University of Nottingham
UK | CHINA | MALAYSIA

The role of TET2 protein and oxidised forms of 5-methylcytosine in brain tumours

Maria Eleftheriou MSc, BSc
School of Medicine
Division of Cancer and Stem Cells

Thesis submitted to the University of Nottingham for the Degree of
Doctor of Philosophy
June 2020

Acknowledgements

I would like to thank my supervisors Dr Alexey Ruzov and Dr Beth Coyle for their time, support and encouragement over the last four years at University of Nottingham. I am also grateful to the technical and administrative staff of cancer and stem cells department and CBTRC in particular who have provided me with invaluable help and support.

I am especially grateful to Dr Alexander Kondrashov and Dr Robert Marcus for being so generous with their time help and expertise in CRISPR/Cas9 design and confocal microscopy respectively.

A huge thanks to all those who I have truly enjoyed working with over the past few years. In particular, I would like to thank the people that became my family in Nottingham; Dr Lara Lewis, Dr Ashley Ramsawhook and Paige Lowe, their constant help, support, encouragement and willingness to help with anything has made the last four years such an enjoyable experience. I would also like to thank the A54 people for their support, kind words and amazing conversations during the stressful and happy moments over the last years at University of Nottingham. A special thanks to Emma Collins, for being such an amazing colleague and flatmate, her support, motivation and friendship made the COVID-19 lockdown experience more affordable and fun.

On a more personal note, I would like to thank my friends Ioanna Hadjiyianni and Stiliana Georgiou, as well as all my Cypriot friends, for the constant support, friendship and countless drunk nights. Lastly, a huge thanks to my parents and my brothers, I would never have had the privilege to accomplish my goals if it was not for their constant help, support and encouragement.

Thesis statement and contributions

I hereby declare that this thesis has been composed by myself and has not been submitted for any other degree previously. This project was supervised by Dr Alexey Ruzov and Dr Beth Coyle. Acknowledgements of specific procedures not performed by myself are stated; otherwise, the work described is my own.

Dr Lara Lewis performed all the qPCR experiments as well as the library preparation of 5caC DIP for the hiPSCs cell line. All the mass spectrometric analysis of oxi-mC marks was performed at Ryszard Olinsky laboratory at Nicolaus Copernicus University, Poland. Next Generation Sequencing (NGS) on the prepared libraries were performed by DeepSeq staff (University of Nottingham). Bioinformatics analysis of raw NGS sequencing data was performed by Dr Tom Giles at the Advanced Data Analysis Centre (ADAC) (University of Nottingham).

Abstract

DNA methylation (5-methylcytosine, 5mC) is the major epigenetic modification involved in transcriptional regulation. Ten-eleven translocation (TET) proteins can enzymatically oxidise 5mC producing 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). According to several reports, 5hmC levels are reduced in human tumours; however, the distribution and the exact biological role of TET-dependent oxidation to 5fC and 5caC in cancers is poorly studied.

Here, using a range of techniques including, immunocytochemistry, 2 dimensional ultra-performance liquid chromatography coupled with tandem mass-spectrometry (2D-UPLC–MS/MS) and CRISPR/Cas9 gene editing we studied the presence and the distribution of 5caC/5hmC in brain tumours cell lines. Our ultimate goal was to examine the role of TET2 in the oxidation of 5mC to 5hmC/5caC in glioblastoma multiforme (GBM) pathogenesis and the role of 5caC potential readers in the interpretation of active DNA demethylation in medulloblastoma (MB) (paediatric brain tumour) cell lines.

We found that, while GBM cell lines exhibit low levels of 5hmC, they are, rather unexpectedly, characterised by detectable 5caC levels. Remarkably, 5caC content in GBM corresponds to elevated levels of *TET2* transcript. We next used CRISPR/Cas9 to knockout (KO) *TET2* in the LN18 GBM cell line. 5hmC levels were significantly reduced in TET2 KO cell line. Moreover, we showed significantly reduced tumorigenic ability of TET2 KO cells in parallel with the reduction of the transcript levels of key glial cancer stem cell markers. We also showed that SMARCC2 and RCOR2 transcription factors with unique roles in neurogenesis, are highly expressed in paediatric brain tumours, while their transient knockdown resulted at impaired 5caC levels in MB cell lines; indicating their role as 5caC readers. Our compiled data show the unique epigenetic signatures of adult and paediatric brain tumours, the role of TET2 in active DNA demethylation and its potential contribution to tumorigenesis.

Table of contents

List of figures

List of tables

Abbreviations

Abbreviation	Description
(R) - 2- hydroxyglutarate	(R) -2-HG
2D-UPLC-MS/MS	Two dimensional ultra-performance liquid chromatography with tandem mass spectrometry
2-HG	2-hydroxyglutarate
2-OG	2-oxoglutarate
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxylmethylcytosine
5hmU	5-hydroxylmethyluracil
5mC	5-methylcytosine
8-oxodG	8-oxoguanine
aa	Aminoacid
AID	Activation-induced cytidine deaminase
AML	Acute myeloid leukaemia
ANKRD17	Ankyrin Repeat Domain 17
APE1	Apurinic/apyrimidinic endonuclease 1
APOBEC	Apolipoprotein B mRNA Editing Catalytic Polypeptide-like
ARRB2	Beta-arrestin 2
BDNF	Brain-derived neurotrophic factor
BER	Base excision repair
BMP4	Bone morphogenetic protein 4
BP	Base Pairs
BSA	Bovine serum albumin
BSD	Blasticidin
caMAB-Seq	5caC methylase-assisted bisulfite sequencing
CDK6	Cyclin dependent kinase 6
cDNA	Complementary DNA
cfDNA	Circulating free DNA
CNS	Central Nervous System
CRISPR	Clustered regularly interspaced short

	palindromic repeats
crRNA	CRISPR RNA
CSC	Cancer stem cell
CTTNB1	Catenin Beta 1
C	Cytosine
DAPI	4’,6-diamidino-2-phenylindole
DIP	DNA immunoprecipitation
DMEM	Dulbecco's modified eagles medium
DMSO	Dimethyl-sulfoxide
DNA	Deoxyribonucleic acid
DNMT	DNA-methyltransferase
DSB	Double strand break
DSBH	Double stranded β helix
dsDNA	Double stranded DNA
DTT	Dithiothreitol
DYRKB1	Dual specificity tyrosine-phosphorylation-regulated kinase B
EGF	Epidermal derived growth factor
EGFR	Epidermal growth factor receptor
EGL	External granule layer
EHMT1	Euchromatic Histone Lysine Methyltransferase 1
EIA	Biotin-avidin mediated enzyme-ased immunoassay
ELDA	Extreme limited dilution assay
ELISA	Enzyme-Linked immunosorbent assay
EMSA	Electrophoretic mobility shift assay
EMT	Epithelial-mesenchymal transition
EPN	Ependymoma
ERG	Erythroblast transformation-specific related gene
ESCs	Embryonic stem cells
FACS	Fluorescent activated cell sorting
FBS	Foetal bovine serum
FDA	U.S. Food and drug administation
Fe (II)	Iron (II)
FGF	Fibroblast growth factor
FIZ1	FLT3 Interacting Zinc Finger 1
GABRA1	Gamma-Aminobutyric Acid Type A Receptor

	Subunit Alpha1
GAPDH	Glyceraldhyde 3- phospate dehydrogenase
GBM	Glioblastoma multiforme
G-CIMP	CpG island methylator phenotype
GFAP	Glial fibrillary acidic protein
GFI1	Factor Independent 1
GFP	Green fluorescent protein
GIN	Glioblastoma Invasive
GLAST	GLutamate ASpartate Transporter 1
GLI2	GLI family zinc finger 2
Gli3	GLI Family Zinc Finger 3
GNPs	Granule neuron precursors
GO	Gene ontology
GS linker	Glycine-Serine linker
H3K27me3	Trimethylation at lysine 27 of histone H3
H3K4	Lysine 4 of histone 3
H3K4me3	Trimethylation at lysine 4 of Histone H3
H3Kme2/3	Di and trimethylation at lysine of histone H3
HCl	Hydrochloric acid
HDAC2	Histone deacetylase 2
HDR	Homology-directed repair
hESCs	Human embryonic stem cells
HHIP	Hedgehog Interacting Protein
HIF	Hypoxia inducible factors
hiPSCs	Human induced pluripotent stem cells
HPLC	High performance liquid chromatography
HSC	Haematopoietic stem cells
ICM	Inner cell mass
ICRs	Imprinting control regions
IDH	Isocitrate dehydrogenase
IL-6	Interleukin-6
INSM1	Insulinoma-associated 1
iPSCs	Induced pluripotent stem cells
JBP1	Base J binding protein 1
KDM1A	Lysine Demethylase 1A
KDM6A	Lysine demethylase 6A

KLF4	Kruppel Like Factor 4
KO	Knock out
LC-ESI-MS/MS	Liquid chromatography-electrospray ionization-tandem mass spectrometry analysis
LC-MS/MS	Liquid chromatography coupled with mass spectrometry
LHA	Left homology arm
LIF	Leukaemia inhibitor factor
LIG3	Ligase 3
lincRNA	Long non-coding RNA
LINE	Long interpersed nuclear element
LSD1	Lysine specific histone demethylase
LTR	Long terminal repeats
MAB-Seq	Methylase-assisted bisulfite sequencing
MAPK	Mitogen protein kinase
MB	Medulloblastoma
MBD	Methyl-CpG binding domain
MDS	Myelodysplastic syndrome
meCP1	Histone deacetylase complex
MEFs	Mouse embryonic fibroblasts
MEM	Minimal essential medium
mESCs	Mouse Embryonic Stem Cells
MET	Mesenchymal-to epithelial transition
MGMT	O-6-methylguanine- DNA methyltransferase
miRNA	Micro RNA
MLL	Mixed lineage leukaemia
MS	Multiple sclerosis
MSH6	MutS Homologue 6
MS-HRM	Methylation-sensitive high-resolution melting
MS-MCA	Methylation-sensitive melting curve analysis
MUM1	Multiple myeloma 1
NaCl	Sodium chloride
NCBI	National Center of Biotechnology information
NEEA	Non-essential aminoacids
NEFL	Neurofilament light polypeptide
NEIL1	Nei Like DNA Glycosylase 1

NF1	Neurofibromin 1
NGS	Next generation sequencin
NHEJ	Non-homologous end joining
NPC	Neural progenitor cells
NRF1	Nuclear Respiratory Factor 1
NRSF	Neuron-restrictive silencer factor
NSCs	Neural stem cells
OCT4	Octamer binding transcription factor 4
O-GlcNac	O-linked B-N-acetylglucosamine
OGT	O-linked B-N-acetylglucosamine transferase
OLIG2	Oligodendrocyte transcription factor 2
OTX2	Orthodenticle homeobox 2
oxi-mCs	Oxidised-methylcytosines
PAM	Short protospacer-adjacent motif
PATCH1	Patched
Pax6	Paired box 6
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDGF	Platelet Derived Growth Factor
PDGFRA	Plateled Derived growth factor receptor A
PFA	Parafolmaldehyde
PGCs	Primordial germ cells
PI3K	Phosphoinostitide 3- kinase
PNKP	Polynucleotide Kinase 3'-Phosphatase
PolII	RNA polymerase II
PRDM6	PR/SET domain 6
PRP8	Pre-mRNA splicing factor 8
QMC	Queens Medical Centre
RA	Retinoic acid
RBM14	RNA Binding Motif Protein 14
RCOR2	Rest Corepressor 2
RE1	Neuronal repressor element 1
REST	RE1-Silencing Transcription factor
RHA	Right homology arm
RNA	Ribonucleic acid
RP-HPLC	Reversed-phase high-performance liquid

	chromatography
RREB1	Ras Responsive Element Binding Protein 1
rSAP	Shrimp alkaline phosphatase
RT PCR	Reversed transcriptase polymerase chain reaction
SALL4	Spalt-like transcription factor 4
SD	Standard deviation
SEM	Standard error of the mean
SFMBT2	Scm Like With Four Mbt Domains 2
sgRNA	Single guided RNA
SHH	Sonic hedgehog
shRNA	Short hairpin RNA
siRNA	Small interfering RNA
SMARCC2	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator of Chromatin Subfamily C member 2
SMO	Smoothened
SNCAIP	Synuclein alpha interacting protein
Sox2	SRY-Box Transcription Factor 2
SpCas9	*Streptococcus pyogenes* Cas9
SRA domain	RING-associated domain
SSB	Single strand break
ssRNA	Single strand RNA
STR	Short tandem repeats
SYT1	Synaptotagmin 1
TAB-Seq	TET-Assisted Bisulfite Sequencing
TALEN's	Transcription activator-like effector nucleases
Taq	*Thermus aquaticus*
TCA	Tricarboxylic acid
TCGA	Cancer Genome Atlas
TDG	Thymine DNA glycosylase
TERT	Telomerase reverse transcriptase
TET	Ten eleven translocation
TF	Transcription factor
TGF-β	Transforming growth factor -β
THYN1	Thymocyte Nuclear Protein 1
TMZ	Temozolomide

tracRNA	Trans-activating crRNA
TSS	Transcription start sites
UHPLC-MS/MS	Ultra high performance liquid chromatography-with tandem mass spectrometrhy
UHRF1	ubiquitin-like, containing PHD and RING finger domains 1
UV	Ultra Violet
VP16	Viral protein 16
WIZ	WIZ Zinc Finger
WT1	Wilm's tumour 1
ZBTB	Zinc Finger And BTB Domain Containing
ZEB1	Zinc finger E-box-binding homeobox 1
ZFNs	Zinc-finger nucleases
Zn	Zinc cations
ZNF	Zinc finger
ZSCAN26	Zinc Finger and SCAN Domain containing 26
α-KG	α-ketoglutarate

Chapter 1. Introduction

1.1. Epigenetics

Genetics is the branch of biology that studies the changes in the DNA sequence that result in heritable changes in gene activity. These alterations include insertions, deletions, point mutations and translocations (Moore *et al.* 2012). Whereas, epigenetics is the branch of biology that studies the heritable alterations in gene activity without any changes in DNA sequence (Moore *et al.* 2012). Over the years, the term "Epigenetics" has been used to describe events that could not be explained by genetic principles. The first definition of the term was coined by Conrad Waddington in 1942 and defined epigenetics as "The branch of biology which studies the casual interactions between genes and their products, which bring the phenotype into being" (Reviewed by: Waddington 2012). Since then, the meaning and the field of epigenetics has changed and is still evolving.

Although all the cells in an organism share the same genetic material not all the genes are expressed in the same manner in all cell types. This diversity is due to the epigenetic mechanisms that control gene expression; being responsible for the repression or activation of certain genes in different cell types and developmental stages (Moore *et al.* 2012). Epigenetic factors contributing to embryonic development, disease development, pathogenesis and response to treatment include: methylation of DNA at the 5^{th} carbon of cytosine, which is generally thought to be associated with gene repression (Goldberg, Allis and Bernstein, 2007; Li and Zhang, 2014); alterations in histone modifications and regulation through non-coding RNAs (Zoghbi & Arthur L. Beaudet 2004).

1.2. DNA methylation

1.2.1. The role of DNA methylation and its genomic distribution in mammals

The most predominant epigenetic modification in mammals is the covalent addition of a methyl group at the fifth carbon of the pyrimidine ring of cytosine called 5-methylcytosine (5mC). 5mC was first discovered in *tubercle bacillus* DNA in 1925 (Johnson & Coghill 1925) and subsequently in *calf thymus* in 1948 (Hotchkiss 1948); suggesting that this modification exists naturally in DNA. Although it was proposed that it might have a specific biological role on gene regulation, it was a few decades later before it was proven that DNA methylation is involved in gene regulation and cell differentiation (Compere & Palmiter 1981; Holliday & Pugh 1975).

DNA methylation is a critical epigenetic modification that regulates gene expression in plants and animals. Methylation at both adenine and cytosine residues in prokaryotes contributes to the protection of the cell from foreign genome, usually bacterial or viral (Bickle & Kruger 1993). In animals, 5mC occurs primarily at cytosine residues in CG rich regions (Kumar *et al.* 2018). Nevertheless, non-CpG methylation also occurs in embryonic stem cells (ESCs) as well as induced pluripotent stem cells (iPSCs) (Lister *et al.* 2009; Ramsahoye *et al.* 2000).

In the human genomes, 70-80% of CpG sites are methylated mainly in sequences of repetitive elements, satellite DNA, gene bodies and non-repetitive intergenic DNA (Li & Zhang 2014). 5mC is mainly associated with gene repression when present at either promoter or enhancer regions (Charlet *et al.* 2016). However, when present in the gene body it might repress or enhance transcriptional activity (Buck-Koehntop & Defossez 2013; Spruijt & Vermeulen 2014). CpG islands, the CpG-rich regions located in the promoters of mammalian genes associated with tissue-

specific expression, are generally un-methylated (Bird *et al.* 1985; Takai & Jones 2002). DNA methylation in GC-rich gene promoters is generally associated with gene repression, while transcribed genes are associated with low methylation levels around transcription start sites (TSS) and high levels in the gene body (Laurent *et al.* 2010; Lister *et al.* 2009). The position of the methylation in the transcript influences its relationship with regulating gene activity. Methylation close to the TSS blocks the initiation of gene transcription, however methylation at the gene body might stimulate transcription elongation (Jones 2012).

Methylation in repeat regions such as centromeres is crucial for chromosomal stability (chromosome segregation at mitosis) (Moarefi & Chédin 2011). It is also important for suppressing the expression of transposable elements with a role in genome stability (Moarefi & Chédin 2011). Genes that are predominately repressed, such as imprinted genes; genes located on the inactive X chromosome; or genes that are exclusively expressed in germ cells, usually have methylated promoter CpG sites (Jones 2012). DNA methylation has a crucial role in the establishment of genomic imprinting at either parental or maternal alleles (Maupetit-Méhouas *et al.* 2016). Genomic imprinting is the expression of a gene from one allele (paternal or maternal) from a parent in an origin specific manner; meaning that only one of the two inherited alleles is expressed (Renfree *et al.* 2013). It has also been suggested that DNA methylation might have a role in splicing as whole-genome studies showed that exons are more highly methylated than introns and transitions of their methylation status occurs in exon-intron boundaries (Laurent *et al.* 2010). Moreover, exons show increased nucleosome occupancy levels in comparison to introns (Schwartz *et al.* 2009) and nucleosomes are preferential sites for DNA methylation.

1.2.2. DNA methyltransferases

The addition and the maintenance of the methyl group into DNA is catalysed by the members of DNA methyltransferase (DNMT) family: DNMT1, DNMT3a, DNMT3b and DNMT3L (Fig. 1.1) (Bestor 2000). They are highly conserved and have similar amino acid sequences. The N-terminal domain is a regulatory domain that allows the enzymes to enter the nucleus and recognize the specific loci whereas the C- terminal domain contains the catalytic domain which is responsible for the enzymatic activity of the proteins (Zhang & Xu 2017).

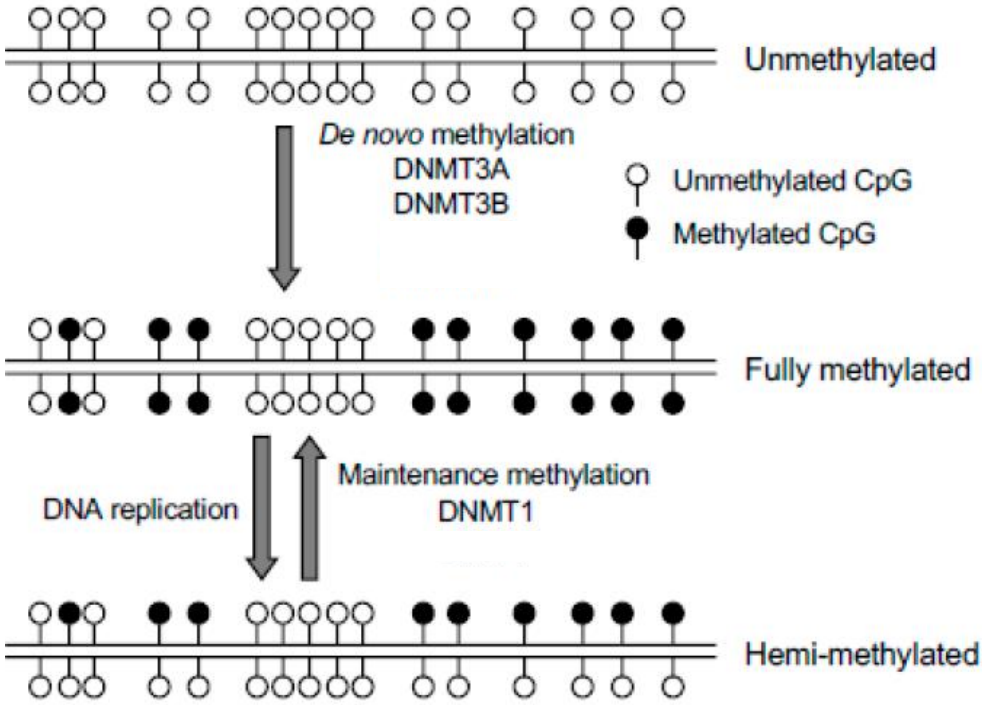


Figure 1.1: The principle of *de novo* and maintenance DNA methylation.

DNMT3a and DNMT3b are responsible for *de novo* methylation and methylate un-methylated CpG sites to establish methylation patterns, whereas DNMT1 maintenance methyltransferase, methylates hemi-methylated DNA strands during DNA replication. Figure adapted from (Zeng & Chen 2019).

DNMT1 is highly expressed in all the mammalian tissues it is called the maintenance methyltransferase as it is responsible for the maintenance of the original DNA methylation patterns in a cell. It preferentially binds and methylates hemi-methylated DNA (Pradhan *et al.* 1999; Ramsahoye *et al.* 2000). It is present at high levels in dividing cells and it is activated by cell- cycle-dependent transcription factors (Goll & Bestor 2005; Kishikawa *et al.* 2003), indicating its role in DNA replication by maintaining the existing methylation patterns. DNMT1 forms a complex with Ubiquitin Like With PHD And Ring Finger Domains 1 (UHRF1) which contains a RING-associated (SRA) domain that binds to hemi-methylated CpG dinucleotides forming protein complexes, providing maintenance of DNA methylation (Achour *et al.* 2008; Bostick *et al.* 2007). UHRF1 is critical for the recruitment of DNMT1 in hemi-methylated CpG sites as it co-localizes at the replication fork during DNA replication, where newly synthesised hemi-methylated DNA is formed (Leonhardt *et al.* 1992) and methylates the new DNA strand in order to precisely mimic the parental methylation pattern (Hermann *et al.* 2004). In the absence of UHRF1, DNMT1 fails to localize at these regions (Bostick *et al.* 2007; Liu *et al.* 2013b; Sharif *et al.* 2007); as a result there is a replication dependent dilution of 5mC.

Experiments with *Dnmt1* knockout mice show lethality of mice between E8.0 and E10.5 with loss of most DNA methylation (Li *et al.* 1992). Moreover, *in vitro* differentiation experiments with stem cells lacking *Dnmt1* resulted in massive cell death mimicking the phenotype observed in mice; indicating the importance of *Dnmt1* in cellular differentiation and cell division (Jackson-Grusby *et al.* 2001). *DNMT1* overexpression results in aberrant DNA methylation in tumours and correlates with poor prognosis in patients with pancreatic cancer (Peng *et al.* 2006). Whereas *DNMT1* knockdown has been shown to have an inhibitory effect on cell proliferation in oesophageal squamous cell carcinoma, suggesting that increased methylation levels promote cell proliferation (Zhao *et al.* 2011).

In contrast, DNMT3a and DNMT3b are called *de novo* DNMTs as they do not show any preference for hemi-methylated DNA (Okano *et al.* 1999) and can introduce methylation into naked DNA (Fig. 1.1). Although both are highly expressed in early mammalian embryos their expression decreases over the course of cell differentiation displaying different spatial

and temporal distribution throughout embryonic development (Zhang & Xu 2017). DNMT3a is expressed in almost all tissues whereas DNMT3b is expressed in early embryonic development and its expression is limited to the thyroid, testes and bone marrow in the differentiated tissues (Xie *et al.* 1999). DNMT3a is important for maternal imprinting at differentially methylated regions while DNMT3b is necessary for the methylation of CpG islands on inactive X- chromosomes and pericentromeric repeats (Kim *et al.* 2009).

Dnmt3b knockout in mice leads to embryonic lethality similar to what is observed with *Dnmt1* knockout (Okano et al. 1999); therefore it is required for early embryonic development. Although double knockout of *Dnmt3a* and *Dnmt3b* in mouse embryonic stem cells (mESCs) is not lethal, *in vitro* experiments show gradual loss of methylation when cultured (Chen *et al.* 2003). In human embryonic stem cells (hESCs) *DNMT3* depletion does not affect pluripotency and the cells retain their ability to differentiate into three germ layers (Liao *et al.* 2015); indicating that maintenance methylation might be responsible for the differentiation. It has also been reported that DNMT3b cooperates with DNMT1 in order to maintain methylation in human cancer cell lines (Rhee *et al.* 2000, 2002) as well as during development of the intestinal epithelium (Elliott *et al.* 2016). These observations indicate that under certain conditions DNMT3b can also act as a maintenance methyltransferase. Additionally, *Dnmt3b* depletion leads to impaired neuronal differentiation and maturation (Martins-Taylor *et al.* 2012), indicating its crucial role during development.

Finally, another member of DNMT family is DNMT3L which lacks the catalytic domain of the enzyme (Aapola *et al.* 2000; Hata *et al.* 2002). Its expression is restricted to the germ cells and thymus in adulthood, as it is mainly expressed in early development (Aapola *et al.* 2000). It has been shown that DNMT3L forms a protein complex with the other two *de novo* methyltransferases targeting this complex at DNA sequences associated with histones that are un-methylated at lysine 4 of Histone H3 (H3K4) which are found at inactive promoters and methylated DNA (Ooi *et al.* 2007). In mice it is required for the establishment of genomic imprinting, for methylation of retrotransposons and for X chromosome compaction (Hata *et al.* 2002; Kaneda *et al.* 2004). Deletion of the *DNMT3L* gene leads

to a failure of *de novo* methylation in growing oocytes (Bourc'his *et al.* 2001) and in prospermatogonia (Bourc'his & Bestor 2004). Interestingly, a dual role of Dnmt3L in mouse ESC differentiation has been suggested, as it acts as a positive regulator of DNA methylation at gene bodies of housekeeping genes and a negative regulator at promoters of bivalent genes (Neri *et al.* 2013).

1.2.3. The dynamics of DNA methylation during mammalian development

DNA methylation is a dynamic process during the early stages of development, displaying changes in methylation patterns and levels as it erases before the implantation of the embryo, allowing a new profile of methylation to be established (Smith *et al.* 2012, 2014). There are two waves of DNA demethylation and re-methylation during mammalian development. The first one occurs during germ cell development while the other one occurs during the early stages of embryogenesis (Figure 1.2) (Messerschmidt *et al.* 2014; Smallwood & Kelsey 2012).

The first wave involves the erasure of DNA methylation during expansion and migration in primordial germ cells (PGCs) followed by the *de novo* re-establishment of methylation patterns in sex specific germ cells, including the methylation in regions that control genomic imprinting (Messerschmidt *et al.* 2014; Smallwood & Kelsey 2012). In mouse PGCs, demethylation starts as PGCs develop and mainly targets the promoters of genes responsible for pluripotency. Only when cells enter the gonads does demethylation in imprinting control regions (ICRs) occur, followed by long interspersed nuclear elements (LINE) repeats and X-linked loci (Guibert *et al.* 2012; Seisenberger *et al.* 2012). The downregulation of DNMT3a DNMT3b and UHRF1 allows passive demethylation to occur (Kagiwada *et al.* 2013; Seisenberger *et al.* 2012). In the female germline *de novo* methylation occurs in the postnatal growth phase of oocytes arrested in meiotic prophase; whereas in males, methylation initiates before birth in prospermatogonia (Smallwood & Kelsey 2012). The second wave of this dynamic process occurs after fertilisation and involves the erasure of methylation inherited from the gametes in pre-implantation embryos and

the *de novo* re-establishment of methylation after the formation of the inner cell mass (ICM) (Zeng & Chen 2019).

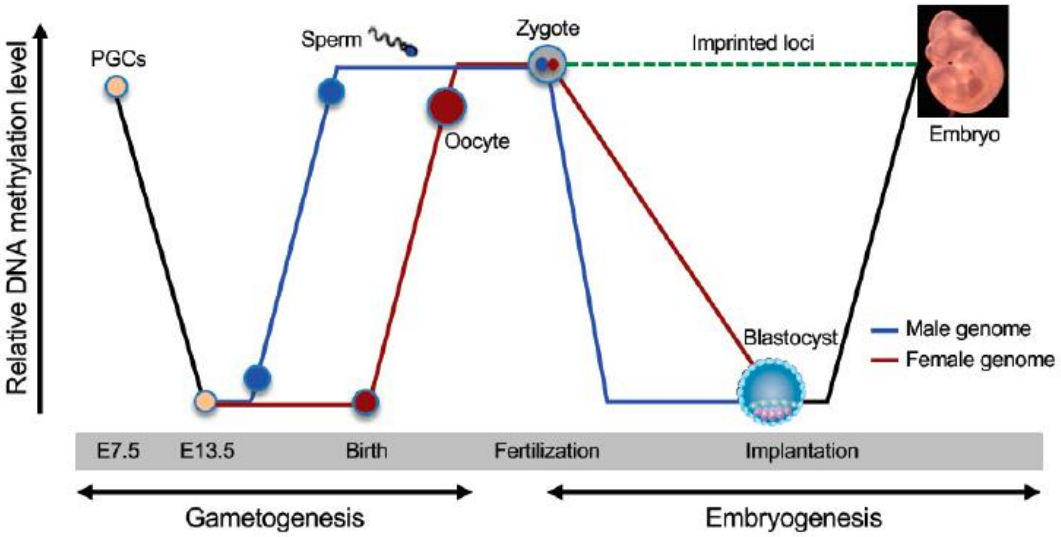


Figure 1.2: The dynamics of DNA methylation during mammalian development.

Global DNA demethylation occurs during PGC expansion and migration followed by *de novo* methylation which results in the establishment of sex-specific germ cell methylation patterns. The second wave of demethylation occurs shortly after fertilisation and upon implantation, when a wave of *de novo* methylation establishes the embryonic methylation pattern. Figure adapted from (Zeng & Chen 2019).

1.3. DNA demethylation

1.3.1. DNA replication dependent passive DNA demethylation

DNA demethylation can occur either passively through successive rounds of DNA replication or actively via enzymatic activity. As mentioned earlier, during DNA replication DNMT1 in complex with UHRF1 is responsible for the maintenance of DNA methylation (Leonhardt *et al.* 1992). The absence of these enzymes can result in replication dependent dilution of 5mC (Fig. 1.3). In pre-implantation embryos, erasure of DNA methylation mainly occurs through DNA replication dependent passive DNA demethylation as

DNMT1 is excluded from the nuclei (Messerschmidt *et al.* 2014; Smallwood & Kelsey 2012). In addition to that, the first wave of DNA demethylation in PGCs is also achieved by DNA replication dependent passive mechanism, as both DNMT1 and UHRF1 are silenced at that stage (Kagiwada *et al.* 2013).

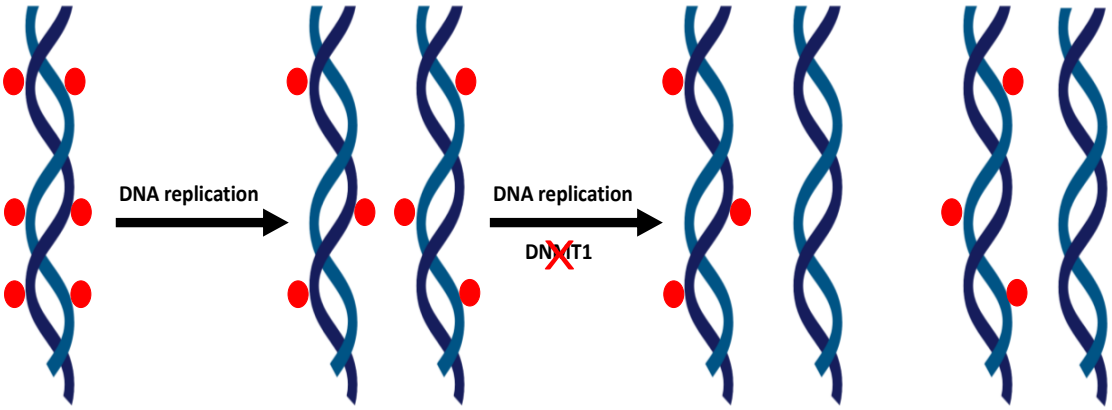


Figure 1.3: Passive DNA demethylation, schematic representation.

When DNMT1 is inhibited or absent the newly synthesised strand of DNA will not be methylated, therefore cell division will eventually lead to passive demethylation.

1.3.2. Active DNA demethylation

Although active DNA demethylation could be achieved by the direct removal of the methyl-group from DNA *in vitro,* this reaction requires a lot of energy and is a thermodynamically unfavourable reaction (Wu & Zhang 2014). Therefore, a well-organized stepwise oxidation of 5mC into oxidative intermediates has been proposed for the removal of DNA methylation, enabling a more favourable thermodynamic reaction. Active DNA demethylation involves the indirect removal of 5mC involving the enzymatic oxidation of 5mC followed by the replacement with unmodified cytosine.

1.3.2.1. TET enzyme dependent active DNA demethylation

5-hydroxymethylcytosine (5hmC) was first reported in 1952 (Wyatt & Cohen 1952), however its role and significance remained unknown until 2009 when it was shown that Ten eleven translocation (TET) enzymes can convert 5mC into 5hmC (Kriaucionis & Heintz 2009; Tahiliani *et al.* 2009). There are three members in the TET family, TET1, TET2 and TET3; all of which have a similar structure comprising of a C-terminal catalytic domain and an N-terminal regulatory domain. The C terminal catalytic domain contains a cysteine rich region (Cys) and a double stranded β-helix (DSBH) domain. The DSBH domain brings together the co- factors at the 5mC site while the Cys region stabilises the complex. They catalyse the oxidation reaction in an iron (II)/α-ketuglutarate (Fe (II)/α-KG) dependent pathway (Wu & Zhang 2017). Specifically, TET1 was identified as the mammalian homologue of base J binding protein 1 (JBP1) and 2 (JBP2) that naturally exists in *Trypanosoma* and has been proposed to oxidise the 5 methyl group of thymine (Tahiliani *et al.* 2009). Tahiliani et al., proved that TET1, similar to its JBP1 and JBP2 homologs, has the ability to mediate 5mC hydroxylation in mammalian DNA generating 5hmC (Tahiliani *et al.* 2009). Similarly, TET2 and TET3 have the ability to oxidise 5mC into 5hmC in mESCs (Ito *et al.* 2010).

Soon after the discovery of 5hmC as a result of 5mC TET1/2/3 mediated oxidation; liquid chromatography based experiments led to the identification of further oxidation of 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (He *et al.* 2011; Ito *et al.* 2011). Indeed, Ito et al. showed that TET2 is able to convert more than 95% of 5mC to 5hmC, 30% to 5fC and 5% to 5caC in an *in vitro* liquid chromatography experiment using HEK293 embryonic kidney cancer cells (Ito *et al.* 2011). Nevertheless, the efficiencies were significantly lowered when 5hmC was used as a substrate (Ito *et al.* 2011). Their findings were confirmed when the catalytic domain of TET2 was fused with green fluorescent protein (GFP) and transfected into HEK293. Fluorescent activated cell sorting (FACS) of the cells revealed that cells expressing TET2 exhibited not only high levels of 5hmC but also detectable levels of 5fC and 5caC (Ito *et al.*

2011). The presence of 5fC and 5caC was also confirmed in the genomic DNA of mESCs. Quantification of these cytosine (C) derivatives using mass spectrometry revealed that $1.3x10^3$ 5hmC, 20 5fC and 4 5caC bases exist in every $10x10^6$ Cs in mESCs (Ito et al. 2011). Specifically, 5hmC represented around 5% of total Cs in mESCs whereas 5fC and 5caC levels are 0.06-0.6% and 0.01% respectively (Ito *et al.* 2011; Pfaffeneder *et al.* 2011).

Tet1 knockdown in mESCs resulted in reduced levels of 5mC oxidative derivatives indicating that TET1 is also responsible for the generation of these marks (Ito *et al.* 2011). Similar observations were also made by He et al., where they showed that TET2 catalyses the oxidation of 5mC to 5hmC and 5caC in DNA using high performance liquid chromatography (HPLC) in HEK 293 cells (He *et al.* 2011). Therefore, TET enzymes are responsible for the oxidation of 5mC into 5hmC, 5fC and 5caC (He *et al.* 2011; Ito *et al.* 2011; Tahiliani *et al.* 2009) (Fig. 1.4). According to *in vitro* biochemical and structural studies, TET enzymes show preferential binding and activity for 5mC compared to 5hmC or 5fC to 5caC and the conversion of 5mC to 5hmC is faster than 5hmC to 5fC and 5fC to 5caC (Hu *et al.* 2015; Ito *et al.* 2011), indicating that TET proteins might be more efficient in utilizing 5mC as a substrate. It has been also proposed that TET enzymes can initiate active DNA demethylation in proliferating cells followed by replication dependent passive dilution of 5hmC, 5fC or 5caC (Kohli & Zhang 2013). Due to substrate preference for hemi-methylated DNA of DNMT1, it is still unclear whether it can recognize 5caC and 5fC (Kohli & Zhang 2013).

5caC and 5fC, but not 5hmC, can be recognized and excised from DNA through thymine DNA glycosylase (TDG) activity, leaving an abasic site which can be replaced by the base excision repair (BER) pathway whose function is to correct mismatched DNA base pairs (Hashimoto *et al.* 2012a; He *et al.* 2011; Maiti & Drohat 2011; Zhang *et al.* 2012). Once TDG excises either 5fC or 5caC an abasic site is created, this site will then be converted into a single strand break (SSB) via apurinic/apyrimidinic endonuclease 1 (APE1)-mediated incision. Addition of DNA polymerase b will follow in order to insert a deoxycytidine monosposphate at the break and DNA ligase 3

(LIG3) ligates the SSB to restore DNA (Weber *et al.* 2016). In addition to TDG, two other glycosylases, NEIL like DNA glysoylase 1 (NEIL1) and 2 (NEIL2) are able to recognise and excise 5fC and 5caC oxidative derivatives creating a single SSB (Schomacher *et al.* 2016). Hence, 5hmC 5fC and 5caC serve as intermediates in active DNA demethylation.

