural rosette structure; inhibition of this pathway results in breakdown of the apical polarity of the neural rosette and accelerates the neuronal differentiation (Main et al., 2013). Since the use of Notch inhibitor is critically required in this method as only a few

have protrusions without it (Nakayama et al., 2014), the formation of cell aggregates was unavoidable.

Interestingly, both the neural progenitor marker Nestin, and the neuronal markers Map2 and NF160, were most highly expressed at day 6, which may suggest the conversation of NPCs to neurons occurs very rapidly. This may be caused by the use of γ-secretase inhibitor that was shown to accelerate the neuronal differentiation by repressing the *Hes1/5* that promote the proliferation of NPCs (Ohtsuka and Kageyama, 2010; Main et al., 2013). The highest level of Nestin at day 6 is followed by a sharp reduction at days 8 and 10. There are differences in Nestin level compared to what seen in the study by Nakayama et al, (2014), who found that Nestin reaches a plateau at day 4 and is then sharply reduced at day 6. It was not clear why this variation was obtained but many factors can influence the neuronal differentiation. The presence of some $Nestin^+$ cells at day 10 indicates some of cells are still proliferative, as noticed by the increase in cell confluence and confirmed by EdU staining. Having proliferative cells at a late stage of neural induction may be necessary to drive production of glial cells. It is well known that during neural development, gliogenesis occurs toward the end of neurogenesis (Temple, 2001; Nagao et al., 2014). Nevertheless, $GFAP^+$ cells were not detected till day 16.

In general, the neural induction of P19 cells using monolayer adherent culture with a serum-free defined medium is robust, rapid and selectively produced neurons. Although it can be applied to study the role of HMGN2 during neural development, particularly during neurogenesis, there were some drawbacks that include the formation of extensive cell aggregates, and the heterogeneous nature of the cultures.

6.6.2 Characterise HMGN1 and 2 expressions during *in vitro* neural differentiation

The level of HMGN1 and 2 proteins changed dynamically during the neuronal differentiation. These proteins increased during neural induction, but then sharply reduced at the later stage of neural induction when the culture mostly contains cell aggregates and mature neurons. The dynamics of HMGN1 and HMGN2 expression during neural induction may suggest that HMGN1 and 2 are

required at a high level during the neurogenesis, but are not required after neuronal production or maturation. Consistent with this hypothesis, Nagao et al, (2014) demonstrated that the Dcx+ immature neurons highly express HMGN1 and 2, whereas the NuN-positive mature neurons have very low expression of these proteins. In addition, several studies have demonstrated the reduction of HMGN1/2 proteins at the late stage of differentiation in several tissues including chondrocytes, myoblasst and erythropoietic cells, and it was shown that overexpression of HMGN proteins can halt the differentiation process (Crippa et al., 1991; Begum et al., 1990; Furusawa et al., 2006). Therefore, reduction of HMGN1 and 2 at the late stage of neurogenesis is comparable with other lineage differentiation studies.

It is known that the CNS stem cells during neurogenesis initially divide symmetrically to expand their population (reviewed in Egger et al., 2010). As the HMGN2 protein is increased after the neural induction and most of $Nestin^{+}$ cells are also $HMGN2^{+}$, it seems that the HMGN2 is important in NPCs. Performing the EdU proliferation assay, in combination with HMGN2 immunostaining revealed that only high HMGN2 positive cells are also positive for EdU incorporation at day 10 after induction, indicating the association between proliferative cells and HMGN2. These observations clearly suggest the correlation between HMGN2 and the proliferation of NPCs. It was recently demonstrated that HMGN family proteins are expressed in NPCs, glial progenitors, and astrocytes in the perinatal and adult brains (Nagao et al., 2014). Broadly, the relationship between high proliferative cells and HMGN2 was previously suggested when the level of HMGN2 transcript was found to be higher in rapidly dividing Hela cells comparing to slowly dividing cells from liver tissue (Bustin et al., 1987). In addition, it was shown that the removal or reduction of HMGN5 can decrease the proliferation in some cancer cells (Chen et al., 2012; Weng et al., 2015).

Thus, it appears that level of HMGN1 and 2 is tightly regulated during neuronal differentiation and follows the nation expression pattern as cellular differentiation of other tissue. Furthermore, the HMGN1 and 2 are highly associated with proliferative NPCs.

6.6.3 Loss of HMGN2 increases spontaneous neuronal differentiation of undifferentiated P19 EC cells

Several proneural- and neuronal-related genes were induced in the HMGN2 knockout lines (B8ΔHMGN2, B19ΔHMGN2 and B38ΔHMGN2) but not in the mutant ΔNBD lines and in controls. However, the expression of these genes was not as high as when all cells are directly induced down the neural lineage. Consequently, the increased expression of these genes could be occurring in all cells equally, or sporadically in a smaller proportion of cells. The immunostaining for neuronal specific markers (ß-III tubulin and Map2) demonstrates that a small proportion of cells expressed these markers and had a neuronal morphology in the HMGN2 knockout cultures. Therefore, the induction of neuronal genes occurs in some cells in culture but not in all. Accordingly, we refer this phenomenon as increased spontaneous neuronal differentiation.

Forced expression of *Zfp521* in ESCs induces the production of neural progenitors in culture conditions that would normally prevent the neural differentiation (Kamiya et al., 2011). Although more replicates are required to be confident of the changes in *Zfp521* expression, this could be the first change in HMGN2 knockout cells that triggers differentiation into neural lineage. The proneural genes (*Ngns* and *Mash1*) also perform a neurogenic determination function, and are expressed in neuronal precursors before overt neuronal differentiation in precursors. *NeuroD* tends to be expressed mostly in more committed NPCs, post-mitotic precursors or differentiated neurons (Lee et al., 1995; Kim et al., 2004). As the expression of the proneural genes inhibits the astrogenesis, the loss of HMGN2 enhance specifically neuronal fate. *Pax6* is also a crucial intrinsic determinant for the neurogenic fate of cortical radial glia precursors (Heins et al., 2002). All these genes that promote the neurogenic fate were up-regulated in HMGN2 knockout cells.

ES and EC cells are known to be able to spontaneously differentiate into cells of different lineages, and spontaneous differentiation is triggered by withdrawal the leukaemia inhibitory factor (LIF) or serum, respectively (Heo et al., 2005). Expression of mesoderm and endoderm marker genes (*Brachyury* and *Gata4*) did not show differential expression between the HMGN2 knockout cells and controls (Garza-Manero S, from our lab).

Taken together, this data indicates that the complete loss of HMGN2 protein, but not the NBD-mutant proteins with less ability to bind to chromatin, leads to increased expression of neuronal-fate determination genes, and consequently, an increase in the rate of spontaneous neuronal differentiation of P19 cells.

6.6.3.1 What is the mechanism of increased spontaneous neuronal differentiation in HGM2 knockout cells?

Increased spontaneous neuronal differentiation could either result from the direct activation of specific pro-neural genes, or from the loss of self-renewal, leading indirectly to more spontaneous differentiation. In other words, HMGN2 could minimise the spontaneous differentiation in ECCS by either blocking neuronal differentiation or sustaining self-renewal or both. These two possibilities are discussed below.

6.6.3.2 HMGN2 may directly mediate the induction of neuronal determination genes

The induction of neuronal-specific genes in HMGN2 knockout lines may be due to direct regulation of these genes by HMGN2. It is worth remembering that HMGN expression increases during neuronal induction in the P19 system, and is known to be high in areas of neurogenesis during development. Therefore, it is unlikely that HMGN proteins inhibit pro-neural gene expression in all cell types. However, HMGNs are known to affect the gene expression in a tissue specific manner (Kugler et al., 2013), so it is possible that they inhibit pro-neural gene expression in ECCs, but do not play this role in NSC/NPCs, when the signalling pathways driving the neuronal lineage are dominant.

Interestingly, although the level of differential induction in gene expressions was variable depending on the gene and the knockout lines, the most highly induced genes in the HMGN2 knockout lines were *Ngn1* and *Mash1*, with 20-60-fold higher expression than the parental cells. The differential induction of other genes did not exceed 7-fold. It has been reported that transient ectopic expression of *Ngn1* in P19 cells and ESCs is sufficient to induce neuronal differentiation in the absence of other inducing factors (Kim et al., 2004; Velkey et al., 2013). Conversely, forced expression of *Mash1* is not sufficient to induce phenotypical changes (Johnson et al., 1992; Kim et al., 2004). Therefore, the loss HMGN2 may

directly regulate the expression of *Ngn1*, which could then drive the neuronal differentiation. It might be worth studying the binding of HMGN2 and different histone marks at the regulatory sites of this gene to determine if there is a direct correlation. In addition, the DNase I hypersensitive sites around this gene could be studied, as there has been shown to be an association between DHSs and HMGN protein binding (Deng et al., 2013; Deng et al., 2015).

In the study by Velkey et al, (2013), inducible-forced expression of *Ngn1* induced the neuronal differentiation of ESCs in both ES medium, that inhibits the spontaneous differentiation, and a defined neural medium which supports neural differentiation of ESCs, but differentiation progressed more rapidly in the latter condition. Thus, it would be expected that more neuronal differentiation of $HMGN2^{-/-}$ P19 cells would be obtained if the cells were grown in defined neural medium and all cells have higher *Ngn1*. Additionally, the forced expression of *Ngn1* in ESCs can reduce the proliferation and increase G1 phase of the cell cycle (Velkey et al., 2013). However, there was no apparent effect on the proliferation and cell cycle by knocking out HMGN2, as seen previously in Chapter 5. This discrepancy may be because the up-regulation of *Ngn1* may not occur in all cells; consequently, the frequency of these cells could be too low to detect a differential effect on the proliferation rate and cell cycle distribution. In addition, the regulation of cell cycle and proliferation could be different in the embryonal carcinoma cells compared to ESCs.

6.6.3.3 The up-regulation of neuronal genes could be indirectly linked to loss of HMGN2

Alternatively, it is possible that pro-neural gene expression is not directly regulated by HMGN2. It is possible that the reduction in *Nanog* expression, which was observed in the HMGN2 knockout lines, results in less efficient maintenance of self renewal, and the cells choose the neural lineage because the neural induction is known as the default model (Temple, 2001; Munnoz-Sanjuan and Brivanlou, 2002). Another possibility is that the effect of HMGN2 on the neural and neuronal genes is a secondary effect of changes in other self-renewal or differentiation signalling pathways. One study showed that HMGN2 interact with the PITX2 homeodomain (HD) protein to form an inactive complex; upon Wnt signalling, ß-catenin would form a ternary complex with PITX2/HMGN2 and

switch this complex to a highly active form (Amen et al., 2008). They speculate that ß-catenin binding to the PITX2/HMGN2 complex causes a conformational change in HMGN2 that allows for the PITX2 protein to recognise its DNA target sequence and promotes the association of a higher ordered chromatin-associated transcription complex. Wnt/ß-catenin signalling pathway is important in controlling the mouse development. Increasing the level of ß-catenin as a result of mutation in *Apc* gene can inhibit the differentiation capacity of ESCs into three germ layers (Kielman et al., 2002). Similarly, inhibiting of the GSK3ß in the mESCs promotes the self-renewal (Sato et al., 2004).

It is striking that most of the changes in gene expression and phenotypes reported here were only observed in the HMGN2 knockout cells, and not in the NBD-mutant cells, 19ΔNBD. Loss part of the NBD is expected to reduce the binding of HMGN2 to nucleosome, and this was confirmed by the ChIP experiments. The functionality of HMGN proteins is reported to be dependent on the binding of HMGN into chromatin (Crippa et al., 1993; Ding et al., 1994: Trieschmann et al.,1995). It is possible that the residual nucleosome binding of the 19ΔNBD protein is sufficient to prevent the induction of pro-neural factors and spontaneous differentiation. Alternatively, the increased *Nanog* expression in the 19ΔNBD cells may counteract any direct or indirect effect of loss HMGN2 on pro-neural gene expression.

6.6.3.4 Why do only some of the HMGN2-knockout cells differentiate into neurons?

Regardless of how the loss of HMGN2 stimulates the neuronal-specific genes, it is interesting to ask why not all the cells differentiated into neurons. Although the expression of genes associated with neurogenesis, including *Ngn1*, *Mash1*, *NeuroD*, *Pax6* and *Zfp521*, plus *Nestin*, *Map2* and *Nf160,* were induced, the negative regulator of neurogenesis *Rest* (RE1-silencing factor) was also slightly and unexpectedly up-regulated in the HMGN2$^{-/-}$ lines. This increase in Rest expression might be the reason of why not all cells committed to neurons. It is unclear if the increase of the early neuronal fate determination genes (specifically *Ngn1*, *Mash1* and *Rest*) happened only in some cells. In order to answer this, it would be worth performing immunostaining for these markers in the different cell lines. However, even if *Ngn1* was expressed in all cells, it

would not necessarily mean all the cells should mature to form neurons. This is because it was observed that not all ES or EC P19 cells matured into neurons although all cells had over-expression of *Ngn1*, even after 5 days (Kim et al., 2004; Velkey et al., 2014). Velkey and colleagues (2014) referred this to the Notch-Delta signalling pathway, that plays a role in maintaining the NSCs, as they observed up-regulation in Notch ligand (*Jag1*), and *Dll1*, *3*, *4*, as well as repressor-type bHLH genes (*Hes1*, *5* and *6*) that are downstream targets of Notch signalling pathway. Therefore, loss of HMGN2 may induce the expression of neuronal fate determination genes in all cells but the up-regulation of *Rest* and possibly Notch-related genes may prevent any further neuronal differentiation in the many of the HMGN2 knockout cells.

To conclude, HMGN2 affects the spontaneous differentiation of EC P19 cells into neuronal lineage by regulating, either directly or indirectly, the expression of neuronal-associated genes and/or self-renewal genes (*Nanog*).

6.6.4 Loss of HMGN2 effectively induces the neuronal differentiation of P19 cells after induction

Although the expression of several neural and neuronal-specific genes is higher in the undifferentiated HMGN2 knockout cells, and is associated with more spontaneous neuronal differentiation, proper neural induction is required to achieve maximum differentiation. The induction of parental, HMGN2$^{-/-}$ and NBD-mutant lines down the neuronal lineage revealed differential expression of most of the neural and neuronal genes at day 3 after induction. Specifically, the expression of *Mash1*, *NeuroD*, *Map2* and *Nf160* was consistently higher after induction in all three HMGN2$^{-/-}$ lines than in parental cells or 19ΔNBD cells. Expression of *Ngn1*, *Zfp521* and *Pax6* was higher in B8ΔHMGN2 and B19ΔHMGN2 cells, but not in the B38ΔHMGN line. However, *Pax6* expression was also more highly induced in 19ΔNBD cells at day 3. Regardless of this conflict, the main theme of the preliminary data implies that the loss of HMGN2 results in more efficient, or more rapid induction down the neuronal lineage at day 3. However, further replicates and experiments are required before drawing a final conclusion. In particular, analysis of several time points following induction would reveal whether the pro-neural genes are really expressed at higher levels

in the HMGN2 knockout cells compared to the parental cells, or whether the time course of induction is different.

The inefficient induction of *Nestin* expression at day 3 in the knockout lines is unexpected because it is opposite to the increase in expression of most of other neural-related gene. However, as this intermediate filament is expressed in the transition state (particularly at progenitor cells) and is often reduced with the concomitant up-regulation of other tissue-specific intermediate filament protein. Therefore, the reduction in nestin might be correlated with the increase of neuron-specific intermediate filaments such as NF160.

By day 6, the expression of many of the pro-neural and neuronal factors was decreased, and there were few differences between parental and mutant cell lines. This suggests that the final production of neurons is not affected by the HMGN2 knockout. However, the unexpected increase in HMGN2 protein in the B8ΔHMGN2 knockout lines at day 6 means that this data must be interpreted with caution, and further experiments are required.

Consistent with our preliminary findings, a recent study on HMGN proteins in neocortical NPCs revealed that overexpression of any HMGN1, 2 or 3 *in vitro* or *in vivo* at postnatal stages increases the fraction of astrocytes and reduces the neurons, while the knockdown system reveals the opposite (Nagao et al., 2014). This means that HMGNs promote astrocyte differentiation of NPCs over the neuronal lineage. The higher production of neurons in the knockdown models may be similar to our observations of having higher expression of neuronal genes at day 3. However, because our neural induction system does not induce the cells into astrocytic lineage, we could not estimate the effect of complete loss of HMGN2 on the astrocyte production. Unlike of HMGNs, another study showed that HMGA family proteins could increase the neurogenic potential of NPCs into the neuronal lineage at the early stages of neocortical development, possibly through induction of an open chromatin state (Kishi et al., 2012).

Chapter 7 Conclusion and future work

7.1 Summary and conclusion

Although there have been many studies on the biochemical mechanisms and biological role of HMGN proteins, the importance of these proteins in ESCs and their malignant counterpart ECCs, as well as during neural induction, has not been comprehensively investigated. Our preliminary data indicates that these proteins, especially HMGN2, may affect the ECCs self-renewal and neuronal differentiation. In this study, we used the CRISPR paired nickase system to knockout or mutate HMGN2 in order to study its role in ECCs and during neuronal differentiation. We targeted exon I in order to prevent the expression of *Hmgn2* completely, and targeted exon IV to disrupt the nucleosome binding domain. Although CRISPR-Cas9D10A is considered to be an efficient method for gene editing, only five lines with biallelic mutations were obtained from these two CRISPR-Cas9D10A experiments. This indel mutation rate can be affected by several factors such as transfection efficiency and type of cells.

