THE UNIVERSITY
of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Dissecting the function of *Myef2* in neural stem cells and glioblastoma

Katrina Anne McCarten

Supervisor: Prof. Steven M Pollard

A thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

MRC Centre for Regenerative Medicine and Edinburgh Cancer Research UK Centre

2019

Declaration

I hereby declare that this thesis and the work presented in it was composed by and originated entirely from me and has not been submitted in any form for another degree or diploma at any other institution. Any information derived from the published or unpublished work of others has been acknowledged in the text and references are listed in the bibliography.

Katrina Anne McCarten

29th August 2019

Copyright declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivations license. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the license terms of this work.

Abstract

Glioblastoma (GBM) is the most common malignant primary brain tumour in adults, with a median survival time of approximately ~15 months after diagnosis. GBM stem cells (GSCs) underpin GBM's aggressiveness, resistance to radiation and chemotherapy, and disease recurrence. Glioblastoma stem cells (GSCs) are capable of tumour formation upon orthotopic transplantation into immunocompromised mice and share many properties of neural stem cells (NSCs). One feature common to both GSCs and NSCs is high expression of the transcription factor SOX2. SOX2 is known to be essential for GCS self-renewal. To understand further how SOX2 exerts its function in GSCs, the Pollard laboratory studied the interactome of SOX2. Using SICAP-MS – a method that looks at the on-chromatin interactors of the protein of interest – the Pollard laboratory have identified MYEF2 as a candidate interacting partner. MYEF2 contains multiple RNA binding domains and is expressed primarily in the brain and testis. Here I explored the function of MYEF2 within NSCs and GSCs.

I confirm *Myef2* is expressed within the brain, however was not specifically enriched in stem cell containing regions. CRISPR/Cas9 mediated knock-in of mCherry and HA to *Myef2* was successfully carried out in mouse NSCs. We observed that MYEF2 is a nuclear protein which retains nuclear localisation and expression in both proliferative and quiescence NSCs, as well as differentiating progeny. During mitosis, unlike SOX2, MYEF2 is not retained on the mitotic chromatin indicating it is not a "bookmarking" factor. CRISPR/Cas9 mediated knock-out of *Myef2* suggests that although it is not essential for the continued proliferation of GSCs, it has a role in regulating the exit from quiescence. Consistent with this, in *Myef2* knock-out mouse

GSCs, the rate of tumour progression is slower, and mice have a significant survival advantage, suggesting there is an important role for *Myef2* in driving tumour growth.

In this thesis, I also describe the use of SMASh tag degron technology to precisely control the degradation of SOX2 in mouse NSCs. SMASh, a drug degradable self-cleaving degron, was fused to the C-terminal of endogenous SOX2 in mouse NSCs. We find that this works well and can therefore be a powerful tool in future studies of GSCs.

Lay Summary

Glioblastoma is the most common primary brain cancer in adults. Prognosis is poor, with only 5% of patients surviving more than 5 years after diagnosis. A major contributor to this poor prognosis is the high level of disease recurrence, that occurs from a small number of tumour cells which evade current therapies. These tumour cells behave in an aberrant manner, capable of proliferating excessively and forming a tumour. These cells are controlled by many molecular pathways that are used in the developing brain.

One factor involved in the proliferation of glioblastoma cells is a protein known as SOX2. SOX2 works in collaboration with many other proteins, many of which are still unknown. It is very difficult to generate therapies which directly target SOX2 using standard drug discovery methods. However, SOX2 interaction partners may be easier to block. To address this, we are identifying the binding partners of SOX2 within glioblastoma cells. Currently, one new factor that has been identified is MYEF2.

During my PhD, I assessed the role MYEF2 plays in glioblastoma and if it is an essential partner of SOX2. In addition to studying MYEF2, I am expanding our experimental tools to study the functions of proteins in glioblastoma. One such tool is the SMASh tag system. Once established, this tool should allow us to understand better whether a particular protein is essential to tumour formation, and hence a good therapeutic target. These studies will further our understanding of glioblastoma and contribute to the development of new targeted therapies against this aggressive disease.

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Steven M Pollard for all of his support and guidance over the past few years. His knowledge, professionalism and constant enthusiasm are inspiring, and have aided me to grow as a scientist.

Secondly, I would like to thank all of the members of the Pollard lab, especially Vivien Grant, Carla Blin and Pooran Dewari for all of their technical guidance. Thanks, as well to SCRM staff for all the assistance. A special thanks also goes to Kirsty Ferguson and Ute Koeber for all of your support, suggestions, and excellent company during lunch and late in the lab. Without both of you, my PhD would have been a much duller experience.

Thirdly on a personal note, I would like to thank my family and friends for the constant encouragement. Most importantly I want to thank my partner, Stephen Du. You were supportive of my choice to do a PhD, even if it was one that put us 370 miles apart. Since, you have always been supportive and understanding throughout all of the bumps along my PhD journey.

Finally, I would like to thank the MRC for providing me with the financial support to carry out my PhD.

List of abbreviations

ASV	Asunaprevir
BMP4	Bone morphogenic protein 4
Cas9	CRISPR associated protein 9
ChIP	Chromatin immunoprecipitation
CNS	Central nervous system
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
DAPI	4',6-Diamidino-2-Phenylindole, Dihydrochloride
dCas9	Nuclease dead Cas9
DMEM	Dulbecco's Modified Eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSB	Double-strand break
EDTA	Ethylenediamine tetra acetic acid
EdU	5-ethynyl-2'-deoxyuridine
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
FACS	Fluorescence-activated cell sorting
FGF2	Fibroblast growth factor 2
GBM	Glioblastoma multiforme
GFP	Green fluorescent protein
GSC	Glioblastoma stem cell
HCV	Hepatitis C virus
HDR	Homology-directed repair
HR	Homologous recombination
ICC	Immunocytochemistry
IP	Immunoprecipitation

MS	Mass spectrometry
NaCl	Sodium chloride
NaF	Sodium fluoride
nCas9	Nickase Cas9
NEPC	Neuroepithelial cell
NHEJ	Non-homologous end joining
NSC	Neural stem cell
P2A	Peptide 2A
PAM	Protospacer adjacent motif
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
POI	Protein of interest
PuroR	Puromycin resistance cassette
RG	Radial glial progenitors
RNP	Ribonucleoprotein
SDS	Sodium Dodecyl Sulphate
SGZ	Subgranular zone
SICAP	Selective isolation of chromatin-associated proteins
SMASh	Small molecule-assisted shutoff
ssODN	Single-stranded oligo DNA nucleotides
SVZ	Subventricular zone
TALEN	Transcription activator-like effector nuclease
TCGA	The Cancer Genome Atlas
TME	Tumour microenvironment
tracrRNA	Trans-activating CRISPR RNA
UTR	Untranslated region

List of figures and tables

Contents

Chapter 1 Introduction

1.1 Overview

Understanding the fundamental biology of cancer enables the generation of new rationally designed therapies. Glioblastoma multiforme (GBM) is the most common adult primary brain tumour and has a poor survival rate. The disease is driven by a glioma cancer stem cell (GSC) population, which has characteristics of neural stem cells (NSCs). One feature of GSCs, is high level expression of the core transcriptional network protein SOX2. Understanding the SOX2 interactome, and its function, may reveal new ways to target this protein for therapeutic gain. Major advancements in mammalian genome engineering, coupled with improved cell culture models, have made this more achievable than ever, allowing rapid large-scale investigations of proteins that have been previously understudied.

This introduction will cover current literature on several key topics. Firstly, the biology of GBM, including discussion of the known common genetic mutations, the core transcriptional network and experimental models. Secondly, current mammalian editing methods with a focus on CRISPR/Cas9. The recent advancements in this field have led to an ever-expanding repertoire of sophisticated genetic manipulations that can be performed efficiently. Thirdly, the use of degrons to post-translationally control protein levels and thereby confirm key therapeutic targets.

1.2 The biology of GBM

1.2.1 Glioma, GBM and current treatment strategy

The World Health Organisation (WHO) has classified over 130 types of brain cancer[1] which are graded on a scale of I-IV, with grade I-II corresponding to benign/low grade tumours and grade III-IV corresponding to high grade/malignant tumours. Gliomas

represent the most common brain tumours observed and are named due to the observed glial cell histology of the tumour. Gliomas are categorised into astrocytoma (from astrocytes), oligodendroglioma (from oligodendrocytes) or ependymoma (from ependymal cells). Grade IV astrocytoma, most frequently referred to as glioblastoma multiforme (GBM), is the most common high-grade primary brain tumour in adults, with tumours most often arising in the cerebral hemisphere, particularly the frontal and temporal lobes.

Treatment for GBM comprises of surgery to remove the bulk tumour, radiotherapy and chemotherapy with temozolomide (TMZ). Whilst this intensive regime increases survival time[2], almost all patient relapse and have a poor median survival rate of just ~12-15 months. The blood-brain barrier, whilst partially disrupted in patients with GBM[3], complicates drug delivery. More significantly, GBM is highly invasive making it impossible to remove all of the malignant cells during surgery[4]. The residual tumour is thought to contain a population of glioma cancer stem cells (GSCs), which trigger re-growth of the tumour. GSCs are relatively resistant to radiation therapy, through activation of the DNA damage checkpoint in response to radiation and efficient repair of radiation induced DNA damage[5], and a subpopulation of these GSCs are in a quiescent state making them resistant to classical chemotherapy treatments such as temozolomide (TMZ)[6–9].

1.2.2 The cancer stem cell model and GSCs

Adult stem cells, also known as somatic stem cells, have been identified in a range of tissues throughout the body[10]. Normal adult stem cells are tissue specific and exhibit the ability to self-renew and to differentiate into cell types of the tissue of origin. The cancer stem cell (CSC) hypothesis proposes that tumours are initiated and maintained by a subpopulation of tumour cells that share biological properties of

normal adult stem cells. This is based on observations that some human cancers have cells that express stem cell and progenitor cell markers, and have the ability to self-renew and differentiate, giving rise to tumours that phenotypically resemble their tissue of origin. While CSCs share properties of normal stem cells, the self-renewal of CSCs is deregulated, leading to upregulated proliferation. Proliferative signalling in cancer arises from many sources, with many of the genes that promote self-renewal being oncogenes, and often, signalling pathways involved in normal stem cell development (such as Wnt, Shh and Notch) are classically associated with cancer[11]. In addition to acquiring proliferative signalling, cancer must overcome the inherent tumour suppressive mechanisms of senescence (a state of permanent cell cycle arrest) and apoptosis (cell death), that provide tight control over uncontrolled proliferation. In response to DNA damage and oncogene activation, cells can undergo apoptosis or cellular senescence. These processes are regulated by Retinoblastoma protein (Rb) and p53 pathways, which are often inactivated in cancer. Additional, to avoid replicative senescence in response to telomere erosion, cells can either activate telomerase or extend telomeres through alternative lengthening of telomeres (ALT). These alterations, enable escape from control and hence tumour formation[12,13].

Some of the first evidence for the existence of CSC came from studies of carcinoma[14] and of human acute myeloid leukaemia (AML), where a subpopulation of cells initiated AML when transplanted into severe combined immune-deficient (SCID) mice[15,16]. Subsequently CSCs have been shown to exist within several solid tumours, including breast cancer[17] and brain cancer[18,19,20]. Cells expressing the human neural stem cell marker CD133[21] were found across a range of primary paediatric brain tumours, including astrocytomas. These $CD133^{+}$ cells were found to have a capacity for proliferation, self-renewal and differentiation *in vitro*[18,20]. Subsequently, CD133 expressing cells were shown to generate tumours *in vivo*[19] with a xenograft phenotype

very similar to the original tumour, including recapitulation of the cellular heterogeneity observed in GBM. In the same year, independent studies showed that GBM generated cell lines could also generate tumours *in vivo*[22]. Subsequently, many other markers of GSCs have been identified including SSEA1[23],CD44[24] and A2B5[25], though no single marker defines a universal GSC population[26,27].

GSCs express a range of immature neural stem and progenitor cell markers, including NESTIN and SOX2, which are considered as stemness markers. Within GBM GSCs can be found in range of different cellular states[28] including proliferative, quiescent[6–9] and more differentiated states[29]. In the proliferative state cells express proliferation markers such as Ki67, and this state can be easily studied *in vitro* using proliferative NSC culture conditions. Quiescent label-retaining GSCs are associated with several features including: low RNA content; downregulation of cell-cycle progression genes such as *CCNB1;* upregulation of several tumour suppressor genes such as *CDKN1A*[30]*,* and expression of CD9[31], a feature of quiescent neural stem cells[32,33]. The ability to mimic the quiescent state of GSCs *in vitro* is not yet well established, but evidence indicates that signalling from BMP4 and FGF2 can induce quiescence in NSCs (Maria Angeles Marques, Pollard lab), whilst BMP alone may induce quiescence in GSCs[34]. A range of more differentiated cell types can be observed within GBM, including astrocytic, neuronal and OPC/oligodendrocytic like cells, in line with the hierarchical CSC model. Exposure to differentiation signals both *in vitro,* such as serum or BMP4, and those in the *in vivo* environment, can induced the expression of differentiation markers within GSCs including, neuronal markers ß-tubulin/TUJ1 and MAP2, astrocytic marker GFAP and oligodendroglial marker GalC[35–37], with the relative proportions of each marker reflecting the original tumour phenotype. Some lines of evidence suggest that the differentiation of GSCs is not terminal, and the cells

are plastic able to return to less differentiated states[27,38,39], a cavate for the potential use of differentiation therapy in GBM.

Experimental findings continue to support the function and existence of GSCs, including those of *in vivo* lineage tracing experiments which support the proliferative hierarchy[40] expected from CSC models, though is it unclear whether the cells undergo a truly terminal differentiation. The study of CSCs can be complex. CSCs are not a single definitive cell type or state, and they can undergo epigenetic and genetic changes during tumour development and treatment.

1.2.3 The genetics of GBM

GBM tumours typically consist of morphically diverse cells, with heterogeneity being observed in both an intratumoral and intertumoral manner. This complexity occurs not only at the phenotypic level, but also on a genetic and epigenetic scale, with tumours similar in morphological and phenotypical appearance varying in treatment response, prognosis and can harbour a distinct spectrum of mutations (Figure 1.1).

Understanding the biology of GBM at the phenotypic, genetic, epigenetic and transcriptional level may provide new opportunities for therapies. Consortiums, such as The Cancer Genome Atlas (TCGA)[41], aid in providing a wealth of information on genetic disruptions – nucleotide substitutions, copy number aberrations, chromosomal rearrangements – as well as DNA methylation patterns and gene expression data from multiple GBM patients. Collating these large-scale analyses has expanded our knowledge of the catalogue of mutations associated with these tumours; however, we still have a poor understanding of how these operate functionally.

The first cancer reported by the TCGA was GBM in 2008[15], with an expanded study performed in 2013[16]. In the GBM samples, mutations within the PI3K/RTK, p53 and RB1 pathways were consistently identified, with 90%, 86% and 79% of samples containing mutations within each pathway, respectively[16]. Mutations occur at several points throughout of these pathways: with mutations within *MDM2, MDM4, CDKN2A* and *TP53* altering the p53 pathway; mutations within *CDK4, CDK6, CCCND2, CDKN2A/B and RB1* altering the RB1 pathway, and mutations within *PI3KCA, PIK3R1, PTEN, EGFR, PDGFRA* and *NF1* altering the PI3K/RTK pathway. The frequency of which a gene is mutated varies, with *EGFR* mutations found in 57.4% of GBM samples, whilst *PDGFRA* mutations are only seen in 13% of GBMs. These finding highlight, that highly heterogenous somatic mutations often converge onto 3 core pathways. However, there is no single actionable driver mutation that encompasses all GBM patients.

One aim of understanding the genetics, epigenetics and transcriptome of GBM is to identify biomarkers, to aid in the prediction of prognosis and response to treatment. For GBM, only a few functional biomarkers have been identified including IDH1/IDH2 mutation status and the methylation prolife of MGMT. The isocitrate dehydrogenase enzyme IDH1 has been found to be mutated in 12% of GBMs. This mutation is primarily found in young patients and most patients with secondary GBMs– the progressive form of the disease that arises from lower grade tumours. Mutations in IDH1, and IDH2, are associated with an increase in overall survival[42] and the WHO[1] now subdivides GBM into IDH mutant and IDH WT classes. O(6)-methylguanine DNA methyltransferase (MGMT), is a DNA repair enzyme. Promoter methylation mediated silencing of MGMT can lead lower level expression and correlates with better response to treatment with TMZ[43].

1.2.4 Transcriptional subtypes of GBM

Transcriptional profiling of GBM tumours has led to a proposed classification of 3 transcriptional subtypes of GBM: Proneural (PN), Classical (CL) and Mesenchymal (MES)[44,45], based on the mRNA profiles of bulk tumours. High levels of intra tumour heterogeneity are often observed, with multiple subtypes coexisting in the same tumour. Additionally, the bulk tumour subtype can change over time and after therapy[28,45,46]. Several genetic mutations correlate with the subtypes. The Pro-neural subtype correlates with *PDGFRA* amplification and *TP53* mutations. The classical subtype with *EGFR* amplification, Chr10 loss, *CDKN2A* deletion and absence of *TP53* mutations. Finally, the mesenchymal subtype, which has elevated invasive potential, poor clinical prognosis, and significantly shortened time to recurrence following initial treatment when compared to other subtypes, is associated with *NF1* loss and this NF1 deficiency may drive the higher levels of tumour associated macrophages/microglial infiltration associated with this subtype[45].

However, these are not robust classifications, and single cell analysis reveals more of a continuum of cell states and varied intratumoral genetic drivers. The subtypes are reminiscent of distinct neural cell types[44]. Recent single cell RNA analysis of IDH WT GBM[29], in fact has found that malignant cells in GBM recapitulate different cellular states, including neural progenitor like (NPC), oligodendrocyte progenitor like (OPC), astrocyte like (AC) and mesenchymal like (MES) states. The abundance of each state within a tumour, correlates with the transcriptional subtypes and the associated genetic mutations. High abundance of the OPC and NPC like states are associated with the proneural subtype and *PDGFRA/CDK4* mutations, respectively. High abundance of MES like state correlates with the mesenchymal subtype, NF1 loss and Chr5q deletions. However, this does not mirror any specific cell type and may instead be a transcriptionally corrupted state, with GSCs harbouring NF1 loss, PTEN loss and

EGFRVIII overexpression, expressing myeloid-specific master transcription factor interferon regulatory factor 8 (IRF8) via epigenetic immunoediting (Pollard lab, unpublished). Finally, high abundance of the astrocytic cellular state correlates with the classical subtype and *EGFR* amplification. Cells can span over more than one cellular state and, through *in vivo* barcoding, plasticity was identified, with a single cell capable of generating all four cellular states. These transcriptional signatures therefore are highly plastic[29].

1.2.5 The tumour microenvironment and epithelial-mesenchymal transitions

In addition to cancer cells, glioblastoma tumours contain: non-cancerous cells including microglial/macrophages, dendritic cells, neutrophils, astrocytes, neurons, pericytes and endothelial cells; a brain distinct extracellular matrix, interstitial fluid, and the blood brain barrier, all of which constitute the tumour microenvironment (TME). The TME is dynamic, with therapeutic treatments such as temozolomide triggering TME remodelling, and plays a key role in tumour progression. One mechanism by which the TME contributes to tumour progression is through crosstalk between tumour cells and tumour associated microglial/macrophages (TAMs), with tumour cells secreting cytokines and chemoattractants whilst TAMs release pro-survival and pro-tumorigenic factors which can contribute to epithelial-mesenchymal transitions (as discussed below)[47,48].

Epithelial-mesenchymal transitions, in which epithelial cells undergo multiple (often reversible) biological changes to switch to a mesenchymal cell type, occur during embryogenesis, organ development, wound healing and cancer, where the associated production of ECM components and altered migratory capacity, increase invasiveness and enable metastasis of cancer[49,50]. Whilst GBM rarely form

metastases outside of the CNS, they are highly infiltrative and can have diffuse growth[51], with evidence indicating an EMT-like process occurs[52] with the mesenchymal subtype having highest upregulation of EMT signature genes[53]. This EMT-like process, also referred to as GMT (glial-mesenchymal transition)[54], is non-classical with neural cells already assuming a mesenchymal phenotype and classical E-cadherin to N-cadherin switches are unlikely to correlate with EMTs, due to the absence of E-cadherin expression in GBM cultures[55]. Several core activators of EMT, including *Twist, ZEB1/2, Snail* and *Slug,* are upregulated during GMT and are triggered through WNT/ß-catenin, which is elevated in human GBM[55–57], and TGF-ß, which is produced by tumour associated microglial/macrophages and in response to radiation therapy[54,58].

In summary, GBM is highly heterogenous at both intratumoral and intertumoral levels. Multiple factors contribute to the cell phenotype, including genetics, epigenetics, the microenvironment and dynamic changes in cell cycle or epithelial-mesenchymal transitions.

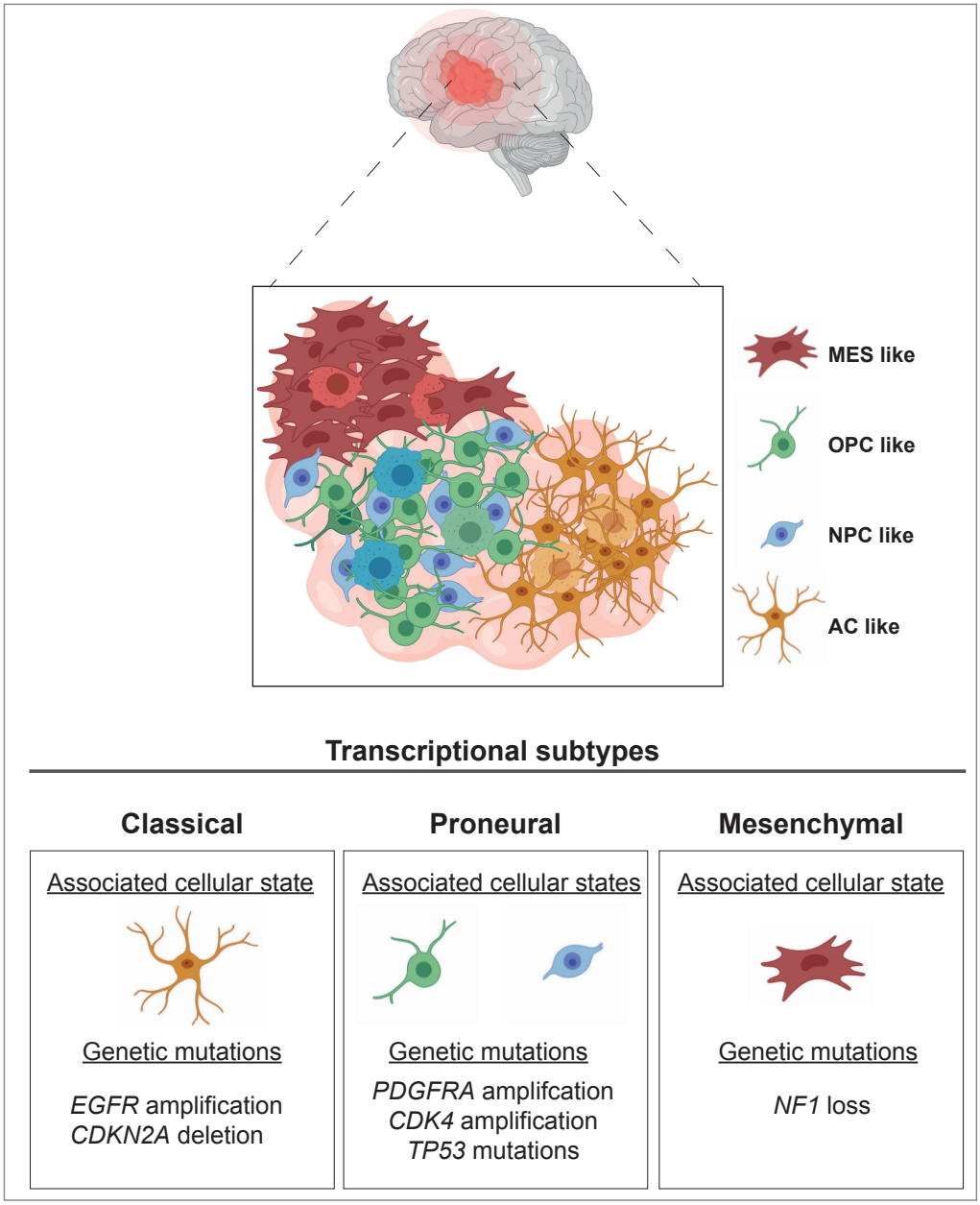


Figure 1.1 | GBM intratumoral heterogeneity

Depiction of the intratumoral heterogeneity of GBM patient samples. The transcriptional subtypes, associated plastic cellular states (MES like = mesenchymal like, OPC like = oligodendrocyte progenitor like, NPC like = neural progenitor like and AC like= astrocyte like) and classically found genetic mutations are highlighted. This figure was generated using Biorender.

1.2.6 The cellular origins of GBM

The cancer stem cell hypothesis does not propose to assume that the cell of origin of a CSC is a stem cell. However, this is an attractive theory, especially when combined with the multi-hit mutational theory. For this to occur in GBM, there must be a population of adult NSCs. Below is a brief summary of the development of NSCs with the mammalian brain, with evidence derived mainly from rodents[59,60].

Within development, neuroepithelial cells (NEPCs) are the earliest cells to arise which are considered as NSCs. NEPCs arise from the neural plate, following neural induction, and undergo symmetric division expanding their numbers. At the onset of neurogenesis, around E11[61], NEPCs transition to a distinct progenitor/stem cell referred to as radial glial progenitors (RGs). RGs express a range of markers including *Sox2, Blbp, Vcam1, Nestin* and *Glast*. RGs undergo asymmetrical division from E11 – E17, generating neurons directly or through generation of transient amplifying intermediated progenitor cells (IPCs)[62]. At later stages of development, RGs/IPCs switch from neurogenesis to gliogenesis, and finally at the end of embryonic development most RGs convert to astrocytes or ependymal cells.

Some RGs, now referred to as adult NSCs, persist with at least two specific regions of the brain, the subgranular zone (SGZ) of the hippocampus and the subventricular zone (SVZ) of the forebrain. NSCs of the SGZ express RG associated markers *Sox2, Gfap* and *Nestin*[59], and their differentiation potential is largely restricted to the neuronal lineage[63]. In the SVZ, quiescent NSCs[64] (also known as type B1 cells) express RG associated markers including *Vcam1, Nestin, Gfap, Glast*, and *Sox2*[59] and have the potential to make neurons[65], astrocytes, and to a lesser extent oligodendrocytes[66].

The cell of origin of GBM is a controversial topic. Multiple views and lines of evidence exist, with some proposing that dedifferentiation leads to the generation of the cell of GBM origin. Cultured early postnatal astrocytes, which may not fully represent *in vivo* mature adult astrocytes, overexpressing EGFR coupled with the deletion of *Ink4a/Arf* are capable of developing tumours in immunodeficient mice[67]. A more commonly held view is that adult NSCs form the cell of origin. Induction of high grade gliomas in mice ablated of conditional alleles *Nf1, Trp53 and Pten,* either within the SVZ or within specific cell populations, support the proposition that neural progenitor cells or NSCs, including quiescent NSCs, are the cell of origin[68–72]. In line with this the susceptibility to malignant transformation from the loss *of Nf1, Trp53 and Pten* decreases with increased lineage restriction[73]. Recent genetic analysis of human GBM samples and the corresponding tumour free SVZ, found the macroscopically normal tissue contained key GBM driver mutations, providing the most direct support to date that cells within the SVZ are the cell of origin[74].

1.2.7 Modelling GSCs and glioblastoma

Until relatively recently, human glioma cell lines, such as U87MG, U251 and T98G, were commonly used. These lines, which are cultured in serum, do not reflect the NSC pathways that underlie GSCs and tumours formed upon xenotransplantation do not resemble GBM tumours[75]. On top of this U87MG, one of the most popular cell lines, does not match the original stock[76]. Because of this, the current gold standard for GBM models are now patient-derived lines and primary cultures. Culture conditions used to expand adult NSCs[77], are applied to primary GBM specimens. These can be expanded continuously in culture and retain the genetic and transcriptional states of the parental tumour[75]. Patient derived lines can be grown as either adherent monolayers[78], or using suspension culture conditions (spheres or

organoid conditions)[79]. Efforts are ongoing to derive repositories of well annotated high quality patient derived glioma cell lines, as well as isogenic pairs with driver mutations reverted to wild-type (http://www.gcgr.org.uk).

Whilst patient-derived lines are highly valuable, they can be slow to expand and, like the patient disease, genetically heterogeneous making deconvolution of observations more challenging. GCSs recapitulate many signatures of NSCs, making NSCs the obvious cell for *in vitro* modelling, *in vivo* modelling, and dissecting the downstream effectors of the common driver mutations.

Genetically engineered mouse models are a major experimental model and a tool for preclinical models and for studying the aetiology of glioma. Often, mice are engineered to mimic key driver mutations including loss of function of tumour suppressor gene such as Cdnk2a (*Ink4A/Arf), Nf1, Trp53, Pten* and *Rb1,* combined with expression of key oncogenes such as Ras, although this is not typically mutated in GBM. In this thesis, the mouse tumour initiating cell line referred to as IENS was used: this line was derived from the retroviral introduction of constitutively active mutant EGFRvIII to NSCs derived from, *cdkn2a* encoding, $p16^{Ink4A}$*/Arf* deficient mice[80]. Genetically engineered mouse models have a key function but are expensive and time consuming to breed and maintain, particularly when compound mutations are required. A complementary approach is to engineer the glioma driver mutations into normal NSCs.

CRISPR/Cas9 mediated gene editing allows for the rapid and precise editing of both human and mouse NSCs, both *in vitro* and *in vivo*[81]. Combined with transgenic methods, such as piggyBac integration, multiple modifications can therefore be engineered simultaneously, rapidly enabling modelling of different genetic subtypes.

One novel example used within this thesis is the mouse GSC model referred to as NPE (Pollard lab, in preparation). Using sgRNAs against the corresponding genes, *Nf1* and *Pten* were ablated, while transgenic constitutively human active EGFR (EGFRvIII) is overexpressed using transposase (piggyBac) mediated integration. This model most closely recapitulates the mesenchymal subtype of GSCs and models simultaneous activation of PI3K and MAPK signalling pathways, without interference with p53, thereby limiting genome instability. Models generated through this manner can be produced quickly and enable comparisons of isogenic pairs of mutant and parental cells (Figure 1.2).

1.3 The core transcriptional network of GSCs

GSCs, like NSCs, have a core transcriptional network that is responsible for their cell type identity. The induction of four core transcription factors (TFs), POUF3F2, SOX2, SALL2 and OLIG2, has been shown to be sufficient to reprogram differentiated glioma cells into GSCs capable of *in vivo* tumour formation[82]. These factors are all aberrantly upregulated in GBM[83], across a range of genetic subtypes.

SOX2 and OLIG2, in combination with ZEB1 can also transform tumour suppressor deficient astrocytes into glioma initiating cells, even in the absence of PI3K/RTK overexpression[84]. In line with the ability of SOX2 to reprogram mature cell types[85], overexpression of SOX2 cooperates with FOXG1 to dedifferentiate astrocytes, mediating acquisition of NSC like characteristics and re-entry to a proliferative state[86]. In summary, GSCs express a wide number of core neurodevelopmental, 'master regulators' and these seem to be critical for their sustained self-renewal. SOX2 is a key component of this gene regulatory network.

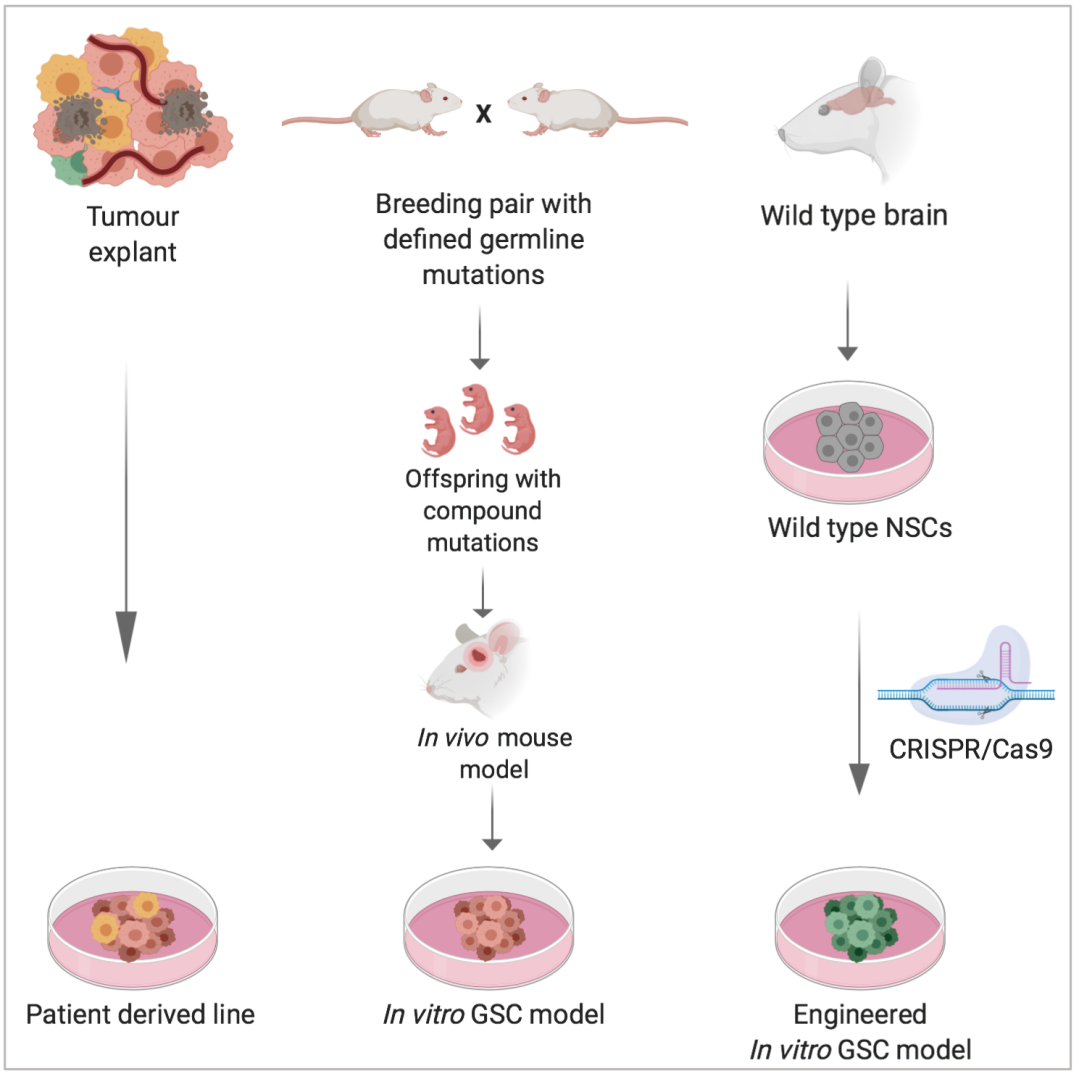


Figure 1.2 | Review of commonly used GBM models

Figure depicting common GBM models. GBM patient tumour samples can be used to derive patient lines, which retain the heterogeneity reflected within the original patient sample. Breeding mouse models can be used for both *in vivo* studies and may also be used to derive *in vitro* models. *In vitro* models of GBM may also be obtained through CRISPR/Cas9 mediated engineering of normal NSCs. This figure was generated using Biorender.

1.3.1 Sry-related HMG box transcription factor, SOX2

In mice and humans, ~20 *Sox* (SRY-related HMG-box) family members have been identified, based on the presence of a highly conserved high-mobility group (HMG) domain which mediates DNA binding[87]. *Sox* members are divided, based on HMG sequence identity, into the groups A-H, with factors from different groups having distinct biological functions in a wide range of stem cells and tissues[88]. The SoxE family, composed of *Sox8, Sox9* and *Sox10*, is capable of forming homo- and hetero-dimers with members of its family. Sox8, Sox9 and Sox10 are all involved in neural crest specification[89], with Sox10 playing a role in the maintenance of migratory neural crest stem cells within mice[90]. Sox9 is also involved in: chondrogenesis[91]; sex determination, with evidence also indicating a role for Sox8 in Sertoli cell function[92–95]; pancreatic progenitor maintenance[96], and neural stem cell specification and maintenance[97–99]. This versatility of SOX9 is achieved through a combination of post-translational modifications and varying binding partners. For example, phosphorylation of SOX9, by protein kinase A, can enhance DNA binding affinity and cause SOX9 translocation into the nucleus in testis cells, while in neural crest cells phosphorylation is required for SOX9-SNAIL interaction. The varying binding partners of SOX9 are relatively well studied with many examples existing, including SOX9 interacting with NFIA during glial initiation in the CNS, whilst during sex determination SOX9 can either homodimerize to regulate prostaglandin D synthase or SOX9 can interact with SF1 to upregulate anti-Mullerian hormone[100]. The SOXB1 family is composed of *Sox1, Sox2* and *Sox3*. Sox1 has been shown to play a role in lens development[101], and all of the SOXB1 family are involved in the specification and maintenance of neural progenitor cells[102]. Sox3 has also been shown to be involved in the progression spermatogenesis[103]. Of particular interest to this thesis is the SOXB1 family member *Sox2* (Fig 1.3) which has key roles throughout development, in adult tissue homeostasis and in disease progression[104].

Sox2 is initially expressed in the inner cell mass (ICM) and trophectoderm of blastocysts, with the zygotic deletion of *Sox2* resulting in a failure to form the pluripotent epiblast and early embryonic lethality[105]. Supporting the key role of *Sox2* in the establishment and maintenance of pluripotency, embryonic stem cells lacking SOX2 inappropriately differentiate into trophectoderm-like cells[106], whilst the forced expression of *Sox2*, *Oct4*, *Klf4* and *c-Myc* enables conversion of fibroblasts to induced pluripotent stem cells[85,107].

After exit from pluripotency, *Sox2* plays critical roles in foetal development of several endodermal tissues, ectodermal tissues and foetal germ cells[104]. In development, *Sox2* regulates early neural specification[108] and continues to play major roles, with other SOXB1 members, in the developing CNS and PNS, controlling the proliferation and differentiation of foetal progenitor cells[109,110]. Generally, overexpression of SOXB1 members (*Sox1, Sox2* and *Sox3*) promotes NPC proliferation, whilst the depletion of SOX2 causes exit from the cell cycle and premature differentiation[102,111].

In the adult, SOX2 has been found to be expressed in within cells of multiple tissues, including the stomach, cervix, anus, testes, lens and multiple glands, with several of the tissues that require *Sox2* in development containing *Sox2* expressing adult stem cells/progenitors. Lineage tracing experiments have shown these SOX2+ adult stem cells originate from foetal SOX2+ tissue progenitors and are required for tissue homeostasis and regeneration[104,112]. *Sox2* also marks adult NSCs[63,113] found within neurogenic regions of the adult brain, where it is required for the maintenance of NSCs properties and functions[114–116]. Consistent with the core function of *Sox2* in NSCs/NPCs, overexpression of *Sox2* alone is sufficient to generate induced NSCs (iNSCs) from mouse embryonic fibroblasts[117].

SOX2 is genetically amplified in ~4% GBM[16], and SOX2 is often highly expressed in GBM[118], with upregulated expression observed in patient derived xenografts irrespective of the oncogenic driver[84]. Reflecting *Sox2*'s function in normal NSCs, *Sox2* is key for GSCs migration/invasion[119], *in vivo* tumour formation and proliferation[84,119–121], with a recently published genome-wide CRISPR viability screen identifying SOX family members as top hits[122].

Despite its potential importance as a therapeutic target, the lack of enzymatic sites makes SOX2 difficult to 'drug' therapeutically using small molecules. SOX2's activity is controlled in multiple ways; through control of expression including post-transcriptionally by miRNAs[123], co-partner expression and post-translational modifications (PTMs), offering potential alternative SOX2 targeting routes. Several PTMs, including phosphorylation, methylation, acetylation, SUMOylation, ubiquitination, PARPylation and O-GlcNAcylation, have been identified to occur on SOX2. The function of many of these PTMs is still being eluted, but evidence shows they can mediate effects on protein turnover, nucleo-cytoplasmic transport, DNA-binding and protein interactions (Figure 1.3)[124,125]. As SOX proteins general exhibit gene regulatory functions through complexes, with target genes often having partner protein binding sites adjacent to functional SOX-binding sites, the protein interactors of SOX2 effect activity of SOX2 and contribute to the activation of alternative transcriptional targets[126,127]. For example, the *UTF1* enhancer[128] and the *Fgf4* enhancer are activated by SOX2-POU5F1/OCT4 heterodimers, but not SOX2 alone, in ESCs[129], while the chicken δ -crystallin gene minimal enhancer, DC5, is activated by SOX2 and PAX6 during lens development[130]. Understanding the transcriptional targets and SOX2 protein interaction partners within GBM may provide novel alternative strategies to disrupt SOX2's function. Recent integration of genetic and proteomic approaches led to the identification of CHD7 as a functional interactor of

SOX2 in mouse NSCs, highlighting the capability of this approach[131]. Through proteomic approaches the Pollard laboratory has recently identified MYEF2 (Myelin expression factor 2) to be an interactor of SOX2 and understanding the function of MYEF2 in GBM will form the basis of a subsection of this thesis.

MYEF2 is a ~65kDa paralogue of hnRNPM with known ssDNA[132] and RNA binding[133,134] properties, most likely conferred through its three RRM domains. A N-terminal truncated version of MYEF2 was identified to be capable of binding to a single stranded probe generated from the myelin basic protein (*mbp*) proximal promoter region MB1, repressing expression from the myelin basic protein (MBP) promoter in non-glial cells but not in glial cells[132]. No validation of the full-length protein has been performed. MYEF2 has also been identified to form a complex with RUNX1 in mouse erythroleukemia cells, contributing to the repression of several RUNX1 targets and morpholino knockdown of *myef2* in zebrafish leads to a reduction in the numbers of HSCs, highlighting a role for myef2 in the regulation of haematopoiesis[135]. Intriguingly, a link between sox2 and myef2 has previously been identified within mouse squamous skin tumours. Whilst SOX2 is absent from the normal epidermis, SOX2 is the most upregulated transcription factor in the cancer stem cells (CSCs) of mouse squamous skin tumours. Within the SOX2 expressing CSC population *myef2* levels increased while knockdown of SOX2 in pre-existing tumour cells led to a reduction of *myef2,* with SOX2 ChIP identifying *myef2* as a direct target of SOX2 in squamous skin CSCs. It should be noted that SOX2 ChIP in glioblastoma did not identify *myef2* as a direct SOX2 target and KD of SOX2 did not lead to a reduction in *myef2* in one human glioblastoma sample[136], highlighting the differential regulation of myef2 within different cell types.

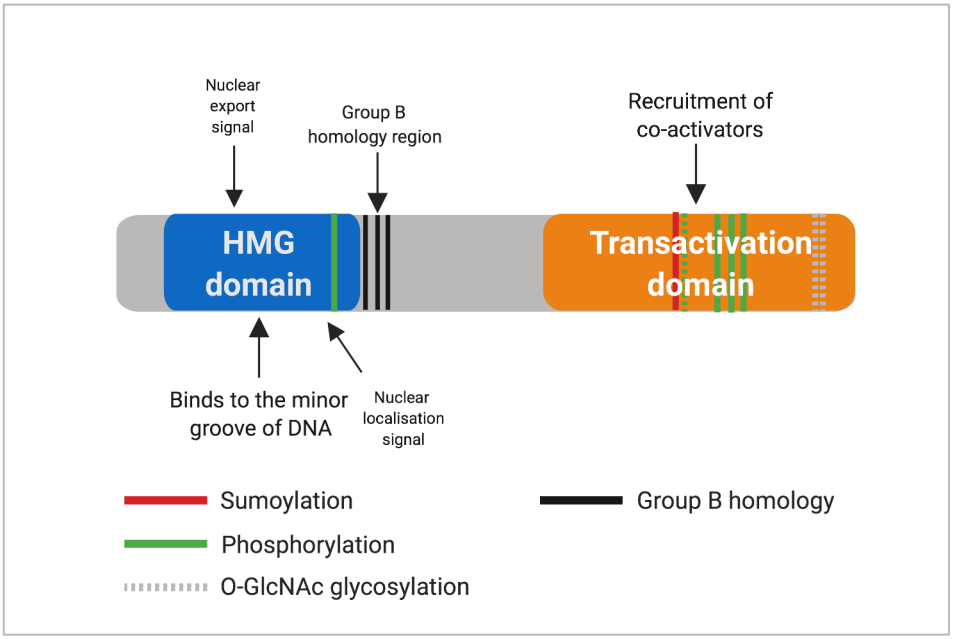


Figure 1.3 | SOX2 and post-translational modifications

SOX2 (SRY-related-HMG-box 2) contains an HMG domain and a C-terminal transactivation domain, as depicted. SOX2 can undergo a range of post-translational modifications including sumoylation (red), phosphorylation (green) and O-GlcNAc glycosylation[137] (grey, dotted). Other modifications such as acetylation and methylation can also occur. O-GlcNAcylation of SOX2 on Ser248,Thr258 and Ser259 (mouse) has been observed to occur in mESCs[137] and the rat forebrain[138], and has been shown to alter SOX2's protein interactions and genomic occupancy[139].AKT-dependant phosphorylation at Thr118 (mouse)/THR116 (human) is associated with increased stability[140], while phosphorylation at residues SER249-251(human)[141] in hESCs, leads to the sumoylation of residue Lys245 (human), which is associated with inhibition of transcriptional activation[142]. Acetylation of residue Lys75 (mouse) within the nuclear export sequence of SOX2 effecting subcellular localisation and protein decay[143]. Methylation of Arg113 (mouse), has been proposed to effect SOX2 homo-dimerisation[144] and methylation of Lys119 (mouse) can effect SOX2 ubiquitination and hence decay of SOX2[145,146]. Figure was adapted from Kamachi and Kondoh, 2013[126].