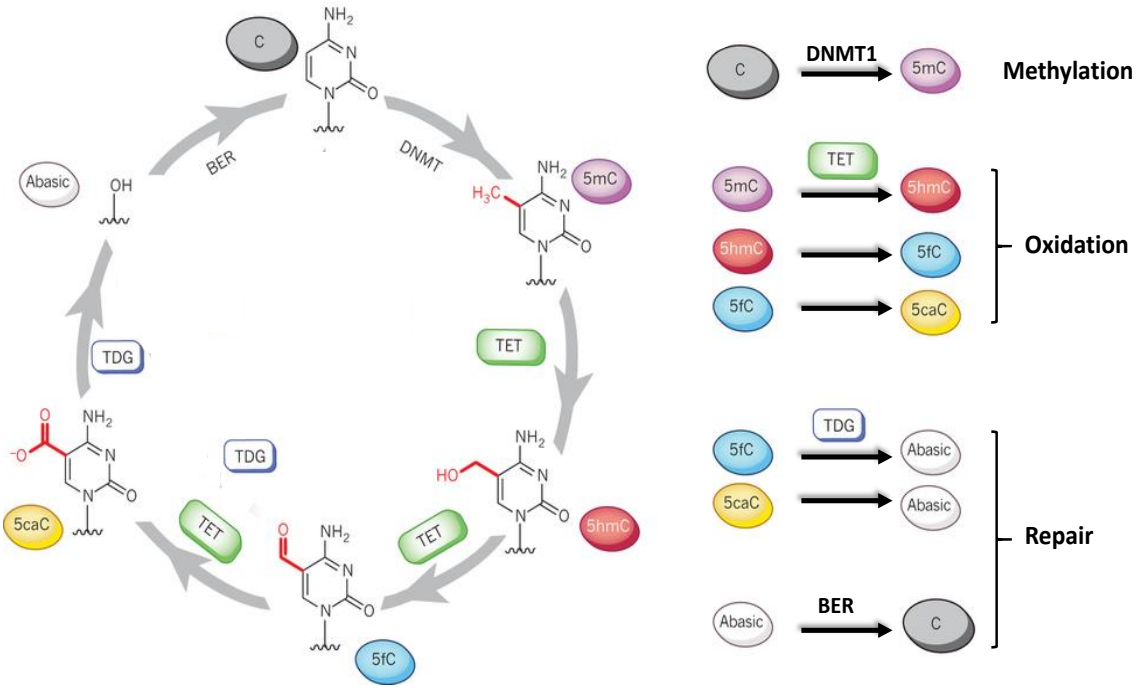


Figure 1.4: Pathway of DNA methylation dynamics through stepwise oxidation.

5mC is introduced to DNA by DNMT enzymes and can be oxidised to 5hmC, 5fC and 5caC by TET proteins. 5fC and 5caC are excised by TDG generating an abasic site as part of the BER pathway which eventually generates an un-modified Cytosine (C). Figure modified from (Kohli & Zhang 2013).

1.4. The genomic distribution of 5mC oxidative derivatives and its function

1.4.1. 5hmC genomic distribution and function

Around 60-80% of the human genome CpG sites are methylated at the fifth carbon position of cytosine (Smith & Meissner 2013). 5mC is evenly distributed across different tissues whereas 5hmC levels vary significantly (Kriaucionis & Heintz 2009; Nestor *et al.* 2012; Ruzov *et al.* 2011). 5hmC dynamics and distribution are highly regulated upon differentiation and are cell type specific. Investigation of global and locus-specific distribution of 5hmC in human tissues and cell lines revealed inter-tissue variation of global 5hmC levels in normal human tissues that is variable and does not correlate with 5mC content (Nestor *et al.* 2012). It has been shown that 5hmC levels are 10-fold more abundant in the central nervous system (CNS) and ESCs compared to peripheral tissues (Branco *et al.* 2012).

5hmC is detectable in all mouse cell types and tissues with the highest levels (0.3-0.7%) found in the CNS (Globisch *et al.* 2010; Kriaucionis & Heintz 2009), namely with 0.6 and 0.2% nucleotides in Purkinje and granule cells respectively (Kriaucionis & Heintz 2009). Gobisch et al., used Liquid Chromatography-Mass Spectrometry (LC-MS) coupled with immunohistochemistry for the precise quantification of 5hmC and its oxidative derivatives in mouse tissues (Globisch *et al.* 2010). 5hmC was detectable in all the tissues tested with the highest levels (0.3%-0.7%) detected in the CNS followed by the spinal cord at 0.47% (Globisch *et al.* 2010). Interestingly, the highest 5hmC levels in the CNS were detected in differentiated neurons of the dentate gyrus. Kidney, nasal epithelium, bladder, lung and skeletal muscle had 0.15%-0.17% 5hmC whereas liver, spleen and endocrine glands possessed the lowest levels of 5hmC (0.03%-0.06%) (Globisch *et al.* 2010).

Similarly, examination of 5hmC distribution during mammalian development, using immunohistochemistry based experiments revealed

that 5hmC is enriched in embryonic tissues compared to adult tissues (Figure 1.5) (Ruzov *et al.* 2011). Specifically, 5hmC firstly appears at the zygote stage during embryonic development and coincides with reduced 5mC levels in the paternal pronucleus (Figure 1.5) (Ruzov *et al.* 2011). Later on in development, 5hmC levels are high in the ICM of the blastocyst and decreased during differentiation (Figure 1.5) (Ruzov *et al.* 2011). Nevertheless, 5hmC is strongly enriched in bone marrow and brain tissues and it appears to be a feature of both neuronal progenitors cells (NPCs) and post-mitotic neurons (Ruzov *et al.* 2011). 5hmC not only represents an intermediate for active DNA demethylation but current evidence also supports an important role in the maintenance of pluripotency in ESCs, in neuronal development, as well as in cancer and disease development (Thomson & Meehan 2017).

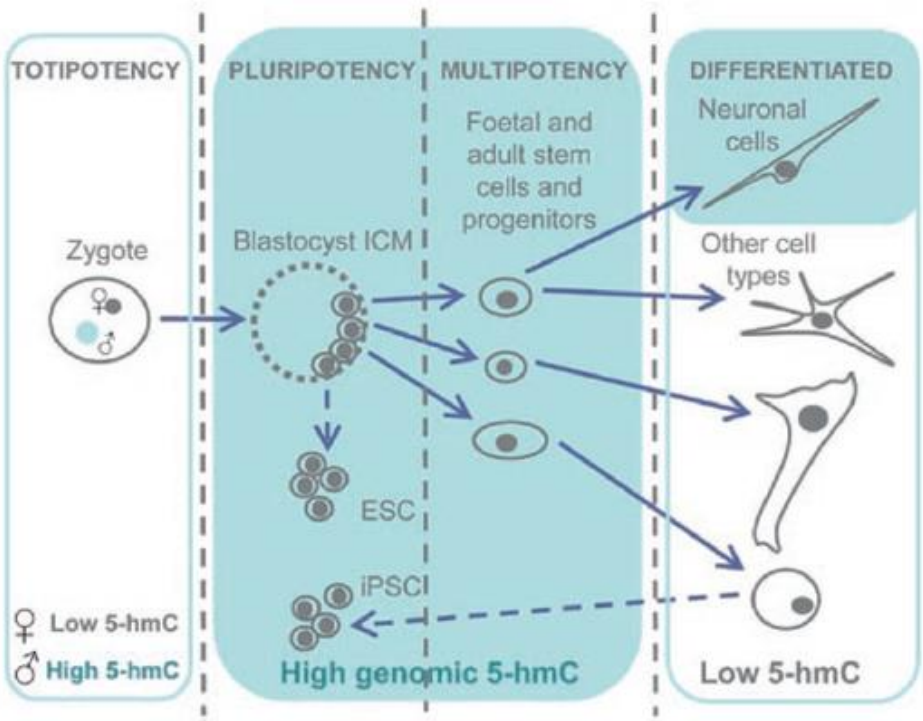


Figure 1.5: The genomic distribution of 5hmC during mammalian development; schematic representation.

High 5hmC levels first appear in the paternal pronucleus. The ICM in the blastocyst also exhibits high genomic 5hmC levels, whereas upon differentiation 5hmC levels are reduced. Figure adapted from (Ruzov *et al.* 2011).

It has been shown that 5hmC in ESCs is mainly located at distal regulatory elements including enhancers; near transcription factor (TF) binding sites (Yu *et al.* 2012); and in gene bodies of active genes and promoters (Nestor *et al.* 2012). Interestingly, another study showed that 5hmC enrichment at promoters and TSS is most likely to be associated with TET1 mediated oxidation while its presence in gene bodies and boundaries of exons in active genes is correlated with TET2 enzymatic activity (Huang *et al.* 2014).

It has been suggested that 5hmC acts as a transcriptional activator or suppressor depending on its genomic position. In neurons, it is mostly enriched at gene bodies of active neuronal function-related genes suggesting that it acts as a transcription activator (Mellen *et al.* 2012). A study based on 5hmC- DNA Immunoprecipitation (5hmC-DIP) showed that genes with active transcription have low levels of 5hmC at their TSS while genes with low expression 5hmC levels are elevated at their promoters in both ESCs and NPCs (Tan *et al.* 2013). Global 5hmC levels in NPCs are much lower than ESCs levels indicating that differentiation requires global reduction of 5hmC levels (Tan *et al.* 2013) .

There is evidence that 5hmC might regulate gene expression through its association with various regulatory elements and processes (Szulwach *et al.* 2011a). Specifically, the distribution of 5hmC is affected by histone modifications and chromatin alterations during cell differentiation (Pastor *et al.* 2011). Interestingly, 5hmC is present at TSS of genes whose promoters have dual histone modifications; such as tri-methylation at lysine 27 of histone H3 (H3K27me3) for transcription repression and tri-methylation at lysine 4 of histone H3 (H3K4me3) for transcription activation (Pastor *et al.* 2011). Moreover, immunostaining based experiments revealed that 5hmC accumulates in gene-rich regions marked by di- and tri-methylation at lysine of histone H3 (H3Kme2/3) indicating its association with chromatin remodelling factors (Ficz *et al.* 2011; Szulwach *et al.* 2011b). 5hmC is associated with gene bodies of active neuronal function-related genes suggesting its role is to promote brain development (Hahn *et al.* 2013).

1.4.2. 5caC and 5fC genomic distribution and function

Unlike 5hmC levels, 5fC and 5caC levels are significantly lower in all mammalian tissues and cells (Ito *et al.* 2011). Mass spectrometry experiments revealed that 5fC and 5caC are detectable in mESCs however their levels are significantly lower than 5hmC (He *et al.* 2011; Ito *et al.* 2011; Pfaffeneder *et al.* 2011). Moreover, according to recent reports based on 5fC and 5caC distribution in mESCs; 5fC regions reside in 5hmC enriched regions however this accounts for only 30% of 5hmC regions (Shen *et al.* 2013; Song *et al.* 2013a).

Specifically, in mESCs 5fC is mainly enriched at distal regulatory elements with a preference for poised enhancers and promoters of genes with low expression (Song *et al.* 2013a). Upon *Tdg* depletion in mESCs 5fC and 5caC levels increase while 5mC and 5hmC levels remain unaltered (He *et al.* 2011; Shen *et al.* 2013; Song *et al.* 2013a); leading to an accumulation of these marks at enhancers of TF-binding sites, low methylated regions, and promoters of genes with low expression (Shen *et al.* 2013; Song *et al.* 2013a). Therefore, 5fC and 5caC might contribute in the maintenance of these key gene elements in a poised or active state for differentiation (Shen *et al.* 2013; Song *et al.* 2013a).

Interestingly, it has been shown that 5fC and 5caC can reduce the rate and substrate specificity of RNA polymerase II (PolII) transcription (Wang *et al.* 2015a). Accumulation of 5fC and 5caC upon TDG depletion at distal regulatory elements coordinates with the binding of transcriptional coactivator p300 (Song *et al.* 2013a) which can serve as a scaffold for TF and transcription machinery for the activation of gene transcription (Chen & Li 2011).

In order to examine 5fC and 5caC abundance at single-base resolution a method based on bisulfite sequencing (methylation-assisted bisulfite sequencing) has been developed (this method will be discussed later in this thesis) (Neri *et al.* 2015). It shows that 5fC and 5caC are located in hypo-methylated promoters of highly expressed genes in mESCs and that

Tdg silencing leads to increased levels, indicating a role of active DNA demethylation at these promoters (Neri *et al.* 2015). In agreement with publications based on affinity purification methods (Shen *et al.* 2013; Song *et al.* 2013a), this study shows 5caC and 5fC enrichment on enhancers, exons and repetitive elements (Neri *et al.* 2015). 5caC and 5fC are also enriched at TSS and are correlated with progressive loss of H3K27me3 repressive mark and gain of H3K4me3. Suggesting that active DNA demethylation occurs at the promoters of actively transcribed genes in ESCs and likely regulating their expression through development (Neri *et al.* 2015).

Increasing evidence, suggest a role for active DNA demethylation in the regulation of mammalian development. 5fC and 5caC are detectable at low levels in both the human and mouse brain (Liu *et al.* 2013a). Studies, in mice showed that 5caC and 5fC are transiently accumulated during differentiation of neural stem cells (NSCs) *in vitro* and *in vivo* and that 5caC is enriched at promoters of neural specific genes during the differentiation of NSCs, indicating the involvement of the active DNA demethylation pathway in lineage specification of NSCs (Fig. 1.6) (Wheldon *et al.* 2014). In line with this finding, a more recent publication by Lewis et al. showed, using immunocytochemistry, that during the differentiation of hiPSCs into endodermal hepatic lineages 5caC levels increase in the specification of foregut and drop in differentiating cells (Lewis *et al.* 2017); similar to neural differentiation (Wheldon *et al.* 2014). Moreover, 5caC-DIP at different stages of differentiation showed that 5caC accumulates at the promoter regions of several genes expressed during hepatic differentiation (Lewis *et al.* 2017). These results indicate that active DNA demethylation might be a feature of lineage specification during mammalian development.

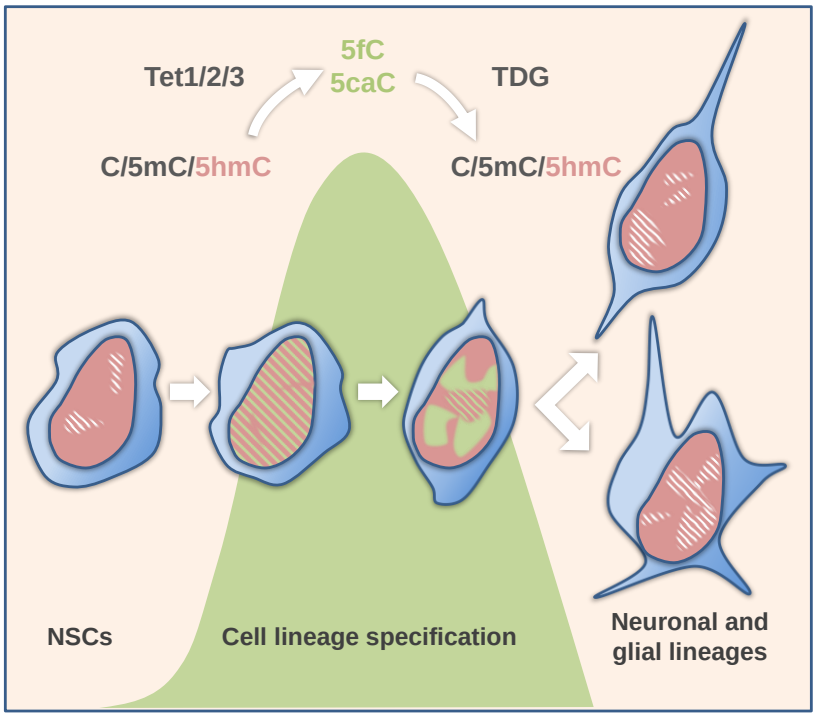


Figure 1.6: Transient accumulation of 5fC/5caC during neuronal differentiation.

TET enzymes are responsible for the oxidation of 5mC/5hmC into 5caC and 5fC during cell lineage specification of NSCs followed by their excision through TDG in differentiated neuronal and glial cells. Figure adapted from (Wheldon *et al.* 2014).

1.5. The role of DNA methylation/demethylation in cancers

Epigenetic alterations and abnormalities in DNA methylation patterns have been observed in almost all types of tumours and are anticipated to be of key importance in the initiation and progression of cancer (Jones & Baylin 2002). Cancer cells are characterised by a global loss of DNA methylation; this hypo-methylation is found in highly repeated DNA sequences including retrotransposons and heterochromatic DNA repeats (Ehrlich 2002). Global loss of methylation at repetitive sequences causes chromosomal instability in the cell and disrupts gene activity (Gaudet 2003). It has also been shown that hypo-methylation at certain promoters can lead to aberrant

expression of oncogenes. At the same time, there is hyper-methylation at the promoter regions of many genes at CpG islands affecting their expression (Portela & Esteller 2010).

5hmC levels are significantly affected in cancer; reduction of 5hmC levels are correlated with tumorigenesis and tumour progression in hematopoietic and solid tumours (Wu, Y.C., and Ling 2014). In breast, colon and lung cancer 5hmC levels are significantly reduced compared with their levels in normal tissues (Jin *et al.* 2011). A similar phenomenon has also been found in myeloid malignancies as studies have shown the reduced 5hmC levels correlate with TET2 mutations in these types of tumours (Ko *et al.* 2010; O. Abdel-Wahab, A. Mullally 2009). Strong reductions in 5hmC levels have been observed in melanoma, where wide-spread loss of 5hmC is present in parallel with 5hmC increase in gene bodies and promoters of genes (Lian *et al.* 2012). Moreover, in pancreatic cancer cells a global loss and re-distribution of 5hmC at exons, TSS and oncogenic promoters was shown (Bhattacharyya *et al.* 2013). Loss of 5hmC is also an epigenetic mark for papillary thyroid carcinoma associated with malignant behaviour (Tong *et al.* 2019).

A comparison between prostate cancer and normal prostate tissue samples using immunohistochemistry showed that both 5mC and 5hmC levels were significantly reduced in cancer tissues (Storebjerg *et al.* 2018). 5fC was also detectable in both cancer and normal tissues and displayed significantly higher levels in erythloblast transformation-specific related gene (ERG)+ prostate cancer tissues compared with normal prostate tissue samples (Storebjerg *et al.* 2018), indicating that 5fC might contribute to tumorigenesis in ERG+ prostate cancer. 5caC elevated levels were observed in both ERG + and ERG- cancer samples (Storebjerg *et al.* 2018). In a more recent study, ultra-high performance liquid chromatography-with tandem mass spectrometry (UHPLC-MS/MS) analysis for the detection of DNA cytosine modifications in hepatocellular carcinoma showed that, 5hmC and 5fC levels were decreased at early stages of the tumour, while 5fC content further decreased in the later stages of hepatocellular carcinoma (Liu *et al.* 2019). However, 5caC was not detectable in the experimental conditions of the study (Liu *et al.* 2019).

The decrease in the levels of 5hmC and 5fC was in agreement with decreased TET enzymatic activity (Liu *et al.* 2019). These data further support the hypothesis that 5fC and 5caC might display unique epigenetic signatures in carcinogenesis; highlighting that active DNA demethylation may contribute to tumorigenesis and cellular transformation through the regulation of the epigenetic landscape.

1.5.1. DNA demethylation in brain cancers

5hmC is enriched in the brain compared to other differentiated tissues such as bladder, kidney and liver (Globisch *et al.* 2010). However, in contrast to normal brain tissues, brain tumours exhibit significantly lower amounts of 5hmC and most importantly these levels are correlated with tumour grade and patient survival (Kraus *et al.* 2012).

5hmC is depleted in glioblastoma multiforme (GBM) compared to prefrontal cortex tissue with an average 3.5 fold reduction (Johnson *et al.* 2016). Moreover, hyper-methylation is observed across GBM promoter regions while 5hmC levels are constantly lower at promoter regions when compared to healthy tissues (Johnson *et al.* 2016). Nevertheless, high 5hmC levels were found in intronic regions as well as in enhancer regions of actively transcribed genes in GBM (Johnson *et al.* 2016). This data indicates that 5hmC genomic distribution may be related to transcription in GBM and the subsequent pathogenesis of the tumour, leading to worse prognosis (Johnson *et al.* 2016). Moreover, a recent study in medulloblastoma (MB) primary tissues showed that global 5hmC levels were decreased in MB compared to the cerebellum control and these levels were associated with the presence of metastasis (Bezerra Salomão *et al.* 2018).

Correspondingly, in our previous studies in breast cancer tissues we found that the levels of 5caC are elevated in 28% of human breast cancers including the samples which produced low 5hmC intensity (Eleftheriou *et al.* 2015). This suggests that, in some cases, depleted 5hmC does not necessarily correlate with low TET dependent oxidation of 5mC. However, it is likely that TET proteins determine the preferential oxidation of 5mC to

5caC in tumours. We also analysed 74 samples of human glioma tissues, 5caC was detectable in a considerable number of gliomas, however, the presence of this modification was not associated with glioma grade (Fig. 1.7 A) (Eleftheriou *et al.* 2015). In addition to that, in order to further extend these results our group determined the level of TET proteins and oxidised-methylcytosines (oxi-mCs) in cell lines of paediatric brain tumours. We found immunochemically detectable levels of 5caC and elevated TET1 expression in MB and ependymoma (EPN) cell lines compared to HeLa cells, which do not seem to have any detectable 5caC levels (Fig. 1.7 B). Interestingly 5caC levels did not correlate with 5hmC levels and TDG expression in paediatric brain tumours (Ramsawhook *et al.* 2017).

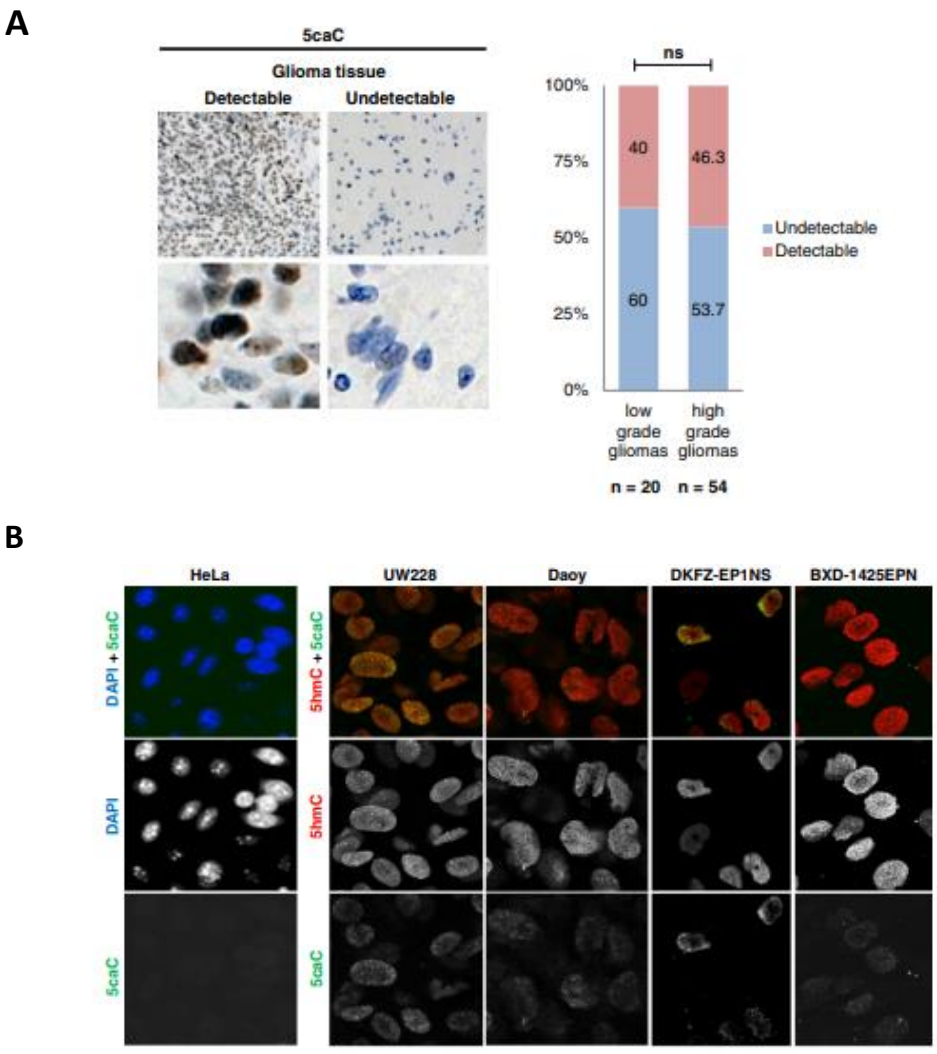


Figure 1.7: 5caC is detectable in primary glioma tissues and in EPN and MB cell lines.

A. Glioma tissue samples with different levels of 5caC signal (designated as detectable or undetectable). Figure adapted from (Eleftheriou *et al.* 2015). B. Paediatric MB (UW228, Daoy) and EPN (DKFZ-EP1NS, BXD-1425EPN) cell lines exhibit immunocytochemically detectable levels of 5hmC and 5caC. Figure adapted from (Ramsawhook *et al.* 2017).

1.6. The role of TET enzymes in development

1.6.1. The role of TET enzymes during embryonic development

All three TET enzymes exhibit cell type specific expression patterns at different developmental stages (Fig. 1.8) (Ito *et al.* 2010; Szwagierczak *et al.* 2010). TET1 is highly expressed in mESCs, the ICM and in developing PGCs in mice while TET2 and TET3 are highly expressed in adult tissues in mice (Dawlaty *et al.* 2011; Ito *et al.* 2010; Yamaguchi *et al.* 2012).

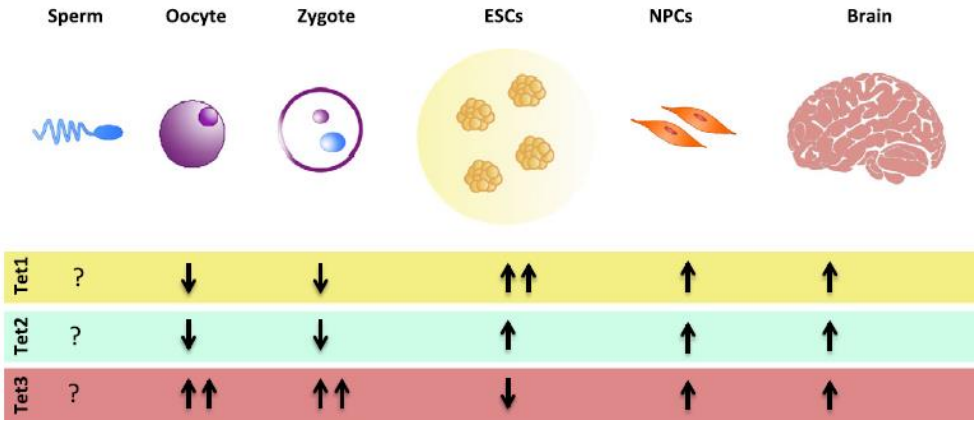


Figure 1.8: Schematic representation of TET enzymes' levels at different developmental stages.

Arrow-up indicates increased expression levels; arrow-down indicates decreased expression levels. Figure modified from (Santiago *et al.* 2014).

TET3 is the only TET enzyme detected in mouse oocytes and one-cell zygotes, indicating that it has a crucial role in the first wave of DNA demethylation during development and is responsible for the erasure of DNA methylation (Iqbal *et al.* 2011; Wossidlo *et al.* 2011). TET3 deletion causes neonatal lethality coupled with the absence of 5hmC oxidative derivatives in the paternal pronucleus (Gu et al. 2011). Global DNA demethylation of both the maternal and paternal genomes occurs shortly after fertilisation (Lee *et al.* 2014; Saitou *et al.* 2012). According to current knowledge the demethylation of the maternal genome mainly occurs via

passive dilution whereas the demethylation of the paternal genome happens through a combination of TET3-mediated active DNA methylation and 5mC replication dependent passive dilution. 5mC decreases drastically soon after fertilisation (Mayer *et al.* 2000). This is accompanied by the generation of oxidative derivatives of 5mC which is mediated by TET3 (Inoue *et al.* 2011; Iqbal *et al.* 2011; Wossidlo *et al.* 2011). Nevertheless, depletion of zygotic TET3 does not have any significant effects on 5mC levels, indicating that passive demethylation is the driving force for paternal genome demethylation whereas TET3 mediated oxidation only contributes to some extent (Guo *et al.* 2014; Shen *et al.* 2014b).

Increasing evidence suggests a role of TET1 and TET2 enzymes in the maintenance of pluripotency in ESCs. According to Dawlaty et al., *Tet1* and *Tet2* double knockout (KO) mESCs remain pluripotent; however, 5hmC levels were significantly reduced leading to developmental defects and chimeric embryos (Dawlaty *et al.* 2013). This study also showed that, although a fraction of double *Tet1* and *Tet2* KO mouse embryos remain viable, embryos with mid-gestation abnormalities leading to perinatal lethality were also obtained (Dawlaty *et al.* 2013). Viable mice, displayed reduced 5hmC levels and abnormal methylation patterns at imprinted loci (Dawlaty *et al.* 2013). Indicating a critical role for TET enzymes during development and genomic imprinting.

Triple *Tet* KO mESCs and murine embryonic fibroblasts (MEFs) displayed differentiation and de-differentiation defects, however they were still viable (Dawlaty *et al.* 2014; Hu *et al.* 2014). In mESCs, triple *Tet* KO led to poorly differentiated embryoid bodies and teratoma formation (Dawlaty *et al.* 2014). Global gene-expression and methylation analysis revealed deregulation of genes important for development and differentiation as well as promoter hyper-methylation (Dawlaty et al. 2014). Moreover, triple *Tet* KO lost their capacity for reprogramming due to impaired mesenchymal-to epithelial transition (MET) (Hu et al. 2014). An additional study, revealed the importance of active DNA demethylation in reprogramming, showing that 5hmC enrichment through TET1 oxidation is involved in demethylation and subsequent reactivation of genes important for pluripotency (Gao *et al.* 2013). Therefore, TET1 can replace octamer

binding transcription factor (OCT4) in induced pluripotent stem cells (iPSCs) induction (Gao *et al.* 2013). Similarly, another study in mESCs revealed that TET1 and TET2 regulate OCT4 expression in order to sustain 5hmC levels and subsequently pluripotency (Koh *et al.* 2012).

It has been also shown that *Tet1* knockdown leads to the downregulation of Nanog in mESCs correlating with methylation of the Nanog promoter (Ito *et al.* 2010). Therefore indicating an important role for TET1 in mESCs maintenance (Ito *et al.* 2010). In the same study it was also shown that knockdown of *Tet1* in pre-implantation mouse embryos leads to impaired differentiation towards trophoectoderm (Ito *et al.* 2010). In addition to TET1 synergy with NANOG, TET2 also associates with NANOG during the establishment of reprogramming (Costa *et al.* 2013). Moreover, another study showed a correlation between TET1 depletion and impaired Leukaemia inhibitor factor (LIF) signalling known to promote self-renewal and pluripotency in mESCs (Freudenberg *et al.* 2012). Thus, further supporting the evidence of TET1 importance in the maintenance of pluripotency.

Transcriptional control of TET genes is mediated by cell-type specific TFs. The promoter region of TET1 gene contains binding sites for TFs related to pluripotency (Ficz *et al.* 2011), indicating that TFs maintain TET1 expression in the early stages of development. It is also important to note that *Tet1* is highly expressed in mESCs and it is downregulated upon differentiation (Ito *et al.* 2010; Ko *et al.* 2011). TET1 depletion results in transcriptional regulation of TET1 direct targets (Williams *et al.* 2011; Wu *et al.* 2011a; Xu *et al.* 2011) suggesting TET proteins have a role in transcriptional regulation. It has been also shown to participate in the repression of Polycomb-targeted developmental regulators and in the regulation of transcription of pluripotency factors (Wu *et al.* 2011b). It is possible that the role of TET enzymes in transcriptional regulation is dependent on its oxidative activity on 5mC to maintain the un-methylated state of active promoters or distal enhancers (Ficz *et al.* 2011; Shen *et al.* 2013; Wu & Zhang 2011).

Emerging evidence suggests the role of TET enzymes during brain development and neurogenesis. All TET enzymes are present in developing mouse brains (Hahn *et al.* 2013; Szwagierczak *et al.* 2010). It has been shown that during the differentiation of NPCs into neurons there is an elevation of 5hmC levels in parallel with increased levels of TET2 and TET3 (Hahn *et al.* 2013). Therefore, *Tet2/3* knockdown results in impaired neuronal differentiation, whereas their overexpression promotes embryonic cortical neurogenesis (Hahn *et al.* 2013). TET3 depletion in Xenopus embryos results in impaired eye and neural development (Xu *et al.* 2012). In mESC, *Tet3* KO leads to impaired neuronal differentiation; as they fail to terminally differentiate into neurons (Li *et al.* 2014). Moreover, another study showed that *Tet1* promotes adult hippocampal neurogenesis via the regulation of NPC proliferation (Zhang *et al.* 2013). TET2 has also been implicated in neurogenesis as it has been shown that AF9, a histone methyltransferase and mixed lineage leukaemia (MLL) fusion partner, recruit TET2 to 5mC loci in order to direct 5mC to 5hmC conversion, with subsequent activation of neural target genes in neural differentiation (Qiao *et al.* 2015).

1.6.2. Regulators of TET activity

TET enzymatic activity is highly dependent on metabolites and co-factors. TET requires oxygen, Fe (II) and α-KG in order to catalyse the oxidation of 5mC into oxi-mCs. A-KG is produced by isocitrate in the tricarboxylic acid (TCA) cycle through isocitrate dehydrogenase 1 (IDH1) and 2 IDH2 activity in the mitochondria (Kaelin & McKnight 2013). It has been shown that, tumours with inactivating mutations in TCA cycle enzymes or in IDH1/2 have an accumulation of fumarate, succinate and oncometabolite 2-hydroxyglutarate (2-HG), which are competitors of α-KG in TET enzymes inhibition (Kaelin & McKnight 2013).

Another important regulator of TET enzymes is vitamin C which has been shown to stimulate their catalytic activity (Blaschke *et al.* 2013; Minor *et al.* 2013; Yin *et al.* 2013). It has been proposed that Vitamin C facilitates conformational changes and/or recycling of the co-factor Fe (II) resulting

in positive regulation of TETs activity (Yin et al., 2013). Vitamin C also significantly increases the levels of 5mC oxidative derivatives in mESCs (Blaschke et al., 2013) and MEFs (Minor *et al.* 2013).

Moreover, TET2 and TET3 directly interact with O-linked B-N-acetylglucosamine (O-GlcNac) transferase (OGT), recruiting the complex to CpG-rich promoters (Chen *et al.* 2013; Deplus *et al.* 2013; Vella *et al.* 2013). Therefore, TET proteins might promote transcriptional activation by enhancing histone modifications such as O-GlcNac.

In addition, TETs mRNA might be post-transcriptionally regulated by microRNAs. TET2 for instance is regulated by miR-22 in myelodysplastic syndrome (MDS) and leukaemia where its expression is correlated with reduced global levels of 5hmC (Song *et al.* 2013b). Indeed, TET2 downregulation correlates with miR-22 overexpression in MDS patients (Song *et al.* 2013b). In breast cancer development, miR-22 directly targets TET proteins resulting in enhancement of epithelial-mesenchymal transition (EMT) promoting metastasis in mouse xenografts (Song *et al.* 2013c).

1.7. The role of TET enzymes in carcinogenesis

Impaired DNA methylation is one of the main hallmarks of cancer (Hanahan & Weinberg 2011; Jones & Baylin 2002). TET proteins function as epigenetic regulators modulating transcriptional regulation and cellular actions (Pastor *et al.* 2013). The first indication of a role for TET proteins in carcinogenesis, came from the discovery of TET1 as a fusion partner of MLL in patients with acute myeloid leukaemia (AML), whereby the fusion protein acts as an oncogene (Ono *et al.* 2002). Since then, several studies investigating the dual activities of TET proteins in cancer pathogenesis in different types of cancer have shown that TET might have tumour promoting or tumour suppression functions; summarised at Table 1.1.

Table 1.1: The dual role of TET genes in malignancies

Experimental evidence	**Main finding**	**Reference**
TET2 mutations and deletions promote the development of MDS AML	Tumour suppressor	(Delhommeau et al. 2009; Ko et al. 2011; Li et al. 2011; Moran-Crusio et al. 2011).
TET1 deficiency promotes the development of B cell lymphoma	Tumour suppressor	(Cimmino et al. 2015)
Low *TET1* levels correlate with increase metastasis in prostate cancer	Tumour suppressor	(Spans et al. 2016)
TET2 reduced levels are found in human melanoma	Tumour suppressor	(Lian et al. 2012).
Methylated *TET2* promoter in low-grade gliomas	Tumour suppressor	(Kim et al. 2011)
TET2 reduced levels in primary glioma tissues	Tumour suppressor	(Chen et al. 2017).
TET2 reduced levels in GBM	Tumour suppressor	(García et al. 2018)
High levels of *TET1* in proneural GBM	Tumour oncogene	(Takai et al. 2014).
TET2 high levels in cancer stem cells of colorectal carcinoma, melanoma and GBM	Tumour oncogene	(Puig et al. 2018)

1.7.1. Haematological malignancies

Impairment of TET2 function predisposes to the development of haematological malignancies probably due to its critical role in haematopoiesis (Quivoron *et al.* 2011). *In vitro* studies using short-hairpin-RNA (shRNA) mediated knockdown of TET2 suggest a role of TET2 in the regulation of myeloid differentiation (Ko *et al.* 2010). Loss of 5hmC in myeloid malignancies is often related to mutations in *TET* genes (Kraus *et al.* 2015). *TET2* mutations and deletions are a frequent phenomenon in multiple forms of myeloid malignancies, including MDS and AML (Ko *et al.* 2011; Li *et al.* 2011). Previous studies in myeloid malignancies have shown somatic deletions and loss-of-function mutations in TET2 in around 10%-20% of patients with MDS (Delhommeau *et al.* 2009).

Tet2 KO in mice results in drastic reductions of 5hmC levels leading to increased hematopoietic repopulating capacity and altered haematopoietic differentiation (Ko *et al.* 2011; Li *et al.* 2011); suggesting that TET2 acts as a tumour suppressor and is responsible for the maintenance of haematopoietic cell homeostasis (Li et al. 2011). Another study showed that *Tet2* loss in mice leads to increased stem cell self-renewal and myeloid transformation including myeloproliferation resulting in splenomegaly and monocytic proliferation (Moran-Crusio et al. 2011). Moreover, patients with MDS and TET2 defects exhibit enhanced repopulating abilities compared to patients without TET2 defects (Delhommeau et al. 2009). Apart from TET2, TET1 has also been found to be deregulated in haematological malignancies. *Tet1* deficiency in mice promotes the development of B cell lymphoma regulating the distribution of 5hmC and the expression of genes encoding for B cell lineage (Cimmino et al. 2015). Therefore, indicating the role of TET1 as a tumour suppressor in haematopoietic malignancies (Cimmino et al. 2015).