Analysis of the HMGN2 mutant and knockout cells indicated that, in the undifferentiated state, loss of HMGN2 does not appear to affect the proliferation rate and the cell cycle progression. However, loss of HMGN2 does result in more spontaneous differentiation into the neural lineage in culture conditions that normally promote self-renewal. This observation was seen in all HMGN2$^{-/-}$ lines but not in the ΔNBD-mutant lines or negative control lines. The increase in the rate of spontaneous neuronal differentiation indicates that HMGN2 may promote self-renewal or suppress neural induction in ECCs. Neural induction is known to be a default pathway that only requires the inhibition of pathways that drive self-renewal of ESCs. However, our preliminary data suggests that there is no increase in the rate of spontaneous differentiation into non-neuronal lineages in HMGN2$^{-/-}$ lines, suggesting that the effect of loss HMGN2 may be limited to regulating the neuronal lineage.

Interestingly, both the down-regulation of pluripotency gene, *Nanog*, and the up-regulation of neuronal-fate determination genes such as pro-neural genes (*Ngn1* and *Mash1*) and *Zfp521* were observed in the HMGN2$^{-/-}$ cells but not in mutant ΔNBD lines, which is compatible with the finding of more spontaneous neural

differentiation in HMGN2$^{-/-}$ cells. Therefore, HMGN2 affects the spontaneous differentiation of EC P19 cells by regulating, either directly or indirectly, the expression of neuronal-associated genes and/or self-renewal genes (*Nanog*). The role of HMGN proteins in regulating gene expression has been studied extensively, and these proteins were found to have variant-specific and tissue-specific effects on the fidelity of the cellular transcription profile (Rochman et al., 2011; Kugler et al., 2013; Deng et al., 2015). Therefore, it is possible that loss of HMGN2 results in up- and down-regulation of other genes that impact on self-renewal and neural induction.

It is not clear why the complete knockout of HMGN2 would down-regulate *Nanog* expression, yet the ΔNBD mutant would lead to *Nanog* upregulation. It is possible that the interaction of the ΔNBD mutant with nucleosomes might be different to that of the wild type protein, altering the recruitment of co-factors, nucleosome remodelling complexes or histone modifying enzymes, which could differentially regulate *Nanog* expression.

We showed that HMGN2 binds equally to the promoter and gene body of the *Nanog* and *Glyt1* genes in HMGN2$^{+/+}$ cells, but the enrichment is lost or reduced in complete knockout lines or NBD-mutant lines, respectively. Our preliminary ChIP data suggests an effect of HMGN2 on H3K27ac enrichment, and possibly H3K27me3, at the *Nanog* and *Glyt1* genes. Specifically, removal of HMGN2 in P19 EC cells led to a reduction in H3K27ac at the *Nanog and Glyt1* loci whereas the partial binding of the NBD-mutant HMGN2 to these loci did not change the enrichment of this histone modification mark. This is also well correlated with expression status of these two genes in HMGN2$^{-/-}$ and HMGN2ΔNBD cells. The effect of HMGN proteins on histone acylation (H3K14ac) was shown previously. For example, HMGN1 enhances the ability of P300/CBP-associated factor (PCAF) to acetylate H3K14, which is linked to transcriptional activation and gene expression (Herrera et al., 1999; Lim et al., 2005; Ueda et al., 2006). Notably, the reduction of H3K27ac at genome-specific sites is in agreement with the global reduction of H3K27ac. Although H3K27me3 appeared to be reduced at the *Nanog* and *Glyt1* loci in HMGN2 knockout cells, western blotting indicated a global increase of H3K27me3 in HMGN2$^{-/-}$ cells. H3K27ac and H3K27me3 are on the same lysine reside, and it was reported that the active transcriptional mark

H3K27ac can antagonise the transcriptional silencing by H3K27me3 that is mediated by PRC2 (Tie et al., 2009). Thus, HMGN2 may directly or indirectly regulate the interplay between the two marks. However, this experiment was performed on one biological chromatin sample and more replicates are required. It is also important to note that the genomic locations that were studied might not be representative of other genes, and ChIP-seq is required in order to obtain a comprehensive picture of the changes in each histone mark.

In conclusion, HMGN2 seems to be an important intrinsic factor that preserves the expression of pluripotency gene (*Nanog*) and neurogenic determination genes that may work together in maintaining the self-renewal of pluripotent P19 EC cells and preventing the spontaneous differentiation into neuronal lineage. HMGN2 may also affect the transcription profile during the neural induction of P19 EC cells. Therefore, HMGN2 plays role in regulating the ECCs and neural differentiation and a further investigation are still required.

7.2 Possible future works

Several aspects can be investigated in order to confirm results obtained in this project, and many of these were suggested during the chapters. However, some critical suggestions are mentioned here.

Several genes are differentially expressed in undifferentiated HMGN2 knockout lines and NBD-mutant lines, including *Nanog*, *Nestin*, *Ngn1*, *Mash1*, *NeuroD*, *Zfp521*, *Rest*, *Map2* and *Nf160*. It is possible that the effect of HMGN2 on these neural and neuronal genes is a secondary effect of changes in other self-renewal or differentiation signalling pathways, for example Wnt or Notch signalling. Genome-wide gene expression analysis such as RNA-seq is required to identify other genes affected by HMGN2. This may also assist in identifying if there are specific pathways regulated by HMGN2 that are involved in the maintenance of ESCs self-renewal or neural differentiation.

The spontaneous neuronal differentiation was observed using immunofluorescent staining for βIII-tubulin and Map2. Using the same technique, it would be helpful to investigate whether the reduction in Nanog protein and the increase in Ngn1

and Mash1 (most induced in mRNA data) proteins occurs in just few cells or globally.

It is also important to investigate whether HMGN2 has a direct effect on the histone modifications studied here, and how this relates to the changes in gene expression. A first step could be to study HMGN2 binding and histone modifications at genes such as *Ngn1* and *Mash1*, but more extensive data would be obtained from ChIP-seq analysis, and comparing this to RNA-seq data. One histone mark that appears to be affected by loss of HMGN2 but not by the NBD-mutant HMGN2 is the H3K27ac. H3K27ac enrichment is apparently reduced globally and at the specific *Nanog* and *Glyt1* loci in HMGN2$^{-/-}$ cells. Therefore, it might be useful to focus on H3K27ac initially, by performing ChIP-seq for this mark and for HMGN2 to investigate the effect of HMGN2 on this mark quantitatively and globally.

Finally, the findings in this project suggest that HMGN2 is likely to be involved in regulating the self-renewal and probably neural induction of embryonic carcinoma stem cells (P19 line). It is important to find out whether HMGN2 plays a similar role in non-malignant stem cells such as mouse ESCs (E14). Performing this experiment in mESCs would also enable further investigation of the role of HMGN2, as the gene regulatory networks involved in self renewal and neuronal differentiation of mESCs are well understood. Using CRISPR-Cas9 to create double knockouts of HMGN2 and HMGN1 in mESCs would also extend this investigation, and might create stronger phenotypes for further study.

List of References

Abranches, E., Silva, M., Pradier, L., Schulz, H., Hummel, O., Henrique, D., & Bekman, E. (2009). Neural Differentiation of Embryonic Stem Cells In Vitro: A Road Map to Neurogenesis in the Embryo. *PLoS One, 4*(7). doi:ARTN e6286

10.1371/journal.pone.0006286

Abuhatzira, L., Shamir, A., Schones, D. E., Schaffer, A. A., & Bustin, M. (2011). The Chromatin-binding Protein HMGN1 Regulates the Expression of Methyl CpG-binding Protein 2 (MECP2) and Affects the Behavior of Mice. *Journal of Biological Chemistry, 286*(49), 42051-42062. doi:10.1074/jbc.M111.300541

Alfonso, P. J., Crippa, M. P., Hayes, J. J., & Bustin, M. (1994). The Footprint of Chromosomal-Proteins Hmg-14 and Hmg-17 on Chromatin Subunits (Vol 236, Pg 189, 1994). *Journal of Molecular Biology, 239*(3), 436-436.

Aloia, L., Di Stefano, B., & Di Croce, L. (2013). Polycomb complexes in stem cells and embryonic development. *Development, 140*(12), 2525-2534. doi:10.1242/dev.091553

Amen, M., Espinoza, H. M., Cox, C., Liang, X., Wang, J., Link, T. M. E., . . . Amendt, B. A. (2008). Chromatin-associated HMG-17 is a major regulator of homeodomain transcription factor activity modulated by Wnt/beta catenin signaling. *Nucleic Acids Res, 36*(2), 462-476. doi:10.1093/nar/gkm1047

Andreason, G. L., & Evans, G. A. (1989). Optimization of electroporation for transfection of mammalian cell lines. *Anal Biochem, 180*(2), 269-275.

Andrews, P. W., Casper, J., Damjanov, I., DugganKeen, M., Giwercman, A., Hata, J., . . . Stern, P. (1996). Comparative analysis of cell surface antigens expressed by cell lines derived from human germ cell tumours. *International Journal of Cancer, 66*(6), 806-816. doi:Doi 10.1002/(Sici)1097-0215(19960611)66:6<806::Aid-Ijc17>3.0.Co;2-0

Andrews, P. W., Matin, M. M., Bahrami, A. R., Damjanov, I., Gokhale, P., & Draper, J. S. (2005). Embryonic stem (ES) cells and embryonal carcinoma (EC) cells: Opposite sides of the same coin. *Biochemical Society Transactions, 33*, 1526-1530.

Ang, Y. S., Tsai, S. Y., Lee, D. F., Monk, J., Su, J., Ratnakumar, K., . . . Lemischka, I. R. (2011). Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell, 145*(2), 183-197. doi:10.1016/j.cell.2011.03.003

Angello, J. C., Stern, H. M., & Hauschka, S. D. (1997). P19 embryonal carcinoma cells: A model system for studying neural tube induction of skeletal myogenesis. *Developmental Biology, 192*(1), 93-98. doi:DOI 10.1006/dbio.1997.8722

Ariff, I. M., Mitra, A., & Basu, A. (2012). Epigenetic regulation of self-renewal and fate determination in neural stem cells. *Journal of Neuroscience Research, 90*(3), 529-539. doi:10.1002/jnr.22804

Atkinson, S., & Armstrong, L. (2008). Epigenetics in embryonic stem cells: regulation of pluripotency and differentiation. *Cell and Tissue Research, 331*(1), 23-29. doi:10.1007/s00441-007-0536-x

Attanasio, C., David, A., & Neerman-Arbez, M. (2003). Outcome of donor splice site mutations accounting for congenital afibrinogenemia reflects order of intron removal in the fibrinogen alpha gene (FGA). *Blood, 101*(5), 1851-1856. doi:10.1182/blood/2002-03-0853

Aubin, J., Davy, A., & Soriano, P. (2004). In vivo convergence of BMP and MAPK signaling pathways: impact of differential Smad1 phosphorylation on development and homeostasis. *Genes Dev, 18*(12), 1482-1494. doi:10.1101/gad.1202604

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N., & Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev, 17*(1), 126-140. doi:10.1101/gad.224503

Backdahl, L., Herberth, M., Wilson, G., Tate, P., Campos, L. S., Cortese, R., . . . Beck, S. (2009). Gene body methylation of the dimethylarginine dimethylamino-hydrolase 2 (Ddah2) gene is an epigenetic biomarker for neural stem cell differentiation. *Epigenetics, 4*(4), 248-254.

Bain, G., & Gottlieb, D. I. (1994). Expression of Retinoid-X Receptors in P19 Embryonal Carcinoma-Cells and Embryonic Stem-Cells. *Biochemical and Biophysical Research Communications, 200*(3), 1252-1256. doi:DOI 10.1006/bbrc.1994.1585

Bain, G., Kitchens, D., Yao, M., Huettner, J. E., & Gottlieb, D. I. (1995). Embryonic Stem-Cells Express Neuronal Properties in-Vitro. *Developmental Biology, 168*(2), 342-357. doi:DOI 10.1006/dbio.1995.1085

Ballas, N., Battaglioli, E., Atouf, F., Andres, M. E., Chenoweth, J., Anderson, M. E., . . . Mandel, G. (2001). Regulation of neuronal traits by a novel transcriptional complex. *Neuron, 31*(3), 353-365. doi:Doi 10.1016/S0896-6273(01)00371-3

Ballas, N., Grunseich, C., Lu, D. D., Speh, J. C., & Mandel, G. (2005). REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell, 121*(4), 645-657. doi:10.1016/j.cell.2005.03.013

Barkess, G., Postnikov, Y., Campos, C. D., Mishra, S., Mohan, G., Verma, S., . . . West, K. L. (2012). The chromatin-binding protein HMGN3 stimulates histone acetylation and transcription across the Glyt1 gene. *Biochemical Journal, 442*, 495-505. doi:10.1042/Bj20111502

Bavykin, S. G., Usachenko, S. I., Lishanskaya, A. I., Shick, V. V., Belyavsky, A. V., Undritsov, I. M., . . . Mirzabekov, A. D. (1985). Primary Organization of Nucleosomal Core Particles Is Invariable in Repressed and Active Nuclei from Animal, Plant and Yeast-Cells. *Nucleic Acids Res, 13*(10), 3439-3459. doi:DOI 10.1093/nar/13.10.3439

Begum, N., Pash, J. M., & Bhorjee, J. S. (1990). Expression and Synthesis of High Mobility Group Chromosomal-Proteins in Different Rat Skeletal Cell-Lines during Myogenesis. *Journal of Biological Chemistry, 265*(20), 11936-11941.

Belova, G. I., Postnikov, Y. V., Furusawa, T., Birger, Y., & Bustin, M. (2008). Chromosomal protein HMGN1 enhances the heat shock-induced remodeling of Hsp70 chromatin. *Journal of Biological Chemistry, 283*(13), 8080-8088. doi:10.1074/jbc.M709782200

Bergel, M., Herrera, J. E., Thatcher, B. J., Prymakowska-Bosak, M., Vassilev, A., Nakatani, Y., . . . Bustin, M. (2000). Acetylation of novel sites in the nucleosomal binding domain of chromosomal protein HMG-14 by p300 alters its interaction with nucleosomes. *Journal of Biological Chemistry, 275*(15), 11514-11520. doi:DOI 10.1074/jbc.275.15.11514

Bernstein, B. E., Mikkelsen, T. S., Xie, X. H., Kamal, M., Huebert, D. J., Cuff, J., . . . Lander, E. S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell, 125*(2), 315-326. doi:10.1016/j.cell.2006.02.041

Besirli, C. G., Deckwerth, T. L., Crowder, R. J., Freeman, R. S., & Johnson, E. M. (2003). Cytosine arabinoside rapidly activates Bax-dependent apoptosis and a delayed Bax-independent death pathway in sympathetic neurons. *Cell Death and Differentiation, 10*(9), 1045-1058. doi:10.1038/sj.cdd.4401259

Bibikova, M., Beumer, K., Trautman, J. K., & Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. *Science, 300*(5620), 764. doi:10.1126/science.1079512

Bibikova, M., Carroll, D., Segal, D. J., Trautman, J. K., Smith, J., Kim, Y. G., & Chandrasegaran, S. (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol, 21*(1), 289-297. doi:10.1128/MCB.21.1.289-297.2001

Bibikova, M., Golic, M., Golic, K. G., & Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics, 161*(3), 1169-1175.

Birger, Y., Catez, F., Furusawa, T., Lim, J. H., Prymakowska-Bosak, M., West, K. L., . . . Bustin, M. (2005). Increased tumorigenicity and sensitivity to ionizing radiation upon loss of chromosomal protein HMGN1. *Cancer Research, 65*(15), 6711-6718. doi:Doi 10.1158/0008-5472.Can-05-0310

Birger, Y., Ito, Y., West, K. L., Landsman, D., & Bustin, M. (2001). HMGN4, a newly discovered nucleosome-binding protein encoded by an intronless gene. *DNA and Cell Biology, 20*(5), 257-264. doi:10.1089/104454901750232454

Birger, Y., West, K. L., Postnikov, Y. V., Lim, J. H., Furusawa, T., Wagner, J. P., . . . Bustin, M. (2003). Chromosomal protein HMGN1 enhances the rate of DNA repair in chromatin. *Embo Journal, 22*(7), 1665-1675. doi:Doi 10.1093/Emboj/Cdg142

Bloushtain-Qimron, N., Yao, J., Shipitsin, M., Maruyama, R., & Polyak, K. (2009). Epigenetic patterns of embryonic and adult stem cells. *Cell Cycle, 8*(6), 809-817. doi:DOI 10.4161/cc.8.6.7938

Blumberg, B., Bolado, J., Moreno, T. A., Kintner, C., Evans, R. M., & Papalopulu, N. (1997). An essential role for retinoid signaling in anteroposterior neural patterning. *Development, 124*(2), 373-379.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., . . . Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science, 326*(5959), 1509-1512. doi:10.1126/science.1178811

Boeuf, H., Hauss, C., DeGraeve, F., Baran, N., & Kedinger, C. (1997). Leukemia inhibitory factor-dependent transcriptional activation in embryonic stem cells. *Journal of Cell Biology, 138*(6), 1207-1217. doi:DOI 10.1083/jcb.138.6.1207

Bouma, G. J., Hart, G. T., Washburn, L. L., Recknagel, A. K., & Eicher, E. M. (2004). Using real time RT-PCR analysis to determine multiple gene expression patterns during XX and XY mouse fetal gonad development. *Gene Expression Patterns, 5*(1), 141-149. doi:10.1016/j.modgep.2004.05.001

Bourillot, P. Y., Aksoy, I., Schreiber, V., Wianny, F., Schulz, H., Hummel, O., . . . Savatier, P. (2009). Novel STAT3 Target Genes Exert Distinct Roles in the Inhibition of Mesoderm and Endoderm Differentiation in Cooperation with Nanog. *Stem Cells, 27*(8), 1760-1771. doi:10.1002/stem.110

Brewer, G. J., Torricelli, J. R., Evege, E. K., & Price, P. J. (1993). Optimized Survival of Hippocampal-Neurons in B27-Supplemented Neurobasal(Tm), a New Serum-Free Medium Combination. *Journal of Neuroscience Research, 35*(5), 567-576. doi:DOI 10.1002/jnr.490350513

Brinster, R. L. (1974). Effect of Cells Transferred into Mouse Blastocyst on Subsequent Development. *Journal of Cell Biology, 63*(2), A37-A37.