1.4 Genome editing

Genome editing is the site-specific deletion, insertion or replacement of DNA within an organism. The first examples of targeted genomic editing came from targeted gene disruption studies in yeast[147] and subsequently mammalian cells[148,149]. In these studies, the directed integration of the exogenously provided DNA was dependant on homologous recombination (HR). Despite this major advance, this approach had several limitations including, low efficiency of targeted integration[150] and random integration of the exogenous DNA, which could occur at a frequency higher than the targeted events[151]. Introduction of a double stranded break (DSB) within the DNA of the target site, promotes HR-directed DNA repair (HDR), improving the efficiency of targeted integration by orders of magnitude[152,153]. DSBs can also be repaired through the alterative pathway of non-homologous end joining (NHEJ). NHEJ is an error prone repair pathway which can result in the insertion and deletions of nucleotides at the break site (Figure 1.4).

The first enzymes used to induce DSBs were the rare cutting endonucleases, such as the 18bp cutter I-SceI[153]. These meganucleases recognise long stretched of DNA (14-40bp), hence the probability of a recognition site occurring at the desired genomic location is low. Overcoming this, fusion of DNA recognising zinc finger proteins and transcription activator-like effector (TALE) proteins, to the DNA cleavage domain of Fok I endonuclease, led to the generation of zinc finger nucleases (ZFNs)[154] and TALENs[155], respectively. As Fok I functions as a homodimer two separate ZFN or TALENs are required mediate a DSB. Each ZFN recognises a 3-bp DNA sequence[156], while each TALEN recognises a single base pair[157,158]. Combinations of zinc fingers and TALENs can be combinatorially assembled to target unique genomic sequences. These 'designer nucleases' successfully increase genome editing efficiency[159–162].

However, targeting of new genomic locations requires re-design and synthesis of extensive and repetitive structures. This protein engineering is burdensome and difficult for non-experts.

1.4.1 Discovery of CRISPR/Cas9

Due to its flexibility and simplicity, the development of CRISPR/Cas editing technology has changed the field of genome editing. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)[163] and their CRISPR associated (Cas) genes[164] form a prokaryotic adaptive immune mechanism[165]. CRISPR repeat clusters are separated by non-repeating spacer DNA sequences. These spacers are derived from DNA of the invading pathogen and dictate the targeting specificity of Cas enzymes, which mediate the introduction of DSBs into the pathogen's DNA upon reinfection.

Several types of CRISPR/Cas systems have been reported across different bacteria and archaea[166]. All systems require a short CRISPR RNA (crRNA), transcribed from the spacer sequence[167], and a transactivating crRNA (tracrRNA)[168] to form an active Cas protein-RNA complex. A single chimeric RNA, termed single guide RNA (sgRNA), formed by the fusion of crRNA and tracrRNA may also be used[169]. The crRNA acts as the guiding sequence, hybridising to the target DNA sequence, which is positioned next to a protospacer adjacent motif (PAM)[170], a critical component of the system that mediates the distinction of self and non-self DNA[171].

Currently, the most commonly used CRISPR/Cas system for genome editing is CRISPR/Cas9, a type II CRISPR system from *Streptococcus pyogenes.* This system has a single effector protein (SpCas9), a 20nt long crRNA and simple NGG PAM requirement, with the cut site generated ~3bp 5' of the PAM[172]. Following proof of

principle *in vitro* experiments[169], the CRISPR/Cas9 system was successfully used for mammalian genomic editing[173–175] and continues to be extensively used.

Compared to previously used ZFNs and TALENs, the CRISPR/Cas9 system is simple, cost effective and efficient. By altering the sequence of the crRNA, Cas9 can be programmed to new target sites with the genome, without the burdensome design and synthesis associated with ZFNs and TALENs. The simple NGG PAM requirements of the CRISPR/Cas9 allow for appropriate targets to be identified at a high density throughout the genome[176], although this is obviously not as flexible as TALENs which in theory should be capable of targeting any genomic region. In attempts to expand the application of CRISPR researches are still exploring other CRISPR systems, identifying systems which varying in their PAM requirements, substrate preferences and effector protein size[177]. An overview of the versatile uses of CRISPR/Cas technology are discussed below (Figure 1.5).

1.4.2 CRISPR/Cas9 for gene knockouts

The co-delivery of Cas9 with an sgRNA targeting an exonic region within the gene of interest, is the simplest method to knockout the protein of interest. Indels introduced by the NHEJ mediated repair of the DSB can disrupt the reading frame, leading to generation of a non-functional protein product[174]. Alternatively, two guide RNAs can be used to target two sites in a gene, leading to a deletion of the intervening sequence, which can be up to a megabase in size[178]. The high efficiency of this method, coupled with the easy large scale generation of multiple sgRNAs, have subsequently allowed for CRISPR/Cas9 mediated high-throughput loss of function screens[179–181].

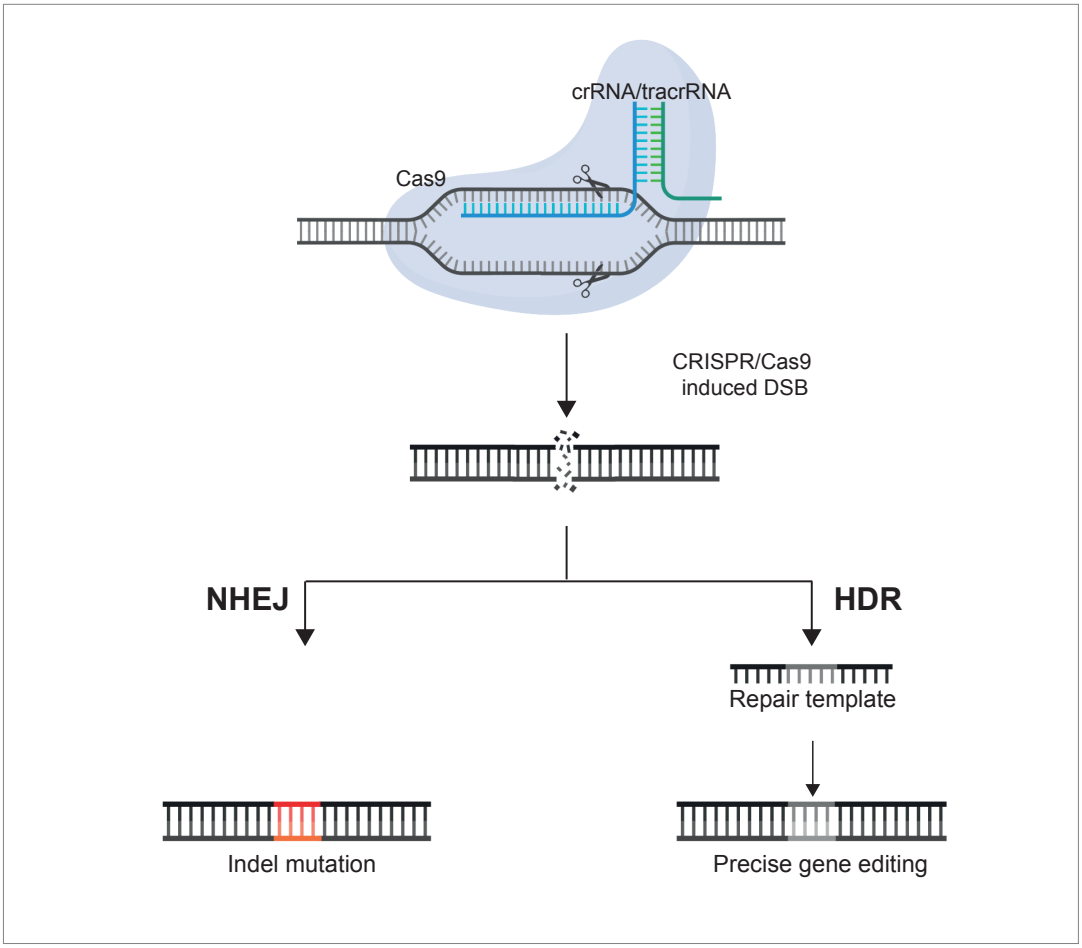


Figure 1.4 | Site specific DSB facilitates precise genome editing

Programmable nucleases, such as Cas9, are recruited to specific regions of the genome where they mediate the generation of double strand breaks (DSBs). These DSBs can be repaired through two pathways: NHEJ (non-homologous end joining) or HDR (homology directed repair). In NHEJ, the ends of the DBS are process and re-joined. This can result in random indel mutations at the junction site. In the presence of a DNA sequence with homology to the damage region, HDR can seal the DSB in an error free manner and can be used for precise gene editing.

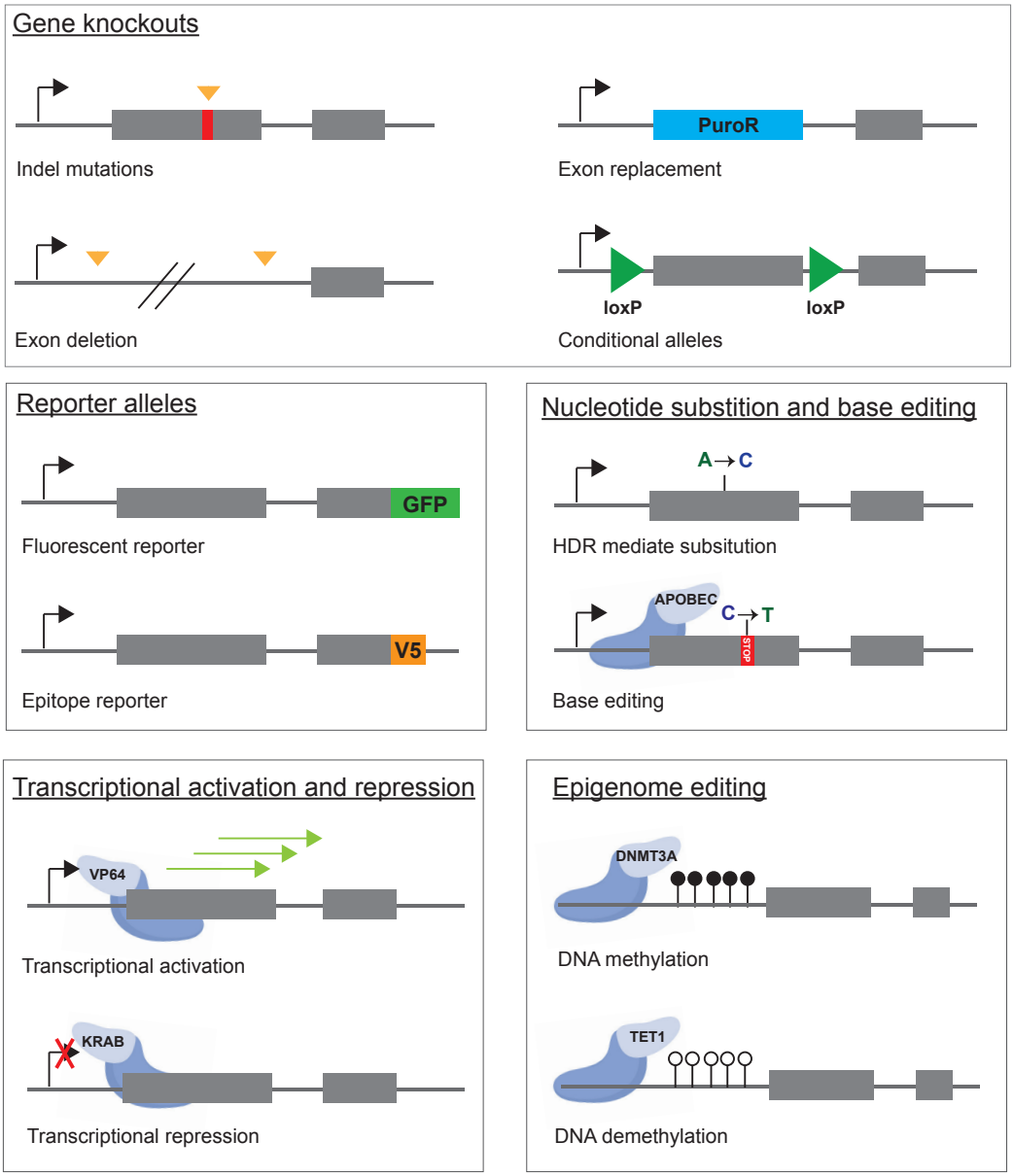


Figure 1.5 | The CRISPR/Cas "tool kit"

Depiction of some of the many uses of CRISPR/Cas for genome editing and beyond. CRISPR/Cas can be used for the generation of genetic knockouts through multiple methods including NHEJ mediated indel formation, exon drop out, exon replacement or the insertion of recombination sites. CRISPR/Cas can also be used to insert fluorescent tags or epitope tags to endogenous genes. Besides genome editing, the fusion of Cas9 mutants to a range of proteins can mediate a multitude of outcomes including base editing, transcriptional control and epigenetic modification.

Alternatively, the Cas9/sgRNA can be delivered in the presence of exogenous DNA with homology to the targeted region. Upon DSB induction, HR-directed repair may occur, integrating the exogenously supplied DNA. This can allow for the replacement of a specific region, generating non-functional protein and a method to select for targeted cells. CRISPR/Cas9 may also be used to integrate flanking loxP or FRT sites, allowing for the generation of Cre-LoxP or Flp-FRP conditional knockout alleles[182–184]

1.4.3 CRISPR/Cas9 for gene knock-ins of tags and reporters

HR-directed repair, following CRISPR/Cas9 generated DSBs, can be used to introduce a wide variety of sequences to protein-encoding genes. Proteins of interest can be engineered at the N-terminal, C-terminal or at internal regions such as within flexible loops. CRISPR/Cas9 can be utilised to insert large DNA sequences encoding fluorescent markers, such as GFP or mCherry[81], or for the insertion of small epitope tags, such as HA, FLAG or V5[185]. The generated in-frame fusion proteins can be used for biochemical studies and as an alternative for low quality or unavailable antibodies. Fluorophores may also be used for real-time observations, such as live-cell imaging or live isolation of specific cell populations.

Other uses of CRISPR/Cas9 mediated knock-ins include HDR mediated insertion of point mutations within the coding sequence. These can be disease specific mutations, expanding the availability of disease models *in vitro* and *in vivo*[186]. More recently, there has been interest in base editing. A ‘nickase’ Cas9 (nCas9) can be fused to APOBEC1 deaminase enzyme and Uracyl Glycosylase inhibitor (UGI), mediating the conversion of cytosine (C) to Thymine (T)[187]. These base editing tools can be used to introduce single base pair mutations, including generating the introduction of stop codons in method known as CRISPR-STOP[188].

1.4.4 Transcriptional and epigenome editing with CRISPR/Cas9

Catalytically inactive Cas9 (dCas9), cannot cleave DNA but still binds to its target sequence[189]. In fact, the strong binding of dCas9 to DNA can interfere with the binding of DNA binding proteins such as RNA polymerase II. This can block transcription and lead to a knockdown in protein expression in a method now referred to as CRISPRi[189]. The fusion of a repressor complex, such as Kruppel-associated Box (KRAB) enhances this effect[190]. In the reverse manner, fusion of dCas9 or nCas9 with strong transcriptional activators, such as VP64, can induce transcription from the target gene[191–193].

dCas9 fusion proteins may also be exploited to introduce epigenetic modifications to specific regions of the genome. dCas9 fused to the catalytic domain of DNA methyltransferase 3A (DNMT3A) is capable of depositing DNA methylation at the targeted locus. This can recruit additional repressive components, such as DNMT3L, causing gene suppression[194,195]. By contrast, dCas9 can be fused to the catalytic domain of methyl cytosine dioxygenase TET1, resulting in the demethylation of the targeted sequence and reversal of gene silencing[196].

1.4.5 Targeting RNA

It has recently been shown that the CRISPR/Cas9 system can be used to bind or cleave specific RNAs. Through the addition of a PAM presenting oligonucleotide (PAMmer), Cas9 can be directed to ssDNA. If the PAMmer sequence is designed for RNA sequences that lack PAMs at the corresponding genomic site, Cas9 will specifically target RNA. This RNA-Cas9 targeting system is referred to as RCas9[197]. RCas9 can be used in manners similar to those discussed above including cleavage

of RNA and RNA visualisation[198] and is being increasingly used to explore the RNA epigenome.

1.4.6 Strategies to deliver CRISPR/Cas9 to cells

As only one CRISPR/Cas9 complex is required to generate a DSB, compare to the two required for ZFNs and TALENs, CRISPR/Cas is considered to be more susceptible to off-target effects. This has led to the generation of several design tools to predict such off-target sites (such as GUIDE-seq, BLESS, HTGTS, and Digenome-seq). However, one of the easiest methods to reduce off-targets and increase the efficiency of targeting is to alter the exact components delivered. For example: The catalytic domain of SpCas9 can be mutated to produce a 'nicking' enzyme, nCas9[174,199,200]. Unlike wild-type spCas9, nCas9 requires the use of 2 guide RNAs to introduce nicks on both strands at a proximal distance. This significantly reduces off target effects.

Most commonly, Cas9 is delivered into cells in either an expression plasmid-based manner or as a purified protein pre-complexed with crRNA/tracrRNA; this is referred to as an RNP (ribonucleoprotein). Delivery of Cas9 as an RNP results in more transient Cas9 activity, compared to the more prolonged expression observed when delivered as a plasmid. This more transient nature of the RNP leads to fewer off-target effects[201,202], though may not be suitable if long term expression of Cas9 is required. Delivery of RNP is also associated with lower toxicity in primary cells and the efficiency of targeting can be very high[202,185,203].

To mediate HR-directed repair, a donor template with a region of homology to the target is required. Typically, the repair template is provided as a plasmid, a dsDNA block or as ssODNs (single stranded oligonucleotides). However for primary cells

which are difficult to transfect the donor template may be delivered in the form of viral vectors, such as adeno-associated virus[204]. Synthetic ssODN are easy to obtain, though currently limited in size, and are not associated with plasmid related toxicities. Recently, it has been shown to be possible to generate long ssODNs using nickase enzymes expanding their potential use[184].

In summary, CRISPR/Cas9 has transformed the genetic analysis of human cancer. The possibilities to engineer specific mutations in patient derived cells or primary stem cells, and the ability to generate new tools for use in cell biology, imaging, biochemical and genomics methods, is opening up a wealth of new discoveries.

1.5 Degron mediated post-translational control of protein levels

Loss of function experiments are extremely useful in aiding the characterisation of the biological functions of a protein. To achieve these loss of function studies, genes and the corresponding proteins they encode can be disrupted at either the DNA, RNA or protein level. Traditional DNA modifying methods facilitate the generation of gene knockouts mediating the constitutive loss of encoded protein[205]. The generation of gene knockouts has been greatly eased by the emergence of genome editing technologies such as TALEN, Zn finger nucleases and in particular CRISPR/Cas9. Gene knockout experiments can be problematic when studying the function of proteins which are essential to cellular homeostasis, proliferation or during specific developmental periods. To overcome this, conditional genetic knockouts can be generated through the use of recombinases such as Cre. However, a relatively long time is required to see protein depletion, and once induced these genetic changes are permanent. To overcome some of these caveats, protein expression can be controlled reversibly at the DNA, RNA and protein level.

CRISPRi[189], as discussed previously, is one method of controlling expression conditionally and reversibly at the pre-transcriptional level. Catalytically dead dCas9 co-expressed with sgRNA, can prevent the initiation of transcription from the targeted locus. This system can be inducible through controlled expression of dCas9 or sgRNA, although the method is limited by the requirement to have a PAM at the right site.

At the RNA level siRNA/shRNA can be used. However these methods are associated with off-target effects and the efficiency of siRNAs mediated knockdown can be highly variable[206]**.** Delivery and stability of siRNA is also an issue for whole animal studies[207].

All the methods described above involve targeting DNA and RNA. This meditates control at the pre-translational level; hence a long time is required before a loss of protein is observed and does not realistically model the goal of a successful small molecule or biological therapeutic, which would knockdown or ablate protein activity/function acutely. For example, siRNA requires 2-3 days to issue an effect, and the exact rate relies on the half-life of the protein studied. This can be an issue if rapid degradation is required, for example if studying cell cycle proteins. This issue can be overcome by mediating control at the post-translation level. To achieve this, investigators are exploiting sequences known as degrons, which can be tethered to the protein of interest (POI)[208].

1.5.1 What is a degron?

Within eukaryotic cells the levels of proteins can be controlled through degradation by the 26S proteasome or the lysosome. This degradation is highly selective, and control is mediated through degradation signals knowns as degrons. Degrons can be generally classified into two categories: acquired or inherent. Acquired degrons are

generally transient, and mediated through the addition of post-translation modifications, most frequently phosphorylation, to a protein. Acquired degrons are most commonly associated with controlling the levels of regulatory proteins, for example the rapid temporal degradation of cyclins/proteins through the cell cycle[209].

Inherent degrons are an internal feature of the protein, and often associated with the protein quality control system which exists to degrade misfolded or damaged proteins[210]. Internal degrons can be composed of hydrophobic stretches of amino acids which are normally buried within the protein but are exposed when proteins fail to correctly fold or assemble into relevant protein complexes. Alternatively, they can be specific amino acid sequences such as those observed in the N-degron and C-degron pathways[211]. The first degron identified was an N-degron. The N-degron is composed of destabilising N-terminal residues, and in the case of eukaryotes an internal lysine substrate, which are recognised by E3 ubiquitin ligases and mediate 26S proteasomal degradation effecting the proteins half-life[212,213]. With destabilisation initially only thought to occur through N-terminal residues, destabilising C-terminal residues have been subsequently identified and are referred to as C-degrons[214].

A key feature of degrons is that they are transferable, with the engineering of a degron to a protein of interest (POI) conferring a shorter half-life[215]. The combination of the transferable nature of degrons and the identification of N-degrons, lead to generation of the first conditional degron[216]. This degron could be engineered to a POI allowing for post-translation protein control. Subsequently, more conditional degrons have been generated. Combined with CRISPR/Cas9 mediated ease of genetic engineering, it is now possible to achieve the rapid, reversible controlled protein degradation using conditional degrons controlled by varying features including light, temperature, and small molecules as discussed below (Figure 1.6).

1.5.2 Temperature controlled degron

The first conditional degron was generated by the Varshavsky laboratory[216]. Using their knowledge of N-degrons the temperature sensitive dihydrofolate reductase, Arg-DHFRts, as fused to Ura3. Arg-DHFRts is a variant of DHFR in which the wild type N-terminal Valine is replaced by the destabilising amino acid Arginine (Arg). At 23^{o}C degree, in spite of a destabilising N residue, Arg-DHFRts is long lived. Increasing the temperature to 37^{o}C causes unfolding of Arg-DHFRts, with the increased exposure of destabilising Arg and/or increased mobility of its internal lysine residue, resulting in rapid ubiquitination and 26S proteasome mediated degradation of the Arg-DHFRts-Ura3 fusion protein. It should be noted that this system is only suitable for N-terminal tagging and its use has been limited to budding yeast and fission yeast due to the temperature sensitivity of plants (too high) and mammalian cells (too low).

1.5.3 Light controlled degrons

Phototrophins, a group of light activated kinases, allow plants to respond to external light cues. The light oxygen voltage 2 (LOV2) domain of phototropins contain a flavin mononucleotide binding core domain and a C-terminal Jα helix. Upon irradiation with blue light, a cysteine of core domain forms a covalent adduct with the excited flavin mononucleotides, inducing dissociation and unfolding of the Jα helix[217]. Fusion of the LOV2 domain with cODC, the C-terminal ubiquitin independent degron of ornitihine decarboxylase, led to the generation of a photosensitive degron (PSD). The PSD is stable in the dark, but upon irradiation with blue light, conformational changes within the LOV2 domain unmask the cODC degron, hence causing degradation in response to blue light[218]. Another light sensitive degron, B-LID, which consists of LOV2 - RRRG degron fusion protein has also been shown to be activated by blue light[219]. These photosensitive degrons offer short POI depletion times ($t_{1/2}$=30 min), and both

temporal and, unlike other degrons, spatial control of POI degradation as exposure to light can be restricted. It should be noted that the use of these degrons requires specialised illumination devices.

1.5.4 Small ligand/ molecule controlled degrons

This is the largest group of conditional degrons. Many have been identified which can be used on the N- or C-terminal and avoid the use of heat shock or expensive illuminators. This type of degron is often relatively easy and cheap for *in vitro* studies, hence is extensively used. Below we discuss three commonly used degrons.

Perhaps, one of the best known conditional degrons is the auxin inducible degron (AID), which is based on the plant auxin dependant degradation pathway. The auxin family hormones, such as IAA (naturally occurring auxin) and NAA (synthetic), bind to the F-box transport inhibitor response 1 (TIR1) protein, promoting the interaction of the E3 ubiquitin ligase SCF-TIR1(a form of SCF complex containing TIR1) and the degron domain of target auxin/IAA transcription repressor proteins. SCF-TIR1 recruits an E2 ubiquitin conjugating enzyme which leads to the ubiquitylation and subsequent proteasome mediated degradation of AUX/IAA protein. The SCF complex is highly conserved among eukaryotes, allowing plant TIR1 to form complex with non-plant SCF. Orthologues of TIR1 and AUX/IAA on the other hand, are only found in plant species. Fusion of the AID degron, derived from the auxin dependant degron sequence of IAA17, to the POI leads to rapid auxin-dependant degradation ($t_{1/2}$=30min) in all non-plant eukaryotes which have been engineered to express high levels of TIR1[220].

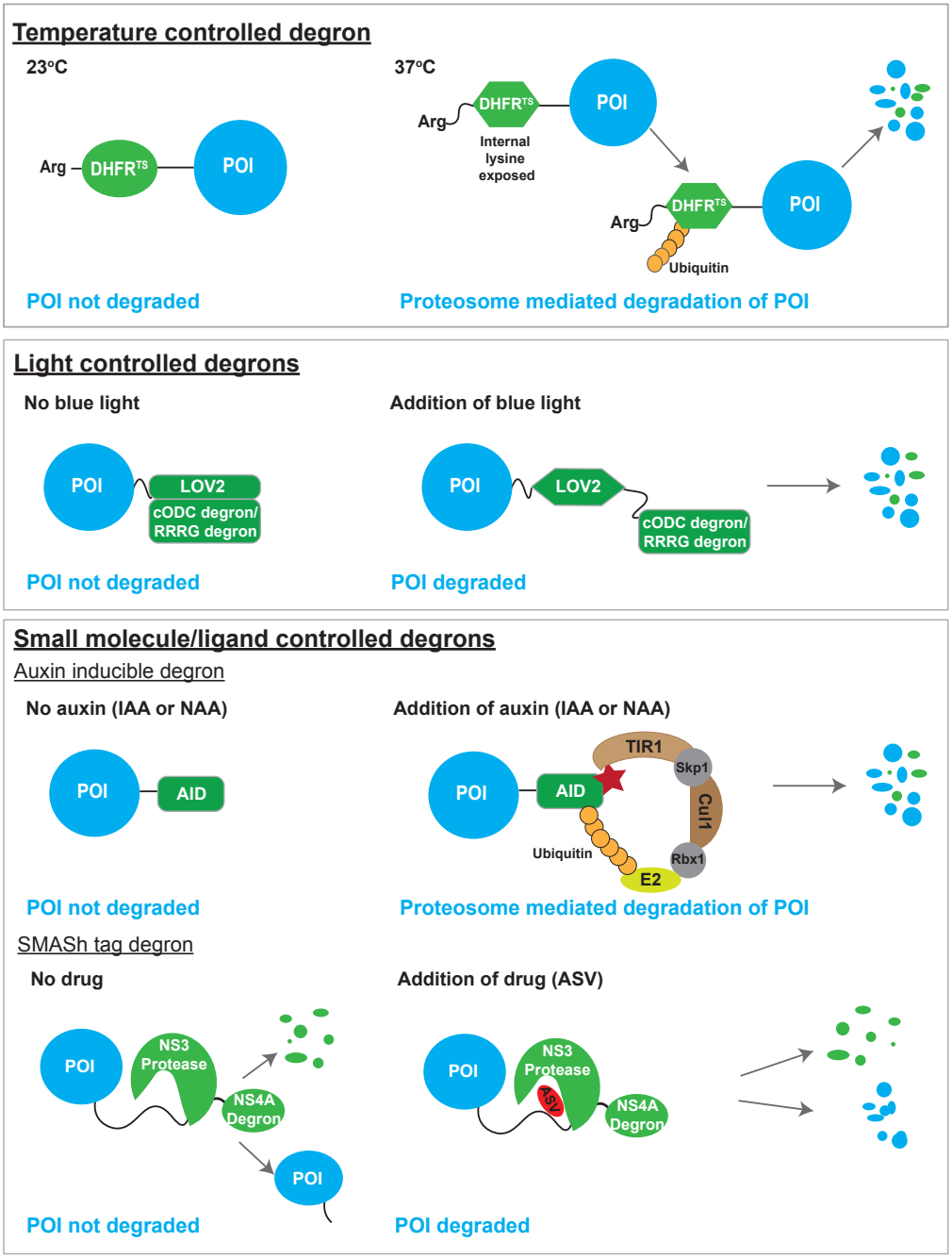


Figure 1.6 | Commonly used conditional degrons

This figure depicts several of the most commonly used conditional degrons, which when fused to the protein of interest (POI) allow for post-translational control. These degrons can be controlled through varying factors including: Temperature (ts-degrons); blue light, as is the case for degrons PSD and B-LID, and small molecules which control many degrons including AID and SMASh.

The HaloTag, derived from bacterial dehalogenase, is a highly versatile tag forms specific covalent bonds with compounds that contain alkyl chloride, allowing its use for a range of methods including cell imaging and protein purification[221]. Whilst the Halo tag itself is not a degron the covalent bonding between the HaloTag and specific binding partners can generate a conditional degron. One such example is HyT13, a small hydrophobic molecule which covalently binds to the HaloTag, and mimics a denatured or misfolded protein mediating the decay of HaloTag and fused POI through the protein quality control system[222]. Another example is through the use of a small molecule ligand of VHL, which when bound to the HaloTag mediates the recruitment of E3 ligase, VHL, and subsequent degradation by the proteasome. This degron is known as a HaloPROTAC[223]. It should be noted that one disadvantage of these degrons is that the rate of degradation is relatively slow compared to other degrons such as AID.

Use of these degrons requires the fusion of a large tag which is retained on the POI. This may affect the normal function of the POI. Secondly, the use of AID requires dual engineering, and due to the relatively high amounts of auxin required to induce degradation, the AID system is problematic when it comes to multicellular organisms. The small molecule-assisted shut off SMASh tag[224] is a self-cleaving degron, composed of a hepatitis C virus (HCV) NS3 protease, a HCV NS3 cleavage sequence and HCV NS4A degron. The exact mechanism of the NS4A degron is unknown but it has been proposed that that the hydrophobic NS4A N-terminal, which in the native protein normally inserts into the ER, is incapable of ER insertion in the modified SMASh peptide, hence is free and acts as a degron. When fused to the POI the SMASh tag normally self-cleaves, and only the SMASh tag is degraded. Addition of HCV NS3 protease inhibitors, such as asunaprevir (ASV), prevent the self-cleavage of the SMASh tag, mediating degradation of both the SMASh tag and the POI. This

degradation is dose dependant on FDA approved drugs and can therefore potentially be used *in vivo*. As the SMASh tag is cleaved in the absence of ASV there is a build-up of nontagged, hence non-degron controlled, protein within the cell. This means that the rate of degradation is slow as, unlike other degrons discussed, is depend on the half-life of the POI. Offsetting this disadvantage, is the major advantage that the POI, in the absence of ASV, is minimally modified and unlikely to affect normal POI function.

In summary, advancements in genome engineering allow for precise editing, rapidly and efficiently within multiple organisms. This has opened up the use of degrons to post-translationally control protein levels. Many different degrons exist, each with their own advantages and disadvantages. Whilst the use of degrons do not replace genetic knockouts, they are an excellent alternative to commonly used knockdown methods, such as the use of siRNA, both *in vitro* and *in vivo.*

1.6 Aims

In summary, our goal is to exploit engineered mouse GSC models and the latest genome editing tools to dissect the role of SOX2 and its candidate partner MYEF2.

There were two primary aims:

1. Validate and dissect the function of the MYEF2/SOX2 interaction within NSCs and GSCs.
2. Assess the feasibility of the SMASh degron system to control transgenic and endogenous SOX2 protein levels with NSCs.

Chapter 2 Materials and Methods

2.1 Cell culture

2.1.1 Cell line derivation

Mouse NSC lines were derived from the subventricular zone of adult male C57BL/6J mice in a method similar to previously described[225]. The dissected tissue was collected in PBS on ice and mechanically dissociated by trituration. A second dissociation in accutase (Sigma) was performed for 5 min at 37°C. The sample was transferred to NSC wash media, filtered in a 50 µm strainer and centrifuged at 400g for 5 min. The cell pellet was then resuspended in NSC maintenance media and plated onto uncoated tissue culture plastic. This was performed by Maria Angeles Marques (Pollard lab).

For the generation of mouse GSC (NPE) lines, NSCs were transfected with two sgRNAs against *Nf1* and *Pten*. Simultaneously, constitutively active hEGFRVIII and GFP-luciferase piggyBac constructs were also integrated into the genome. Cells were then sorted by FACS and selected for stable piggyBac integration with 4 days of treatment in 50 µg/ml hygromycin followed by 4 days in 5 µg/ml blasticidine. The bulk population was injected into an NSG mouse and a tumour was allowed to form. The tumour was dissected, tissue dissociated as above, and individual clones picked from the cell colonies formed. Clones were checked for the absence of NF1 and PTEN, and presence of hEGRVIII by western blotting and genotyping. (Ester Gangoso, Pollard lab, unpublished).

2.1.2 Cell line maintenance

Established lines were maintained at 37°C and 5% CO_2 in serum-free DMEM/F12 basal media supplemented as described below and in previous studies[81]. Cells were split, onto uncoated tissue culture plastic, at a ratio of 1:4 to 1:7 using accutase dissociation every 2-3 days. Cryopreservation of lines was performed using 10% DMSO and lines stored in -80°C or liquid nitrogen for short- or long-term storage respectively.

2.1.3 List of cell culture media

Name	Base media	Supplements (final conc.)
NSC maintenance media	DMEM/F12 (Sigma)	120 µg/ml BSA (Gibco) 100 U/ml Penicillin (Gibco) 100 µg/ml Streptomycin (Gibco) 1.5 mg/ml Glucose (Sigma) 1x NEAA (Gibco) 0.5 X N2 (LifeTech) 0.5 X B27 (LifeTech) 100 µM 2-merceptoethanol (Gibco) 10 ng/ml EGF (Peprotech) 10 ng/ml FGF2 (Peprotech) 2 µg/ml Laminin (Cultrex)
NSC wash media	DMEM/F12	150 µg/ml BSA 100 U/ml Pen 100 µg/ml Strep
Freezing media	NSC maintenance media	10% DMSO

Table 1: Cell culture media components

2.1.4 Differentiation of mouse NSCs

Neuronal differentiation

For neuronal differentiation cells were seeded at a density of 5000 cells per cm^2 in maintenance media with EGF removed for 3 days. At day 3 media was changed to maintenance media without EGF and FGF2 and cells differentiated for a further 4 days.

Astrocytic differentiation

For astrocytic differentiation, cells were seeded at a density of 3000 cells per cm^2 in maintenance media without EGF and FGF2, supplemented with hBMP4 (10 ng/ml, R&D) for 3 days. For serum induced differentiation cells were seeded at a density of 3000 cells per cm^2 in 10% foetal calf serum (Gibco) for 3 days.

Oligodendrocytic differentiation

For oligodendrocytic differentiation cells were seeded at a density of 4000 cells per cm^2 in maintenance media without EGF and supplemented with hPDGFαα (10 ng/ml, R&D) for 4 days. At day 4 media was changed to maintenance media without EGF or FGF2 and supplemented with 30 ng/ml T3 (Sigma) and 200 µM ascorbic acid (Sigma).

2.2 Derivation of genetically modified cell lines

2.2.1 Design and construction of sgRNAs and crRNAs

CRISPR gRNAs were designed using either MIT CRISPR design tool (http://crispr.mit.edu) or Desktop genetics (https://www.deskgen.com). For knock-in experiments gRNAs which cut, close to, but after the stop codon were selected. For knock-out experiments, gRNAs were designed to remove a section of the gene which contains a functional domain. Ideally knock-out guides are neither close to the start codon (in case of alternative start sites) nor at end of the coding sequence (to reduce

chance of obtaining a functional truncated protein). The potential off-targets of gRNAs were analysed with the WTSI Genome Editing tool (http://www.sanger.ac.uk/htgt/wge). To reduce potential off-target cleavage, gRNAs selected had 3 or mismatches between on-target sequence and similar sequences in the genome[226]. For the generation of sgRNA plasmids, used in section 3.2.4, single stranded oligos containing the 20 nt guide and 4 bp overhangs were ordered (Sigma). The oligos were annealed, phosphorylated and ligated into the BsaI sites of U6-sgRNA-SpCas9-2A-GFP (Addgene PX458[227]). For crRNAs, used in 3.2.2, 4.2.1, and 5.2.2, Alt-R® CRISPR-Cas9 crRNAs were ordered from IDT (Integrated DNA Technologies).

2.2.2 List of gRNAs used

Gene	Function	DNA sequence (5'-3')	PAM
Mouse *Myef2*	KI of mcherry	TGCGTAATTTCAAGCATGGT	TGG
Mouse *Myef2*	KI of mCherry and HA	GCAATGCGTAATTTCAAGCA	TGG
Mouse *Myef2*	KO exon2	ATTAAAAGGCTTGGACAGTA	TGG
Mouse *Myef2*	KO exon2	TTTGAGTCTGTGTGGCTAAG	AGG
Mouse *Sox2*	KI mcherry and mcherry smash	GTGAGGGCTGGACTGCGAAC	TGG

Table 2: List of gRNAs used

2.2.3 Designing and generating targeting vectors and piggyBac constructs

Myef2-mCherry targeting cassette was generated as described previously[81], by Maria Kalantzaki (Pollard lab). Briefly, 1kbp homology arms were PCR amplified from mouse

NSC genomic DNA. Gibson assembly[228] using standard protocols with homemade enzyme mixes, was then used to assemble the C' terminal *Myef2*-linker-mCherry targeting vector. PCR amplification, to include 1 kbp of 5' and 3' homologous sequence, gave the final double stranded targeting cassette.

The ssODN ultramer, used for section 3.2.2, was ordered from IDT. The ssODN comprises of 86 bp 5' homology arm, HA epitope tag, stop codon, and 82 bp 5' homology arm with PAM mutated from NGG>NGC. The ssODN was from the same strand as the crRNA (known as the PAM-strand).

SMASh tag targeting vectors and piggyBac vectors were assembled using EMMA[229], a modular assembly system which allows for the rapid generation of vectors via a one-tube golden gate reaction. Briefly, each modular part was amplified via PCR from gDNA or plasmid DNA, with the SMASh tag obtained from Addgene #68853[230], to generate a DNA element which is flanked by BsaI restriction sites and the specific positional fusion site which is generated upon cleavage by BsmBI. The fragment was first cloned into a part-entry vector via the BsaI sites and the library part is sequence verified. The one-tube golden gate reaction is then performed. This is composed of: the receiver vector, specific combination of library parts, BsmBI enzyme, and T4 DNA ligase. The final assembled plasmid is the verified by restriction enzyme digest.

For the SMASh tag targeting vectors 500 bp homology arms were PCR amplified from the targeting vector and the product purified using a QIAquick PCR purification kit (Quiagen). The dsDNA was then incubated with 33% DMSO for 5 min at 95^{o}C, then plunged onto ice to promote retention of single stranded DNA. The DNA is kept on ice until used for transfection. This was done to reduce toxicity during transfection (Pollard lab, unpublished observation).

A list of ssODNs, targeting vectors and piggyBac plasmids used is found within the appendix (Appendix, table 8).

2.3 Transfections

Cells were transfected with the Amaxa 4D nucleofection system (Lonza). The transfection solution SG (Lonza) was used with the program DN100. For section 3.2.4, $1.5x10^6$ cells were pre-mixed with 600 ng of targeting cassette and 1 µg of each sgRNA/Cas9 plasmid in 100 µl SG buffer. For section 3.2.2 and chapter 5, $3x10^5$ cells were pre-mixed with 12.4 µg of Cas9/sgRNA ribonucleprotein and 300 ng of targeting cassette or 1.5 µg of ssODN (IDT) in 20 µl SG buffer. The assembly of Cas9/sgRNA complex was performed as previously described[185]. Briefly, 2.2 µg of tracrRNA (IDT) and 1.2 µg of crRNA (IDT) were annealed by heating to 95°C for 5 min and gradually cooled to 25°C at a rate of 0.1°C/sec. 10 µg of recombinant Cas9 protein (expressed and purified from Addgene, #53261) was mixed with the annealed crRNA/tracrRNA at room temperature for 10 min allowing the ribonucleoprotein complex to form. For piggyBac integrations, $3x10^5$ cells were pre-mixed with 300 ng of expression plasmid, 500 ng of hyPBase expression plasmid and 300 ng of pCAG-Tet-On® 3G plasmid.

After transfection cells were plated into uncoated tissue culture plastic and allowed to recover. Targeted cells were then enriched by BD FACSAria™ II (BD bioscience) or treatment with 1 µg/ml puromycin for 7 days.

2.3.1 Colony picking

Single cells were deposited into 96-well plates using BD FACS Aria™ II cell sorter. After 2 weeks approximately 15-30 colonies per 96-well plate were obtained. Colonies

were manually transferred using 5 µl of accutase with a 20 µl pipette. Clones were checked for correct targeting via ICC, PCR and western.

2.3.2 PCR genotyping

Cells from a confluent 96-well were incubated for 2 hrs at 55ºC in 20 µl of lysis buffer (0.2 µg/ml proteinase K, 0.45% NP40, 0.45% Tween20 and 1x LongAmp PCR buffer). Samples were then heated to 95ºC for 10 min. Alternatively, DNA was isolated by DNeasy blood and tissue kit (Qiagen) according to manufacturer's instructions. Genotyping primers were designed outside of the 5' and 3' homology arms. For chapters 3 and 4 LongAmp® Taq (NEB) DNA polymerase was used, whilst PrimerSTAR® MAX was use for chapter 5. The following PCR conditions were used.

Knock-in of HA to Myef2

Step	Temperature (ºC)	Time (sec)	Cycles
Initial denaturation	92	120	1
Denaturation	95	30	
Annealing	60	30	33
Extension	68	50	
Final extension	72	300	1
Storage	4	∞	1

Knock-in of mCherry to Myef2

Step	Temperature (ºC)	Time (sec)	Cycles
Initial denaturation	92	120	1
Denaturation	95	30	
Annealing	60	30	30
Extension	68	180	
Final extension	72	300	1
Storage	4	∞	1

Knock-out of Myef2

Step	Temperature (ºC)	Time (sec)	Cycles
Initial denaturation	95	120	1
Denaturation	95	30	
Annealing	60	30	33
Extension	68	50	
Final extension	72	300	1
Storage	4	∞	1

Knock-in of SMASh to Sox2

Step	Temperature (ºC)	Time (sec)	Cycles
Initial denaturation	94	120	1
Denaturation	98	30	
Annealing (start-final) (A temperature gradient PCR was performed)	65-55 (dropped 1ºC per cycle for 10 cycles prior to final annealing temp)	30	28
Extension	72	45	
Final extension	72	300	1
Storage	4	∞	1

2.3.3 List of genotyping primers

Locus	Function	Forward (5'-3')	Reverse (5'-3')
Myef2	HA knock-in	TGTGGGACAGTCAGGTTTGA	GCTGGTGGGAGTTGTAAAAGTT
Myef2	mCherry knock-in	GCAGTAGTAACCCTGACCGA	ACCAACCTAATCCCCACCAG
Myef2	Knock-out	AGAAGGAAAATGCCTGCTGA	ATTCTTGGTGGCATGCAGAT
Sox2	mCherry and SMASh knock-ins	GTTCATCGACGAGGCCAAG	ACCACGAAAACGGTCTTGC

Table 3: Genotyping primers

2.4 Immunocytochemistry

Cells were fixed in 4% PFA and blocked in blocking solution (PBS with 0.1% BSA and 3% goat serum). For intracellular epitopes, samples were permeabilised with PBST (PBS plus 0.1% triton-x) between the fixation and blocking step. Samples were incubated with primary antibody overnight at 4°C. After three washes, samples were incubated with the respective goat secondary antibody (Invitrogen, 1:1000) for 1 hr. Nuclear counter stain DAPI was used at a final concentration of 1 µg/ml. To acquire images, either a Nikon TiE microscope or a Zeiss fluorescence microscope was used.