In a more recent study, mouse models to investigate the effects of *Tet2* expression in normal haematopoiesis were produced (Cimmino *et al.* 2017). Hence, a transgenic mouse to model the restoration of endogenous *Tet2* expression was generated. It has been shown that *Tet2* knockout

mice display aberrant haematopoietic stem cells (HSC) self-renewal leading to disease (Cimmino *et al.* 2017). Restoration of *Tet2* reverses aberrant hematopoietic self-renewal *in vitro* and *in vivo* promoting cellular differentiation corresponding to 5hmC elevated levels (Cimmino *et al.* 2017). Moreover, in the same study it has been shown that administration of Vitamin C to AML cell lines leads to active DNA demethylation resulting at enhanced 5hmC levels without any apparent alteration at TET levels (Cimmino *et al.* 2017). Similar to these findings, in another study Agathocleous et al., using metabolomics, showed that vitamin C deprivation in mice leads to enhanced HSC self-renewal and reduced 5hmC levels (Agathocleous *et al.* 2017). This phenotype corresponds to TET2 catalytic activity; and administration of vitamin C at mice with AML is capable of restoring TET2 activity leading to enhanced survival (Agathocleous *et al.* 2017). Therefore, these studies suggest that Vitamin C can be used as a combination therapy to target AML.

1.7.2. The role of TET enzymes in brain malignancies

TET enzymes also have dual roles in other solid cancers. In prostate cancer, low TET1 mRNA levels are correlated with increased risk of metastasis suggesting that TET1 catalytic activity has a suppressive role in prostate cancer (Spans *et al.* 2016). According to an immunocytochemistry based study in precancerous breast lesions, 5hmC levels were positively associated with TET2 expression, suggesting that they can be used as biomarkers for breast cancer development (Zhang *et al.* 2019). Reduced levels of IDH2 and TET2 enzymes in human melanoma are responsible for 5hmC loss as the re-introduction of active TET2 in melanoma cell lines results in the global increase in 5hmC levels coupled with suppression of tumour invasion and growth (Lian *et al.* 2012).

In a proportion of low-grade gliomas, *TET2* promoter is methylated, therefore leading to inactivation of the *TET2* gene; suggesting TET2 has a tumour suppressor role in low-grade gliomas (Kim *et al.* 2011). Another study showed the exclusion of TET1 from the nuclei of glioma cells might be responsible for the loss of 5hmC observed in gliomas, contributing to

carcinogenesis (Müller *et al.* 2012). A study in primary glioma tissues compared with non-tumour brain adjacent tissue showed that, TET2 expression and protein levels were decreased in glioma tissues (Chen *et al.* 2017). TET2 downregulation was correlated with increased tumour size and tumour grade (Chen *et al.* 2017). Ectopic overexpression of *TET2* in U87-MG and U251 GBM cell lines resulted in inhibition of growth and invasion *in vitro* and *in vivo* (Chen *et al.* 2017). TET2 promoter physically binds the Zinc finger E-box-binding homeobox 1 *(ZEB1)* gene which is located 2kb upstream of the first exon of *TET2* gene and its knockdown led to increased levels of *TET2,* suggesting that *TET2* expression might be regulated by *ZEB1* (Chen *et al.* 2017). Moreover, ectopic overexpression of TET2 in GBM LN228 cell line *in vitro* resulted in a significant decrease of cell population doubling time as well as in decreased cell viability (García *et al.* 2018). Following transplantation of TET2-overexpressing LN229 GBM cells to nude mice, 2 out of 5 mice injected displayed reduced tumorigenic potential *in vivo* (García *et al.* 2018). This evidence suggests that TET2 might act as a tumour suppressor in GBM.

Interestingly, a study in proneural glioblastoma primary cell lines derived from patients showed that they contain high levels of TET1 resulting in TET-mediated oxidation of 5mC into 5hmC (Takai *et al.* 2014). Small interfering RNA (siRNA)-mediated knockdown of *TET1* led to a significant decrease in cell growth and suppressed sphere formation while re-expression of TET1 could rescue the phenotype (Takai *et al.* 2014). Infection of lentivirus-expressing shRNAs against *Tet1* in nude mice resulted in increased survival compared with mice injected with control lentivirus (Takai *et al.* 2014). Analysis of 5hmC levels in TET1 knockdown cells revealed decreased expression levels of key genes involved in GBM formation (Takai *et al.* 2014). These observations suggest TET1 has a crucial role in proneural GBM tumorigenicity probably through the regulation of cancer stem cells.

Similarly, in a recently published paper it has been shown that TET2 enzymatic activity controls survival and recurrence of slow-cycling cancer cells (Puig *et al.* 2018). These cells are often called cancer stem cells and resist current treatment methods, which mainly includes anti-proliferative

drugs. Cancer stem cells are also responsible for tumour recurrence. TET2 was shown to be highly expressed in the subpopulation of cancer stem cells of different tumour types including colon cancer, myeloma and GBM (Puig *et al.* 2018). Furthermore, patients with colorectal carcinoma that showed enrichment of TET2 levels and treated with chemotherapy relapsed significantly earlier (Puig *et al.* 2018). shRNA mediated knockdown of *TET2* as well as clustered regularly interspaced short palindromic repeats/Cas 9 (CRISPR/Cas9) mediated *TET2* knockout in cancer stem cells showed an increased apoptosis in human colorectal cancer cell lines (Puig *et al.* 2018). Moreover, knockout of *TET2* delayed subcutaneous tumour re-growth *in vivo* and progression-free survival was significantly longer (Puig *et al.* 2018), indicating TET2 regulates the population of cancer stem cells in human colorectal carcinoma.

1.7.3. The non-enzymatic roles of TET enzymes in carcinogenesis

Several studies have shown TET proteins can regulate transcription in malignancies in a non-enzymatic way. Mutations in *IDH1/2* genes are often found at haematopoietic and solid tumours, these mutations result in reduced enzymatic activity of TET, as they are responsible for α-KG production (Dang *et al.* 2010). Instead, mutations at these genes leads to the generation of (R) -2- hydroxyglutarate ((R)-2-HG) which acts as a repressor of TET enzymatic activity (Losman *et al.* 2013; Ye *et al.* 2013). Therefore, tumours with *IDH1/2* mutations display lower 5hmC levels. Another study showed that Wilm's tumour 1 (WT1) interacts with TET2 as mutations in *WT1* gene leads to reduced TET2 enzymatic activity resulting in decreased 5hmC levels in AML (Rampal *et al.* 2014). This study showed that WT1 interacts with TET proteins stimulating their activity at specific genomic regions (Rampal *et al.* 2014). It has also been shown that TET2 recruits histone deacetylase 2 *(HDAC2*) to selectively inhibit Interleukin-6 *(IL-6)* in innate myeloid cells acting independently of its enzymatic activity (Zhang *et al.* 2015). This study suggests that suppression of TET2's non enzymatic activity leads to the elevation of IL-6 levels in tumours implicating another level of regulation of carcinogenesis through TET enzymes (Zhang *et al.* 2015). Moreover Tsai et al., showed that TET1

interacts with hypoxia inducible factors (HIFs) increasing gene expression of genes responsive to hypoxia and to EMT (Tsai *et al.* 2014)**.**

1.8. Adult and Paediatric brain tumours

1.8.1. The concept of cancer heterogeneity: a primary feature of brain tumours

Cancer is characterised by inter-tumour and intra-tumour heterogeneity. Inter-tumour heterogeneity refers to differences between tumours in different as well as in the same organ; while intra-tumour refers to differences in the cells within the tumour (Inda *et al.* 2014). The latter refers to various clones that exist within the same tumour which is most probably caused by stochastic events favouring or disfavouring their maintenance under selective pressure imposed by treatment approaches (Inda *et al.* 2014).

There have been several attempts to explain the ontogenesis of cancers. The first proposed model was based on Darwinian theory of evolution (Nowell 1976). In this model, genetic and epigenetic alterations randomly occurred and are subjected to natural selection. According to this model the clones that can survive environmental changes, treatment approaches in particular, would survive and become dominant (Fig. 1.9) (Nowell 1976). The expanding field of developmental biology and stem cells, led to the cancer stem cell (CSC) theory for cancer initiation and evolution. In this model, the tumour arises from cells with stem cell characteristics; these cells will go under asymmetric division to generate more differentiated cells which will constitute the bulk tumour while CSC remain a small population (Fig. 1.9) (Shibata & Shen 2013). The heterogeneity in the microenvironment within the tumour due to differences in growth factors, oxygen supply, nutrient supply and blood vessel density is also considered to be the cause of genetic and epigenetic alterations observed in tumours (Fig. 1.9) (Marusyk & Polyak 2010). It is more likely that all three models described are responsible for the heterogeneity in tumours.

For instance, changes in the microenvironment of the tumour could induce either epigenetic or genetic changes in CSC altering their "stemness" properties and tumorigenic potential. Thus tumour heterogeneity is complex, affected by both hereditable genetic events and non-hereditable epigenetic alterations.

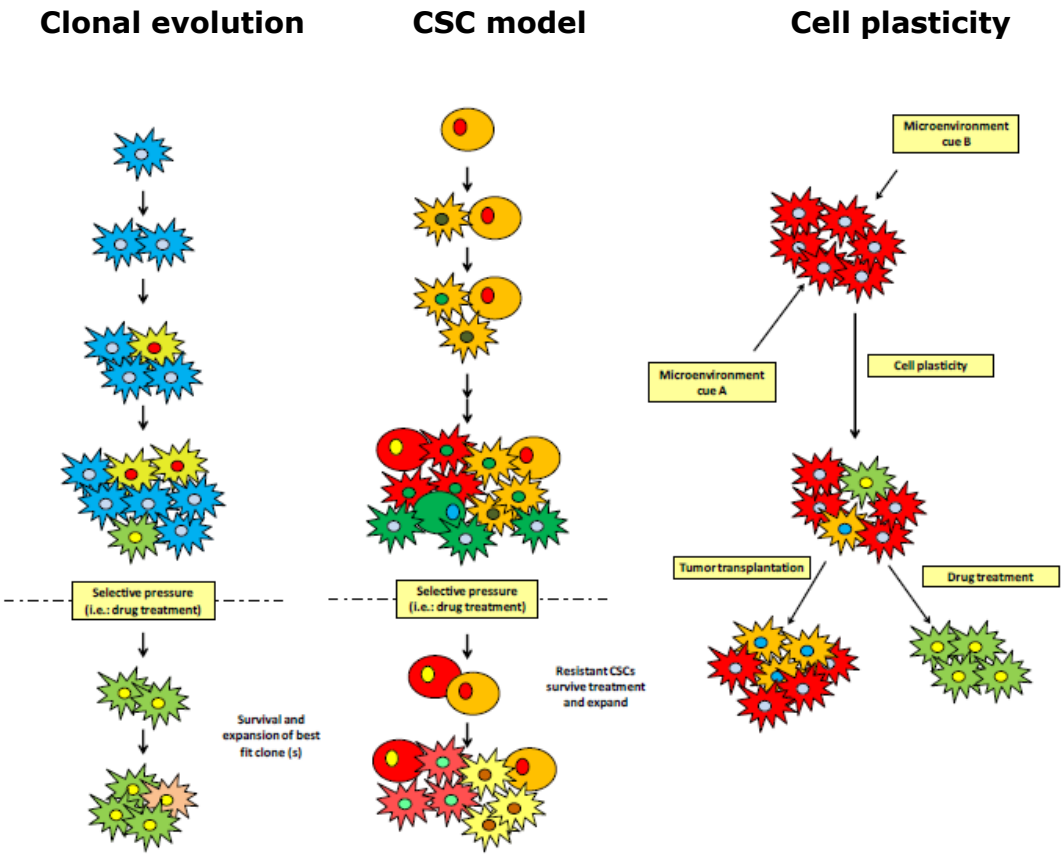


Figure 1.9: Models for the origin of cancer heterogeneity.

In the clonal evolution model, cells acquire mutations under selective pressure (e.g. drug treatment). In the CSC model, cells divide and generate CSCs and more differentiated tumour cells; under pressure CSCs can survive and will lead to tumour relapse. In the last model, heterogeneity is generated by cell plasticity in response to environmental cues. Figure modified from (Inda *et al.* 2014).

1.8.2. The concept of cancer stem cells in malignancies

Brain tumours are characterised by the subpopulation of cells that are able to recapitulate the tumour after treatment; which often called CSCs due to the similarities they have with stem cells. CSCs have been broadly characterised in multiple patient-derived tumours including breast tumours (Zhou *et al.* 2019), colon cancer (Zhou *et al.* 2018b) and brain tumours (Chen *et al.* 2012; Singh *et al.* 2003). Specifically, CSCs are characterised by their self-renewal ability; multi-lineage differentiation capacity; their expression of stemness markers including NANOG, OCT4, sex-determining region Y-box 2 (SOX2) and tumour formation ability *in vivo* (Lathia *et al.* 2015).

The first report of CSCs in brain tumours was in 2003; in this study CSCs that possess self-renewal, differentiation and proliferation capacities were purified from human MB tumours (Singh *et al.* 2003). This is not surprising considering that key pathways involved in NPC proliferation and specification (Wnt, Sonic hedgehog (SHH)) are aberrantly activated in these tumours indicating a link between normal neural progenitor differentiation and MB development (Gibson *et al.* 2010).

The comparison of glial tumours and normal brain cells in their ability to generate neurospheres in non-serum conditions led to the identification of a population of cells in glial tumours with similar characteristics to normal brain stem cells, including the expression of NESTIN and glial fibrillary acidic protein (GFAP) (Ignatova *et al.* 2002). In another study, Chen and colleagues, used a genetically engineered mouse model of GBM to identify a subset of quiescent tumour cells that are able to recapitulate tumour growth after the administration of temozolomide (TMZ) (Chen *et al.* 2012). CSCs' resistance to chemotherapy and radiotherapy causes tumour recurrence (Chen *et al.* 2012). TMZ is a common chemotherapy drug used for the treatment of brain tumours, through the methylation of guanidine DNA residues that would eventually lead to cell death (Beier *et al.* 2008). Interestingly, CSCs that overexpress O-6-methylguanine- DNA methyltransferase (MGMT) are not responsive to TMZ treatment due to the

ability of MGMT to remove the methylation introduced by TMZ leading to DNA repair (Beier *et al.* 2008). Therefore, despite the aggressive therapy that is given to patients, the residual CSCs lead to the recurrence of tumour. A follow up study showed that SOX2, OCT4, Spalt-like 4 (SALL4) and oligodendrocyte 2 (OLIG2) TFs are essential for GBM propagation and these TFs are able to induce de-differentiation of GBM cells to CSCs (Suvà *et al.* 2014).

1.8.3. Glioblastoma multiforme (GBM)

Gliomas are the most common form of primary brain tumour in the CNS (Williams *et al.* 2008). The World Health Organization (WHO) classifies gliomas into grades I-IV according to their malignancy level (Louis *et al.* 2016). Glioblastoma multiforme (GBM) is grade IV glioma and the most common and aggressive primary tumour of the adult brain with a median survival time of 9-12 months (Louis *et al.* 2016; Wesseling *et al.* 2015). It is a primary neuroepithelial tumour of the CNS with an extensively aggressive clinical phenotype due to the heterogeneity of tumour (Szopa *et al.* 2017). Current approaches for therapy involve surgical resection, although complete resection is challenging due to the invasive nature of this tumour, followed by whole brain radiotherapy and chemotherapy (Szopa *et al.* 2017).

There have been several theories about the cell of origin in GBM; one of them proposes that NSCs undergo malignant transformation retaining their stem cell features (Alcantara Llaguno *et al.* 2009). Moreover, transient amplifying progenitors have been suggested as the origin of GBM (Alcantara Llaguno *et al.* 2015). The last theory claims that differentiated cells undergo into de-differentiation and form glioma stem cells. Indeed, it has been shown that mutations in astrocytes, oligodendrocyte progenitors or in neurons are sufficient to confer stem cell properties during malignant transformation (Friedmann-Morvinski *et al.* 2012; Persson *et al.* 2010).

Initial molecular classification studies of GBM were performed using The Cancer Genome Atlas (TCGA) integrated gene expression data from

different platforms. Verhaak and colleagues identified 840 genes that classified GBM into four different subgroups according to the genomic profile of the tumour: classical; proneural; mesenchymal and neural (Verhaak *et al.* 2010).

According to the initial classification, the neural subtype is associated with oligodendrocytic and astrocytic differentiation and it is characterised by the expression of neuron markers such as Neurofilament light polypeptide (NEFL), Gamma-Aminobutyric Acid Type A Receptor Subunit Alpha1 (GABRA1) and Synaptotagmin 1 (SYT1) (Verhaak *et al.* 2010). However, subsequent studies in glioma stem cells and single-cell gene expression profiles showed that GBM can only be classified into proneural, mesenchymal and classical GBM (Behnan *et al.* 2019; Wang *et al.* 2017). It is possible that the neural type identified was due to contamination with normal neuronal cells (Behnan *et al.* 2019; Brennan *et al.* 2013; Sturm *et al.* 2012).

The classical subtype is associated with an astrocytic signature and characterised by high levels of epidermal growth factor receptor (EGFR), impaired levels of pro-apoptotic proteins and mitogen protein kinase (MAPK) (Brennan *et al.* 2013; Verhaak *et al.* 2010). Stem cell markers, neural precursor markers and Notch and SHH signalling pathways are also highly expressed in this subtype (Verhaak *et al.* 2010).

The mesenchymal subtype is characterised by lower levels of the neurofibromin 1 (NF1) gene, which are not correlated with any methylation signature in that area; mutations in *PTEN* gene and activation of MAPK pathway (Verhaak *et al.* 2010). This subtype is characterised by a combination of higher activity of mesenchymal and astrocytic markers and also genes in the tumour necrosis factor super family (Verhaak *et al.* 2010). The mesenchymal subtype tends to have worse prognosis than proneural and classical subtypes (Wang *et al.* 2017). A recently published study, proposed that the interaction between tumour and immune cells can drive the development of GBM (Martinez-Lage *et al.* 2019). Close examination of the immune phenotype of the GBM subtypes identified that mesenchymal GBMs have the highest percentage of microglia,

macrophages and lymphocytes which could explain the worse prognosis observed in this subtype (Martinez-Lage *et al.* 2019).

The proneural class is associated with an oligodendrocytic signature rather than the astrocytic signature and is characterised mainly by alterations of platelet derived growth factor receptor A (*PDGFRA)* and point mutations in *IDH1* (Verhaak *et al.* 2010). It is also characterised by normal *EGFR* and *PTEN* expression and Notch activation (Phillips *et al.* 2006). It has been suggested that patients with proneural GBM tend to be younger and display a favourable prognosis (Verhaak *et al.* 2010). However, this favourable prognosis is more likely to be associated with point mutations in the *IDH* gene in these tumours. Subsequent analysis of proneural GBM patient samples with the exclusion of patients with *IDH* revealed that proneural subtype is characterised by worse prognosis than previously described (Sturm *et al.* 2012).

Since then, a single cell RNA-seq study of GBM showed that it consists of a heterogeneous population of individual cells which correspond to different subtypes and this heterogeneity at the single cell level affects clinical outcome (Patel *et al.* 2014). Moreover, a subsequent study based on gene expression and DNA methylation profiles showed that GBM can be further classified into six methylation groups: clusters M1, M2, M3, M4, glioma CpG island methylator phenotype (G-CIMP) and M6 (Brennan *et al.* 2013). The classical subtype was enriched in the M1 cluster while the mesenchymal in the M1 cluster and the G-CIMP contained mainly the proneural subtype (Brennan *et al.* 2013). Therefore, these results indicate the critical importance of epigenetic marks in GBM classification.

1.8.4. Paediatric brain tumours: Medulloblastoma (MB) and Ependymoma (EPN)

Paediatric brain tumours are the leading cause of cancer mortality in children (Ward *et al.* 2014). MB and EPN represent the two most common forms (Northcott *et al.* 2012a). MB was first identified by James Homer Wright in 1910 (Wright 1910) and is the most common malignant childhood brain tumour accounting for around 20% of paediatric brain tumour cases (Louis *et al.* 2016). It has a 5-years survival rate of around 75-85% (Gajjar *et al.* 2006) and WHO grade IV status, the highest grade of malignancy. MB is a heterogeneous embryonal tumour that arises in the cerebellum and exhibits deregulations in critical pathways known to have a crucial role in cerebellum development (Coluccia *et al.* 2016; Hatten 2011). It is believed to originate from primitive neuroectodermal progenitor cells involved in the formation of the cerebellum and brain stem (Gibson *et al.* 2010). Current approaches to treatment include maximal safe tumour resection, chemotherapy and craniospinal radiotherapy for patients older than 3 years old.

MB is a highly heterogeneous tumour and is classified into four subgroups based on their transcription profiles and DNA methylation patterns: WNT; SHH; Group 3 and Group 4 (Jones *et al.* 2012; Northcott *et al.* 2011; Taylor *et al.* 2012). WNT tumours are rarely metastatic and represent the most favourable prognosis accounting for 10% of all the cases (Gajjar *et al.* 2015; Northcott *et al.* 2011; Taylor *et al.* 2012). Tumours in this category are often located in the fourth ventricle near the brainstem. The cell of origin is considered to be a progenitor in the lower rhombic lip (Fig.1.10) (Gibson *et al.* 2010). They are characterised by somatic mutations of Catenin Beta 1 (*CTTNB1)* leading to stabilisation and nuclear localisation of β-catenin and subsequently aberrant activation of WNT signalling pathway (Clifford *et al.* 2006; Thompson *et al.* 2006). Moreover, mutations in chromatin modifier genes are also common in WNT MB; indicating that deregulation of the epigenome is also of critical importance in the development of MB (Robinson *et al.* 2012).

SHH MB has an intermediate survival rate and represents around 30% of all MB cases (Gajjar *et al.* 2015). These tumours are located in the cerebellar hemispheres and granule neuron precursors (GNPs) which are considered to be the cell of origin (Fig. 1.10) (Gibson *et al.* 2010; Raybaud *et al.* 2015). Common alterations in SHH MB include mutations in genes involved in SHH pathway such as Patched (*PATCH1)* and *SUFU*, and amplifications of *MYCN* and GLI family zinc finger 2 *(GLI2)* (Jones *et al.* 2012; Kool *et al.* 2014; Robinson *et al.* 2012). Interestingly, some SHH MB patients have mutations in the *IDH1* gene which are also common in gliomas and contribute to DNA hyper-methylation (Northcott *et al.* 2017). According to a recent report, SHH tumours are heterogeneous and can be further subdivided into 4 subtypes. The SHH α subtype is characterised by *MYCN* and *GLI2* amplifications and *TP53* mutations and is associated with the worst prognosis among the 4 subtypes. SHH β tumours are often metastatic and mainly affect infants and like γ tumours which have a better prognosis (Cavalli *et al.* 2017). SHH δ subtype is mainly found in adults and contains telomerase reverse transcriptase (TERT) promoter mutations (Cavalli *et al.* 2017).

Group 3 accounts for 25% of all MBs, has the worst outcome and nearly 50% of tumours are metastatic (Kool *et al.* 2012). It is often located in the fourth ventricle near the brainstem (Raybaud *et al.* 2015). It has been suggested that the cell of origin is lineage-negative cerebellar stem cells and GNPs, however it remains unclear (Fig. 1.10) (Kawauchi *et al.* 2012). The most common genetic alteration that characterises group 3 MB is the amplification of MYC proto-oncogene (Northcott *et al.* 2011). Signalling pathways such as Notch and transforming growth factor- β (TGF-β) have also been found to be activated in group 3 MB tumours (Kool *et al.* 2012; Northcott *et al.* 2012b). Similar to SHH tumours, group 3 can be subdivided into three distinct subgroups (Cavalli *et al.* 2017). Group 3α tumours are metastatic and mostly found in infants. Group 3β has *growth factor independent 1 (GFI1)* family oncogene activation and orthodenticle homeobox *2 (OTX2)* amplification (Cavalli *et al.* 2017). Group 3γ is characterised by *MYC* amplification and is often metastatic, representing the worst prognosis among the three subtypes (Cavalli et al., 2017).

Group 4 is the most common subgroup with similar prognosis to SHH (Northcott *et al.* 2011). Tumours that belong to group 4 are frequently metastatic and often invade into the fourth ventricle near the brainstem (Taylor *et al.* 2012). Similar to group 3, the origin of the tumour is unclear; however, it is believed that they arise from primitive progenitors (Fig. 1.10) (Raybaud *et al.* 2015). Common alterations in Group 4 include inactivating mutations in the lysine demethylase 6A (*KDM6A)* gene, duplication of the gene encoding synuclein alpha interacting protein *(SNCAIP*) and amplification of the cyclin dependent kinase 6 *(CDK6)* and *MYCN* genes (Northcott *et al.* 2011; Taylor *et al.* 2012). Group 4 MB also shows inter-tumour heterogeneity. Groups 4α and 4γ have focal CDK6 amplification, chromosome 8p loss, and chromosome 7q gain, while Group 4α also exhibits MYCN amplification (Cavalli *et al.* 2017). Group 4β is characterised by *SNCAIP* duplication and *PR/SET domain 6 (PRDM6)* overexpression (Cavalli *et al.* 2017).

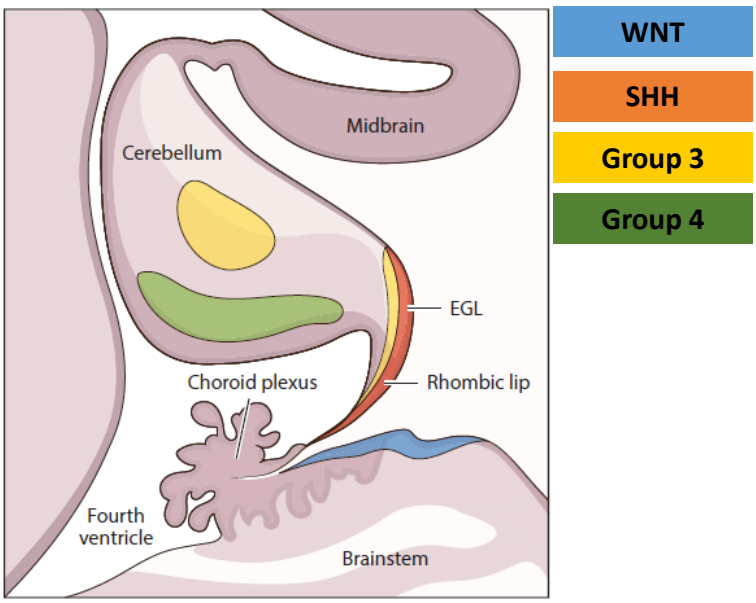


Figure 1.10: Schematic representation of the origin of the MB subtypes.

WNT tumours arise from the lower rhombic lip (blue), SHH is believed to arise from granule neuron precursors in the upper rhombic lip and external granule layer (EGL) (orange). Group 3 and 4 are most likely to arise from primitive progenitors (yellow and green) in different locations inside the cerebellum. Figure modified from (Wang *et al.* 2018).

Another type of malignant paediatric brain tumour is EPN which can arise from the supratentorial region (cerebral hemispheres), the posterior fossa and the spinal cord (Louis *et al.* 2007). The 5-year survival rate varies between 24-82% with some patients experiencing a rapidly growing tumour variants and others harbouring slow growing variants (Liu *et al.* 2017). 90% of EPN occurs intracranially, with 70% of them in the posterior fossa (Kilday *et al.* 2009). A study based on DNA methylation profiles of EPN identified nine distinct molecular subgroups, three in each area (spine, supratentoria and posterior fossa) which are genetically, epigenetically, transcriptionally and clinically distinct (Pajtler *et al.* 2015).

Conventional therapies for MB and EPN include maximal surgical removal, cranio-spinal radiotherapy and chemotherapy (Karajannis *et al.* 2009). However, radiotherapy cannot be applied to infants as the brain is still developing (Raimondi & Tomita 1983). Although in most cases MB and EPN are curable, the majority of patients suffer from long-term side effects, including impairments in information processing speed, IQ levels and visual memory (Mabbott *et al.* 2008; Spiegler *et al.* 2004). Like GBM, paediatric brain tumours also harbour a cancer stem cell population, expressing neural stem cell markers such as *SOX2* and *NESTIN*, which is responsible for tumour recurrence after treatment (Milde *et al.* 2012; Vanner *et al.* 2014). In EPN especially, tumour recurrence is a major issue as 50% of EPNs relapse after treatment even following aggressive radiotherapy (Sowar *et al.* 2006).

1.9. Hypothesis and Aims

Previous studies in our lab showed, for the first time, the presence of 5caC in glioma and breast cancer tissues coupled with its absence in normal brain and breast adjacent tissues (Eleftheriou *et al.* 2015). 5caC levels were elevated in a proportion of breast cancer tissues, however; these levels were not related to the 5hmC levels in the corresponding tissues (Eleftheriou *et al.* 2015). Thus, this was an indication of the preferential oxidation of 5mC into 5caC by TET enzymes in some cancers suggesting a possible distinct biological role of 5caC. A follow up study in our lab revealed that 5caC is also present in EPN and MB paediatric brain tumour cell lines coupled with high *TET1* levels (Ramsawhook *et al.* 2017). Considering the transient accumulation of 5caC during neuronal lineage differentiation (Wheldon *et al.* 2014) and the nature of brain tumours, their possible origin from neural/glial progenitor cell as well as the presence of glioma stem cells we hypothesised that 5caC presence in cancer cells might relate with their differentiation state thereby representing a unique epigenetic mark in brain cancers.

Therefore, the central goal of this thesis was to study the presence the genomic distribution and the possible biological roles of oxidative derivatives of 5mC in brain tumours cell lines. This thesis was split into three main aims:

1) To examine the presence, spatial and genomic distribution of 5caC in GBM and MB cell lines.
2) To identify the role of TET2 in the oxidation of 5mC to 5hmC/5caC/5fC in LN18 GBM cell line, as well as its potential role in cancer pathogenesis.
3) To study 5caC potential readers in adult and paediatric brain tumour cell lines.

Chapter 2. Materials and Methods

2.1. Cell culture

2.1.1. Preparation of complete media and maintenance of cells

LN18, LN229, U87-MG, U251 GBM cell lines BXD-1425EPN, DKFZ-EP1NS EPN cell lines and HeLa cells were cultured in Dulbecco's Modified Eagles medium (DMEM) (Gibco, Life technologies: 31885023) supplemented with 10% heat- inactivated foetal bovine serum (FBS) (ThermoFisher: SH30541.03). For the high glucose experiments; we used DMEM media supplemented with 25 mM glucose (Gibco, Life technologies: 11965092) GIN8 and GIN28 primary GBM cell lines were cultured in DMEM supplemented with 15% and 10% heat-inactivated FBS respectively. DAOY MB cells were cultured in HyClone™ minimal Essential Medium with Earle's MEM/EBSS supplemented with 10% heat-inactivated FBS, 1% MEM non-essential aminoacids (NEAA) 100X (Gibco: 11140-035), 1% glutamine and 1% sodium pyruvate UW228-3 MB cells cultured in DMEM/F12 (Gibco, Life technologies: 11320074) supplemented with 15% heat-inactivated FBS, 1% glutamine and 1% sodium pyruvate. ONS76 and HDMB03 cells were cultured in RPMI (Sigma Aldrich: 1640) supplemented with 10% heat-inactivated FBS. All the above cell lines were cultured in tissue-culture treated flasks. CHLA-01R and CHLA-01 cells were cultured in DMEM supplemented with 10% B27 Supplement 50 X (Gibco: 17504-044), 1% epidermal derived growth Factor (EGF) (Peprotech: 315-09) and 1% fibroblast growth factor 2 (FGF2) (Peprotech: 100-18B) in non-tissue culture coated flasks (NUNC). Information about media components in each cell line are displayed at Table 2.1. Cells were incubated at 37°C with 5% CO_2.

Table 2.1: Media formulation

Cell line	Media components
LN18, LN229, U87MG, U251, BXD-1425EPN, DKFZ-EP1NS, HeLa, GIN28	DMEM + 10% FBS
GIN8	DMEM + 15% FBS
DAOY	MEM/EBSS +10% FBS + 1% Glutamine + 1% sodium pyruvate + non-essential amino acids
UW228-3	DMEM/F12 + 15% FBS +1% Glutamine +1% Sodium pyruvate
ONS76, HDMB03	RPMI + 10% FBS
CHLA-01R, CHLA-01	DMEM + 10% B27 +1% EGF +1% FGF2

2.1.2. Glioma stem cell enrichment culture

LN18 cells cultured in stem cell enrichment conditions were cultured in neurobasal medium contained: DMEM/HAMS-F12 (Sigma: D8437), 7.25 ml Glucose 100 g/L (Sigma G8644), 1% NEAA 100 X (Gibco: 11140-035), 1% Penicillin Streptomycin (Gibco 15140-122), 800 µl bovine serum albumin (BSA) solution 7.5% (Gibco 15260-037) 1% β-mercaptoethanol 50 mM (Gibco 31350-010), 1% B27 Supplement 50 X (Gibco 17504-044) and 0.5% N2 Supplement 100 X (Gibco 17502-048). Complete neurobasal medium was further supplemented with 20 ng/ml bFGF (Peprotech: 100-18B) and 20 ng/ml EGF (Peprotech: 315-09). Cells were incubated at 37°C with 5% CO_2.

2.1.3. Sub-culturing of cells

$1x10^6$ cells were thawed from liquid nitrogen for 2 minutes in a 37°C water bath and seeded into a T25 (NUNC) flask containing 5ml of the appropriate media (Table 2.1). Fresh media was replenished after 24 hours in order to remove any residual dimethyl-sulfoxide (DMSO). Media was replenished

every 3 days and cells were passaged when reaching 70-80% confluence. 2 ml of 1 X 0.25% trypsin (Gibco) dissolved in HBSS (Fisher Scientific: 14170138) was used for the dissociation of the cells for 5-10 min at 37°C with 5% CO_2. Following incubation, the flasks were tapped gently to agitate adherent cells from the tissue culture plastic surface. Trypsin was then neutralised using complete media.

2.1.4. Cryopreservation of cells

Approximately $1x10^6$ cells were centrifuged at 200 g for 5 minutes, pelleted, re-suspended to 1 ml FBS containing 5% DMSO (Sigma: C6164) and immediately placed in a Mr Frosty freezing container. This container enables constant cooling rate at -1˚C per minute. Cells were immediately transferred at 80˚C freezer for 24 hours incubation; followed by their final transfer to liquid nitrogen for long term storage.

2.2. Cell culture assays

2.2.1.1. siRNA mediated transfections

SMARCC2, RCOR2, TET2 and TDG SMART pool 5 nmol siRNAs (Dharmacon: L-008312-00-0005, L-018296-02-0005, L-013776-03-0005, L-003780-01-0005) were dissolved in 500 µl dH_2O in order to make 10 µM stock solutions. Cells were seeded at approximately 40-50% confluence in a 12-well plate for gene expression analysis and in an 8-well chamber slide for immunostaining 24 hours prior transfection. 10 µM siRNA was diluted in serum free Opti-MEM (Gibco: 31985062) in order to achieve the optimal concentration up to a final volume of 100 µl for the transfection in the 12 well plate and 40 µl 8-well chamber slides. In a separate tube 3 µl of DharmaFECT transfection reagent (Dharmacon: T-2001-01/02) was diluted in 97 µl of serum free Opti-MEM, (1.2 µl Dharmafect in Opti-MEM for the 8-well chamber slide). The two mixtures were incubated separately for 5 minutes at room temperature then mixed and left to incubate for another 20 minutes. After the incubation time, serum free medium was

added to the mixtures before adding to UW228-3 and LN18 cells seeded in 12 well plates and 8-well chamber slides. The final siRNA concentration for all the genes was 50 nM per well. Cell were then incubated at 37˚C and 5% $C0_2$ for 48 hours. After 24 hours, transfection medium was replaced by complete medium and 24 hours later cells were harvested for gene expression and immunocytochemical analysis.

2.2.2. Electroporation transfection

Transfection of plasmid into LN18 cell line was achieved using the Nucleofector II AMAXA (Lonza) with the Cell Line Nucleofector™ kit L (Lonza: VCA-1005). We used 1 µg of gRNA and Cas9 plasmids, 2 µg donor and GFP plasmid. 500,000 cells were used per transfection. Briefly, cells were counted and re-suspended in Buffer L (Lonza: VCA-1005) and added to the DNA plasmids. Samples were transferred to cuvettes and placed in Nucleofector II AMAXA electroporator, following by transfection using the U-30 C6 glioma high efficiency settings. The cuvettes where incubated at 37^0C and 5% CO_2 for 10 minutes following their transfer into 6-well tissue culture treated plates.

2.2.3. Growth curves

Cells were seeded in tissue culture treated 6-well plates at a density of 10,000 cells per well. Their growth was monitored over the course of 8 days. Trypan Blue 0.4% solution (ThermoFisher: 15250061) was used to exclude dead cells from the counts; viable cells were counted every day using a haemocytometer. We performed 3 technical replicates in each biological replicate.

2.2.4. Limited dilution assay

Single cell suspensions from LN18 cell line and LN18 TET2 KO cell lines were seeded on low attachment 96-well plates in decreasing numbers (250-5 cells/well, serial dilution) in 200 µl complete neurobasal media in 6

replicates per condition. Cells were incubated at 37˚C 5% CO_2 in a humidified incubator and media was changed every 2-3 days. Spheroids were grown for 14 days in culture until reaching an approximate diameter of 100-200 µm. After 14 days in culture all the spheroids in each well were imaged, counted and measured using Image J. Cells were then dissociated using trypsin and single cells viable cells were re-seeded as previously and analysed in the same way.

2.2.5. Differentiation assay

Cells were conditioned for four days in complete neurobasal medium coated with 1% Laminin (Sigma: L2020-1MG) prior to differentiation. Complete neurobasal medium was then replenished with pro-neuronal medium containing neurobasal medium supplemented with 1 µM retinoic acid (RA) (Sigma: R2625), 10 ng/ml brain-derived neurotrophic factor (BDNF) and 0.5% FBS; or pro-glial containing neurobasal medium supplemented with 10 ng/ml bone morphogenic protein BMP4 (R&D: 414-BP), and 5% FBS. Media was changed every 48 hours and samples were harvested after 2, 4 and 6 days post differentiation induction.

2.3. Molecular biology

2.3.1. RNA extraction

RNA extraction was performed using RNeasy mini kit (Qiagen: 74104/74106) following manufacturer's instructions. Samples were eluted in RNAse-free dH_2O. RNA concentration was measured using NanoDrop-1000 spectrophotometer (ThermoFisher), 260/280 nm and 260/280 nm were also measured to ensure good purity and quality of RNA.