Brons, I. G. M., Smithers, L. E., Trotter, M. W. B., Rugg-Gunn, P., Sun, B. W., Lopes, S. M. C. D., . . . Vallier, L. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature, 448*(7150), 191-U197. doi:10.1038/nature05950

Bruce, A. W., Donaldson, I. J., Wood, I. C., Yerbury, S. A., Sadowski, M. I., Chapman, M., . . . Buckley, N. J. (2004). Genome-wide analysis of repressor element 1 silencing transcription factor/neuron-restrictive silencing factor (REST/NRSF) target genes. *Proc Natl Acad Sci U S A, 101*(28), 10458-10463. doi:10.1073/pnas.0401827101

Burdon, T., Chambers, I., Stracey, C., Niwa, H., & Smith, A. (1999). Signaling mechanisms regulating self-renewal and differentiation of pluripotent embryonic stem cells. *Cells Tissues Organs, 165*(3-4), 131-143. doi:Doi 10.1159/000016693

Burdon, T., Stracey, C., Chambers, I., Nichols, J., & Smith, A. (1999). Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. *Developmental Biology, 210*(1), 30-43. doi:DOI 10.1006/dbio.1999.9265

Burgold, T., Spreafico, F., De Santa, F., Totaro, M. G., Prosperini, E., Natoli, G., & Testa, G. (2008). The histone H3 lysine 27-specific demethylase Jmjd3 is required for neural commitment. *PLoS One, 3*(8), e3034. doi:10.1371/journal.pone.0003034

Bustin, M. (1999). Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. *Mol Cell Biol, 19*(8), 5237-5246.

Bustin, M. (2001a). Chromatin unfolding and activation by HMGN(*) chromosomal proteins. *Trends Biochem Sci, 26*(7), 431-437.

Bustin, M. (2001b). Revised nomenclature for high mobility group (HMG) chromosomal proteins. *Trends Biochem Sci, 26*(3), 152-153.

Bustin, M., Crippa, M. P., & Pash, J. M. (1990). Immunochemical analysis of the exposure of high mobility group protein 14 and 17 surfaces in chromatin. *J Biol Chem, 265*(33), 20077-20080.

Bustin, M., Soares, N., Landsman, D., Srikantha, T., & Collins, J. M. (1987). Cell-Cycle Regulated Synthesis of an Abundant Transcript for Human Chromosomal Protein Hmg-17. *Nucleic Acids Res, 15*(8), 3549-3561. doi:DOI 10.1093/nar/15.8.3549

Byrne, S. M., Mali, P., & Church, G. M. (2014). Genome editing in human stem cells. *Methods Enzymol, 546*, 119-138. doi:10.1016/B978-0-12-801185-0.00006-4

Cai, C., & Grabel, L. (2007). Directing the differentiation of embryonic stem cells to neural stem cells. *Dev Dyn, 236*(12), 3255-3266. doi:10.1002/dvdy.21306

Capecchi, M. R. (1989). Altering the Genome by Homologous Recombination. *Science, 244*(4910), 1288-1292. doi:DOI 10.1126/science.2660260

Caric, D., Gooday, D., Hill, R. E., McConnell, S. K., & Price, D. J. (1997). Determination of the migratory capacity of embryonic cortical cells lacking the transcription factor Pax-6. *Development, 124*(24), 5087-5096.

Carroll, D. (2014). Genome engineering with targetable nucleases. *Annu Rev Biochem, 83*, 409-439. doi:10.1146/annurev-biochem-060713-035418

Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., & Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development, 132*(5), 885-896. doi:10.1242/dev.01670

CastroObregon, S., & Covarrubias, L. (1996). Role of retinoic acid and oxidative stress in embryonic stem cell death and neuronal differentiation. *Febs Letters, 381*(1-2), 93-97. doi:Doi 10.1016/0014-5793(96)00088-9

Catez, F., Brown, D. T., Misteli, T., & Bustin, M. (2002). Competition between histone H1 and HMGN proteins for chromatin binding site. *EMBO Rep, 3*(8), 760-766. doi:10.1093/embo-reports/kvf156

Catez, F., Ueda, T., & Bustin, M. (2006). Determinants of histone H1 mobility and chromatin binding in living cells. *Nature Structural & Molecular Biology, 13*(4), 305-310. doi:10.1038/nsmb1077

Catez, F., Yang, H., Tracey, K. J., Reeves, R., Misteli, T., & Bustin, M. (2004). Network of dynamic interactions between histone H1 and high-mobility-group proteins in chromatin. *Mol Cell Biol, 24*(10), 4321-4328. doi:10.1128/Mcb.24.10.4321-4328.2004

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., & Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell, 113*(5), 643-655.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., . . . Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. *Nature, 450*(7173), 1230-U1238. doi:10.1038/nature06403

Chambers, I., & Smith, A. (2004). Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene, 23*(43), 7150-7160. doi:10.1038/sj.onc.1207930

Chan, H. M., & La Thangue, N. B. (2001). p300/CBP proteins: HATs for transcriptional bridges and scaffolds. *Journal of Cell Science, 114*(13), 2363-2373.

Charpentier, E., & Doudna, J. A. (2013). Biotechnology: Rewriting a genome. *Nature, 495*(7439), 50-51. doi:10.1038/495050a

Chen, P., Wang, X. L., Ma, Z. S., Xu, Z., Jia, B., Ren, J., . . . Yang, J. L. (2012). Knockdown of HMGN5 Expression by RNA Interference Induces Cell Cycle Arrest in Human Lung Cancer Cells. *Asian Pacific Journal of Cancer Prevention, 13*(7), 3223-3228. doi:10.7314/Apjcp.2012.13.7.3223

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., . . . Ng, H. H. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell, 133*(6), 1106-1117. doi:10.1016/j.cell.2008.04.043

Cherukuri, S., Hock, R., Ueda, T., Catez, F., Rochman, M., & Bustin, M. (2008). Cell cycle-dependent binding of HMGN proteins to chromatin. *Molecular Biology of the Cell, 19*(5), 1816-1824. doi:10.1091/mbc.E07-10-1018

Chiang, T. W., le Sage, C., Larrieu, D., Demir, M., & Jackson, S. P. (2016). CRISPR-Cas9(D10A) nickase-based genotypic and phenotypic screening to enhance genome editing. *Sci Rep, 6*, 24356. doi:10.1038/srep24356

Chiba, K., & Hockemeyer, D. (2015). Genome editing in human pluripotent stem cells using site-specific nucleases. *Methods Mol Biol, 1239*, 267-280. doi:10.1007/978-1-4939-1862-1_15

Cho, B., Kim, H. J., Kim, H., & Sun, W. (2011). Changes in the Histone Acetylation Patterns during the Development of the Nervous System. *Exp Neurobiol, 20*(2), 81-84. doi:10.5607/en.2011.20.2.81

Cho, S. W., Kim, S., Kim, J. M., & Kim, J. S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol, 31*(3), 230-232. doi:10.1038/nbt.2507

Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., & Kim, J. S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res, 24*(1), 132-141. doi:10.1101/gr.162339.113

Chuang, J. H., Tung, L. C., & Lin, Y. (2015). Neural differentiation from embryonic stem cells in vitro: An overview of the signaling pathways. *World J Stem Cells, 7*(2), 437-447. doi:10.4252/wjsc.v7.i2.437

Clarke, S. R., Shetty, A. K., Bradley, J. L., & Turner, D. A. (1994). Reactive astrocytes express the embryonic intermediate neurofilament nestin. *Neuroreport, 5*(15), 1885-1888.

Cline, S. D., & Osheroff, N. (1999). Cytosine arabinoside lesions are position-specific topoisomerase II poisons and stimulate DNA cleavage mediated by the human type II enzymes. *Journal of Biological Chemistry, 274*(42), 29740-29743. doi:DOI 10.1074/jbc.274.42.29740

Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., . . . Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes Dev, 26*(15), 1714-1728. doi:10.1101/gad.194209.112

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., . . . Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science, 339*(6121), 819-823. doi:10.1126/science.1231143

Cortese, R., Lewin, J., Backdahl, L., Krispin, M., Wasserkort, R., Eckhardt, F., & Beck, S. (2011). Genome-wide screen for differential DNA methylation associated with neural cell differentiation in mouse. *PLoS One, 6*(10), e26002. doi:10.1371/journal.pone.0026002

Crippa, M. P., Alfonso, P. J., & Bustin, M. (1992). Nucleosome core binding region of chromosomal protein HMG-17 acts as an independent functional domain. *J Mol Biol, 228*(2), 442-449.

Crippa, M. P., Nickol, J. M., & Bustin, M. (1991). Differentiation-Dependent Alteration in the Chromatin Structure of Chromosomal Protein Hmg-17 Gene during Erythropoiesis. *Journal of Molecular Biology, 217*(1), 75-84. doi:Doi 10.1016/0022-2836(91)90612-A

Crippa, M. P., Trieschmann, L., Alfonso, P. J., Wolffe, A. P., & Bustin, M. (1993). Deposition of Chromosomal Protein Hmg-17 during Replication Affects the Nucleosomal Ladder and Transcriptional Potential of Nascent Chromatin. *Embo Journal, 12*(10), 3855-3864.

Crossley, P. H., Martinez, S., & Martin, G. R. (1996). Midbrain development induced by FGF8 in the chick embryo. *Nature, 380*(6569), 66-68. doi:DOI 10.1038/380066a0

Cuddapah, S., Schones, D. E., Cui, K. R., Roh, T. Y., Barski, A., Wei, G., . . . Zhao, K. J. (2011). Genomic Profiling of HMGN1 Reveals an Association with Chromatin at Regulatory Regions. *Mol Cell Biol, 31*(4), 700-709. doi:10.1128/Mcb.00740-10

da Silva, P. M., Soudant, P., Carballal, M. J., Lambert, C., & Villalba, A. (2005). Flow cytometric DNA content analysis of neoplastic cells in haemolymph of the cockle Cerastoderma edule. *Dis Aquat Organ, 67*(1-2), 133-139. doi:10.3354/dao067133

Deb-Rinker, P., Ly, D., Jezierski, A., Sikorska, M., & Walker, P. R. (2005). Sequential DNA methylation of the Nanog and Oct-4 upstream regions in human NT2 cells during neuronal differentiation. *Journal of Biological Chemistry, 280*(8), 6257-6260. doi:10.1074/jbc.C400479200

Dehmelt, L., & Halpain, S. (2005). The MAP2/Tau family of microtubule-associated proteins. *Genome Biol, 6*(1), 204. doi:10.1186/gb-2004-6-1-204

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., . . . Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature, 471*(7340), 602-607. doi:10.1038/nature09886

Deng, T., Zhu, Z. I., Zhang, S. F., Leng, F. F., Cherukuri, S., Hansen, L., . . . Bustin, M. (2013). HMGN1 Modulates Nucleosome Occupancy and DNase I Hypersensitivity at the CpG Island Promoters of Embryonic Stem Cells. *Mol Cell Biol, 33*(16), 3377-3389. doi:10.1128/Mcb.00435-13

Deng, T., Zhu, Z. I., Zhang, S. F., Postnikov, Y., Huang, D., Horsch, M., . . . Bustin, M. (2015). Functional compensation among HMGN variants modulates the DNase I hypersensitive sites at enhancers. *Genome Res, 25*(9), 1295-1308. doi:10.1101/gr.192229.115

Denissov, S., Hofemeister, H., Marks, H., Kranz, A., Ciotta, G., Singh, S., . . . Stewart, A. F. (2014). Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development, 141*(3), 526-537. doi:10.1242/dev.102681

Derveaux, S., Vandesompele, J., & Hellemans, J. (2010). How to do successful gene expression analysis using real-time PCR. *Methods, 50*(4), 227-230. doi: 10.1016/j.ymeth.2009.11.001

Dessi, F., Pollard, H., Moreau, J., Benari, Y., & Charriautmarlangue, C. (1995). Cytosine-Arabinoside Induces Apoptosis in Cerebellar Neurons in Culture. *Journal of Neurochemistry, 64*(5), 1980-1987.

Diez del Corral, R., Olivera-Martinez, I., Goriely, A., Gale, E., Maden, M., & Storey, K. (2003). Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. *Neuron, 40*(1), 65-79.

Ding, H. F., Bustin, M., & Hansen, U. (1997). Alleviation of histone H1-mediated transcriptional repression and chromatin compaction by the acidic activation region in chromosomal protein HMG-14. *Mol Cell Biol, 17*(10), 5843-5855.

Ding, H. F., Rimsky, S., Batson, S. C., Bustin, M., & Hansen, U. (1994). Stimulation of Rna-Polymerase-Ii Elongation by Chromosomal Protein Hmg-14. *Science, 265*(5173), 796-799. doi:DOI 10.1126/science.8047885

Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., . . . Root, D. E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol, 32*(12), 1262-1267. doi:10.1038/nbt.3026

Draper, J. S., Pigott, C., Thomson, J. A., & Andrews, P. W. (2002). Surface antigens of human embryonic stem cells: changes upon differentiation in culture. *Journal of Anatomy, 200*(3), 249-258. doi:DOI 10.1046/j.1469-7580.2002.00030.x

Duda, K., Lonowski, L. A., Kofoed-Nielsen, M., Ibarra, A., Delay, C. M., Kang, Q., . . . Frodin, M. (2014). High-efficiency genome editing via 2A-coupled co-expression of fluorescent proteins and zinc finger nucleases or CRISPR/Cas9 nickase pairs. *Nucleic Acids Res, 42*(10), e84. doi:10.1093/nar/gku251

Dudas, A., & Chovanec, M. (2004). DNA double-strand break repair by homologous recombination. *Mutat Res, 566*(2), 131-167. doi:10.1016/j.mrrev.2003.07.001

Easley, C. A., Faison, M. O., Kirsch, T. L., Lee, J. A., Seward, M. E., & Tombes, R. M. (2006). Laminin activates CaMK-II to stabilize nascent embryonic axons. *Brain Research, 1092*, 59-68. doi:10.1016/j.brainres.2006.03.099

Egger, B., Gold, K. S., & Brand, A. H. (2010). Notch regulates the switch from symmetric to asymmetric neural stem cell division in the Drosophila optic lobe. *Development, 137*(18), 2981-2987. doi:10.1242/dev.051250

Endo, M., Antonyak, M. A., & Cerione, R. A. (2009). Cdc42-mTOR signaling pathway controls Hes5 and Pax6 expression in retinoic acid-dependent neural differentiation. *J Biol Chem, 284*(8), 5107-5118. doi:10.1074/jbc.M807745200

Ericson, J., Rashbass, P., Schedl, A., Brenner-Morton, S., Kawakami, A., van Heyningen, V., . . . Briscoe, J. (1997). Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling. *Cell, 90*(1), 169-180.

Ernst, M., Oates, A., & Dunn, A. R. (1996). gp130-mediated signal transduction in embryonic stem cells involves activation of Jak and Ras/mitogen-activated protein kinase pathways. *Journal of Biological Chemistry, 271*(47), 30136-30143.