2.4.1 List of primary and secondary antibodies used for ICC

Antigen	Species	Subtype	Dilution	Manufacture
mCherry	Rabbit	IgG	1/500	Abcam (ab167453)
HA	Mouse	IgG1	1/100	Cell signalling (2367)
SOX2	Rabbit	IgG	1/100	Abcam (ab92494)
BLBP	Rabbit	IgG	1/200	Santa Cruz (sc30088)
NESTIN	Mouse	IgG1	1/500	R&D (mab1259)
GFAP	Rabbit	IgG	1/1000	Sigma (G9269)
TUJ1	Mouse	IgG2a	1/500	Biolegend (801202)
MBP	Rat	IgG	1/20	Biorad (aa8287)
MYEF2 (non-specific)	Rabbit	IgG	1/200	Sigma (HPA004883)
MYEF2 (non-specific)	Rabbit	IgG	1/200	Sigma (AV32738)
MYEF2 (non-specific)	Rabbit	IgG	1/200	Proteintech (16051-1-AP)

Table 4: List of antibodies used for ICC

2.5 Western immunoblotting

For immunoblotting of tissue samples, a male Bl6 adult mouse was sacrificed and relevant tissues collected. Dissection of the individual brain regions was performed by Maria Angeles Marques (Pollard lab). The tissues suspended in protease inhibitor supplemented lysis buffer (25 mM Tris-HCl pH8, 150 mM NaCl, 1% NP-40, 1% NaDoc, 1% SDS, 20mM NaF and 2.5 mM Na_3VO_4) and manually dissociated through a syringe. Samples were vortexed and incubated on ice for 20 min. Lysates were then cleared by centrifugation at 13000 rpm for 15min at 4°C.

For *in vitro* samples, cell pellets were lysed and incubated for 30 min at 4°C in lysis buffer (150 mM NaCl, 20 mM TrisHCl, 1% NP40, 2 mM EDTA, 1 mM NaF, 0.1% SDS) freshly supplemented with complete protease inhibitors (Roche). The lysate was then cleared by centrifugation at 13000 rpm at 4°C for 15 min. Lysates were then quantified Pierce Assay (Thermo Fisher) according to the manufacturer's instructions. Samples were mixed in loading buffer (LDS with 20 mM DTT) and heated to 95°C for 5 min. Electrophoresis was performed on polyacrylamide gels using the BioRad system and Precision plus protein™ dual Xtra ladder (BioRad). Semi-dry trans-blot Turbo transfer system (BioRad) was used to transfer proteins onto 0.45 µM Immobilon-P PVDF membrane (Millipore).

Membranes were blocked with 5% milk in TBST for 1 hr at room temperature. Membranes where then incubated with the relevant dilution of primary antibody in TBST with 5% milk overnight at 4°C. Washes with TBST were performed at room temperature and membranes were incubated with relevant HPR-conjugated secondary antibodies (Novex) at a dilution of 1:10000 in TBST/5% milk for 1 hr at room temperature. Membranes were then washed with TBST and developed using either homemade ECL or immobilon ECL (Millipore).

2.6 Immunoprecipitations

For immunoprecipitation experiments cells were lysed in protease inhibitor supplemented lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5% NP-40, 5% glycerol, 1 mM DTT, 2 mM $MgCl_2$, 10 mM β-glycerophosphate, 10 mM sodium pyrophosphate, 1 mM NaF and 2 mM Na_3VO_4). Lysates were incubated with 250 units per ml of benzonase (Sigma) for 30 min at 4°C, then cleared by centrifugation. After quantification 0.8 mg–1.5 mg of protein was incubated with magnetic beads conjugated to the relevant primary antibody overnight at 4°C shaking at 1200 rpm. Beads were then washed in wash buffer (lysis buffer without protease inhibitors and NP40). Sample was then mixed with 2x loading buffer and heated to 95°C for 5 min.

2.7 Mass spectrometry

For mass spectrometry, samples were lysed at a concentration of $20x10^6$ cells per ml of lysis buffer (20 mM HEPES pH7.5, 150 mM NaCl,1% triton X-100, 2 mM $MgCl_2$, 1x protease inhibitor cocktail and 250 units/ml benzonase) for 30 min at 4°C. Lysates were then cleared by centrifugation at 13000 rpm for 15 min at 4°C. Samples were flash frozen in liquid nitrogen. This was performed by Carla Blin.

Samples were then sent to collaborator Alex Von Kriegsheim. 500 µl of cleared lysates were incubated with Pierce™ anti-HA magnetic beads for 2 hrs. Samples were then processed as previously described[231]. After obtaining the filtered list, it was analysed within Cluego in cytospace[232]. A *p* value of 0.05 with GO biological processes was used. Data was displayed by group with medium detailed pathways.

2.7.1 List of primary antibodies used for immunoprecipitations and western blotting

Antigen	Species	Sub-type	Application	Dilution	Manufacture
V5	Mouse	IgG2ak	IP	50 µl per IP	MBL (M167-11)
V5	Mouse	IgG2b	WB	1/1000	eBioscience (14679682)
MYEF2	Rabbit	IgG	WB	1/1000	Sigma (AV32738)
SOX2	Mouse	IgG2a	IP WB	2 µg per IP 1/400	R&D (mab2018)
HA	Rabbit	IgG	IP WB	2 ug per IP 1/1000	Abcam (ab 9110)
GAPDH	Mouse	IgG1	WB	1/5000	Ambion (AM4300)
SOX8	Rabbit	IgG	WB	1/1000	Abcam (ab104245)
SOX9	Rabbit	IgG	WB	1/1000	Millipore (ab5535)
OLIG2	Rabbit	IgG	WB	1/3000	Millipore (ab9610)
mCherry	Rabbit	IgG	WB	1/2500	Abcam (ab167453)
GFP	Alpaca	Nano-body	IP	10 µl per IP	Chromotek (gtma-20)
pH3 Ser10	Rabbit	IgG	WB	1/1000	Cell signalling (3458)
HA	Mouse	IgG1	Mass spec	50 µl per IP	Pierce (88836)
EGFR Y1068	Rabbit	IgG	WB	1:500	Cell signalling (3777)

Table 5: List of antibodies used for western blotting and immunoprecipitations

2.8 Flow cytometry

Cells were collected using accutase and resuspended in PBS with 0.1% BSA. A dead live marker of DAPI, DRAQ7 or TO-PRO-3 was used. Samples were analysed using a BD LSR Fortessa cell analyser (BD biosciences) and data analysis was performed with FlowJo analysis software (TreeStar).

2.8.1 Hoechst 3342 staining and cell cycle analysis

$1x10^6$ mouse NS cells were stained with Hoechst 33342 (Tocris), in 1 ml of NSC maintenance media, at a final concentration of 5 µg/ml of Hoechst for 90 min at 37°C. Cell cycle analysis was then performed using FlowJo V10 using the univariate model.

2.9 Cellular assays

2.9.1 Proliferation rate assays

For analysis of proliferation rates cells were seeded at a density of 5000 cells per cm^2. For BMP sensitivity assays, cells were seeded at a density of 3000 cells per cm^2. Samples were then incubated with 10 µM EdU for either 2 hrs or 24 hrs and stained with the Click-iT EdU alexa 647 assay kit (Life technologies) according to the manufacturer's instructions.

2.9.2 Confluency curves

For confluency curve analysis cells were seeded at 5000 cells per cm^2, in triplicate wells, and confluence analysed by the Incucyte live cell imaging system (Essen Bioscience). This assay measures the occupied area of the dish (cell confluency) acting as a read out of cell proliferation, provided cells do not undergo large alterations in morphology.

2.9.3 Colony formation assays

Cells were seeded at density of 10 cells per cm^2 in 10mm dishes and colonies allowed to form for 10-15 days. Fixed cells were stained with 1% methylene blue and colonies counted manually and using the Fiji ColonyArea[233] plugin.

For the de-differentiation assays, cells were seeded at a density of 300 or 500 cells per cm^2 in astrocytic differentiation media for 3 days. After 3 days, media was changed to NSC maintenance media for 10 days.

2.10 Nocodazole treatment

Cells were treated with nocodazole (sigma) at a final concentration of 50 ng/ml for 6 hrs in NSC maintenance media.

2.11 Treatment with HCV NS3/4A serine protease inhibitor

For experiments in Chapter 5, cells were treated with ASV (Cayman Chemical), dissolved in DMSO, at the relevant concentration for the specified time. Samples were supplemented with DMSO, so the final DMSO concentration of all samples was 0.3%.

2.12 Live imaging

Time-lapse images were taken on Nikon TIE microscope at 37ºC and 5% CO_2. Images were taken every 3 min with a 60X oil immersion objective.

2.13 Intracranial transplantation

$2x10^5$ cells, resuspended in 2 µl of NS maintenance media, were transplanted into the frontal cortex of adult male NSG mice (aged 6-8 weeks). A stereotaxic frame was used and cells were injected into the coordinates 1 mm rostral, 2mm lateral to the Bregma and 2.5 mm deep[234]. Mice were monitored for tumour formation through luciferase bioluminescent imaging (IVIS spectrum imager, PerkinElmer) and the

health of the mice was continually monitored. At the onset of symptoms, mice were culled, and tumours were imaged using a leica M165 FC stereomicroscope. Injections were performed by Steven Pollard and Neza Alfazema. Monitoring and bioluminescent imaging was performed by Neza Alfazema and the SCRM animal unit.

Chapter 3 MYEF2, a brain enriched SOX2 interactor

3.1 Introduction

SOX2 is a core transcription factor of NSCs and GSCs, with the downregulation of SOX2 impairing the proliferation and tumour formation capacity of GSCs[84,119–121]. Due to its role in glioblastoma, SOX2 is a clear therapeutic target. However, targeting transcription factors with small molecules is difficult. Identification of key SOX2 interactors may provide an alternative strategy, with the interactor itself or the interaction between SOX2 providing novel therapeutic targets. Also, SOX2 has an important role in many tissue stem cells[112]; hence, there is a need to determine if there are neural or NSC specific mechanisms that could be targeted, without affecting other tissues.

To identify interactors which are most likely to be functionally relevant, Pooran Dewari (Pollard lab) recently used CRISPR/Cas9 to epitope tag several SOX members. As SOX factors are DNA binding proteins, we specifically looked at the on-chromatin interactors of several SOX factors, to increase the probability of obtaining functionally relevant interactors. The well-established mass spectrometry methods, ChIP-MS[235] and RIME[236], crosslink DNA-protein complexes *in-cellulo* prior to immunoprecipitation allowing for stringent recovery under harsh conditions and enrichment for on-chromatin interactions. RIME also includes an initial nuclear isolation step prior, between crosslinking and immunoprecipitation steps, leading to the identification of nuclear only interactors. However, both ChIP-MS and RIME often have contamination issues, from the immunoprecipitation (IP) antibody, and are incapable of actually distinguishing between on-chromatin and off-chromatin interactions. SICAP[237], selective isolation of chromatin associated proteins, combined with mass spectrometry can overcome these issues. In SICAP-MS, as in ChIP-MS and RIME,

DNA-protein complexes are crosslinked *in-cellulo* and the protein of interest and its interactors are isolated under ChIP conditions. In SICAP-MS, unlike ChIP-MS and RIME, DNA is then labelled with biotinylated nucleotides. This allows for the removal of the IP antibody and the isolation of on-chromatin only interactors by streptavidin isolation of the DNA.

3.1.1 Epitope tagging of endogenous *sox* factors

The quality of mass spectrometry data is affected by the availability of high-quality antibodies suitable for IP of the protein of interest. The generation of an epitope tagged fusion protein is one manner to overcome the absence of suitable antibodies. Establishment of nucleofector mediated delivery of a ssODN repair template and recombinant Cas9 pre-complexed with synthetic crRNA/tracrRNA (RNP), enabled for efficient scalable epitope tagging of endogenous loci in NSCs[185]. From this method, V5 knock-in efficiencies of 21% (SOX2), 30% (SOX3), 14% (SOX8) and 26% (SOX9) were obtained for *sox* factors in the mouse GSC line, IENS, without prior selection[185]. Clonal lines for each *sox* factor were subsequently derived and used for mass spectrometry experiments (Pooran Dewari, Pollard lab).

3.1.2 SICAP mass spectrometry of V5 tagged SOX factors in mouse GSC lines

SICAP-MS[237] was performed on the endogenously V5 tagged mouse GSC lines sox2-V5, sox3-V5, sox8-V5 and sox9-v5 by Pooran Dewari, German Monogarov and Jeroen Krijgsveld. All four of these SOX factors are highly expressed within the mouse GSC line, and sox2, sox3 and sox9 are known regulators of NSCs/NPCs[104]. The degree of functional redundancy between SOX factors within the *in vivo* tumour environment is unknown, hence SICAP-MS was performed on all four SOX factors. The data was filtered to remove histones, ribosomal proteins (RPL/RPS),

heterogenous nuclear ribonucleoproteins (HNRNPs) and splicing factors (SRSF), which are common contaminants of mass spectrometry experiments, using Perseus[238]. Many on-chromatin interactors were identified for each SOX factor, including several which were common between all four SOX factors (Fig 3.1, unpublished data provided by Pooran Dewari, German Monogarov and Jeroen Krijgsveld). This included MYEF2, which was the only common interactor of all four SOX factors after CRAPome filtering[239].

3.1.3 Co-immunoprecipitation confirms the interaction of MYEF2 and SOX2

In SICAP-MS protein-DNA complexes are crosslinked, chromatin sonicated to 200-500 bp fragments, protein of interest is immunoprecipitated and DNA bound complexes are isolated. Due to this, SICAP-MS cannot distinguish between protein interactors of the protein of interest and indirect interactors which may be identified as they reside on the same short DNA fragment.

In order to determine if MYEF2 is a protein interactor of SOX factors, co-immunoprecipitations were performed on *sox2*-V5, *sox3*-V5, *sox8*-V5 and *sox9*-V5 mouse GSC lines. Immunoprecipitations for two other TFs, *olig2*-V5 and *foxk2*-V5, were also performed as anticipated negative controls. Using V5 magnetic beads each V5 tagged SOX protein was immunoprecipitated, with the untagged parental line providing an excellent negative control for the V5 immunoprecipitation. Samples were then checked for the co-immunoprecipitation of MYEF2. In the absence of *in-cellulo* crosslinking, used in SICAP-MS, MYEF2 co-precipitated with SOX2; however, this was not the case for SOX3, SOX8 or SOX9. MYEF2 did not co-precipitate with predicted negative controls OLIG2 and FOXK2 (Fig 3.2.A). These data indicate that MYEF2 may interact closely, perhaps even directly, with SOX2, while the interaction

between SOX3, SOX8 and SOX9 is likely to occurs indirectly through their co-localisation in the same chromatin regions.

The interaction between SOX2 and MYEF2 was also observed in the immunoprecipitation of native SOX2 in a mouse NSC line. The interaction was seen in both the absence and presence of ethidium bromide (EtBr). This suggests that the interaction between SOX2 and MYEF2 is not DNA dependant[240], though further work is required to determine this, as this particular interaction may have not been disturbed by the distortions mediated by EtBr. In fact, a higher amount of MYEF2 could be seen with ethidium bromide. It is possible that the distortion cause by ethidium bromide increases the exposure the MYEF2 antibodies epitopes, as the amount of SOX2 precipitated appears similar between the samples (Fig 3.2.B). Finally, the presence or absence of NaF, a serine/threonine phosphatase inhibitor, did not alter the interaction between SOX2 and MYEF2. In the absence of serine/threonine inhibitors it is likely that serine and threonine phosphorylation will be globally lost, suggesting serine/threonine phosphorylation is not required for the interaction of SOX2 and MYEF2. However, it is necessary to confirm the loss of phosphorylation before a solid conclusion can be drawn (Fig 3.2.C).

To conclude, MYEF2 was initially identified as a SOX factor interactor through SICAP-MS. The interaction between SOX2 and MYEF2 occur through the same protein complex, perhaps even directly, while the interaction of MYEF2 and SOX3, SOX8 and SOX9 most like occurs indirectly through co-localisation in the same chromatin regions. Exploration of the function of MYEF2, and the significance of its interaction with SOX2, will form the basis of my next two chapters.

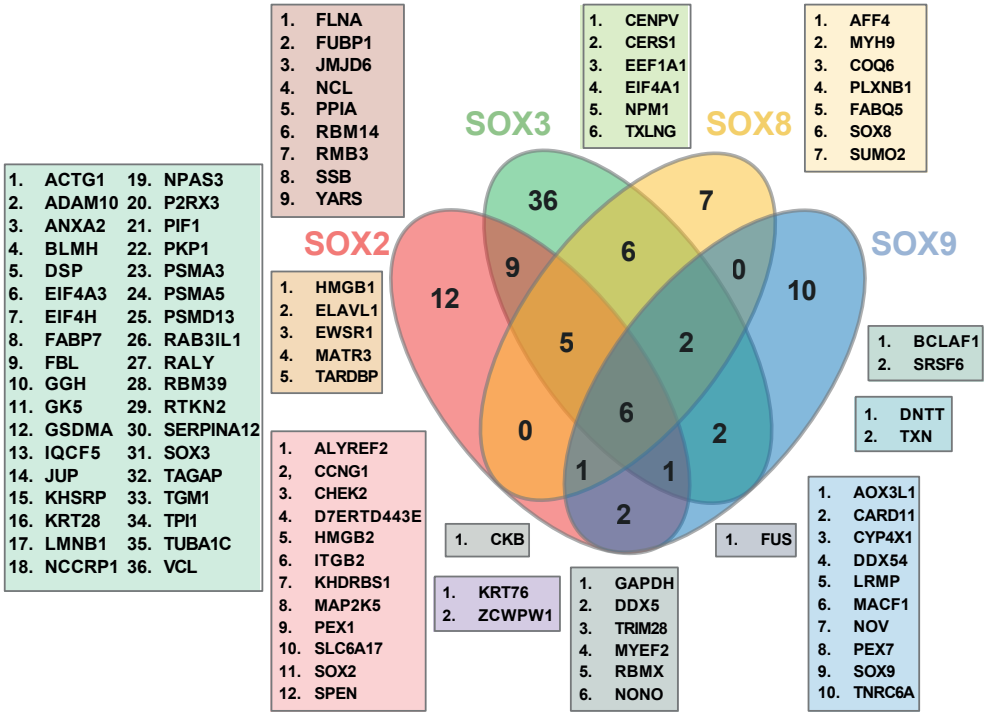


Figure 3.1 | SICAP mass spectrometry of V5 tagged SOX factors in mouse GSC lines

SICAP mass spectrometry data of interactors identified for SOX2-V5, SOX3-V5, SOX8-V5 and SOX9-V5 in mouse GSC cell lines. Data was filtered to remove histones, ribosomal proteins (RPL/RPS), heterogenous nuclear ribonucleoproteins (HNRNPs) and splicing factors (SRSF). Data shown is before CRAPome filtering. N=1. Data for this figure was provided by Pooran Dewari, German Monogarov and Jeroen Krijgsveld.

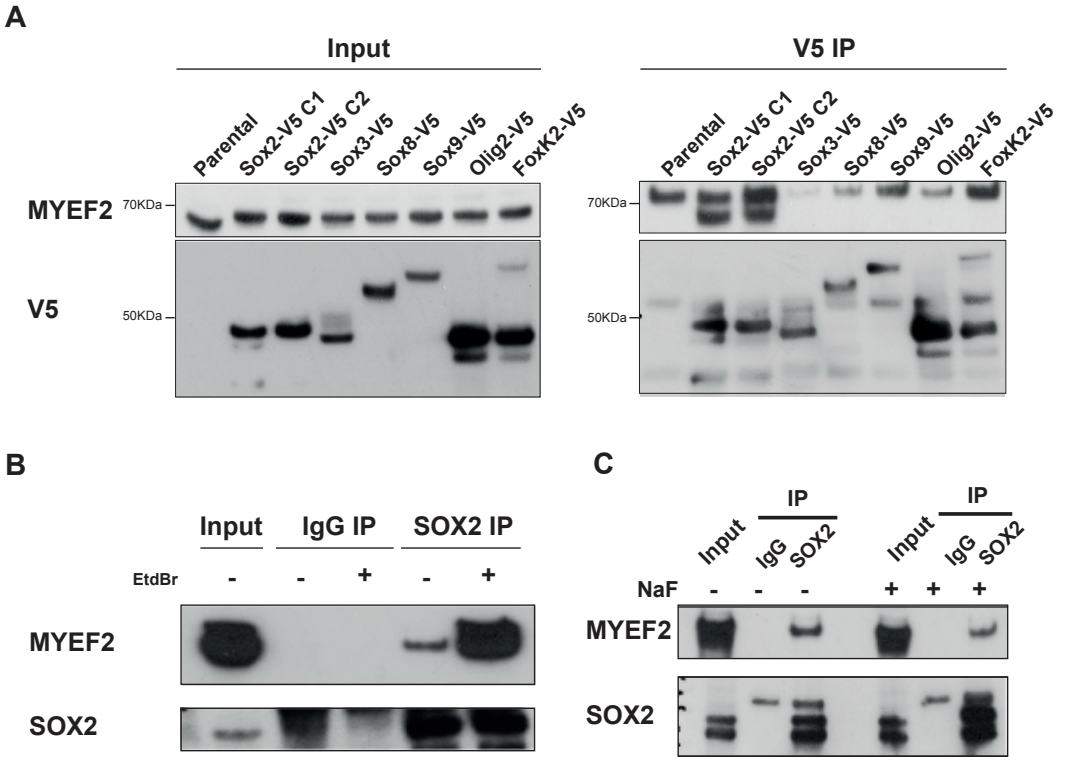


Figure 3.2 | Co-immunoprecipitation of MYEF2 associated with V5 tagged SOX2 and native SOX2

(A) Co-IP of proteins associated with V5 tagged proteins in endogenously tagged mouse GSC lines (IENS). *Sox*2, *sox*3, *sox*8, *sox*9, *olig*2 and *foxk*2 were tagged with V5 as described[185] by Pooran Dewari. Extracts from each line were made with benzonase. Immunoprecipitation against V5 was performed using V5 beads. IPs were blotted for V5 and for the co-IP of MYEF2. Parental untagged line was included as a negative control for the IP. N=1 **(B)** IP of native SOX2 in mouse NSCs in the presence or absence of ethidium bromide**.** Extracts were made from untagged mouse NSC samples (ANS4) with benzonase. Samples were immunoprecipitated with magnetic protein G beads conjugated to anti-SOX2 or anti-IgG antibodies, either in the presence or absence of ethidium bromide. IPs were then blotted for SOX2 and the co-IP of MYEF2. N=1 **(C)** IP of native SOX2 in mouse NSCs in the presence or absence of NaF. Extracts were made from untagged mouse NSC samples (ANS4) with benzonase either in the presence or absence of NaF. Samples were immunoprecipitated with magnetic protein G beads conjugated to anti-SOX2 or anti-IgG antibodies, either in the presence or absence of ethidium bromide. IPs were then blotted for SOX2 or co-IP of MYEF2. N=1.All data in this figure was provided by Carla Blin.

3.2 Results

3.2.1 MYEF2 is a testis and brain enriched protein

MYEF2 was first identified as a single stranded DNA binding protein from a phage based screen to identify proteins, derived from a 15 day postnatal mouse brain cDNA library, that can bind to a single stranded probe generated from the myelin basic protein (*mbp*) proximal promoter region MB1[132]. The proposed sequence of *myef2* from this initial study was a N-terminal truncated version. Subsequently, the full length *myef2* locus was shown to encode a ~60 kDa protein, proposed to have three RRM RNA/single stranded DNA binding domains[241]. Consistent with our SICAP-MS data, MYEF2 has been proposed to be a nuclear localised protein[242]. Human RNA-seq data has shown that *MYEF2* transcripts are enriched in the brain and testes (Fig 3.3.A, data sets taken from the human protein atlas[242]). We therefore performed western blotting to determine in which tissues MYEF2 is found. This confirmed that MYEF2 is enriched at the protein level in the brain and testes of adult mice (Fig 3.3.B). Within the brain, MYEF2 is expressed throughout all core regions tested and was not enriched in the subventricular zone, which harbours the NSCs (Fig 3.3.C). This brain enriched localisation of MYEF2 suggests it could have a neural-specific function and contributed to its potential as an interesting specific target for GBM. We therefore began characterising the expression and functions of this understudied SOX interaction partner.

3.2.2 Generation of epitope tagged *Myef2* by CRISPR-Cas9

To study the function of MYEF2, several commercial antibodies were tested on both *Myef2* wild-type mouse GSCs (IENS) and *Myef2* knockout GSCs (IENS) I generated (Fig 3.4.A - 3.4.D). By ICC, these antibodies gave a strong signal, though this was not always nuclear (MYEF2 antibody 3). Signal was retained in *Myef2* knock-out GSC

lines (IENS), highlighting that the tested commercial antibodies are unsuitable for ICC (Fig 3.4.E). One of these commercial antibodies detects MYEF2 by western, but also identifies additional unspecific proteins.

To overcome these limitations, a C-terminal epitope tag knock-in mouse NSC line was generated. In a wild-type mouse NSC line a single copy of the epitope tag HA was knocked into the endogenous locus of *Myef2* using the RNP based system[185] (Fig 3.5.A). Attempts to tag MYEF2 with V5 were unsuccessful. From the bulk transfected population single cells were sorted into 96 well plates by FACS and subsequent colonies were screened for HA staining. Of the 84 colonies screened, one clone was positive for HA staining (Fig 3.5.B). PCR-genotyping, and sequencing of the relevant PCR product, confirmed that in this clone, one allele of *Myef2* was targeted (Fig 3.5.C, sequencing data not shown) and a band could be observed at the expect molecular weight when immunoblotting against HA (Fig 3.5.D). The HA staining observed was exclusively nuclear. This nuclear staining is in line with current data and the identification of MYEF2 as an on-chromatin partner of SOX2.

3.2.3 Reverse co-immunoprecipitation confirms the interaction of SOX2 and MYEF2

With a tagged *Myef2* NSC line to hand, we were able to next confirm the interaction of MYEF2 and SOX2 co-immunoprecipitation experiments, in the reverse order to that previously performed. MYEF2-HA was immunoprecipitated and the interacting complexes checked for co-interaction of SOX factors and OLIG2.

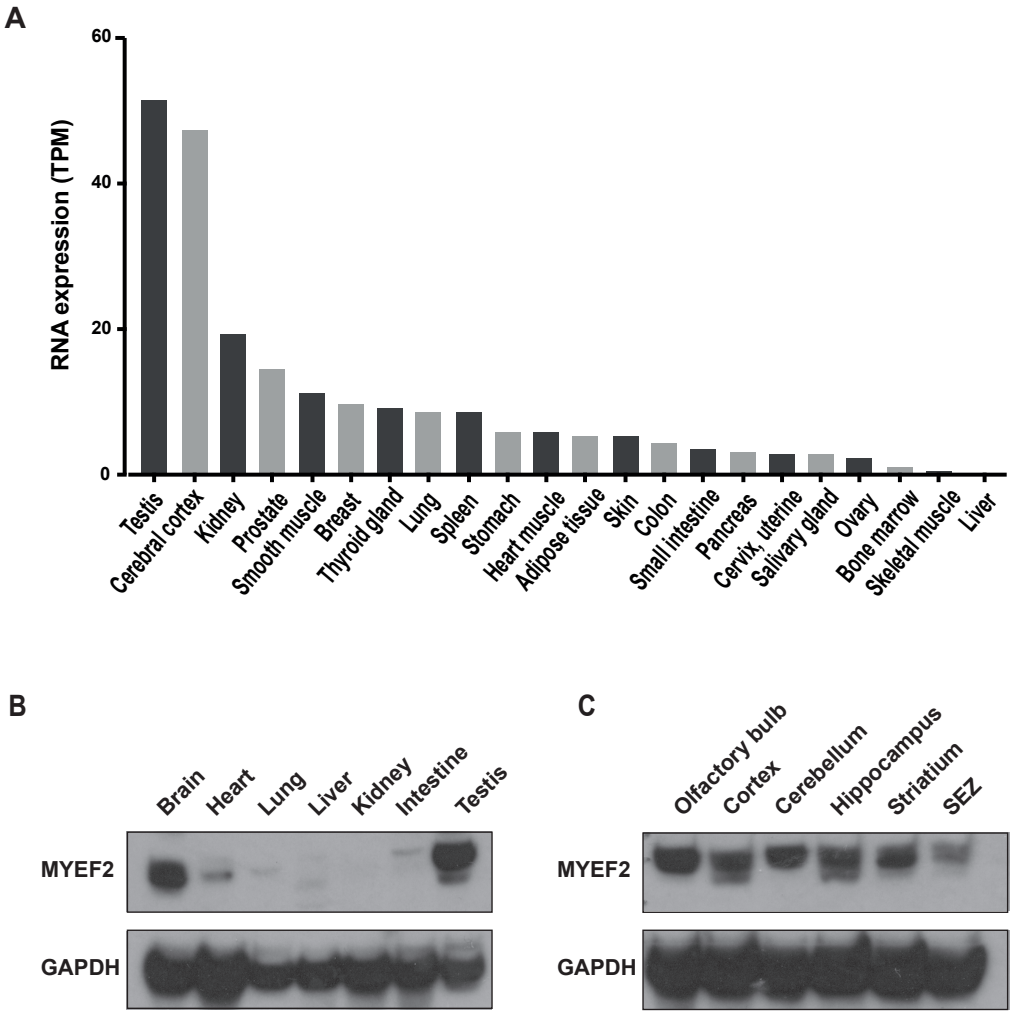


Figure 3.3 | MYEF2 is testis and brain enriched

(A) RNA-seq data of *MYEF2* protein coding transcripts per million (TPM) in various tissues. Data taken from the human protein atlas project[242]. **(B)** Western blotting against MYEF2 on tissue taken from various organs of an adult male mouse (C57BL/6J). N=1. **(C)** Western blotting of MYEF2 on samples taken from various regions of the brain of an adult male mouse (C57BL/6J). N=1.

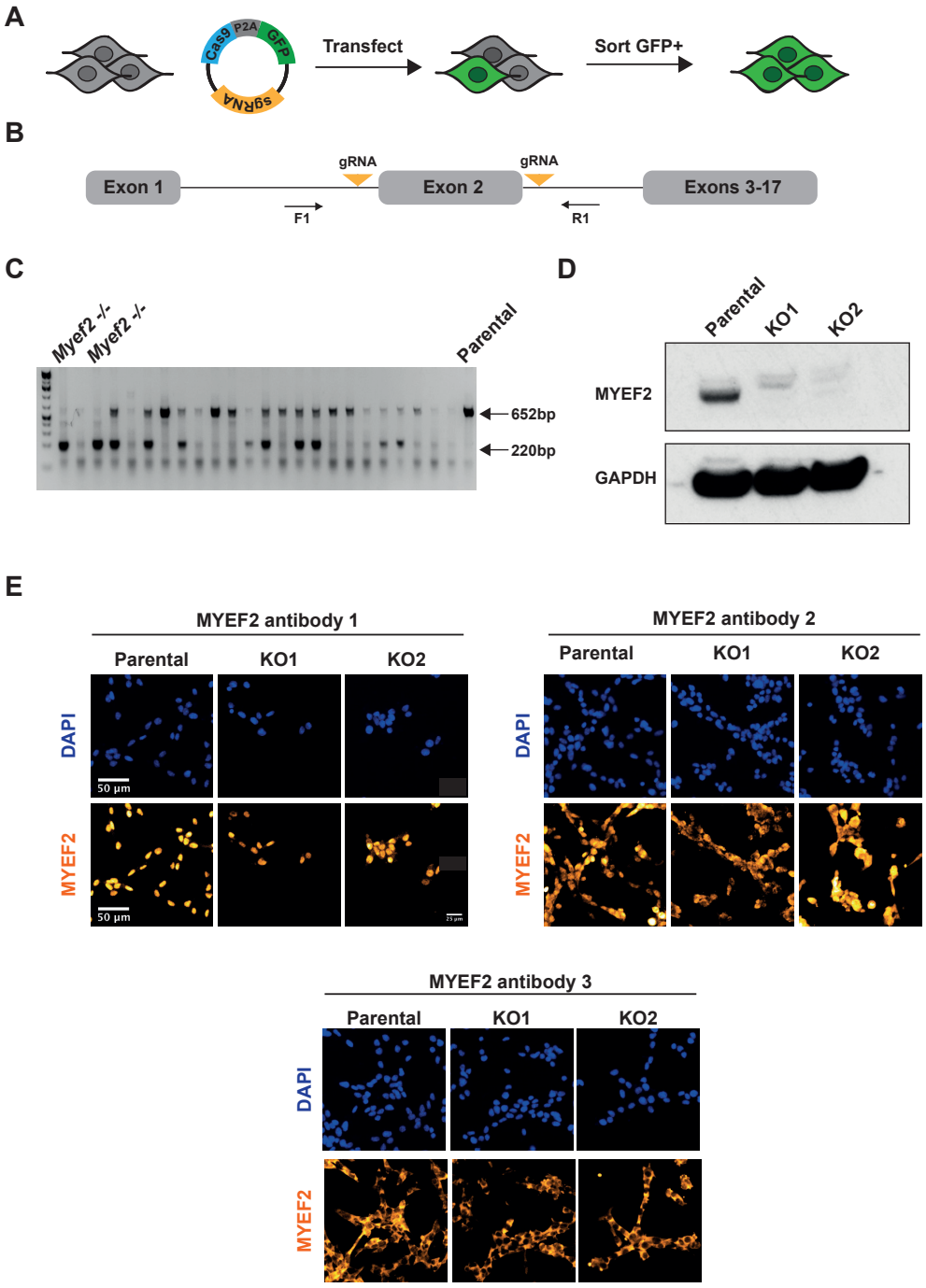


Figure 3.4 | Mouse GSC (IENS) *Myef2* knockout and suitability of commercial antibodies

(A) Experimental method to knockout exon 2 of *Myef2*. Cells were transiently transfected with a plasmid encoding both Cas9-P2A-GFP and sgRNAs. Targeted cells were then enriched by FACS based on transient GFP expression. **(B)** Schematic of *Myef2* locus. Two guide RNAs, depicted in yellow, were used to generate double strand breaks within

the intronic regions flanking either side of exon 2. PCR genotyping primers were designed outside of the cut sites and are depicted as F1 and R1. **(C)** Representative genotyping of IENS clones with primers F1 and R1. Sizes of expected wild-type and knockout bands are highlighted. **(D)** Western blotting of the two clones identified as knockouts by genotyping. **(E)** ICC of *Myef2* wild-type and *Myef2* mutant lines with three commercial antibodies. Antibody 1 corresponds to catalogue number HPA004883, antibody 2 to catalogue number 16051-1-AP and antibody 3 to catalogue number AV32738. Scale bar: 50 μm.

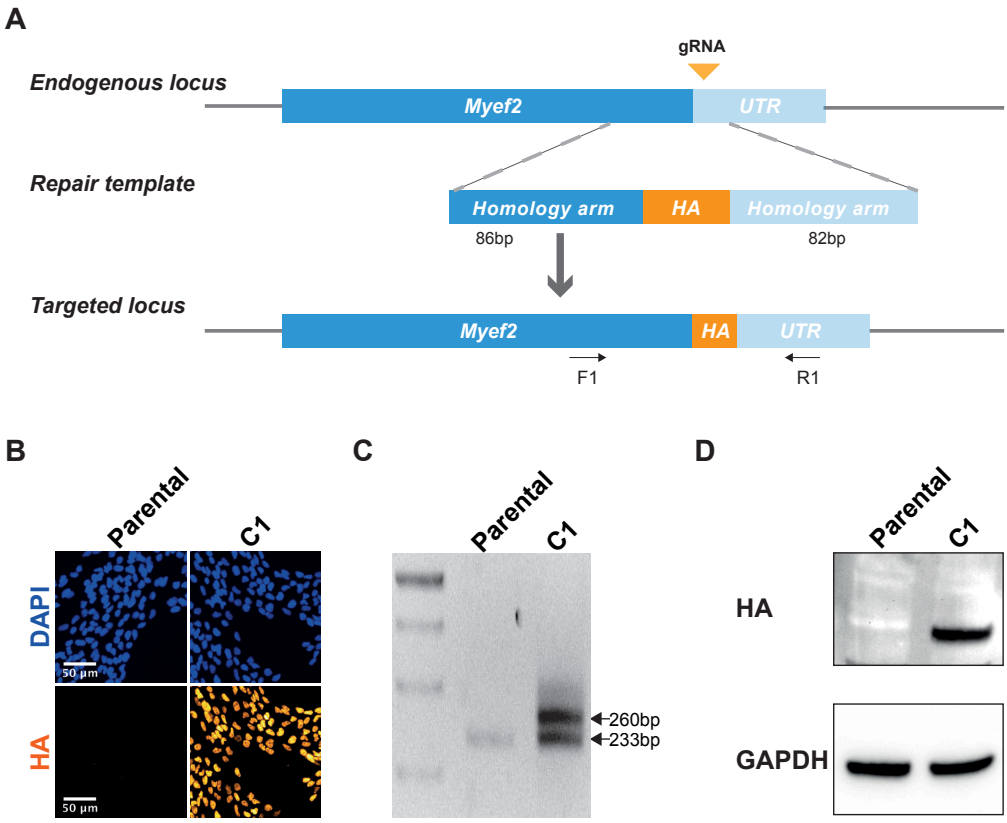


Figure 3.5 | Generation of epitope tagged *Myef2* by CRISPR-Cas9

(A) Schematic of ssODN repair template and targeted *Myef2* locus. CRISPR gRNA, which cuts in the 3'UTR near the stop codon, is depicted by a yellow triangle. PCR genotyping primers were designed outside of the homology arms and are depicted as F1 and R1. **(B)** Representative ICC of wild-type parental line and the targeted clone positive for HA staining. Scale bar: 50 µm. **(C)** Genotyping of the HA positive clone with primers F1 and R1. Expected sizes of wild-type and knock-in bands are shown. **(D)** Western immunoblotting against HA and GAPDH on extracts from the wild-type parental line and the correctly targeted clone.

As previously observed (Fig 3.2), in the absence of *in-cellulo* crosslinking, an interaction between MYEF2 and SOX2 was confirmed, although the result is less striking than previous experiments (Fig 3.6 and Fig 3.3). However, perhaps interestingly, we did not see interaction between with either SOX8 or SOX9. Due to the failure of the SOX3 antibody used to give a signal, information on the interaction between SOX3 and MYEF2 could not be derived from these experiments (Fig 3.6). Overall, this data supports the interaction of MYEF2 and SOX2 within the same protein complex in mouse NSCs, whereas interactions with other SOX members may be rather indirect occurring from co-localisation on the same chromatin regions.

3.2.4 Endogenous knock-in of mCherry to the *Myef2* locus

Fluorophores such as GFP and mCherry are widely used for the study of living cells by microscopy and flow cytometry. To further explore the levels and localisation of MYEF2 we used a plasmid based CRISPR/Cas9 method and knocked in a linker-mCherry encoding sequence to the endogenous *myef2* locus of mouse NSCs (Fig 3.7.A and Fig 3.7.B). By FACS, mCherry positive cells were single cell sorted into 96 wells from a 0.2% mCherry positive population. From the 7 clones which were PCR genotyped, one clone (A8) showed mono-allelic targeting of the *Myef2* locus whilst retaining the other wild-type allele. Two other clones had mono-allelic knock-in of mCherry, but a large deletion in the other *Myef2* allele was observed (Fig 3.7.C). Western blotting against MYEF2 in these two clones gave bands at the expected sizes of both MYEF2-mCherry fusion protein and MYEF2 (data not shown). The normal size of the MYEF2 band suggest the genetic deletion has not dramatically affected the C-terminal of the protein, although sequencing is required to show the exact deletion. These two clones were not taken forward. ICC of clone A8 showed nuclear mCherry expression, as expected for the MYEF2-mCherry fusion protein (Fig 3.7.D). Western blotting of A8 detected a mCherry band of the expected size (~90 kDa) and blotting

using an antibody against MYEF2 identified bands at the predicted size of MYEF2 (~60 kDa) and MYEF2-mCherry (~90 kDa), supporting this is a heterozygous clone as observed by genotyping (Fig 3.7.E).

3.2.5 MYEF2-mCherry levels are retained under differentiation cues

Treatment of NSCs with BMP4/FGF2 or BMP4 promotes entry of NSCs into the quiescence (late G1) and dormancy (G0) respectively, while serum promotes exit from the cell cycle (Maria Angeles Marques, unpublished). When treated with BMP4 or serum for 3 days the levels of SOX2-mCherry drops dramatically (Fig 3.8). This drop is homogenous throughout the population (Fig 3.8.B) with levels dropping to 20% and 7%, correspondingly, of those observed in growth conditions (Fig 3.8.C). The treatment of *sox2*-mCherry NSCs with BMP4/FGF2 for 3 days has a more heterogenous effect on the cells. From live microscopy, in many cells the expression of SOX2-mCherry drops dramatically, though it is retained in a fraction of the population. The morphology of cells appears similar throughout the population (Fig 3.8.A). This heterogeneity can be more clearly seen from flow cytometry data. While the average SOX2-mCherry level within the BMP4/FGF2 treated population is 50% of that in growth factors (Fig 3.8.C) the expression level of SOX2-mCherry in individual cells varies (Fig 3.8.B).

Unlike SOX2, MYEF2-mCherry can be seen by live microscopy in all conditions (Fig 3.8.A). The levels observed are similar to those in growth conditions, with a marginal reduction occurring in BMP4 and serum (Fig 3.8.C). The range of MYEF2 expression within in *Myef2*-mCherry cells in all conditions is similar, with the population showing a similar distribution as to that observed in proliferative conditions (Fig 3.8.B). A small number of mCherry negative cells can be seen in all conditions, though amounts are higher in BMP4/FGF2 and serum treated cells (Fig 3.8.B). MYEF2 therefore is not lost

upon differentiation and therefore could have SOX2 independent roles. It should be noted that the generation of a MYEF2-mCherry fusion protein may alter the normal behaviour of MYEF2. Complementary methods, such as western blotting and qPCR, would circumvent potential fusion protein caveats, but are limited to observations on the population as a whole, rather than for individual cells.

3.2.6 Nuclear MYEF2 is retained upon differentiation

Adherent NSCs are tripotential, capable of differentiating into neurons, astrocytes and oligodendrocytes[243]. The clonal NSCs lines of *Myef2*-mCherry and *Myef2*-HA retained this tripotential capability (Fig 3.9 and Fig 3.10). In line with previous observations (Fig 3.7) nuclear MYEF2 is retained in GFAP positive astrocytic like cells, at a level similar to that of undifferentiated NSCs (Fig 3.9.A and Fig 3.10.A). Differentiation to neurons and oligodendrocytes gave more heterogenous populations, with ß-III tubulin (Tuj1) positive neurons and MBP positive oligodendrocytes also retaining nuclear MYEF2 expression at levels comparable to undifferentiated NSC (Fig 3.9 and 3.10).

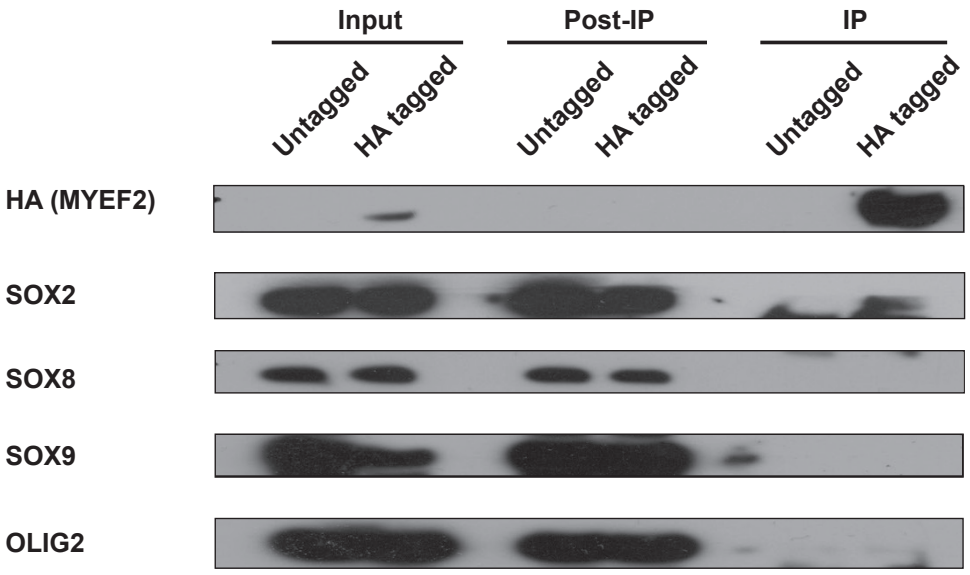


Figure 3.6 | Co-IPs using Myef2-HA confirm the interaction with SOX2

Co-immunoprecipitation of proteins associated with HA tagged MYEF2 in mouse *myef2*-HA NSC line using anti-HA conjugated magnetic beads. Untagged wild-type parental line provides negative control for the HA IP. Between 0.8-1.2 mg of protein extract, pre-treated with benzonase, was used per co-IP. After the IP, blotting against HA and native SOX2, SOX8, SOX9 and OLIG2 was performed. Clean-blot HPR was used (1/500). N=2.