2.3.2. cDNA synthesis

Reverse transcription PCR (RT PCR) was performed for the cDNA synthesis. 2 µg RNA was added to a 20 µl reaction mix containing 1 µl 10 mM dNTPs (New England Biolabs (NEB): N0447S), 2 µl of 10X random hexamers (Thermo Scientific: N8080127) and dH_2O. Reaction mix was heated at 65˚C for 5 minutes followed by immediate ice incubation. 5 X first strand buffer (4 µl) 1 µl Dithiothreitol (DTT) (Invitrogen: D1532) and 0.1 M (1 µl) of Superscript III Reverse Transcriptase (Invitrogen: 180800) were added to the reaction mix followed by 5 minutes incubation at room temperature then 1 hour incubation at 50˚C. Inactivation of the enzyme was performed at 70˚C for 15 minutes. Reaction volume was diluted with dH_2O up to a total volume of 200 µl.

2.3.3. Gene expression analysis via Real-time Polymerase Chain Reaction (qPCR)

Gene expression was measured via qPCR on Applied Biosystems FAST 7500 qPCR machine and using the manufacturer's software. 5 µl GoTaq qPCR master mix (Promega: A6001), 0.25 µl forward and reverse primers (5 µM) (Sigma Aldrich) and 100 ng of cDNA were added to a 10 µl reaction mix. Primer sequences were designed using Primer3 and were verified using NCBI-Blast (Table 2.2). qPCR program was run at 40 cycles with denaturation at 95˚C, annealing temperature at 60˚C for 10 seconds each and extension step at 72˚C for 30 seconds. Gene expression was measured using comparative Ct values and it was normalised to Glyceraldehyde 3-phospate dehydrogenase (GAPDH) endogenous reference gene expression.

Table 2.2: List of qPCR primers

Gene name	Forward primer (5'-3')	Reverse primer (5'-3')
SMARCC2	TGAGAAGAAGGAGCCCAAGG	TCCTTCAGCACTTCCTCCTG
WIZ	AACTGCCTCTCCTCCTCCTA	TCAGGGGTGTCATGTCTTCC
THYN1	GCATCGCAGGACTCATGAAG	GCCAGGGGAATGAAACGTTT
RNF169	GGTGGAAAGGAGTCAGAGCT	TTATGCAGGGAAGCCAAGGA
ANKRD17	TGCTTCTGGTGGCTATGTGA	AGCAGGCTAAAGTAAGGGCA
MUM1	GTTCTTCCGGGTTTGCTGTT	CACAGCAAACACGACACAGA
RCOR2	ATTCCCTGACGAGTGGACAG	CATCACACTAGTTCGGCTGC
KDM1A	GCCACACCTCTCTCAACTCT	CCTGTCGCACTGCTGTATTC
SMARCC1	GGTTAGGATGGGACGGGAAA	GAGCACCCCAAACCTCAAAG
RREB1	GCCTCCTCTTCCCTTTCAGT	CGGCCCCTTCTAAATTCTGC
ZBTB7A	ACCTTCCCCTATCCATGCAC	TTTAGTGCAATCCCCGAGGT
CTCF	TGATGCAGTCGAAGCCATTG	CATCTGAAGAAGGGTGGGGT
HMCES1	GTGATCCTTCCAAGCTGCAG	AGTATGGCTGCCTCTGGTTT
DIDO1	ACGAAGCCTGTTACTCGTGA	GTCGCCTTTGTCGTCCATAC
NSUN5	CCATGTCATCGATGCCTGTG	AAGTCCTCCTCAGCCAGTTC
BRD7	GCTGGCAGAGAGAGAGAGAG	CAGATTCCTTCACGATGCGG
FIZ1	GGCCCTCCAAATCCAAATCC	CACACTCATTCCAGGGACCT
SFMBT2	ACACGAAACAGGAGGAGGAG	GAAGGGTCAGAAGCAGGAGT
TCF25	CTCACCATGTTCCCTGGAGT	TCCCAAGGTACAGGTTCACC
DNMT1	ACCAAGAACGGCATCCTGTA	GCTGCCTTTGATGTAGTCGG
EHMT1	AAGGACGGGGAGGTTTACTG	CATAGTCAAACCCGAGCTGC
ZBTB7B	CTTGAATACCGCACCCACAG	CTCTGGCCCTACAAAGTCCA
NEIL1	GGGCAAAGAGAGACCTTCCT	GAAAGGGTGAGTGCAAGCAA
NRF1	AAGACTCGCCTTCTTCTCCC	ACGTGTTTGTTGCCTCTTCC
ZSCAN26	GAGAAGAAAGCGCTCCGAAG	GTTTCTTTGGCTGGTCTGGG
ZNF462	CCTTTCCTCACCATGCATCG	CACAGCATCCCAATCTGACG
ZNF687	TCCCCTCTTCCTTTGAGCTG	CCCTGGAGACTGCTTGAAGA
EP400	GTATCTGGAGGAGGATGCCC	TGGTGTAAGCTGCTCCATGA
TET1	CAGTGTGTGCTCCTTTTCCC	TAGGAGACGTAGACCCACCA
TET2	GCAAGATCTTCTTCACAG	GCATGGTTATGTATCAAGTA
TET3	ACCGTGAGATGAGTCGTGAG	ACTTGATCTTCCCCTCCAGC

TDG	CAGCTATTCCCTTCAGCA	GGAACTTCTTCTGGCATTTG
SOX2	CACATGAACGGCTGGAGCAAC	GTAGGACATGCTGTAGGTGGG
NANOG	CAGCCTTTACTCTTCCTACCACCAG	GTTATAGAAGGGACTGTTCCAGGCC
NESTIN	GGAAGAGGAAGAGAACCTGGGAAA	CTCAGATTCAGCTCTGCCTCATC
NOTCH1	CTTTGAGACTGGCCCACCTC	CCGCAGAGGGTTGTATTGG
BMP2	CTTCTAGCGTTGCTGCTTCC	AGTGCCTGCGATACAGGTCT
BMP4	TGATACCTGAGACGGGGAAG	CCAGACTGAAGCCGGTAAAG
NTRK4	AGCGTCTGGCTGGACTATGT	TTGATACTGGCGTTCCCATT
EGFR	CATTTAGGGGTGACTCCTTCACAC	GACGACTGCAAGAGAAAACTGACC
DYRKB1	TGTGCCTGGTATTTGAGCTG	CTTGAGGTCGCAGTGAATGA
ARRB2	TGGAGCTGCCTTTTGTTCTT	TCATCCTTCATCCCCTTCAG
GLI3	TGCAGGGTGAATGGTATCAA	TGATTAGCACCTGGGGAAAG
PTCH2	TTCCGGGGTGACTATCAGAC	CATGGTCACACAGGCATAGG
HHIP	TGCTCTCCTTTAAGCTGCTG	CTGGGACATCATCCTCCTGT
OCT4	CAGGCCCGAAAGAGAAAGCG	GTTGTGCATAGTCGCTGCTTGATC
GAPDH	GATGCTGGCGCTGAGTACG	GCAGAGATGATGACCCTTTTGG
GFAP	GCCAGCTACATCGAGAAGGTTC	CTGTGCCAGATTGTCCCTCTC
PDGFR	GAGATGCTTTGGGGAGAGTGAAG	GCCTGCCTTCAAGCTCATTCTC
β-tubulin III	CCTCAAGATGTCCTCCACCTTCAT	GTTCATGTTGCTCTCGGCCTC
GLAST	GTGAATGGAGTCAATGCCCTGG	CTCCACAATCTTCCCAGCAATCAG
CD133	GATGCCTCTGGTGGGGTATTTC	CGACTCCTTTTGATCCGGGTTC

2.3.4. Semi-quantitative PCR reactions

Semi-quantitative PCR reactions were performed in a total 50 µl reaction volume. Reactions were performed using 10X *Thermus aquaticus* (Taq) polymerase reaction buffer (NEB: M0267L), 10 mM dNTPs (NEB: N0447S) 10 µM of each primer and 2.5 Units (U) Taq DNA polymerase (NEB: M0267L) (Table 2.3).

Table 2.3: PCR reaction mix

Component	Concentration
Taq polymerase buffer	10X
dNTPs	10 mM
Forward primer	10 µM
Reverse primer	10 µM
Template DNA	50ng
Taq DNA polymerase	2.5 U
dH_20	Up to 50 µl

Primer annealing temperature was calculated using OligoCalc online software tool (http://biotools.nubic.northwestern.edu/OligoCalc.html). Polymerase extension time at 72˚C was estimated based on the length of the amplified product (1 minute per kilobase). PCR conditions are summarise at Table 2.4, steps 2-4 were repeated for 30 cycles, and were performed on Tecne TC312 thermocycler.

Table 2.4: Thermocycler conditions

Condition	Temperature (˚C)	Minutes/seconds
1. Initial denaturation	94	5 minutes
2. Denaturation	94	30 seconds
3. Annealing	Variable (55-65)	30 seconds
4. Extension	72	Variable
5. Final extension	72	5 minutes
6. Hold	4	Indefinitely

2.3.5. Immunostaining

Immunostaining for 5mC, 5hmC and 5caC was performed as previously described (Abakir *et al.* 2016). In brief, cells were seeded at a density of $1.5x10^4$ per well in 8-well chamber slides and fixed after 48 hours in 4% paraformaldehyde (PFA). Fixed samples were washed in phosphate buffered saline (PBS) (Calbiochem: 524650-IE). PBX (0.5% Triton X-100 (Sigma: T8787-1000ML in PBS) was used for permeabilization for 30 minutes at room temperature, excess PBX was removed with PBS washes. Depurination of DNA was achieved with 1 hour incubation of 4 M hydrochloric acid (HCL) (ThermoFisher: 10316380) at room temperature followed by several PBS washes. Samples were then placed in a humid chamber and incubated for 1 hour with blocking solution (10% FBS in PBS); followed by 1 hour incubation with the primary antibody at room temperature. Antibodies used are summarized in the Table 2.5. Samples were then washed in PBT (0.01% Tween (Sigma: P1379) in PBS). Secondary antibody was applied for 1 hour at room temperature; 1:400 goat anti-rabbit HRP-conjugated antibody (Invitrogen: 65-6120) and 1:400 of goat anti-mouse 633 (Invitrogen: A21052). Following 3x5 minutes washes in PBT; slides that were stained for 5caC were incubated for 2 minutes in 1:200 Tyromide in TSA (Perkin Elmer: NEL741001KT) followed by 3x5 minutes washes with PBS. Slides were then mounted using mounting medium (Vectashield: H-1200) containing DAPI. The same procedure was followed for the staining for protein markers (Nestin, β-tubulin III, GFAP) omitting the HCL depurination step.

Table 2.5: Antibodies list and specifications

Antibody	Specifications	Dilution
5caC	anti-5caC rabbit polyclonal antibody (Active Motif: 61229)	1:500
5hmC	anti-5hmC mouse monoclonal (Active Motif: 39999)	1:1000
5hmC	anti-5hmC rabbit polyclonal antibody (Active Motif: 39791)	1:1000
5mC	anti-5mC mouse monoclonal (Diagenode: C15200081-100)	1:100
Nestin	Anti-Nestin clone 10C2 mouse monoclonal (Millipore: MAB5326)	1:200
β-tubulin III	Anti- β-tubulin III mouse monoclonal (Proteintech: 66240-1-Ig)	1:200
GFAP	Anti-GFAP chicken polyclonal antibody (ThermoFisher: PA1-10004)	1:200

2.3.6. Western blot

Whole protein extraction was performed using RIPA lysis buffer (Sigma: R0278) supplemented with protease inhibitor cocktail (Sigma: P8340) following manufacturer's instructions. Protein concentration was measured using the Bradford reagent (Sigma: B6916) following manufacturer's instructions. A standard curve was produced using BSA standards (Sigma: P0834) absorbance was read using a Tecan Plate reader and the protein concentration was determined based on the standard curve.

6 µg of total protein was used for each reaction. Protein samples were denatured in 3X SDS loading buffer (Thermofisher: NP0002) with 1:10 DTT at 60˚C for 10 minutes. Denatured samples were loaded in NuPAGE™ 4-12% Bis-Tris Protein Gels, 1.5 mm (Invitrogen: NP0335) and electrophoresed in 1 X NuPAGE MES SDS Running buffer (Thermofisher: NP00002) at 200 Volts for approximately 3 hours.

Proteins were transferred from gels onto the Nitrocellulose/Filter Paper Sandwich, 0.45 µm pore size (Abcam: ab133412) and in 1 X NuPAGE Transfer Buffer (ThermoFisher: NP00006) supplemented with 20% (v/v) methanol at 25 Volt for 2 hours.

Nitrocellulose membranes with the bound proteins were then transferred into tubes containing blocking solution (5% skimmed milk diluted in 1 X Tris Buffered Saline 10 X (TBS) (Invitrogen: BP24711) supplemented with 0.1% Tween). Samples were incubated in blocking solution for 1 hour at room temperature; to avoid non-specific antibody binding. Membranes were then washed with TBS supplemented with 0.1% Tween (TBST) followed by 1 hour incubation at room temperature with the primary TET2 Rabbit pAb antibody (Abcam: ab94580) at 1:2000 dilution. Samples were then washed with TBST and incubated with the secondary antibody Goat-anti rabbit IgG (H+L), HRP conjugated (Invitrogen: 65-6120) at 4˚C overnight in a 1:10000 dilution. Following washes with TBST, membranes were developed using the Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare: RPN2232) following the manufacturer's instructions.

2.4. 5caC DNA immunoprecipitation (DIP) Next generation sequencing (NGS)

2.4.1. DNA extraction

DNA immunoprecipitation requires large amounts of high quality genomic DNA; therefore, we first optimised the DNA extraction conditions. We compared DNA extraction using the DNeasy extraction Kit (Qiagen: 69504) following manufacturer's instructions and phenol/chloroform DNA extraction. The latter extraction method led to higher concentration and quality of DNA therefore we proceeded with that approach.

In the phenol/chloroform approach, cell pellets were re-suspended in the appropriate amount of lysis buffer (200 mM Tris Ph 8.0, 4 mM EDTA, 20

mM NaCl, 1% SDS). 3.6 µg/ml RNAse A (Qiagen: 19101) was added to the lysed cells following by 1 hour incubation at 37˚C. Proteinase K (Promega: V3021) 200 mg/ml was added to the samples and left to incubate at 55˚C overnight. Extraction was performed with the addition of 1 volume Phenol: Chloroform: Isoamyl Alcohol 25:24:1 Saturated with 10 mM Tris, pH 8.0, 1 mM EDTA (SIGMA: P2069) followed by vigorous mix and centrifugation at 14,000 g for 5 minutes. Then the aqueous phase was placed in the fresh tube followed by a second round of phenol/chloroform extraction. DNA was precipitated with 1/10 5 M sodium chloride (NaCl) diluted in one volume isopropanol. Samples were incubated at -20˚C for 1 hour followed by centrifugation at 14000 g for 10 minutes. Isopropanol was removed and DNA pellets were washed with 70% ethanol following by further centrifugation. Samples were then left to dry at room temperature to remove residual ethanol and re-suspended in 50 µl nuclease free dH_20.

DNA was quantified using the Qubit® dsDNA BR Assay Kits (ThermoFisher: Q32850) and Qubit® Fluorometer following the manufacturer's instructions. DNA was then diluted and DNA concentration and integrity was further assessed using the Agilent High Sensitivity DNA kit (Agilent: 5067-4626) following the manufacturer's instructions.

2.4.2. DNA sonication

DNA samples were diluted in 700 µl dH_20 and sonicated using Covaris S220 ultrasonicator at a frequency of 550 KHz in order to generate 350 base pairs (bp) fragments. The sonicated samples were analysed using the Agilent High Sensitivity DNA kit (Agilent: 5067-4626) following the manufacturer's instructions in order to ensure correct fragmentation.

2.4.3. Adaptor ligation

Adaptor ligation was performed in the sonicated samples using NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1) (NEB: E7335S) following the manufacturer's instructors from the NEBNext Ultra II DNA library Prep Kit (NEB: E7645S). We used 5 µg for samples prepared for

meDIP and 1 µg of input samples. Clean up of adaptor ligated DNA was performed using size selection using the AMPure XP beads (NEB). DNA was eluted in 20 µl and 50 µl 0.1 X TE (10 mM Tris pH 7.5, 1Mm EDTA) for input and meDIP samples. Libraries were verified for DNA size distribution using Agilent Bioanalyzer and the Agilent High Sensitivity DNA kit (Agilent: 5067-4626) following the manufacturer's instructions.

2.4.4. 5caC DNA immunoprecipitation (meDIP)

DNA immunoprecipitation was performed following the adaptor ligation in all the samples apart from input sample. DNA was denatured at 95˚C for 10 minutes samples were then immediately placed on ice for 10 minutes. 51 µl of 10X IP buffer (100 mM Na-Phospate pH 7.0, 1.4 M NaCl, 0.5% Triton X-100) were added in each sample. 2.5 µl of 5caC rabbit polyclonal antibody (Active Motif: 61229) were added in each sample following by 2 hours incubation whilst spinning at 4˚C. Sheep anti-rabbit Dynabeads M-280 (Invitrogen: 11203D) were used for the precipitation of 5caC modified DNA fragments. 20 µl Dynabeads were washed, re-suspendend in 20 µl 1 X IP buffer and added to each sample. Samples were left to incubate for 2 hours at 4˚C whilst spinning. Following the incubation, samples were placed into the magnetic rack to remove the supernatant. They were then washed with 1 X IP buffer and incubated overnight in a shaking incubator at 55°C with Proteinase K buffer containing 25 mg/ml proteinase K (Promega: MC5005). DNA was purified using QIAquick PCR Purification Kit (Qiagen: 28104) and eluted in 25 µl 0.1 X TE. Immunoprecipitated DNA was quantified using QubitTM dsDNA HS Assay Kit (ThermoFisher: Q32854) following the manufacturer's instructions.

2.4.5. PCR amplification of the meDIP samples

PCR amplification of meDIP and input samples was performed following the instruction from NEBNext Ultra II DNA library Prep Kit (NEB: E7645S). PCR amplification cycles were determined based on the concentration of the samples in order to generate 100 ng of DNA. Samples were eluted in dH_2O and size distribution was examined using Agilent Bioalanyzer and the

Agilent High Sensitivity DNA kit (Agilent: 5067-4626) following the manufacturer's instructions.

2.4.6. Next generation sequencing (NGS)

NGS was performed at DEEPSEQ facility of University of Nottingham on the Illumina NextSeq 500 short read platform using 150bp pair-ended reads. 30 million reads were generated per biological replicate. We used two biological replicates for each sample, one input sample was used per cell line.

2.4.7. Bioinformatics Analysis

Bioinformatics analysis was performed at the Advanced Data Analysis Centre (ADAC) at the University of Nottingham by Dr Tom Giles. The analysis of bioinformatics was performed as previously described (Abakir et al. 2020). In brief, the 150bp Illumina paired-end reads were trimmed using the Skewer tool in order to remove adapters and gain the correct paired ends (Jiang et al. 2014). Trimmed reads were then aligned to the human Ensembl genome (build hg38.89) with the Burrows-Wheeler aligner (Li & Durbin 2009). The alignments were merged using SAMtools (Li et al. 2009); PCR artefacts were removed using de-duplication with picard-tools MarkDuplicates (Chang et al. 2013). Phantompeakqualtools (Marinov et al. 2014) was used to assess the impact of each pulldown and MACS2.1.1 was used to identify highly modified regions (peaks) (Zhang et al. 2008). Consensus peaks were identified using the Bioconductor package DiffBind (Stark, R and Brown 2016). Peak calling was performed against input DNA.

The genomic distribution of the peaks was assessed using the R package ChiQC25. In order to extract gene biotypes, annotated gene features, intergenic regions, introns, exons and sequences 3kb up/downstream of genes the Ensembl Genome GTF file (Homo_sapiens.GRCh38.89.gtf) was used. For the identification of repeat-masked regions, the Repeatmasker Library (db20140131) was downloaded from UCSC genome browser. Regulatory features were identified using the Ensembl regulatory build

(HgGRCh38.Regulatory_Build.regulatory_features.20161111). BedTools intersect was used for the annotation of bedfiles with the list of gene features, gene biotypes, repeat-masked sequences and regulatory regions. Gene ontology (GO) analysis was performed in the protein coding related 5caC peaks (3kb up/downstream to the nearest protein coding gene) using PANTHER with the default settings.

2.5. Generation of plasmids for CRISPR/Cas9 transfections

2.5.1. Design of primers

All primers were designed using SnapGene Software. They were examined for specificity using NCBI Blast tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi). All the primers used for the design of CRISPR/Cas9 plasmids and for genotyping of the clones are listed in the Table 2.6.

Table 2.6: Primer sequences for CRISPR/Cas9 building and genotyping

Guide RNA primers	
L gRNA ex3 (F)	TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCAGATATGTCTGGTCAACAAG
L gRNA ex3 (R)	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCTTGTTGACCAGACATATCT
R gRNA ex3 (F)	TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCCTGCGCAACTTGCTCAGCAA
R gRNA ex3 (R)	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTTGCTGAGCAAGTTGCGCAG
Primers for donor plasmid	
F. LHA	CGTACAGGATCCCCCTCGTTTTCACCAAGC
R. LHA	GAGTCTGGCGCCATCCCTGTAGAACTGAAGCTT
F. RHA	AAAATACGGCGCGCCTTGATACATAACCATGC

R. RHA	GGGACCTCTAGAGATGGGAAATCATATTGAG
F. LHA in fusion	TCAAGGCGCGCCGTATTTTGAGTCTGGCGCCATCCCTGT
R.RHA in fusion	ACAGGGATGGCGCCAGACTCAAAATACGGCGCGCCTTGA
F. Cassette	GAATCTTTGCGCGCTAATGGACCTTCTAGGTCTTGAAAGGA
R. Cassette	TGACGTCAAACAGCTATGACCATGG
Primers to check the donor plasmid	
F. LHA-EF1α	GCAGGAAACAAGACCCAA
R. LHA-EF1α	GCAGCAGGTCATCAAAAAT
F. EF1α-EGFP	AAGTTAGGCCAGCTTGG
R. EF1α-EGFP	TGCACGCCGTAGGTCA
F. EGFP-BSD	GCGATCACATGGTCCTG
R. EGFP-BSD	CGCTGTAGTCTTCAGAG
F. BSD-RHA	TGAATTGCTGCCCTCTG
R. BSD-RHA	CTGTGTTTGCTGCTGTTC
F. 3UTR-RHA	GCCTCAATAAAGCTTGCC
R. 3UTR-RHA	CTGTGTTTGCTGCTGTTC
Primers to check integration with Nickase Cas9	
F. 5' Int.nickase	AGACTGCAGGGACAATGACTG
R. 5' Int.nickase	CGGAGCCAGTACACGACATC
F. 3' Int.nickase	TGGGAGCTCTCTGGCTAACT

R. 3' Int.nickase	CACAAGACACAAGCATCGGT
F. KO	TGTCTGGTCAACAAGCTGCG
R. KO	TCAGTTTGGGGTTGCTGTGT
Primers to check integration with wild type Cas 9	
F. 5' int (B)	GAGACAAGGAGCAAACACGA
R. P. 5' int (A)	CCTAGAAGGTCCATTAGCGC
F. 5' int (A)	CTGAAACCAGGATGGATTGA
R. 5'int (B)	CATAACCCGTAAAGAGGCCA
F. 3' int (A)	GCATCGCATTGTCTGAGTAG
R. 5' int (A)	TCTCAGTACATTTCTGGCAC
R. 5' int (B)	TTACCTACACATCTGCAAGATG
F. ex3 1	AGACTTTTCCTCACCCCCAAA
R. ex6 1	GTAGCTTCTCTTCACTGCTGC
F. ex3 2	TCCCAGAGTTCACATCTCCCT
R. ex6 2	TCACTGCTGCTTCTGCGAAC
F. ex3 3	CCCAAAGCAACAATGATCAGCA
R. ex6 3	CACAATCACTGCAGCCTCACA

2.5.2. Annealing PCR

A PCR reaction was performed in order to achieve annealing of gRNA primers. Reactions were performed in a total volume of 50 µl, 10 µl of 5 X HF fusion buffer (NEB: B0518S) 1O µM of each primer and 10 mM DNTPs (NEB: N0447S). PCR reactions for DNA renature was performed (Table: 2.7). After the initial PCR reaction 2.5 U of Phusion Polymerase (NEB: M0530S) was added to the samples followed by PCR extension at 72˚C for 2 minutes.

Table 2.7: Annealing PCR conditions

Temperature	Time
94^0C	30 seconds
79^0C	5 minutes
52^0C	5 minutes
45^0C	5 minutes
4^0C	Hold

2.5.3. Phusion PCR

Phusion PCR was used for the generation of the intermediate plasmid. PCR oligonucleotides for the arms of homology were designed with 20 bp overlaps to ensure sequence hybridisation. The reaction was performed in three steps: the denature renature PCR, the extension PCR, and the standard PCR for product amplification. For the first step, equimolar DNA of arms of homology with complementary overlaps were mixed with 5 X phusion buffer (NEB: B0518S) in 15 µl reaction and incubated in the thermocycler under the conditions in Table 2.7. In the second step, DNTPs (NEB: N0447S) and 2.5 U Phusion Polymerase (NEB: M0530S) were added to the mixture and incubated at 72˚C for 2 minutes. Lastly the designed primers for the amplification of the fused product were added to the mixture followed by their incubation in the thermocycler for the standard PCR amplification (Table 2.4).

2.5.4. Gibson assembly

Gibson Assemblies was performed using the NEB Gibson assembly master mix (NEB: E2611S) and protocol with a ratio of 1:5 plasmid to insert.

2.5.5. Restriction digestions and ligations

Restriction digests were performed using 1 µg of plasmid DNA and 100 ng of PCR products in a total reaction volume of 50 µl in the appropriate reaction buffers (NEB) depending upon the enzyme. All restriction enzymes were purchased from NEB; 1 U of restriction enzyme was used per µg of DNA and the reaction mix was incubated in a water bath according to manufacturer's conditions for each enzyme; high-fidelity NEB enzymes were preferred for overnight digestions in order to achieve maximal activity. AfIII (NEB: R0541S) enzyme was used for the digestion of the gRNA cloning vector; BAMHI (NEB: R3136) and XbaI (NEB: R0145s) were used for the digestion of the pBluescript and the fused arms of homology; AscI (NEB: R0558S) and BssHII (NEB: R0199S) were used for the digestion of the cassette and the intermediate plasmid. Intermediate plasmid was further treated with shrimp alkaline phosphatase (rSAP) (NEB: M0371S) for one hour to prevent self-ligation. Ligations were performed using T4 DNA ligase (T4) (NEB: M0202S) followed manufacturer's instructions. Vector and insert DNA were added to the mix at a ratio of 1:3.

2.5.6. Agarose gel electrophoresis

Agarose gel electrophoresis was performed to detect the products of PCR reactions and restriction digestion. Agarose gels ranging from 0.8% - 2% were cast (depending on the size of the product). For the visualisation of the DNA bands, ethidium bromide (Invitrogen: 15585011) was added at the gels at a final concentration of 250 µg/ml. DNA samples were mixed with 6 X loading dye (NEB: B7021S) to a final concentration of 1 X dye in each reaction and loaded into the gel with 5 µg of DNA Ladder (NEB: N3231S, N3232L). Gels were run at 90 volts, 400 mA for 1 hour and visualised using the Fujifilm LAS 4000 UV transilluminator; pictures captured in LAS 4000 image acquisition software. DNA was extracted when needed using the QIAquick Gel Extraction Kit (Qiagen: 28704) following

manufacturer's instructors. DNA was eluted in nuclease free dH_20 and quantified in NanoDrop-1000 spectrophotometer (ThermoFisher).

2.5.7. Bacterial transformation

Plasmid transformations were performed using One Shot TOP10 chemically competent E.coli bacteria (Invitrogen: C404003). Plasmid DNA was added to the cells, gently mixed and incubated on ice for 30 minutes followed by heat-shocked at 42˚C for 30 seconds in a water bath. Cells were then immediately placed on ice for 2 minutes and allowed to grow for 1 hour in SOC medium (Invitrogen: 15544034) at 37˚C at 225 RPM. Cells were then plated into pre-made antibiotic selection LB agar plates and incubated for 16 hours at 37˚C. Grown bacterial colonies were picked and placed in 5 ml LB medium supplemented with antibiotic; incubated 12-16 hours at 37˚C at 225 RPM for further bacterial growth.

2.5.8. DNA mini prep and midi prep

DNA miniprep extractions from bacterial cultures were performed using the Wizard® Plus SV miniprep DNA purification kit (Promega: A1460) following the manufacturer's instructions. The DNA was eluted in nuclease-free dH_2O.

When large amounts of DNA plasmid were needed for transfections onto mammalian cells, we performed DNA midiprep extractions with the NucleoBond® Xtra Midi kit (Macherey-Nagel: 12773550) following the manufacturer's instructions. The plasmid DNA was eluted in nuclease-free dH_2O. DNA concentrations were measured using NanoDrop-1000 spectrophotometer (ThermoFisher).

2.5.9. DNA extraction

Genomic DNA was extracted using the DNeasy extraction Kit (Qiagen: 28104) following the manufacturer's instructions. Samples were eluted in nuclease-free dH_20. DNA concentration was measured using Nanodrop

spectrophotometry, 260/280 nm and 260/280 nm were also measured to ensure good purity and quality of DNA.

Alternatively, in the case of DNA extraction from the clones for rapid initial genotyping experiments, cell pellets were lysed with DirectPCR lysis reagent (VWR: 31-302-C) supplemented with 5% proteinase K (Promega: V3021). Pellets were incubated at 55˚C overnight followed by 1 hour incubation at 85˚C for proteinase K inactivation. The lysates were diluted with nuclease free dH_20 and kept for subsequent PCR experiments.

2.5.10. Sequencing

Samples were sequenced using an external Sanger sequencing service (Source Bioscience). Sequencing results were analysed using SnapGene software.

2.6. Confocal microscopy

Fluorescence images were captured with the Zeiss LSM 700 AxioObserver confocal microscope using a Plan-Apochromat 63x/1.40 Oil DIC M27 objective Zeiss. Control staining without primary antibody produced no detectable signal. All images were captured at identical exposure settings and acquisition protocols for all the channels; images processed using Image J.

2.7. Image quantification

We used Image J for the quantification of all the immunocytochemistry images. We measured the signal intensity of each channel in individual nuclei in all the cells for each technical and biological replicate. In brief, we adjusted the threshold in each image individually in order to make sure we select for the nuclei of all the cells in the images and then measured the

intensity values of all the channels in individual nuclei. Resulted values in each image (three in each biological replicate) were averaged in order to gain the values of each biological replicate. We then used GraphPad PRISM 6 software for data visualization and to perform statistical analysis.

2.8. Statistical analysis

All experiments were performed in three biological replicates, unless stated otherwise, the mean and standard deviation or standard error (where stated) was reported and plotted in each figure. Statistical analysis was performed using GraphPad Prism 6 Software. Normality tests were performed in all datasets in order to check for normal distribution. Paired and un-paired two tailed parametric Student's t-tests were performed on normally distributed data. All statistical tests were conducted using 95% confidence interval and were denoted by NS $p \leq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ and **** $p < 0.0001$.

Chapter 3. The global distribution of Oxi-mCs in the DNA of adult and paediatric brain tumours

3.1. Introduction

Whole genome methylation profiling has been used for the study of the events that impact DNA (de)methylation machinery (Guibert *et al.* 2012; Kang *et al.* 2014), response of cells to treatment with compounds (de Conti *et al.* 2014) and epigenetic changes that contribute to tumorigenesis (Jin *et al.* 2011). Several methods have been developed over the last few decades for the detection, examination of the spatial distribution and quantification of DNA methylation. However, our knowledge about the biological role of DNA methylation and demethylation in gene expression, genomic stability and chromosomal remodelling in cancer pathogenesis is still very limited.

Therefore, choosing the right method that is most appropriate to answer a particular biological question is then of crucial importance. There are several factors that need to be taken into consideration when choosing the right method for DNA methylation analysis in order to receive an unbiased answer to the question being asked. These factors include: the aim of the study; the amount and quality of the sample that is available; sensitivity and specificity of the study; robustness and simplicity of the method; as well as availability of equipment and the cost of reagents (Kurdyukov & Bullock 2016).

Digestion based assays were the first approach used to study DNA methylation (Ben-Hattar & Jiricny 1988; Cedar *et al.* 1979). They are based on restriction endonuclease methylation dependent digestion. Restriction enzymes recognize and cleave specific DNA sequences known as their recognition sites; the activity of some of these restriction enzymes is based on the methylation status of the cytosines in the recognition site.

Methylation can either block or enable digestion, since some restriction enzymes are only active when their recognition site is methylated (Umer & Herceg 2013). Endonucleases such as HpaI have the ability to only digest un-methylated sites whereas other enzymes like MspI can digest the same sites regardless of the DNA methylation status of the area. Therefore, digestion of a sample with HpaI and MsPI and analysis of the size of the digested sample can reveal the location of the sites of DNA methylation (Takamiya *et al.* 2006). Nevertheless, methods relying on the use of methylation-sensitive restriction enzymes provide very little information about the whole genome methylation as only 3.9% of all CpGs in human non-repetitive DNA resides at the recognition sites of the HpaII enzyme which is frequently used (Rakyan *et al.* 2004).

One of the earliest gold standard methods established for the quantification of the global levels of DNA methylation is reversed-phase high-performance liquid chromatography (RP-HLPC) first developed by Kuo et al., in 1980 (Kuo *et al.* 1980). This method requires few micrograms of hydrolysed DNA which can then be directly separated by RP-HPLC followed by the detection using ultraviolet (UV) absorption at 254 nm and 280 nm (Kuo *et al.* 1980). The DNA is hydrolysed into nucleoside bases and mC and dC are separated chromatographically; the fractions are then measured and calculated for each sample (Kuo *et al.* 1980). The accuracy as well as the reproducibility of chromatography-based methods have made this technique the gold standard approach for the quantification of global DNA methylation levels (Kurdyukov & Bullock 2016). However, RP-HLPC is not readily available and feasible as it requires large amount of DNA (3-10 µg) which is not always possible especially in the case of patients' samples; it also requires specialised laboratory equipment (Kurdyukov & Bullock 2016).

Techniques based on bisulfite conversion were introduced in 1992 allowing the analysis of DNA methylome at single base resolution (Frommer *et al.* 1992). Bisulfite treatment of DNA mediates the deamination of cytosine into uracil. After PCR amplification and Sanger sequencing analysis the converted bases will be read as thymine while methylated cytosines are resistant to that conversion and will be read as cytosine. Therefore, the

comparison of reads between the treated and untreated samples allows the detection of the methylated cytosines (Frommer *et al.* 1992). Once the samples are converted they are then quantified; sequencing is one of the most widely used methods for the analysis and visualisation of the results, as it provides high-resolution information about the methylation status of that particular DNA fragment. Sequencing techniques include direct sequencing (Paul & Clark 1996), cloning followed by sequencing (Warnecke *et al.* 2002) and a more recently developed method, pyrosequencing (Tost & Gut 2007). Other methods of data quantification include, melting profile analysis based on the methylation status, as due to the differences in base composition (C/T) methylated bisulfite treated DNA melting temperature differs from the un-methylated (Worm *et al.* 2001). Thus, methylation-sensitive melting curve analysis (MS-MCA) (Worm *et al.* 2001) and methylation-sensitive high- resolution melting (MS-HRM) (Wojdacz & Dobrovic 2007) are the main non-sequencing methods used for the quantification of these differences.

Soon after the discovery of 5mC oxidative derivatives (He *et al.* 2011; Ito *et al.* 2011; Kriaucionis & Heintz 2009; Tahiliani *et al.* 2009), several labs developed bisulfite conversion based methods for the identification of these marks (Wu *et al.* 2016; Yu *et al.* 2012). One of the methods that has been developed is the TET-Assisted Bisulfite Sequencing (TAB-Seq); this technique uses a combination of 5hmC glycosylation, 5mC oxidation to 5caC and bisulfite treatment in order to resolve 5hmC and 5mC at single base resolution (Yu *et al.* 2012). For the detection of 5fC and 5caC the Methylase-Assisted Bisulfite sequencing (MAB-seq) technique has been developed; allowing base-resolution mapping of 5caC and 5fC as unmodified C is enzymatically converted to 5mC allowing direct mapping of 5fC and 5caC however, it cannot distinguish between the two marks (Neri *et al.* 2015; Wu *et al.* 2014). Another method that has also been developed for the direct identification of 5caC at single-base resolution, is the caMAB-seq (5caC methylase-assisted bisulfite sequencing) (Lu *et al.* 2013). In this method 5fC is reduced by $NaBH_4$ to 5hmC prior to M.SssI treatment, 5caC is then sequenced as T after bisulfite conversion whereas C, 5mC 5hmC and 5fC are read as C (Lu *et al.* 2013). A combination of MAB-seq and caMAB-seq enables the direct mapping of 5caC.

Nevertheless, due to the inability of the M.SssI enzyme to methylate cytosines at non CpGs, 5fC and 5caC cannot be distinguished within a non-CpG context. However this limitation of the assay does not seem to greatly affect the application of the technique, firstly because 5hmC is found almost exclusively at CpG context (Lister *et al.* 2013; Yu *et al.* 2012) and secondly because TET proteins have a strong preference for 5hmC within the CpG context (Hashimoto *et al.* 2014b; Hu *et al.* 2013a). While these methods can provide single-based maps for these marks on the DNA, they cannot provide any information about the distribution and the levels of 5hmC, 5caC, 5fC.

Although bisulfite-treatment based techniques have been widely used they still present some major downfalls. Firstly, incomplete denaturation of genomic DNA could lead to incomplete conversion of some cytosine residues to uracil resulting in false-positive quantification of methylation (Genereux *et al.* 2008; Grunau 2001; Rein *et al.* 1997). Moreover, due to the prolonged time at high temperatures and acidic pH in the bisulfite conversion step there is a high-degree of DNA degradation (Tanaka & Okamoto 2007). While this issue can be partially resolved through optimising the time of the reaction and by using larger amounts of DNA; it makes this method unsuitable for clinical applications (Grunau 2001; Raizis *et al.* 1995). More importantly, primer design could also be quite challenging due to the reduced complexity of template DNA (Hackenberg *et al.* 2012) and the same primer pair can preferentially amplify either the methylated or the un-methylated sequence (Luo & Preuss 2003; Warnecke *et al.* 2002).

In addition to bisulfite conversion, restriction enzymes and liquid chromatography dependent techniques; antibody based methods have also been used for the detection of DNA methylation. Methylated DNA immunoprecipitation (meDIP) was developed, permitting highly efficient enrichment of methylated DNA (Weber *et al.* 2005). In this method, a monoclonal antibody against 5mC or any other epigenetic modification is used to immunoprecipitate epigenetically modified genomic regions. This method can be combined with large scale analysis using existing DNA microarrays or using whole genome next generation sequencing (NGS)

(Weber *et al.* 2005). MeDIP provides high resolution results that enable not only the quantification of DNA methylation but also its genomic distribution.