Eyer, J., & Peterson, A. (1994). Neurofilament-Deficient Axons and Perikaryal Aggregates in Viable Transgenic Mice Expressing a Neurofilament-Beta-Galactosidase Fusion Protein. *Neuron, 12*(2), 389-405. doi:Doi 10.1016/0896-6273(94)90280-1

Fei, T., Xia, K., Li, Z. W., Zhou, B., Zhu, S. S., Chen, H., . . . Chen, Y. G. (2010). Genome-wide mapping of SMAD target genes reveals the role of BMP signaling in embryonic stem cell fate determination. *Genome Res, 20*(1), 36-44. doi:10.1101/gr.092114.109

Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., . . . Chambers, I. (2012). Esrrb Is a Direct Nanog Target Gene that Can Substitute for Nanog Function in Pluripotent Cells. *Cell Stem Cell, 11*(4), 477-490. doi:10.1016/j.stem.2012.08.002

Finch, B. W., & Ephrussi, B. (1967). Retention of Multiple Developmental Potentialities by Cells of a Mouse Testicular Teratocarcinoma during Prolonged Culture in Vitro and Their Extinction Upon Hybridization with Cells of Permanent Lines. *Proc Natl Acad Sci U S A, 57*(3), 615-&. doi:DOI 10.1073/pnas.57.3.615

Finley, M. F. A., Devata, S., & Huettner, J. E. (1999). BMP-4 inhibits neural differentiation of murine embryonic stem cells. *Journal of Neurobiology, 40*(3), 271-287. doi:Doi 10.1002/(Sici)1097-4695(19990905)40:3<271::Aid-Neu1>3.0.Co;2-C

Fouse, S. D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., . . . Fan, G. (2008). Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell, 2*(2), 160-169. doi:10.1016/j.stem.2007.12.011

Frackman, S., Kobs, G., Simpson, D., & Storts, D. (1998). Betaine and DMSO: enhancing agents for PCR. *Promega notes, 65*(27-29), 27-29.

Franco, P. G., Paganelli, A. R., Lopez, S. L., & Carrasco, A. E. (1999). Functional association of retinoic acid and hedgehog signaling in Xenopus primary neurogenesis. *Development, 126*(19), 4257-4265.

Friede, R. L., & Samorajs.T. (1970). Axon Caliber Related to Neurofilaments and Microtubules in Sciatic Nerve Fibers of Rats and Mice. *Anatomical Record, 167*(4), 379-&. doi:DOI 10.1002/ar.1091670402

Friel, R., van der Sar, S., & Mee, P. J. (2005). Embryonic stem cells: Understanding their history, cell biology and signalling. *Advanced Drug Delivery Reviews, 57*(13), 1894-1903. doi:10.1016/j.addr.2005.08.002

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol, 31*(9), 822-826. doi:10.1038/nbt.2623

Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M., & Joung, J. K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol, 32*(3), 279-284. doi:10.1038/nbt.2808

Fujita, P. A., Rhead, B., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Cline, M. S., . . . Kent, W. J. (2011). The UCSC Genome Browser database: update 2011. *Nucleic Acids Res, 39*(Database issue), D876-882. doi:10.1093/nar/gkq963

Furusawa, T., & Cherukuri, S. (2010). Developmental function of HMGN proteins. *Biochim Biophys Acta, 1799*(1-2), 69-73. doi:10.1016/j.bbagrm.2009.11.011

Furusawa, T., Lim, J. H., Catez, F., Birger, Y., Mackem, S., & Bustin, M. (2006). Down-regulation of nucleosomal binding protein HMGN1 expression during embryogenesis modulates Sox9 expression in chondrocytes. *Mol Cell Biol, 26*(2), 592-604. doi:10.1128/Mcb.26.2.592-604.2006

Gabriel, R., Lombardo, A., Arens, A., Miller, J. C., Genovese, P., Kaeppel, C., . . . von Kalle, C. (2011). An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol, 29*(9), 816-823. doi:10.1038/nbt.1948

Gaiano, N., Nye, J. S., & Fishell, G. (2000). Radial glial identity is promoted by Notch1 signaling in the murine forebrain. *Neuron, 26*(2), 395-404. doi:Doi 10.1016/S0896-6273(00)81172-1

Gajovic, S., St-Onge, L., Yokota, Y., & Gruss, P. (1997). Retinoic acid mediates Pax6 expression during in vitro differentiation of embryonic stem cells. *Differentiation, 62*(4), 187-192.

Gan, Y., Tan, J., Yang, J., Zhou, Y., Dai, Y., He, L., . . . Tang, Y. (2015). Knockdown of HMGN5 suppresses the viability and invasion of human urothelial bladder cancer 5637 cells in vitro and in vivo. *Med Oncol, 32*(4), 136. doi:10.1007/s12032-015-0594-y

Garciaramirez, M., Rocchini, C., & Ausio, J. (1995). Modulation of Chromatin Folding by Histone Acetylation. *Journal of Biological Chemistry, 270*(30), 17923-17928.

Garneau, J. E., Dupuis, M. E., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., . . . Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature, 468*(7320), 67-71. doi:10.1038/nature09523

Gaspar-Maia, A., Alajem, A., Meshorer, E., & Ramalho-Santos, M. (2011). Open chromatin in pluripotency and reprogramming (vol 12, pg 36, 2011). *Nature Reviews Molecular Cell Biology, 12*(4). doi:10.1038/nrm3077

Geller, H. M., Cheng, K. Y., Goldsmith, N. K., Romero, A. A., Zhang, A. L., Morris, E. J., & Grandison, L. (2001). Oxidative stress mediates neuronal DNA damage and apoptosis in response to cytosine arabinoside. *Journal of Neurochemistry, 78*(2), 265-275. doi:DOI 10.1046/j.1471-4159.2001.00395.x

Gifford, C. A., Ziller, M. J., Gu, H. C., Trapnell, C., Donaghey, J., Tsankov, A., . . . Meissner, A. (2013). Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells. *Cell, 153*(5), 1149-1163. doi:10.1016/j.cell.2013.04.037

Golebiewska, A., Atkinson, S. P., Lako, M., & Armstrong, L. (2009). Epigenetic Landscaping During hESC Differentiation to Neural Cells. *Stem Cells, 27*(6), 1298-1308. doi:10.1002/stem.59

Gomez-Lopez, S., Wiskow, O., Favaro, R., Nicolis, S. K., Price, D. J., Pollard, S. M., & Smith, A. (2011). Sox2 and Pax6 maintain the proliferative and developmental potential of gliogenic neural stem cells In vitro. *Glia, 59*(11), 1588-1599. doi:10.1002/glia.21201

Gonzalez-Romero, R., Eirin-Lopez, J. M., & Ausio, J. (2015). Evolution of high mobility group nucleosome-binding proteins and its implications for vertebrate chromatin specialization. *Mol Biol Evol, 32*(1), 121-131. doi:10.1093/molbev/msu280

Gotz, M., & Huttner, W. B. (2005). The cell biology of neurogenesis. *Nature Reviews Molecular Cell Biology, 6*(10), 777-788. doi:10.1038/nrm1739

Greenway, D. J., Street, M., Jeffries, A., & Buckley, N. J. (2007). RE1 Silencing transcription factor maintains a repressive chromatin environment in embryonic hippocampal neural stem cells. *Stem Cells, 25*(2), 354-363. doi:10.1634/stemcells.2006-0207

Grove, E. A., Williams, B. P., Li, D. Q., Hajihosseini, M., Friedrich, A., & Price, J. (1993). Multiple Restricted Lineages in the Embryonic Rat Cerebral-Cortex. *Development, 117*(2), 553-561.

Guilinger, J. P., Pattanayak, V., Reyon, D., Tsai, S. Q., Sander, J. D., Joung, J. K., & Liu, D. R. (2014). Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nat Methods, 11*(4), 429-435. doi:10.1038/nmeth.2845

Guo, G., Yang, J., Nichols, J., Hall, J. S., Eyres, I., Mansfield, W., & Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development, 136*(7), 1063-1069. doi:10.1242/dev.030957

Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., . . . Terns, M. P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell, 139*(5), 945-956. doi:10.1016/j.cell.2009.07.040

Halilagic, A., Zile, M. H., & Studer, M. (2003). A novel role for retinoids in patterning the avian forebrain during presomite stages. *Development, 130*(10), 2039-2050. doi:10.1242/dev.00423

Hall, J., Guo, G., Wray, J., Eyres, I., Nichols, J., Grotewold, L., . . . Smith, A. (2009). Oct4 and LIF/Stat3 Additively Induce Kruppel Factors to Sustain Embryonic Stem Cell Self-Renewal. *Cell Stem Cell, 5*(6), 597-609. doi:10.1016/j.stem.2009.11.003

Handler, M., Yang, X. D., & Shen, J. (2000). Presenilin-1 regulates neuronal differentiation during neurogenesis. *Development, 127*(12), 2593-2606.

Hatano, S. Y., Tada, M., Kimura, H., Yamaguchi, S., Kono, T., Nakano, T., . . . Tada, T. (2005). Pluripotential competence of cells associated with Nanog activity. *Mech Dev, 122*(1), 67-79. doi:10.1016/j.mod.2004.08.008

Hattori, N., Imao, Y., Nishino, K., Hattori, N., Ohgane, J., Yagi, S., . . . Shiota, K. (2007). Epigenetic regulation of Nanog gene in embryonic stem and trophoblast stem cells. *Genes to Cells, 12*(3), 387-396. doi:10.1111/j.1365-2443.2007.01058.x

Hattori, N., Nishino, K., Ko, Y. G., Hattori, N., Ohgane, J., Tanaka, S., & Shiota, K. (2004). Epigenetic control of mouse Oct-4 gene expression in embryonic stem cells and trophoblast stem cells. *Journal of Biological Chemistry, 279*(17), 17063-17069. doi:10.1074/jbc.M309002200

Haubensak, W., Attardo, A., Denk, W., & Huttner, W. B. (2004). Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: A major site of neurogenesis. *Proc Natl Acad Sci U S A, 101*(9), 3196-3201. doi:10.1073/pnas.0308600100

Hawkins, R. D., Hon, G. C., Lee, L. K., Ngo, Q., Lister, R., Pelizzola, M., . . . Ren, B. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell, 6*(5), 479-491. doi:10.1016/j.stem.2010.03.018

Hebbes, T. R., Clayton, A. L., Thorne, A. W., & Cranerobinson, C. (1994). Core Histone Hyperacetylation Co-Maps with Generalized Dnase-I Sensitivity in the Chicken Beta-Globin Chromosomal Domain. *Embo Journal, 13*(8), 1823-1830.

Heins, N., Malatesta, P., Cecconi, F., Nakafuku, M., Tucker, K. L., Hack, M. A., . . . Gotz, M. (2002). Glial cells generate neurons: the role of the transcription factor Pax6. *Nature Neuroscience, 5*(4), 308-315. doi:10.1038/nn828

Hellemans, J., Preobrazhenska, O., Willaert, A., Debeer, P., Verdonk, P. C., Costa, T., . . . Mortier, G. R. (2004). Loss-of-function mutations in LEMD3 result in osteopoikilosis, Buschke-Ollendorff syndrome and melorheostosis. *Nature Genetics, 36*(11), 1213-1218. doi: 10.1038/ng1453

Hemmatibrivanlou, A., Kelly, O. G., & Melton, D. A. (1994). Follistatin, an Antagonist of Activin, Is Expressed in the Spemann Organizer and Displays Direct Neuralizing Activity. *Cell, 77*(2), 283-295. doi:Doi 10.1016/0092-8674(94)90320-4

Henke, W., Herdel, K., Jung, K., Schnorr, D., & Loening, S. A. (1997). Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Res, 25*(19), 3957-3958. doi:DOI 10.1093/nar/25.19.3957

Heo, J., Lee, J. S., Chu, I. S., Takahama, Y., & Thorgeirsson, S. S. (2005). Spontaneous differentiation of mouse embryonic stem cells in vitro: Characterization by global gene expression profiles. *Biochemical and Biophysical Research Communications, 332*(4), 1061-1069. doi:10.1016/j.bbrc.2005.04.173

Herrera, J. E., Sakaguchi, K., Bergel, M., Trieschmann, L., Nakatani, Y., & Bustin, M. (1999). Specific acetylation of chromosomal protein HMG-17 by PCAF alters its interaction with nucleosomes. *Mol Cell Biol, 19*(5), 3466-3473.

Hezroni, H., Tzchori, I., Davidi, A., Mattout, A., Biran, A., Nissim-Rafinia, M., . . . Meshorer, E. (2011). H3K9 histone acetylation predicts pluripotency and reprogramming capacity of ES cells. *Nucleus-Austin, 2*(4), 300-309. doi:10.4161/nucl.2.4.16767

Hill, D. A., Peterson, C. L., & Imbalzano, A. N. (2005). Effects of HMGN1 on chromatin structure and SWI/SNF-mediated chromatin remodeling. *Journal of Biological Chemistry, 280*(50), 41777-41783. doi:10.1074/jbc.M509637200

Hitoshi, S., Alexson, T., Tropepe, V., Donoviel, D., Elia, A. J., Nye, J. S., . . . van der Kooy, D. (2002). Notch pathway molecules are essential for the maintenance, but not the generation, of mammalian neural stem cells. *Genes Dev, 16*(7), 846-858. doi:10.1101/gad.975202

Ho, L., & Crabtree, G. R. (2010). Chromatin remodelling during development. *Nature, 463*(7280), 474-484. doi:10.1038/nature08911

Hock, R., Furusawa, T., Ueda, T., & Bustin, M. (2007). HMG chromosomal proteins in development and disease. *Trends in Cell Biology, 17*(2), 72-79. doi:10.1016/j.tcb.2006.12.001

Hough, S. R., Clements, I., Welch, P. J., & Wiederholt, K. A. (2006). Differentiation of mouse embryonic stem cells after RNA interference-mediated silencing of OCT4 and Nanog. *Stem Cells, 24*(6), 1467-1475. doi:10.1634/stemcells.2005-0475

Hruscha, A., Krawitz, P., Rechenberg, A., Heinrich, V., Hecht, J., Haass, C., & Schmid, B. (2013a). Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development, 140*(24), 4982-4987. doi:10.1242/dev.099085

Hruscha, A., Krawitz, P., Rechenberg, A., Heinrich, V., Hecht, J., Haass, C., & Schmid, B. (2013b). Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development, 140*(24), 4982-4987. doi:10.1242/dev.099085

Hsieh, J., Nakashima, K., Kuwabara, T., Mejia, E., & Gage, F. H. (2004). Histone deacetylase inhibition-mediated neuronal differentiation of multipotent adult neural progenitor cells. *Proc Natl Acad Sci U S A, 101*(47), 16659-16664. doi:10.1073/pnas.0407643101

Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell, 157*(6), 1262-1278. doi:10.1016/j.cell.2014.05.010

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., . . . Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol, 31*(9), 827-832. doi:10.1038/nbt.2647

Hu, D. Q., Garruss, A. S., Gao, X., Morgan, M. A., Cook, M., Smith, E. R., & Shilatifard, A. (2013). The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nature Structural & Molecular Biology, 20*(9), 1093-+. doi:10.1038/nsmb.2653

Hu, X. L., Wang, Y. P., & Shen, Q. (2012). Epigenetic control on cell fate choice in neural stem cells. *Protein & Cell, 3*(4), 278-290. doi:10.1007/s13238-012-2916-6

Huang, G. Y., Ye, S. D., Zhou, X. L., Liu, D. H., & Ying, Q. L. (2015). Molecular basis of embryonic stem cell self-renewal: from signaling pathways to pluripotency network. *Cellular and Molecular Life Sciences, 72*(9), 1741-1757. doi:10.1007/s00018-015-1833-2

Humphrey, R. K., Beattie, G. M., Lopez, A. D., Bucay, N., King, C. C., Firpo, M. T., . . . Hayek, A. (2004). Maintenance of pluripotency in human embryonic stem cells is

STAT3 independent. *Stem Cells, 22*(4), 522-530. doi:DOI 10.1634/stemcells.22-4-522

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Kaini, P., Sander, J. D., . . . Yeh, J. R. (2013). Heritable and precise zebrafish genome editing using a CRISPR-Cas system. *PLoS One, 8*(7), e68708. doi:10.1371/journal.pone.0068708

Hyslop, L., Stojkovic, M., Armstrong, L., Walter, T., Stojkovic, P., Przyborski, S., . . . Lako, M. (2005). Downregulation of NANOG induces differentiation of human embryonic stem cells to extraembryonic lineages. *Stem Cells, 23*(8), 1035-1043. doi:10.1634/stemcells.2005-0080

Isagawa, T., Nagae, G., Shiraki, N., Fujita, T., Sato, N., Ishikawa, S., . . . Aburatani, H. (2011). DNA Methylation Profiling of Embryonic Stem Cell Differentiation into the Three Germ Layers. *PLoS One, 6*(10). doi:ARTN e26052

10.1371/journal.pone.0026052

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol, 169*(12), 5429-5433.

Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., & Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature, 466*(7310), 1129-U1151. doi:10.1038/nature09303

Ito, Y., & Bustin, M. (2002). Immunohistochemical localization of the nucleosome-binding protein HMGN3 in mouse brain. *Journal of Histochemistry & Cytochemistry, 50*(9), 1273-1275.

Itoh, F., Nakane, T., & Chiba, S. (1997). Gene expression of MASH-1, MATH-1, neuroD and NSCL-2, basic helix-loop-helix proteins, during neural differentiation in P19 embryonal carcinoma cells. *Tohoku J Exp Med, 182*(4), 327-336.