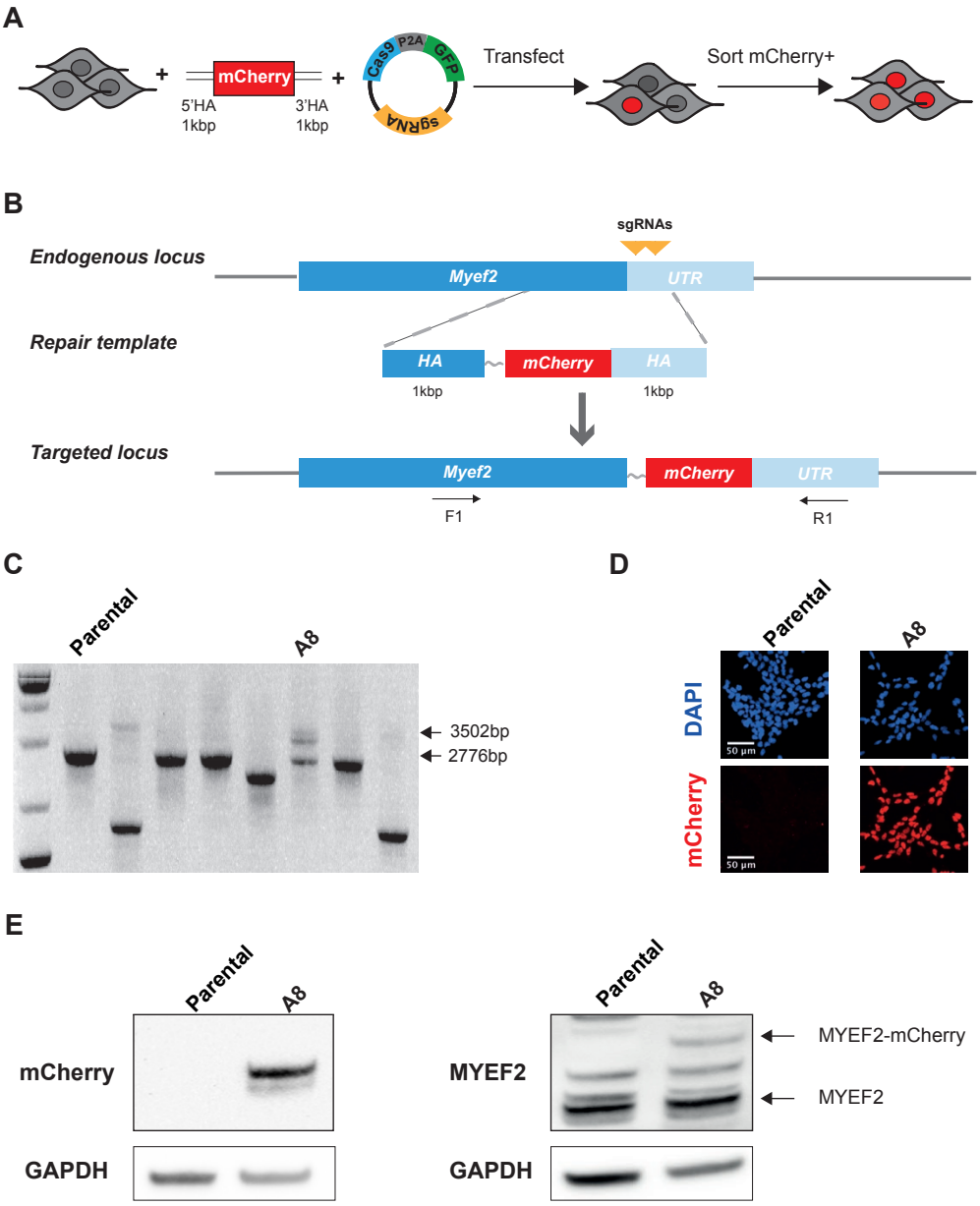


Figure 3.7 | Endogenous knock-in of mCherry to the *Myef2* locus

(A) Experimental method to endogenously knock-in mCherry to *Myef2*. Cells were transiently transfected with a mCherry targeting cassette (provided by Maria Kalantzaki, Pollard lab) and plasmids encoding both Cas9-P2A-GFP and sgRNAs. Targeted cells were then selected by FACS based on mCherry expression. **(B)** Schematic of dsDNA mCherry targeting cassette and the targeted *Myef2* locus. CRISPR sgRNAs, which cut in the 3'UTR near the stop codon, are depicted by yellow triangles. 2 different sgRNAs were used. PCR genotyping primers were designed outside of the 1 kbp homology arms and

are depicted as F1 and R1. **(C)** Genotyping of mCherry positive sorted clones with primers F1 and R1. Expected sizes of wild-type and knock-in bands are shown. **(D)** Representative ICC of wild-type parental line and the targeted clone A8. Scale bar: 50 µm. **(E)** Western immunoblotting against mCherry, MYEF2 and GAPDH on extracts from the wild-type parental line and the clone A8.

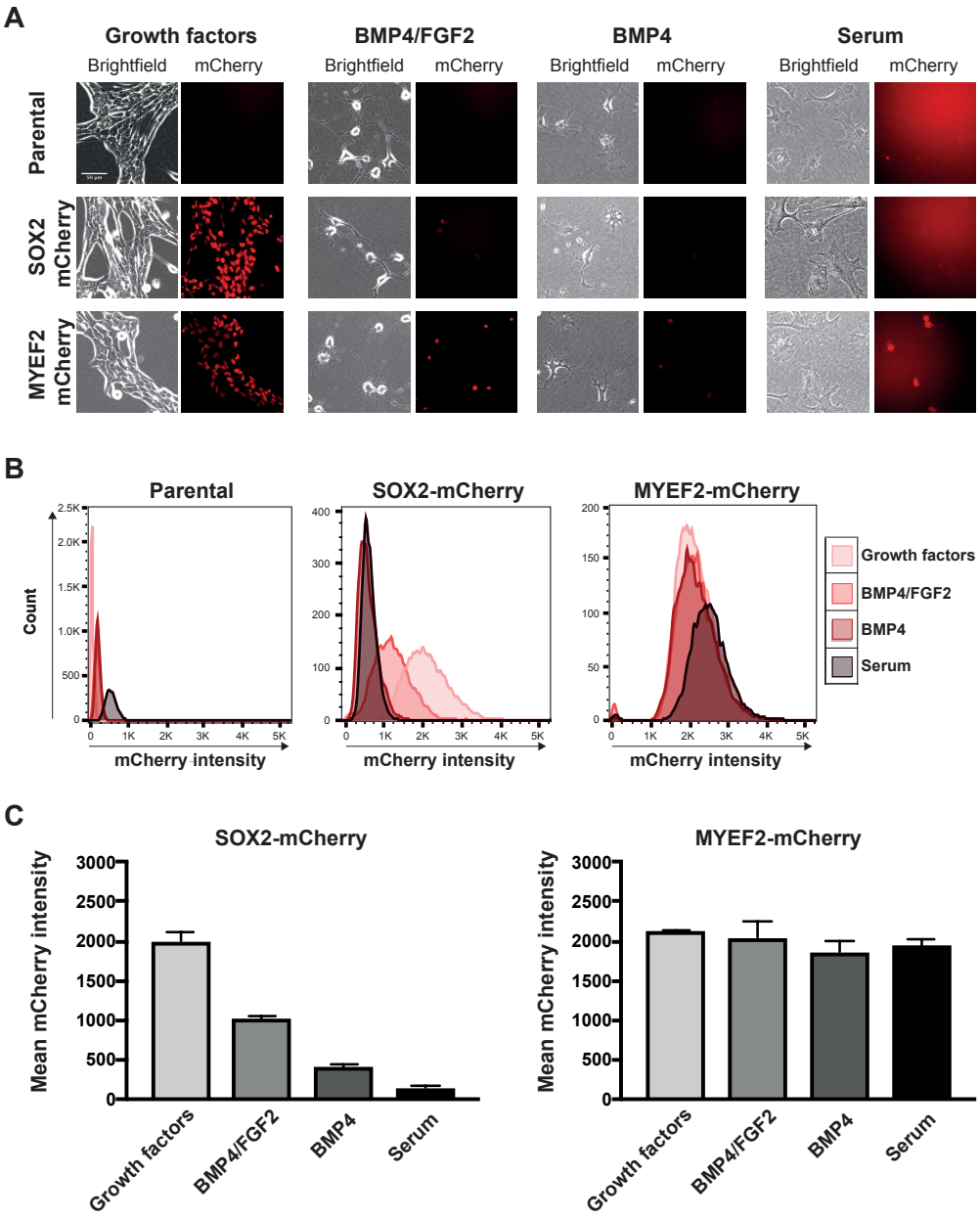


Figure 3.8 | MYEF2-mCherry levels are retained under differentiation cues

(A) Representative microscopy images of live mouse NSCs treated for 3 days in complete media with growth factors, BMP4 and FGF2, BMP4 or 10% serum. (Bulk sorted *Sox2*-mCherry NSC line provided by Pooran Dewari, Pollard lab). Scale bar: 50 μm, n=3 biological replicas. **(B)** Representative flow cytometry data of lines treated for 3 days in complete media with growth factors, BMP4 and FGF2, BMP4 or 10% serum. Parental wild-type line provided gating controls. Data is represented on a linear scale. n=3 biological replicas. **(C)** Flow cytometry data of lines treated for 3 days in complete media

with growth factors, BMP4 and FGF2, BMP4 or 10% serum. Due to the varying levels of autofluorescence between conditions data is normalised to parental wild-type line in the corresponding treatment. Graph shows mean mCherry intensity ± SD from 3 biological replicas.

3.2.7 MYEF2 does not bind to mitotic chromosomes

During mitosis major changes occurs within the cell. These include a general repression in transcription, nuclear envelope disassembly, histone phosphorylation and chromatin reorganisations including changes to 3D topology and its condensation[244]. These chromosomal changes are associated with a general displacement of transcription factors from chromatin. Recently, however, it has been shown that several transcription factors, including GATA1[245], FOXA1[246] and ESRRB[247], remain bound to many of their interphase targets on the mitotic chromatin. This feature is known as mitotic bookmarking and is thought to be involved in rapid post-mitotic reactivation of genes.

Live cell imaging of SOX2 fusion proteins has shown that SOX2 is bound to the mitotic chromatin in embryonic stem cells and there is some evidence this is occurring, at least in part, through the interphase targets of SOX2[248–250]. Through live cell imaging of several NSC *Sox2* fusion lines we have observed that SOX2 is retained on the mitotic chromatin in NSCs (Kalantzaki, Pollard lab, unpublished).

This was first observed in our lab by Maria Kalantzaki using *Sox2*-mCherry NSC lines (Pollard lab, data not shown) and recently confirmed in both a *Sox2*-GFP NSC line (Charles Williams, Pollard lab, data not shown) and an independent *Sox2*-mCherry NSC line (Fig 3.11.A). It remains unclear if this retention in NSCs is occurring through specific interphase sites and has bookmarking roles.

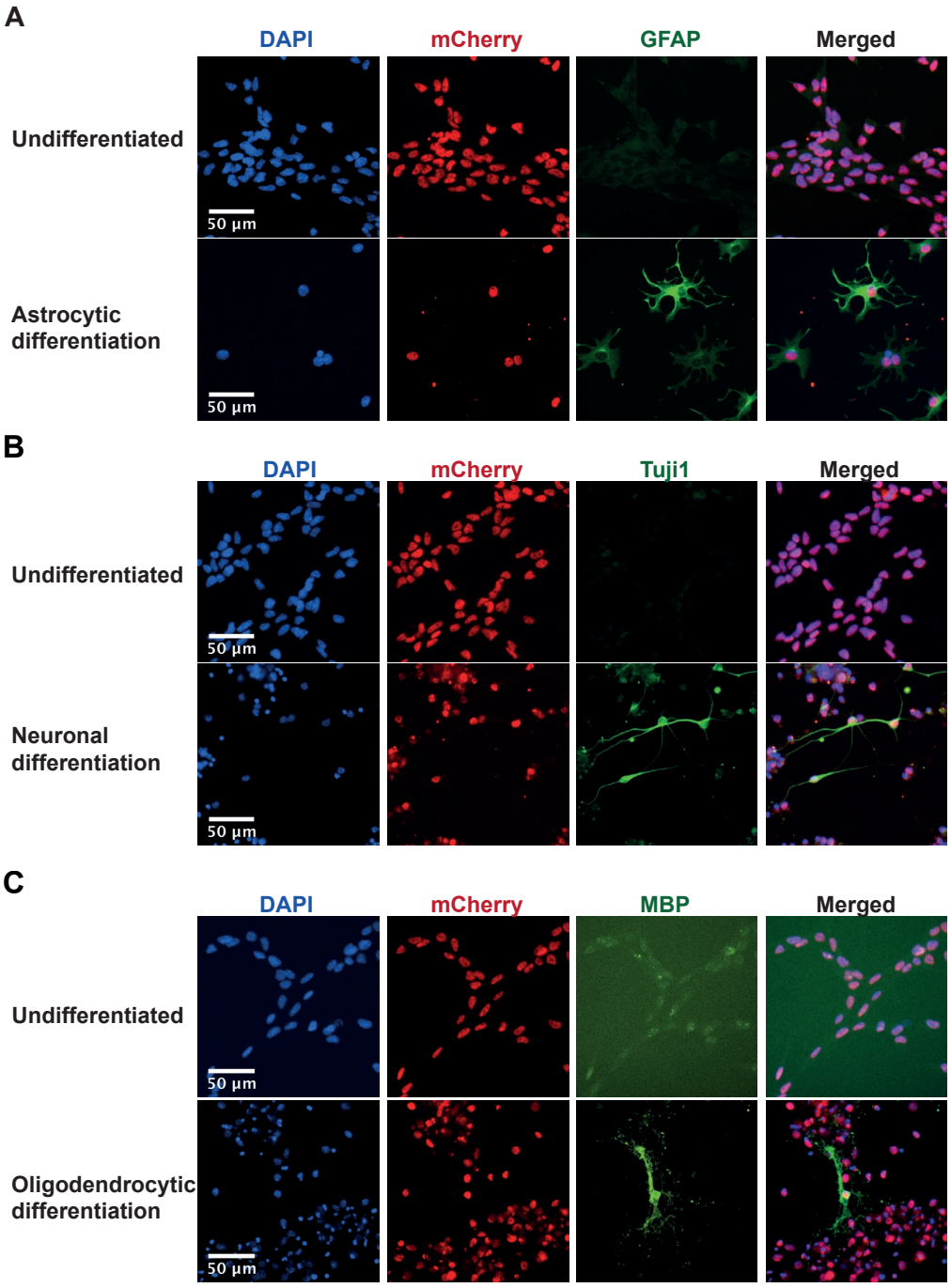


Figure 3.9 | Nuclear MYEF2-mCherry is retained upon differentiation

Representative ICC images of the mouse NSC *Myef2*-mCherry line A8 differentiated to **(A)** GFAP positive astrocytes **(B)** Tuj1 positive neurons and **(C)** MBP positive oligodendrocytes. Undifferentiated cells provided ICC control, with all cells observed in focus expressing MYEF2-mCherry. All cells positive for differentiation markers (GFAP/ Tuj1/ MBP) were also positive for MYEF2-mCherry. Scale bar: 50 µm. n=1.

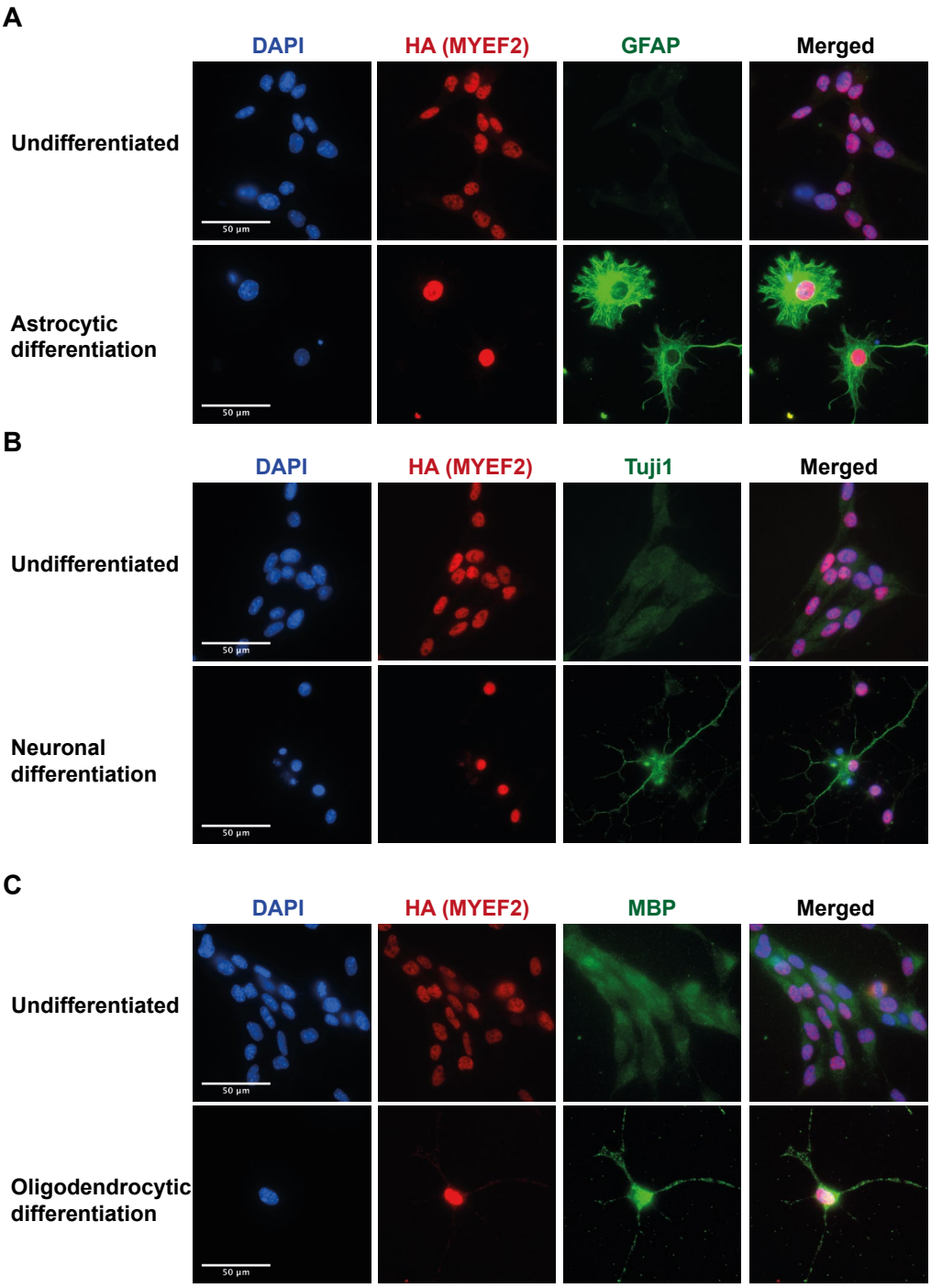


Figure 3.10 | Nuclear MYEF2-HA expression is retained upon differentiation

Representative ICC images of the mouse NSC *Myef2*-HA line differentiated to **(A)** GFAP positive astrocytes **(B)** Tuj1 positive neurons and **(C)** MBP positive oligodendrocytes. Undifferentiated cells provided ICC control, with all cells observed in focus expressing MYEF2-HA. All cells positive for differentiation markers (GFAP/ Tuj1/ MBP) were also positive for MYEF2-HA. Scale bar: 50 µm. n=1.

As fixation with paraformaldehyde can lead to artificial exclusion of factors from the mitotic chromatin[249], live cell imaging of *Myef2*-mCherry NSC line was performed to asses if, like SOX2, MYEF2 may be a bookmarking factor. MYEF2-mCherry was not observed on the mitotic chromatin (Fig 3.11.B). Instead MYEF2 is clearly absent from mitotic chromosomes, and is present diffusely throughout the cell, then returning to the nucleus once the nuclear envelop reforms (Fig 3.11.B). This suggests MYEF2 is not a mitotic bookmarking factor and becomes dissociated from SOX2 in mitosis.

From live cell imaging the MYEF2 signal appears weaker in mitotic cells. It is, however, unclear if this observed drop in levels of MYEF2 during mitosis is due to the nuclear envelope disassembly in mitosis, and hence a more diffuse signal. To quantify the levels of MYEF2-mCherry throughout the cell cycle, *Myef2*-mcherry NSCs were incubated with Hoechst 33342 and analysed by flow cytometry. Using Hoescht 33342 determined cellular DNA content and ploidy, we observed that the levels of MYEF2-mCherry were similar between G0/G1 (2N), S phase and G2/M (4N). The same was also observed for SOX2-mCherry (Fig 3.12). This indicates that there is a redistribution, rather than downregulation, of MYEF2 during mitosis. These findings suggest that although MYEF2 and SOX2 are interaction partners during interphase, the interaction with SOX2 during mitosis is broken.

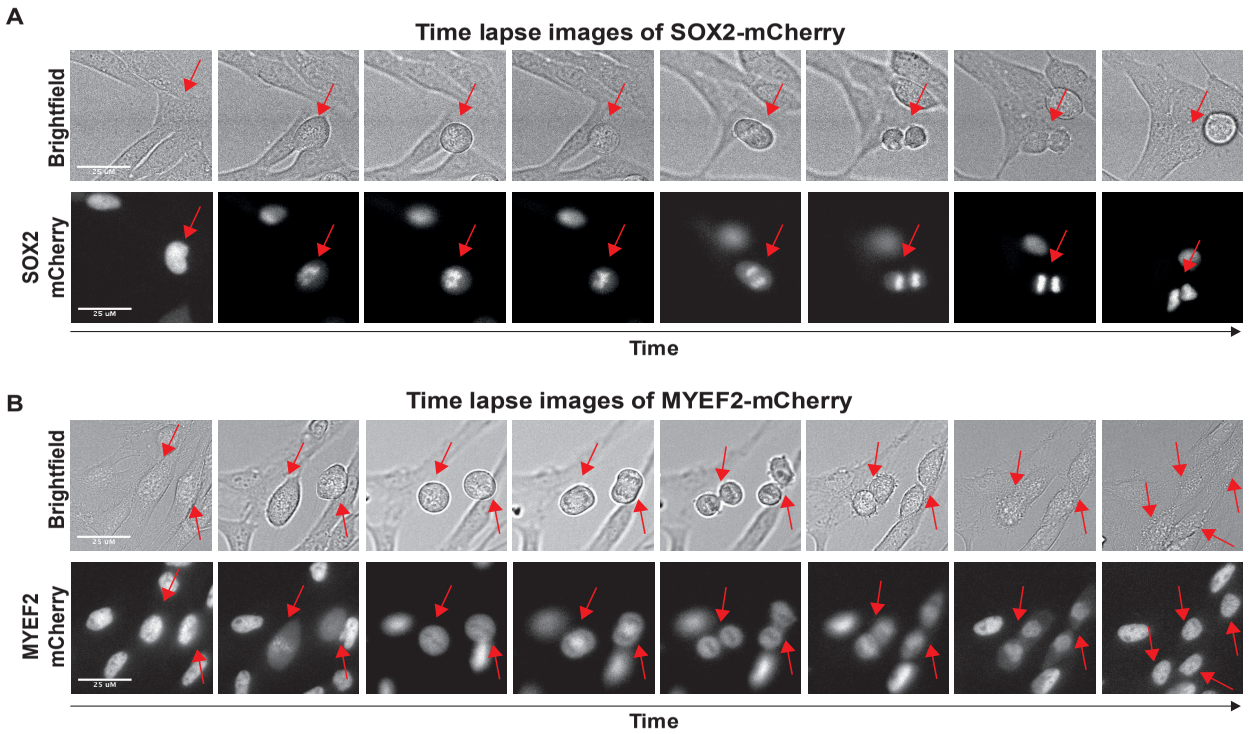


Figure 3.11 | Live imaging of *sox2*-mCherry and *myef2*-mCherry NSC lines

Representative live cell images of **(A)** SOX2-mCherry and **(B)** MYEF2-mCherry throughout the cell cycle. To highlight the differences observed between SOX2, which can be retained on mitotic chromatin, and MYEF2, which cannot, the mean mCherry intensity was roughly quantified. For images 3-5 in the panel, which represent cells in varying stages of mitosis, mCherry signal of a non-chromatin and a chromatin region (based on brightfield images) was calculated. For SOX2-mCherry, an enrichment of 2.2, 2.3 and 2.3-fold over non-chromatin regions was observed for images 3, 4 and 5 respectively. For MYEF-mCherry, a depletion of 0.8, 0.8 and 0.6-fold over non-chromatin regions was observed for images 3, 4 and 5 respectively. Scale bar: 25 μm.

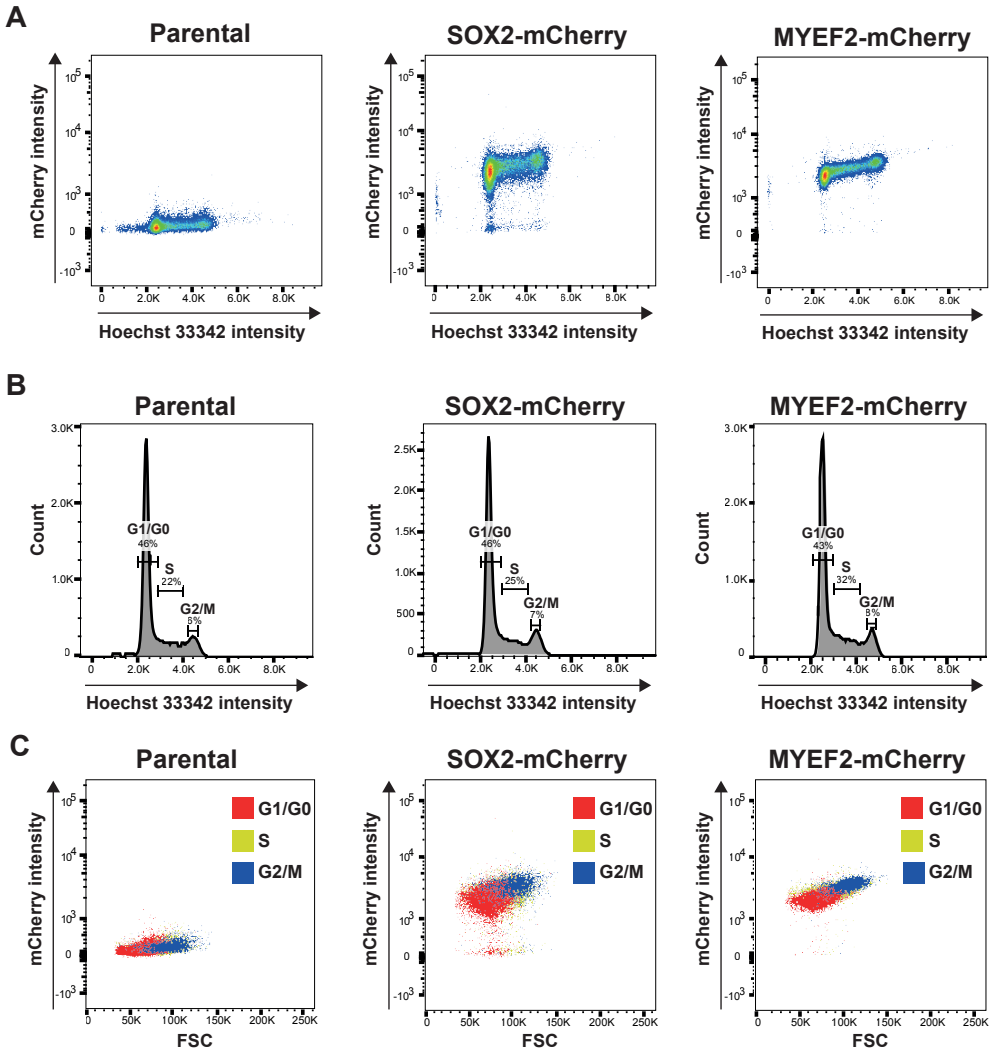


Figure 3.12 | MYEF2-mCherry levels are consistent throughout the cell cycle

Flow cytometry analysis of parental wild-type, *Sox2*-mCherry and *Myef2*-mCherry NSC lines incubated with Hoechst 33342. TO-PRO-III was used as a live-dead marker and parental line used for gating controls, n=2. **(A)** Flow cytometry plot of fluorescent intensity (mCherry vs Hoechst 33342) for each individual cell. **(B)** FlowJo V10 univariate cell cycle modelling to identify cellular ploidy. **(C)** Plot shows mCherry intensity of cells from the corresponding cell cycle population, with G1/G0 (2N) shown in red, S phase in yellow and G2/M (4N) in blue.

3.2.8 Interaction of MYEF2 and SOX2 in mitosis

The earlier data revealed that SOX2 and MYEF2 interact in interphase under proliferative conditions (section 3.1.3). However, MYEF2 unlike SOX2 does not bookmark (section 3.2.7). We, therefore, next further explored the interaction between SOX2 and MYEF2 during mitosis.

NSCs overexpressing a wild-type SOX2-GFP fusion protein (provided by Charles Williams) were incubated either with nocodazole, which disrupts microtubules and arrests cells in mitosis, or without nocodazole (Fig 3.13.A). From the anti-GFP nanobody immunoprecipitation of SOX2-GFP, we observed that MYEF2 co-immunoprecipitated with SOX2 in the absence of nocodazole, but did not co-immunoprecipitate in the nocodazole treated mitotically enriched sample (Fig 3.13.B). This confirms our earlier imaging data, that the interaction of SOX2 and MYEF2 is disrupted in mitosis.

SOX2 undergoes mitotic phosphorylation[251]. We hypothesised that this phosphorylation may disrupt the SOX2/MYEF2 interaction. Mass spectrometry of NSCs identified several specific serine/threonine phosphorylation sites (Charles Williams, unpublished data). NSC lines with doxycycline inducible mutant SOX2-GFP were generated by piggyBac integration (cell lines by Charles Williams). Two variants of particular interested were the lines referenced to as 3A-SOX2, in which serine residues 251, 252 and 253 were mutated to alanine, and B5-MK-SOX2, in which all serine/threonine residues within amino acid stretch 240-280 were mutate to alanine. Each of these mutant NSC lines showed nuclear SOX2-GFP when induced with doxycycline (Fig 3.13.A), confirming the transgene was expressed and retained nuclear localisation at interphase (which is ~95%) of the population. IP of each of the two SOX2-GFP phosphorylation mutants failed to show the co-immunoprecipitation

of MYEF2 (Fig 3.13.B). This suggests that the phosphorylation of SOX2 impairs the binding to MYEF2. These data support a model in which mitotic kinases phosphorylate Sox2, and these phosphorylation sites are involved in the interaction of SOX2 with MYEF2.

3.2.9 Mass spectrometry of MYEF2-HA in mouse NSCs

To identify the interactome of MYEF2, mass spectrometry was performed on HA immunoprecipitated complexes from the *Myef2*-HA NSC line. HA immunoprecipitation and mass spectrometry in an untagged parental line was also carried out, providing a negative control. This was performed in three technical replicas by Alexander von Kriegsheim at the Edinburgh CRUK Centre Mass Spec core facility[252].

Proteins with an average intensity over untagged control ratio higher than 2, and a t-test *p* value ≤ 0.05, were considered to be specific. From the filtered list, 75 candidate protein interactors of MYEF2 were identified (Appendix, table 9). One of these interactors was SOX2, further supporting previous observations. No other SOX factors were within the filtered list, though SOX8 was observed just after the cut off criteria with a ratio of 5.5 and *p* value of 0.058. This also supports our conclusions and data from the SICAP experiments. The filtered set of interactors were analysed in the gene ontology enrichment software, ClueGO[232]. Interactors of MYEF2 were over-represented in biological pathways associated with RNA processing, especially splicing (Fig 3.14, *p* value ≤0.05). This is consistent with the RNA binding (RRM) domains present within *Myef2* and indicates that this protein is indeed enriched in RNA associated pathways. Further experiments are required to identify if MYEF2 directly interacts with RNA and, if so, the biological relevance of interaction. However, these findings suggest it may have broad binding to the splicing machinery.

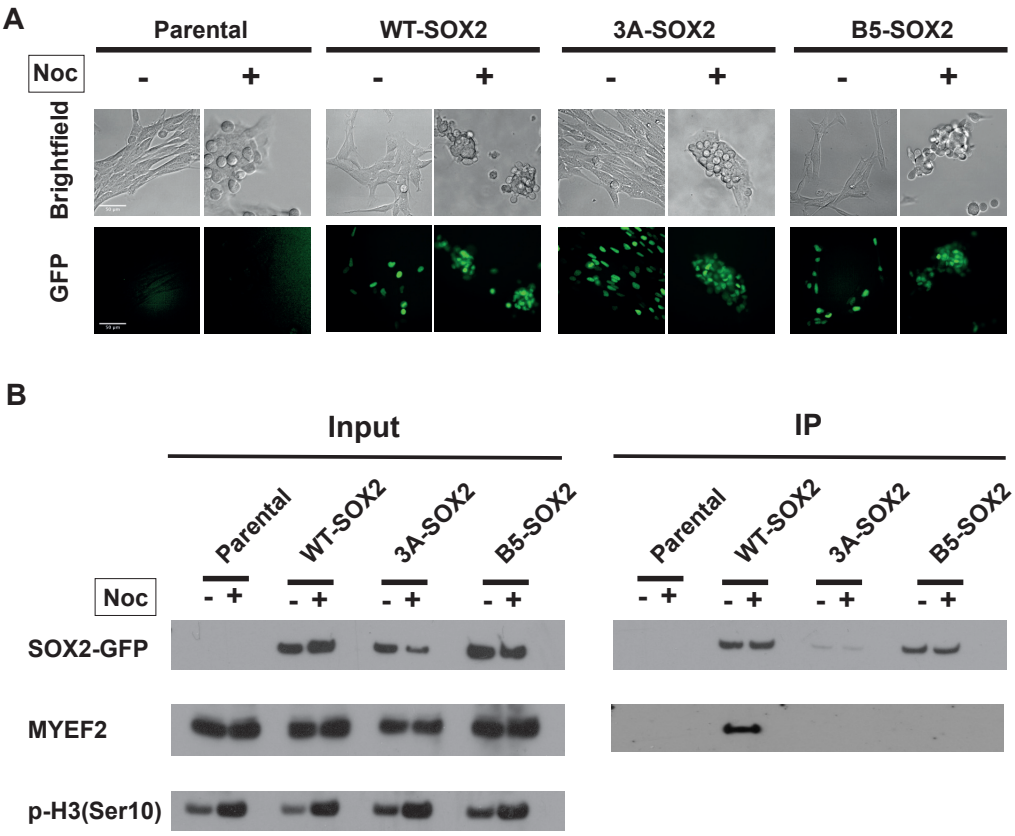


Figure 3.13 | Interaction of SOX2 and MYEF2 in mitosis

(A) Representative live cell images of SOX2-GFP and mutant SOX2-GFP lines. Lines were treated with doxycycline for 48 hrs and nocodazole for 6 hrs. Parental wild-type line provided negative control. All GFP inducible lines were made by Charles Williams. Scale bar: 50 μm. **(B)** IP of SOX2-GFP and mutant SOX2-GFP lines with anti-GFP nanobody. Lines were treated with doxycycline for 48hrs and nocodazole for 6hrs. Parental wild-type line provided the negative control for the IP of SOX2-GFP. Phospho-histone H3 (Ser10) acts as a mitotic marker, indicating the mitotic enrichment mediated by nocodazole treatment. Immunoprecipitated SOX2-GFP was then probed for the co-immunoprecipitation of MYEF2. N=1.

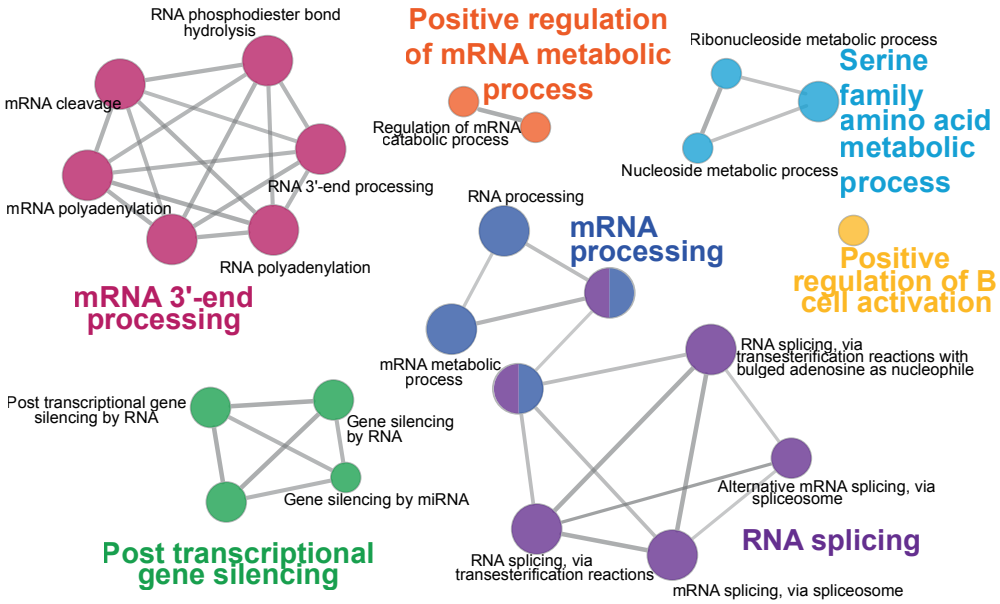


Figure 3.14 | Mass spectrometry of MYEF2-HA in mouse NS cells

Mass spectrometry data of MYEF2 interactors. MYEF2-HA was immunoprecipitated and complexes analysed by mass spectrometry. Identified interactors were then filtered based on the average intensity over control ratio and a t-test *p* value ≤ 0.05. n=3, technical replicas (performed by Edinburgh Cancer Centre, Mass Spec core). GO enrichment analysis, based on biological processes, was performed on the filtered data set using the cytoscape plugin ClueGO[232] (*P* value ≤ 0.05).

3.3 Discussion

SICAP coupled mass spectrometry was previously performed, and identified MYEF2 as a candidate on-chromatin interactor of SOX2 in mouse GSCs. Co-immunoprecipitation experiments confirmed that the interaction between MYEF2 and SOX2 occurs within normal mouse NSCs and mouse GSCs. The interaction between MYEF2 and SOX2 occurred both in the presence and absence of crosslinking, suggesting that MYEF2 interacts with SOX2 either directly or through a common complex. To conclusively conclude that MYEF2 and SOX2 directly interact, the interaction between purified proteins would need to be tested. The interaction of MYEF2 with other SOX factors (SOX3, SOX8 and SOX9) appears to rely on crosslinking, suggesting that the interaction observed in SICAP-MS maybe indirect, due to their co-localisation at certain chromatin regions or regulatory elements. Alternatively, the interaction of MYEF2 and other SOX factors, within the same protein complex, may be more transient or only occur in a small subsection of MYEF2 containing complexes, escaping detection in the performed co-immunoprecipitation experiments.

The preference of SOX2 binding compared to other SOX members is intriguing. This might be consistent with the anticipated role of SOX2 as a pioneer factor, that lies upstream of other members. We speculate that SOX2 and MYEF2 interactions may be important in recruitment of the other SOX members to key target loci. Preliminary SICAP-MS data on human glioma neural stems cells also identified MYEF2 as an on-chromatin SOX2 interactor (Pooran Dewari, Pollard lab, unpublished data). Supporting our observations, independent studies have also previously identified MYEF2 as a SOX2 interactor. In ESCs overexpressing SOX2-Flag, which promotes ESC differentiation, the interaction between SOX2 and MYEF2 was observed[253]. The interaction between SOX2 and MYEF2 was also enriched in medulloblastoma DAOY

cells overexpressing inducible SOX2 compared to uninduced cells[254]. Finally, MYEF2 has been observed to co-immunoprecipitate with flag-SOX2 purified from the nuclear extract of NSCs[255]. From our data, and that of others, we show that MYEF2 is an interactor of SOX2 in NSCs. The interaction between MYEF2 and SOX2 appears to be ablated in mitosis. From live cell microscopy on *Myef2*-mCherry NSCs, MYEF2-mCherry was not observed on the mitotic chromatin, though we cannot rule out that a small fraction of MYEF2 is not retained on specific sites during mitosis. This contrasts with SOX2, which we have identified as a clear binder to mitotic chromosomes and a potential bookmarking factor.

Our most recent experiments suggest that it is the mitotic phosphorylation of SOX2 that effects the interaction with MYEF2. Further experiments are required to confirm, that the mitotic phosphorylation of SOX2 disrupts the interaction of MYEF2 and SOX2, and the functional importance of this for continued self-renewal. The behaviour of MYEF2 does not mimic that of SOX2 when cells are exposed to differentiation cues. Upon exposure to BMP4/FGF2, BMP4 and serum, SOX2 levels drop, whilst levels of MYEF2 remain similar to those observed under proliferative conditions. MYEF2 is also still expressed within the nucleus of terminally differentiated cells. Thus, it is not as simple as MYEF2 simply being a key partner that is expressed only in the stem cells; it likely has broader roles in gene regulation in the differentiating progeny as well.

Several tools to study MYEF2 have been generated in the experiments described in this chapter. *Myef2*-HA NSC line is useful for biochemical experiments and is likely to be used in future experiments such as ChIP-seq. When generating the *myef2*-mCherry NSC line a high level of random integration was observed. Several clones which showed nuclear mCherry staining failed to show targeted knock-in of mCherry

by PCR genotyping. This occurred in 4 out of 7 clones screened. Some studies suggest that dsDNA templates undergo NHEJ in addition to HDR hence may contribute to the high level of random integration observed[256]**.** In future experiments ssODN templates will be used when possible. In addition to the high levels of random integration observed, 2 clones which showed targeted knock-in had large deletions within the untargeted *Myef2* locus. The longer term plasmid based expression of Cas9 and sgRNAs may contribute to this observation, as the recombinant Cas9 complex delivered for HA tagging is likely to reside within cells for a shorter duration**.**

3.3.1 Concluding remarks

To conclude, we observe that the RNA-binding protein MYEF2 interacts specifically with SOX2 in NSCs, and although co-localised with SOX3, SOX8 and SOX9 through SICAP-MS, the interaction with these other SOX members is likely to occur through co-localisation on similar chromatin regions. The interaction between MYEF2 and SOX2 is dynamic, with different subcellular localisation emerging in mitosis, possibly because of SOX2 phosphorylation. In the next chapter we used genetic ablation to determine the function of *Myef2* in NSCs and GSCs.

Chapter 4 The function of MYEF2 in glioblastoma

4.1 Introduction

In the previous chapter we validated the interaction of SOX2 and MYEF2 in both NSC and GSC lines, as well as the initial characterisation of MYEF2 expression in NSCs. This chapter focuses on our experiments to understand the function of MYEF2 in mouse GSCs through loss of function studies, with *Myef2* knockout lines obtained through genetic ablation (CRISPR knock out).

RNA-seq datasets from pan-cancer patient samples, indicate there is high levels of *SOX2* transcripts within GBM and glioma patients (Fig 4.1.A, data from the TCGA[41] and analysis from the cBioPortal[257,258]). The high levels of *SOX2* are observed within almost all GBM and glioma samples, supporting proposals that SOX2 is a core transcription factor in all molecular sub-types of GBM. *MYEF2* transcripts levels are also high within almost all GBM and glioma samples, although levels are also high across a wider range of cancers types than *SOX2* (Fig 4.1.B, data from the TCGA[41] and analysis from the cBioPortal[257,258]). A strong positive correlation between *SOX2* and *MYEF2* transcript levels is also observed within GBM patient samples (Fig 4.2, data from TCGA[259] and analysis from the cBioPortal[257,258]). This supports that MYEF2 may be functioning alongside SOX2 in GBM, and if MYEF2 is therapeutically relevant, likely to be so in a wide range of GBM and glioma patients.

To study the function of MYEF2 in glioblastoma, *Myef2* mutant lines were generated using CRISPR/Cas9. The mouse *Myef2* locus is composed of 17 coding exons, which encode for a protein of approximately 60 kDa predicated to have three RRM RNA/single stranded DNA binding domains (Fig 4.3). Several splice variants of *Myef2*

exist. All splice variants retain exons encoding the RRM domains and vary in the retention of exons 11 and 14. To knockout *Myef2*, guide RNAs which cut within the intronic regions either side of exon 2 were used. This strategy should affect all expressed splice variants of *Myef2*. *Myef2* knockout lines were generated from a new mouse GBM model referred to as NPE (Ester Gangoso, unpublished data). This model contains mutation of *Nf1*, *Pten* and overexpression of human EGFRVIII, best modelling the mesenchymal sub-type of glioblastoma. It should be noted that this exact combination of mutations is rarely observed in human patients and this is a potential limitation of this model. In this chapter the model was generated through a multiplex transfection of multiple guide RNAs, against *Nf1* and *Pten*, and piggyBac overexpression plasmids for EGFRVIII and GFP-Luciferase. Transfected cells were then sorted and transplanted into an NSG mouse. The subsequent tumour was then dissected, and single cell clones were then derived and screen for the relevant modifications. These clones are referred to as NPE-MX-TD ($NF1^{-/-}$, $PTEN^{-/-}$, $EGFRVIII^{+}$ - multiplex - tumour derived) cells with a corresponding clonal ID (Cell line derivation was performed by Ester Gangoso, Pollard lab). This model was chosen for three main reasons: firstly, tumours form rapidly and reliably, making the model useful for *in vivo* studies; secondly, by avoiding the deletion of P53 it is more likely that genomic integrity is maintained and that knockout lines remain isogenic to the parental line, and thirdly this model corresponds to the mesenchymal subtype of GBM.

Generated *Myef2* knockout lines were subsequently used in a range of functional assays to assess the effect that the loss of MYEF2 has on the proliferation rate, colony formation potential, BMP sensitivity and tumour formation potential of GSC cells.