Antibody-based enrichment methods enrich for DNA fragments containing high CpG densities and repetitive regions (Ficz *et al.* 2011; Thomson *et al.* 2013). However, it has been shown that this bias of meDIP towards repetitive regions, including LINE1 elements and simple repeats, is more likely to be an artefact of the method and not a real biological result (Thomson *et al.* 2013). Recent evidence, suggest that oxi-mCs enrichment at simple repeats identified by meDIP is more likely to be due to the non-specific binding of the antibody (Lentini *et al.* 2018). It has been also shown that analysis of the samples against the input control leads to false positive enrichment resulting in inconsistent findings (Lentini *et al.* 2018). While analysis of meDIP results against IgG control reduces the number of enriched regions increasing specificity and sensitivity (Lentini *et al.* 2018). Therefore, choosing the right controls in meDIP experiments is of crucial importance.

Another sensitive method for the estimation of global DNA methylation levels is liquid chromatography coupled with mass spectrometry (LC-MS/MS) (Song *et al.* 2005). The combination of liquid chromatography methods with mass spectrometry led to the identification of oxidised forms of 5mC; 5hmC, 5caC and 5fC (Ito *et al.* 2011; Kriaucionis & Heintz 2009; Tahiliani *et al.* 2009). It has been validated to detect levels of methylation ranging from 0.05%-10% and can detect differences between samples as small as 0.25% (Song *et al.* 2005). More recently, a sensitive Isotope-dilution automated online two dimensional ultra-performance liquid chromatography with tandem mass spectrometry-based technology (2D-UPLC-MS/MS) in combination with isotopically labelled internal standards for the simultaneous detection of 5mC, 5hmC, 5caC, 5fC, and 8-oxoguanine (8-oxo) has been developed (Gackowski *et al.* 2016). This method is suitable for rapid, accurate and complex assessment of an array of endogenously generated DNA modifications providing high sensitivity and selectivity (Gackowski *et al.* 2016). However, LC-MS/MS based

methods requires specific expensive equipment and expertise to run; therefore they are not been widely used.

The development of immunochemical techniques using specific antibodies against DNA methylation marks, led to the robust and rapid analysis of DNA modifications in individual cells of different tissues (Santos & Dean 2005). The incorporation of peroxidase-conjugated secondary antibody and tyramide signal amplification step also permitted the detection of 5fC and 5caC which are present at extremely low levels in the genome (Abakir *et al.* 2016; Wheldon *et al.* 2014). Therefore providing information about the spatial distribution of these marks in different cell types in various developmental stages and conditions.

3.2. Aims

The aim of this chapter was to examine the potential presence of 5hmC/5caC oxidative derivatives of 5mC in adult and paediatric brain tumours cell lines. Previous experiments in our lab using a highly sensitive immunocytochemistry method indicated the presence of 5caC in MB and EPN (paediatric brain tumours) cell lines (Ramsawhook *et al.* 2017); we next aim to examine the presence of this mark in GBM (adult brain tumour) cell lines. Our aim was to examine the spatial distribution, nuclear localisation and genomic distribution of oxi-mCs (5hmC, 5caC) in GBM and MB cell lines. Considering the advantages and the downfalls of each method described as well as the cost and the equipment available; we first used our developed immunocytochemistry protocol for the study of 5hmC/5caC spatial distribution and nuclear localisation. We also used 2D-UPLC-MS/MS for the quantification of these marks and meDIP followed by NGS for the examination of their genomic distribution. Finally using loss-of function experiments we examined the role of TDG and TET2 in the regulation of active DNA demethylation in adult and paediatric brain tumour cell lines.

3.3. Results

3.3.1. GBM cell lines exhibit immunocytochemically detectable levels of 5caC with unique signatures

Our previous studies showed that there is a proportion of glioma tissues derived from individual patients, that exhibit immunohistochemistry detectable levels of 5caC oxidative derivative of 5mC in active DNA demethylation pathway (Eleftheriou *et al.* 2015). Thus, we examined the levels of both 5hmC and 5caC in grade IV gliomas GBM in four different cell lines. We performed immunocytochemistry, as a quick, reliable qualitative method to examine the presence, the distribution and the levels of both 5caC and 5hmC. 4 M HCL was used to denature the DNA in order to optimise the retrieval of the antigen in the nucleus. As a consequence, there is a disruption of base stacking forces at purinic sites and nuclear DAPI cannot be visualised.

Our immunocytochemistry experiments showed a homogeneous presence of 5hmC across the cells in each individual cell line (Fig. 3.1 A) confirming previous research based on 5hmC levels in GBM (Kraus *et al.* 2012). Quantification of the 5hmC signal in the 5caC positive cells showed that U87-MG and U251 cell lines exhibit significantly higher levels of 5hmC compared to LN18 and LN229 cell lines ($p < 0.001$) (Fig. 3.1 B). Interestingly, 5caC was present in all the cell lines examined presenting a more heterogeneous distribution across different cancer cells in the individual cell lines (Fig. 3.1 A). In LN229, LN18 and U87-MG cell lines there is a distinct subpopulation of cells with 5caC positive staining, thus 65% of LN229, 71% of LN18 and 60% of U87-MG stained positive for 5caC (Fig. 3.1 C). In U251 cell line however, 5caC seems to be more homogeneous and present in almost all the cancer cells (85% 5caC positive cells) (Fig. 3.1 C). Remarkably, the intensity of the 5caC staining differed rather extensively between the GBM cell lines. 5caC signal quantification indicates that unlike 5hmC distribution, the highest intensity levels of 5caC are present in LN229 and U87-MG cell lines followed by LN18 and U251 (Fig. 3.1 B).

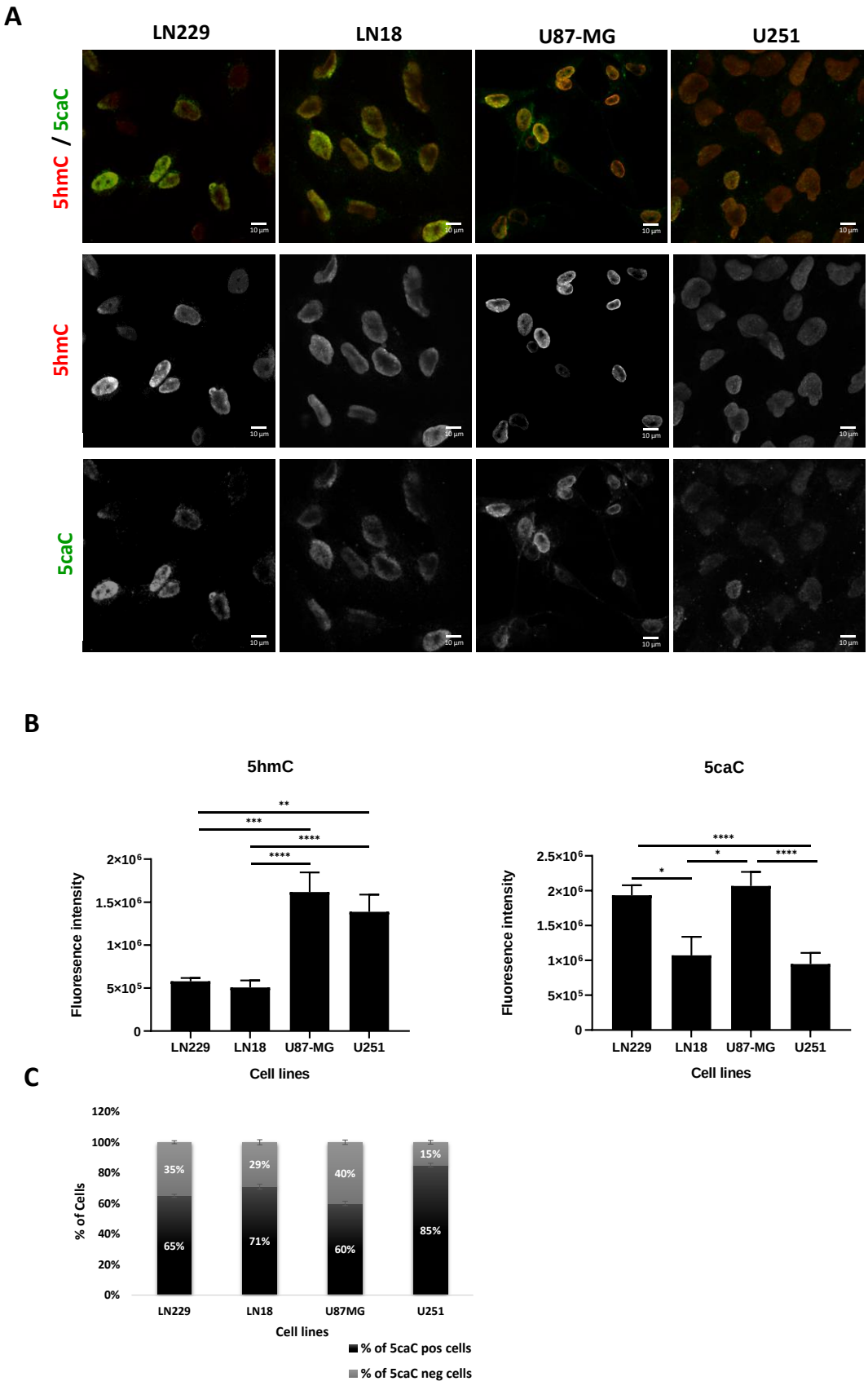

A
LN229
LN18
U87-MG
U251
5hmC / 5caC
5hmC
5caC
B
5hmC
5caC
Fluoresence intensity
Cell lines
LN229
LN18
U87-MG
U251
C
% of Cells
Cell lines
LN229
LN18
U87MG
U251
35%
65%
29%
71%
40%
60%
15%
85%
% of 5caC pos cells
% of 5caC neg cells

Figure 3.1: Glioblastoma cell lines exhibit immunohistochemically detectable levels of 5hmC/5caC.

A. Co-detection of 5caC with 5hmC. Merged views and individual channels are shown. The cell cultures were stained in parallel at the same experimental conditions and were imaged at identical settings. Scale bar = 10 μm. B. Quantification of 5hmC and 5caC signals in the indicated glioblastoma cell lines in the 5caC positive cells. Mean values of the average intensities of each individual cell in each cell line/population are shown in three individual experiments (n = 3). Error bars are shown as ± SEM. Statistical analysis was performed using un-paired t test. $p \leq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ and **** $p < 0.0001$. C. Percentage of 5caC positive cells in each cell line.

Single cell images were also taken from the two populations of cells (5caC positive and 5caC negative) in each cell line and 5caC/5hmC spatial distribution and magnitude profile were plotted (Fig. 3.2). Interestingly, all cell lines displayed different levels of 5caC enrichment. Correlation analysis showed a significant moderate correlation between 5hmC and 5caC in U251 cell line with R = 0.54 for cells high in 5caC and R = 0.56 for cells with low 5caC levels ($p < 0.0001$) indicating a moderate correlation in spatial distribution of 5caC and 5hmC in U251 (Fig. 3.2 D). Remarkably, while LN229 5caC positive subpopulation of cells display a strong correlation between 5caC and 5hmC (R = 0.76 $p < 0.0001$) 5caC negative cells have a distinct spatial distribution indicated by the negative correlation (R = -0.19 $P < 0.0001$) (Fig. 3.2 A). The strongest correlation of 5hmC and 5caC spatial distribution was present in LN18 (Fig. 3.2 B) and U87-MG (Fig. 3.2 C) (R = 0.61, R = 0.86, R = 0.83, R = 0.78 $p < 0.0001$).

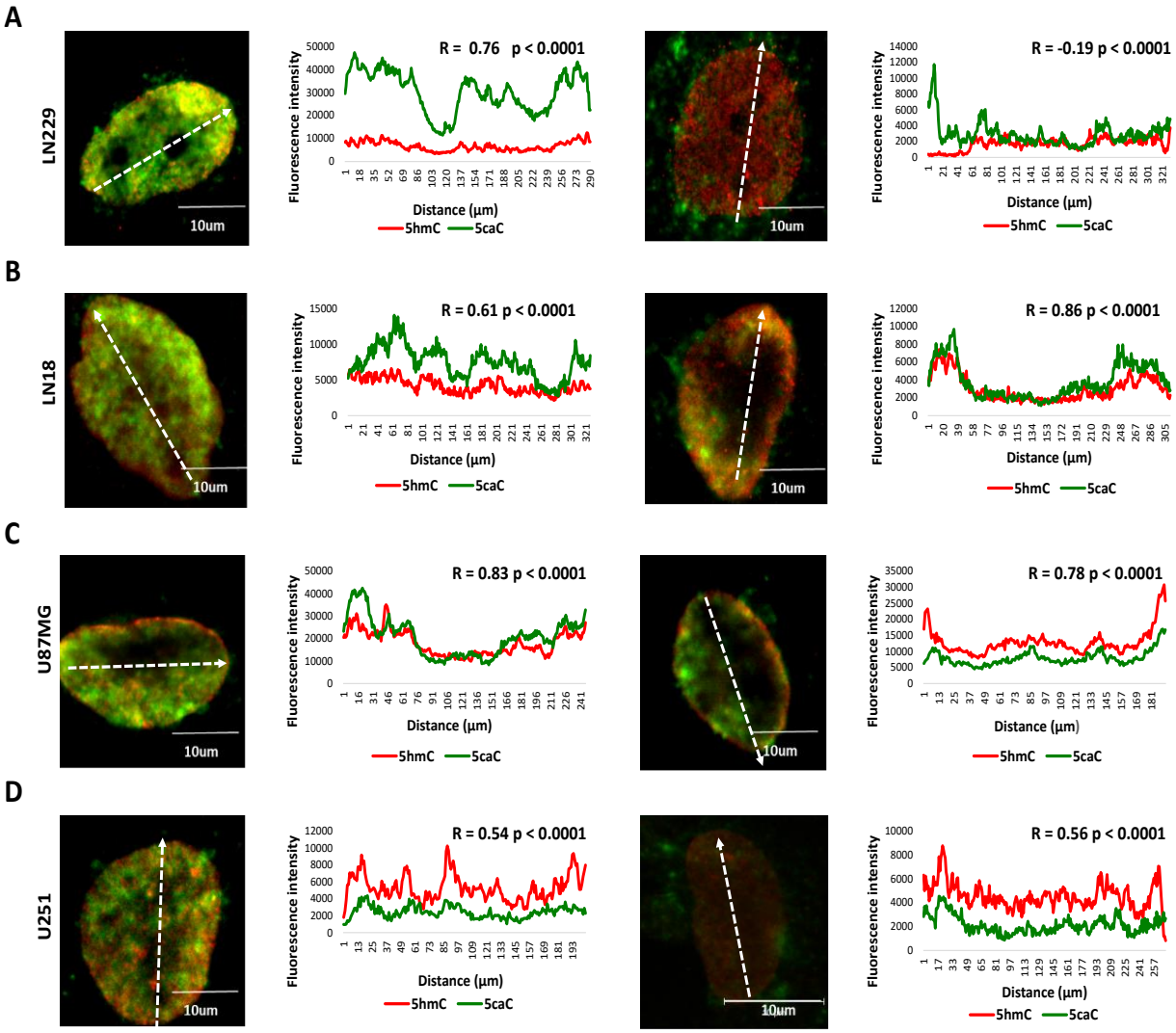


Figure 3.2: Nuclear localisation of 5caC/5hmC in GBM cell lines.

Merged views of the confocal images and graphs of the fluorescence intensity of 5caC/5hmC in representative 5caC positive and negative nuclei of A.LN229 B. LN18 C. U87-MG D. U251 cell lines. Graphs show the signal intensity of 5hmC/5caC across each cells. Scale bar = 10 µm, images were taken at identical settings. Pearson correlation was performed in all the cells of each sample in three individual experiments. Pearson's correlations coefficient test was performed to check the correlation of the spatial distribution of 5hmC and 5caC. Mean R values are displayed in each graph. *Correlation coefficients were transformed in to a new variable called Fisher Z value to be able to calculate an average value and then was transformed back to R value. Statistical analysis was then performed using un-paired t test: $p \leq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ and **** $p < 0.000$.

3.3.2. GBM cell lines exhibit high expression levels of *TET2/3* accompanied with low levels of *TDG*

We next examined the transcript levels of the key genes involved in active DNA demethylation pathway. We first decided to examine the absolute transcript levels of *TET1/2/3* using semi-quantitative PCR with three different cycle numbers in order to better evaluate which TET is the most highly expressed in each cell line (20, 25 and 30 based on our qPCR data) (Fig. 3.3). Our results confirmed that *TET1* is not present (or present at low levels) in U87-MG cell line whereas in LN229, LN18 and U251 *TET1* only appears after 30 cycles of PCR (Fig. 3.3). On the other hand, *TET2* appears at 25 cycles in all the cell lines apart from LN229, whereas *TET3* is present in GBM cell lines but in lower levels compared to *TET2 (*Fig. 3.3).

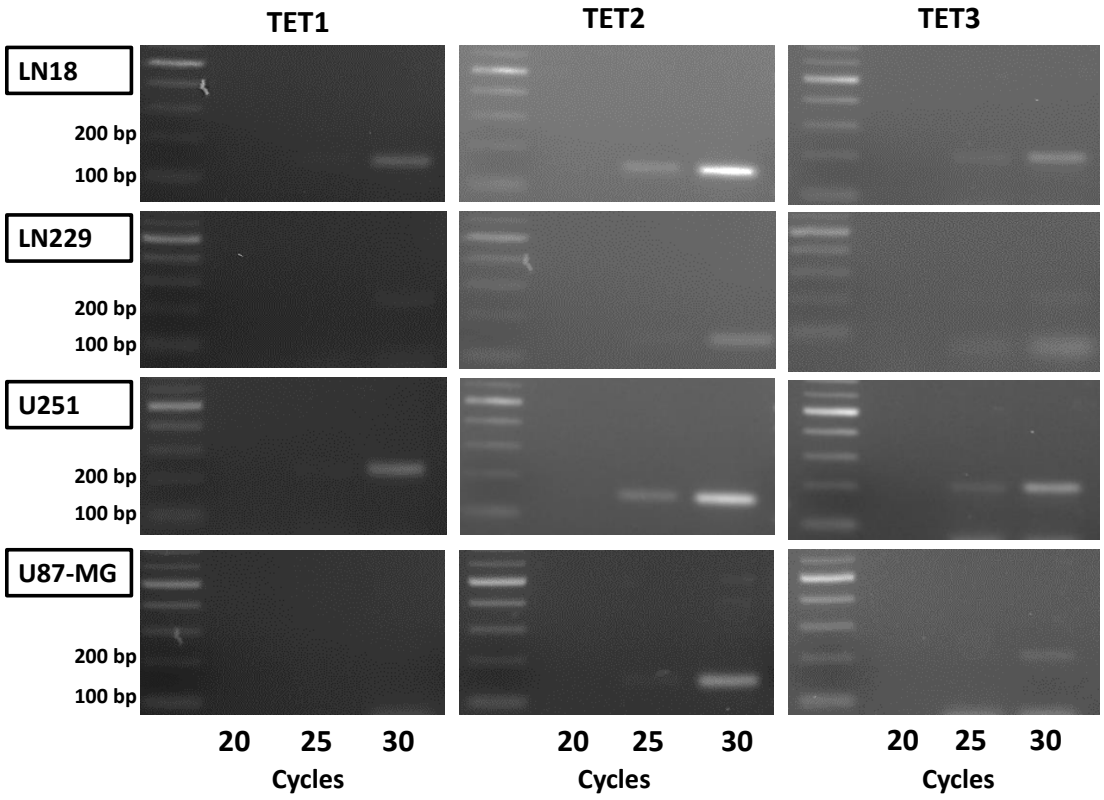


Figure 3.3: Semi-quantitative PCR analysis of TET1/2/3 transcripts in GBM cell lines.

PCR was performed using 20 25 and 30 PCR cycles the same sample was used in each condition. *TET1* at 230 bp, *TET2* at 150 bp and *TET3* at 201 bp.

In order to gain a better understanding on the transcript levels, we next performed real time PCR (qPCR) analysis in *TET1/2/3* and *TDG* in all GBM cell lines. The samples were compared against human induced pluripotent stem cells (hiPSCs) the Rebl-pat cell line which was reprogrammed at University of Nottingham while a human mature astrocytic cell line would be a better control for examining the transcript levels between healthy and cancer conditions; however this was not possible at this time. The Ct values obtained in the experiment were normalised against GAPDH gene, as previously described (Dickson *et al.* 2009), and are displayed at Table 3.1. We observed significantly lower levels of *TET1* transcripts in all GBM cell lines compared to hiPSCs (Fig. 3.4 A) where *TET1* is highly expressed (Bartoccetti *et al.* 2020; Koh *et al.* 2012; Piccolo *et al.* 2013). Specifically, U87-MG cell line displayed 1000-fold ($p < 0.0001$) lower levels of *TET1* compared to hiPSCs while LN18, LN229 and U251 had 100-fold ($p < 0.0001$) lower *TET1* levels (Fig. 3.4 A). While all GBM cell lines exhibit significantly lower levels of *TET1, TET2* and *TET3* transcripts were significantly higher compared to hiPSCs (Fig. 3.4 B and C). Specifically *TET2* displayed 3-4 fold higher levels in all GBM cell lines compare to hiPSCs (Fig. 3.4 B). An even higher increase was observed with *TET3* transcripts which present 5-100-fold higher levels (Fig. 3.4 C) that could be explained due to the fact that TET3 is mainly present at PGS in the formation of the zygote and is repressed at the later stages of development (Wossidlo *et al.* 2011). Finally, *TDG* was significantly down regulated ($p < 0.0001$) in GBM with the lower levels being observed in LN229 (12- fold) and U87-MG (13- fold) cell lines (Fig. 3.4 D). Our results indicate that TET2 is the most highly expressed TET genes in GBM cell lines and might be responsible for the 5caC enrichment observed in our immunocytochemistry experiments.

Table 3.1: Normalised Ct values of *TET1/2/3, TDG* genes in the indicative GBM cell lines.

	TET1	**TET2**	**TET3**	**TDG**
hiPSCs	12.38865	18.26622	29.14796	13.93526
LN18	17.469	16.51667	20.71817	16.09662
LN229	29.41167	15.43767	19.57267	17.23752
U87MG	22.68233	16.122	20.85267	17.25139
U251	18.80967	14.66567	19.31867	16.13021

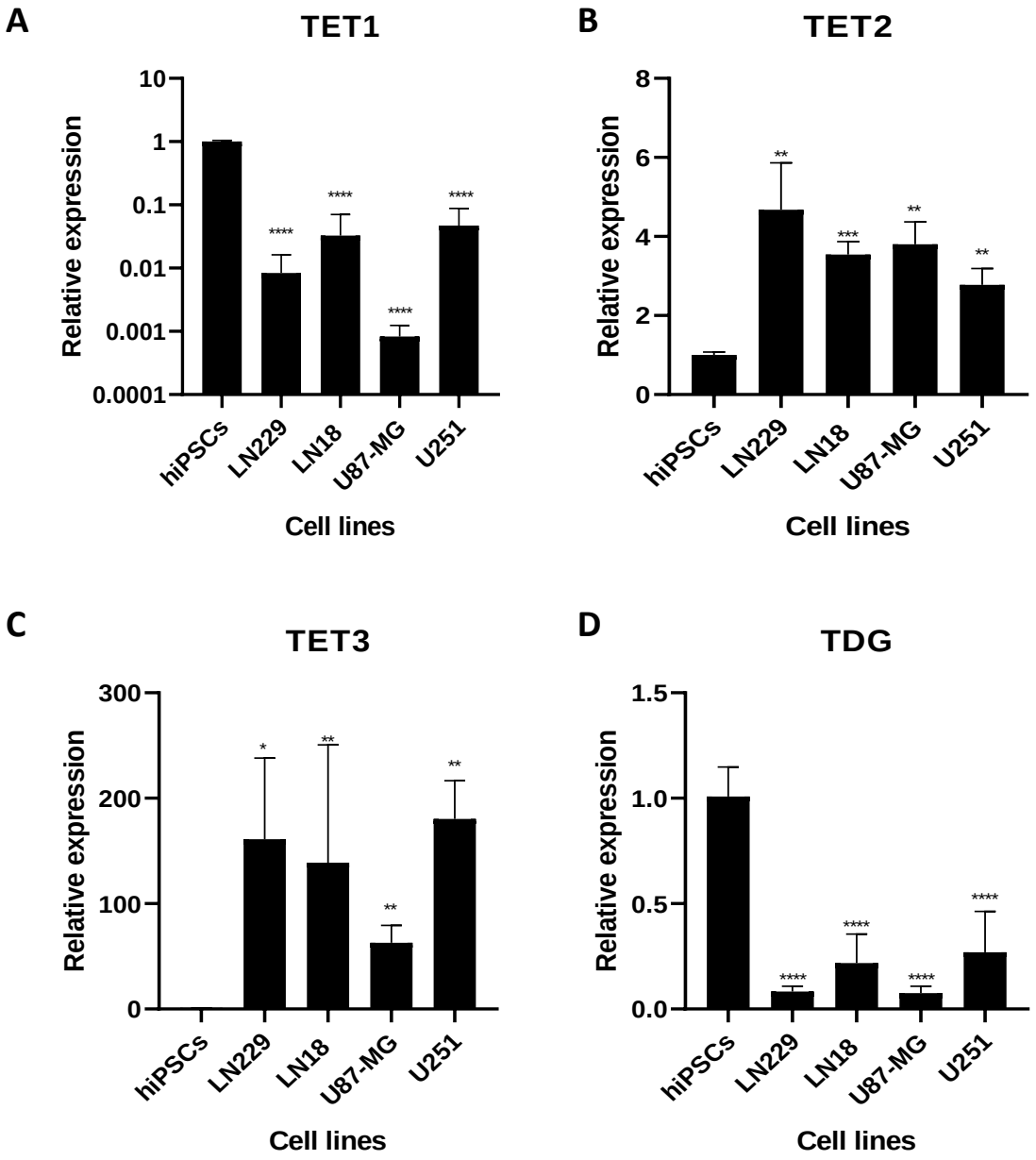


Figure 3.4: *TET2/TET3* levels are elevated at GBM cell lines while *TDG* is present at low levels.

Relative expression of *TET1/2/3 TDG* in LN18 LN228 U87-MG U251 cell lines relative to hiPSCs. Mean values of three individual experiments are shown (n = 3). Experimental error is shown as ± SD Statistical analysis was performed using un-paired t test. $p < 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ and **** $p < 0.0001$.

3.3.3. Primary GBM cell lines display immunocytochemically detectable levels of 5caC, similar to continuous cell lines, in response to high *TET2* transcript levels

We next examined the levels of 5caC/5hmC in primary GBM cell lines in order to confirm that 5caC presence in GBM cell lines is a real biological result and not a consequence of the continuous culture of the cell lines. We used two primary GBM cell lines, isolated from the invasive margin of the tumour, GIN8 and GIN31 (Glioblastoma INvasive) obtained from patients at Nottingham Queens Medical Centre (QMC). GIN8 cell line was isolated from the left frontal of the cerebellum from a 54 years old female patient. GIN31 was isolated from the right temporal of the cerebellum from a 57 years old female patient both of the cell lines are IDH wild type. Our immunocytochemistry experiments confirmed homogenous staining of 5hmC across all the cells in both cell lines (Fig. 3.5 A), similar to the continuous cell lines. Quantification of 5hmC nuclear signal showed that both cell lines display similar levels of 5hmC (Fig. 3.5 B). As expected, primary GBM cell lines present high levels of 5caC in a subpopulation of cells displaying a heterogeneous spatial distribution, while cytoplasmic staining was also observed we reasoned that it was probably a background staining and not real biological effect (Fig. 3.5 A). Interestingly, while in both cell lines the intensity of the 5caC nuclear staining in 5caC positive cells does not display any difference among the cell lines (Fig. 3.5 B); the percentage of 5caC positive cells differs. GIN31 has 78% 5caC positive cells whereas in GIN8 only 37.5% of the cells are positive for 5caC (Fig. 3.5 C). These differences could reflect the differences in the proliferation rate of the cell lines. Moreover, it is also possible that these cell lines would exhibit a population of senescent cells making the direct comparison with continuous cell lines impossible.

Further to that, single cell images were taken from the two populations of cells in both cell lines and spatial distribution and magnitude profile were plotted (Fig. 3.5 D). Correlation analysis showed that in both 5caC positive and negative nuclei, 5hmC and 5caC display a moderate to weak correlation (Fig. 3.5 D). Specifically, in GIN8 the correlation between 5hmC and 5caC is moderate $R = 0.51$ ($p < 0.0001$) for 5caC positive cells and weak $R = 0.3$ ($P = 0.0002$) for 5caC negative cells (Fig. 3.5 D). Similarly in GIN31 the

correlation is moderate R = 0.6 ($p < 0.00001$) for 5caC positive cells and weak R = 0.37 ($p < 0.0001$) for 5caC negative cells (Fig. 3.5 D); indicating that 5hmC and 5caC are present at close proximity regions in the genome.

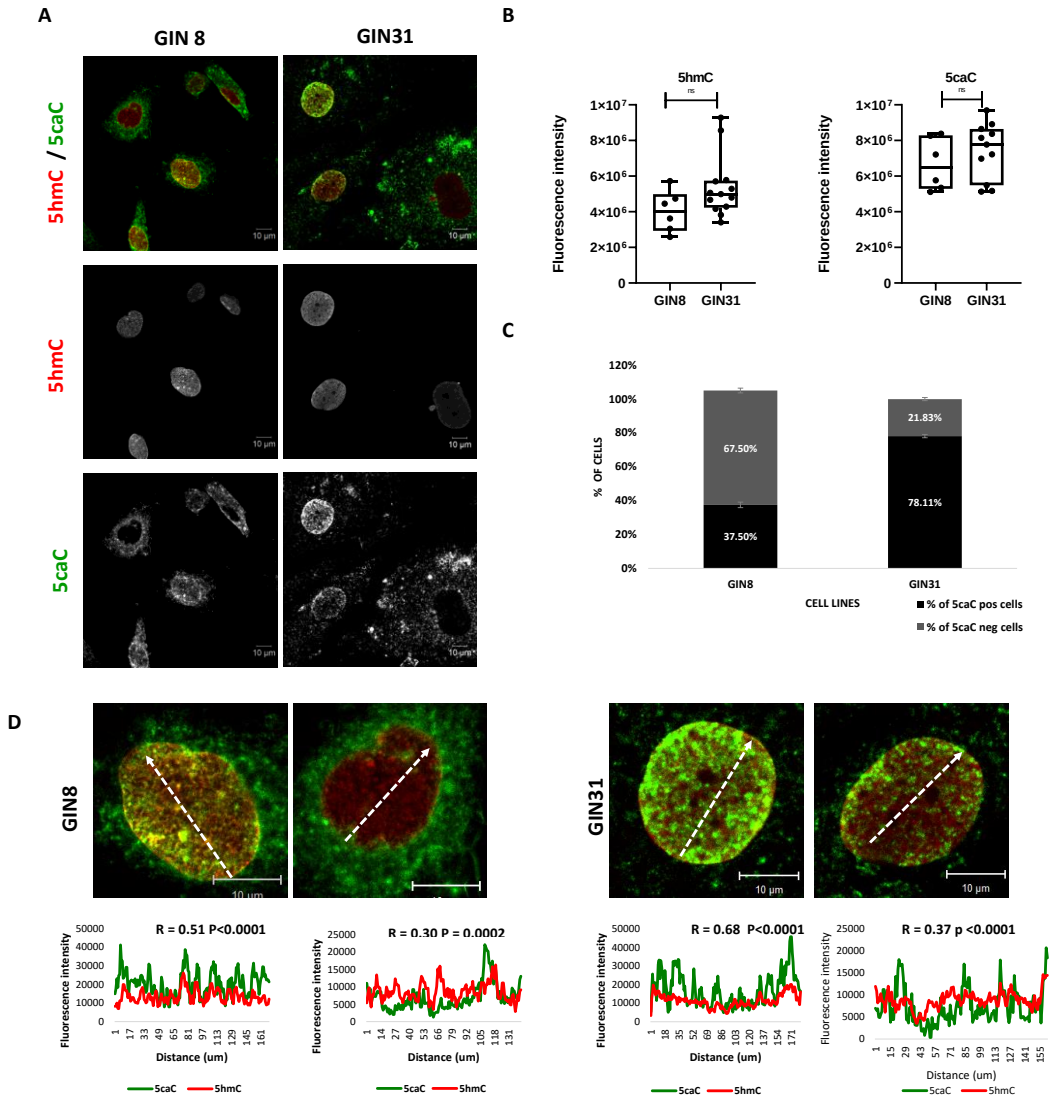


Figure 3.5: Primary GBM cell lines exhibit immunohistochemically detectable levels of 5hmC/5caC similar to continuous GBM cell lines.

A. Co-detection of 5caC with 5hmC. Merged views and individual channels are shown. The cell cultures were stained in parallel at the same experimental conditions and were imaged at identical settings. Scale bar = 10 µm. B. Quantification of 5hmC and 5caC signals in the indicated primary GBM cell lines. Mean values of the average intensities of each individual 5caC positive cell in each cell line/population are shown in three individual experiments (n = 3). Error bars are shown as ± SEM. C. Percentage of 5caC positive cells in each cell line. Error bars are shown as ± SEM. D. Nuclear localisation of 5caC and 5hmC in GBM cell lines. Merged views of the confocal images and graphs of the fluorescence intensity of 5caC and 5hmC in representative 5caC positive and 5caC negative nuclei of GIN8 and GIN31. Graphs shows the signal

intensity of 5hmC/5caC across each cell. Scale bar = 10 μm, images were taken at identical settings. Pearson correlation was performed in all the cells in three individual experiments (n = 3). Statistical analysis was performed using un-paired t test. $p \leq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ and **** $p < 0.0001$.

We next examined the levels of the key genes involved in active DNA demethylation using qPCR. As explained earlier in this chapter, we used hiPSCs cell line as a reference as the only available option we had at the time. The Ct values obtained in the experiment were normalised against GAPDH (Dickson *et al.* 2009)gene and are displayed at Table 3.2. Similar to continuous cell lines, TET1 has 100-fold lower levels in both GBM cell lines examined ($p < 0.0001$) (Fig. 3.6. A). TET2 displayed 4-fold higher levels in GIN8 ($p < 0.0001$) and GIN31 ($p = 0.0075$) compared to hiPSCs (Fig. 3.6 B). Quite surprisingly, and in contrast with the continuous cell lines, TET3 was not detectable in our conditions in these two cell lines. However, that might not be a biological result as at that point we were facing issues with TET3 amplification because of primer dimerization as we were not able to detect TET3 in other samples which are characterised by TET3 high levels. At that point we did not further investigate TET3 expression in our cell lines. TDG levels, similar to the other GBM cell lines were 10-fold lower in both primary cell lines ($p < 0.0001$) (Fig. 3.6 C). Our results point to the same direction as our previous studies in the continuous cell lines; indicating the possible role of TET2 in the active DNA demethylation in GBM most possibly being responsible for the 5caC enrichment.

Table 3.2: Normalised Ct values of *TET1/2, TDG* genes in the indicative primary GBM cell lines.

	TET1	**TET2**	**TDG**
hiPSCs	12.38865	18.26622	13.93526
GIN8	18.65809	13.81023	19.08049
GIN31	18.84488	14.42954	19.83359

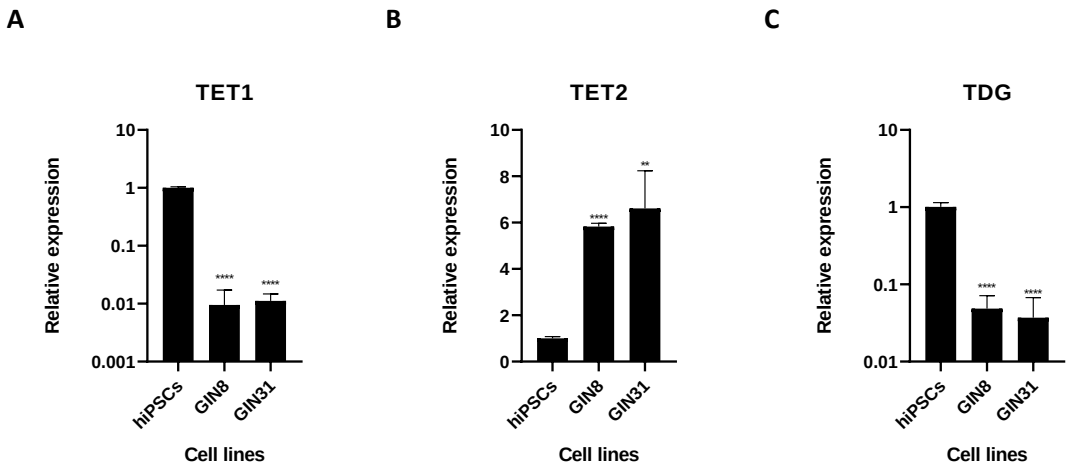


Figure 3.6: Relative expression of the key genes involved in active DNA demethylation in primary GBM cell lines relative to hiPSCs.

A. *TET1*, B. *TET2* C. *TDG*. Average values of three individual experiments are shown (n = 3). Experimental error is shown as ± SD. Statistical analysis was performed using un-paired t test. $p \leq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ and **** $p < 0.0001$.

3.3.4. Glucose concentration does not influence active DNA demethylation in GBM cell lines

It has been previously shown that hyperglycaemia induces demethylation via the activation of TET enzymes. Specifically Dhilwayo et al., used a zebrafish model and provided evidence that high glucose levels induce the activity of TET enzymes resulting is an increase in the intermediates of active DNA demethylation pathway including 5hmC and 5fC (Dhliwayo *et al.* 2014). In addition to that, another study, showed that intraperitoneal injection of glucose in mice results in a rapid increase in blood glucose levels with a corresponding increase in α-KG levels. Therefore suggesting that, an induction of TET enzymes activity results at increased levels of 5hmC and 5fC (Yang *et al.* 2014). Nevertheless, 5caC levels were not measured in either study as they were below the threshold levels.

Given our previous findings that 5caC is present in a subpopulation of GBM cells we next sought to examine whether we could enrich this subpopulation of 5caC positive cells using different culture conditions. Thus, we cultured two GBM cell lines in low and high glucose media and performed qPCR and immunocytochemistry experiments in order to identify any differences in genes involved in active DNA demethylation pathway as well as in 5hmC and 5caC levels. We chose LN18 and U251 as representative cell lines because they are derived from different GBM subtypes with different genetic background and we were interested to examine whether we could find a similar phenotype between the two GBM subtypes. Our results showed that administration of high glucose in the media in LN18 and U251 cells did not impair the morphology of the cells nor alter their behaviour in cell culture (Fig. 3.7 A). The levels of *TET2* transcripts were 10 times lower in high glucose culture condition in U251 compared to low glucose whereas the levels of TET1 and TET3 remain the same in both cell lines examined (Fig. 3.7 B). Further to that levels of TDG were significantly increased (2- fold) in U251 cell line (Fig. 3.7 B).