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., . . . Lemischka, I. R. (2006). Dissecting self-renewal in stem cells with RNA interference. *Nature, 442*(7102), 533-538. doi:10.1038/nature04915

Jacobs, S., Lie, D. C., DeCicco, K. L., Shi, Y., DeLuca, L. M., Gage, F. H., & Evans, R. M. (2006). Retinoic acid is required early during adult neurogenesis in the dentate gyrus. *Proc Natl Acad Sci U S A, 103*(10), 3902-3907. doi:10.1073/pnas.0511294103

Janesick, A., Wu, S. C., & Blumberg, B. (2015). Retinoic acid signaling and neuronal differentiation. *Cellular and Molecular Life Sciences, 72*(8), 1559-1576. doi:10.1007/s00018-014-1815-9

Jang, E. S., & Goldman, J. E. (2011). Pax6 expression is sufficient to induce a neurogenic fate in glial progenitors of the neonatal subventricular zone. *PLoS One, 6*(6), e20894. doi:10.1371/journal.pone.0020894

Jansen, R., Embden, J. D., Gaastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol, 43*(6), 1565-1575.

Jasin, M., & Rothstein, R. (2013). Repair of strand breaks by homologous recombination. *Cold Spring Harb Perspect Biol, 5*(11), a012740. doi:10.1101/cshperspect.a012740

Jiang, H., Shukla, A., Wang, X., Chen, W. Y., Bernstein, B. E., & Roeder, R. G. (2011). Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. *Cell, 144*(4), 513-525. doi:10.1016/j.cell.2011.01.020

Jiang, J. M., Chan, Y. S., Loh, Y. H., Cai, J., Tong, G. Q., Lim, C. A., . . . Ng, H. H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nature Cell Biology, 10*(3), 353-U103. doi:10.1038/ncb1698

Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol, 31*(3), 233-239. doi:10.1038/nbt.2508

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., & Weeks, D. P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res, 41*(20), e188. doi:10.1093/nar/gkt780

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science, 337*(6096), 816-821. doi:10.1126/science.1225829

Johnson, K. R., Cook, S. A., Bustin, M., & Davisson, M. T. (1992). Genetic-Mapping of the Murine Gene and 14 Related Sequences Encoding Chromosomal Protein Hmg-14. *Mammalian Genome, 3*(11), 625-632. doi:Doi 10.1007/Bf00352479

Johnson, K. R., Cook, S. A., Wardbailey, P., Bustin, M., & Davisson, M. T. (1993). Identification and Genetic-Mapping of the Murine Gene and 20 Related Sequences Encoding Chromosomal Protein Hmg-17. *Mammalian Genome, 4*(2), 83-89. doi:Doi 10.1007/Bf00290431

Jonesvilleneuve, E. M. V., Mcburney, M. W., Rogers, K. A., & Kalnins, V. I. (1982). Retinoic Acid Induces Embryonal Carcinoma-Cells to Differentiate into Neurons and Glial-Cells. *Journal of Cell Biology, 94*(2), 253-262. doi:DOI 10.1083/jcb.94.2.253

Josling, G. A., Selvarajah, S. A., Petter, M., & Duffy, M. F. (2012). The Role of Bromodomain Proteins in Regulating Gene Expression. *Genes, 3*(2), 320-343. doi:10.3390/genes3020320

Kabadi, A. M., Ousterout, D. G., Hilton, I. B., & Gersbach, C. A. (2014). Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Res, 42*(19), e147. doi:10.1093/nar/gku749

Kaiser, T. N., Lojewski, A., Dougherty, C., Juergens, L., Sahar, E., & Latt, S. A. (1982). Flow cytometric characterization of the response of Fanconi's anemia cells to mitomycin C treatment. *Cytometry, 2*(5), 291-297. doi:10.1002/cyto.990020505

Kalcheva, N., Albala, J., O'Guin, K., Rubino, H., Garner, C., & Shafit-Zagardo, B. (1995). Genomic structure of human microtubule-associated protein 2 (MAP-2) and characterization of additional MAP-2 isoforms. *Proc Natl Acad Sci U S A, 92*(24), 10894-10898.

Kalmar, T., Lim, C., Hayward, P., Munoz-Descalzo, S., Nichols, J., Garcia-Ojalvo, J., & Arias, A. M. (2009). Regulated Fluctuations in Nanog Expression Mediate Cell Fate Decisions in Embryonic Stem Cells. *Plos Biology, 7*(7). doi:ARTN e1000149

10.1371/journal.pbio.1000149

Kamiya, D., Banno, S., Sasai, N., Ohgushi, M., Inomata, H., Watanabe, K., . . . Sasai, Y. (2011a). Intrinsic transition of embryonic stem-cell differentiation into neural progenitors. *Nature, 470*(7335), 503-U592. doi:10.1038/nature09726

Kamiya, D., Banno, S., Sasai, N., Ohgushi, M., Inomata, H., Watanabe, K., . . . Sasai, Y. (2011b). Intrinsic transition of embryonic stem-cell differentiation into neural progenitors. *Nature, 470*(7335), 503-509. doi:10.1038/nature09726

Karimova, M., Beschorner, N., Dammermann, W., Chemnitz, J., Indenbirken, D., Bockmann, J. H., . . . Hauber, J. (2015). CRISPR/Cas9 nickase-mediated disruption of hepatitis B virus open reading frame S and X. *Sci Rep, 5*, 13734. doi:10.1038/srep13734

Kato, H., van Ingen, H., Zhou, B. R., Feng, H. Q., Bustin, M., Kay, L. E., & Bai, Y. W. (2011). Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. *Proc Natl Acad Sci U S A, 108*(30), 12283-12288. doi:10.1073/pnas.1105848108

Katsetos, C. D., Herman, M. M., & Mork, S. J. (2003). Class III beta-tubulin in human development and cancer. *Cell Motil Cytoskeleton, 55*(2), 77-96. doi:10.1002/cm.10116

Katsetos, C. D., Legido, A., Perentes, E., & Mork, S. J. (2003). Class III beta-tubulin isotype: a key cytoskeletal protein at the crossroads of developmental

neurobiology and tumor neuropathology. *J Child Neurol, 18*(12), 851-866; discussion 867. doi:10.1177/088307380301801205

Kehler, J., Tolkunova, E., Koschorz, B., Pesce, M., Gentile, L., Boiani, M., . . . Tomilin, A. (2004). Oct4 is required for primordial germ cell survival. *EMBO Rep, 5*(11), 1078-1083. doi:10.1038/sj.embor.7400279

Khavari, D. A., Sen, G. L., & Rinn, J. L. (2010). DNA methylation and epigenetic control of cellular differentiation. *Cell Cycle, 9*(19), 3880-3883. doi:10.4161/cc.9.19.13385

Khokha, M. K., Yeh, J., Grammer, T. C., & Harland, R. M. (2005). Depletion of three BMP antagonists from Spemann's organizer leads to a catastrophic loss of dorsal structures. *Developmental Cell, 8*(3), 401-411. doi:DOI 10.1016/j.devcel.2005.01.013

Kielman, M. F., Rindapaa, M., Gaspar, C., van Poppel, N., Breukel, C., van Leeuwen, S., . . . Fodde, R. (2002). Apc modulates embryonic stem-cell differentiation by controlling the dosage of beta-catenin signaling. *Nature Genetics, 32*(4), 594-605. doi:10.1038/ng1045

Kim, S., Kim, D., Cho, S. W., Kim, J., & Kim, J. S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res, 24*(6), 1012-1019. doi:10.1101/gr.171322.113

Kim, S., Yoon, Y. S., Kim, J. W., Jung, M., Kim, S. U., Lee, Y. D., & Suh-Kim, H. (2004a). Neurogenin1 is sufficient to induce neuronal differentiation of embryonal carcinoma P19 cells in the absence of retinoic acid. *Cellular and Molecular Neurobiology, 24*(3), 343-356.

Kim, S., Yoon, Y. S., Kim, J. W., Jung, M. Y., Kim, S. U., Lee, Y. D., & Suh-Kim, H. (2004b). Neurogenin1 is sufficient to induce neuronal differentiation of embryonal carcinoma P19 cells in the absence of retinoic acid. *Cellular and Molecular Neurobiology, 24*(3), 343-356. doi:Doi 10.1023/B:Cemn.0000022767.74774.38

Kimura, A. P., Liebhaber, S. A., & Cooke, N. E. (2004). Epigenetic modifications at the human growth hormone locus predict distinct roles for histone acetylation and methylation in placental gene activation. *Molecular Endocrinology, 18*(4), 1018-1032. doi:10.1210/me.2003-0468

Kishi, Y., Fujii, Y., Hirabayashi, Y., & Gotoh, Y. (2012). HMGA regulates the global chromatin state and neurogenic potential in neocortical precursor cells. *Nature Neuroscience, 15*(8), 1127-+. doi:10.1038/nn.3165

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., & Joung, J. K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature, 529*(7587), 490-495. doi:10.1038/nature16526

Kolm, P. J., & Sive, H. L. (1995). Regulation of the Xenopus Labial Homeodomain Genes, Hoxa1 and Hoxd1 - Activation by Retinoids and Peptide Growth-Factors. *Developmental Biology, 167*(1), 34-49. doi:DOI 10.1006/dbio.1995.1005

Korner, U., Bustin, M., Scheer, U., & Hock, R. (2003). Developmental role of HMGN proteins in Xenopus laevis. *Mechanisms of Development, 120*(10), 1177-1192. doi:10.1016/j.mod.07.001

Krajewski, W. A., & Becker, P. B. (1998). Reconstitution of hyperacetylated, DNase I-sensitive chromatin characterized by high conformational flexibility of nucleosomal DNA. *Proc Natl Acad Sci U S A, 95*(4), 1540-1545. doi:DOI 10.1073/pnas.95.4.1540

Kraner, S. D., Chong, J. A., Tsay, H. J., & Mandel, G. (1992). Silencing the type II sodium channel gene: a model for neural-specific gene regulation. *Neuron, 9*(1), 37-44.

Kroll, T. T., & O'Leary, D. D. (2005). Ventralized dorsal telencephalic progenitors in Pax6 mutant mice generate GABA interneurons of a lateral ganglionic eminence fate. *Proc Natl Acad Sci U S A, 102*(20), 7374-7379. doi:10.1073/pnas.0500819102

Kruyt, F. A. E., Vanderveer, L. J., Mader, S., Vandenbrink, C. E., Feijen, A., Jonk, L. J. C., . . . Vandersaag, P. T. (1992). Retinoic Acid Resistance of the Variant Embryonal

Carcinoma Cell-Line Rac65 Is Caused by Expression of a Truncated Rar-Alpha. *Differentiation, 49*(1), 27-37. doi:DOI 10.1111/j.1432-0436.1992.tb00766.x

Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., . . . Bernstein, B. E. (2008). Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. *Plos Genetics, 4*(10). doi:ARTN e1000242

10.1371/journal.pgen.1000242

Kufe, D. W., Major, P. P., Egan, E. M., & Beardsley, G. P. (1980). Correlation of Cyto-Toxicity with Incorporation of Ara-C into DNA. *Journal of Biological Chemistry, 255*(19), 8997-9000.

Kugler, J. E., Horsch, M., Huang, D., Furusawa, T., Rochman, M., Garrett, L., . . . Bustin, M. (2013). High Mobility Group N Proteins Modulate the Fidelity of the Cellular Transcriptional Profile in a Tissue- and Variant-specific Manner. *Journal of Biological Chemistry, 288*(23), 16690-16703. doi:10.1074/jbc.M113.463315

Kulkeaw, K., Inoue, T., Mizuochi, C., Horio, Y., Ishihama, Y., & Sugiyama, D. (2012). Ectopic expression of Hmgn2 antagonizes mouse erythroid differentiation in vitro. *Cell Biol Int, 36*(2), 195-202. doi:10.1042/CBI20110169

Kuroda, T., Tada, M., Kubota, H., Kimura, H., Hatano, S., Suemori, H., . . . Tada, T. (2005). Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. *Mol Cell Biol, 25*(6), 2475-2485. doi:10.1128/Mcb.25.6.2475-2485.2005

Kuwabara, T., Hsieh, J., Nakashima, K., Taira, K., & Gage, F. H. (2004). A small modulatory dsRNA specifies the fate of adult neural stem cells. *Cell, 116*(6), 779-793.

Lai, E. C. (2004). Notch signaling: control of cell communication and cell fate. *Development, 131*(5), 965-973. doi:10.1242/dev.01074

Landsman, D., Mcbride, O. W., & Bustin, M. (1989). Human Non-Histone Chromosomal Protein Hmg-17 - Identification, Characterization, Chromosome Localization and Rflps of a Functional Gene from the Large Multigene Family. *Nucleic Acids Res, 17*(6), 2301-2314. doi:DOI 10.1093/nar/17.6.2301

Langston, A. W., Thompson, J. R., & Gudas, L. J. (1997). Retinoic acid-responsive enhancers located 3' of the Hox A and Hox B homeobox gene clusters - Functional analysis. *Journal of Biological Chemistry, 272*(4), 2167-2175.

Lee, J. E. (1997). Basic helix-loop-helix genes in neural development. *Current Opinion in Neurobiology, 7*(1), 13-20. doi:Doi 10.1016/S0959-4388(97)80115-8

Lee, J. E., Hollenberg, S. M., Snider, L., Turner, D. L., Lipnick, N., & Weintraub, H. (1995). Conversion of Xenopus ectoderm into neurons by NeuroD, a basic helix-loop-helix protein. *Science, 268*(5212), 836-844.

Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., . . . Young, R. A. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell, 125*(2), 301-313. doi:10.1016/j.cell.2006.02.043

Leeb, M., Pasini, D., Novatchkova, M., Jaritz, M., Helin, K., & Wutz, A. (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev, 24*(3), 265-276. doi:10.1101/gad.544410

Leeb, M., & Wutz, A. (2007). Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. *Journal of Cell Biology, 178*(2), 219-229. doi:10.1083/jcb.200612127

Lehtonen, S., & Lehtonen, E. (2001). HMG-17 is an early marker of inductive interactions in the developing mouse kidney. *Differentiation, 67*(4-5), 154-163. doi:DOI 10.1046/j.1432-0436.2001.670407.x

Lehtonen, S., Olkkonen, V. M., Stapleton, M., Zerial, M., & Lehtonen, E. (1998). HMG-17, a chromosomal non-histone protein, shows developmental regulation during organogenesis. *International Journal of Developmental Biology, 42*(6), 775-782.

Lessard, J., Wu, J. I., Ranish, J. A., Wan, M., Winslow, M. M., Staahl, B. T., . . . Crabtree, G. R. (2007). An essential switch in subunit composition of a chromatin remodeling

complex during neural development. *Neuron, 55*(2), 201-215. doi:10.1016/j.neuron.2007.06.019

Li, D., Guo, B., Wu, H. J., Tan, L. N., & Lu, Q. J. (2015). TET Family of Dioxygenases: Crucial Roles and Underlying Mechanisms. *Cytogenetic and Genome Research, 146*(3), 171-180. doi:10.1159/000438853

Li, E., Bestor, T. H., & Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell, 69*(6), 915-926.

Li, H. L., Gee, P., Ishida, K., & Hotta, A. (2016). Efficient genomic correction methods in human iPS cells using CRISPR-Cas9 system. *Methods, 101*, 27-35. doi:10.1016/j.ymeth.2015.10.015

Li, W., Teng, F., Li, T., & Zhou, Q. (2013). Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. *Nat Biotechnol, 31*(8), 684-686. doi:10.1038/nbt.2652

Lim, J. H., Catez, F., Birger, Y., West, K. L., Prymakowska-Bosak, M., Postnikov, Y. V., & Bustin, M. (2004). Chromosomal protein HMGN1 modulates histone H3 phosphorylation. *Mol Cell, 15*(4), 573-584. doi:DOI 10.1016/j.molcel.2004.08.006

Lim, J. H., West, K. L., Rubinstein, Y., Bergel, M., Postnikov, Y. V., & Bustin, M. (2005). Chromosomal protein HMGN1 enhances the acetylation of lysine 14 in histone H3. *Embo Journal, 24*(17), 3038-3048. doi:10.1038/sj.emboj.7600768

Linker, C., & Stern, C. D. (2004). Neural induction requires BMP inhibition only as a late step, and involves signals other than FGF and Wnt antagonists. *Development, 131*(22), 5671-5681. doi:10.1242/dev.01445

Loh, Y. H., Wu, Q., Chew, J. L., Vega, V. B., Zhang, W. W., Chen, X., . . . Ng, H. H. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genetics, 38*(4), 431-440. doi:10.1038/ng1760

Lord, C. J., Garrett, M. D., & Ashworth, A. (2006). Targeting the double-strand DNA break repair pathway as a therapeutic strategy. *Clin Cancer Res, 12*(15), 4463-4468. doi:10.1158/1078-0432.CCR-06-1269

Lowell, S., Benchoua, A., Heavey, B., & Smith, A. G. (2006). Notch promotes neural lineage entry by pluripotent embryonic stem cells. *Plos Biology, 4*(5), 805-818. doi:ARTN e121

10.1371/journal.pbio.0040121

Ma, Q. F., Sommer, L., Cserjesi, P., & Anderson, D. J. (1997). Mash1 and neurogenin1 expression patterns define complementary domains of neuroepithelium in the developing CNS and are correlated with regions expressing notch ligands. *Journal of Neuroscience, 17*(10), 3644-3652.