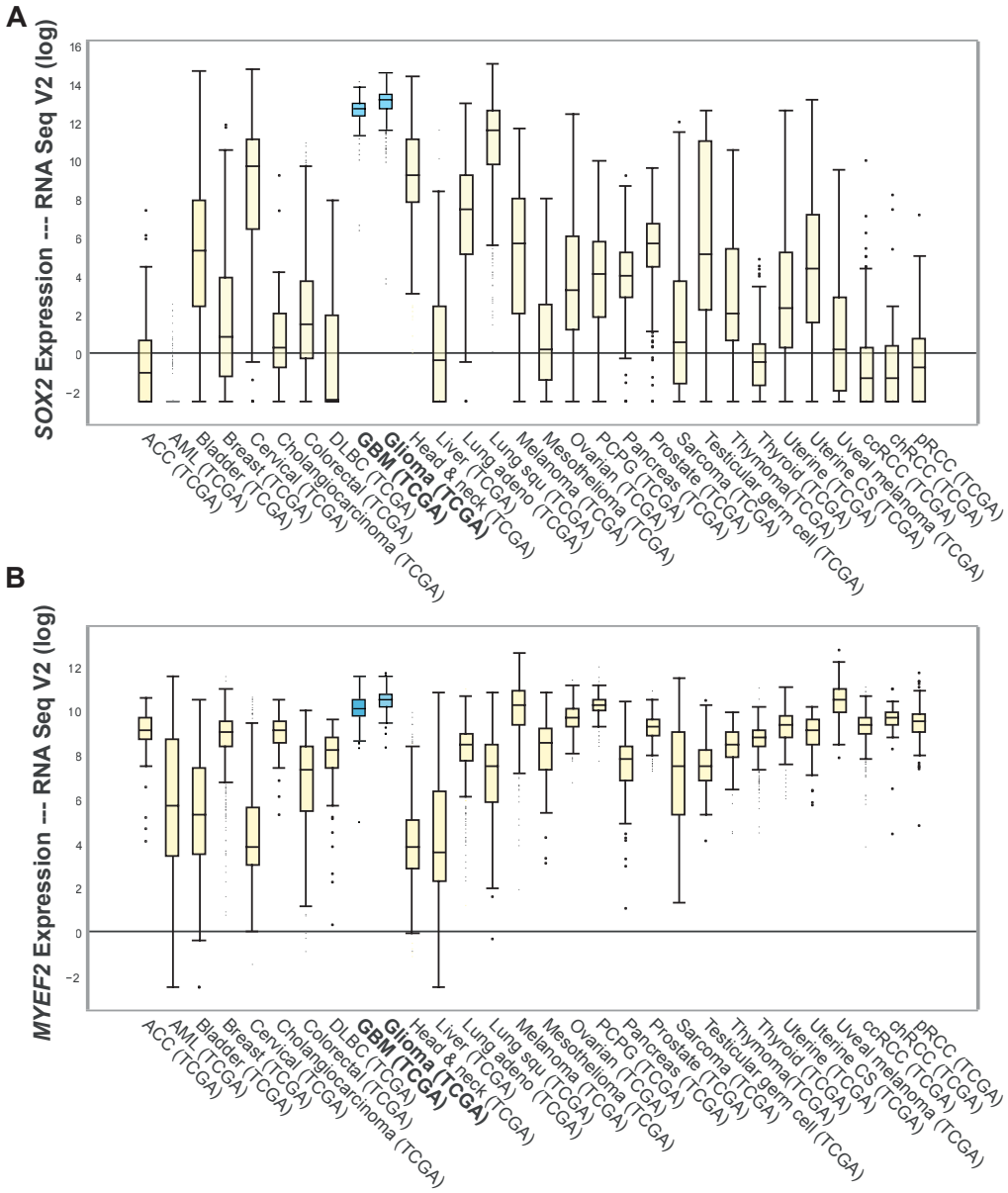


Figure 4.1 | Expression of *SOX2* and *MYEF2* in cancer

(A) *SOX2* RNA-seq data from pan-cancer TCGA data. GBM and glioma are highlighted in blue and in bold. **(B)** *MYEF2* expression from RNA-seq analysis (TCGA data). GBM and glioma are highlighted in blue and in bold. Data for this panel was analysed by the cBioPortal[257,258] and raw data sets from TCGA research network[41].

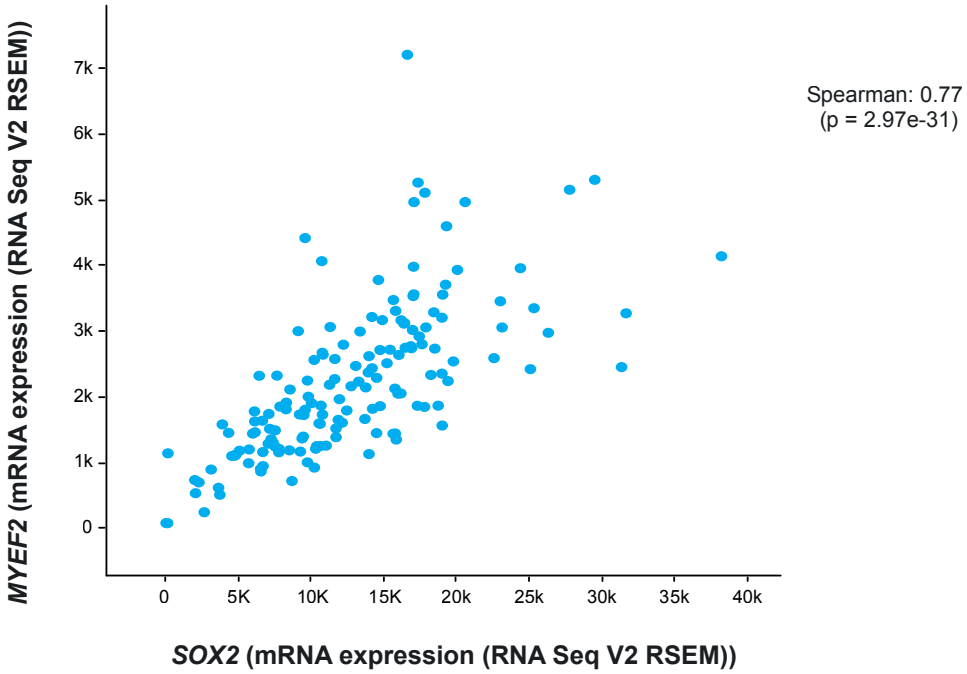


Figure 4.2 | Correlation of *SOX2* and *MYEF2* transcript levels in glioblastoma

RNA-seq data comparing transcript levels of *SOX2* and *MYEF2* in GBM patient samples. Data for this panel is from a TGCA data set[259] and analysed by cBioPortal[257,258].

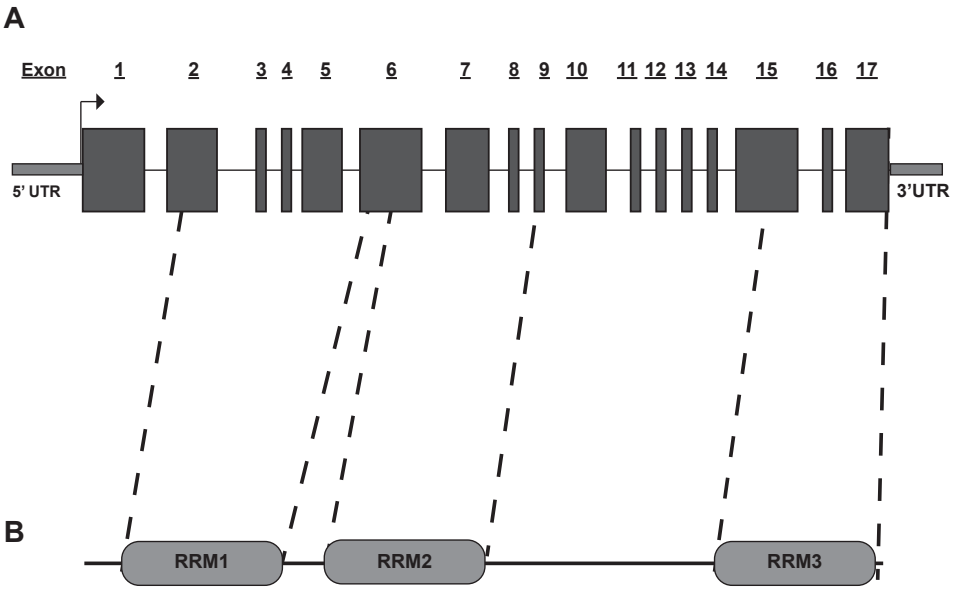


Figure 4.3 | Locus of *Myef2* and predicted protein domains of MYEF2

(A) Depiction of the locus of *Myef2* within the mouse genome. Exons 1 to 17 are depicted as dark grey boxes. **(B)** Depiction of the protein's domains of MYEF2. Corresponding exons encoding the RRM domains are highlighted by dashed lines. Based on exon numbers from *Myef2*-209 transcript exons 2-6 contribute to RRM1, exons 6-9 contribute to RRM2 and exons 15-17 to RRM3. Data for this panel is from Ensembl[241].

4.2 Results

4.2.1 CRISPR-Cas9 mediated generation of mouse GSC *Myef2* knockout lines

Two NPE-MX-TD clonal lines, clone 4 and clone 6, were used to derive MYEF2 knockouts. Two different gRNAs were co-transfected into mouse GSC lines NPE-MX-TD C4 and NPE-MX-TD-C6. The gRNAs were predicted to target intronic regions flanking exon 2, which encodes the majority of the proposed RNA binding domain RRM1, which is retained in all splice variants of *myef2* (Fig 4.3). Single cells were then sorted into 96 well plates by FACS and ~120 clones for each line were picked and genotyped. As two gRNAs were used, the PCR genotyping strategy was designed to identify the deletion of a 403 bp segment, of which 209 bp corresponds to exon 2, of DNA between the gRNAs (Fig 4.4.B). Fragment sizes were based on the assumption that the Cas9: gRNA complex cuts 3 bp upstream of the PAM sequence. While this strategy may underestimate the number of targeted clones; for example, if larger deletions occur which also remove the sequences encoding the genotyping primers, it is simple, relatively fast and sufficient to screen targeted clones.

From the isolated clones, 7 clones of the 125 clones picked from parental line NPE-MX-TD C4 and 8 clones of the 129 clones picked from parental line NPE-MX-TD C6, were identified as knockouts based on PCR genotyping (Fig 4.4B). Clones positive for targeting by genotyping were then verified for the loss of MYEF2 protein by western blotting (Fig 4.4.C). Clones with the exception of two false positives, clone 47 and clone 88, were identified to be *Myef2* knockout lines.

The expression of EGFRVIII is key to the tumour forming potential of the NPE line, with isogenic lines containing only the deletion of NF1 and PTEN (NP line) failing to form tumours. To ensure that the piggyBac integrated human EGFRVIII had not been silenced within individual *Myef2* knockout clones, immunoblotting against phospho-

tyrosine1068 EGFR was performed (Fig 4.4.C). In response to ligand binding and dimerization, intracellular tyrosine phosphorylation occurs within EGFR. As EGFRVIII is truncated and constitutively active[260], the total EGFRvIII protein expression is detected by the pY1068 antibody and can be identified from endogenous EGFR by its reduced size. All mutant lines retained the expression of EGFRVIII, though the levels varied. The MYEF2 knockout clones highlighted in red were used in the subsequent experiments and were selected so range of EGFRVIII levels were represented.

4.2.2 *Myef2* knockout GSCs retain NSC markers and differentiation potential

Myef2 knockout mouse GSC lines can be maintained *in vitro* over several passages, with cells retaining the expression of neural stem cell markers BLBP, NESTIN and SOX2 (Fig 4.5.A). Based on GFAP (astrocytes) and TUJ1 (neurons) (Fig 4.5.B), we found that *Myef2* knockout clonal lines retained the capacity to differentiate into astrocytes and neurons. From initial observations the *in vitro* astrocytic and neuronal differentiation capacity of *Myef2* knockout lines was not dramatically different from the parental line. However, more experiments would be required to determine this, especially due to the low frequency of neuronal differentiation and the high heterogeneity of differentiation experiments. The low efficiency is likely due to a differentiation block imposed by the transforming mutations that the NPE line has been engineered with. Neither the parental line nor the *Myef2* knockout lines differentiated to MBP positive oligodendrocytes.

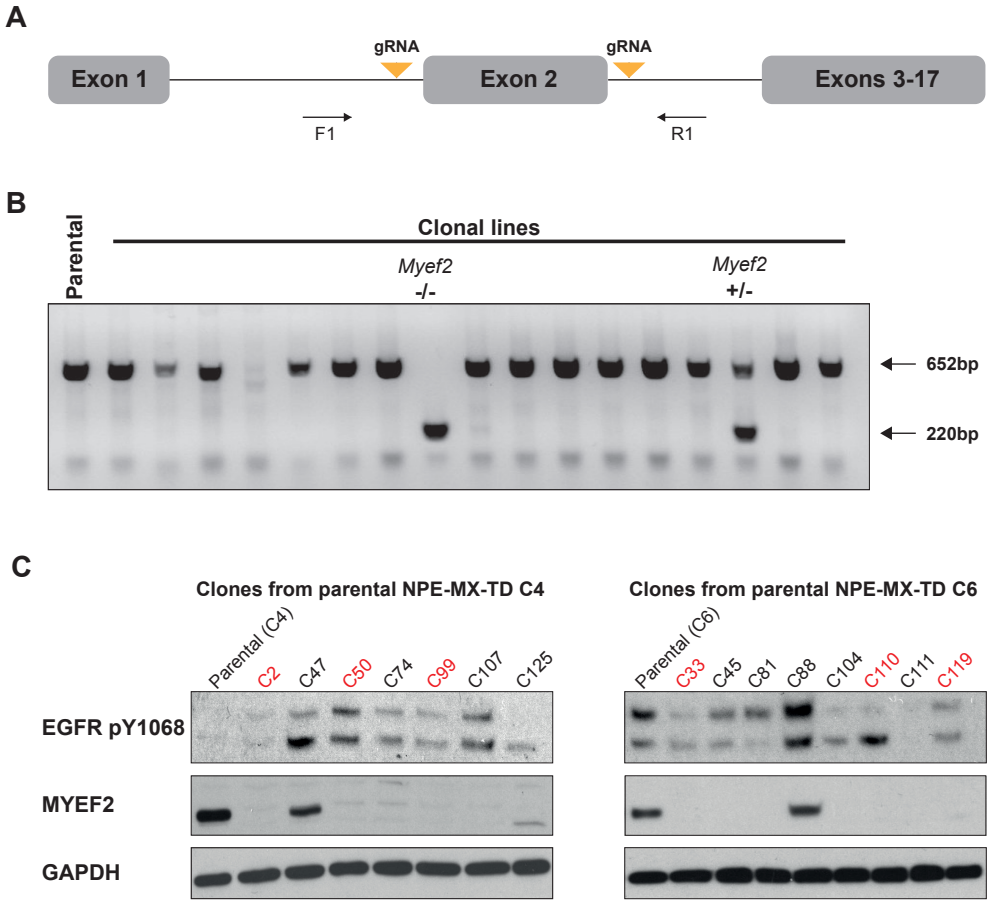


Figure 4.4 | Generation of mouse GSC *Myef2* knockout lines

(A) Schematic of *Myef2* locus. Two guide RNAs, depicted in yellow, were used to generate double strand breaks within the intronic regions flanking either side of exon 2. PCR genotyping primers were designed outside of the cut sites and are depicted as F1 and R1. **(B)** Representative genotyping of NPE-MX-TD C6 clones 26-42 with primers F1 and R1. Sizes of expected wild-type and knockout bands are highlighted. **(C)** Western blotting of clones identified as knockouts by genotyping. Lower band of pY1068 plot corresponds to hEGFRVIII, whilst the higher band to endogenous EGFR. Clones highlighted in red were used in subsequent experiments. Data for panel C was provided by Carla Blin.

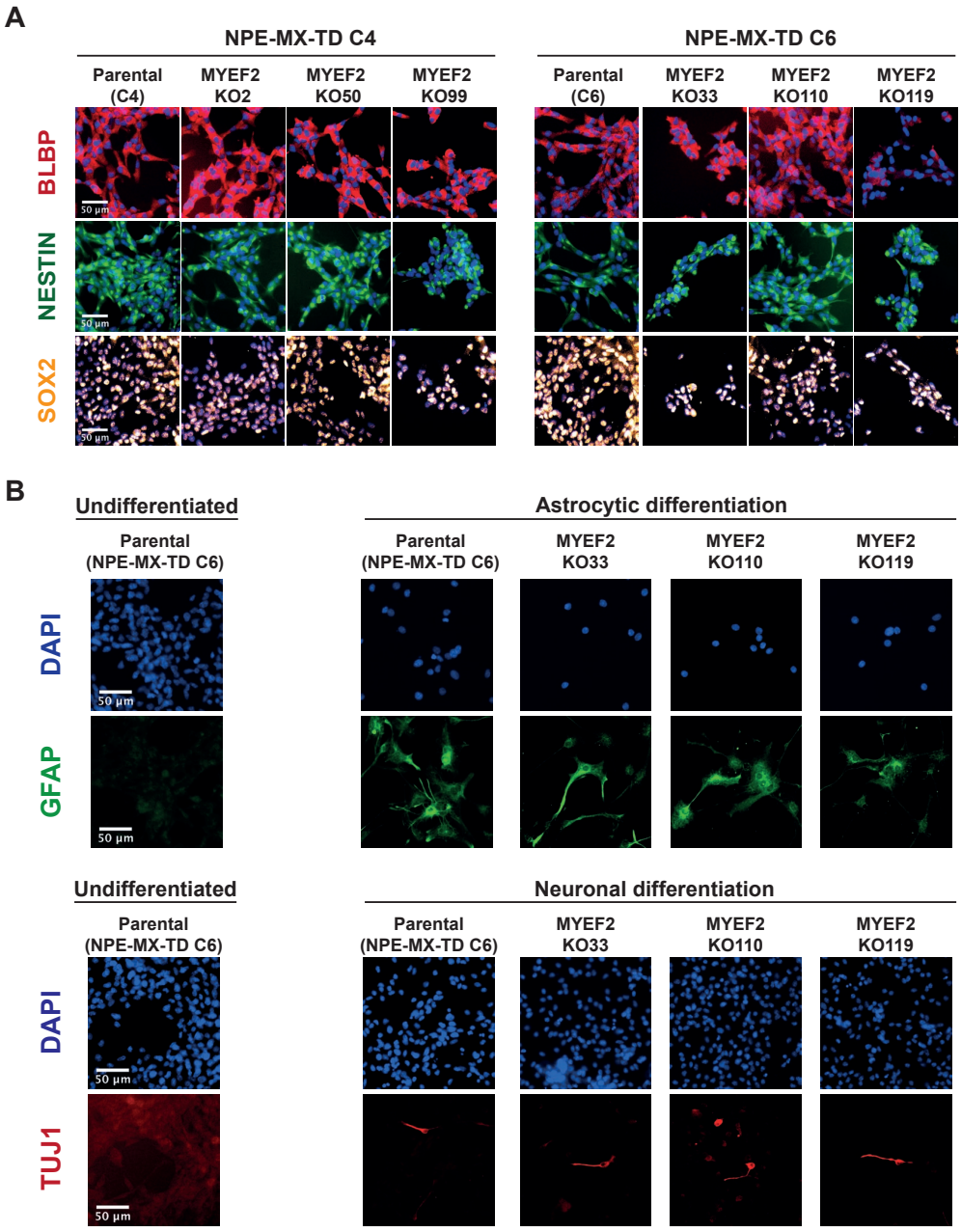


Figure 4.5 | *Myef2* knockout lines retain NSC marker expression and differentiate with low efficiency

(A) Representative ICC images of parental NPE-MX-TD and *Myef2* knockout lines stained for NSC markers BLBP, NESTIN and SOX2 (n=1). Scale bar:50 µm. **(B)** Representative ICC images of parental NPE-MX-TD and *Myef2* knockout lines differentiated to the astrocytic lineage (above) or neuronal lineage (below). Cells were stained for GFAP (astrocytic marker) or Tuj1 (neuronal marker). Undifferentiated parental line provided negative staining control. Scale bar: 50 µm. N=1.

4.2.3 Investigating the proliferation rate of GSC *Myef2* knockout lines

The recovery of mutant *Myef2* cell lines suggested that MYEF2 is not essential for NS cell self-renewal. This is in contrast to SOX2 knockout, which has previously been shown to be critical, and no knockout lines can be recovered[81,261]. We therefore characterised, more carefully, the proliferation rates of the MYEF2 knockout cells.

Using thymidine analogue EdU incorporation assays, we directly compared the growth rates of parental lines to those of *Myef2* knockout lines. Three different clones were tested for both of the independent parental NPE-MX clonal lines. EdU was incorporated into newly synthesised DNA of an asynchronous population during a 2-hour pulse period. Cells were cultured in complete NSC media supplemented with EGF and FGF2. The percentage of cells with EdU incorporation observed in *Myef2* knockout clones 2 (KO2) and 50 (KO50) was very similar to that observed for the parental NPE-MX-TD C4 line, indicating these lines have a similar proliferation rate. However, one clone, the *Myef2* knockout line KO99 had a more drastic reduction in percentage of EdU incorporation compared to the parental NPE-MX-TD C4 line, though unfortunately this was not statistically significant (p value of 0.3, nested t test)

Myef2 knockout clone 33 (KO33) derived from the independent parental line NPE-MX-TD C6 displayed a significant reduction in EdU incorporation compared to the parental line. *Myef2* knockout clones 110 (KO110) and 119 (KO119) did not have a statistically significant reduction in EdU percentage with nested t test p values of 0.24 and 0.11 respectively. Whilst not always statistically significant all *Myef2* knockout clones showed a clear trend of reduction in EdU incorporation, with the fold reductions (based on mean values) varying from 1.02 to 1.6 (Fig 4.6). Whilst there is a high level of variability between knockout clones, the data indicates that the loss of *Myef2* leads to a slight reduction in the proliferation rate.

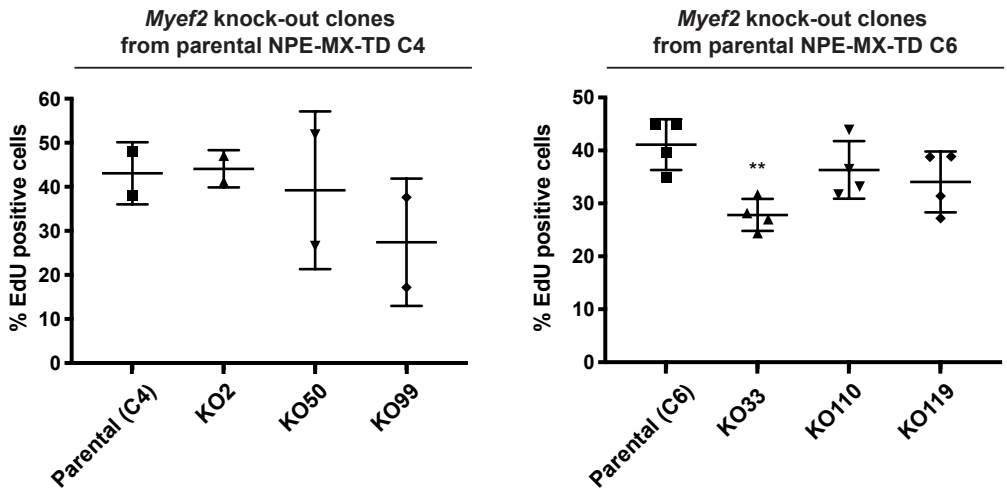


Figure 4.6 | Investigating the proliferation rate of GSC *Myef2* knockout lines

Quantification of the percentage of cells with the integration of thymidine analogue EdU following a 2 hour pulse period. Cells are grown in complete NSC media supplemented with growth factors EGF and FGF2. Plots show mean percentage of EdU positive cells ± SD, with individual technical mean data points shown. For parental line NPE-MX-TD C4 and corresponding knockout lines n=2 biological replicas, with a minimum of 2 technical replicas within each experiment. For parental line NPE-MX-TD C6 and corresponding knockout lines n=4 biological replicas, with a minimum of 2 technical replicas within each experiment. For statistical analysis Nested t tests were performed with each knockout being compared to the relevant parental line. One star represents a *p* value ≤ 0.05, two stars a *p* value ≤0.01 and three stars a *p* value ≤0.001.

4.2.4 Colony formation potential of *Myef2* knockout GSCs

A more stringent test of self-renewal and proliferation capacity is to examine the capability of a single GSCs to produce colonies when plated at a low density. 500 cells were seeded into 10 mm dishes in complete NSC maintenance media with EGF and FGF2. Colonies were allowed to form over 10-15 days and were the quantified. The number of colonies observed was highly variable between experiments, however there was a clear trend, within each experiment, in which NPE-MX-TD *Myef2* knockout GSCs form fewer colonies than the corresponding parental NPE-MX-TD (Fig 4.7. A). In addition to forming fewer colonies, the colonies of *Myef2* knockout GCSs appeared to be smaller. In an attempt to represent and quantify this the Fiji plugin ColonyArea was used, which calculates the plate area occupied by the colonies, giving an indication of the size of the colonies, and the intensity, which indicates the number of cells within a colony. For parental NPE-MX-TD C4 and the corresponding *Myef2* knockouts this emphases the differences when MYEF2 is lost while, for C6 the differences were similar to those observed by the colony number alone (Fig 4.7.B and C). To conclude, preliminary experiments suggest that the loss of MYEF2 causes GSCs to be less clonogenic.

4.2.5 BMP4 sensitivity of *Myef2* knockout GSCs

In NSCs and GSCs, BMP4 can mediate the upregulation of astrocytic markers, such as GFAP, and cell cycle exit. The degree of response varies both between NSCs and GSCs. GSCs generally considered to be more resistant to the cytostatic signals of BMP4 than NSCs, and response level between each GSC line can vary dramatically[262]. In response to BMP4 signalling, a high percentage of NPE-MX-TD and *Myef2* knockout NPE-MX-TD cells increase the expression of astrocytic marker GFAP (Fig 4.8). To assess the sensitivity of cells to the cytostatic response of BMP4

signalling, the degree of EdU incorporation in varying BMP4 concentrations was measured. Following a 2 hr pulse of EdU, we observed that, whilst not significant (Mann Whitney U test, p value <0.05), there is a general trend for *Myef2* knockout cells to have a lower percentage of EdU incorporation relative to the corresponding parental NPE-MX-TD cells (Fig 4.8.A), though as previously observed there is some variability in degree of response between clones. This suggests that *Myef2* knockout GSCs are more sensitive to BMP4, though EdU incorporation levels did not drop to those of normal NSCs. NPE-MX-TD C6 had a higher percentage of EdU positive cells relative to NPE-MX-TD C4, indicating NPE-MX-TD C6 is less responsive to the cytostatic signals of BMP4. This observation is in line with the higher levels of *Bmpr1* expressed in NPE-MX-TD C4 (*Bmpr1* expression levels observed by qPCR, Kirsty Ferguson, Pollard lab).

To ensure that slow cycling cells were also scored, a 24hr pulse with EdU was performed (Fig 4.8.B). Levels of EdU incorporation were lower in *Myef2* knockouts compared to the corresponding parental line, though this trend was not statistically significant (Mann Whitney U test, p value <0.05). Parental NPE-MX-TD C6 had a 14% incorporation rate of EdU, while NPE-MX-TD C6 *Myef2* knockouts a range of 3.5%, 2.1% and 4.6%. NPE-MX-TD C4 had a 4.3% incorporation rate, while *Myef2* knockouts a 1%, 2.3% and 1.5% incorporation. NSCs had a 0.5% incorporation rate of EdU. This data suggests that, the loss of MYEF2 increases the sensitivity of GSCs to the cytostatic signals of BMP, though further replicas are required to confirm this.

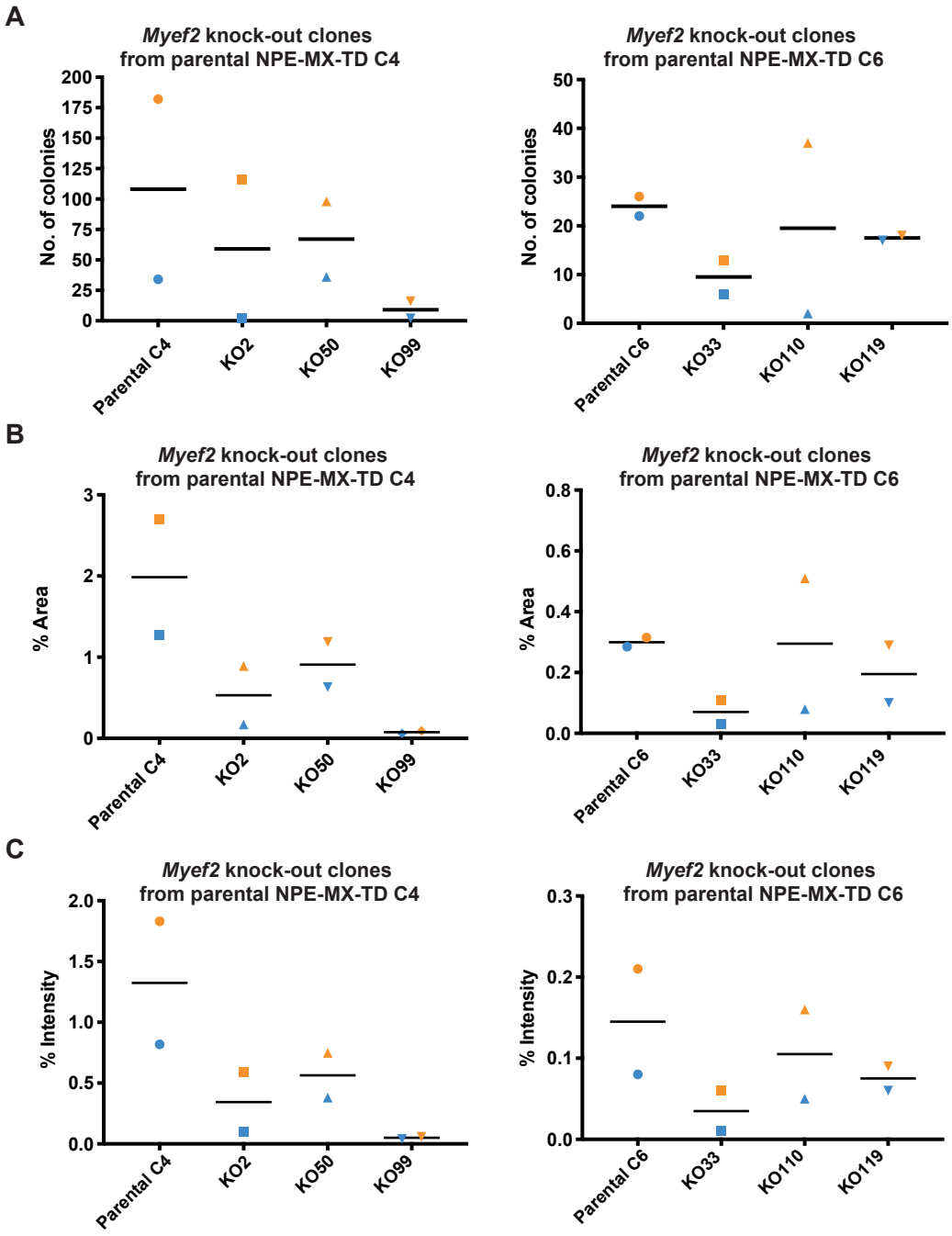


Figure 4.7 | Colony formation potential of *Myef2* knockout GSCs

(A) Quantification of the number of colonies formed from the corresponding cell lines. Single cells were seeded and colonies that had formed after 10-15 days were stained with 0.01% of methyl blue. Graph shows mean colony number from 2 replicas, with individual data points shown (orange= replica 1, blue= replica 2 data points). **(B)** and **(C)** show the % area that colonies occupy and the intensity, correspondingly, calculated using the fiji plugin ColonyArea[233]. Graph shows mean from 2 replicas, with individual data points shown (orange= replica 1, blue= replica 2 data points).

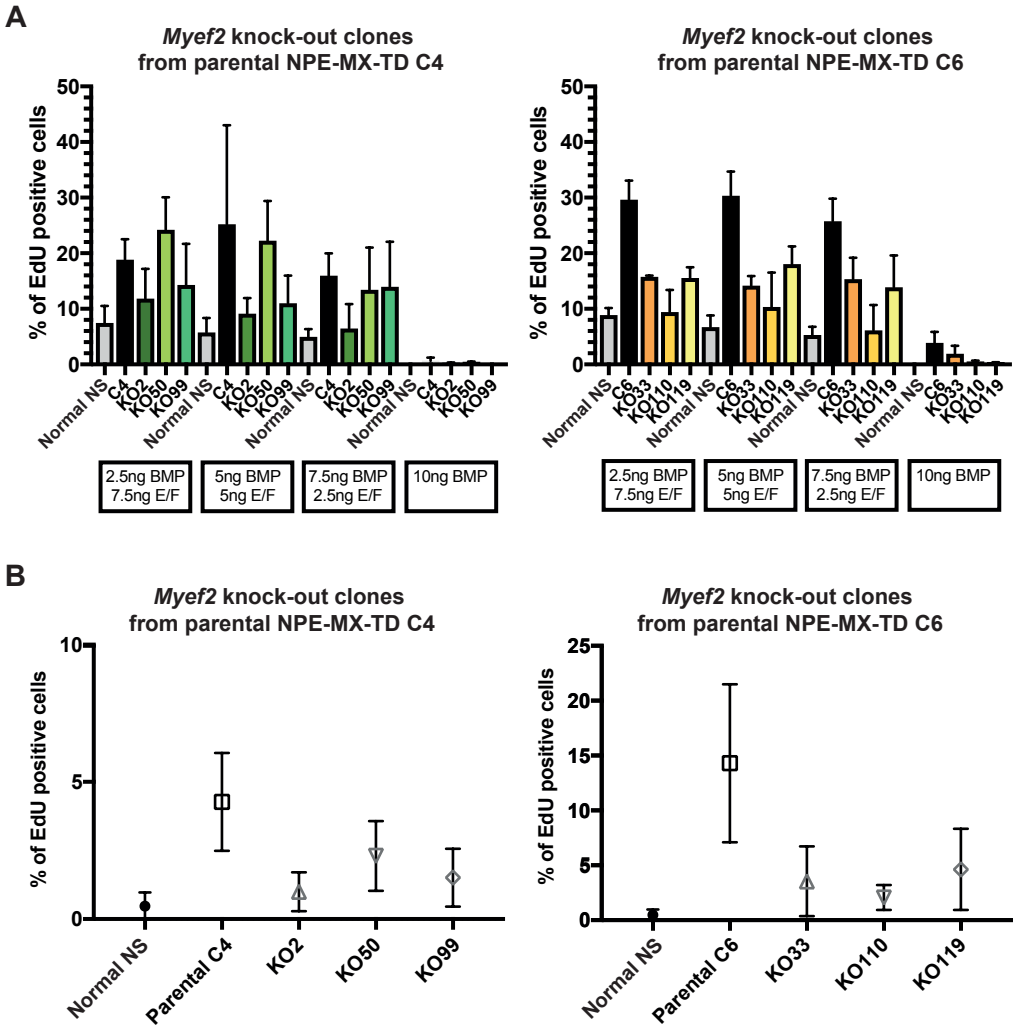


Figure 4.8 | BMP4 sensitivity of *Myef2* knockout GSCs

(A) Quantity of proliferative cells in varying concentrations of EGF, FGF2 and BMP4. Graph shows percentage of EdU positive cells after a 2 hr pulse period in normal NS, parental NPE-MX-TD and corresponding *Myef2* knockout cells treated for 3 days in varying concentrations (ng/ml) of growth factors EGF/FGF2 and differentiation promoting BMP4. Plot shows the mean percentage ± SD from 2 replicas. **(B)** Quantification of proliferating cells after treatment with BMP4. Normal NS, parental NPE-MX-TD and *Myef2* knockout cells were treated with BMP4 (10 ng/ml) for 3 days. Percentage of EdU after a 24 hr pulse period was then calculated. Graph shows mean percentage of EdU positive cells ± SD from 3 replicas.

4.2.6 Re-entering the cell cycle from dormancy (G0)

Previous studies have shown that GSCs are not terminally differentiated by BMP4[34] and are capable of returning back into the cell cycle after BMP4 treatment[262]. Recent studies from Maria Angeles Marques in fact suggest NSCs treated with BMP4 enter a dormant (G0) state (unpublished observation), To assess if GSCs lacking MYEF2 are in a deeper G0 state than cells with MYEF2, cells were treated with BMP4 for 3 days, then re-added to NSC maintenance media supplemented with growth factors EGF/FGF2 for 10-15 days, till colonies formed (Figure 4.9). As expected, no colonies were observed from normal NSCs (non-cancerous). We observe that parental NPE-MX-TD C4 GSCs forms few colonies and the corresponding *Myef2* knockouts almost none (Fig 4.10.A). Parental NPE-MX-TD C6 GSCs form many more colonies when returning to proliferative conditions, whilst *Myef2* knockouts derived from the NPE-MX-TD C6 parental lines form very few colonies (Fig 4.10.B).

As there are differences in the BMP sensitivity (4.2.5) and colony formation potential (4.2.4) of *Myef2* knockout lines it is possible the differences observed in this assay arise from a combination of these factors, as opposed to differences in re-entry from dormancy. In an attempt to quantify this, the differences between BMP sensitivity (Fig 4.8) and colony formation potential (Fig 4.7) of *Myef2* knockout lines compared to the parental lines were calculated and combined. This gave rise to the expected ratio, the difference in colony number expected to be observed between parental and *Myef2* knockouts assuming no differences in the ability to re-enter cell cycle exists. Expected ratio differences of parental compared to *Myef2* knockouts were then compared to the observed ratio differences (the differences in experimentally observed colonies numbers of *Myef2* knockout compared to the parental line). These are presented in table 4.1 for C4 and table 4.2 for C6.

Table 6: Predicted and observed colony numbers for NPE-MX-TD C4 and corresponding *Myef2* knockouts GSCs

Cell line	Observed colony number (average)	Expected ratio (relative to parental)	Observed ratio (relative to parental)
NPE-MX-TD C4	11	1	1
KO2	1	1/8	1/11
KO50	6	1/3	6/11
KO99	1	1/33	1/11

Table 7: Predicted and observed colony numbers for NPE-MX-TD C6 and corresponding *Myef2* knockouts GSCs

Cell line	Observed colony number (average)	Expected ratio (relative to parental)	Observed ratio (relative to parental)
NPE-MX-TD C6	37	1	1
KO33	3	1/10	3/37
KO110	2	1/8	2/37
KO119	2	1/3	2/37

Generally, the observed reduction in the colony numbers of *Myef2* knockout clones is higher than the expected ratio. This suggests that the loss of MYEF2 reduces the ability of GSCs to re-enter the cell cycle after BMP4 treatment. However, these data were highly variable. We can therefore not yet conclude that re-entry from dormancy requires MYEF2. Future experiments using increased replicates and/or rescue experimentations are required for this. Additionally, exploration of changes to the epigenetic landscape including the methylation and chromatin accessibility profile of cells, especially at genes associated with astrocytic differentiation (such as *S100A6*),

may provide alternative methods to indicate the extent of altered differentiation/G0 depth vs de-differentiation/re-entry to the cell cycle[38,86].

4.2.7 Tumour formation potential of *Myef2* knockout GSCs

Parental NPE-MX-TD lines and corresponding NPE-MX-TD MYEF2 knockout lines were transplanted into the frontal cortex of adult male NOD scid gamma (NSG) mice. NPE-MX-TD lines, and corresponding *Myef2* knockouts, constitutively express a randomly integrated GFP-luciferase construct, allowing for *in-vivo* imaging and identification of injected cells *ex-vivo*. Upon the development of symptoms, such as fitting, mice were culled and brains dissected to asses for the formation of tumours. All mice injected formed a tumour at the site of injection. Tumours from parental lines and *Myef2* knockout lines appeared to be similar in size and in the degree of infiltration (Fig 4.11.A and Fig 4.12.A). Survival curves indicate that *Myef2* knockout lines have slower tumour growth than corresponding parental lines. The median survival time of parental NPE-MX-TD C4 was 13 days, whilst the median survival times of MYEF2 knockout clones 2 and 50 were 21 and 23 days respectively. With parental NPE-MX-TD C6 a median survival time of 15 days was observed, with median survival times of 17 and 18 days observed for *Myef2* knockout clones 33 and 110 respectfully (Fig 4.11.B and Fig 4.12.B). The observed differences between *Myef2* knockout lines and the isogenic parental line are significant, suggesting that the loss of *Myef2* slows the tumour formation rate of mouse glioma neural stem cells.

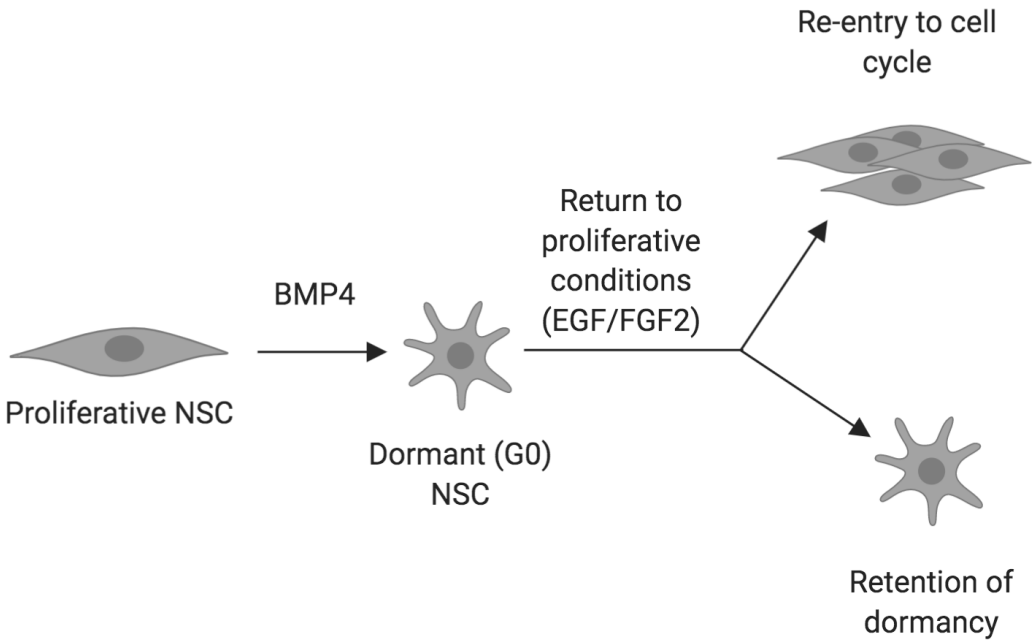


Figure 4.9 | Schematic of re-entry to the cell cycle from dormancy (G0)

Schematic of the re-entry into the cell cycle assay performed in section 4.2.6 and figure 4.10. Cells are treated for 3 days with 10 ng/ml BMP4 inducing the expression of astrocytic marker GFAP, changes in cellular morphology and a dormant state (G0). After 3 days, cells are returned to proliferative conditions (NSC maintenance media with growth factors).

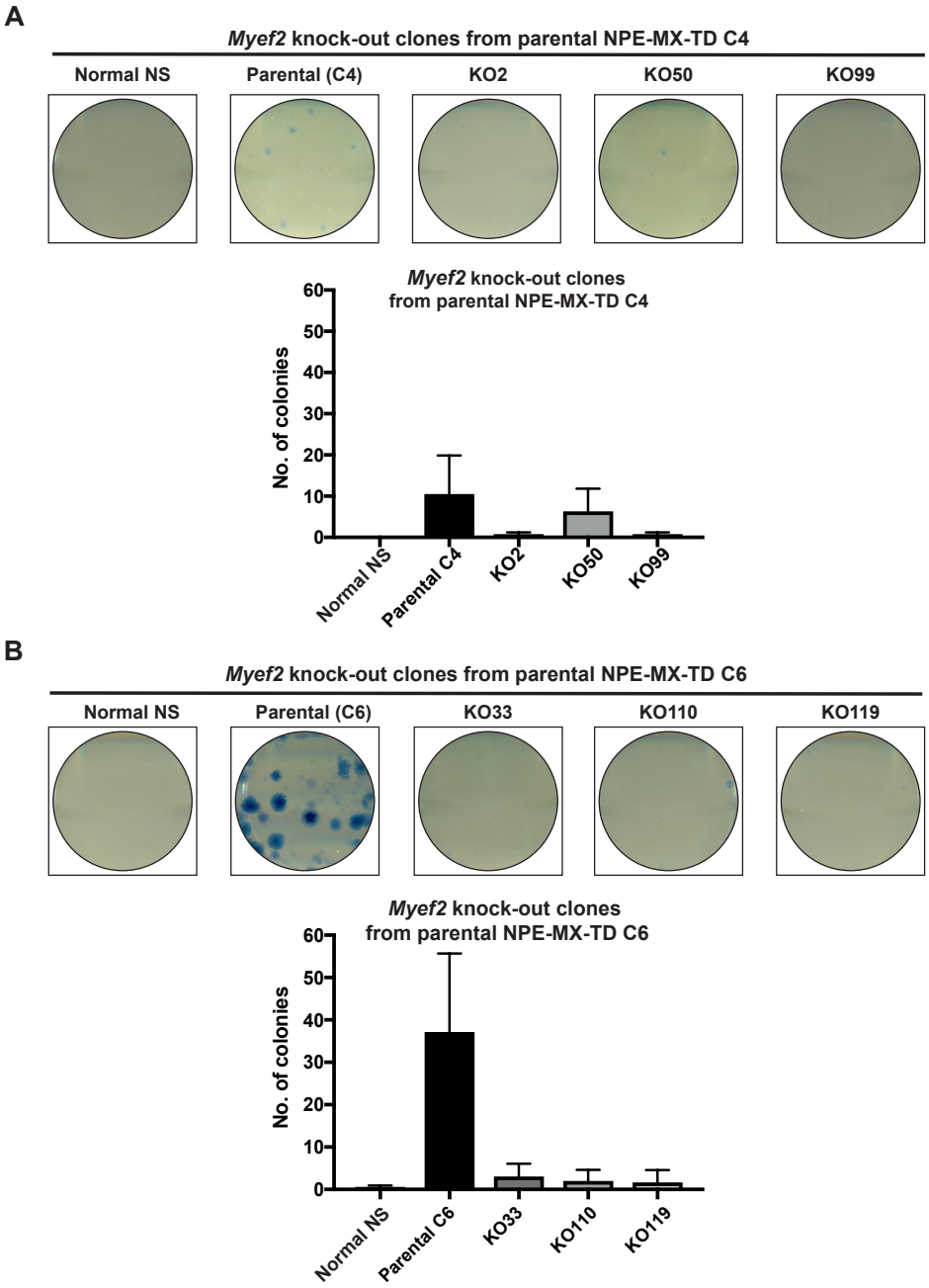


Figure 4.10 | Exit from dormancy (G0) of mouse *Myef2* knockout GSCs

Cells were plated in BMP4 for 3 days to induce differentiation and exit from the cell cycle. Subsequently, cells were reintroduced to proliferative NSC conditions (NSC maintenance media supplemented with EGF/FGF2) and colonies were allowed to form. Colonies were then fixed, stained with 0.01% methyl blue and colony numbers quantified. Example images and colony quantification for **(A)** NPE-MX-TD C4 GSCs and corresponding knockouts and **(B)** NPE-MX-TD C6 GSCs and corresponding knockouts. Graph shows mean number of colonies ± SD from 3 replicas.