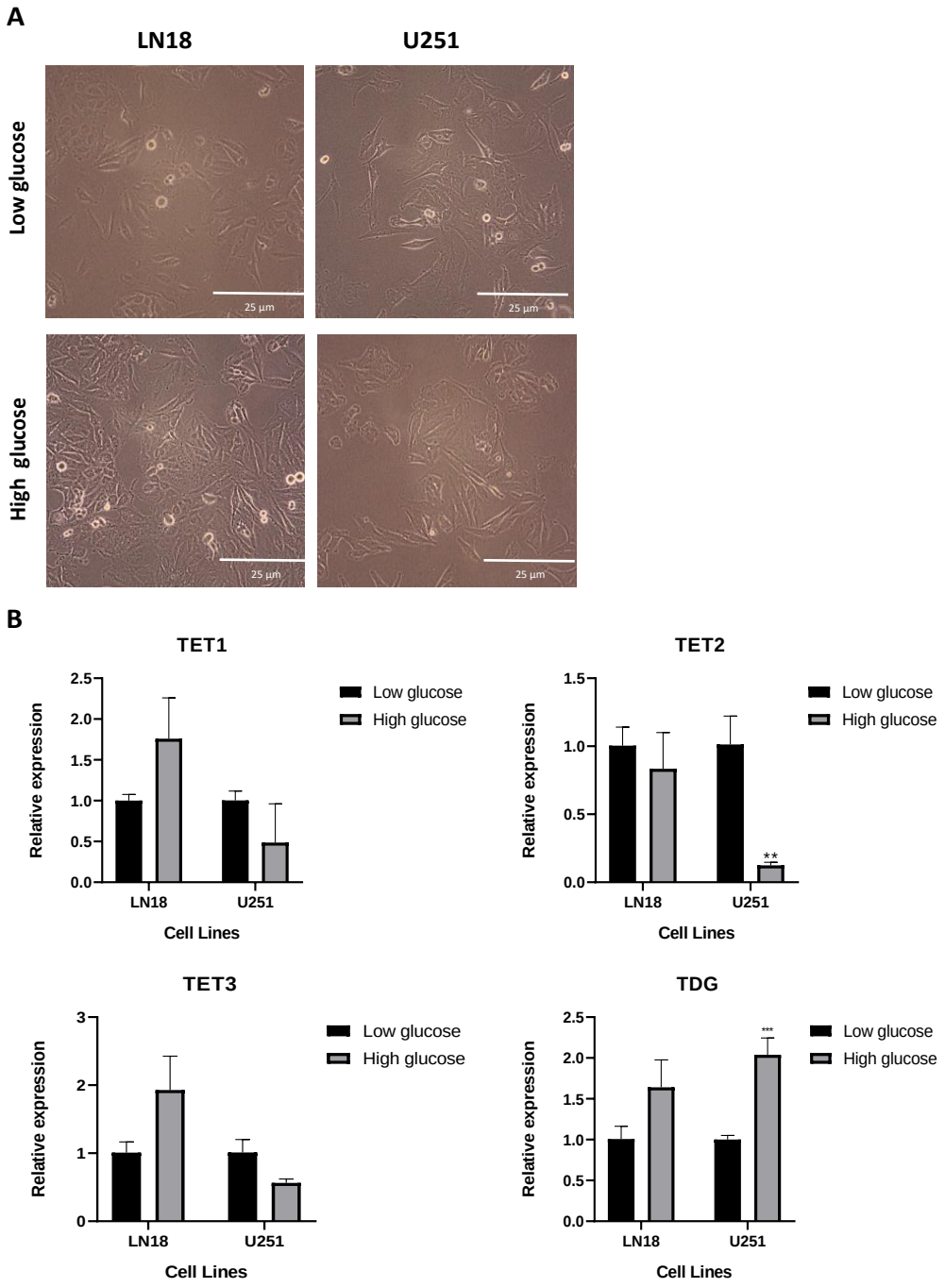


Figure 3.7: GBM cell lines cultured in high and low glucose do not show any significant changes morphologically while U251 shows reduced *TET2* levels.

A. Light microscope images at high and low glucose conditions. Morphology of the cells remain the same under the different culture conditions in LN18 and U251. Scale bar = 25 µm. B. Relative expression of *TET1/2/3* and *TDG* transcripts in low versus high glucose culture conditions in LN18 and U251. Mean values of three individual experiments are shown (n = 3). Experimental error is shown as ± SD. Statistical analysis was performed using un-paired t test. $p \leq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ and **** $p < 0.0001$.

We then examined the levels of 5hmC and 5caC epigenetic modifications using our sensitive immunocytochemistry method followed by image quantification. While, in LN18 cell line the levels of *TET1/2/3* and *TDG* transcripts were unaltered, 5hmC levels are significantly reduced ($p < 0.0001$) and 5caC levels are significantly increased ($p < 0.0001$) (Fig. 3.8 A and B). This observation led us to hypothesise that there might be a preferential direct oxidation of 5mC into 5caC in LN18 cell line in high glucose conditions without a significant change at *TET1/2/3* levels; however further experiments need to be done in order to examine this possibility. Rather unexpectedly and in contrast with qPCR results and the *TET2* statistically significant lower levels in high glucose culture conditions; 5hmC levels were significantly ($p < 0.001$) increased in U251 cell line in high glucose conditions whereas 5caC levels did not show any statistically significance difference between the two different culture conditions (Fig. 3.8 C and D). Thus, this experiment did not serve our primary goal of enrichment of the 5caC positive population of cells, leading to inconclusive results. We therefore, decided to proceed with our experiments using our default glucose conditions.

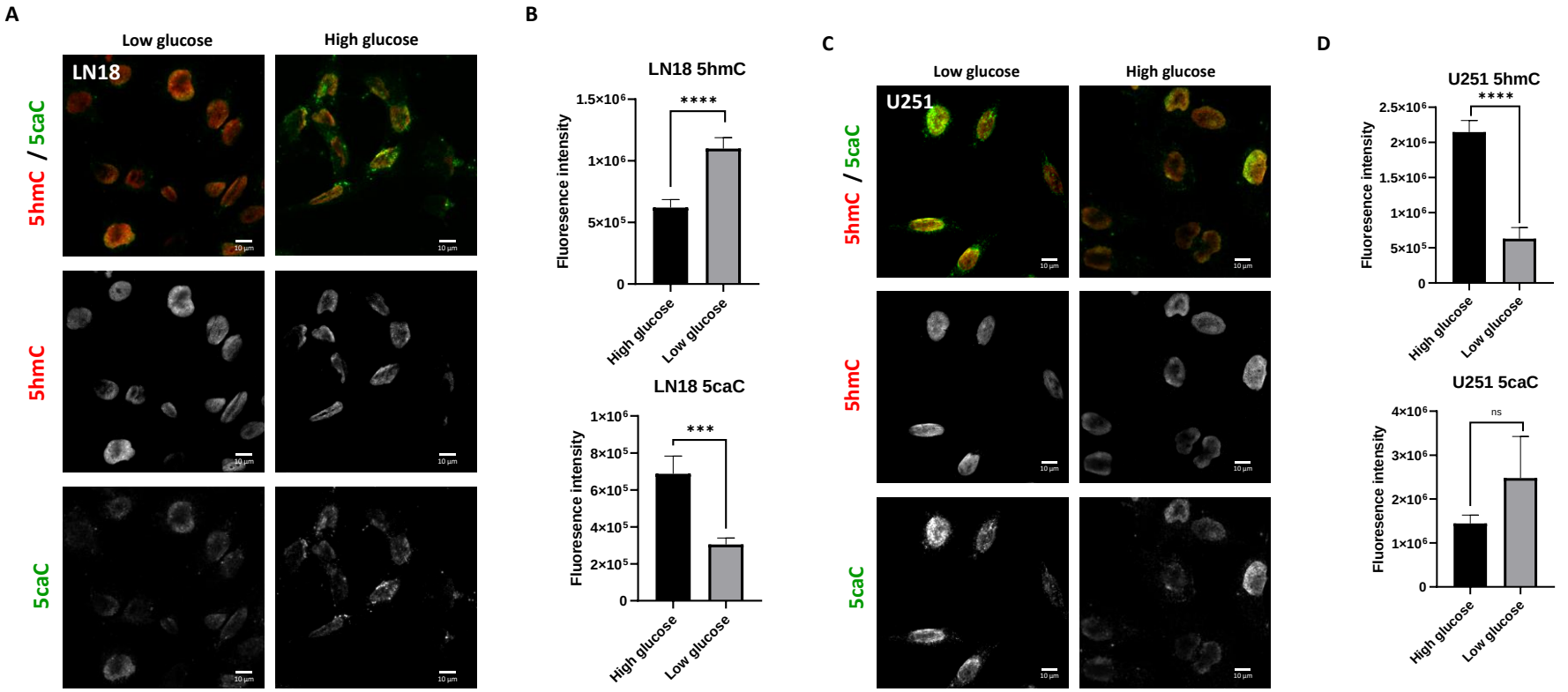


Figure 3.8: Glioblastoma cell lines exhibit immunocytochemically detectable levels of 5caC/5hmC with no specific pattern in high and low glucose conditions.

A. Co-detection of 5caC with 5hmC in LN18 cell line and C. U251 cell line in low and high glucose culture conditions. Merged views and individual channels are shown. The cell cultures were stained in parallel at the same experimental conditions and were imaged at identical settings. Scale bar = 10µm. B. Quantification of 5hmC and 5caC signals in LN18 cell line and D. U251 cell line. Mean values of the average intensities of each individual cell in three individual experiments are shown (n = 3). Error bars are shown as ± SEM Un-paired t test has been performed*:p < 0.05, **: p < 0.01, ***:p < 0.01, ****:p<0.001.

3.3.5. Mass spectrometry analysis reveals that different MB subgroups are characterised by distinct 5hmC levels and elevated levels of 5fC and 5caC

We next sought to study the absolute levels of 5mC oxidative derivatives in MB cell lines using a quantitative method. It has been previously shown that MB cell lines display immunocytochemically detectable levels of 5hmC/5caC (Ramsawhook *et al.* 2017), therefore we sought to expand this screen at MB cell lines derived from different molecular subgroups. We used Isotope-dilution automated online two dimensional ultra-performance liquid chromatography with tandem mass spectrometry-based technology (2D-UPLC-MS/MS) in combination with isotopically labelled internal standards for the simultaneous detection of 5mC, 5hmC, 5caC and 5fC as well as 5-hydroxymethyluracil (5hmU) and 8-oxo-2'-deoxyguanosine (8-oxodG). This method is suitable for rapid accurate and complex assessment of an array of endogenously generated DNA modifications providing high sensitivity and selectivity (Gackowski *et al.* 2016). We first obtained calibration curves on the internal standards identifying limits of detection (LOD), limits of quantification (LOQ) and recovery of the corresponding epigenetic modifications (Fig. 3.9). The ratio for the unlabelled nucleoside to the internal standard (A/A) was plotted against the ratio of their molar concentrations (C/C) following by the calculation of R^2 coefficients (Fig. 3.9). The chemical purity of the samples and internal standards was assessed with UV chromatograms obtained in 1D-UPLC-UV mode (Fig. 3.10).

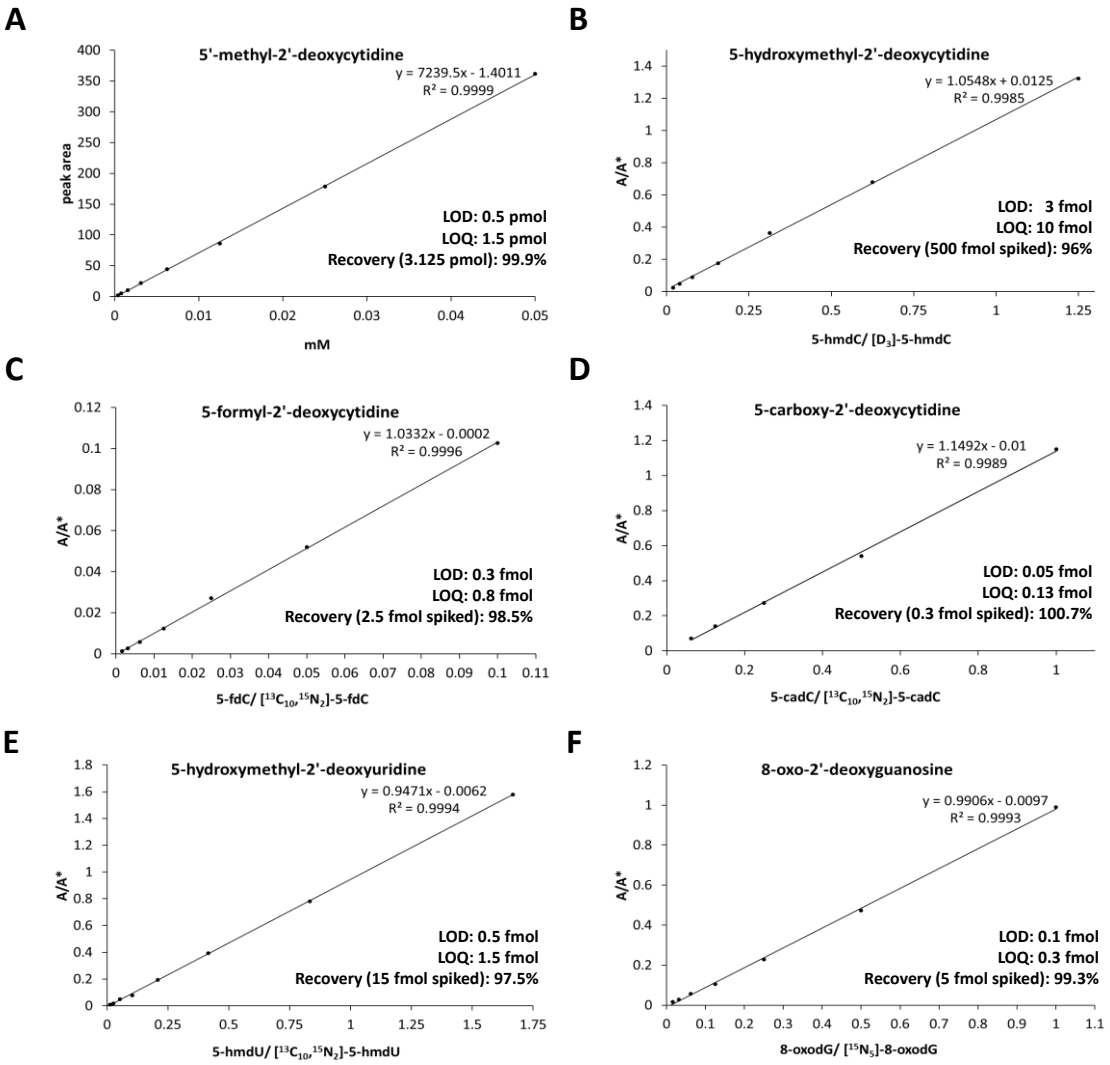


Figure 3.9: Calibration curves, limits of detection (LOD), limits of quantitation (LOQ) and recovery for analysed modified deoxynuclosides.

A. 5-methyl-2'-deoxycytidine B. 5-(hydroxymethyl)-2'-deoxycytidine C. 5-formyl-2'-deoxycytidine D. 5-carboxy-2'-deoxycytidine. E. 5-(hydroxymethyl)-2'-deoxyuridine and F. 8-oxo-2'-deoxyguanosine.

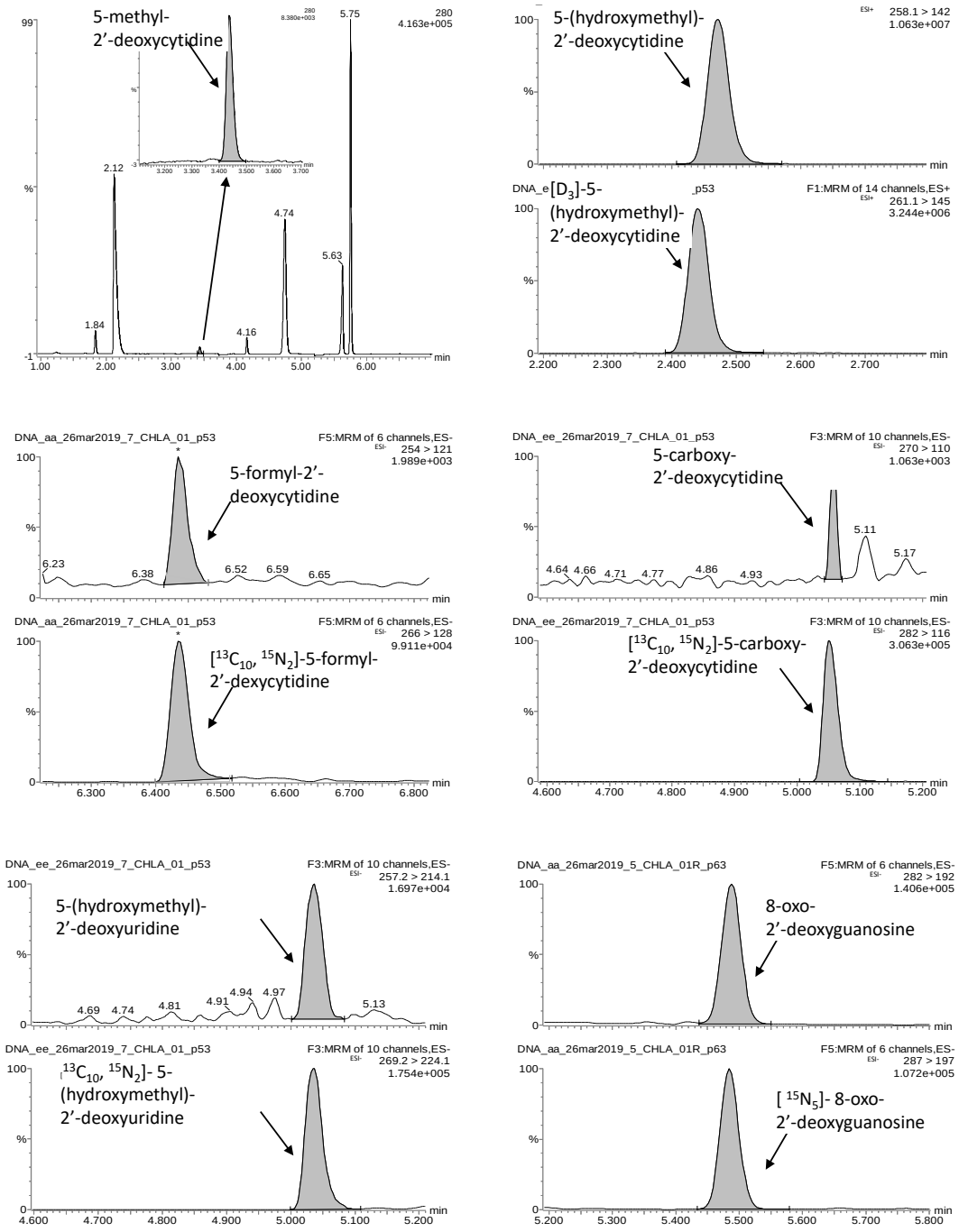


Figure 3.10: Representative traces of internal standards and corresponding modifications in a representative CHLA-01-MED MB cell line.

Representative UV (280nm, 5-methyl-2'-deoxycytidine, and MRM traces (5-(hydroxymethyl)-2'-deoxycytidine , 5-formyl-2'-deoxycytidine, 5-carboxy-2'-deoxycytidine, 5-(hydroxymethyl)-2'-deoxyuridine, and 8-oxo-2'-deoxyguanosine with corresponding internal standards) for CHLA-01-MED MB cell line.

We next wanted to examine whether there are any differences in the levels of 5hmC and 5caC in different molecular subgroups of MB. Initially, we

assessed the levels of cytosine modifications in eight MB cell lines: CHLA-01R-MED and CHLA-01-MED derived from group 4 tumours; UW228-3, DAOY and ONS-76 belonging to SHH subgroup; as well as D458, D283 and HD-MB03 cell lines classified as group 3 MBs (Xu *et al.* 2015). Although the levels of 5mC displayed variability across the tested MB cell lines they did not correlate with molecular subgroups of MB (Fig. 3.11 A). Interestingly, we found out that distinct levels of 5hmC correlated with SHH, group 3 and group 4 MB subgroups (Fig. 3.11 B). Specifically, group 4 MBs (CHLA-01R-MED, CHLA-01-MED) displayed the highest levels of 5hmC (0.03 modifications per 1000 DNA bases), cell lines belonging to SHH subgroup (UW228-3, DAOY and ONS-76) were characterised by the lowest 5hmC content (around 0.005 modified bases per 1000 DNA nucleotides) and group 3 cell lines (D458, D283 and HD-MB03) exhibited intermediate levels of this modification (0.01 modified nucleotides per 100 DNA bases) (Fig. 3.11 B).

Interestingly, 5fC and 5caC were also detectable in the majority of the tested primary MB cell lines showing no specific correspondence to different molecular subgroups of MB (Fig. 3.11 C and D). Thus, UW228-3 and DAOY cell lines belonging to SHH subgroup displayed the highest levels of 5caC (around 60 modifications and approximately 50 modifications per 10^9 DNA bases correspondingly), however 5caC was not detectable in another SHH cell line, ONS-76 (Fig. 3.11 D). Surprisingly, although CHLA-01 possessed detectable levels of 5caC, this modification was not detectable in the corresponding recurrent metastatic cell line CHLA-01R (Fig. 3.11 D). Moreover, SHH cell lines (DAOY and UW228) exhibited relatively low levels of 5fC paralleled by relatively high 5caC content (Fig. 3.11 C and D) indicating that 5mC is likely to be preferentially converted to 5caC in these tumours.

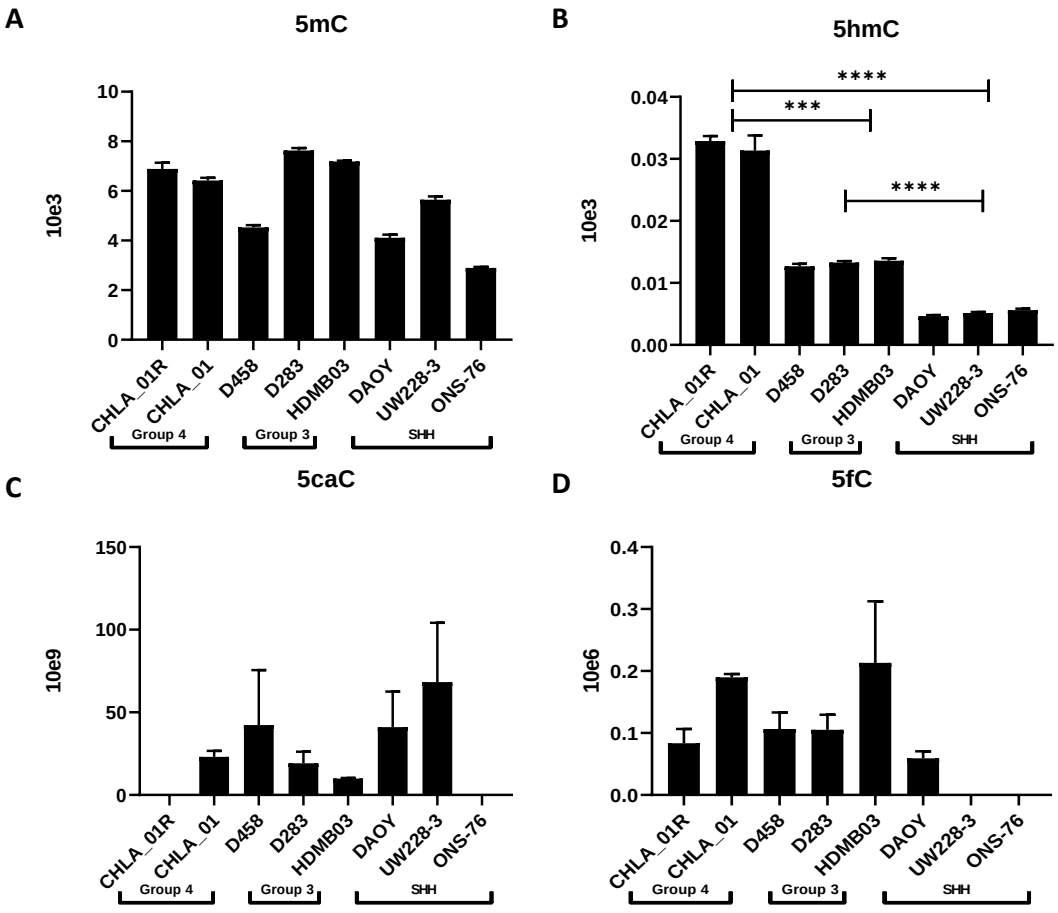


Figure 3.11: Different molecular subgroups of MB are characterised by distinct levels of 5-hmC.

The results of 2D-UPLC–MS/MS quantification of the indicated forms of modified cytosine in the cells lines derived from different MB molecular subgroups. A. 5mC levels are comparable across the cell lines. B. 5hmC distinct levels at different subgroups of MB. C 5caC and D. are detectable in the majority of the cell lines tested. Graphs show mean values of three individual experiments (n = 3). Experimental error is shown as ± SD. Statistical analysis was performed using un-paired t test. $p \leq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ and **** $p < 0.0001$.

Next, we assessed the levels of 5-hydroxymethyluracil (5hmU) (Fig. 3.12 A), a DNA modification with currently unknown function, previously implicated with DNA damage (Olinski *et al.* 2016; Pfaffeneder *et al.* 2014), and a product of DNA oxidation, 8-oxo-2'-deoxyguanosine (8-oxodG) (De Souza-Pinto *et al.* 2001) (Fig. 3.12 B) in the same cell lines. Unlike the species of modified cytosine, both these modifications displayed comparable levels in all the tested MBs suggesting that the levels of oxidative DNA damage were similar in all our specimens and the differences in 5hmC/5fC/5caC we observed were unlikely to be due to non-enzymatic oxidation events that may have taken place during preparation and processing of the samples (Fig. 3.12).

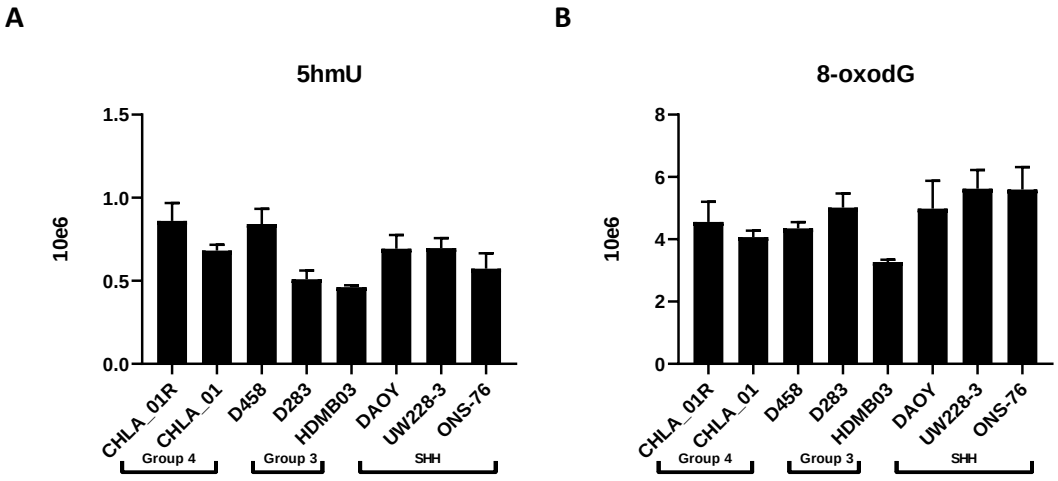


Figure 3.12: The levels of oxidative DNA damage are comparable across the tested MB cell lines.

The results of 2D-UPLC-MS/MS quantification of A. and B. 8-oxodG in the indicated MB cell lines. Graphs show mean values of three individual experiments (n = 3). Experimental error is shown as ± SD.

3.3.6. Mass spectrometry confirms the presence of 5caC in the DNA of brain cancer cell lines.

We then examined the levels of 5mC, 5hmC, 5caC and 5fC in the GBM cell lines (LN18 and U251), the MB cell line (UW228-3), the human colorectal carcinoma cell line (HCT-116) and human induced pluripotent stem cell line (hiPSCs) the latter has been previously shown to have high levels of 5mC oxidative derivatives. We observed lower levels of 5mC in all the brain cancer cell lines examined (6 modifications per 10^3 DNA bases) compared to hiPSCs (9 modifications per 10^3 DNA bases) (Fig. 3.13 A). 5hmC levels were drastically decreased in all the cancer cell lines examined (Fig. 3.13 B). hiPSCs displayed 01 5hmC modifications per 10^3 DNA bases while 5hmC levels were 10 fold lower than cancer cell lines examined, confirming that reduced 5hmC levels is a characteristic of cancer tissues (Ficz & Gribben 2014; Jin *et al.* 2011). Moreover, all cancer cell lines examined, presented mass spectrometry detectable levels of 5caC and 5fC modifications (Fig. 3.13 C and D). In summary, our mass spectrometry experiments confirmed our previous findings regarding 5caC enrichment in both GBM and MB cell lines (Ramsawhook *et al.* 2017).

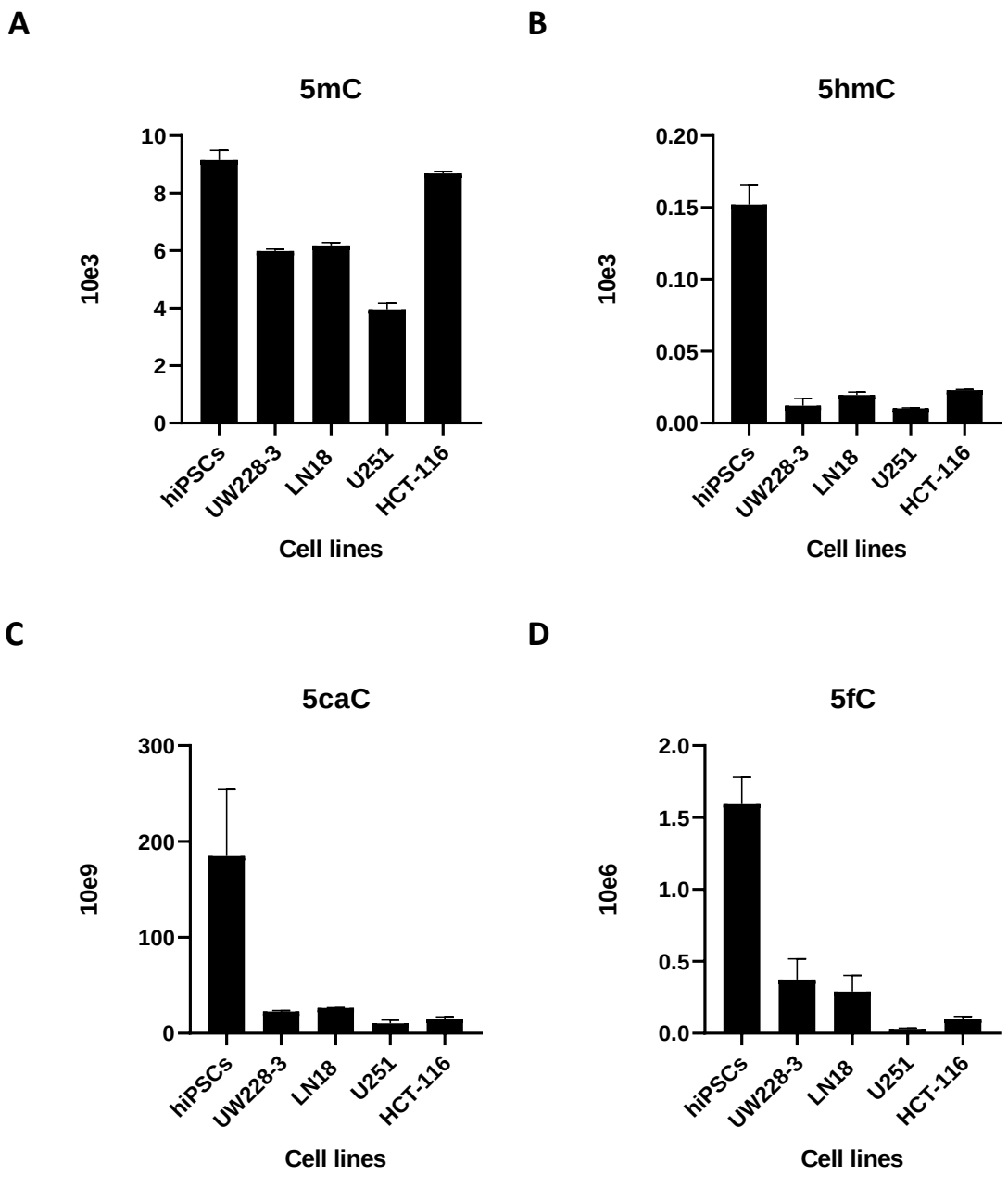


Figure 3.13: Mass spectrometry determined levels of DNA methylation marks in cancer cell lines.

The results of 2D-UPLC–MS/MS quantification of the indicated forms of modified cytosine in the cells lines derived from different cell lines hiPSCs, UW228 (MB), LN18 and U251 (GBM) HCT-116 (colorectal carcinoma). A. 5mC levels are significantly lower in cancer cell lines compared to hiPSCss. B. 5hmC levels are reduced in cancer cell lines. C. 5caC and D. 5fC are detectable in MB and GBM cell lines tested. Mean values of three individual biological replicates for hiPSCs, UW228-3 and LN18 and three technical replicates for U251 and HCT-116 are shown (n = 3).Experimental error is shown as ± SD.

3.3.7. 5caC DNA immunoprecipitation followed by NGS reveals 5caC unique genomic distribution in adult and paediatric brain tumours

We next examined the genomic distribution of 5caC using DNA immunoprecipitation (DIP) followed by NGS analysis. We used LN18 and U251 GBM and UW228-3 MB cell lines. We used two biological replicates of each sample and an input control, the summary of the quality control report is displayed in Figure 3.14. We also, used previously generated data derived from hiPSCs cell line for the analysis and the visualisation of our results.

	Reads	Map%	Filt%	Dup%	ReadL	FragL	RelCC	SSD	RiP%
LN18_5caC_1	31783507	100	12.1	0	40	196	2.95	1.15	4.81
LN18_5caC_2	52808138	100	10.8	0	64	129	2.66	1.12	3
Input	50450609	100	10	0	68	159	1.7	0.787	NA

	Reads	Map%	Filt%	Dup%	ReadL	FragL	RelCC	SSD	RiP%
UW228-3_5caC_1	37784768	100	13.2	0	40	168	2.98	1.51	6.31
UW228-3_5caC_2	44885866	100	12.9	0	46	124	2.15	1.54	5.25
Input	77182706	100	9.69	0	69	150	2.26	1.03	2.56

	Reads	Map%	Filt%	Dup%	ReadL	FragL	RelCC	SSD	RiP%
U251_5caC_1	45830100	100	8.98	0	73	174	3	0.811	2.37
U251_5caC_2	44486925	100	9.48	0	71	148	2.06	0.852	2.46
Input	16966827	100	12.1	0	71	144	1.8	1.85	NA

Figure 3.14: NGS quality control report

Number of mapped reads generated in each sample and biological replicate after NGS are shown.

We first mapped the reads to the human genome and identified 5caC DIP peaks; the consensus peaks were then identified for each sample by comparing the corresponding replicates (Fig. 3.15 A and B). 5caC peaks were normalised against the input sample; IgG control failed to amplify during library preparation therefore it was not possible to normalise against it. We next analysed the 5caC genomic enrichment throughout genomic features in our GBM and MB samples along with hiPSC's. As expected the 5caC called out

peaks were reduced in brain tumours compared to hiPSCs (Fig. 3.16). More specifically, while in hiPSCs 5caC peaks are at the range of 10000 in UW228-3 they are in the 1000 range while LN18 and U251 cell lines the peaks are only at 100 (Fig. 3.16). These results confirmed our previous observations based on immunocytochemistry and mass spectrometry experiments indicating that 5caC is only present in a sub-population of cells in brain tumours and we were unable to enrich for that subpopulation of cells. Interestingly, hiPSCs and UW228-3 MB cell lines displayed virtually similar genomic distribution across various genomic features, repetitive sequences, regulatory features and gene biotypes (Fig. 3.16). In repetitive sequences, 5caC enrichment is observed at LINE, Sine/Alu, Long terminal repeats (LTR) in both UW228-3 and hiPSCs; whereas in GBM cell lines enrichment was mainly observed in satellites and LTR sequences (Fig. 3.16 A). Within gene features, greatest enrichment was observed in transcripts in hiPSCs and UW228-3 followed by intragenic regions, introns and 3kb up/down stream transcription (Fig. 3.16 B). In contrast, in both GBM cell lines examined, highest enrichment was observed in intragenic regions followed by transcripts, fewer peaks related to other gene features were identified (Fig. 3.16 B). Amongst regulatory features, 5caC enrichment in U251 and LN18 was predominantly found in CTCF and transcription factor- (TF) binding sites (Fig. 3.16 C). While within UW228-3 and hiPSCs, genomic distribution was spread among CTCF-binding sites, open chromatin, promoter and TF-binding sites (Fig. 3.16 C). Finally, within gene biotypes, 5caC peaks in GBM cell lines were exclusively found in protein-coding regions whereas their distribution in UW228-3 and hiPSCs was higher in protein coding regions followed by lincRNAs and antisense RNA (Fig. 3.16 D).

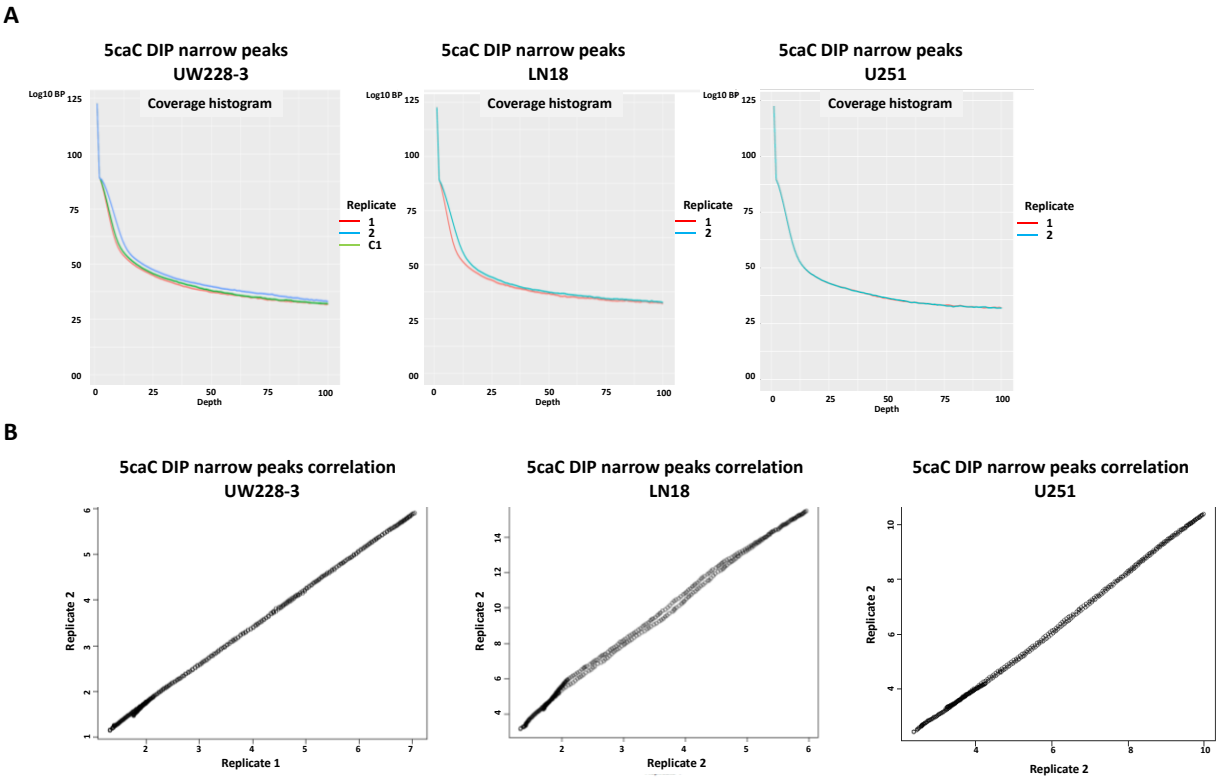


Figure 3.15: 5caC DIP replicates' correlation analysis.