MacDonald, J. L., & Roskams, A. J. (2008). Histone deacetylases 1 and 2 are expressed at distinct stages of neuro-glial development. *Dev Dyn, 237*(8), 2256-2267. doi:10.1002/dvdy.21626

Maden, M., Gale, E., Kostetskii, I., & Zile, M. (1996). Vitamin A-deficient quail embryos have half a hindbrain and other neural defects. *Current Biology, 6*(4), 417-426. doi:Doi 10.1016/S0960-9822(02)00509-2

Maeder, M. L., Thibodeau-Beganny, S., Osiak, A., Wright, D. A., Anthony, R. M., Eichtinger, M., . . . Joung, J. K. (2008). Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell, 31*(2), 294-301. doi:10.1016/j.molcel.2008.06.016

Main, H., Radenkovic, J., Jin, S. B., Lendahl, U., & Andersson, E. R. (2013). Notch Signaling Maintains Neural Rosette Polarity. *PLoS One, 8*(5). doi:ARTN e62959

10.1371/journal.pone.0062959

Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., . . . Church, G. M. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol, 31*(9), 833-838. doi:10.1038/nbt.2675

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., . . . Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science, 339*(6121), 823-826. doi:10.1126/science.1232033

Malik, N., Smulson, M., & Bustin, M. (1984). Enrichment of Acetylated Histones in Polynucleosomes Containing High Mobility Group Protein-17 Revealed by Immunoaffinity Chromatography. *Journal of Biological Chemistry, 259*(2), 699-702.

Marei, H. E. S., Ahmed, A. E., Michetti, F., Pescatori, M., Pallini, R., Casalbore, P., . . . Elhadidy, M. (2012). Gene Expression Profile of Adult Human Olfactory Bulb and Embryonic Neural Stem Cell Suggests Distinct Signaling Pathways and Epigenetic Control. *PLoS One, 7*(4). doi:ARTN e33542

10.1371/journal.pone.0033542

Markowetz, F., Mulder, K. W., Airoldi, E. M., Lemischka, I. R., & Troyanskaya, O. G. (2010). Mapping Dynamic Histone Acetylation Patterns to Gene Expression in Nanog-Depleted Murine Embryonic Stem Cells. *Plos Computational Biology, 6*(12). doi:ARTN e1001034

10.1371/journal.pcbi.1001034

Martello, G., Bertone, P., & Smith, A. (2013). Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. *Embo Journal, 32*(19), 2561-2574. doi:10.1038/emboj.2013.177

Martynoga, B., Drechsel, D., & Guillemot, F. (2012). Molecular Control of Neurogenesis: A View from the Mammalian Cerebral Cortex. *Cold Spring Harb Perspect Biol, 4*(10). doi:ARTN a008359

10.1101/cshperspect.a008359

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., . . . Niwa, H. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature Cell Biology, 9*(6), 625-U626. doi:10.1038/ncb1589

Matsuda, T., Nakamura, T., Nakao, K., Arai, T., Katsuki, M., Heike, T., & Yokota, T. (1999). STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *Embo Journal, 18*(15), 4261-4269. doi:DOI 10.1093/emboj/18.15.4261

Mattout, A., & Meshorer, E. (2010). Chromatin plasticity and genome organization in pluripotent embryonic stem cells. *Current Opinion in Cell Biology, 22*(3), 334-341. doi:10.1016/j.ceb.2010.02.001

Matus, A. (1990). Microtubule-associated proteins and the determination of neuronal form. *J Physiol (Paris), 84*(1), 134-137.

Maye, P., Becker, S., Siemen, H., Thorne, J., Byrd, N., Carpentino, J., & Grabel, L. (2004). Hedgehog signaling is required for the differentiation of ES cells into neurectoderm. *Developmental Biology, 265*(1), 276-290. doi:10.1016/j.ydbio.2003.09.027

Mcburney, M. W. (1993). P19 Embryonal Carcinoma-Cells. *International Journal of Developmental Biology, 37*(1), 135-140.

Medler, T. R., Craig, J. M., Fiorillo, A. A., Feeney, Y. B., Harrell, J. C., & Clevenger, C. V. (2016). HDAC6 Deacetylates HMGN2 to Regulate Stat5a Activity and Breast Cancer Growth. *Molecular Cancer Research, 14*(10), 994-1008. doi:10.1158/1541-7786.Mcr-16-0109

Menezes, J. R., & Luskin, M. B. (1994). Expression of neuron-specific tubulin defines a novel population in the proliferative layers of the developing telencephalon. *J Neurosci, 14*(9), 5399-5416.

Miller, J., McLachlan, A. D., & Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J, 4*(6), 1609-1614.

Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., . . . Rebar, E. J. (2011). A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol, 29*(2), 143-148. doi:10.1038/nbt.1755

Min, T. H., Kriebel, M., Hou, S. R., & Pera, E. M. (2011). The dual regulator Sufu integrates Hedgehog and Wnt signals in the early Xenopus embryo. *Developmental Biology, 358*(1), 262-276. doi:10.1016/j.ydbio.2011.07.035

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., . . . Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell, 113*(5), 631-642. doi:Doi 10.1016/S0092-8674(03)00393-3

Miyata, T., Kawaguchi, A., Saito, K., Kawano, M., Muto, T., & Ogawa, M. (2004). Asymmetric production of surface-dividing and non-surface-dividing cortical progenitor cells. *Development, 131*(13), 3133-3145. doi:10.1242/dev.01173

Mohan, G. (2012). Chromatin-binding HMGN proteins and the neuronal differentiation of embryonal carcinoma cells in vitro. PhD thesis, University of Glasgow, Institute of Cancer Sciences, Glasgow

Mohamed, O. A., Bustin, M., & Clarke, H. J. (2001). High-mobility group proteins 14 and 17 maintain the timing of early embryonic development in the mouse. *Developmental Biology, 229*(1), 237-249. doi:10.1006/dbio.2000.9942

Mori, N., Schoenherr, C., Vandenbergh, D. J., & Anderson, D. J. (1992). A common silencer element in the SCG10 and type II Na+ channel genes binds a factor present in nonneuronal cells but not in neuronal cells. *Neuron, 9*(1), 45-54.

Morton, R. L., David, H., OConnor, P. M., & Bustin, M. (1996). Chromosomal proteins HMG-14 and HMG-17 are synthesized throughout the S-Phase in Burkitt's lymphoma. *Biochemical and Biophysical Research Communications, 222*(2), 368-373. doi:DOI 10.1006/bbrc.1996.0798

Mummery, C. L., Vandenbrink, C. E., & Delaat, S. W. (1987). Commitment to Differentiation Induced by Retinoic Acid in P19 Embryonal Carcinoma-Cells Is Cell-Cycle Dependent. *Developmental Biology, 121*(1), 10-19. doi:Doi 10.1016/0012-1606(87)90133-3

Munoz-Sanjuan, I., & Brivanlou, A. H. (2002). Neural induction, the default model and embryonic stem cells. *Nat Rev Neurosci, 3*(4), 271-280. doi:10.1038/nrn786

Nacher, J., Varea, E., Blasco-Ibanez, J. M., Castillo-Gomez, E., Crespo, C., Martinez-Guijarro, F. J., & McEwen, B. S. (2005). Expression of the transcription factor Pax 6 in the adult rat dentate gyrus. *Journal of Neuroscience Research, 81*(6), 753-761. doi:10.1002/jnr.20596

Nagao, M., Lanjakornsiripan, D., Itoh, Y., Kishi, Y., Ogata, T., & Gotoh, Y. (2014). High Mobility Group Nucleosome-Binding Family Proteins Promote Astrocyte Differentiation of Neural Precursor Cells. *Stem Cells, 32*(11), 2983-2997. doi:10.1002/stem.1787

Nakatake, Y., Fukui, N., Iwamatsu, Y., Masui, S., Takahashi, K., Yagi, R., . . . Niwa, H. (2006). Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells. *Mol Cell Biol, 26*(20), 7772-7782. doi:10.1128/Mcb.00468-06

Nakayama, Y., Wada, A., Inoue, R., Terasawa, K., Kimura, I., Nakamura, N., & Kurosaka, A. (2014). A rapid and efficient method for neuronal induction of the P19 embryonic carcinoma cell line. *Journal of Neuroscience Methods, 227*, 100-106. doi:10.1016/j.jneumeth.2014.02.011

Nemeth, A., & Langst, G. (2004). Chromatin higher order structure: opening up chromatin for transcription. *Brief Funct Genomic Proteomic, 2*(4), 334-343.

Newcomb, R., Sun, X. Y., Taylor, L., Curthoys, N., & Giffard, R. G. (1997). Increased production of extracellular glutamate by the mitochondrial glutaminase following neuronal death. *Journal of Biological Chemistry, 272*(17), 11276-11282.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., . . . Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell, 95*(3), 379-391. doi:Doi 10.1016/S0092-8674(00)81769-9

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., . . . Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell, 156*(5), 935-949. doi:10.1016/j.cell.2014.02.001

Niwa, H., Burdon, T., Chambers, I., & Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev, 12*(13), 2048-2060. doi:DOI 10.1101/gad.12.13.2048

Niwa, H., Miyazaki, J., & Smith, A. G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genetics, 24*(4), 372-376. doi:Doi 10.1038/74199

Nixon, R. A., & Shea, T. B. (1992). Dynamics of neuronal intermediate filaments: a developmental perspective. *Cell Motil Cytoskeleton, 22*(2), 81-91. doi:10.1002/cm.970220202

Noctor, S. C., Flint, A. C., Weissman, T. A., Dammerman, R. S., & Kriegstein, A. R. (2001). Neurons derived from radial glial cells establish radial units in neocortex. *Nature, 409*(6821), 714-720. doi:Doi 10.1038/35055553

Noctor, S. C., V, M.-C., Ivic, L., & Kriegstein, A. R. (2004). Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. *Nature Neuroscience, 7*(2), 136-144. doi:10.1038/nn1172

Nunez, R. (2001). DNA measurement and cell cycle analysis by flow cytometry. *Curr Issues Mol Biol, 3*(3), 67-70.

Obier, N., Lin, Q., Cauchy, P., Hornich, V., Zenke, M., Becker, M., & Muller, A. M. (2015). Polycomb Protein EED is Required for Silencing of Pluripotency Genes upon ESC Differentiation. *Stem Cell Reviews and Reports, 11*(1), 50-61. doi:10.1007/s12015-014-9550-z

Ogawa, K., Matsui, H., Ohtsuka, S., & Niwa, H. (2004). A novel mechanism for regulating clonal propagation of mouse ES cells. *Genes to Cells, 9*(5), 471-477. doi:10.1111/j.1365-9597.2004.00736.x

Ohse, M., Takahashi, K., Kadowaki, Y., & Kusaoke, H. (1995). Effects of plasmid DNA sizes and several other factors on transformation of Bacillus subtilis ISW1214 with plasmid DNA by electroporation. *Biosci Biotechnol Biochem, 59*(8), 1433-1437. doi:10.1271/bbb.59.1433

Ohtsuka, T., & Kageyama, R. (2010). The Basic Helix-Loop-Helix Transcription Factors in Neural Differentiation. *Cell Cycle Regulation and Differentiation in Cardiovascular and Neural Systems*, 15-34. doi:10.1007/978-1-60327-153-0_2

Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell, 99*(3), 247-257.

Olynik, B. M., & Rastegar, M. (2012). The genetic and epigenetic journey of embryonic stem cells into mature neural cells. *Front Genet, 3*, 81. doi:10.3389/fgene.2012.00081

Onishi, K., & Zandstra, P. W. (2015). LIF signaling in stem cells and development. *Development, 142*(13), 2230-2236. doi:10.1242/dev.117598

Ooi, L., & Wood, I. C. (2007). Chromatin crosstalk in development and disease: lessons from REST. *Nature Reviews Genetics, 8*(7), 544-554. doi:10.1038/nrg2100

Orr-Weaver, T. L., & Szostak, J. W. (1983). Yeast recombination: the association between double-strand gap repair and crossing-over. *Proc Natl Acad Sci U S A, 80*(14), 4417-4421.

Osumi, N., Shinohara, H., Numayama-Tsuruta, K., & Maekawa, M. (2008). Concise review: Pax6 transcription factor contributes to both embryonic and adult neurogenesis as a multifunctional regulator. *Stem Cells, 26*(7), 1663-1672. doi:10.1634/stemcells.2007-0884

Ousterout, D. G., Kabadi, A. M., Thakore, P. I., Perez-Pinera, P., Brown, M. T., Majoros, W. H., . . . Gersbach, C. A. (2015). Correction of Dystrophin Expression in Cells From Duchenne Muscular Dystrophy Patients Through Genomic Excision of Exon

51 by Zinc Finger Nucleases. *Molecular Therapy, 23*(3), 523-532. doi:10.1038/mt.2074.234

Ozair, M. Z., Kintner, C., & Brivanlou, A. H. (2013). Neural induction and early patterning in vertebrates. *Wiley Interdisciplinary Reviews-Developmental Biology, 2*(4), 479-498. doi:10.1002/wdev.90

Pabo, C. O., Peisach, E., & Grant, R. A. (2001). Design and selection of novel Cys2His2 zinc finger proteins. *Annu Rev Biochem, 70*, 313-340. doi:10.1146/annurev.biochem.70.1.313

Pachernik, J., Bryja, V., Esner, M., Kubala, L., Dvorak, P., & Hampl, A. (2005). Neural differentiation of pluripotent mouse embryonal carcinoma cells by retinoic acid: Inhibitory effect of serum. *Physiological Research, 54*(1), 115-122.

Papaioannou, V. E., Mcburney, M. W., Gardner, R. L., & Evans, M. J. (1975). Fate of Teratocarcinoma Cells Injected into Early Mouse Embryos. *Nature, 258*(5530), 70-73. doi:Doi 10.1038/258070a0

Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., . . . Tessier-Lavigne, M. (2016). Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature, 533*(7601), 125-129. doi:10.1038/nature17664

Paquin, J., Danalache, B. A., Jankowski, M., McCann, S. M., & Gutkowska, J. (2002). Oxytocin induces differentiation of P19 embryonic stem cells to cardiomyocytes. *Proc Natl Acad Sci U S A, 99*(14), 9550-9555. doi:10.1073/pnas.152302499

Paranjape, S. M., Krumm, A., & Kadonaga, J. T. (1995). Hmg17 Is a Chromatin-Specific Transcriptional Coactivator That Increases the Efficiency of Transcription Initiation. *Genes Dev, 9*(16), 1978-1991. doi:DOI 10.1101/gad.9.16.1978

Park, D., Xiang, A. P., Mao, F. F., Zhang, L., Di, C. G., Liu, X. M., . . . Lahn, B. T. (2010). Nestin Is Required for the Proper Self-Renewal of Neural Stem Cells. *Stem Cells, 28*(12), 2162-2171. doi:10.1002/stem.541

Pash, J. M., Alfonso, P. J., & Bustin, M. (1993). Aberrant Expression of High-Mobility Group Chromosomal Protein-14 Affects Cellular-Differentiation. *Journal of Biological Chemistry, 268*(18), 13632-13638.

Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M., & Helin, K. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol Cell Biol, 27*(10), 3769-3779. doi:10.1128/MCB.01432-06

Pasini, D., Bracken, A. P., Jensen, M. R., Lazzerini Denchi, E., & Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J, 23*(20), 4061-4071. doi:10.1038/sj.emboj.7600402

Pera, E. M., Ikeda, A., Eivers, E., & De Robertis, E. M. (2003). Integration of IGF, FGF, and anti-BMP signals via Smad1 phosphorylation in neural induction. *Genes Dev, 17*(24), 3023-3028. doi:10.1101/gad.1153603

Pere, E. M., Acosta, H., Gouignard, N., Climent, M., & Arregi, I. (2014). Active signals, gradient formation and regional specificity in neural induction. *Experimental Cell Research, 321*(1), 25-31. doi:10.1016/j.yexcr.2013.11.018

Perez-Pinera, P., Ousterout, D. G., & Gersbach, C. A. (2012). Advances in targeted genome editing. *Curr Opin Chem Biol, 16*(3-4), 268-277. doi:10.1016/j.cbpa.2012.06.007

Petek, L. M., Russell, D. W., & Miller, D. G. (2010). Frequent endonuclease cleavage at off-target locations in vivo. *Mol Ther, 18*(5), 983-986. doi:10.1038/mt.2010.35

Petersen, P. H., Zou, K. Y., Krauss, S., & Zhong, W. M. (2004). Continuing role for mouse Numb and Numbl in maintaining progenitor cells during cortical neurogenesis. *Nature Neuroscience, 7*(8), 803-811. doi:10.1038/nn1289

Popescu, N., Landsman, D., & Bustin, M. (1990). Mapping the Human Gene Coding for Chromosomal Protein Hmg-17. *Human Genetics, 85*(3), 376-378.