4.2.8 Rescue of *Myef2* knockout GSCs

To more robustly asses that previous observations from the derived knockout clonal lines are only due to the loss of MYEF2, and not additional unintentional difference between KO and parental GSCs, rescue lines were generated in knockout clones KO110 and KO119. A *Myef2*-HA-IRES-puromycin selection cassette under control of the TRE3G promoter (Fig 4.13.A) was randomly integrated into cells co-transfected with a constitutive TET-On-3G vector. Selection for transfected cells in doxycycline and puromycin was performed for 5 days. All control cells, treated with puromycin in the absence of doxycycline, died by day 3 of selection. Surviving cells were expanded and checked for the inducibility of HA. A very low percentage of HA positive cells was observed, approx. 3% for KO110 and 5% for KO119 (Fig 4.13.B). This was unexpected based on the selection. The cells were then plated in BMP4 in the presence or absence of doxycycline for 3 days, then subsequently retuned to growth factors in presence or absence of doxycycline and colonies allowed to form for 10 days (as described in section 4.2.6). In all lines colonies formed at a similar rate in the presence and absence of doxycycline (Fig 4.13.C). Some colonies formed in *Myef2* knockouts in both presence and absence of doxycycline, but none of these stained positive for HA (data not shown). From these observations it was not possible to conclude that the reintroduction of exogenous *Myef2* had rescued the previously observed phenotype.

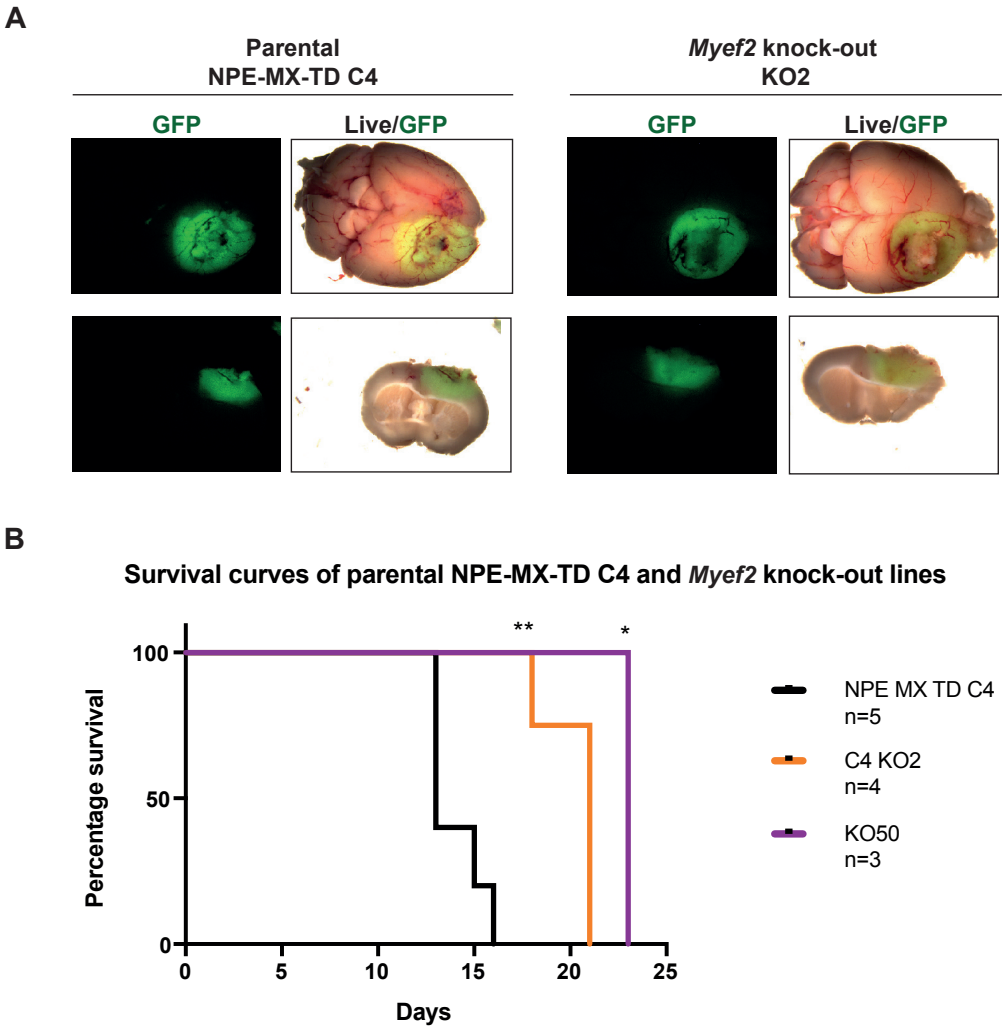


Figure 4.11 | Tumour formation of *Myef2* knockout lines derived from parental NPE-MX-TD C4 line

(A) Representative stereo microscope images of brains dissected from post-transplantation symptomatic NSG mice. NPE-MX-TD C4 line and corresponding *Myef2* knockout lines are engineered to constitutively express GFP-luciferase, hence tumours are labelled with GFP. **(B)** Survival curves of NSG mice after transplantation with parental NPE-MX-TD C4 line (black), *Myef2* knockout line clone 4 (KO4, orange) or *Myef2* knockout clone 50 (KO50, purple). P values were generated from the log-rank (Mantel-Cox) statistical test. One star represents a *p* value ≤ 0.05, two stars a *p* value ≤0.01 and three stars a *p* value ≤0.001.

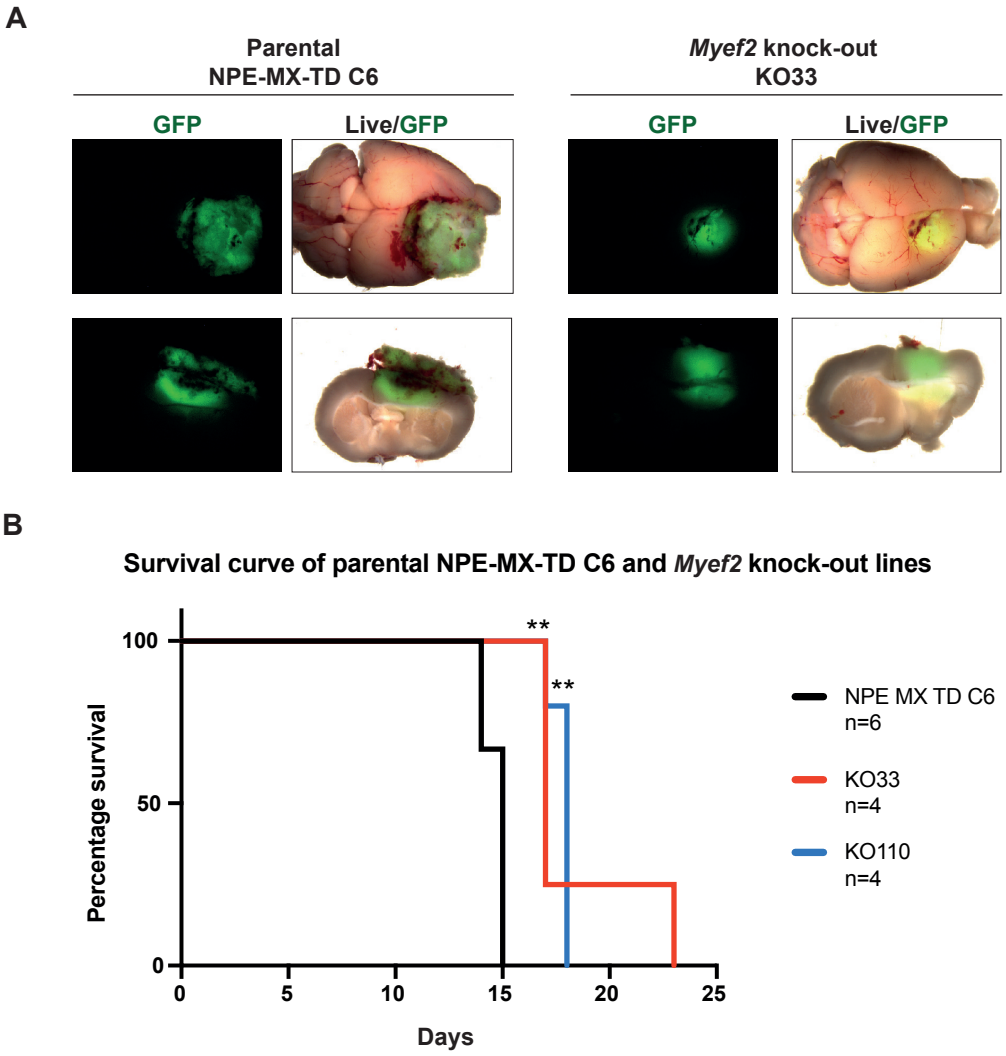


Figure 4.12 | Tumour formation of *Myef2* knockout lines derived from parental NPE-MX-TD C6 line

(A) Representative stereo microscope images of brains dissected from post-transplantation symptomatic NSG mice. NPE-MX-TD C6 line and corresponding *Myef2* knockout lines are engineered to constitutively express GFP-luciferase, hence tumours are labelled with GFP. **(B)** Survival curves of NSG mice after transplantation with parental NPE-MX-TD C6 line (black), *Myef2* knockout line clone 33 (KO33, red) or *Myef2* knockout clone 110 (KO110, blue). P values were generated from the log-rank (Mantel-Cox) statistical test. One star represents a *p* value ≤ 0.05, two stars a *p* value ≤0.01 and three stars a *p* value ≤0.001.

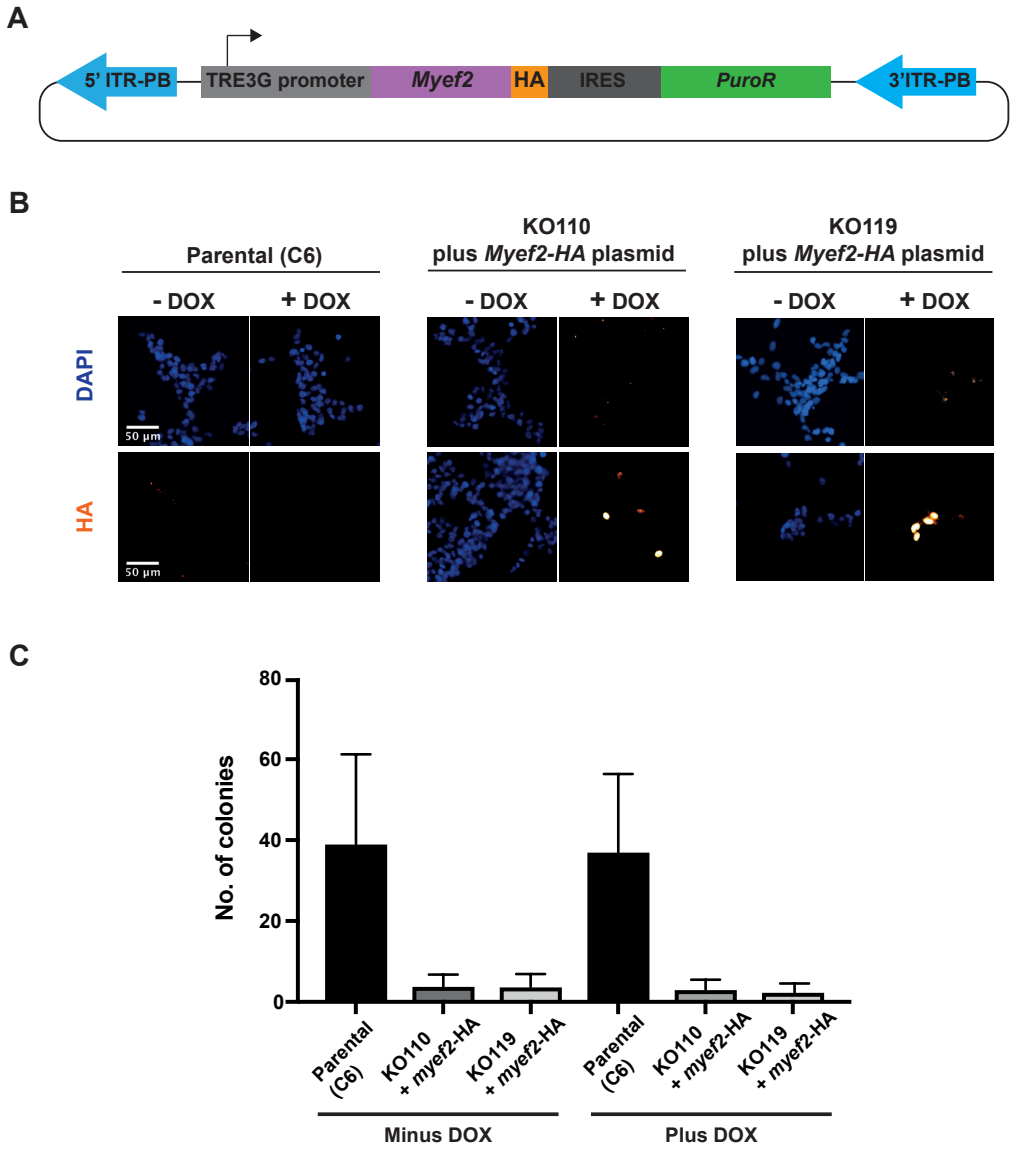


Figure 4.13 | Inconsistent *Myef2* expression hampered attempted exogenous rescue of MYEF2 knockout GSCs

(A) Depiction of the piggyBac integration plasmid used to exogenously express a single *myef2* isoform. Plasmid was assembled using EMMA. **(B)** Representative ICC images of MYEF2 knockout lines KO110 and KO119 transfected with *myef2* plasmid and selected for stable piggyBac integration with puromycin. Parental line was not transfected and provided a negative control. Scale bar: 50 µm. **(C)** Number of colonies observed after cells were treated with BMP4 for 3 days and subsequently returned to growth factors EGF/FGF2 for 10 days. Graph shows mean± SD from 3 replicas.

4.3 Discussion

From the generation of isogenic *Myef2* knockout GSCs we were able to begin assessing the function of MYEF2 in glioblastoma. We observed that the loss of MYEF2: reduces the *in vitro* proliferation rate; reduces the colony formation potential and increases the BMP4 sensitivity of GSCs *in vitro*. *In vivo* tumour formation potential is reduced when MYEF2 is ablated in mouse GSCs.

Quantification of EdU incorporation highlighted that, the loss of MYEF2 causes a minor reduction in the proliferation rate of GSCs. This observation was only significant (*p* value <0.05 nested t test) for one clone, with all clones showing a mean decrease in EdU percentages. The degree of reduction in proliferation varied between each individual knockout clone, which may either reflect inherent cell variability acquired throughout cell culture or, more likely, the acquisition of different CRISPR/Cas9 off-target effects when generating the *Myef2* knockout lines. The observed reduction in proliferation was generally a relatively modest reduction, while the loss of SOX2 is predicted to mediate a greater effect on the proliferation rate, though the effect of SOX2 loss on the NPE model has not been directly tested. This indicates that either MYEF2 is not essential to all functions of SOX2 or there is functional redundancy between MYEF2 and another unknown protein.

Preliminary experiments indicated that the colony formation potential of *Myef2* knockout lines is reduced. In addition to reduced colony numbers, the colony sizes of *Myef2* knockout cells were also generally smaller, which is likely to reflect the reduced proliferation rates observed upon the loss of MYEF2. Unfortunately, the differences in clonogenic potential observed between parental and *Myef2* knockout were not significant (Mann U test *p*<,0.05). The clonogenic assay used was highly variable between experiments and failed to form colonies during attempts to obtain further

replicas. In order to conclusively state that the loss of MYEF2 causes GSCs to form fewer colonies, it will be necessary to further optimise the clonogenic assay most likely by varying: the plating density; tissue culture plastic, and pre-coating treatment of tissue culture plastic.

GSCs are often less sensitive than normal NSCs to the cytostatic effects of BMP4. In the absence of MYEF2 we observe fewer EdU positive cells when cells are treated with BMP4, indicating that MYEF2 contributes to cytostatic resistance. Fold differences of between 2-fold and 7-fold were observed when 10ng/ml BMP4 were used (Fig 4.8). In general, NPE-MX-TD C4 and the corresponding *Myef2* knockouts lines were more sensitive to the cytostatic cues of BMP4 than NPE-MX-TD C6 and the derived knockouts. This is likely to reflect the higher initial expression levels of *Bmpr1* that are found in NPE-MX-TD C4. The increased sensitivity of *Myef2* knockout lines, may make GSCs lacking MYEF2 more sensitive to differentiation therapy, though the levels of cells which had exited the cell cycle did not drop to those observed in normal NSCs (Fig 4.8).

The loss of MYEF2 significantly increased the survival times of transplanted mice. This slower rate of tumour formation was observed for all *Myef2* knockout clones tested. The degree to which survival is affected by MYEF2 loss does not appear to correlate with some of the *in vitro* observations. *Myef2* knockout clone KO50 had the greatest increase in median survival time compared to the parental line NPE-MX-TD C4, whilst from the *in vitro* data this clone had the most modest reductions in proliferation, colony formation potential and BMP sensitivity. *Myef2* knockout clone KO33 on the other hand had some of the strongest *in vitro* differences compared to the parental line but showed the smallest increase in median survival times. The observed differences between *in vitro* and *in vivo* responses, highlight some of the

limitations of *in vitro* studies, which fail to investigate many components involved in tumour formation capacity including immune evasion, angiogenesis, invasiveness, the microenvironment and the brain extracellular matrix (ECM). Histology and immunohistochemistry (IHC) on control and *Myef2* knockout tumours will provide insight into several of these factors, with staining of blood vessels, e.g. through CD31 staining, indicating angiogenesis and staining for GFP, which marks the injected cancer cells, highlighting the degree of infiltration and invasion. As NSG mice, which only have functional neutrophils and monocytes and lack functional macrophages, dendritic cells, mature B/T cells, natural killer (NK) cells, were used for the *in vivo* tumour formation capacity experiments described in this thesis, investigations into the effect of MYEF2 loss on immune evasion will be limited. However, as the NPE-MX-TD model is capable of forming tumours within the immune competent C57BL/6J strain, additional orthotopic transplantations experiments would enable for better investigation into the potential effects of MYEF2 loss on immune evasion and the tumour microenvironment. Overall in the absence of MYEF2, GSCs form tumours at a modestly slower rate.

To support that the observations occurring are purely from the loss of MYEF2, and no other off-target effect of CRISPR/Cas9, we attempted rescue experiments to return MYEF2 expression to the knockout lines. Puromycin selection only gave a small population of cells that could induce immunocytochemistry observable MYEF2-HA protein – perhaps because *Myef2* overexpression cannot be tolerated. As expected, based on the low percentage of the inducible population, no clear differences in colony numbers were observed between induced and uninduced samples when cells were tested for their ability to exit from the dormant state. It is necessary to repeat this experiment with a purer population in which more cells are inducible. To achieve this, it may be best to produce a new line in which fluorophore, so can sort without keeping

levels on for a long time in case of dosage issues. Other issues including our use of non-physiological levels and the use of only one isoform. These may also contribute to the negative result of this preliminary experiment.

4.3.1 Concluding remarks

To conclude, whilst not always statistically significant, all data indicates that the loss of MYEF2 is unfavourable for GSCs. This is in line with the high levels observed within GBM and suggest that the interaction of MYEF2 with SOX2 may be functionally important.

Chapter 5 Tuneable degradation of transcription factors using SMASh

5.1 Introduction

By engineering a degron onto a protein of interest (POI), the levels of the engineered protein can be controlled post-translationally. In this chapter I focus on of the drug tuneable SMASh[230] degron (Figure 1.6)

Whilst many different experimental degrons have been reported, the SMASh degron has many advantages. Firstly, self-cleavage of SMASh from the C-terminal of the POI occurs in the non-drug treated state, thereby leaving the target protein minimally modified with only 8 residual amino acids (PGDEMEEC). This is unlike other degrons, such as the auxin system, which require the permanent addition of large tags to the POI, which can often interfere with protein function. Secondly, the HCV NS3/4A inhibitor drugs, which inhibit the self-cleavage of the SMASh and hence mediate degradation of the POI, are FDA approved drugs which can be used *in vivo* studies. This is a major advantage when transitioning to *in vivo* systems as dosing and toxicity issues have already been studied. Recent studies have shown that, through the introduction of multiple mutations to the NS3 protease, variable sensitivity to different NS3 protease inhibitors can be generated[263]. These NS3 protease variants can be used to generate SMASh tags inhibited by different drugs, highlighting the potential to specifically target multiple proteins within the same population by using different inhibitors. Thirdly, only one engineering event is required when using the SMASh degron. Other systems are often bipartite and require significant optimisation of the levels. Due to these advantages, in addition to those generally achieved by degron

control, we were highly interested in assessing the control achieved by the SMASh degron.

Using *Sox2* I assessed the suitability of SMASh as a degron in mouse NSCs. SOX2 is a key transcription factor in GBM and we have established methods to knock-in small and large tags to this locus [81,185], making *Sox2* an obvious choice for this proof of principle study.

We tested the ability of SMASh to control both transgenic and endogenous SOX2 protein levels *in vitro.* Finally, to further understand the biological effect of endogenous SOX2 loss, in NSCs and GSCs, bi-allelic knock-in of the SMASh tag to *Sox2* locus is required. Hence, I will describe a strategy which should enrich for the selection of cells with bi-allelic SMASh knock-in.

5.2 Results

5.2.1 SMASh control of a SOX2-mCherry transgene confirms controllable protein degradation in mouse neural stem cells

We first generated two new plasmids containing either a doxycycline inducible *Sox2*-mCherry-SMASh transgene or a control doxycycline inducible *Sox2*-mCherry transgene. A SOX2-fluorophore fusion protein was chosen in order to allow for more quantitative results, such as data obtained through flow cytometry. These plasmids were used to generate a stable transgenic *Sox2*-mCherry-SMASh mouse NSC line (termed, NSC-iS2C-SM) and a stable transgenic *Sox2*-mCherry mouse NSC control line (termed, NSC-iS2C). The addition of NS3/4A protease inhibitor asunaprevir (ASV) to these two lines, when induced to express the transgene, lead to the absence of SOX2-mCherry within the NSC-iS2C-SM line, but not within the control NSC-iS2C

line (Fig 5.1.A). This validated that SMASh is capable of controlling transgenic SOX2-mCherry within mouse NSCs.

Next, to determine the concentrations of ASV which give a tuneable range of SOX2 degradation, NSC-iS2C-SM and NSC-iS2C lines were induced with doxycycline, and simultaneously treated with ASV concentrations varying from 0 μM to 1 μM (Fig 5.1.B and Fig 5.1.C). From western blotting, the addition of 1 nM ASV caused levels of SOX2-mCherry to drop and, at 1 μM of ASV SOX2-mCherry appears to be absent from the NSC-iS2C-SM line (Fig 5.1.B). It should also be noted that, only a very faint band was observed at 100 kDa within the NSC-iS2C-SM line in the absence of ASV. This is the predicted size of non-cleaved SOX2-mCherry-SMASh fusion protein, highlighting the efficient self-cleavage of SMASh from SOX2-mCherry in the absence of ASV.

To obtain more quantitative results flow cytometry was performed (Fig 5.1.C). Once again, the tuneable range appears to be between 1 nM and 1 μM of ASV, with a drop in protein levels of 6% and 89% respectively, being observed in the NSC-iS2C-SM line, when normalising to the induced line without ASV. 1 μM and 10 μM of ASV caused a drop of 87.4% and 89% in SOX2-mCherry protein levels within the NSC-iS2C-SM line, though levels were not reduced to those observed under uninduced conditions. This highlights the presence of residual protein not observed by western blotting (Fig 5.1.C). Whilst 10 μM of ASV gave the largest drop in SOX2-mCherry from the NSC-iS2C-SM line, toxicity was indicated by a 57.4% reduction of SOX2-mCherry in the control NSC-iS2C line. This toxicity of ASV at 10 μM, was further validated by growth rate analysis in unmodified mouse NSC and GSC lines (Fig 5.2). No alteration in growth rates were observed in either NSCs nor GSCs at concentrations of ASV up to 1 μM, while a drop in proliferation rate was seen at 10 μM of ASV (Fig 5.2). An

optimal ASV concentration of 1 μM was therefore selected for subsequence experiments.

5.2.2 Endogenous knock-in of mCherry SMASh to the *Sox2* locus

The previous experiments were carried out using SOX2-mCherry overexpression, and therefore represent an excessive amount of SOX2. To confirm the SMASh tag is effective for endogenous genes, a SOX2-linker-mCherry-SMASh fusion protein was generated by CRISPR/Cas9 mediated knock-in to the endogenous *Sox2* locus in mouse NSCs (Fig 5.3.A). Following transfection of NSCs with the targeting cassette and Cas9/crRNA/tracrRNA complex (RNP), mCherry positive cells were FACS sorted from a 0.05% positive population. 17/23 clones stained positive for nuclear mCherry and showed cassette integration by PCR (data not shown). 3/3 of these clones were positive for correct targeting by genotyping (Fig 5.3.B). All clones screened had wild-type band by PCR, indicating all were heterozygous knock-ins.

As expected, the addition of 1 μM of ASV, for 48 hrs, lead to a reduction of SOX2-mCherry protein from the knock-in lines. From high exposure ICC images, residual SOX2-mCherry can be observed. However, the residual protein does not appear to be exclusively nuclear (Fig 5.3.C). Residual SOX2-mCherry protein, in addition to the larger non-cleaved SOX2-mCherry-SMASh fusion protein, can also be observed after 48 hrs of ASV treatment by western blotting (Fig 5.3.D). In the absence of ASV no un-cleaved SOX2-mCherry-SMASh fusion protein is observed indicating efficient self-cleavage of SMASh from SOX2 (Fig 5.3.D).

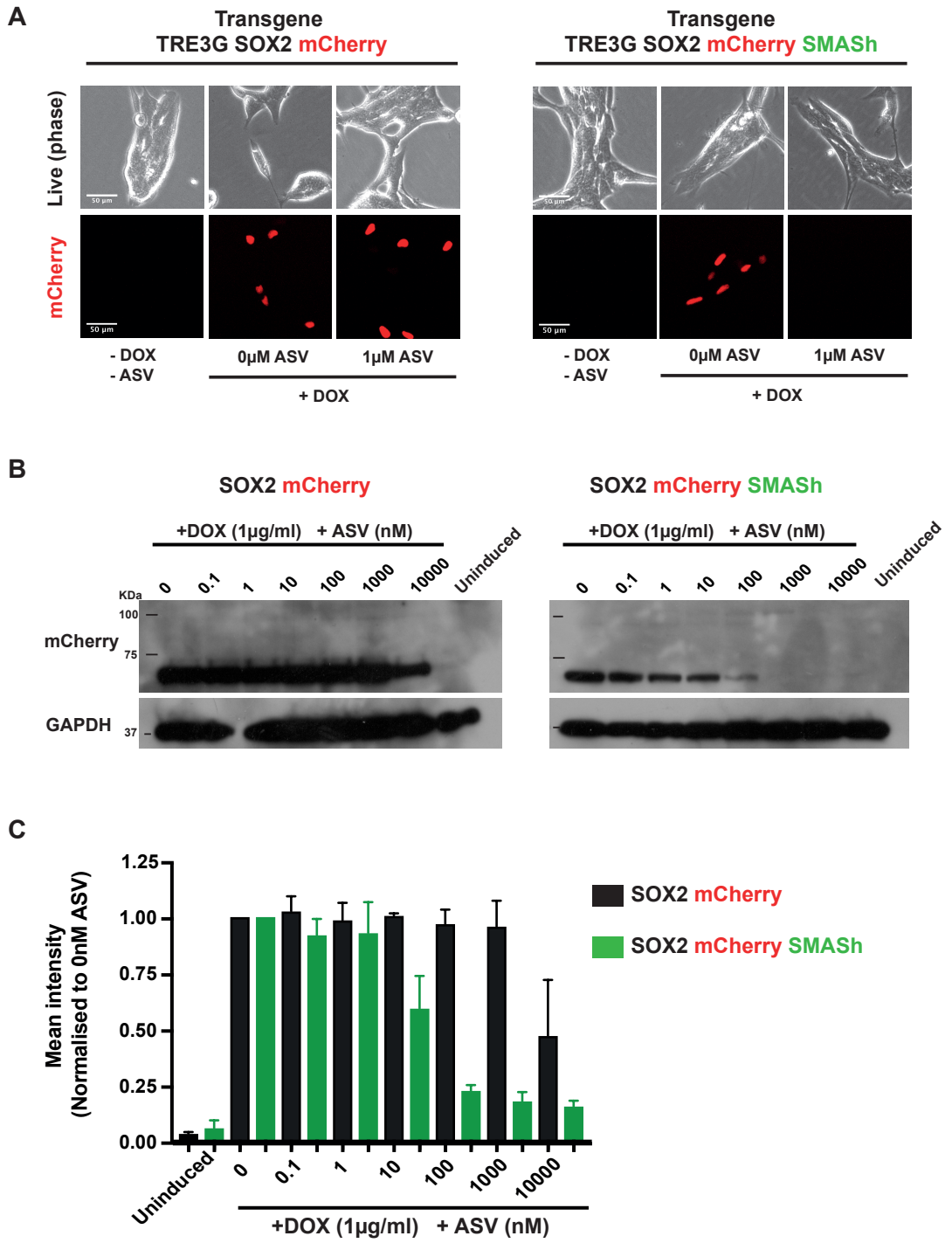


Figure 5.1 | SMASh controls transgenic SOX2-mCherry protein levels

(A) Representative live images 48 hrs after doxycycline (1 µg/ml) mediated transgene induction and 1 µM ASV treatment. ASV treatment was started on day 0. n=3. Scale bar: 50 µm. **(B)** Western blotting for mCherry and GAPDH. Samples, with the exception of

uninduced control, were treated with 1µg/ml of doxycycline and ASV (0, 0.1, 1, 10, 100, 1000 and 10000 nM) for 48hrs prior to collection. All samples contained 0.3% DMSO. n=1. **(C)** Flow cytometry for mCherry in samples, with the exception of uninduced control, treated with 1 µg/ml of doxycycline and ASV (0, 0.1, 1, 10, 100, 1000 and 10000 nM) for 48hrs prior to collection. All samples contained 0.3% DMSO. Data from each line is normalised to induced with 0 nM ASV. Plots show mean ± SD from 3 biological replicas.

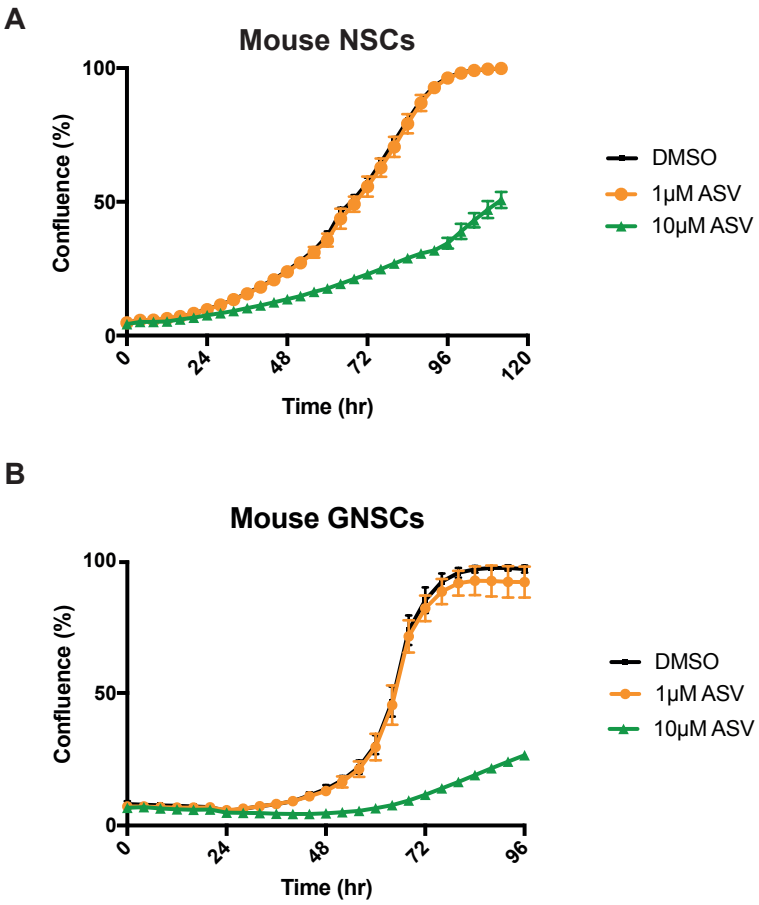


Figure 5.2 | Confluency curves of mouse NSCs and GSCs in ASV to asses toxicity

Confluency curves of mouse **(A)** NSCs and **(B)** GSCs treated with 0 μM (black) ,1 μM (orange) or 10 μM (green) of ASV in 0.3% DMSO. Plots show mean percentage confluence ± SD from 4 replica wells. Biological replicas show same trend (n=3).

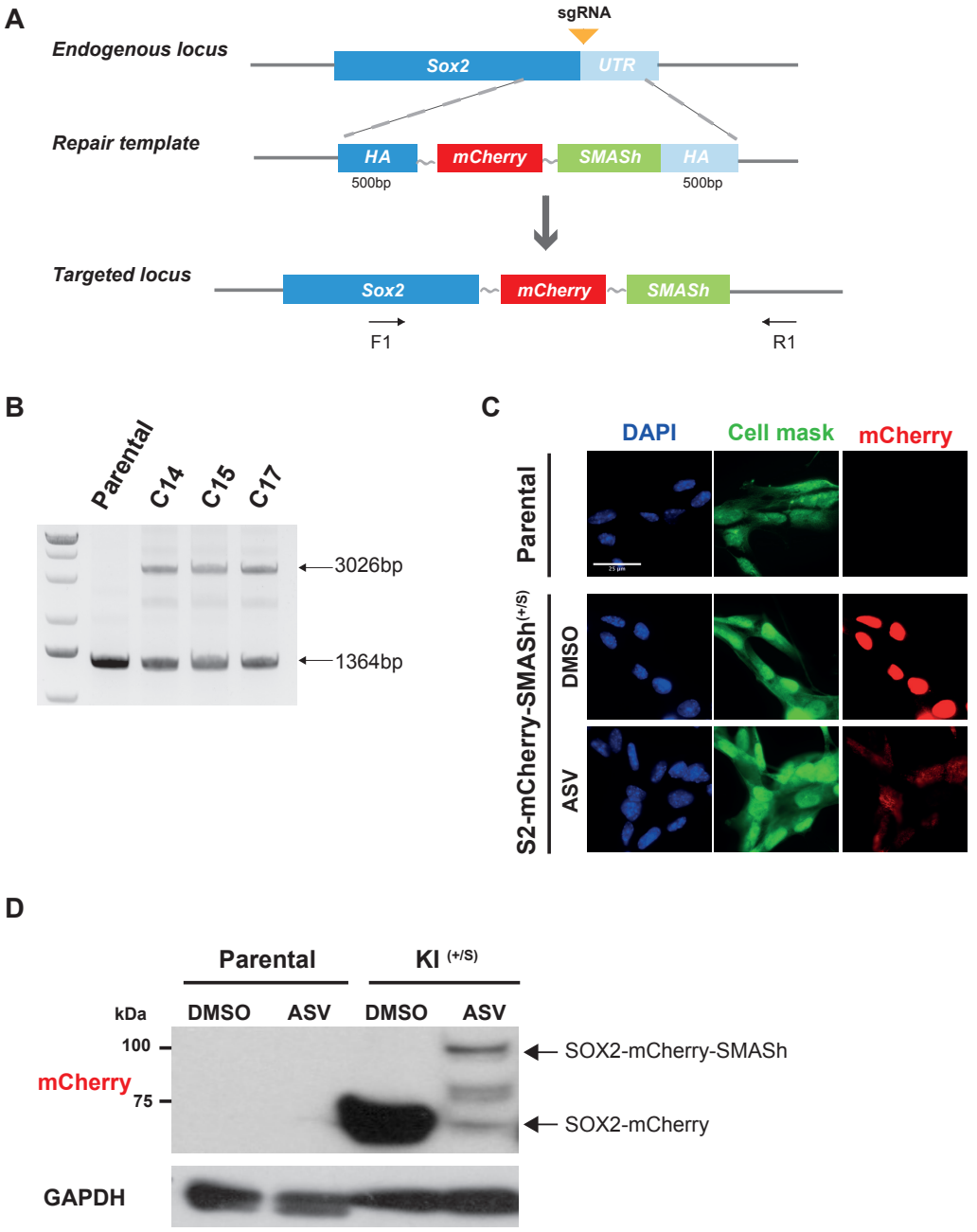


Figure 5.3 | Endogenous knock-in of mCherry-SMASh to the *Sox2* locus

(A) Schematic of targeting vector (repair template) and targeted *Sox2* locus. CRISPR sgRNA, which cuts in the 3'UTR near the stop codon, is depicted by a yellow triangle. Using 500 bp homology arms, *Sox2*-mCherry-SMASh fusion protein sequence was generated. Grey lines represent flexible peptide linkers. PCR genotyping primers were designed outside of the homology arms and are depicted as F1 and R1. **(B)** Genotyping

of 3 clones with primers F1 and R1. Sizes of expected wild-type and knock-in bands are highlighted. **(C)** ICC in wild-type parental line and a heterozygous *Sox2*-mCherry-SMASh line (+/s). *Sox2*-mCherry-SMASh line was treated for 48 hrs with DMSO or 1 μM of ASV (n=3). Scale bar: 25 μm. **(D)** Western blotting in wild-type parental and *Sox2*-mCherry-SMASh lines. Cells were treated with either DMSO or 1 μM ASV for 48 hrs. (n=3)

5.2.3 Rate of SOX2-mCherry control by SMASh

In the endogenous knock-in line unexposed to ASV there is an accumulation of SOX2-mCherry which has been separated from the self-cleaving SMASh degron. This differs from section 5.2.1 where ASV was added simultaneously to doxycycline, hence the SMASh tag should not self-cleave, and SOX2-mCherry-SMASh will be degraded as it is generated.

Hence, to determine the optimal duration of ASV treatment to induce the clearance of SOX2-mCherry a time course was performed using the knock-in reporter/degron line. By ICC a significant drop in SOX2-mCherry was observed after 16 hrs of ASV treatment. The additional reductions in protein levels occurring after 16 hrs appear comparatively minor by ICC, though after 48 hrs of ASV treatment the residual SOX2-mCherry present is no longer exclusively nuclear (Fig 5.4.A). The localisation of residual SOX2-mCherry suggests it may not be functional, though further studies are required to determine this. By flow cytometry a 18% drop in SOX2-mCherry levels can be observed 6 hrs after initial ASV treatment. At 16 hrs of ASV treatment a drop of 57.5% is seen and this increases to a maximal drop of 79% by 48 hrs of ASV treatment (Fig 5.4.B). This time span required to see a drop in SOX2-mCherry, under SMASh control, needs to be considered when making future observations on the phenotypical effects caused by the loss of SOX2.

After self-cleavage of the SMASh tag, the resulting SOX2-mCherry is no longer under SMASh control, and hence should not be affected by ASV addition. The loss of this protein is now dependant on its half-life. Assuming that, when adding ASV the rate of degradation of newly formed SMASh controlled protein is constant, the half-life of SOX2-mCherry in mouse NSCs is estimated to be around 14 hrs (Fig 5.4.B). This

highlights how the SMASh tag can be used to estimate protein half-lives without the use of traditional protein synthesis inhibitors such as cycloheximide (CHX).

5.2.4 Rapid post-translational reversibility of SMASh

Control mediated by the SMASh tag is reversible, with the removal of ASV allowing for the re-accumulation of SOX2-mCherry. To determine the rate at which SOX2-mCherry is re-established, cells were treated with 1 μM of ASV for 48 hr then ASV was washed out. By ICC, the return of nuclear protein can be observed 2hrs after ASV removal, with protein levels similar to pre-ASV treatment being observed 8hrs after ASV removal (Fig 5.5.A). This rapid return of protein fits with the post-translational control that the SMASh degron offers. Western blotting confirmed the reappearance of SOX2-mCherry 2 hrs after ASV removal. From ImageJ quantification, and normalisation to DMSO treatment, it can be estimated that 15 hrs after ASV removal SOX2-mCherry levels returned to 50% the amount to prior ASV treatment. It will be necessary to perform a longer time course to determine when SOX2-mCherry levels return to pre-ASV treatment levels as, 24 hrs after ASV removal SOX2-mCherry levels were only 63.8% of those observed in DMSO only treatment (Fig 5.5.B). This data demonstrates the reversible nature of SMASh, with exact timings likely to vary between proteins.

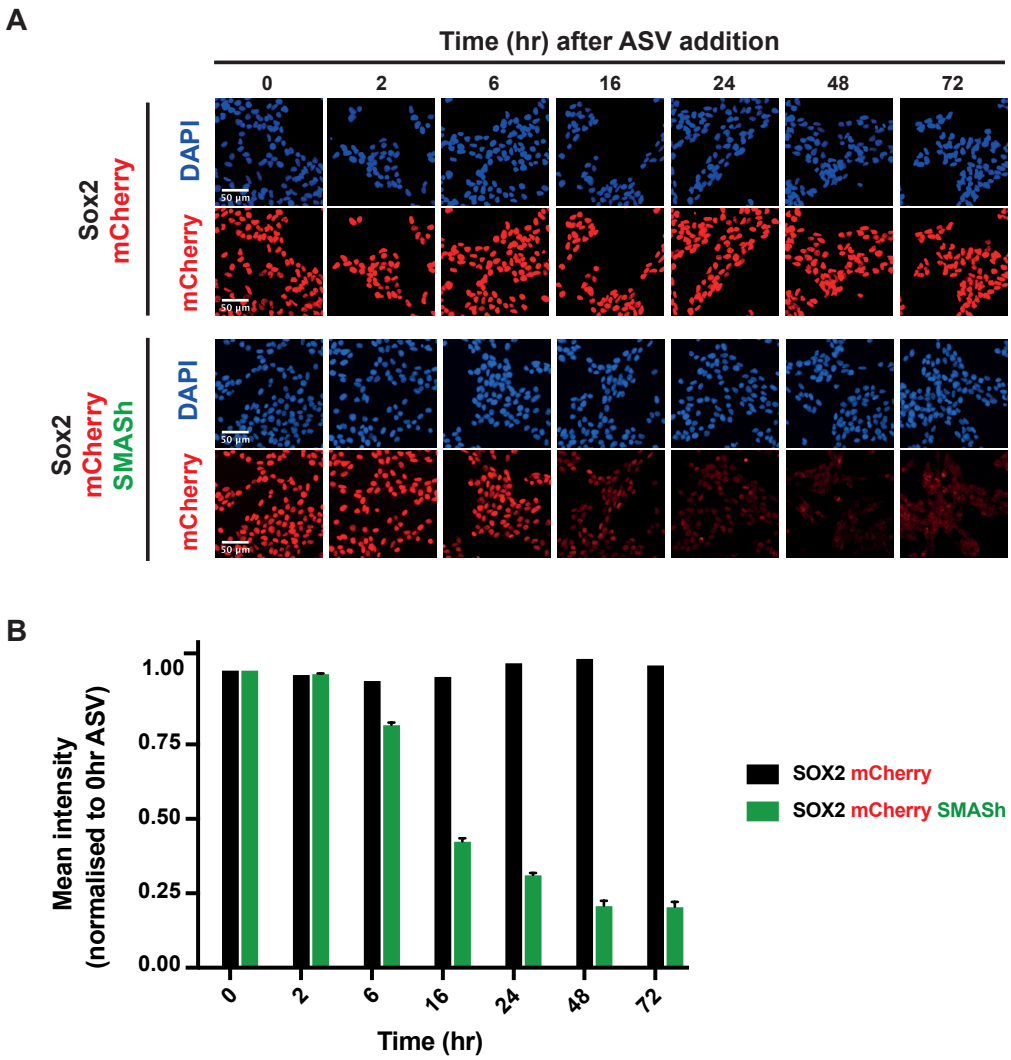


Figure 5.4 | Rate of SOX2-mCherry control by SMASh

(A) ICC of *Sox2*-mCherry and *Sox2*-mCherry-SMASh endogenous knock-in lines. Cells were treated with 1 µM of ASV for varying time points (0,2,6,16,24,48 and 72 hrs). Representative images are shown (n=3). Scale bar: 50 µm. **(B)** Flow cytometry of *Sox2*-mCherry and *Sox2*-mCherry-SMASh lines. Cells were treated with 1 µM of ASV for varying time points (0, 2, 6, 16, 24, 48 and 72 hrs). Wild-type parental line was used to set gates. Fluorescent mean intensity was normalised to 0hr time point. Plot shows mean ± SD, n=3 for *Sox2*-mcherry-SMASh and n=1 for *Sox2*-mCherry.