A. Replicates' coverage histograms for the narrow peaks of the indicated samples. B. Peaks correlation histogram for the indicated samples.

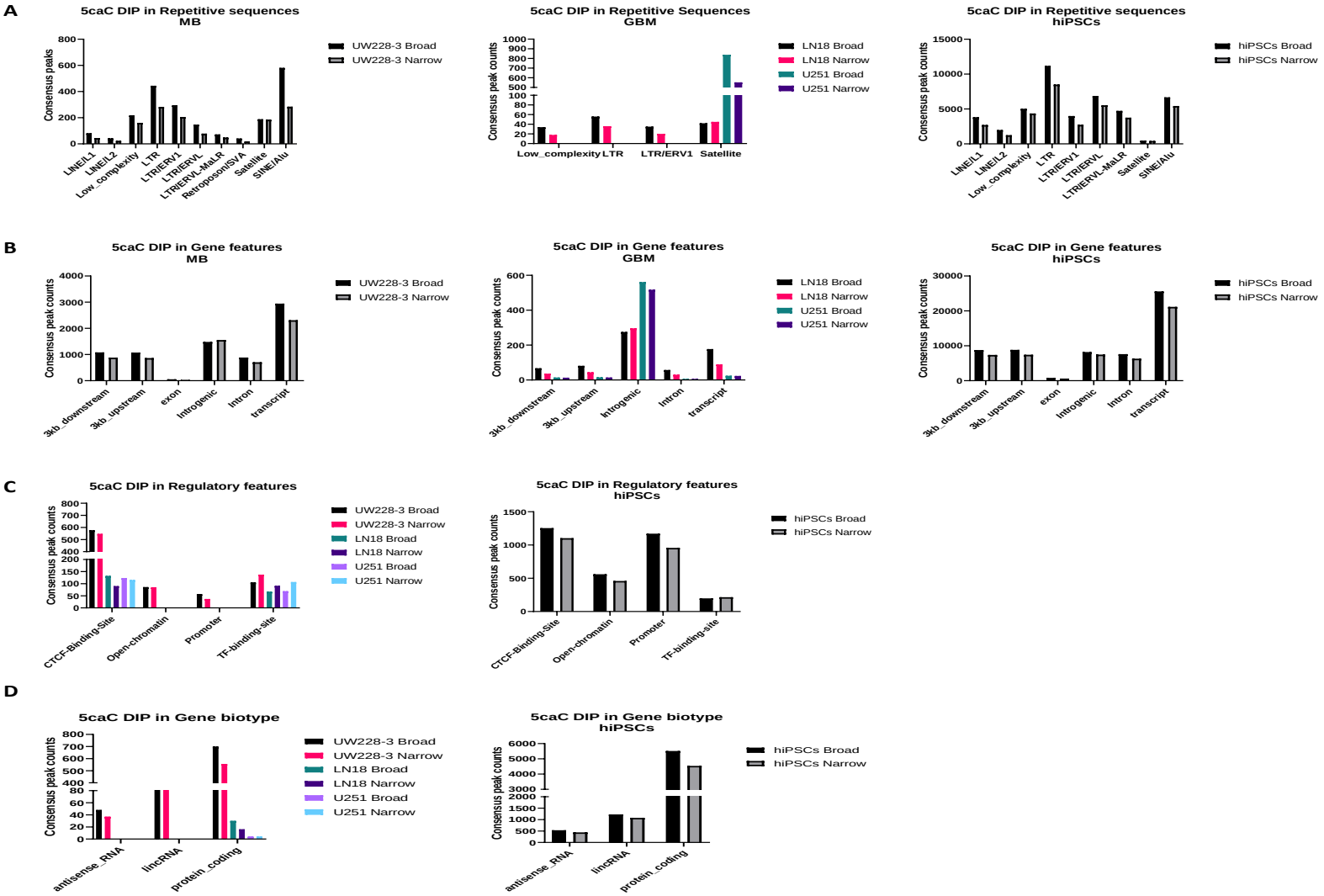


Figure 3.16: Genomic distribution of 5caC peaks in UW228-3 MB, U251 LN18 GBM and hiPSCs cell lines.

Distribution of the 5caC consensus peaks across the indicated gene categories. A. Repetitive sequences B. Gene features C. Regulatory features and D. Gene biotype.

Moreover, analysis of the distribution of the peaks around TSS showed that most of the peaks in UW228-3 and LN18 cell line are at around 10-100 kb distance in either sites of TSS with UW228-3 presenting many more peaks than LN18 (Fig. 3.17). Nevertheless due to the extremely low levels of 5caC in our examined cell lines, further validation of our results should be performed using independent single based resolution analysis like caMAB-seq.

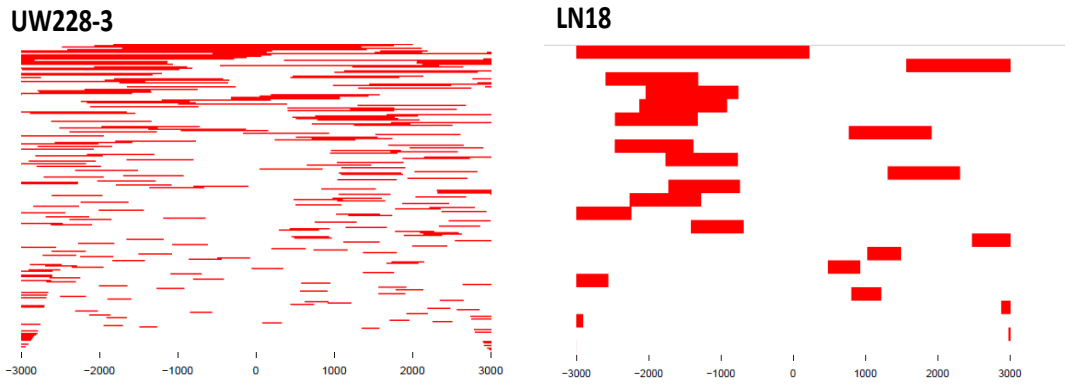


Figure 3.17: Genomic distribution of 5caC peaks relative to TSS.

A. The graphs illustrate the distribution of 5caC DIP reads relative to TSS in UW228-3 and LN18 cell line. Each line displays individual read.

We next decided to closely examine the 5caC peaks at genomic regions that are related with protein coding genes (3kb up/downstream) in order to have a better understanding of 5caC potential role in the regulation of transcription. Thus, we exported the gene lists of 5caC peaks related to protein coding genes; U251 cell line was excluded from our analysis due to the limited number of peaks we observed in that regions. We performed Gene Ontology (GO) analysis using the online platform "Panther" (http://www.pantherdb.org/geneListAnalysis.do) in hiPSCs, UW228-3 and LN18 cell lines. According to the biological process, the majority of the genes in all cell lines were related to cellular, biological and metabolic activity; however more similarities were displayed between hiPSCs and UW228-3 (Fig. 3.18 A). Regarding the cellular component, the place in the cell that the gene is active, the majority of genes identified in all cell lines

are present in cell, organelle and membrane (Fig. 3.18 B). Finally, regarding molecular functions and the biochemical activity of those genes they are mainly responsible for binding and catalytic activity in the three cell lines analysed (Fig. 3.18 C).

We then compared the UW228-3 5caC protein coding peaks with hiPSCs 5caC peaks; we chose UW228-3 as the cancer cell line with the highest 5caC enrichment. We found that only 14.8% of the genes with 5caC peaks were unique for UW228-3 whereas the rest of the genes overlap with the hiPSCs (Fig. 3.19 A); indicating that most likely 5caC is present at almost same genomic loci in both hiPSCs and UW228-3 presenting much higher abundance in hiPSCs. Interestingly, among the genes that displayed a unique peak at UW228-3 was *PAX5* (Fig. 3.19 B) which has been previously associated with MB development (Czapiewski *et al.* 2016; Kozmik *et al.* 1995).

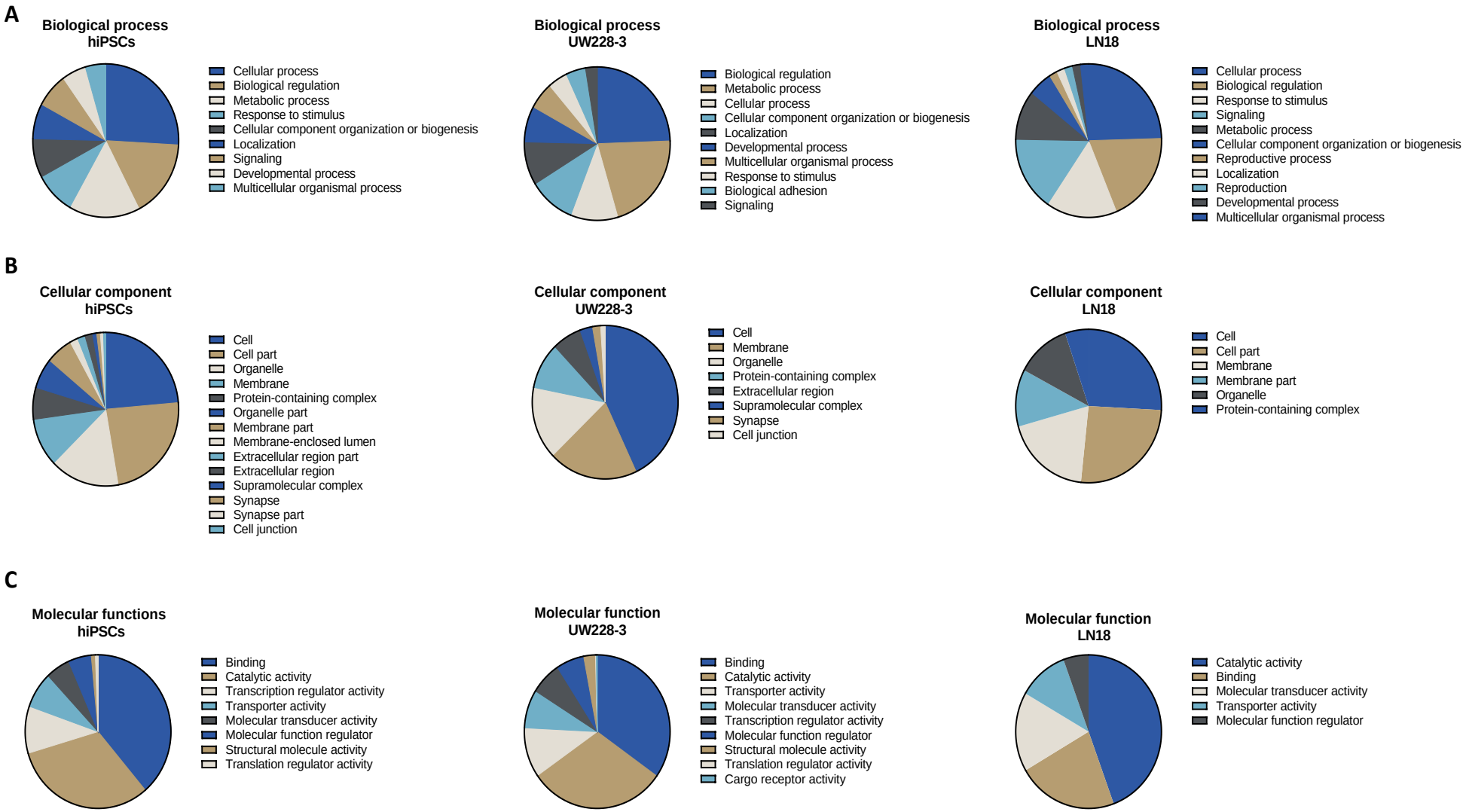


Figure 3.18: Gene ontology (GO) analysis of the protein coding related 5caC peaks (3 kb up/downstream of each protein coding gene) at the indicated cell lines.

Individual pie charts for each sample have been created for A. Biological process B. Cellular component C. Cellular functions.

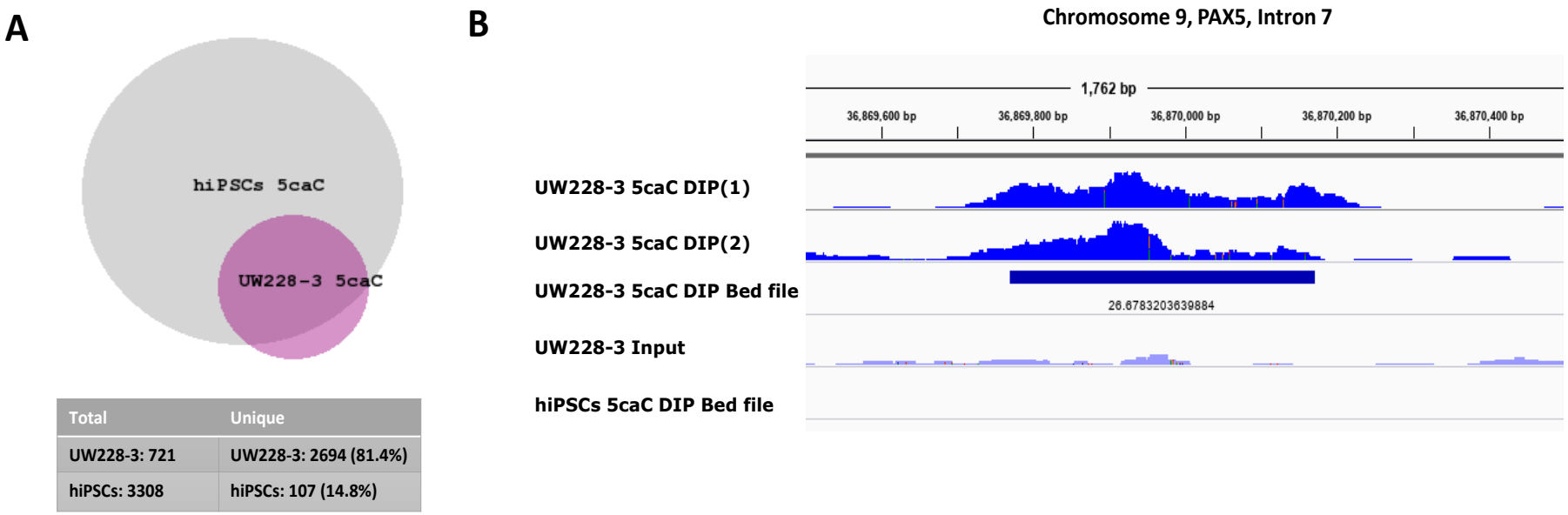


Figure 3.19: 5caC peak distribution within genes in hiPSCs and UW228-3 cell lines.

A. A venn-diagram for the protein coding 5caC peaks identified in hiPSCs and UW228-3; 14.8 % of the genes identified in UW228-3 were unique to this cell line. B. The coverage plots of 5caC DIP at intron 7 of PAX5 gene at the indicated samples.

3.3.8. *TDG* siRNA mediated knockdown leads to elevated 5hmC/5caC levels in both GBM and MB cell lines

Previous studies in ES cells showed that TDG is responsible for the recognition and excision of 5caC from the DNA; as a result BER repair pathway is then activated to replace the abasic site with the unmodified C (He *et al.* 2011). Therefore, we next examined the effects of *TDG* knockdown in our system in adult and paediatric brain tumours. We performed siRNA mediated knockdown of *TDG* in UW228-3 MB cell line and in U251 GBM cell line. We chose U251 GBM cell line as it has the highest levels of *TDG* among the GBM cell lines examined.

Cells were chemically transfected with the TDG siRNA. We first optimised the concentration of the TDG siRNA and the transfection reagent in order to obtain the highest efficiency without causing cell death to the culture; optimisation was performed in one cell line and we used the optimised conditions (50 nM of siRNA and 3 µl Dharmafect) for both cell lines (Fig. 3.20 A and B). We then performed siRNA mediated *TDG* knockdown in UW228-3 MB cell line and in U251 cell line achieving 72% and 80% efficiency respectively (Fig. 3.20 C).

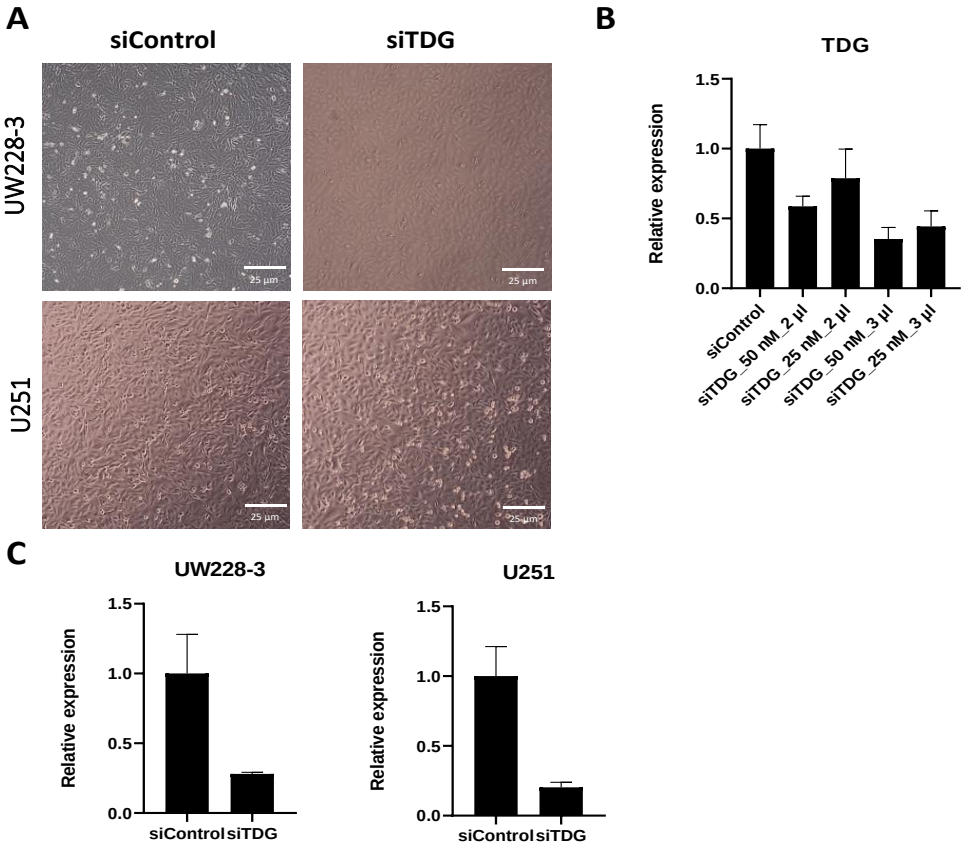


Figure 3.20: Optimisation of the siRNA mediated knockdown of TDG in GBM and MB cell lines.

A. Light microscope images for siControl and siTDG cells no cell death nor changes in morphology is observed. Scale bar = 25 µm. B. Relative expression of TDG in 2 different concentrations of the siRNA and the transfection reagent; 4 different conditions in total. 50 nM of siRNA and 3 µl of Dharmafect lead to 65 % knockdown of TDG. B. Relative expression of TDG transcripts at the siRNA treated sample compare with the control; 72 % knockdown of TDG in UW228-3 MB cell line and 80 % knockdown of TDG in U251 GBM cell line. Error bar is shown as ±SD (n = 3).

The levels of *TET1/2/3* genes were then examined using qPCR, observing no significant changes in the TDG knockdown samples compared with the control in both cell lines (Fig. 3.21 C). Our immunocytochemistry experiment showed that both 5hmC and 5caC levels were significantly elevated upon *TDG* depletion in both cell lines (Fig. 3.21 A and B). These results were also confirmed by 2D-UPLC–MS/MS quantification of 5hmC 5caC and 5fC whose absolute levels were elevated after the TDG depletion (Fig. 3.21 B). Our results further confirm the role of TDG in active DNA demethylation in both GBM and MB.

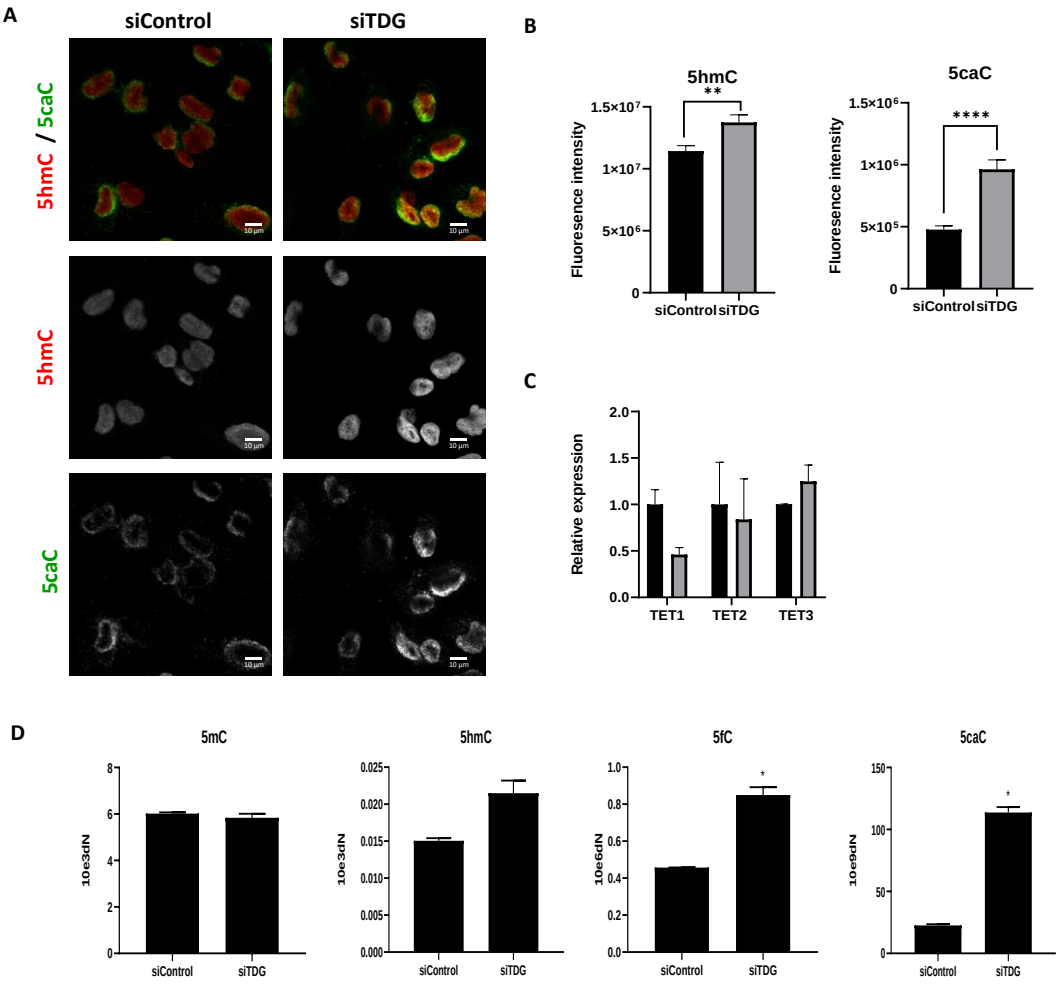


Figure 3.21: siRNA mediated depletion of TDG in GBM and MB cell lines leads at elevated levels of both 5hmC and 5caC.

A. Co-detection of 5caC/5hmC in U251 GBM cell line in siRNA treated cells and in control. Merged views and individual channels are shown. The cell cultures were stained in parallel at the same experimental conditions and were imaged at identical settings. Scale bar = 10 μm. B. Quantification of 5hmC and 5caC signals in the siRNA treated cells and in control. Mean values of the average intensities of all the individual cells in three different experiments for each cell condition are shown. Error bars are shown as ± SEM (n = 3). Un-paired t test was performed *:p < 0.05, **: p < 0.01, ***:p < 0.01, ****:p < 0.001. C. Relative expression of *TET1/2/3* transcripts after the depletion of TDG; no significant changes at the levels of TETs were found in the siRNA treated samples. Mean values of three individual samples are shown(n = 3) . Experimental error is shown as ± SD. D. 2D-UPLC–MS/MS quantification of the indicated forms of modified cytosine in siRNA mediated TDG knockdown; 5fC and 5caC levels were significantly elevated. Mean values of two biological replicates are shown (three technical replicates in each sample) (n = 2). Experimental error is shown as ± SD.

3.3.9. TET2 protein levels are elevated in GBM cell lines and *TET2* siRNA mediated knockdown leads to diminished levels of 5hmC and 5caC in GBM but not in MB.

Our previous experiments showed that the *TET2* gene is highly expressed in GBM cell lines, thus we examined the protein levels of TET2 in order to confirm its abudance in GBM as well as in MB cell lines. We first optimised TET2 and β-tubulin (used as a reference protein) antibodies using different dilutions of the antibody and different protein extracts concentrations (Fig. 3.22).

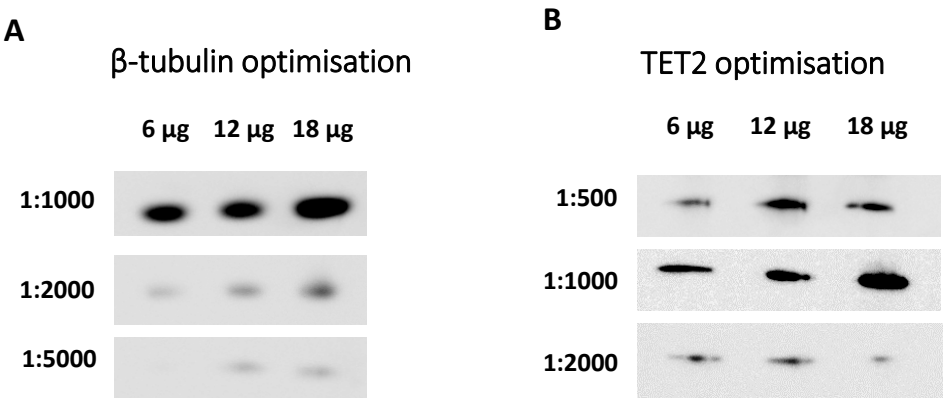


Figure 3.22: β-tubulin and TET2 antibodies optimisation for western blot analysis.

A. β-tubulin B. TET2

Once we optimised all the conditions we performed western blot experiment in LN18, LN229, U251, U87-MG GBM cell lines and in UW228-3 MB cell line. The results were normalised against β-tubulin (Fig. 3.23). Our preliminary western blot experiment shows that TET2 protein levels are elevated in LN18, LN229 GBM and UW228-3 MB cell lines.

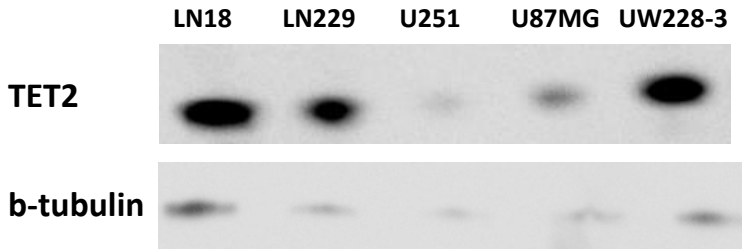


Figure 3.23: Preliminary data on TET2 protein levels in GBM and MB cell lines.

TET2 levels were normalised against β-tubulin (n = 1).

We next performed siRNA mediated knockdown of *TET2* in both GBM and MB cell lines in order to examine our hypothesis that TET2 high levels might be responsible for the oxidation of 5mC to 5hmC and 5caC in GBM and MB. Similarly to TDG knockdown, we chemically transfected LN18 GBM and UW228-3 MB cell lines and optimised for the best concentrations; to achieve the highest efficiency (90 % and 80 % efficiency respectively) with the lowest cell death (Fig. 3.24).

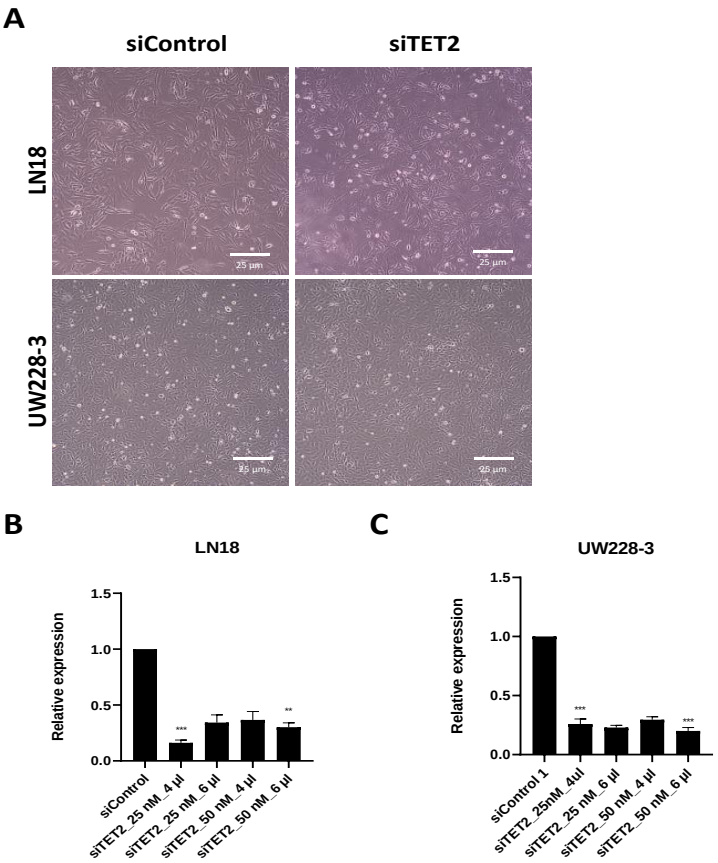


Figure 3.24: Optimisation of siRNA mediated knockdown of TET2.

A. Light microscope images for siControl and siTDG cells in UW228-3 MB cell line and LN18 GBM cell line, no cell death or changes in morphology is observed. Scale bar = 25 µm. B, C. Relative expression of TET2 in 2 different concentrations of the siRNA and the transfection reagent; 4 different conditions in total. B. 25 nM of siRNA and 4 µl of Dharmafect lead to 90 % knockdown of TET2 in LN18. C. 50 nM of siRNA and 6 µl of Dharmafect lead to 80 % knockdown of TET2 in UW228-3. Error bars shown as ±SD (n=3). Paired t test was performed *:$p<0.05$, **: $p<0.01$, ***:$p<0.01$, ****:$p<0.001$.

We then performed immunocytochemistry to examine the levels of 5hmC/5caC after the siRNA mediated depletion of TET2. Rather unexpectedly, 5hmC and 5caC levels were almost completely diminished in LN18 cell line in the siRNA treated samples (Fig. 3.25 A and B). While in control cells 5caC was present in a subset of nuclei at relatively high levels, in siRNA treated cells 5caC levels were almost not detectable (Fig. 3.25 A and B). Signal quantification of the images taken confirmed our observations; 5hmC levels were almost 50 % reduced in siTET2 transfected cells ($p < 0.001$) (Fig. 3.25 C). More drastic effects were observed in 5caC where its levels were reduced

by almost 75 % ($p < 0.0001$) (Fig. 3.25 C). In order to examine whether our findings were solely due to the TET2 knockdown we performed qPCR experiment examining the levels of the other *TETs* as well as *TDG*. Our results showed that there was no significant changes in the levels of *TET1* and *TDG* (Fig. 4.27 A); indicating that TET2 knockdown did not impair their expression.

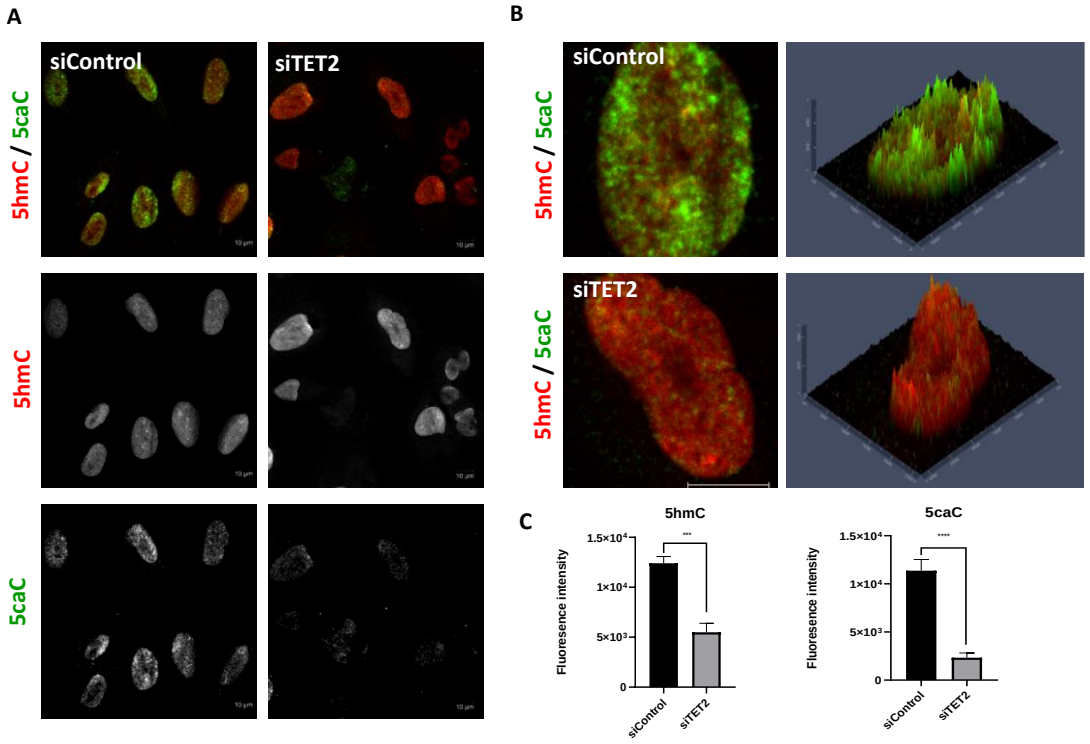


Figure 3.25: siRNA mediated knockdown of TET2 in LN18 GBM cell line leads to diminished levels of both 5hmC and 5caC.

A. Co-detection of 5caC with 5hmC in LN18 control cells and in TET2 siRNA mediated knockdown cells. Merged views and individual channels are shown. B. Single cell images and 2.5D signal intensity plots of 5caC/5hmC staining in representative nuclei in siControl and siTET2 treated cells. The cell cultures were stained in parallel at the same experimental conditions and were imaged at identical settings. Scale bar = 10 µm. C. Quantification of 5hmC and 5caC signals. Mean values of the average intensities of each individual cell for each cell population at three individual experiments are shown (n = 3). Error bars are shown as ± SEM. Paired t tests of three independent experiments have been performed. *:$p < 0.05$, **: $p < 0.01$, ***:$p < 0.01$, ****:$p < 0.001$.

Previous studies from our lab showed that elevated *TET1* levels are related to the accumulation of 5caC in MB and EPN cell lines (Ramsawhook *et al.* 2017). Thus, we wanted to test whether TET2 knockdown would result in altered 5hmC /5caC levels in UW228-3 MB cell line. In contrast to GBM cell line, in UW228-3 TET2 siRNA mediated knockdown did not cause any significant alterations in 5caC/5hmC levels nor their spatial distribution (Fig. 3.26 A and B). Signal quantification of the images taken confirm our observations (Fig. 3.26 C). Similarly to LN18, the levels of *TET1/3* and *TDG* were not changed in the TET2 knockdown samples (Fig. 3.27 B). Our observations lead to the hypothesis that TET2 might be crucial for the establishment of 5hmC and 5caC epigenetic modifications in GBM but not in MB.

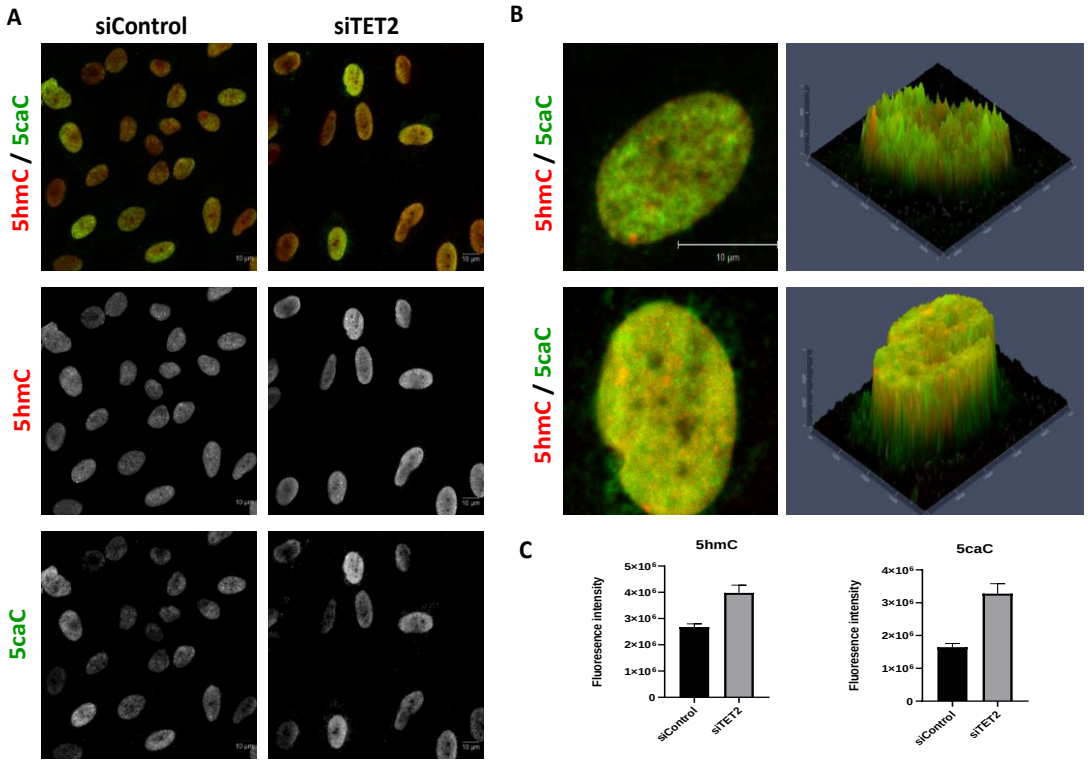


Figure 3.26: siRNA mediated knockdown of TET2 in UW228-3 MB cell line does not alter the levels of 5hmC.