Postel-Vinay, S., Vanhecke, E., Olaussen, K. A., Lord, C. J., Ashworth, A., & Soria, J. C. (2012). The potential of exploiting DNA-repair defects for optimizing lung cancer treatment. *Nat Rev Clin Oncol, 9*(3), 144-155. doi:10.1038/nrclinonc.2012.3

Postnikov, Y. V., & Bustin, M. (2016). Functional interplay between histone H1 and HMG proteins in chromatin. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms, 1859*(3), 462-467. doi:10.1016/j.bbagrm.2015.10.006

Postnikov, Y. V., Lehn, D. A., Robinson, R. C., Friedman, F. K., Shiloach, J., & Bustin, M. (1994). The Cooperative Binding of Chromosomal Protein Hmg-14 to Nucleosome Cores Is Reduced by Single-Point Mutations in the Nucleosomal Binding Domain. *Nucleic Acids Res, 22*(21), 4520-4526. doi:DOI 10.1093/nar/22.21.4520

Postnikov, Y. V., Shick, V. V., Belyavsky, A. V., Khrapko, K. R., Brodolin, K. L., Nikolskaya, T. A., & Mirzabekov, A. D. (1991). Distribution of High Mobility Group Proteins 1/2,E and 14/17 and Linker Histone-H1 and Histone-H5 on Transcribed and Nontranscribed Regions of Chicken Erythrocyte Chromatin. *Nucleic Acids Res, 19*(4), 717-725. doi:Doi 10.1093/Nar/19.4.717

Postnikov, Y. V., Trieschmann, L., Rickers, A., & Bustin, M. (1995). Homodimers of chromosomal proteins HMG-14 and HMG-17 in nucleosome cores. *J Mol Biol, 252*(4), 423-432. doi:10.1006/jmbi.1995.0508

Potter, H. (2003a). Transfection by electroporation. *Curr Protoc Mol Biol, Chapter 9*, Unit 9 3. doi:10.1002/0471142727.mb0903s62

Potter, H. (2003b). Transfection by electroporation. *Curr Protoc Cell Biol, Chapter 20*, Unit 20 25. doi:10.1002/0471143030.cb2005s19

Pratt, M. A. C., Kralova, J., & Mcburney, M. W. (1990). A Dominant Negative Mutation of the Alpha-Retinoic Acid Receptor Gene in a Retinoic Acid-Nonresponsive Embryonal Carcinoma Cell. *Mol Cell Biol, 10*(12), 6445-6453.

Prymakowska-Bosak, M., Hock, R., Catez, F., Lim, J. H., Birger, Y., Shirakawa, H., . . . Bustin, M. (2002). Mitotic phosphorylation of chromosomal protein HMGN1 inhibits nuclear import and promotes interaction with 14.3.3 proteins. *Mol Cell Biol, 22*(19), 6809-6819. doi:10.1128/Mcb.22.19.6809-6819.2002

Prymakowska-Bosak, M., Misteli, T., Herrera, J. E., Shirakawa, H., Birger, Y., Garfield, S., & Bustin, M. (2001). Mitotic phosphorylation prevents the binding of HMGN proteins to chromatin. *Mol Cell Biol, 21*(15), 5169-5178. doi:Doi 10.1128/Mcb.21.15.5169-5178.2001

Quinn, J. C., Molinek, M., Martynoga, B. S., Zaki, P. A., Faedo, A., Bulfone, A., . . . 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), 50-65. doi:10.1016/j.ydbio.2006.08.035

Ramirez, J. M., Gerbal-Chaloin, S., Milhavet, O., Qiang, B., Becker, F., Assou, S., . . . De Vos, J. (2011). Brief Report: Benchmarking Human Pluripotent Stem Cell Markers During Differentiation Into the Three Germ Layers Unveils a Striking Heterogeneity: All Markers Are Not Equal. *Stem Cells, 29*(9), 1469-1474. doi:10.1002/stem.681

Ran, F. A., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., . . . Zhang, F. (2015). In vivo genome editing using Staphylococcus aureus Cas9. *Nature, 520*(7546), 186-191. doi:10.1038/nature14299

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., . . . Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell, 154*(6), 1380-1389. doi:10.1016/j.cell.2013.08.021

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat Protoc, 8*(11), 2281-2308. doi:10.1038/nprot.2013.143

Rattner, B. P., Yusufzai, T., & Kadonaga, J. T. (2009). HMGN Proteins Act in Opposition to ATP-Dependent Chromatin Remodeling Factors to Restrict Nucleosome Mobility. *Mol Cell, 34*(5), 620-626. doi:10.1016/j.molcel.2009.04.014

Reeves, R. (2010). Nuclear functions of the HMG proteins. *Biochim Biophys Acta, 1799*(1-2), 3-14. doi:10.1016/j.bbagrm.2009.09.001

Rehbini, O. (2016). The role of high mobility nucleosomal binding protein (Hmgn2) in the undifferentiated mouse epiblast carcinoma stem cells. PhD thesis, University of Glasgow, Institute of Cancer Scinces, Glasgow

Reversade, B., & De Robertis, E. M. (2005). Regulation of ADMP and BMP2/4/7 at opposite embryonic poles generates a self-regulating morphogenetic field. *Cell, 123*(6), 1147-1160. doi:10.1016/j.cell.2005.08.037

Reversade, B., Kuroda, H., Lee, H., Mays, A., & De Robertis, E. M. (2005). Depletion of Bmp2, Bmp4, Bmp7 and Spemann organizer signals induces massive brain formation in Xenopus embryos. *Development, 132*(15), 3381-3392. doi:10.1242/dev.01901

Reyon, D., Tsai, S. Q., Khayter, C., Foden, J. A., Sander, J. D., & Joung, J. K. (2012). FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol, 30*(5), 460-465. doi:10.1038/nbt.2170

Rhinn, M., & Dolle, P. (2012). Retinoic acid signalling during development. *Development, 139*(5), 843-858. doi:10.1242/dev.065938

Rochman, M., Postnikov, Y., Correll, S., Malicet, C., Wincovitch, S., Karpova, T. S., . . . Bustin, M. (2009). The Interaction of NSBP1/HMGN5 with Nucleosomes in Euchromatin Counteracts Linker Histone-Mediated Chromatin Compaction and Modulates Transcription. *Mol Cell, 35*(5), 642-656. doi:10.1016/j.molcel.2009.07.002

Rochman, M., Taher, L., Kurahashi, T., Cherukuri, S., Uversky, V. N., Landsman, D., . . . Bustin, M. (2011). Effects of HMGN variants on the cellular transcription profile. *Nucleic Acids Res, 39*(10), 4076-4087. doi:10.1093/nar/gkq1343

Rodda, D. J., Chew, J. L., Lim, L. H., Loh, Y. H., Wang, B., Ng, H. H., & Robson, P. (2005). Transcriptional regulation of Nanog by Oct4 and Sox2. *Journal of Biological Chemistry, 280*(26), 24731-24737. doi:10.1074/jbc.M502573200

Rohwedel, J., Guan, K., Hegert, C., & Wobus, A. M. (2001). Embryonic stem cells as an in vitro model for mutagenicity, cytotoxicity and embryotoxicity studies: present state and future prospects. *Toxicology in Vitro, 15*(6), 741-753. doi:Doi 10.1016/S0887-2333(01)00074-1

Rosenbloom, K. R., Dreszer, T. R., Long, J. C., Malladi, V. S., Sloan, C. A., Raney, B. J., . . . Kent, W. J. (2012). ENCODE whole-genome data in the UCSC Genome Browser: update 2012. *Nucleic Acids Res, 40*(Database issue), D912-917. doi:10.1093/nar/gkr1012

Ross, S. E., Greenberg, M. E., & Stiles, C. D. (2003). Basic helix-loop-helix factors in cortical development. *Neuron, 39*(1), 13-25. doi:Doi 10.1016/S0896-6273(03)00365-9

Rudin, N., Sugarman, E., & Haber, J. E. (1989). Genetic and physical analysis of double-strand break repair and recombination in Saccharomyces cerevisiae. *Genetics, 122*(3), 519-534.

Sahara, S., & O'Leary, D. D. M. (2009). Fgf10 Regulates Transition Period of Cortical Stem Cell Differentiation to Radial Glia Controlling Generation of Neurons and Basal Progenitors. *Neuron, 63*(1), 48-62. doi:10.1016/j.neuron.2009.06.006

Sakuma, T., Nishikawa, A., Kume, S., Chayama, K., & Yamamoto, T. (2014). Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci Rep, 4*, 5400. doi:10.1038/srep05400

San Filippo, J., Sung, P., & Klein, H. (2008). Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem, 77*, 229-257. doi:10.1146/annurev.biochem.77.061306.125255

Sander, J. D., & Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol, 32*(4), 347-355. doi:10.1038/nbt.2842

Saraiva, N. Z., Oliveira, C. S., & Garcia, J. M. (2010). Histone acetylation and its role in embryonic stem cell differentiation. *World J Stem Cells, 2*(6), 121-126. doi:10.4252/wjsc.v2.i6.121

Saretzki, G., Armstrong, L., Leake, A., Lako, M., & von Zglinicki, T. (2004). Stress defense in murine embryonic stem cells is superior to that of various differentiated murine cells. *Stem Cells, 22*(6), 962-971. doi:DOI 10.1634/stemcells.22-6-962

Sasai, Y., Lu, B., Steinbeisser, H., Geissert, D., Gont, L. K., & Derobertis, E. M. (1994). Xenopus Chordin - a Novel Dorsalizing Factor-Activated by Organizer-Specific Homeobox Genes. *Cell, 79*(5), 779-790. doi:Doi 10.1016/0092-8674(94)90068-X

Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., & Brivanlou, A. H. (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nature Medicine, 10*(1), 55-63. doi:10.1038/nm979

Schaniel, C., Ang, Y. S., Ratnakumar, K., Cormier, C., James, T., Bernstein, E., . . . Paddison, P. J. (2009). Smarcc1/Baf155 Couples Self-Renewal Gene Repression with Changes in Chromatin Structure in Mouse Embryonic Stem Cells. *Stem Cells, 27*(12), 2979-2991. doi:10.1002/stem.223

Schmitges, F. W., Prusty, A. B., Faty, M., Stutzer, A., Lingaraju, G. M., Aiwazian, J., . . . Thoma, N. H. (2011). Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Mol Cell, 42*(3), 330-341. doi:10.1016/j.molcel.2011.03.025

Schneider, R. A., Hu, D., Rubenstein, J. L. R., Maden, M., & Helms, J. A. (2001). Local retinoid signaling coordinates forebrain and facial morphogenesis by maintaining FGF8 and SHH. *Development, 128*(14), 2755-2767.

Schwarz, B. A., Bar-Nur, O., Silva, J. C. R., & Hochedlinger, K. (2014). Nanog Is Dispensable for the Generation of Induced Pluripotent Stem Cells. *Current Biology, 24*(3), 347-350. doi:10.1016/j.cub.2013.12.050

Seo, S., Lim, J. W., Yellajoshyula, D., Chang, L. W., & Kroll, K. L. (2007). Neurogenin and NeuroD direct transcriptional targets and their regulatory enhancers. *Embo Journal, 26*(24), 5093-5108. doi:10.1038/sj.emboj.7601923

Shakoori, A. R., Owen, T. A., Shalhoub, V., Stein, J. L., Bustin, M., Stein, G. S., & Lian, J. B. (1993). Differential Expression of the Chromosomal High Mobility Group Protein-14 and Protein-17 during the Onset of Differentiation in Mammalian Osteoblasts and Promyelocytic Leukemia-Cells. *Journal of Cellular Biochemistry, 51*(4), 479-487.

Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., . . . Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science, 343*(6166), 84-87. doi:10.1126/science.1247005

Sharpe, C. R., & Goldstone, K. (1997). Retinoid receptors promote primary neurogenesis in Xenopus. *Development, 124*(2), 515-523.

Shen, B., Zhang, J., Wu, H., Wang, J., Ma, K., Li, Z., . . . Huang, X. (2013). Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Res, 23*(5), 720-723. doi:10.1038/cr.2013.46

Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., . . . Skarnes, W. C. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods, 11*(4), 399-402. doi:10.1038/nmeth.2857

Shen, S., Pu, J., Lang, B., & McCaig, C. D. (2011). A zinc finger protein Zfp521 directs neural differentiation and beyond. *Stem Cell Res Ther, 2*(2), 20. doi:10.1186/scrt61

Shen, X., Liu, Y., Hsu, Y. J., Fujiwara, Y., Kim, J., Mao, X., . . . Orkin, S. H. (2008). EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol Cell, 32*(4), 491-502. doi:10.1016/j.molcel.2008.10.016

Shimahara, H., Hirano, T., Ohya, K., Matsuta, S., Seeram, S. S., & Tate, S. (2013). Nucleosome structural changes induced by binding of non-histone chromosomal proteins HMGN1 and HMGN2. *Febs Open Bio, 3*, 184-191. doi:10.1016/j.fob.2013.03.002

Shirakawa, H., Rochman, M., Furusawa, T., Kuehn, M. R., Horigome, S., Haketa, K., . . . Bustin, M. (2009). The Nucleosomal Binding Protein NSBP1 Is Highly Expressed in the Placenta and Modulates the Expression of Differentiation Markers in

Placental Rcho-1 Cells. *Journal of Cellular Biochemistry, 106*(4), 651-658. doi:10.1002/jcb.22046

Siegenthaler, J. A., Ashique, A. M., Zarbalis, K., Patterson, K. P., Hecht, J. H., Kane, M. A., . . . Pleasure, S. J. (2009). Retinoic Acid from the Meninges Regulates Cortical Neuron Generation. *Cell, 139*(3), 597-609. doi:10.1016/j.cell.2009.10.004

Sikorska, M., Sandhu, J. K., Deb-Rinker, P., Jezierski, A., LeBlanc, J., Charlebois, C., . . . Walker, P. R. (2008). Epigenetic modifications of SOX2 enhancers, SRR1 and SRR2, correlate with in vitro neural differentiation. *Journal of Neuroscience Research, 86*(8), 1680-1693. doi:10.1002/jnr.21635

Silva, G., Poirot, L., Galetto, R., Smith, J., Montoya, G., Duchateau, P., & Paques, F. (2011). Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr Gene Ther, 11*(1), 11-27.

Singh, A. M., Hamazaki, T., Hankowski, K. E., & Terada, N. (2007). A heterogeneous expression pattern for nanog in embryonic stem cells. *Stem Cells, 25*(10), 2534-2542. doi:10.1634/stemcells.2007-0126

Singh, R. P., Shiue, K., Schomberg, D., & Zhou, F. C. (2009). Cellular epigenetic modifications of neural stem cell differentiation. *Cell Transplant, 18*(10), 1197-1211. doi:10.3727/096368909X12483162197204

Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., & Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science, 351*(6268), 84-88. doi:10.1126/science.aad5227

Smith, A. G. (2001). Embryo-derived stem cells: Of mice and men. *Annual Review of Cell and Developmental Biology, 17*, 435-462. doi:DOI 10.1146/annurev.cellbio.17.1.435

Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M., & Rogers, D. (1988). Inhibition of Pluripotential Embryonic Stem-Cell Differentiation by Purified Polypeptides. *Nature, 336*(6200), 688-690. doi:Doi 10.1038/336688a0

Smith, J., Grizot, S., Arnould, S., Duclert, A., Epinat, J. C., Chames, P., . . . Duchateau, P. (2006). A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res, 34*(22), e149. doi:10.1093/nar/gkl720

Smith, W. C., & Harland, R. M. (1992). Expression Cloning of Noggin, a New Dorsalizing Factor Localized to the Spemann Organizer in Xenopus Embryos. *Cell, 70*(5), 829-840. doi:Doi 10.1016/0092-8674(92)90316-5

Solter, D. (2006). Timeline - From teratocarcinomas to embryonic stem cells and beyond: a history of embryonic stem cell research. *Nature Reviews Genetics, 7*(4), 319-327. doi:10.1038/nrg1827

Song, M. R., & Ghosh, A. (2004). FGF2-induced chromatin remodeling regulates CNTF-mediated gene expression and astrocyte differentiation. *Nature Neuroscience, 7*(3), 229-235. doi:10.1038/nn1192

Spangrude, G. J., Heimfeld, S., & Weissman, I. L. (1988). Purification and Characterization of Mouse Hematopoietic Stem-Cells. *Science, 241*(4861), 58-62. doi:DOI 10.1126/science.2898810

Stella, S., & Montoya, G. (2016). The genome editing revolution: A CRISPR-Cas TALE off-target story. *Bioessays, 38 Suppl 1*, S4-S13. doi:10.1002/bies.201670903

Stern, C. D. (2006). Neural induction: 10 years on since the 'default model'. *Current Opinion in Cell Biology, 18*(6), 692-697. doi:10.1016/j.ceb.2006.09.002

Stoykova, A., Fritsch, R., Walther, C., & Gruss, P. (1996). Forebrain patterning defects in Small eye mutant mice. *Development, 122*(11), 3453-3465.