5.2.5 A strategy to enrich for the homozygous knock-in of SMASh

Following the endogenous knock-in of mCherry-SMASh to the *sox2* locus, we did not obtain any homozygous clonal lines. To enrich for the selection of cells with bi-allelic engineering, an alternative strategy was designed using two independent targeting vectors (Fig 5.6.A). Whilst a fluorophore has obvious advantages, the addition of a large tag to SOX2 maybe unfavourable. Instead of making a SOX2-mCherry fusion protein the mCherry was replaced with the small epitope tag HA. Secondly, the vectors contain fluorescent proteins mCherry or BFP, co-expressed from *sox2* via a P2A "self-cleaving" peptide. Thus, the two final vectors that were constructed were: HA-SMASh-P2A-mCherry or HA-SMASh-P2A-BFP with homology arms directed to the C-terminal of SOX2. This would enable us to identify and sort red and blue double positive cells but have both with the same C-terminal HA fusion tag. This strategy might avoid the need to pick and screen clonal lines, as we would be confident working with the population.

From a preliminary multiplex transfection experiment, dual positive cells could be observed by FACS, albeit it at low efficiency (25 double positives from 1353709 corresponding to 0.002% of the live population). There appears to be a higher rate of integration of the mCherry targeting cassette (0.06%) over the BFP cassette (0.007%) within the multiplex sample (Fig 5.6.B). Similar integration efficiencies, 0.04% for mCherry and 0.01% for BFP, were observed by FACS from single cassette transfections (data not shown). Subsequent ICC (Fig 5.6.C) and genotyping (Fig 5.6.D) indicate correct targeting of the single colours; however, the sorted population is not 100% targeted (Fig 5.6.C). Unfortunately, the dual positive cells did not survive and expand, and could not be further characterised. Future studies to obtain bi-allelic knock-ins will be performed using larger numbers of starting cells.

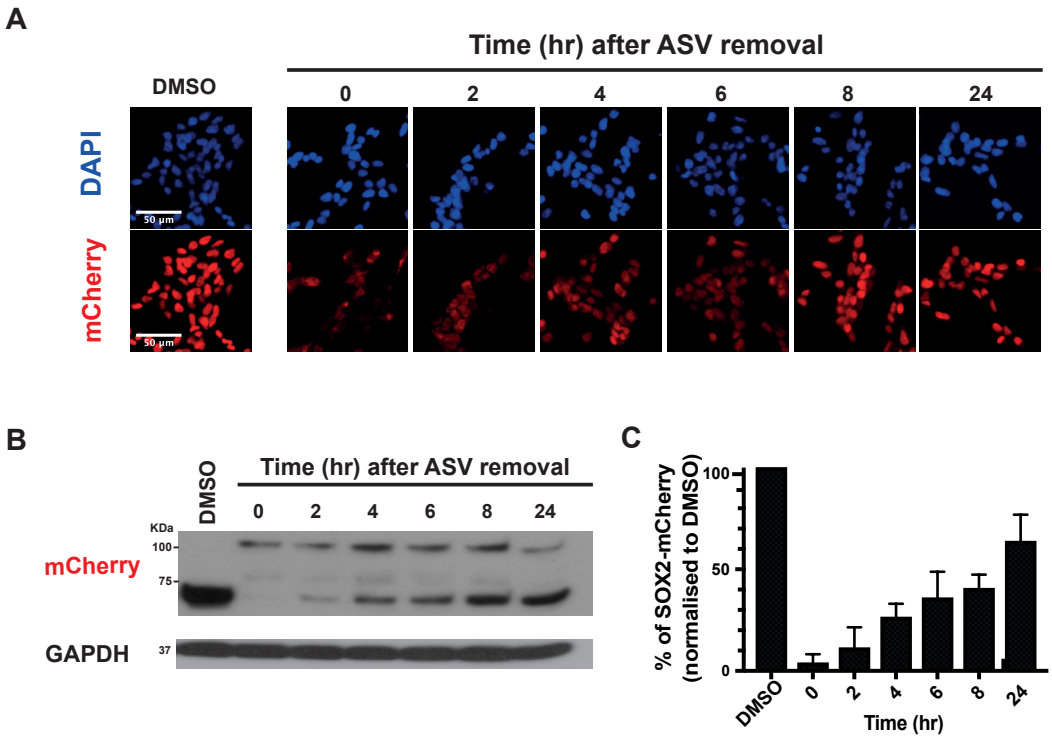


Figure 5.5 | Rapid post-translational reversibility of SMASh

(A) Representative ICC of *Sox2*-mCherry-SMASh knock-in lines. Cells were treated with 1 uM of ASV for 48 hrs. ASV was washed out and samples collected at varying points (0, 2, 4, 6, 8 and 24 hrs) post-ASV removal. Scale bar :50 µm. n=3 **(B)** Representative Western blotting. *Sox2*-mCherry-SMASh lines were treated with 1 uM ASV for 48 hrs. ASV was washed out and samples collected at varying time points (0, 2, 4, 6, 8 and 24 hrs) post ASV removal, n=3. **(C)** ImageJ analysis of SOX2-mCherry protein levels from western blotting. SOX2-mCherry-SMASh fusion was excluded from analysis. Data is normalised to DMSO treated samples. Plot shows mean ± SD from 3 biological replicas.

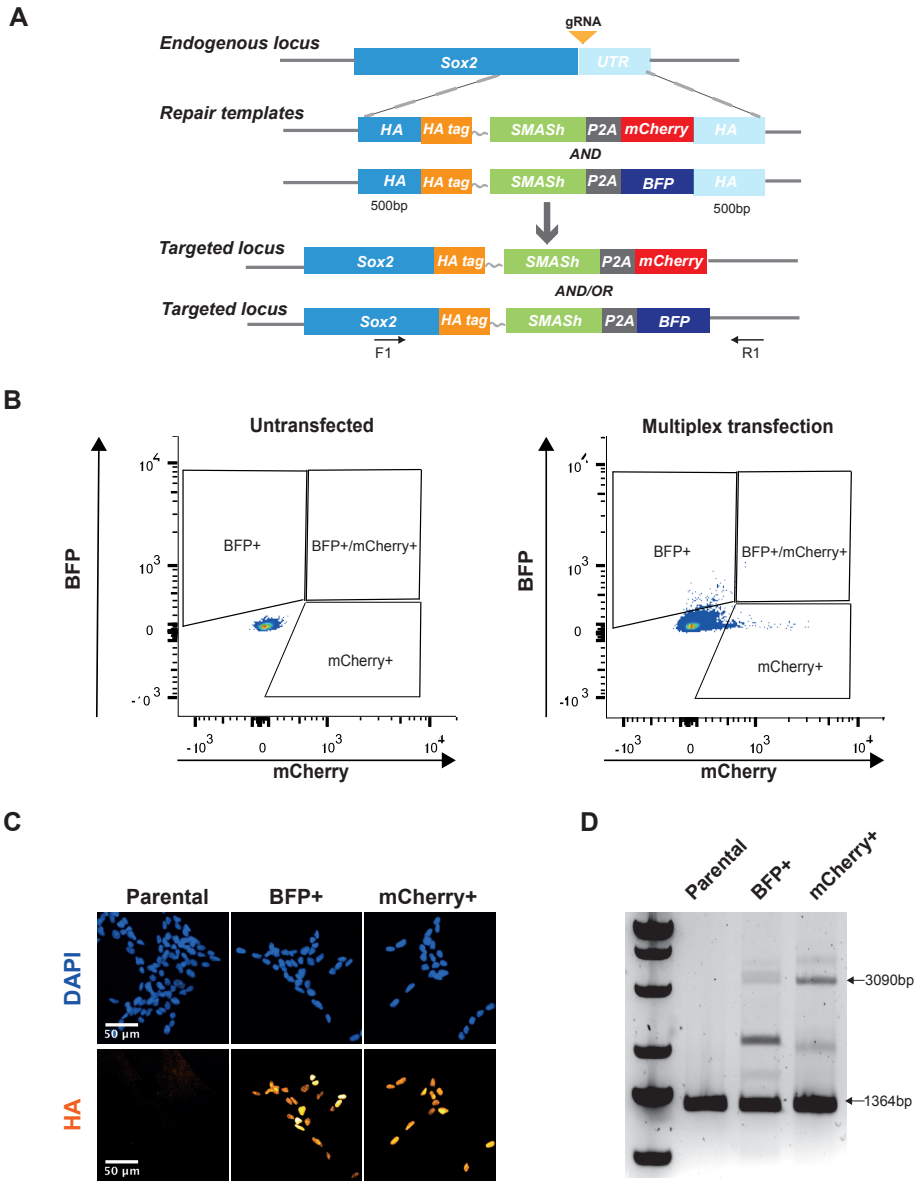


Figure 5.6 | Bi-allelic knock-in of SMASh

(A) Schematic of targeting vectors (repair template) and targeted *Sox2* locus. CRISPR sgRNA is depicted by a yellow triangle. PCR genotyping primers were designed outside of the homology arms and are depicted as F1 and R1. Vector was designed by Katrina McCarten and made by Rachel White. **(B)** FACS of mouse NSCs transfected using both targeting vectors. Non-transfected parental line was used to set gates. **(C)** ICC of BFP positive and mCherry positive sorted cells. Nuclear HA staining indicates correct targeting. **(D)** Genotyping from the bulk population of BFP positive and mCherry positive cells. Genotyping primers F1 and R1 were used. Sizes of expected wild-type and knock-in bands are highlighted.

5.3 Discussion

SMASh is a relatively recently reported degron, with several advantages over more widely used degrons, such as AID. At the time of this study, no one had yet reported on the use of SMASh for neural stem cells or the transcription factor, *Sox2*. From our experimental data, we observed for the first time, that SMASh functions for both transgenes and knock-in reporters within mouse NSCs. Specifically, SMASh enables post-translation control of SOX2, with endogenously SMASh tagged SOX2 retaining its nuclear localisation even with the residual amino acids left by the self-cleaving tag. In agreement with our findings, a newly published study found SMASh be an effective degron for SOX2 in hPSCs and for FOXG1 in hPSC derived NPCs[264].

The self-cleavage of SMASh from the protein of interest leaves the protein minimally modified and hence, loss of the protein of interest is dependent on its half-life. While the kinetics of the SMASh degron are slower than some other degrons, we observed that the rate is still suitable for most studies. This feature also enables the SMASh system to be used to study the protein's half-life, without the need for global protein synthesis inhibitors or complicated pulse-chase SILAC experiments. From our studies we estimated the half-life of SOX2-mCherry to be approximately 14 hrs. This estimation is longer than 5-6 hr SOX2 half-life predicted by the use of the protein synthesis inhibitor, cycloheximide [145,146]. From SNAP-tag® and Halo-tag® SOX2 fusion proteins the half-life of SOX2 was estimated to be around 9 hrs in mouse ESCs[265]. This estimate is closer to the one we observe. The addition of the fluorophore mCherry may alter the degradation of SOX2 or using the SMASh system may give an overestimation of protein half-life.

In order to use SMASh to investigate the phenotypic effects of SOX2 loss, it will be necessary to derive a line with bi-allelic knock-in of SMASh. By using tagging

strategies which include P2A-conjugated fluorophores, we observed that the knock-in efficiency to the *Sox2* locus is low (~ 0.05% for vectors with mCherry and 0.01% for vectors with BFP). This is not unusual, with similar low efficiency rates inferred by the use of antibiotic selection cassettes to obtain knock-in lines[264,266], with dual antibiotic selection methods further enriching for bi-allelic knock-ins[264]. To obtain bi-allelic knock-in using our vectors either a large sample must be sorted or, as an alternative to multiplex transfection, transfection with each vector can be performed sequentially, enriching for targeted cells after each transfection by FACS. Either way, even from the efficiencies observed it should be possible to obtain bi-allelic knock-ins. Assessing the feasibility of this approach is also important for future studies with heterogenous human GBM patient lines (www.gcgr.org.uk), where the generation of clonal populations of engineered cells should be avoided.

The translation of SMASh mediated control from *in vitro* to *in vivo* systems will be highly valuable but has been beyond the scope of this current thesis. Using current mouse GBM models (Ester Gangoso, unpublished) our lab has established protocols for tumour formation and monitoring; hence it should be feasible to allow a tumour to form and then look at the effect of knocking down the protein of interest. The observed *in vitro* concentration of ASV required to get maximal knock down of SOX2-mCherry is 1 μM, though further studies using between 100 nM and 1 μM of ASV should be performed to see if a lower dose will still give maximal knock down. The concentrations of ASV required *in vitro* are in line with those often observed in the liver of mice treated with ASV [267,268] and one study has shown the reduction of the protein BRM under SMASh control within a subcutaneous tumour when orally administering ASV[269]. A major potential limitation to using SMASh for *in vivo* studies is the inability of the NS3 inhibitors, including ASV[270], to cross the blood brain barrier (BBB). It remains to be seen if this will be an issue for future studies as our mouse

GBM models create tumours which are thought to have a leaky BBB in the tumour bulk.

Similar to previous reports[264,269] we do not observe a complete loss of SOX2 when inducing degradation by ASV addition. It is unclear if this residual protein is still functional as it is composed of both SMASh fusion and non-fusion fractions and, more importantly, is not exclusively nuclear. Further exploration is required to determine if the residual SOX2 observed after ASV treatment is sufficient for normal cellular function. Due to the inability to completely remove all protein, the SMASh system is unlikely to replace knockout studies, but is a technology to supplement them. In fact, whilst the inability to ablate all of SOX2 initially appears like a negative feature of SMASh, the situation is much more likely to mimic acute knockdowns of proteins observed by therapeutic drugs. This, combined with the easy tuneability of protein levels achieved by SMASh, is particularly useful when assessing proteins for their therapeutic potential.

5.3.1 Concluding remarks

In summary, we demonstrate that the SMASh tag is an efficient degron offering easy tuneable control of SOX2 levels in neural stem cells. Future studies will soon be able to determine if it retains value *in vivo and* is capable of acutely ablating SOX2 in established tumours.

Chapter 6 General discussion

In this thesis we aimed to understand the biological functions of the understudied SOX2 interacting protein, MYEF2. Additionally, we sought to explore the efficiency of the SMASh tag degron at degrading SOX2 in NSCs. Below, I will summarise the key observations and future directions of this investigation.

6.1 The function of MYEF2 in NSCs and glioblastoma

6.1.1 MYEF2 is SOX2 interactor, but does it interact with other SOX factors?

Work from this thesis, the Pollard laboratory and others[271,272] strongly support that MYEF2 is an interactor of SOX2 within NSCs and GSCs. In line with this we observed that global MYEF2 expression is high in the brain, suggesting a functional role which was subsequently studied. MYEF2 protein was also observed to be particularly high with the testis of adult mice. Recent studies have indicated that whilst, SOX3 is classically associated with spermatogonia stem cells (SSCs)[103], a SOX2 expressing population of SSCs also resides in the testis[112]. It would be interesting to investigate if MYEF2 is a protein interactor of SOX2 or SOX3 within SSCs. Lower levels of MYEF2 were also seen in the intestine, liver and heart. Within the intestine and liver, SOX9[273] is the main SOX factor predominantly associated with adult stem cells or progenitors cells. It would be extremely interesting to investigate, though RNA-FISH and IHC (if a good antibody is identified), if MYEF2 co-localises within the SOX expressing stem cell populations of multiple tissues and, through western blotting, if MYEF2 interacts with the SOX factors predominantly associated with these tissues. This will indicate if MYEF2 is involved in a global mechanism with varying SOX proteins or if the function is specific only to SOX2.

6.1.2 The biological function of MYEF2

One of the major finding of this thesis was the effect that the loss of MYEF2 had on GSCs. Loss of MYEF2 generally reduced the competitiveness of GSCs, with reduced proliferation rates, reduced colony formation potential, increased BMP sensitivity and most dramatically increased survival rates of transplanted mice being observed. This was perhaps not unsurprising due to its interaction with SOX2 and high levels of expression within NSCs and GSCs but adds strong support that MYEF2 is playing a functional role in GBM. The phenotype of *Myef2* knockout GSCs was mild compare to that expected for GSCs lacking SOX2, suggesting that MYEF2 is not essential for the function of SOX2 but aids it to work efficiently. Currently all observations on the function of MYEF2 are within mice and a MES specific CSC model. It will be both interesting and necessary, to determine if the phenotypic observations observed from MYEF2 loss are maintained within multiple CSCs, modelling a range of transcriptional subtypes, and within human GBM samples, hence providing information as to if MYEF2 is relevant to clinical cases of GBM. It should also be noted that within this thesis emphasis was placed on studying the interaction of SOX2 and MYEF2. However, MYEF2 has previously been shown to interact with RUNX1 within mouse erythroleukemia cells[135]. RUNX1 is overexpressed in the MES subtype of GBM and is thought to contribute to the MES subtype and increased invasion capacity, via the TGFß pathway[274]. The potential links of MYEF2 and RUNX1 within GBM is still to be explored, with this link highlighting the requirement for more extensive characterisation of *Myef2* knockout tumours as previously discussed.

The next major step for this work is to identify the molecular mechanisms of MYEF2 and determine how they link with SOX2. Based on the structure and functional domains of MYEF2, we propose MYEF2 is likely to function as an RNA binding protein. Supporting this several studies have identified that MYEF2 binds to a range

of RNAs including mRNAs[275,276] and the long noncoding RNA (lncRNA) XIST[134,277]. Based on current data we tentatively propose two potential mechanisms for MYEF2. The first hypothesis is that MYEF2 functions through lncRNA. lncRNA (>200nt) are generally transcribed by RNAPII, evolve more rapidly than protein encoding gene and tend to have restricted expression in specific tissues.

The mechanism of action of many lncRNAs is still elusive. However, several have been found to mediate transcriptional regulation through recruitment of chromatin modifying complexes[278–280], transcription factors[281], preventing transcriptional factor access[282] and potentially through chromatin organisation[283–285]. Through interacting with lncRNAs, we propose that MYEF2 may mediate the efficient recruitment of SOX2 to its targets, mediating rapid expression of SOX2 target factors upon exit from the quiescent state. In line with this, MYEF2 expression is not reduced *in vitro* when treating with BMP4 and *Myef2* knockout cells are potentially less efficient at exiting the quiescent state, though further experiments, such as epigenetic profiling, are required to confirm this. Targeting quiescent CSCs, which are considered responsible for tumour relapse after surgery, radiotherapy and chemotherapy[6–9], is an attractive therapy. If the core molecules and signals involved in the exit from quiescence can be identified, it may be possible to inhibit cells from waking up, making tumour recurrence is unlikely.

The second hypothesis is that MYEF2 is functioning as a splicing factor mediating the rapid and correct co-transcriptional splicing[286] of SOX2 target genes. In line with this, our mass spectrometry data of MYEF2 interactors was enriched with interactors associated with mRNA processing and splicing, and others have shown mRNA binding of MYEF2[275,276]. Additionally, MYEF2 does not act as bookmarking factor and dissociated from the mitotic chromatin, and possibly SOX2, during mitosis. It should

be noted that during mitosis, transcription is reduced on a global scale[287] hence, MYEF2 target mRNAs may not be present. Finally levels of MYEF2, and many other RNA binding proteins, are high in both the brain and testis, the two tissues associated with highest levels of alternative splicing[288,289]. In support of a potential RNA splicing mechanism, there is growing evidence indicating that alterations in RNA splicing play tumorigenic roles, with global splicing changes and mutations in several splicing factors observed in multiple types of cancer[290–292]. Furthermore, the inhibition of arginine methyltransferase PRMT5 can alter the splicing of detained introns and mediate suppression of *in vivo* GBM tumour growth [293].

To distinguish between these mechanisms, and understand the biology, several experiments could be performed. RNA-seq in lines with and without MYEF2 should allow us to look at alternative splicing of transcripts. If we were to observe a clear enrichment in alternative splicing of transcripts from genes regulated by SOX2, this would support that MYEF2 functions as a splicing factor. ChIP-seq would allow assessment of the overlap of MYEF2 and SOX2, and help to asses if MYEF2 does enable the rapid recruitment of SOX2 to its G1 targets upon re-entry to the cell cycle. In addition, if MYEF2 is a splicing factor binding would be enriched over exon-intron boundaries. Finally, RNA immunoprecipitation methods, such as RIP and CLIP[294], coupled with sequencing can identify the exact RNAs bound to MYEF2, hence providing potential explanation for the observed *Myef2* knockout phenotype.

The proposed mechanisms of MYEF2 functioning as an RNA binding protein are likely to make it difficult to directly therapeutically target. Some proposed methods for therapeutically targeting splicing[295], including splice-switch oligonucleotides and spliceosome inhibitors, do exist however they are not well established, though alternatively, the transcripts and corresponding proteins which are altered by

alternative splicing could be therapeutically targeted. The combination of the strength of the phenotype observed in *Myef2* knockouts and the difficulty in targeting an RNA binding protein, mean it is unlikely one would invest in MYEF2 alone as a therapeutic candidate. Whilst this may be the case, understanding the interplay of SOX2 and MYEF2 is likely to providing exciting insight into the fundamental behaviour of SOX2 and RNA biology.

Whilst disrupting NSC, or more ideally GSC, specific SOX2 interactors is a potential therapeutic approach, the SOX2 interactors responsible for key functions of SOX2 are yet to be identified and it is possible that many interactors of SOX2, like MYEF2, will individually only mildly effect SOX2's function. Hence, it is important that other avenues for targeting SOX2 are explored, including exploration of the effects of post-translational modifications, and the upstream signalling responsible, on SOX2[125].

6.2 The SMASh degron is successful at modulation of the nuclear protein SOX2 in NSCs

Conditional degrons provide alternative methods to classical genetic gene knockout and siRNA knockdowns, to study a proteins role through loss of function studies. Many conditional degrons exist, each with different features including those such as sensitivity to light or control by small molecules. Of particular interest, is the relatively recently derived SMASh tag degron[224]. Unlike many other degrons, the rate of degradation is dependent on the half-life of the protein and hence slow. However, the SMASh degron has the key advantages of easy dose dependant response and most importantly, the protein of interest (POI) to which it is fused is left minimally modified. This feature is especially important if the POI has known functional domains at the N- or C-terminal, as the addition of tags may affect protein function. In this thesis I

describe the successful use of the SMASh tag degron to control the protein levels of exogenous and endogenous SOX2 within mouse NSCs.

From the expression of exogenous SOX2-mCherry-SMASh we observed a clear dose dependant reduction in SOX2-mCherry levels. A maximal ~90% reduction in protein could observed in response to 1μM HCV N3S protease inhibitor ASV. At a slight lower rate than transgene, endogenously tagged SOX2-mCherry-SMASh was depleted to 20% of the endogenous level. This reduction was slow to achieve, taking 48hrs, due to the dependence on the proteins half-life. The absence of bi-allelic knock in clones has hampered the ability to test if this 80% knockdown is sufficient to investigate the phenotypic effects of SOX2 loss. Ongoing attempts are underway to derive a bi-allelic SMASh knock-in clones within NSCs and GSCs. Once obtained, bi-allelic SMASh tagged lines should allow both *in vitro* and *in vivo* experimentation in order to model the effects that the loss of SOX2 has on proliferation, senescence, BMP sensitivity, clonogenicity and on tumour initiation and maintenance. This approach will allow for rigorous evaluation of the function of SOX2, and the effect of varying SOX2 levels, in GBM, without the off-target effects associated with siRNA and CRISPR of previous studies.

The SMASh tag appears to act on a wide range of proteins[224,263,264,266]. This versatile degron may provide an excellent tool for further investigations of the function of *myef2.* In particular the generation of a bi-allelic *Myef2*-SMASh knock-in may complement knockout experiments and provide an alternative to the requirement of MYEF2 rescue experiments as described in chapter 4. Using endogenously tagged *Myef2*-SMASh assays can be performed in the exact same line in the absence and presence of ASV. Use of a control untagged line in the presence of ASV can account for any phenotypic observations by ASV. This alternative can avoid several of the issues associated with

piggyBac rescue experiments including non-physiological protein expression levels and limited isoform expression.

While the SMASh degron has the distinct advantage of leaving the protein of interest minimally modified, hence minimising the chance of altering normal protein functions, new degrons are continually being identified and developed, and current ones improved, expanding the repertoire of the degron toolbox. For example, the AID system has recently been improved through the development of mini-AID (mAID), a 68 aa fragment of the original 229 aa long AID[296], and improvements are being made to improve the leaky degradation often observed with the AID system[297–299]. Many of the degrons developed require the genetic modification of the protein of interest, providing a valuable research tool but are unsuitable for direct therapeutic use in patients. An exception to this are proteolysis-targeting chimeric molecules (PROTACs), which do not require modification of the target protein but change the specificity of E3 ligases (induce polyubiquitylation and proteasomal degradation) by tethering the protein of interest to an E3 ligase. As PROTACs are specific, they must be identified for each protein of interest and at the time of writing this thesis we were not aware of any that are capable of targeting SOX2, however it is clear that PROTACs are a potentially viable method to clinically degrade transcription factors, such as SOX2, which were previously considered "undruggable"[300–302].

6.3 Concluding remarks

To conclude, the RNA binding protein MYEF2 is pro-tumorigenic in GSCs, and most likely operates with SOX2. Dissecting the functions and mechanism of MYEF2 should expand our understanding of SOX2. The SMASh tag, combined with the efficiency of CRISPR/Cas9 genome editing, should provide a key tool for therapeutic target validation and support drug discovery efforts.

References

1. Louis, D. N. *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol.* **131**, 803–820 (2016).
2. Stupp, R. *et al.* Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. *N. Engl. J. Med.* **352**, 987–996 (2005).
3. Sarkaria, J. N. *et al.* Is the blood–brain barrier really disrupted in all glioblastomas? A critical assessment of existing clinical data. *Neuro. Oncol.* **20**, 184–191 (2018).
4. Young, R. M., Jamshidi, A., Davis, G. & Sherman, J. H. Current trends in the surgical management and treatment of adult glioblastoma. *Annals of Translational Medicine* (2015) doi:10.3978/j.issn.2305-5839.2015.05.10.
5. Bao, S. *et al.* Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* (2006) doi:10.1038/nature05236.
6. Chen, J. *et al.* A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature* (2012) doi:10.1038/nature11287.
7. Deleyrolle, L. P. *et al.* Evidence for label-retaining tumour-initiating cells in human glioblastoma. *Brain* (2011) doi:10.1093/brain/awr081.
8. Richichi, C., Brescia, P., Alberizzi, V., Fornasari, L. & Pelicci, G. Marker-independent method for isolating slow-dividing cancer stem cells in human glioblastoma. *Neoplasia (United States)* (2013) doi:10.1593/neo.13662.
9. Ishii, A. *et al.* Histological Characterization of the tumorigenic 'peri-necrotic niche' harboring quiescent stem-like tumor cells in glioblastoma. *PLoS One* (2016) doi:10.1371/journal.pone.0147366.
10. Clevers, H. What is an adult stem cell? *Science (80-.).* **350**, 1319–1320 (2015).
11. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* (2001) doi:10.1038/35102167.
12. Lowe, S. W., Cepero, E. & Evan, G. Intrinsic tumour suppression. *Nature* (2004) doi:10.1038/nature03098.
13. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* (2011) doi:10.1016/j.cell.2011.02.013.
14. KLEINSMITH LJ, P. G. J. MULTIPOTENTIALITY OF SINGLE EMBRYONAL CARCINOMA CELLS. *October* (1964).

15. Lapidot, T. *et al.* A cell initiating human acute myeloid leukaemia after transplantation into S. *Nature* **367**, 645–648 (1994).
16. Dick, J. E. & Bonnet, D. Human Acute Myeloid Leukaemia is organised as a heirarchy that originates from the primitive haematopoetic cell. *Nat Med* **3**, 730–737 (1997).
17. Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* (2003) doi:10.1073/pnas.0530291100.
18. Sheila K. Singh, Ian D. Clarke, Mizuhiko Terasaki, Victoria E. Bonn, Cynthia Hawkins, Jeremy Squire, A. *et al.* Identification of a cancer stem cell in human brain tumors. *Cancer Res* **63**, 5821–5828 (2003).
19. Singh, S., Hawkins, C., Clarke, I. & Squire, J. Identification of human brain tumour initiating cells. *Nature* **17**, 133 (2004).
20. Hemmati, H. D. *et al.* Cancerous stem cells can arise from pediatric brain tumors. *Proc. Natl. Acad. Sci. U. S. A.* (2003) doi:10.1073/pnas.2036535100.
21. Uchida, N. *et al.* Direct isolation of human central nervous system stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 14720–14725 (2000).
22. Galli, R. *et al.* Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. *Cancer Res.* (2004) doi:10.1158/0008-5472.CAN-04-1364.
23. Son, M. J., Woolard, K., Nam, D. H., Lee, J. & Fine, H. A. SSEA-1 Is an Enrichment Marker for Tumor-Initiating Cells in Human Glioblastoma. *Cell Stem Cell* (2009) doi:10.1016/j.stem.2009.03.003.
24. Anido, J. *et al.* TGF-β Receptor Inhibitors Target the CD44high/Id1high Glioma-Initiating Cell Population in Human Glioblastoma. *Cancer Cell* **18**, 655–668 (2010).
25. Ogden, A. T. *et al.* Identification of A2B5+CD133- tumor-initiating cells in adult human gliomas. *Neurosurgery* (2008) doi:10.1227/01.neu.0000316019.28421.95.
26. Chen, R. *et al.* A Hierarchy of Self-Renewing Tumor-Initiating Cell Types in Glioblastoma. *Cancer Cell* (2010) doi:10.1016/j.ccr.2009.12.049.
27. Dirkse, A. *et al.* Stem cell-associated heterogeneity in Glioblastoma results from intrinsic tumor plasticity shaped by the microenvironment. *Nat. Commun.* (2019) doi:10.1038/s41467-019-09853-z.
28. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in

primary glioblastoma. *Science (80-.).* **344**, 1396–1401 (2014).

29. Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* 835–849 (2019) doi:10.1016/j.cell.2019.06.024.
30. Aulestia, F. J. *et al.* Quiescence status of glioblastoma stem-like cells involves remodelling of Ca2+ signalling and mitochondrial shape. *Sci. Rep.* (2018) doi:10.1038/s41598-018-28157-8.
31. Marques-Torrejon, M. A., Gangoso, E. & Pollard, S. M. Modelling glioblastoma tumour-host cell interactions using adult brain organotypic slice co-culture. *Dis. Model. Mech.* (2018) doi:10.1242/dmm.031435.
32. Llorens-Bobadilla, E. *et al.* Single-Cell Transcriptomics Reveals a Population of Dormant Neural Stem Cells that Become Activated upon Brain Injury. *Cell Stem Cell* (2015) doi:10.1016/j.stem.2015.07.002.
33. Shin, J. *et al.* Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis. *Cell Stem Cell* (2015) doi:10.1016/j.stem.2015.07.013.
34. Sachdeva, R. *et al.* BMP signaling mediates glioma stem cell quiescence and confers treatment resistance in glioblastoma. *Sci. Rep.* (2019) doi:10.1038/s41598-019-51270-1.
35. Piccirillo, S. G. M. *et al.* Bone morphogenetic proteins inhibit the tumorigenic potential of human brain tumour-initiating cells. **444**, 761–765 (2006).
36. Pollard, S. M. *et al.* Glioma Stem Cell Lines Expanded in Adherent Culture Have Tumor-Specific Phenotypes and Are Suitable for Chemical and Genetic Screens. *Cell Stem Cell* **4**, 568–580 (2009).
37. Park, N. I. *et al.* ASCL1 Reorganizes Chromatin to Direct Neuronal Fate and Suppress Tumorigenicity of Glioblastoma Stem Cells. *Cell Stem Cell* (2017) doi:10.1016/j.stem.2017.06.004.
38. Carén, H. *et al.* Glioblastoma stem cells respond to differentiation cues but fail to undergo commitment and terminal cell-cycle arrest. *Stem Cell Reports* **5**, 829–842 (2015).
39. Hu, J. *et al.* Neutralization of terminal differentiation in gliomagenesis. *Proc. Natl. Acad. Sci. U. S. A.* (2013) doi:10.1073/pnas.1308610110.
40. Lan, X. *et al.* Fate mapping of human glioblastoma reveals an invariant stem cell hierarchy. *Nature* **549**, 227–232 (2017).
41. TCGA Research Network. https://www.cancer.gov/tcga.
42. Parsons, D. W. *et al.* An Integrated Genomic Analysis of Human Glioblastoma

Multiforme. *Science (80-.).* **321**, 1807–1812 (2008).

43. Hegi, M. E. *et al.* MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma. *N. Engl. J. Med.* **352**, 997–1003 (2005).
44. Verhaak, R. G. W. *et al.* Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–110 (2010).
45. Wang, Q. *et al.* Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell* **32**, 42-56.e6 (2017).
46. Sottoriva, A. *et al.* Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl. Acad. Sci.* **110**, 4009–4014 (2013).
47. Quail, D. F. & Joyce, J. A. The Microenvironmental Landscape of Brain Tumors. *Cancer Cell* (2017) doi:10.1016/j.ccell.2017.02.009.
48. Wolf, K. J., Chen, J., Coombes, J. D., Aghi, M. K. & Kumar, S. Dissecting and rebuilding the glioblastoma microenvironment with engineered materials. *Nature Reviews Materials* (2019) doi:10.1038/s41578-019-0135-y.
49. Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *Journal of Clinical Investigation* (2009) doi:10.1172/JCI39104.
50. Brabletz, T. To differentiate or not-routes towards metastasis. *Nature Reviews Cancer* (2012) doi:10.1038/nrc3265.
51. Watanabe, M., Tanaka, R. & Takeda, N. Magnetic resonance imaging and histopathology of cerebral gliomas. *Neuroradiology* (1992) doi:10.1007/BF00598951.
52. Kubelt, C., Hattermann, K., Sebens, S., Mehdorn, H. M. & Held-Feindt, J. Epithelial-to-mesenchymal transition in paired human primary and recurrent glioblastomas. *Int. J. Oncol.* (2015) doi:10.3892/ijo.2015.2944.
53. Zarkoob, H., Taube, J. H., Singh, S. K., Mani, S. A. & Kohandel, M. Investigating the Link between Molecular Subtypes of Glioblastoma, Epithelial-Mesenchymal Transition, and CD133 Cell Surface Protein. *PLoS One* (2013) doi:10.1371/journal.pone.0064169.
54. Mahabir, R. *et al.* Sustained elevation of Snail promotes glial-mesenchymal transition after irradiation in malignant glioma. *Neuro. Oncol.* (2014) doi:10.1093/neuonc/not239.
55. Kahlert, U. D. *et al.* Activation of canonical WNT/β-catenin signaling enhances in vitro motility of glioblastoma cells by activation of ZEB1 and other activators

of epithelial-to-mesenchymal transition. *Cancer Lett.* (2012) doi:10.1016/j.canlet.2012.05.024.

56. Yang, W. *et al.* Sortilin promotes glioblastoma invasion and mesenchymal transition through GSK-3β/β-catenin/twist pathway. *Cell Death Dis.* (2019) doi:10.1038/s41419-019-1449-9.
57. Sandberg, C. J. *et al.* Comparison of glioma stem cells to neural stem cells from the adult human brain identifies dysregulated Wnt- signaling and a fingerprint associated with clinical outcome. *Exp. Cell Res.* (2013) doi:10.1016/j.yexcr.2013.06.004.
58. Ye, X. *et al.* Tumor-Associated Microglia/Macrophages Enhance the Invasion of Glioma Stem-like Cells via TGF-β1 Signaling Pathway. *J. Immunol.* (2012) doi:10.4049/jimmunol.1103248.
59. Kriegstein, A. & Alvarez-Buylla, A. The Glial Nature of Embryonic and Adult Neural Stem Cells. *Annu. Rev. Neurosci.* **32**, 149–184 (2009).
60. Yuzwa, S. A. *et al.* Developmental Emergence of Adult Neural Stem Cells as Revealed by Single-Cell Transcriptional Profiling. *Cell Rep.* **21**, 3970–3986 (2017).
61. Haubensak, W., Attardo, A., Denk, W. & Huttner, W. B. From The Cover: Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: A major site of neurogenesis. *Proc. Natl. Acad. Sci.* **101**, 3196–3201 (2004).
62. Noctor, S. C., Martínez-Cerdeño, V., Ivic, L. & Kriegstein, A. R. Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. *Nat. Neurosci.* **7**, 136–144 (2004).
63. Suh, H. *et al.* In Vivo Fate Analysis Reveals the Multipotent and Self-Renewal Capacities of Sox2+ Neural Stem Cells in the Adult Hippocampus. *Cell Stem Cell* **1**, 515–528 (2007).
64. Morshead, C. M. *et al.* Neural stem cells in the adult mammalian forebrain: A relatively quiescent subpopulation of subependymal cells. *Neuron* **13**, 1071–1082 (1994).
65. Doetsch, F., Caillé, I., Lim, D. A., García-Verdugo, J. M. & Alvarez-Buylla, A. Subventricular Zone Astrocytes Are Neural Stem Cells in the Adult Mammalian Brain. *Cell* **97**, 703–716 (1999).
66. Menn, B. *et al.* Origin of Oligodendrocytes in the Subventricular Zone of the Adult Brain. *J. Neurosci.* **26**, 7907–7918 (2006).

67. Bachoo, R. M. *et al.* Epidermal growth factor receptor and Ink4a/Arf. *Cancer Cell* **1**, 269–277 (2002).
68. Alcantara Llaguno, S. *et al.* Malignant Astrocytomas Originate from Neural Stem/Progenitor Cells in a Somatic Tumor Suppressor Mouse Model. *Cancer Cell* **15**, 45–56 (2009).
69. Galvao, R. P. *et al.* Transformation of quiescent adult oligodendrocyte precursor cells into malignant glioma through a multistep reactivation process. *Proc. Natl. Acad. Sci.* **111**, E4214–E4223 (2014).
70. Alcantara Llaguno, S. R. *et al.* Adult Lineage-Restricted CNS Progenitors Specify Distinct Glioblastoma Subtypes. *Cancer Cell* **28**, 429–440 (2015).
71. Chow, L. M. L. *et al.* Cooperativity within and among Pten, p53, and Rb Pathways Induces High-Grade Astrocytoma in Adult Brain. *Cancer Cell* **19**, 305–316 (2011).
72. Jacques, T. S. *et al.* Combinations of genetic mutations in the adult neural stem cell compartment determine brain tumour phenotypes. *EMBO J.* **29**, 222–235 (2010).
73. Alcantara Llaguno, S. *et al.* Cell-of-origin susceptibility to glioblastoma formation declines with neural lineage restriction. *Nat. Neurosci.* **22**, 545–555 (2019).
74. Lee, J. H. *et al.* Human glioblastoma arises from subventricular zone cells with low-level driver mutations. *Nature* **560**, 243–247 (2018).
75. Lee, J. *et al.* Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer Cell* **9**, 391–403 (2006).
76. Allen, M., Bjerke, M., Edlund, H., Nelander, S. & Westermark, B. Origin of the U87MG glioma cell line: Good news and bad news. *Sci. Transl. Med.* **8**, 354re3-354re3 (2016).
77. Reynolds, B. & Weiss, S. Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. *Science (80-.).* **255**, 1707–1710 (1992).
78. Pollard, S. M. *et al.* Glioma Stem Cell Lines Expanded in Adherent Culture Have Tumor-Specific Phenotypes and Are Suitable for Chemical and Genetic Screens. *Cell Stem Cell* (2009) doi:10.1016/j.stem.2009.03.014.
79. Robertson, F. et al. Disease model and mechanisms,. *Press* (2019).
80. Bruggeman, S. W. M. *et al.* Bmi1 Controls Tumor Development in an Ink4a/Arf-

Independent Manner in a Mouse Model for Glioma. *Cancer Cell* **12**, 328–341 (2007).

81. Bressan, R. B. *et al.* Efficient CRISPR/Cas9-assisted gene targeting enables rapid and precise genetic manipulation of mammalian neural stem cells. *Development* **44**, dev.140855 (2017).
82. Suvà, M. L. *et al.* Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell* **157**, 580–594 (2014).
83. Rheinbay, E. *et al.* An Aberrant Transcription Factor Network Essential for Wnt Signaling and Stem Cell Maintenance in Glioblastoma. *Cell Rep.* **3**, 1567–1579 (2013).
84. Singh, D. K. *et al.* Oncogenes Activate an Autonomous Transcriptional Regulatory Circuit That Drives Glioblastoma. *Cell Rep.* **18**, 961–976 (2017).
85. Takahashi, K. *et al.* Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* **131**, 861–872 (2007).
86. Bulstrode, H. *et al.* Elevated FOXG1 and SOX2 in glioblastoma enforces neural stem cell identity through transcriptional control of cell cycle and epigenetic regulators. *Genes Dev.* **31**, 757–773 (2017).
87. Schepers, G. E., Teasdale, R. D. & Koopman, P. Twenty Pairs of Sox. *Dev. Cell* **3**, 167–170 (2002).
88. Wegner, M. All purpose Sox: The many roles of Sox proteins in gene expression. *Int. J. Biochem. Cell Biol.* **42**, 381–390 (2010).
89. O'Donnell, M., Hong, C. S., Huang, X., Delnicki, R. J. & Saint-Jeannet, J. P. Functional analysis of Sox8 during neural crest development in Xenopus. *Development* (2006) doi:10.1242/dev.02558.
90. Kim, J., Lo, L., Dormand, E. & Anderson, D. J. SOX10 maintains multipotency and inhibits neuronal differentiation of neural crest stem cells. *Neuron* (2003) doi:10.1016/S0896-6273(03)00163-6.
91. Akiyama, H., Chaboissier, M. C., Martin, J. F., Schedl, A. & De Crombrugghe, B. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes Dev.* (2002) doi:10.1101/gad.1017802.
92. Chaboissier, M. C. *et al.* Functional analysis os Sox8 and Sox9 during sex determination in the mouse. *Development* (2004) doi:10.1242/dev.01087.
93. O'Bryan, M. K. *et al.* Sox8 is a critical regulator of adult Sertoli cell function and male fertility. *Dev. Biol.* (2008) doi:10.1016/j.ydbio.2008.01.042.

94. Portnoi, M. F. *et al.* Mutations involving the SRY-related gene SOX8 are associated with a spectrum of human reproductive anomalies. *Hum. Mol. Genet.* (2018) doi:10.1093/hmg/ddy037.
95. Croft, B. *et al.* Human sex reversal is caused by duplication or deletion of core enhancers upstream of SOX9. *Nat. Commun.* (2018) doi:10.1038/s41467-018-07784-9.
96. Seymour, P. A. *et al.* SOX9 is required for maintenance of the pancreatic progenitor cell pool. *Proc. Natl. Acad. Sci. U. S. A.* (2007) doi:10.1073/pnas.0609217104.
97. Scott, C. E. *et al.* SOX9 induces and maintains neural stem cells. *Nat. Neurosci.* **13**, 1181–1189 (2010).
98. Martini, S. *et al.* A critical role for Sox9 in Notch-induced astrogliogenesis and stem cell maintenance. *Stem Cells* (2013) doi:10.1002/stem.1320.
99. Cheng, L. C., Pastrana, E., Tavazoie, M. & Doetsch, F. MiR-124 regulates adult neurogenesis in the subventricular zone stem cell niche. *Nat. Neurosci.* (2009) doi:10.1038/nn.2294.
100. Jo, A. *et al.* The versatile functions of Sox9 in development, stem cells, and human diseases. *Genes Dis.* **1**, 149–161 (2014).
101. Nishiguchi, S., Wood, H., Kondoh, H., Lovell-Badge, R. & Episkopou, V. Sox1 directly regulates the γ-crystallin genes and is essentiat for lens development in mice. *Genes Dev.* (1998) doi:10.1101/gad.12.6.776.
102. Bylund, M., Andersson, E., Novitch, B. G. & Muhr, J. Vertebrate neurogenesis is counteracted by Sox1-3 activity. *Nat Neurosci* **6**, 1162–1168 (2003).
103. Raverot, G., Weiss, J., Park, S. Y., Hurley, L. & Jameson, J. L. Sox3 expression in undifferentiated spermatogonia is required for the progression of spermatogenesis. *Dev. Biol.* **283**, 215–225 (2005).
104. Sarkar, A. & Hochedlinger, K. The Sox family of transcription factors: Versatile regulators of stem and progenitor cell fate. *Cell Stem Cell* **12**, 15–30 (2013).
105. Avilion, A. A. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17**, 126–140 (2003).
106. Masui, S. *et al.* Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol.* **9**, 625–635 (2007).
107. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676 (2006).