Strichman-Almashanu, L. Z., Bustin, M., & Landsman, D. (2003). Retroposed copies of the HMG genes: A window to genome dynamics. *Genome Res, 13*(5), 800-812. doi:10.1011/gr.893803

Sun, Y., Nadal-Vicens, M., Misono, S., Lin, M. Z., Zubiaga, A., Hua, X. X., . . . Greenberg, M. E. (2001). Neurogenin promotes neurogenesis and inhibits glial differentiation

by independent mechanisms. *Cell, 104*(3), 365-376. doi:Doi 10.1016/S0092-8674(01)00224-0

Tahiliani, M., Koh, K. P., Shen, Y. H., Pastor, W. A., Bandukwala, H., Brudno, Y., . . . Rao, A. (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science, 324*(5929), 930-935. doi:10.1126/science.1170116

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell, 126*(4), 663-676. doi:DOI 10.1016/j.cell.2006.07.024

Taylor, D. W., Zhu, Y., Staals, R. H., Kornfeld, J. E., Shinkai, A., van der Oost, J., . . . Doudna, J. A. (2015). Structural biology. Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. *Science, 348*(6234), 581-585. doi:10.1126/science.aaa4535

Temple, S. (2001). The development of neural stem cells. *Nature, 414*(6859), 112-117. doi:Doi 10.1038/35102174

ten Berge, D., Kurek, D., Blauwkamp, T., Koole, W., Maas, A., Eroglu, E., . . . Nusse, R. (2011). Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nature Cell Biology, 13*(9), 1070-1075. doi:http://www.nature.com/ncb/journal/v13/n9/abs/ncb2314.html#supplementary-information

Tesar, P. J., Chenoweth, J. G., Brook, F. A., Davies, T. J., Evans, E. P., Mack, D. L., . . . McKay, R. D. G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature, 448*(7150), 196-U110. doi:10.1038/nature05972

Theus, M. H., Wei, L., Francis, K., & Yu, S. P. (2006). Critical roles of Src family tyrosine kinases in excitatory neuronal differentiation of cultured embryonic stem cells. *Experimental Cell Research, 312*(16), 3096-3107. doi:10.1016/j.yexcr.2006.06.022

Tie, F., Banerjee, R., Stratton, C. A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., . . . Harte, P. J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development, 136*(18), 3131-3141. doi:10.1242/dev.037127

Tomioka, M., Nishimoto, M., Miyagi, S., Katayanagi, T., Fukui, N., Niwa, H., . . . Okuda, A. (2002). Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. *Nucleic Acids Res, 30*(14), 3202-3213. doi:DOI 10.1093/nar/gkf435

Trieschmann, L., Alfonso, P. J., Crippa, M. P., Wolffe, A. P., & Bustin, M. (1995). Incorporation of Chromosomal-Proteins Hmg-14/Hmg-17 into Nascent Nucleosomes Induces an Extended Chromatin Conformation and Enhances the Utilization of Active Transcription Complexes. *Embo Journal, 14*(7), 1478-1489.

Trieschmann, L., Martin, B., & Bustin, M. (1998). The chromatin unfolding domain of chromosomal protein HMG-14 targets the N-terminal tail of histone H3 in nucleosomes (vol 95, pg 5468, 1998). *Proc Natl Acad Sci U S A, 95*(15), 9059-9059.

Truong, L. N., Li, Y., Sun, E., Ang, K., Hwang, P. Y., & Wu, X. (2014). Homologous recombination is a primary pathway to repair DNA double-strand breaks generated during DNA rereplication. *J Biol Chem, 289*(42), 28910-28923. doi:10.1074/jbc.M114.576488

Tse, C., Sera, T., Wolffe, A. P., & Hansen, J. C. (1998). Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. *Mol Cell Biol, 18*(8), 4629-4638.

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., . . . Okano, M. (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes to Cells, 11*(7), 805-814. doi:10.1111/j.1365-2443.2006.00984.x

Ueda, T., Catez, F., Gerlitz, G., & Bustin, M. (2008). Delineation of the protein module that anchors HMGN proteins to nucleosomes in the chromatin of living cells. *Mol Cell Biol, 28*(9), 2872-2883. doi:10.1128/MCB.02181-07

Ueda, T., Furusawa, T., Kurahashi, T., Tessarollo, L., & Bustin, M. (2009). The Nucleosome Binding Protein HMGN3 Modulates the Transcription Profile of Pancreatic beta Cells and Affects Insulin Secretion. *Mol Cell Biol, 29*(19), 5264-5276. doi:10.1128/Mcb.00526-09

Ueda, T., Postnikov, Y. V., & Bustin, M. (2006). Distinct domains in high mobility group n variants modulate specific chromatin modifications. *Journal of Biological Chemistry, 281*(15), 10182-10187. doi:10.1074/jbc.M600821200

van der Lugt, N. M., Domen, J., Linders, K., van Roon, M., Robanus-Maandag, E., te Riele, H., . . . et al. (1994). Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. *Genes Dev, 8*(7), 757-769.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol, 3*(7), RESEARCH0034.

Vasil'eva, E. A., Melino, D., & Barlev, N. A. (2015). [CRISPR/Cas system for genome editing in pluripotent stem cells]. *Tsitologiia, 57*(1), 19-30.

Vastenhouw, N. L., & Schier, A. F. (2012). Bivalent histone modifications in early embryogenesis. *Current Opinion in Cell Biology, 24*(3), 374-386. doi:10.1016/j.ceb.2012.03.009

Velkey, J. M., & O'Shea, K. S. (2013). Expression of Neurogenin 1 in Mouse Embryonic Stem Cells Directs the Differentiation of Neuronal Precursors and Identifies Unique Patterns of Down-stream Gene Expression. *Developmental Dynamics, 242*(3), 230-253. doi:10.1002/dvdy.23920

Vestner, B., Bustin, M., & Gruss, C. (1998). Stimulation of replication efficiency of a chromatin template by chromosomal protein HMG-17. *Journal of Biological Chemistry, 273*(16), 9409-9414. doi:DOI 10.1074/jbc.273.16.9409

Vignali, M., Hassan, A. H., Neely, K. E., & Workman, J. L. (2000). ATP-dependent chromatin-remodeling complexes. *Mol Cell Biol, 20*(6), 1899-1910. doi:Doi 10.1128/Mcb.20.6.1899-1910.2000

Voigt, P., Tee, W. W., & Reinberg, D. (2013). A double take on bivalent promoters. *Genes Dev, 27*(12), 1318-1338. doi:10.1101/gad.219626.113

Vouillot, L., Thelie, A., & Pollet, N. (2015). Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3 (Bethesda), 5*(3), 407-415. doi:10.1534/g3.114.015834

Waddington, C. H. (2012). The epigenotype. 1942. *Int J Epidemiol, 41*(1), 10-13. doi:10.1093/ije/dyr184

Wakayama, T., Nakata, H., Kumchantuek, T., Gewaily, M. S., & Iseki, S. (2015). Identification of 5-Bromo-2 '-Deoxyuridine-Labeled Cells during Mouse Spermatogenesis by Heat-Induced Antigen Retrieval in Lectin Staining and Immunohistochemistry. *Journal of Histochemistry & Cytochemistry, 63*(3), 190-205. doi:10.1369/0022155414564870

Walther, C., & Gruss, P. (1991). Pax-6, a murine paired box gene, is expressed in the developing CNS. *Development, 113*(4), 1435-1449.

Wan, M., Liang, J., Xiong, Y., Shi, F., Zhang, Y., Lu, W., . . . Songyang, Z. (2013). The trithorax group protein Ash2l is essential for pluripotency and maintaining open chromatin in embryonic stem cells. *J Biol Chem, 288*(7), 5039-5048. doi:10.1074/jbc.M112.424515

Wang, C., Xia, C., Bian, W., Liu, L., Lin, W., Chen, Y. G., . . . Jing, N. (2006). Cell aggregation-induced FGF8 elevation is essential for P19 cell neural differentiation. *Molecular Biology of the Cell, 17*(7), 3075-3084. doi:10.1091/mbc.E05-11-1087

Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., & Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell, 153*(4), 910-918. doi:10.1016/j.cell.2013.04.025

Wang, J., Mager, J., Schnedier, E., & Magnuson, T. (2002). The mouse PcG gene eed is required for Hox gene repression and extraembryonic development. *Mamm Genome, 13*(9), 493-503. doi:10.1007/s00335-002-2182-7

Wang, K., Chen, Y., Chang, E. A., Knott, J. G., & Cibelli, J. B. (2009). Dynamic Epigenetic Regulation of the Oct4 and Nanog Regulatory Regions during Neural Differentiation in Rhesus Nuclear Transfer Embryonic Stem Cells. *Cloning and Stem Cells, 11*(4), 483-496. doi:10.1089/clo.2009.0019

Wang, T., Lander, E. S., & Sabatini, D. M. (2016). Large-Scale Single Guide RNA Library Construction and Use for CRISPR-Cas9-Based Genetic Screens. *Cold Spring Harb Protoc, 2016*(3), pdb top086892. doi:10.1101/pdb.top086892

Wang, X., Wang, Y., Wu, X., Wang, J., Wang, Y., Qiu, Z., . . . Yee, J. K. (2015). Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat Biotechnol, 33*(2), 175-178. doi:10.1038/nbt.3127

Wang, X. M., Nagl, N. G., Wilsker, D., Van Scoy, M., Pacchione, S., Yaciuk, P., . . . Moran, E. (2004). Two related ARID family proteins are alternative subunits of human SWI/SNF complexes. *Biochemical Journal, 383*, 319-325. doi:10.1042/Bj20040524

Ware, C. B., Wang, L. L., Mecham, B. H., Shen, L. L., Nelson, A. M., Bar, M., . . . Blau, C. A. (2009). Histone Deacetylase Inhibition Elicits an Evolutionarily Conserved Self-Renewal Program in Embryonic Stem Cells. *Cell Stem Cell, 4*(4), 359-369. doi:10.1016/j.stem.2009.03.001

Wataya, T., Ando, S., Muguruma, K., Ikeda, H., Watanabe, K., Eiraku, M., . . . Sasai, Y. (2008). Minimization of exogenous signals in ES cell culture induces rostral hypothalamic differentiation. *Proc Natl Acad Sci U S A, 105*(33), 11796-11801. doi:10.1073/pnas.0803078105

Weigmann, N., Trieschmann, L., & Bustin, M. (1997). Enhancement of the transcription potential of nascent chromatin by chromosomal proteins HMG-14/-17 is coupled to nucleosome assembly and not DNA synthesis. *DNA and Cell Biology, 16*(10), 1207-1216. doi:DOI 10.1089/dna.1997.16.1207

Weisbrod, S., & Weintraub, H. (1979). Isolation of a Subclass of Nuclear Proteins Responsible for Conferring a Dnase I-Sensitive Structure on Globin Chromatin. *Proc Natl Acad Sci U S A, 76*(2), 630-634. doi:DOI 10.1073/pnas.76.2.630

Weng, M. Z., Song, F. B., Chen, J. Y., Wu, J. Y., Qin, J., Jin, T., & Xu, J. M. (2015). The high-mobility group nucleosome-binding domain 5 is highly expressed in breast cancer and promotes the proliferation and invasion of breast cancer cells. *Tumor Biology, 36*(2), 959-966. doi:10.1007/s13277-014-2715-1

West, K. L., Castellini, M. A., Duncan, M. K., & Bustin, M. (2004). Chromosomal proteins HMGN3a and HMGN3b regulate the expression of glycine transporter 1. *Mol Cell Biol, 24*(9), 3747-3756.

West, K. L., Ito, Y., Birger, Y., Postnikov, Y., Shirakawa, H., & Bustin, M. (2001). HMGN3a and HMGN3b, two protein isoforms with a tissue-specific expression pattern, expand the cellular repertoire of nucleosome-binding proteins. *J Biol Chem, 276*(28), 25959-25969. doi:10.1074/jbc.M101692200

West, K. L., Postnikov, Y. V., Birger, Y., & Bustin, M. (2003). Chromatin decompaction method by HMGN proteins. *Methods Enzymol, 371*, 521-536. doi:10.1016/S0076-6879(03)71039-8

White, J., & Dalton, S. (2005). Cell cycle control of embryonic stem cells. *Stem Cell Reviews, 1*(2), 131-138. doi:Doi 10.1385/Scr:1:2:131

Wiese, C., Rolletschek, A., Kania, G., Blyszczuk, P., Tarasov, K. V., Tarasova, Y., . . . Wobus, A. M. (2004). Nestin expression - a property of multi-lineage progenitor

cells? *Cellular and Molecular Life Sciences, 61*(19-20), 2510-2522. doi:10.1007/s00018-004-4144-6

Williams, R. L., Hilton, D. J., Pease, S., Willson, T. A., Stewart, C. L., Gearing, D. P., . . . Gough, N. M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature, 336*(6200), 684-687.

Wills, A. E., Choi, V. M., Bennett, M. J., Khokha, M. K., & Harland, R. M. (2010). BMP antagonists and FGF signaling contribute to different domains of the neural plate in Xenopus. *Developmental Biology, 337*(2), 335-350. doi:10.1016/j.ydbio.2009.11.008

Wilson, S. W., & Rubenstein, J. L. R. (2000). Induction and dorsoventral patterning of the telencephalon. *Neuron, 28*(3), 641-651. doi:Doi 10.1016/S0896-6273(00)00171-9

Woodcock, C. L., Skoultchi, A. I., & Fan, Y. H. (2006). Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Research, 14*(1), 17-25. doi:10.1007/s10577-005-1024-3

Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., & Smith, A. (2011). Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nature Cell Biology, 13*(7), 838-U246. doi:10.1038/ncb2267

Wu, M., Zhang, Y., Wu, N. H., & Shen, Y. F. (2009). Histone Marks and Chromatin Remodelers on the Regulation of neurogenin1 Gene in RA Induced Neuronal Differentiation of P19 Cells. *Journal of Cellular Biochemistry, 107*(2), 264-271. doi:10.1002/jcb.22122

Wu, X., Scott, D. A., Kriz, A. J., Chiu, A. C., Hsu, P. D., Dadon, D. B., . . . Sharp, P. A. (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol, 32*(7), 670-676. doi:10.1038/nbt.2889

Xue, Y. Z., Gao, X., Lindsell, C. E., Norton, C. R., Chang, B., Hicks, C., . . . Gridley, T. (1999). Embryonic lethality and vascular defects in mice lacking the Notch ligand JAGGED1. *Human Molecular Genetics, 8*(5), 723-730. doi:DOI 10.1093/hmg/8.5.723

Yang, H., Wang, H. Y., Shivalila, C. S., Cheng, A. W., Shi, L. Y., & Jaenisch, R. (2013). One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell, 154*(6), 1370-1379. doi:10.1016/j.cell.2013.08.022

Ye, S. D., Li, P., Tong, C., & Ying, Q. L. (2013). Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1. *Embo Journal, 32*(19), 2548-2560. doi:10.1038/emboj.2013.175

Ying, Q. L., Nichols, J., Chambers, I., & Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell, 115*(3), 281-292. doi:Doi 10.1016/S0092-8674(03)00847-X

Ying, Q. L., Stavridis, M., Griffiths, D., Li, M., & Smith, A. (2003). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol, 21*(2), 183-186. doi:10.1038/nbt780

Ying, Q. L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., . . . Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature, 453*(7194), 519-U515. doi:10.1038/nature06968

Yoon, K. J., & Gaiano, N. (2005). Notch signaling in the mammalian central nervous system: insights from mouse mutants (vol 8, pg 709, 2005). *Nature Neuroscience, 8*(10), 1411-1411. doi:DOI 10.1038/nn1005-1411b

Yuan, A., Rao, M. V., Veeranna, & Nixon, R. A. (2012). Neurofilaments at a glance. *Journal of Cell Science, 125*(Pt 14), 3257-3263. doi:10.1242/jcs.104729

Yum, S. W., Zhang, J. X., Mo, K., Li, J., & Scherer, S. S. (2009). A Novel Recessive NEFL Mutation Causes a Severe, Early-Onset Axonal Neuropathy. *Annals of Neurology, 66*(6), 759-770. doi:10.1002/ana.21728

Yun, K., Fischman, S., Johnson, J., de Angelis, M. H., Weinmaster, G., & Rubenstein, J. L. R. (2002). Modulation of the notch signaling by Mash1 and DIx1/2 regulates sequential specification and differentiation of progenitor cell types in the subcortical telencephalon. *Development, 129*(21), 5029-5040.

Zhan, Y., Paolicelli, R. C., Sforazzini, F., Weinhard, L., Bolasco, G., Pagani, F., . . . Gross, C. T. (2014). Deficient neuron-microglia signaling results in impaired functional brain connectivity and social behavior. *Nature Neuroscience, 17*(3), 400-406. doi:10.1038/nn.3641

Zhang, J. W., & Li, L. H. (2005). BMP signaling and stem cell regulation. *Developmental Biology, 284*(1), 1-11. doi:10.1016/j.ydbio.2005.05.009

Zhang, S., Zhu, I., Deng, T., Furusawa, T., Rochman, M., Vacchio, M. S., . . . Bustin, M. (2016). HMGN proteins modulate chromatin regulatory sites and gene expression during activation of naive B cells. *Nucleic Acids Res, 44*(15), 7144-7158. doi:10.1093/nar/gkw323

Zhong, W. M., Jiang, M. M., Schonemann, M. D., Meneses, J. J., Pedersen, R. A., Jan, L. Y., & Jan, Y. N. (2000). Mouse numb is an essential gene involved in cortical neurogenesis. *Proc Natl Acad Sci U S A, 97*(12), 6844-6849. doi:DOI 10.1073/pnas.97.12.6844

Zhu, N., & Hansen, U. (2007). HMGN1 modulates estrogen-mediated transcriptional activation through interactions with specific DNA-binding transcription factors. *Mol Cell Biol, 27*(24), 8859-8873. doi:10.1128/Mcb.01724-07

Zupkovitz, G., Tischler, J., Posch, M., Sadzak, I., Ramsauer, K., Egger, G., . . . Seiser, C. (2006). Negative and positive regulation of gene expression by mouse histone deacetylase 1. *Mol Cell Biol, 26*(21), 7913-7928. doi:10.1128/Mcb.01220-06