108. Takemoto, T. *et al.* Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells. *Nature* **470**, 394–398 (2011).
109. Pevny, L. H. & Nicolis, S. K. Sox2 roles in neural stem cells. *Int. J. Biochem. Cell Biol.* **42**, 421–424 (2010).
110. Wegner, M. & Stolt, C. C. From stem cells to neurons and glia: a Soxist's view of neural development. *Trends Neurosci.* **28**, 583–588 (2005).
111. Graham, V., Khudyakov, J., Ellis, P. & Pevny, L. SOX2 functions to maintain neural progenitor identity. *Neuron* **39**, 749–765 (2003).
112. Arnold, K. *et al.* Sox2 + adult stem and progenitor cells are important for tissue regeneration and survival of mice. *Cell Stem Cell* **9**, 317–329 (2011).
113. Ellis, P. *et al.* SOX2, a Persistent Marker for Multipotential Neural Stem Cells Derived from Embryonic Stem Cells, the Embryo or the Adult. *Dev. Neurosci.* **26**, 148–165 (2004).
114. Cavallaro, M. *et al.* Impaired generation of mature neurons by neural stem cells from hypomorphic Sox2 mutants. *Development* **135**, 541–557 (2008).
115. Ferri, A. L. M. Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development* **131**, 3805–3819 (2004).
116. Favaro, R. *et al.* Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. *Nat. Neurosci.* **12**, 1248–1256 (2009).
117. Ring, K. L. *et al.* Direct Reprogramming of Mouse and Human Fibroblasts into Multipotent Neural Stem Cells with a Single Factor. *Cell Stem Cell* **11**, 100–109 (2012).
118. Schmitz, M. *et al.* Erratum: Identification of SOX2 as a novel glioma-associated antigen and potential target for T cell-based immunotherapy. *Br. J. Cancer* **96**, 1928–1928 (2007).
119. Gangemi, R. M. R. *et al.* SOX2 silencing in glioblastoma tumor-initiating cells causes stop of proliferation and loss of tumorigenicity. *Stem Cells* **27**, 40–48 (2009).
120. Alonso, M. M. *et al.* Genetic and epigenetic modifications of Sox2 contribute to the invasive phenotype of malignant gliomas. *PLoS One* **6**, (2011).
121. Ikushima, H. *et al.* Autocrine TGF-β Signaling Maintains Tumorigenicity of Glioma-Initiating Cells through Sry-Related HMG-Box Factors. *Cell Stem Cell* **5**, 504–514 (2009).
122. MacLeod, G. *et al.* Genome-Wide CRISPR-Cas9 Screens Expose Genetic

Vulnerabilities and Mechanisms of Temozolomide Sensitivity in Glioblastoma Stem Cells. *Cell Rep.* **27**, 971-986.e9 (2019).

123. Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J. A. & Kosik, K. S. MicroRNA-145 Regulates OCT4, SOX2, and KLF4 and Represses Pluripotency in Human Embryonic Stem Cells. *Cell* (2009) doi:10.1016/j.cell.2009.02.038.
124. Schaefer, T. & Lengerke, C. SOX2 protein biochemistry in stemness, reprogramming, and cancer: the PI3K/AKT/SOX2 axis and beyond. *Oncogene* (2020) doi:10.1038/s41388-019-0997-x.
125. Williams, C. A. C., Soufi, A. & Pollard, S. M. Post-translational modification of SOX family proteins: Key biochemical targets in cancer? *Seminars in Cancer Biology* (2019) doi:10.1016/j.semcancer.2019.09.009.
126. Kamachi, Y. & Kondoh, H. Sox proteins: regulators of cell fate specification and differentiation. *Development* **140**, 4129–4144 (2013).
127. Kamachi, Y., Uchikawa, M. & Kondoh, H. Pairing SOX off: With partners in the regulation of embryonic development. *Trends Genet.* **16**, 182–187 (2000).
128. Nishimoto, M., Fukushima, A., Okuda, A. & Muramatsu, M. The Gene for the Embryonic Stem Cell Coactivator UTF1 Carries a Regulatory Element Which Selectively Interacts with a Complex Composed of Oct-3/4 and Sox-2. *Mol. Cell. Biol.* (1999) doi:10.1128/mcb.19.8.5453.
129. Ambrosetti, D. C., Basilico, C. & Dailey, L. Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. *Mol. Cell. Biol.* (1997) doi:10.1128/mcb.17.11.6321.
130. Kamachi, Y., Uchikawa, M., Tanouchi, A., Sekido, R. & Kondoh, H. Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes Dev.* (2001) doi:10.1101/gad.887101.
131. Engelen, E. *et al.* Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. *Nat. Genet.* **43**, 607–611 (2011).
132. Haas, S., Steplewski, A., Siracusa, L. D., Amini, S. & Khalili, K. Identification of a sequence-specific single-stranded DNA binding protein that suppresses transcription of the mouse myelin basic protein gene. *Journal of Biological Chemistry* vol. 270 12503–12510 (1995).
133. Baltz, A. G. *et al.* The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Mol. Cell* (2012)

doi:10.1016/j.molcel.2012.05.021.

134. McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–6 (2015).
135. van Riel, B. *et al.* A novel complex, RUNX1-MYEF2, represses hematopoietic genes in erythroid cells. *Mol. Cell. Biol.* **32**, 3814–22 (2012).
136. Boumahdi, S. *et al.* SOX2 controls tumour initiation and cancer stem-cell functions in squamous-cell carcinoma. *Nature* **511**, 246–253 (2014).
137. Jang, H. *et al.* O-GlcNAc Regulates Pluripotency and Reprogramming by Directly Acting on Core Components of the Pluripotency Network. *Cell Stem Cell* **11**, 62–74 (2012).
138. Khidekel, N., Ficarro, S. B., Peters, E. C. & Hsieh-Wilson, L. C. Exploring the O-GlcNAc proteome: Direct identification of O-GlcNAc-modified from the brain. *Proc. Natl. Acad. Sci. U. S. A.* (2004) doi:10.1073/pnas.0403471101.
139. Myers, S. A. *et al.* SOX2 O-GlcNAcylation alters its protein-protein interactions and genomic occupancy to modulate gene expression in pluripotent cells. *Elife* (2016) doi:10.7554/eLife.10647.
140. Jeong, C. H. *et al.* Phosphorylation of Sox2 cooperates in reprogramming to pluripotent stem cells. *Stem Cells* (2010) doi:10.1002/stem.540.
141. Van Hoof, D. *et al.* Phosphorylation Dynamics during Early Differentiation of Human Embryonic Stem Cells. *Cell Stem Cell* **5**, 214–226 (2009).
142. Tsuruzoe, S. *et al.* Inhibition of DNA binding of Sox2 by the SUMO conjugation. *Biochem. Biophys. Res. Commun.* **351**, 920–926 (2006).
143. Baltus, G. A. *et al.* Acetylation of Sox2 induces its nuclear export in embryonic stem cells. *Stem Cells* (2009) doi:10.1002/stem.168.
144. Zhao, H. yong, Zhang, Y. jun, Dai, H., Zhang, Y. & Shen, Y. fei. CARM1 mediates modulation of Sox2. *PLoS One* (2011) doi:10.1371/journal.pone.0027026.
145. Fang, L. *et al.* A Methylation-Phosphorylation Switch Determines Sox2 Stability and Function in ESC Maintenance or Differentiation. *Mol. Cell* **55**, 537–551 (2014).
146. Cui, C. P. *et al.* Dynamic ubiquitylation of Sox2 regulates proteostasis and governs neural progenitor cell differentiation. *Nat. Commun.* **9**, (2018).
147. ROTHSTEIN, R. J. One-step gene disruption in yeast. *Methods Enzym.* **101**, 202–211 (1983).
148. Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A. & Kucherlapati, R.

S. Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination. *Nature* **317**, 230–234 (1985).

149. THOMAS, K. High frequency targeting of genes to specific sites in the mammalian genome. *Cell* **44**, 419–428 (1986).
150. Capecchi, M. R. Altering the Genome Homologous Recombination by From ES Cells to Germ Line Chimera. *Science (80-.).* **244**, 1288–1292 (1989).
151. Lin, F. L., Sperle, K. & Sternberg, N. Recombination in mouse L cells between DNA introduced into cells and homologous chromosomal sequences. *Proc. Natl. Acad. Sci.* **82**, 1391–1395 (1985).
152. Rudin, N., Sugarman, E. & Haber, J. E. Genetic and physical analysis of double-strand break repair and recombination in Saccharomyces cerevisiae. *Genetics* **122**, 519–34 (1989).
153. Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.* **14**, 8096–8106 (1994).
154. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci.* **93**, 1156–1160 (1996).
155. Christian, M. *et al.* Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics* **186**, 757–761 (2010).
156. Klug, A. & Rhodes, D. Zinc fingers: a novel protein fold for nucleic acid recognition. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 473–82 (1987).
157. Moscou, M. J. & Bogdanove, A. J. A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science (80-.).* **326**, 1501–1501 (2009).
158. Boch, J. *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science (80-.).* **326**, 1509–1512 (2009).
159. Porteus, M. H. Chimeric Nucleases Stimulate Gene Targeting in Human Cells. *Science (80-.).* **300**, 763–763 (2003).
160. Bibikova, M. *et al.* Stimulation of Homologous Recombination through Targeted Cleavage by Chimeric Nucleases. *Mol. Cell. Biol.* **21**, 289–297 (2001).
161. Li, T. *et al.* Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res.* **39**, 6315–6325 (2011).
162. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29**, 143–148 (2011).

163. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.* **169**, 5429–33 (1987).
164. Jansen, R., Embden, J. D. A. van, Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575 (2002).
165. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science (80-.).* **315**, 1709–1712 (2007).
166. Koonin, E. V, Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **37**, 67–78 (2017).
167. Brouns, S. J. J. *et al.* Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science (80-.).* **321**, 960–964 (2008).
168. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
169. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-.).* **337**, 816–821 (2012).
170. Deveau, H. *et al.* Phage Response to CRISPR-Encoded Resistance in Streptococcus thermophilus. *J. Bacteriol.* **190**, 1390–1400 (2008).
171. Mojica, F. et al. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
172. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci.* **109**, E2579–E2586 (2012).
173. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, (2013).
174. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (80-.).* **339**, 819–823 (2013).
175. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science (80-.).* **339**, 823–826 (2013).
176. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
177. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **9**, (2018).
178. Essletzbichler, P. *et al.* Megabase-scale deletion using CRISPR/Cas9 to

generate a fully haploid human cell line. *Genome Res.* **24**, 2059–2065 (2014).

179. Shalem, O. *et al.* Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science (80-.).* **343**, 84–87 (2014).
180. Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).
181. Hart, T. *et al.* High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **163**, 1515–1526 (2015).
182. Yang, H. *et al.* One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell* **154**, 1370–1379 (2013).
183. Chen, Y. *et al.* Engineering Human Stem Cell Lines with Inducible Gene Knockout using CRISPR/Cas9. *Cell Stem Cell* **17**, 233–244 (2015).
184. Quadros, R. M. *et al.* Easi-CRISPR: A robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. *Genome Biol.* **18**, 1–15 (2017).
185. Dewari, P. S. *et al.* An efficient and scalable pipeline for epitope tagging in mammalian stem cells using Cas9 ribonucleoprotein. *Elife* **7**, 1–29 (2018).
186. Paquet, D. *et al.* Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125–129 (2016).
187. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
188. Billon, P. *et al.* CRISPR-Mediated Base Editing Enables Efficient Disruption of Eukaryotic Genes through Induction of STOP Codons. *Mol. Cell* **67**, 1068-1079.e4 (2017).
189. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).
190. Gilbert, L. A. *et al.* CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154**, 442–451 (2013).
191. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 (2013).
192. Maeder, M. L. *et al.* CRISPR RNA–guided activation of endogenous human genes. *Nat. Methods* **10**, 977–979 (2013).

193. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nat. Methods* **10**, 973–976 (2013).
194. Amabile, A. *et al.* Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. *Cell* **167**, 219-232.e14 (2016).
195. Liu, X. S. *et al.* Editing DNA Methylation in the Mammalian Genome. *Cell* **167**, 233-247.e17 (2016).
196. Choudhury, S. R., Cui, Y., Lubecka, K., Stefanska, B. & Irudayaraj, J. CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. *Oncotarget* **7**, (2016).
197. O'Connell, M. R. *et al.* Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **516**, 263–266 (2014).
198. Nelles, D. A. *et al.* Programmable RNA Tracking in Live Cells with CRISPR/Cas9. *Cell* **165**, 488–496 (2016).
199. Ran, F. A. *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154**, 1380–1389 (2013).
200. Shen, B. *et al.* Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods* **11**, 399–402 (2014).
201. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, (2014).
202. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24**, 1012–1019 (2014).
203. Hendel, A. *et al.* Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat. Biotechnol.* **33**, 985–989 (2015).
204. Eyquem, J. *et al.* Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection. *Nature* **543**, 113–117 (2017).
205. Capecchi, M. Altering the genome by homologous recombination. *Science (80-.).* **244**, 1288–1292 (1989).
206. Carthew, R. W. & Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–655 (2009).
207. Whitehead, K. A., Langer, R. & Anderson, D. G. Knocking down barriers: advances in siRNA delivery. *Nat. Rev. Drug Discov.* **8**, 129–138 (2009).
208. Varshavsky, A. Naming a targeting signal. *Cell* **64**, 13–5 (1991).
209. Davey, N. E. & Morgan, D. O. Building a Regulatory Network with Short Linear

Sequence Motifs: Lessons from the Degrons of the Anaphase-Promoting Complex. *Mol. Cell* **64**, 12–23 (2016).

210. Ella, H., Reiss, Y. & Ravid, T. The Hunt for Degrons of the 26S Proteasome. *Biomolecules* **9**, 230 (2019).
211. Varshavsky, A. N-degron and C-degron pathways of protein degradation. *Proc. Natl. Acad. Sci.* **116**, 358–366 (2019).
212. Bachmair, A., Finley, D. & Varshavsky, A. In vivo half-life of a protein is a function of its amino-terminal residue. *Science (80-.).* **234**, 179–186 (1986).
213. Bartel, B., Wünning, I. & Varshavsky, A. The recognition component of the N-end rule pathway. *EMBO J.* **9**, 3179–89 (1990).
214. Chatr-aryamontri, A., van der Sloot, A. & Tyers, M. At Long Last, a C-Terminal Bookend for the Ubiquitin Code. *Mol. Cell* **70**, 568–571 (2018).
215. Hochstrasser, M. UBIQUITIN-DEPENDENT PROTEIN DEGRADATION. *Annu. Rev. Genet.* **30**, 405–439 (1996).
216. Dohmen, R., Wu, P. & Varshavsky, A. Heat-inducible degron: a method for constructing temperature-sensitive mutants. *Science (80-.).* **263**, 1273–1276 (1994).
217. Harper, S. M. Structural Basis of a Phototropin Light Switch. *Science (80-.).* **301**, 1541–1544 (2003).
218. Renicke, C., Schuster, D., Usherenko, S., Essen, L.-O. & Taxis, C. A LOV2 Domain-Based Optogenetic Tool to Control Protein Degradation and Cellular Function. *Chem. Biol.* **20**, 619–626 (2013).
219. Bonger, K. M., Rakhit, R., Payumo, A. Y., Chen, J. K. & Wandless, T. J. General Method for Regulating Protein Stability with Light. *ACS Chem. Biol.* **9**, 111–115 (2014).
220. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods* **6**, 917–922 (2009).
221. Los, G. V. *et al.* HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **3**, 373–382 (2008).
222. Neklesa, T. K. *et al.* Small-molecule hydrophobic tagging-induced degradation of HaloTag fusion proteins. *Nat. Chem. Biol.* **7**, 538–543 (2011).
223. Buckley, D. L. *et al.* HaloPROTACS: Use of Small Molecule PROTACs to Induce Degradation of HaloTag Fusion Proteins. *ACS Chem. Biol.* **10**, 1831–1837 (2015).

224. Chung, H. K. *et al.* Tunable and reversible drug control of protein production via a self-excising degron. *Nat. Chem. Biol.* **11**, 713–720 (2015).
225. Pollard, S. M., Conti, L., Sun, Y., Goffredo, D. & Smith, A. Adherent neural stem (NS) cells from fetal and adult forebrain. *Cereb. Cortex* **16**, (2006).
226. Cho, S. W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Cold Spring Harb. Lab. Press Method* 132–141 (2014) doi:10.1101/gr.162339.113.Freely.
227. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
228. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–5 (2009).
229. Martella, A., Matjusaitis, M., Auxillos, J., Pollard, S. M. & Cai, Y. EMMA: An Extensible Mammalian Modular Assembly Toolkit for the Rapid Design and Production of Diverse Expression Vectors. *ACS Synth. Biol.* acssynbio.7b00016 (2017) doi:10.1021/acssynbio.7b00016.
230. Chung, H. K. *et al.* Tunable and reversible drug control of protein production via a self-excising degron. *Nat. Chem. Biol.* **11**, 713 (2015).
231. Turriziani, B. *et al.* On-Beads Digestion in Conjunction with Data-Dependent Mass Spectrometry: A Shortcut to Quantitative and Dynamic Interaction Proteomics. *Biology (Basel).* **3**, 320–332 (2014).
232. Bindea, G. *et al.* ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
233. Guzmán, C., Bagga, M., Kaur, A., Westermarck, J. & Abankwa, D. ColonyArea: An ImageJ plugin to automatically quantify colony formation in clonogenic assays. *PLoS One* **9**, 14–17 (2014).
234. Pollard, S. M. *et al.* Glioma Stem Cell Lines Expanded in Adherent Culture Have Tumor-Specific Phenotypes and Are Suitable for Chemical and Genetic Screens. *Cell Stem Cell* **4**, 568–580 (2009).
235. Engelen, E. *et al.* Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. *Nat. Commun.* **6**, (2015).
236. Mohammed, H. *et al.* Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. *Nat. Protoc.* **11**, 316–326 (2016).

237. Rafiee, M. R., Girardot, C., Sigismondo, G. & Krijgsveld, J. Expanding the Circuitry of Pluripotency by Selective Isolation of Chromatin-Associated Proteins. *Mol. Cell* **64**, 624–635 (2016).
238. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
239. Mellacheruvu, D. *et al.* The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Meth* **10**, 730–736 (2013).
240. Lai, J. S. & Herr, W. Ethidium bromide provides a simple tool for identifying genuine DNA-independent protein associations. *Proc. Natl. Acad. Sci.* **89**, 6958–6962 (2006).
241. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).
242. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science (80-.).* **347**, (2015).
243. Glaser, T., Pollard, S. M., Smith, A. & Brüstle, O. Tripotential differentiation of adherently expandable Neural Stem (NS) cells. *PLoS One* **2**, 2–5 (2007).
244. McIntosh, J. R. Mitosis. *Cold Spring Harb. Perspect. Biol.* **8**, a023218 (2016).
245. Kadauke, S. *et al.* Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. *Cell* **150**, 725–737 (2012).
246. Caravaca, J. M. *et al.* Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. *Genes and Development* vol. 27 251–260 (2013).
247. Festuccia, N. *et al.* Mitotic binding of Esrrb marks key regulatory regions of the pluripotency network. *Nat. Cell Biol.* **18**, 1139–1148 (2016).
248. Deluz, C. *et al.* A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. 2538–2550 (2016) doi:10.1101/gad.289256.116.5.
249. Teves, S. S. *et al.* A dynamic mode of mitotic bookmarking by transcription factors. *Elife* **5**, 1–24 (2016).
250. Liu, Y. *et al.* Widespread Mitotic Bookmarking by Histone Marks and Transcription Factors in Pluripotent Stem Cells. *Cell Rep.* **19**, 1283–1293 (2017).
251. Qi, D. *et al.* Mitotic phosphorylation of SOX2 mediated by Aurora kinase A is critical for the stem-cell like cell maintenance in PA-1 cells. *Cell Cycle* **15**, 2009–2018 (2016).
252. Turriziani, B. *et al.* On-Beads Digestion in Conjunction with Data-Dependent Mass Spectrometry: A Shortcut to Quantitative and Dynamic Interaction

Proteomics. *Biology (Basel).* **3**, 320–332 (2014).

253. Mallanna, S. K. *et al.* Proteomic analysis of Sox2-associated proteins during early stages of mouse embryonic stem cell differentiation identifies Sox21 as a novel regulator of stem cell fate. *Stem Cells* **28**, 1715–1727 (2010).
254. Cox, J. L. *et al.* The SOX2-Interactome in Brain Cancer Cells Identifies the Requirement of MSI2 and USP9X for the Growth of Brain Tumor Cells. *PLoS One* **8**, (2013).
255. Engelen, E. *et al.* Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. *Nat. Genet.* **43**, 607–611 (2011).
256. Jacobi, S. A. *et al.* Efficient homology-directed repair using single-stranded DNA templates. 7 (2017).
257. Jianjiong, G. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).
258. Cerami, E. *et al.* The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
259. Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013).
260. Nishikawa, R. *et al.* A mutant epidermal growth factor receptor common in human glioma confers enhanced tumorigenicity. *Proc. Natl. Acad. Sci.* **91**, 7727–7731 (1994).
261. Gómez-López, S. *et al.* Sox2 and Pax6 maintain the proliferative and developmental potential of gliogenic neural stem cells in vitro. *Glia* **59**, 1588–1599 (2011).
262. Carén, H. *et al.* Glioblastoma stem cells respond to differentiation cues but fail to undergo commitment and terminal cell-cycle arrest. *Stem Cell Reports* **5**, 829–842 (2015).
263. Jacobs, C. L., Badiee, R. K. & Lin, M. Z. StaPLs: Versatile genetically encoded modules for engineering drug-inducible proteins. *Nat. Methods* **15**, 523–526 (2018).
264. Zhu, W. *et al.* Precisely controlling endogenous protein dosage in hPSCs and derivatives to model FOXG1 syndrome. *Nat. Commun.* **10**, (2019).
265. Alber, A. B., Paquet, E. R., Biserni, M., Naef, F. & Suter, D. M. Single Live Cell Monitoring of Protein Turnover Reveals Intercellular Variability and Cell-Cycle Dependence of Degradation Rates. *Mol. Cell* **71**, 1079-1091.e9 (2018).

266. Lemmens, B. *et al.* DNA Replication Determines Timing of Mitosis by Restricting CDK1 and PLK1 Activation. *Mol. Cell* **71**, 117-128.e3 (2018).
267. McPhee, F. *et al.* Preclinical Profile and Characterization of the Hepatitis C Virus NS3 Protease Inhibitor Asunaprevir (BMS-650032). *Antimicrob. Agents Chemother.* **56**, 5387–5396 (2012).
268. Mosure, K. W. *et al.* Preclinical Pharmacokinetics and in Vitro Metabolism of Asunaprevir (BMS-650032), a Potent Hepatitis C Virus NS3 Protease Inhibitor. *J. Pharm. Sci.* **104**, 2813–2823 (2015).
269. Rago, F. *et al.* Biochemical and Biophysical Research Communications Degron mediated BRM / SMARCA2 depletion uncovers novel combination partners for treatment of BRG1 / SMARCA4-mutant cancers. *Biochem. Biophys. Res. Commun.* **508**, 109–116 (2019).
270. Government, T. D. of H. and A. A. Australian Public Assessment Report for asunaprevir. (2015).
271. Cox, J. L. *et al.* The SOX2-Interactome in Brain Cancer Cells Identifies the Requirement of MSI2 and USP9X for the Growth of Brain Tumor Cells. *PLoS One* **8**, (2013).
272. Engelen, E. *et al.* Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. *Nat. Genet.* **43**, 607–611 (2011).
273. Furuyama, K. *et al.* Continuous cell supply from a Sox9-expressing progenitor zone in adult liver, exocrine pancreas and intestine. *Nat. Genet.* **43**, 34–41 (2011).
274. Zhao, K. *et al.* RUNX1 contributes to the mesenchymal subtype of glioblastoma in a TGFβ pathway-dependent manner. *Cell Death Dis.* (2019) doi:10.1038/s41419-019-2108-x.
275. Liao, Y. *et al.* The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and Heart Disease. *Cell Rep.* **16**, 1456–1469 (2016).
276. Kwon, S. C. *et al.* The RNA-binding protein repertoire of embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1122–30 (2013).
277. Chu, C. *et al.* Systematic Discovery of Xist RNA Binding Proteins Resource Systematic Discovery of Xist RNA Binding Proteins. *Cell* **161**, 404–416 (2015).
278. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–126 (2011).
279. Da Rocha, S. T. & Heard, E. Novel players in X inactivation: Insights into Xist-mediated gene silencing and chromosome conformation. *Nat. Struct. Mol. Biol.*

24, 197–204 (2017).

280. Tsai, M. C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science (80-.).* **329**, 689–693 (2010).
281. Ng, S. Y., Bogu, G. K., Soh, B. & Stanton, L. W. The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. *Mol. Cell* **51**, 349–359 (2013).
282. Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. & Akoulitchev, A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* **445**, 666–670 (2007).
283. Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497–501 (2013).
284. Ariel, F. *et al.* Noncoding transcription by alternative rna polymerases dynamically regulates an auxin-driven chromatin loop. *Mol. Cell* **55**, 383–396 (2014).
285. Böhmdorfer, G. & Wierzbicki, A. T. Control of Chromatin Structure by Long Noncoding RNA. *Trends Cell Biol.* **25**, 623–632 (2015).
286. Herzel, L., Ottoz, D. S. M., Alpert, T. & Neugebauer, K. M. Splicing and transcription touch base: Co-transcriptional spliceosome assembly and function. *Nat. Rev. Mol. Cell Biol.* **18**, 637–650 (2017).
287. Palozola, K. C. *et al.* Mitotic transcription and waves of gene reactivation during mitotic exit. **122**, 119–122 (2017).
288. Yeo, G., Holste, D., Kreiman, G. & Burge, C. B. Variation in alternative splicing across human tissues. **5**, 1–15 (2004).
289. Grosso, A. R. *et al.* Tissue-specific splicing factor gene expression signatures. *Nucleic Acids Res.* **36**, 4823–4832 (2008).
290. Venables, J. P. *et al.* Cancer-associated regulation of alternative splicing. *Nat. Struct. Mol. Biol.* (2009) doi:10.1038/nsmb.1608.
291. Shuai, S. *et al.* The U1 spliceosomal RNA is recurrently mutated in multiple cancers. *Nature* (2019) doi:10.1038/s41586-019-1651-z.
292. Seiler, M. *et al.* Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. *Cell Rep.* (2018) doi:10.1016/j.celrep.2018.01.088.
293. Braun, C. J. *et al.* Coordinated Splicing of Regulatory Detained Introns within Oncogenic Transcripts Creates an Exploitable Vulnerability in Malignant Glioma. *Cancer Cell* **32**, 411-426.e11 (2017).

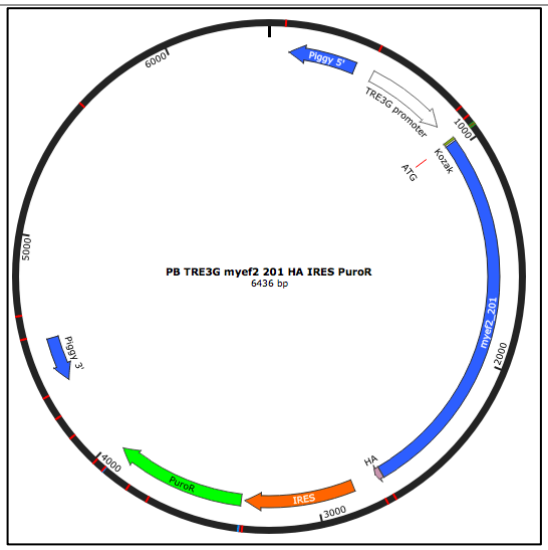
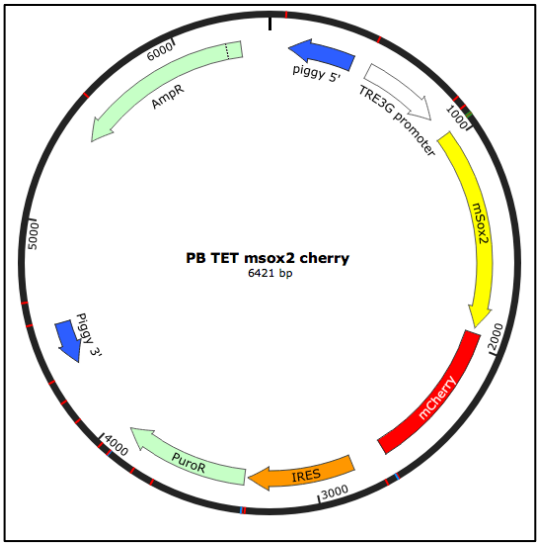
294. Ule, J. *et al.* CLIP identifies Nova-regulated RNA networks in the brain. *Science (80-.).* **302**, 1212–1215 (2003).
295. Lee, S. C. W. & Abdel-Wahab, O. Therapeutic targeting of splicing in cancer. *Nat. Med.* **22**, 976–986 (2016).
296. Kubota, T., Nishimura, K., Kanemaki, M. T. & Donaldson, A. D. The Elg1 Replication Factor C-like Complex Functions in PCNA Unloading during DNA Replication. *Mol. Cell* (2013) doi:10.1016/j.molcel.2013.02.012.
297. Li, S., Prasanna, X., Salo, V. T., Vattulainen, I. & Ikonen, E. An efficient auxin-inducible degron system with low basal degradation in human cells. *Nat. Methods* (2019) doi:10.1038/s41592-019-0512-x.
298. Sathyan, K. M. *et al.* An improved auxin-inducible degron system preserves native protein levels and enables rapid and specific protein depletion. *Genes Dev.* (2019) doi:10.1101/gad.328237.119.
299. Yesbolatova, A., Natsume, T., Hayashi, K. ichiro & Kanemaki, M. T. Generation of conditional auxin-inducible degron (AID) cells and tight control of degron-fused proteins using the degradation inhibitor auxinole. *Methods* (2019) doi:10.1016/j.ymeth.2019.04.010.
300. Sun, X. *et al.* PROTACs: great opportunities for academia and industry. *Signal Transduct. Target. Ther.* (2019) doi:10.1038/s41392-019-0101-6.
301. Bushweller, J. H. Targeting transcription factors in cancer — from undruggable to reality. *Nature Reviews Cancer* (2019) doi:10.1038/s41568-019-0196-7.
302. Guo, J., Liu, J. & Wei, W. Degrading proteins in animals: “PROTAC”tion goes in vivo. *Cell Research* (2019) doi:10.1038/s41422-019-0144-9.

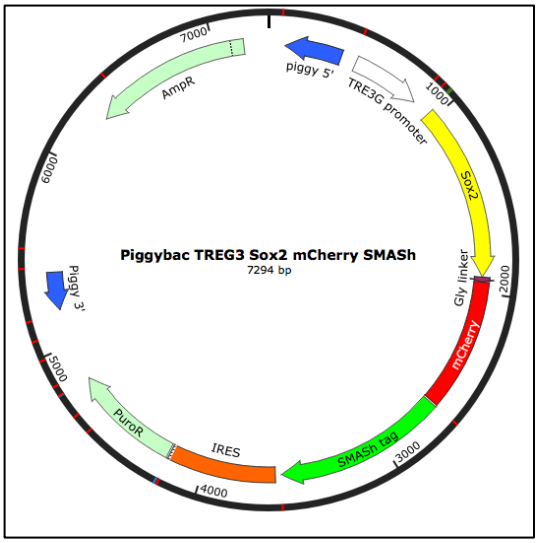
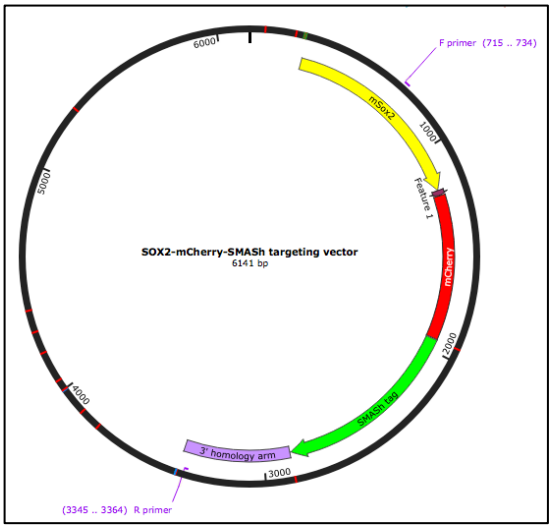
Chapter 7 Appendix

7.1 List of ssODN, targeting vectors and piggyBac plasmids

Gene	Use/Type	Sequence (5'-3') or plasmid map
Myef2	HA knock-in/ ssODN	CAGAGTCAGCGGAAAAGGCCTGCAGGATCATGAATGGCA TCAAGATCAGCGGCAGGGAAATCGATGTGCGCTTGGACC GCAATGCGTACCCATACGATGTTCCAGATTACGCTTAATTT CAAGCATGCTTGCAACCTTTCCTCATCTGTTTATGACTCTC CTAGTAAAAGTCATTTTTAGTAATGTTGTATGCTTACA
Myef2	mCherry knock-in/ dsDNA	ACATACTAAAGTTAGCACAGTAATCTAGCTTCTGTAATAGT CAAAGGCTAGCATGATTTGATAGGTGCACCAGCATGGGG GAGTATCGCAGGCTACGGCCATCTGCCTGGGTGTGTCTC AGTTGGCCTGAGCTGGAGTCTTACTCGTTTACATGGGATC ACTACTCCCTCCCTCCTGTTTAGTCTCAGCAGTGCTCTTG TTTAAAACTACTAGTTTCTCTGAGATGTCCCTTTAACAGCA TTGTTTTTATCGTTAATGTTTTAAATCTGCTGTCGTTTTATA ACCGTTTCCTTATTTGATGCCATCCTTAAAAATTTGAATCT AAATTGTCACAGTATTTGAAAAGTCAAATTGGTGGGCCCTT TTATGAACTGACTGAAACTATAGTTACTCTTGTGAGGTGCA AAGAGAAGATAAAAACAGAAGGGAGAAATATTCAGGAGAT TGGACAGTGAGCTACTGTATCATTTTCATTAAAGGTGTAAT GTCACCAGGGGTCATAATGAGTTGTTTGGCATTCACGCTA GCTAGCCTTTTTGCTATAACATCTACTTCAGAAATTGGGGA CTTAAAACCAAACTCCATGTCCAATTCCAGTTTCATGTAAG CTAATATGTATGTATATGTAGTCATTATGTAAGAATTGTATT CAGACTGTAAAATGATAATGCTCTCATGCATTTTATAGTAT TAAATTAATATGTTTAACCAAATTACCTGTGATCTCTCTCCA TTGCTGTGAAGAGTTGAAAAAATTTGTCTAGTGAAATGGAA CCCATTTAAATGGGCTCGTAATTGTGTAACTCTCCTTTCCT AGGTCATGTAATGTTTGCAGAGATAAAGATGGAGAATGGC AAGTCAAAAGGCTGTGGGACAGTCAGGTTTGAATCTGCAG AGTCAGCGGAAAAGGCCTGCAGGATCATGAATGGCATCA AGATCAGCGGCAGGGAAATCGATGTGCGCTTGGACCGCA ATGCGGGCAGCGGCGGCGGCAGCATGGTGAGCAAGGGC GAGGACAAGTGGGAGCGCGTGATGAAGGAGTTCATGCGC TTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAG TTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGA GGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTG

		GCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGT TCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCG ACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTT CAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCG TGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCG AGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCC CTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTG GGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCG CCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGG ACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACA AGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAAC GTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACT ACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCC ACTCCACCGGCGGCATGGACGAGCTGTACAAGTGATAAT TTCAAGCATGGTTGGAACCTTTCCTCATCTGTTTATGACTC TCCTAGTAAAAGTCATTTTTAGTAATGTTGTATGCTTACAA ATGCTGTAAAAATGAACTTTTACAACTCCCACCAGCTATTA ACAGGATAGTGTGGAAAATGTACTGTGAGTTTTTTGTTTTT TGTTTTTTGTTTTTTTTTTTTTTTTTTTTTGTTCTCAAGTTT GAGTTTCTAAAGACAGCACATCTGATCATTCAGTTTCAGTG GATGGACATACCATTTTTAATGAAATAAGCCATTTTGTTAT TTTCAGTAGCAGAGTTTGTCTGTTCCAGTTTTCCCGACCTT GGTGTAAGCTCACAGATACTTTGTTTTCTTTTAAAATCTTT GCTTTGTATGTGAGTTAAAGGAAAGAAAGGGCTCAGACAA ATTAGGATGGATTTTGGTTTGGTTTAAATTACTTTTCTCTG CTTATAAAGAATAGTAATTAAGTGTTAGACTCTAAAGTTGA AGATGCTTTTTATTTGACCTAAATGAAGACATGAATTTTCTT CTTTTTGCCCCTCTACCATTCCATCTTTCCACATAACACTA TTAAAAAATATCAAACTCCACAACCCCTTATTCTATTATTTC CAATAATTCCAATTTCATATAGAACTGATAAAGTAGCAAGT CCTAAGTATAACACTAGGCAGACCACCCCCAACTTTCGGT CTAGTTTCCAGCCATTAAAATGAACTGCTAAGAACAGAAAT AAAATTGAAATGTTGAGAGAGATTGTTATATATGTGAGTCC CTTGCTATTCACTTCTATAGGAGCAGATGCATTCGTAAATG CAGTCTTAATAAACCAGGGAAGACCTAGGCATAGCATATC AAACACATTGGATCCCACGATGTTAGATATAGCCATGTCT CCTTTACCTGCAGAAGAATAGCATGT

Myef2	DOX inducible overexpression/ piggyBac plasmid	 Plasmid contains: • TRE3G • *Myef2* (isoform 201) plus C terminal HA tag • IRES • Puromycin selection cassette
Sox2	DOX inducible overexpression/ piggyBac plasmid	 Plasmid contains: • TRE3G • *Sox2*-mCherry • IRES • Puromycin

Sox2	DOX inducible overexpression/ piggyBac plasmid	 Plasmid contains: • TRE3G • *Sox2*-mCherry-SMASh • IRES • Puromycin
Sox2	Targeting vector/ plasmid (PCR product is used for transfection)	 Plasmid contains: • 500bp homology arms • *Sox2*-mCherry-SMASh

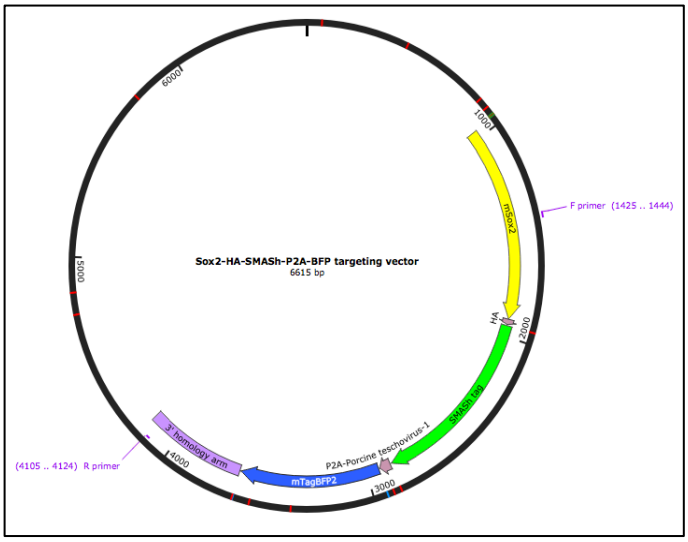
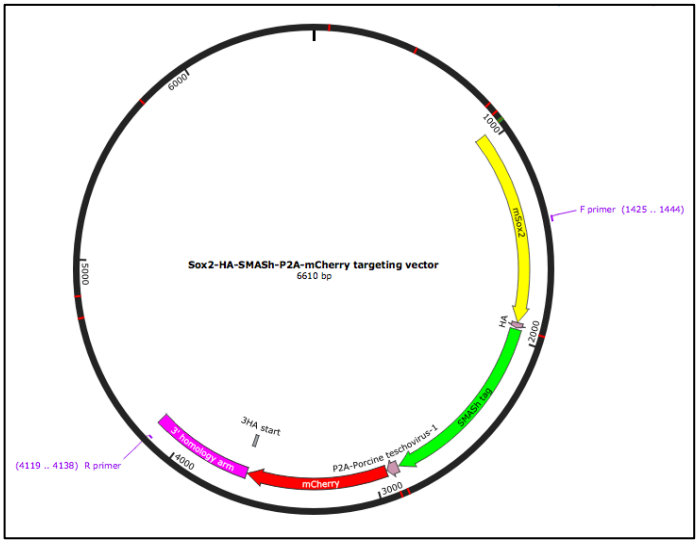
Sox2	Targeting vector/ plasmid (PCR product is used for transfection)	 Plasmid contains: • 500bp homology arms • *Sox2*-HA-SMASh-P2A-BFP
Sox2	Targeting vector/ plasmid (PCR product is used for transfection)	 Plasmid contains: • 500bp homology arms • *Sox2*-HA-SMASh-P2A-mCherry

Table 8: List of ssODN, targeting vectors and piggyBac plasmids used

7.2 Protein interactors of MYEF2-HA identified by mass spectrometry

Gene.names	Q.value	Score	log2.av.ratio	gm.pval
A0A140T8P5	0	92.4790	11.4440	0.0124
Myef2	0	323.3100	8.9526	0.0087
Rsrc2	0	68.6730	8.9341	0.0379
Rab3b	0	22.3760	7.8571	0.0236
Snip1	0	70.3630	6.2991	0.0149
Fam64a	0	16.9570	6.2912	0.0011
Dido1	0	3.9610	6.2731	0.0004
Ddx46	0	323.3100	6.2113	0.0136
Ahcyl2	0.0014058	3.0926	6.1672	0.0021
Naca	0.0041247	2.8036	5.9168	0.0038
Rsrc1	0	67.7650	5.7752	0.0062
Gldc	0	3.9697	5.6515	0.0356
Mmtag2	0	86.1790	5.6004	0.0272
Rab3a	0	16.0050	5.5660	0.0020
Tnrc6a	0	32.0450	5.3826	0.0005
Rab7a	0.00048239	3.4820	5.2746	0.0004
Ighv1-18	1	-2.0000	5.2676	0.0015
Ctsc	0	19.2890	5.2540	0.0020
Rfx4	0	14.5560	5.1448	0.0048
Rrm1	0.0027701	2.9195	5.1393	0.0005
Tfrc	0	157.7200	5.0252	0.0001
Sf3b6	0	29.1720	5.0170	0.0417
Dock10	0	5.2659	5.0047	0.0048
Hnrnph3	0	27.4220	4.8105	0.0012
Zcchc6	0	5.8302	4.8095	0.0020
Ide	0	169.4200	4.7995	0.0107
Olfm2	0	8.5537	4.7198	0.0014
Tmed10	0.0071652	2.4774	4.5463	0.0015
AU019823	0	44.6830	4.5299	0.0128
Taf2	0	323.3100	4.4166	0.0432
Commd6	0	5.4391	4.2750	0.0017
Hnrnpm	0	323.3100	4.0639	0.0004

Rbm39	0	323.3100	3.8430	0.0395
S100a13	0	11.8060	3.7567	0.0194
Nudt21	0	112.4800	3.7337	0.0033
Hexim1	0	173.5400	3.7332	0.0318
Scaf11	0	100.9100	3.5904	0.0431
Rtn1	0.0079752	2.4003	3.5804	0.0090
Fip1l1	0	284.5400	3.5256	0.0136
Cpsf1	0	268.3100	3.3915	0.0177
Nusap1	0	72.1200	3.3740	0.0346
Ncan	0	262.5400	3.3264	0.0322
Apobec1	0	32.5530	3.1330	0.0144
Wdr33	0	215.9800	3.0568	0.0087
Cactin	0	132.1500	2.9973	0.0415
U2af1l4	0	120.5600	2.9956	0.0285
Cdk13	0	230.8500	2.8650	0.0394
Snrnp27	0	124.2800	2.8180	0.0252
Srek1ip1	0	17.5310	2.7664	0.0343
Cpsf2	0	149.4000	2.7262	0.0168
Marcks	0	11.8360	2.7237	0.0022
Sf3b1	0	323.3100	2.6808	0.0003
Zrsr1	0	9.3144	2.6761	0.0088
Surf6	0	15.6150	2.5916	0.0475
Mlf2	0	37.5790	2.5685	0.0170
Rpl29;Gm17669;Gm3550	0	43.6430	2.5385	0.0486
Rbm3	0	27.6520	2.5370	0.0128
Riok2	0	15.8520	2.5367	0.0028
Map2	0	45.4580	2.4893	0.0000
Nol12	0	33.6890	2.4524	0.0261
Cpsf3	0	173.4500	2.4266	0.0141
Mark3	0	32.3970	2.4186	0.0260
Sf1	0	22.2500	2.4161	0.0303
Prkcb	0	54.9540	2.4142	0.0014
Pcbp1	0	323.3100	2.3792	0.0184

Mrpl38	0	109.9100	2.3212	0.0377
Calu	0	123.9300	2.3122	0.0166
Cpsf7	0	50.7140	2.2404	0.0039
Sap30bp	0	29.7260	2.2385	0.0038
Ahcyl1	0	34.2660	2.2004	0.0069
Rbm7	0	29.6210	2.1350	0.0308
Hmgxb4	0	26.5050	2.0897	0.0394
Vav3	0	112.0300	2.0781	0.0114
Hnrnpd	0	73.2870	2.0638	0.0011
Cluh	0	25.7090	2.0232	0.0310
Sox2	0	170.9200	2.0123	0.0222

Table 9: Protein interactors of MYEF2-HA by mass spectrometry