A. Co-detection of 5caC with 5hmC in UW228-3 control cells and in TET2 siRNA mediated knockdown cells. Merged views and individual channels are shown. B. Single cell images and 2.5D signal intensity plots of 5caC/5hmC staining in representative nuclei in siControl and siTET2 treated cells. The cell cultures were

stained in parallel at the same experimental conditions and were imaged at identical settings. Scale bar = 10 µm. C. Quantification of 5hmC and 5caC signals. Mean values of the average intensities of each individual cell for each cell population in three individual experiments are shown (n = 3). Error bars are shown as ± SEM. Un-paired t tests have been performed. *:p < 0.05, **: p < 0.01, ***:p < 0.01, ****:p < 0.001.

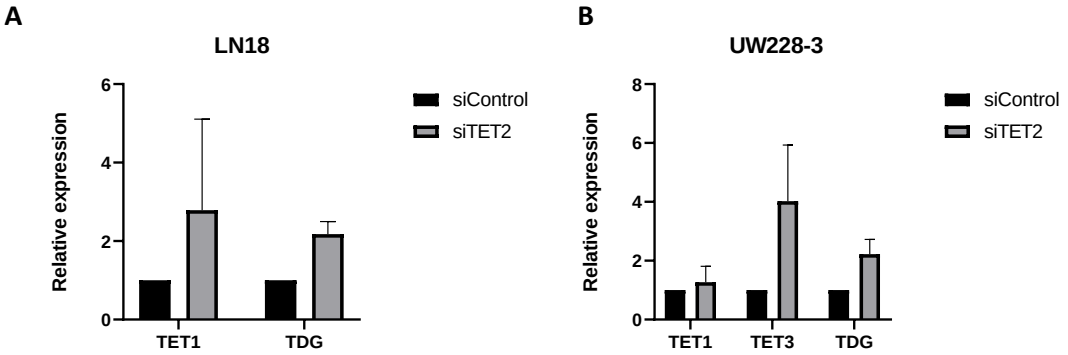


Figure 3.27: Expression of *TET1/3, TDG* in TET2 siRNA mediated knockdown.

Relative expression of *TET1/3, TDG* in siControl compared with siTET2 in A. LN18 GBM cell line B. UW228-3 MB cell line.Mean values of three individual experiments are shown (n = 3). Experimental error is shown as ± SD.

3.4. Technical discussion

The aim of this chapter was to identify and study the presence of 5mC oxidative derivatives in adult and paediatric brain tumours. In summary, we used immunocytochemistry, 2D-UPLC-MS/MS and meDIP followed by NGS for the qualitative and quantitative analysis of 5caC in GBM and MB. Our results reveal several novel findings: (1) GBM primary and continuous cell lines exhibit immunocytochemically and mass-spectrometry detectable levels of 5caC accompanied with high levels of *TET2* and low levels of *TDG* transcripts; (2) Different subgroups of MB are characterised by distinct 5hmC levels and elevated levels of 5fC/5caC; (3) We confirmed that TDG is responsible for the removal of these marks and; (4) TET2 is responsible for the oxidation of 5mC into 5hmC and 5caC in GBM but not in MB.

The methods that have been widely used for the detection of these epigenetic modifications have been described earlier in this chapter. We chose to use our sensitive immunocytochemistry protocol for the initial examination of 5caC presence in GBM cell lines. This method allows us to study the spatial distribution of the marks as well as the heterogeneity that might exist in each sample. This is the first study to show the presence of 5caC in GBM sub-population of cells. Specifically, we show a homogeneous 5hmC nuclei staining at low levels in all the cells; however 5caC distribution is heterogeneous in the nuclei and is not present in all the cells (Fig. 3.1). We performed co-immunostaining of 5hmC and 5caC in order to ensure that HCl-mediated antigen retrieval in the nuclei has been successful, thus the heterogeneity we observed in 5caC staining is not due to technical issues of the method used but a real biological event. Nevertheless, the method used does not allow the direct comparison of the levels of 5hmC and 5caC as they only give a semi-quantitative assessment of proportions of the signal in one cell relative to another. Unfortunately, the tyromide amplification step we used for the detection of 5caC permits a non-linear amplification of the 5caC signal making the signal intensities of 5caC and 5hmC directly incompatible.

One limitation of our method is that we used high resolution confocal microscopy for the visualisation of the results with 63 X magnification for the

quantification and analysis of the images; creating bias as we might have selected for the sub-population of cells that have higher levels of 5caC. Another limitation of that experiment is the lack of controls; unfortunately we did not have access to normal human astrocytic or neural progenitor cells. This experiment would give an indication whether 5caC presence in GBM cell lines is due to the fact that this subpopulation of cells is in a neural progenitor state or the cells are in a de-differentiation state; and active DNA demethylation is operative in this system in a similar way that has been described in different systems during the differentiation into neural/glial lineages and hepatocytes (Lewis *et al.* 2017; Wheldon *et al.* 2014). We tried to overcome these limitations and examine whether the presence of 5caC correlates with the cancer stem cell population exists in these tumours (data not shown). Studies in our lab based on immunocytochemistry showed that there is no correlation between 5caC staining and several cancer stem cell markers (e.g. Nestin, SOX2) (data not shown) thus we could not perform further experiments using fluorescent activated cell sorting (FACS) to isolate the 5caC sub-population.

Several reports showed that high glucose can induce α-KG levels as a co-factor of TET enzymes (Killian *et al.* 2013; Yang *et al.* 2014). Therefore, we next sought to examine whether glucose enriched culture conditions can induce TETs activity encouraging 5caC elevated levels. Rather unexpectedly, we observed significantly reduced TET2 levels in U251 cells cultured in high glucose accompanied by higher TDG levels (Fig. 3.7). These results are consistent with our immunocytochemistry experiments that showed lower levels of 5hmC. Our results indicate that high glucose cannot promote TETs activity in U251 cell line and that 5caC levels are not affected by the changes in glucose conditions. Thus, we did not proceed with these experiments.

5hmC is enriched in the brain compared to other differentiated tissues such as bladder, kidney and liver (Globisch *et al.* 2010). However, in contrast to normal brain tissues, brain tumours exhibit significantly lower amounts of 5hmC and most importantly these levels are correlated with tumour grade and patient survival (Kraus et al. 2012). Correspondingly, 5caC is detectable in breast cancer tissues as well as in paediatric brain tumours but is not

detectable in the corresponding healthy tissues (Eleftheriou et al. 2015; Ramsawhook et al. 2017). Interestingly, the levels of 5caC do not correlate to 5hmC levels indicating that 5caC and active DNA demethylation have a distinct role in carcinogenesis.

Although, 5hmC levels are reduced in gliomas it has been suggested that there is no correlation of these levels with TET proteins alterations in contrast to myeloid malignancies (Kraus et al. 2015). Therefore, we next examined whether 5caC elevated levels corresponded to high TET levels in GBM. We first chose to examine the relative expression levels of *TETs* and *TDG* in GBM cell lines compared to hiPSCs, we also performed semi-quantitative PCR at different cycles and observed that *TET2* levels are elevated in two out of four cell lines examined leading us to propose that 5caC elevated levels might correspond to elevated TET2 levels and to oxidation of 5mC to 5caC. Our siRNA mediated knockdown of TET2 showed significantly reduced levels of 5hmC and 5caC in the absence of TET2 in LN18 cell line confirming our hypothesis that TET2 is responsible for the elevated levels of 5caC observed in some GBM cell lines. These results are in contrast with previous study in gliomas (Kraus *et al.* 2015). They rather indicate that the expression of TETs in these cells may be involved in their pathogenesis, promoting the oxidation of 5mC into 5hmC and 5caC in GBM cell lines.

Chapter 4. Generation and functional characterisation of a TET2 knockout GBM cell line

4.1. Introduction

The recent advances in targeted gene editing techniques include engineered nucleases, zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALEN's) and clustered regularly interspaced short palindromic repeats (CRISPR) have provided a simple and robust method for gene editing (Kim & Kim 2014). Specifically, engineered nucleases create double-strand DNA break (DSB) at the targeted genomic region activating non-homologous end joining (NHEJ) or homology-directed repair (HDR) depending upon the presence of a donor template with homology to the locus of interest (Ma *et al.* 2014).

Megalonucleases, including ZFNs and TALENs, have been widely used for genome editing before the emergence of CRISPR/Cas; they are engineered versions of naturally occurring restriction enzymes (Khan 2019). ZFNs are fusions between a custom designed zinc-finger protein and the cleavage domain of the FokI restriction endonuclease (Kim *et al.* 1996); they act as dimers and each monomer recognizes a specific DNA sequence through the zinc-finger DNA-binding domain. The FokI cleavage domain mediates the dimerization of the ZFN proteins and cuts DNA within 5-7 bp spacer sequence that separates the two flanking zinc-finger binding sites (Smith 2000). The discovery of TALE proteins, which are bacterial effectors that can recognize DNA (Boch *et al.* 2009; Moscou & Bogdanove 2009), enabled the creation of custom TALENs capable of gene manipulation. TALENs are comprised of an amino-terminal TALE DNA binding domain fused to a carboxyl-terminal FokI cleavage domain (Christian *et al.* 2010; Miller *et al.* 2011). Similar to ZFNs, dimerization is achieved through the FokI cleavage domain (Miller *et al.* 2011). One of the challenges that this method has is that TALENs are

substantially larger than ZFNs with highly repetitive structures which makes their delivery into cells much more challenging (Holkers *et al.* 2013).

The CRISPR/Cas9 system was first discovered in bacteria and archaeon as it is an adaptive immune system mechanism against invading DNA contaminants. In bacteria, the CRISPR system protects against DNA from invading organisms through RNA guide DNA cleavage by Cas proteins (Sorek et al. 2013; Wiedenheft et al. 2012). The type II microbial CRISPR system (Chylinski et al. 2014) consists of CRISPR-associated (Cas) genes and noncoding repetitive elements interspaced by short variable sequences (spacers). The spacers, known as protospacers, are short sequences (around 30 bp). They are often derived from foreign genetic elements (Barrangou et al. 2007) and each protospacer is flanked by a short protospacer-adjacent motif (PAM). Once short segments of foreign DNA are integrated within the CRISPR locus and transcribed into CRISPR RNA (crRNA) they then anneal to trans-activating crRNA (tracRNA) directing the degradation of pathogenic DNA by the Cas9 protein (Jinek *et al.* 2012). About a decade ago it was shown that specific recognition by the Cas9 protein requires only a target sequence within the crRNA and a conserved PAM upstream of the binding site (Jinek *et al.* 2012). Since then, the system has been simplified and widely used for genome engineering; consisting of only Cas9 nuclease and a single guide RNA (sgRNA containing the essential crRNA and tracrRNA elements (Cho *et al.* 2013; Jinek *et al.* 2012).

The most widely used Cas9 enzyme for genome editing is derived from *Streptococcus pyogenes* (SpCas9) and requires the 5'-NGG PAM sequence for recognition (Mojica *et al.* 2009). Alteration of the 20 bp guide sequence of the sgRNA enables targeting of Cas9 to any desired genomic locus that harbours the PAM sequence (Cong *et al.* 2013; Wang *et al.* 2013). Cas9 protein is comprised of two nuclease domains RuvC (consisting of three subdomains) and HNH (Jinek *et al.* 2012) which are responsible for the cleavage of the DNA strands at the target sites (Nishimasu *et al.* 2014). The sgRNA is responsible for site targeting as it recognises the sites next to the PAM sequence, while Cas9 is responsible for cleavage and the introduction of DSB. Introduction of DSB by Cas9 enables stimulation of endogenous DNA

damage repair pathways. Homology-directed repair (HDR), which requires a homologous template for recombination, and the non-homologous end-joining (NHEJ), which does not require a template and often produces insertions of deletions as a result of repair (Trevino & Zhang 2014). Thus, exogenous templates can be designed and introduced with Cas9 and sgRNA in order to promote homologous recombination in the area of interest (Trevino & Zhang 2014).

One of the main concerns that needs to be taken into consideration when planning to genetically modify a system using CRISPR/Cas9 is the potential for off target events. It has been shown that mismatches at the 5' end of the target site are generally better tolerated than 3' end (Cong *et al.* 2013; Jinek *et al.* 2012). Nevertheless mismatches of fewer than three nucleotides can induce off-target events (Fu *et al.* 2013). Moreover, it has been reported, that high GC contents in the target site can improve hybridisation and allow more mismatches to be tolerated (Pattanayak *et al.* 2013). Several online engineering design and off-target search tools have been developed for the accurate prediction of sgRNA regions with lower probabilities of off-target events. Aside from sgRNA design optimisation, it has also been suggested that reduction of Cas9 and gRNA concentrations can decrease off-target effects; however it might also have an effect on-target cleavage (Hsu *et al.* 2013). The use of Cas9 and gRNA mRNA/protein instead of an expression plasmid can also reduce off targets as it remains active for a shorter time in the cell (Kim & Kim 2014).

A mutated variant of Cas9 has also been developed; it contains D10A or an H840A mutation at RuvC and HNH nuclease domain respectively (Jinek *et al.* 2012; Ran *et al.* 2013). This mutant version of Cas9 can induce a single stranded nick capable of stimulating HDR but at lower efficiency. However, when used with two gRNAs driving two nicks in close proximity to each other in the area of interest, DSBs with a 3' or 5' overhangs are created, in contrast to wild-type Cas9 that introduces a blunt DSB (Mali *et al.* 2013; Ran *et al.* 2013). The advantage of this technique is that it minimises the off-target sites making the whole Cas9 system more precise (Ran *et al.* 2013; Sander & Joung 2014; Shen *et al.* 2014a).

This system has also been adapted for transcriptional modulation via fusion of effector domains to a catalytically inactivated variant of the Cas9 protein (Qi *et al.* 2013). The introduction of two amino acids into the RuvC and HNH endonuclease domains of Cas9 can lead to the deactivation of Cas9 generating a mutant dCas9 which is unable to produce DSB. However, dCas9 can bind to DNA and modulate gene expression from either strand of the targeted DNA sequence (Farzadfard *et al.* 2013; Maeder *et al.* 2013; Perez-Pinera *et al.* 2013). Fusion of dCas9 to transcriptional repressor domains can lead to efficient silencing of the promoter for instance (Gilbert *et al.* 2013; Zalatan *et al.* 2015).

In summary, the development of targeted, efficient genome editing using RNA-guided Cas9 system has enabled the systematic manipulation of genetic elements in a variety of cells and organisms (Hsu et al. 2014). One of the advantages of this system is that it does not require the engineering of specific protein pairs for each target site, unlike ZNFs and TALEN's (Boch & Bonas 2010; Klug 2010). The site recognition is entirely mediated by the sgRNA, making CRISPR/Cas9 system the most flexible and adaptable platform for genome editing, eliminating the need for engineering of new proteins for the recognition of each target site.

4.2. Aims and rational

Our previous results showed that TET2 levels are elevated in GBM while the transient knockdown of TET2 leads to decreased immunocytochemically detectable levels of 5caC/5hmC in GBM cell lines. Thus, given our indications that TET2 might be the enzyme that mediates the oxidation of 5mC into 5hmC and 5caC in GBM we decided to use CRISPR/Cas9 to knockout the *TET2* gene in one GBM cell line. Our aim was to target and disrupt the catalytic domain of *TET2* gene in order to study the effects of TET2 enzymatically inactive protein in active DNA demethylation in GBM.

4.3. TET2 gene structure and catalytic domain

According to the UCSC genome browser (https://genome.ucsc.edu/) the *TET2* gene is situated on chromosome 4 (chr4 q24) and is 133,524 bp in length (Fig. 4.1 A). The TET2 transcript is comprised of 11 exons, 9 of which are translated (Fig. 4.1 B). The TET2 protein is 2002 aminoacids (aa) long and belongs to the TET gene family. All three TET enzymes are comprised of a C- terminal catalytic domain that encompasses a Cys-Rich domain and a DSBH domain with a large low-complexity insert that is responsible for the oxidation of 5mC in a 2-oxoglutarate- (2-OG) and Fe (II)- dependent manner (Hu *et al.* 2013a; Pastor *et al.* 2013) (Fig. 4.1 C). *TET1* and *TET3* also have a CXXC domain which has the ability to bind to DNA (Pastor *et al.* 2013; Williams *et al.* 2012). Nevertheless, the *TET2* gene underwent a chromosomal gene inversion during evolution resulting in the separation of the CXXC domain from the catalytic domain, thus IDAX gene encodes CXXC domain of TET2 regulating its expression (Ko *et al.* 2013). Specifically, it has been shown that CXXC domain of IDAX binds at CpG dinucleotides and directly interacts with the catalytic domain of TET2 (Ko *et al.* 2013).

In order to identify the minimum catalytically active fragment in TET2 enzyme; the catalytic activity of various N-terminal and C-terminal truncations and deletions in TET2 protein were measured, identifying TET2 1129-1936 aa positions as the minimum catalytically active fragment (Hu *et al.* 2013a) (Fig. 4.1 C). Nevertheless, full length TET2 resulted in much higher levels of 5hmC than any truncated versions of TET2 suggesting that full-length TET2 is more active than the catalytic domain only, which was in agreement with previous study (He *et al.* 2011).

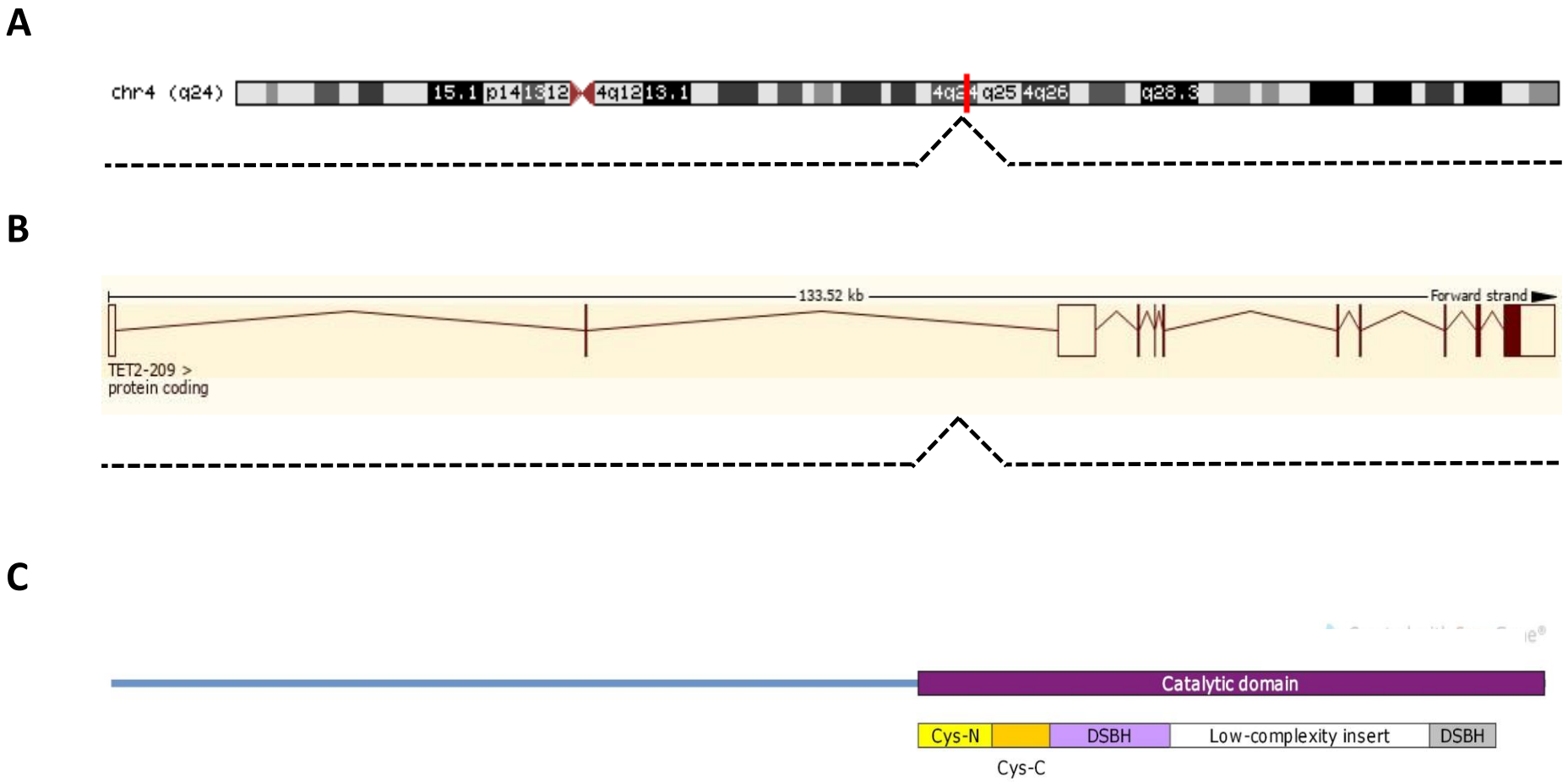


Figure 4.1: Schematic illustration of TET2 gene and protein structure.

A. Genomic location of the *TET2* gene at chromosome 4. B. TET2 protein coding transcript which comprises of 11 exons, 9 of which are translated. C TET2 protein consists of the catalytic domain which is responsible for oxidation of epigenetic marks.

A study revealing the crystal structure of TET2 showed that two zinc fingers bring together the DSBH and Cys-Rich domains to form a compact catalytic domain (Hu *et al.* 2013a) (Fig. 4.2). TET2 targets DNA and forms a compact globular fold of catalytic domain in complex with DNA. The DSBH core, contains two antiparallel β sheets with the Fe (II) and NOG co-factors localised in the centre (Hu *et al.* 2013a). The methyl group at the C-5 position of DNA is then flipped to the catalytic domain and oxidation happens once the NOG is replaced by 2-OG (Hu *et al.* 2013a) (Fig. 4.2). Interestingly, the residues involved in this reaction are highly conserved across the TET enzymes family as is the Cys-Rich domain which is also essential for their enzymatic activity (Iyer *et al.* 2009; Tahiliani *et al.* 2009). Moreover, the TET2 overall structure is further stabilised by the coordination of three zinc cations (Zn1, Zn2, and Zn3) located at separated corners of TET2 protein (Hu *et al.* 2013a). Specifically, Zn1 coordinated only with residues within the Cys-N subdomain and deletions in N terminus domain results in significantly decreased TET2 activity, indicating the crucial role of Zn1 coordination in stabilising TET2 structure (Hu *et al.* 2013a). The Cys-Rich domain is important for the integrity of the TET2-DNA complex to ensure proper catalytic activity of TET2. Moreover, two of the three zinc fingers bring the Cys-Rich and DSBH domains together in order to stabilise the complex.

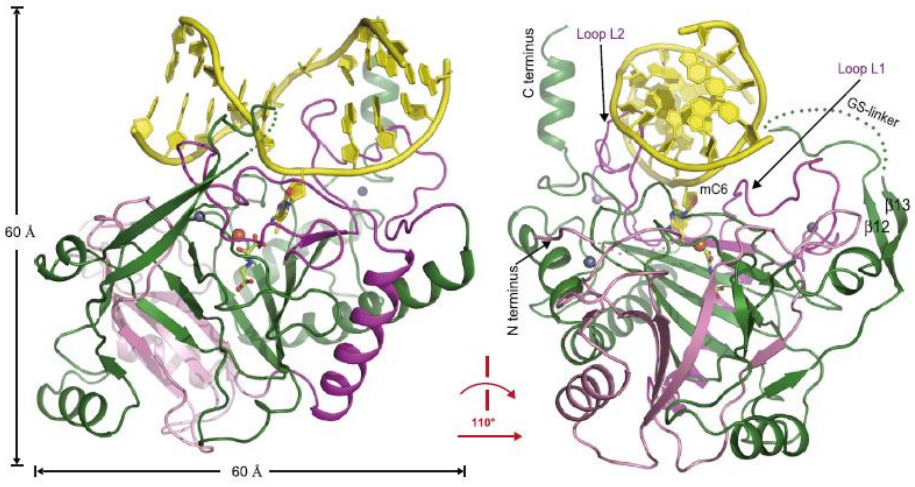


Figure 4.2: Structure of TET2-DNA interaction.

Representation of TET2-DNA structure in two different views. The DNA is shown in yellow; the three zinc citations are shown as grey balls and the iron is shown as red. The mC6 DNA-interacting loops (L1 and L2), N and C termini, and GS linker are indicated. Adapted from (Hu *et al.* 2013a).

4.4. Results

4.4.1. Design of TET2 targeting strategy

With the aim of achieving complete disruption of TET2 protein enzymatic activity, we decided to target the genomic sequence at exon 3 which encodes for the protein just before the start of the catalytic domain (Fig. 4.3 A). Overall, the plan for TET2 knockout is to direct Cas9 to the locus of interest and introduce DSBs in the DNA which will then be repaired by homologous recombination. A cassette with sequences homologous to the locus of interest will be inserted at the site of the DNA break in order to select for targeted clones. The cassette we chose to use encodes for the EF1-α promoter which drives the transcription of the GFP and BSD (Blasticidin resistance) sequences downstream; it is derived from a parental vector plasmid pSIN-GFP-IRES-BSD (provided by Dr Alexander Kondroshov) (Fig. 4.3 B). We chose a cassette with two selection markers (GFP and BSD) in order to increase the selection efficiency of the correctly targeted clones. After homologous recombination the cassette will then be incorporated into the genome at the Cas9 cut site resulting in disruption of exon 3 of the *TET2* gene (Fig. 4.3 C).

In order to direct Cas9 nuclease to the locus chosen, we designed two sgRNA sequences using an online tool provided by the Sanger database (https://wge.stemcell.sanger.ac.uk/find_crisprs?). The database is designed to find unique sequences of 20 nucleotides (nt) in the locus of interest containing a PAM site at the 3' end. It also gives information on the off target counts, giving possible mismatches in the hybridisation of the two sequences (Table 4.1). We decided to reduce the possibility of off-target events to 0 by using a pair of bidirectional sgRNAs with a mutant Cas9 D10A; this way Cas9 would introduce single strand DNA breaks 3-4 nt upstream of the PAM sequence on each DNA strand. The two guide RNA sequences should be placed at the two opposite strands of DNA and be in close proximity of each other (around 30 nt) in order to increase efficiency. We then double checked that the sgRNAs chosen did not recognise any other regions in the genome using the NCBI Blast online tool.

Table 4.1: sgRNA sequences as identified by the online Sanger tool at exon 3 of TET2 gene.

The table shows the sequence of the sgRNAs, the DNA strand they are on, and the off target counts (Exonic, Intronic and Intergenic). The first number indicates the number of mismatches and the second indicates the number of off target events.

Guide RNA	Sequence	Strand	Off target counts	Exonic off target	Intronic off target	Intergenic
Guide RNA A	CCAAGATATGTCTGGTC AACAAG	-	124 {0: 1, 1: 0, 2: 0, 3: 6, 4: 116}	4 {0:1, 1:0, 2:0, 3:0, 4:3}	51 {0:0, 1:0, 2:0, 3:4, 4:47}	68 {0:0, 1:0, 2:0, 3:2, 4:66}
Guide RNA B	CTGCGCAACTTGCTCA GCAAAGG	+	81 {0: 1, 1: 0, 2: 0, 3: 9, 4: 71}	9 {0:1, 1:0, 2:0, 3:1, 4:7}	32 {0:0, 1:0, 2:0, 3:4, 4:28}	40 {0:0, 1:0, 2:0, 3:4, 4:36}

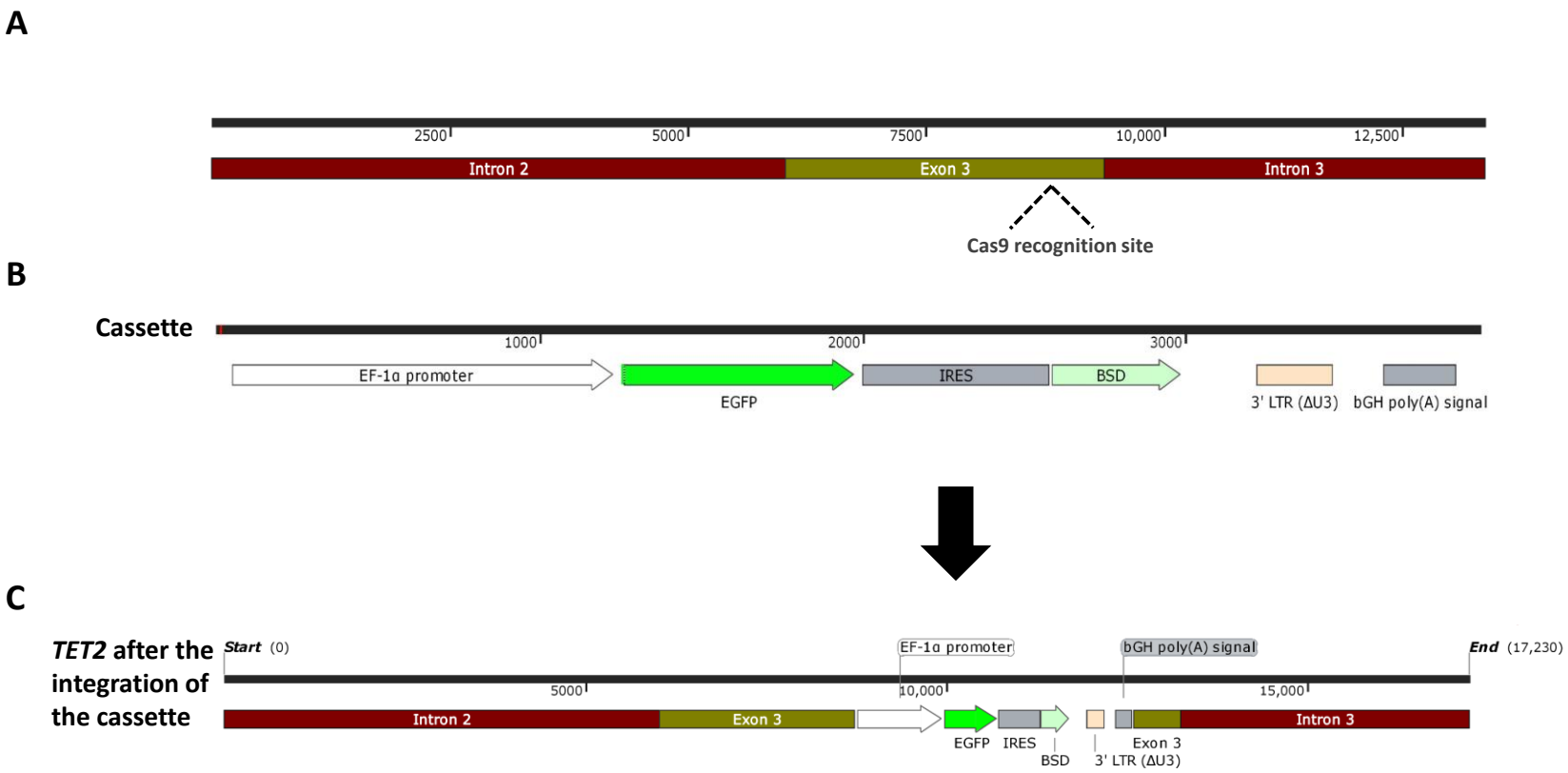


Figure 4.3: Schematic illustration of TET2 gene disruption at the 3' end of exon 3.

A. Illustration of the targeting site of *TET*2 (Intron2-Intron 3); Cas9 recognition site is at the 3' end of exon 3. B. Illustration of the cassette that is going to be inserted via homologous recombination, disrupting TET2 function. This cassette is comprised of two selection markers, EGFP fluorescent protein and BSD resistance gene. C *TET2* gene (Intron 2-Intron 3) after the integration of the cassette at the Cas9 recognition cut site.

4.4.2. Building of the guide RNA plasmids

After selection of the sgRNA sequences we then designed oligos for the building of the gRNA vectors. We used Gibson Assembly cloning method for the integration of guide sequences into the guide RNA (gRNA) vector; the workflow of the building is presented in Figure 4.4 A. In brief, Gibson Assembly is developed for the easy assembly of multiple linear overlapping DNA fragments in one isothermal reaction using three different enzymes: T5 exonuclease creates single strand 3' overhangs; Phusion DNA polymerase incorporates the nucleotides in the annealed DNA fragments; and Taq DNA ligase joins the annealed complementary DNA fragments (Gibson *et al.* 2009). Thus, we designed primers containing 40 bp of the RNA pol III U6 promoter upstream (at the 5' end) followed by the sgRNA sequence in one strand and the reverse compliment of the sgRNA sequence followed by 40 bp of small nuclear RNA scaffold (at the 3' end) (Fig. 4.4 A). We then annealed the gRNA primers using PCR (Fig. 4.4 B), in order to achieve a double stranded DNA sequence with the promoter region at the 5' end and the RNA scaffold at the 3' end of the sequence. The gRNA vector plasmid was linearised using AfIII restriction enzyme which is located between the U6 promoter and the gRNA scaffold sequence (Fig. 4.4 C). Finally, we performed the Gibson Assembly reaction for the assembly of the two DNA fragments. The product of the reaction was transformed in *E.Coli* TOP10 competent cells to make multiple copies, which were isolated and sent for sequencing. The results of the sequencing are shown in Figure 4.4; displaying the correct integration of the two gRNA sequences in the gRNA vector (Fig. 4.4 D).

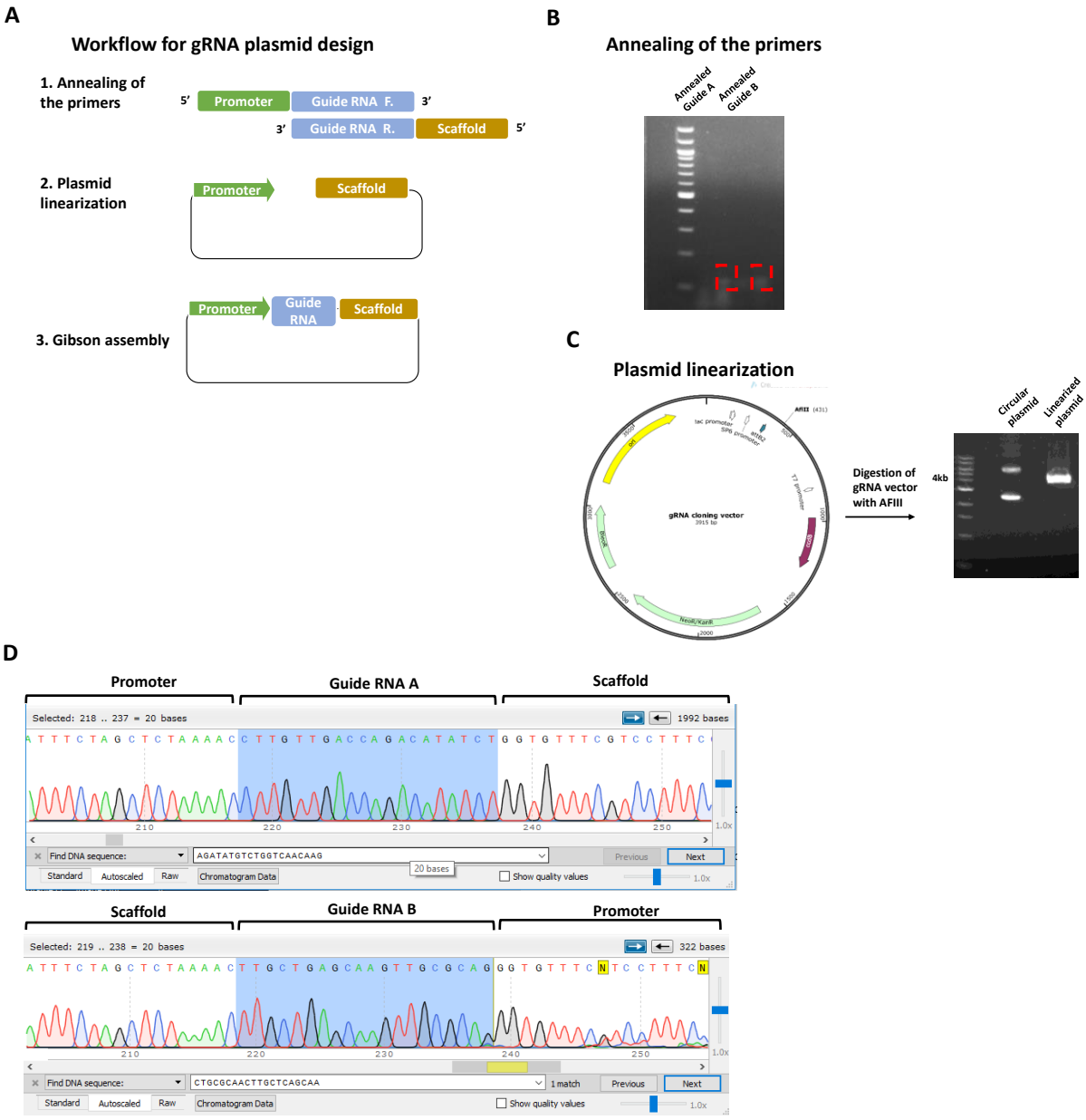


Figure 4.4: Design and building of gRNA plasmids.

A. Schematic illustration of the steps for the design of the plasmids including the annealing of the primers; plasmid linearisation and Gibson Assembly. B. Annealing of the primers designed for Gibson Assembly strategy. The primers contain the gRNA sequence as well as part of the promoter region at the 5′ end and the scaffold at 3′ end of the gRNA vector plasmid to ensure the successful integration of the gRNA into the plasmid. C. gRNA cloning vector linearisation using the AFIII enzyme D. Sequencing results as illustrated by the chromatogram after the integration of the gRNAs into the plasmid.

4.4.3. Design of the intermediate and final donor plasmids

We next proceeded with the building of the donor plasmid which would contain the arms of homology and the selection cassette. We designed the arms of homology upstream (Left homology arm (LHA)) and downstream (Right homology arm (RHA)) of the Cas9 recognition sites. We first designed primers for amplification of the two fragments containing unique restriction enzyme sites at both ends. Specifically, BamHI, and KasI were added at the 5' and 3' end of the LHA respectively; resulting in a PCR product of 1013 bp (Fig. 4.5 A). For the RHA, AscI and XbaI restriction sites were added at the 5' and 3' end respectively, resulting in a 630 bp PCR product (Fig. 4.5 B). The incorporation of the restriction sites at both sites of RHA and LHA is essential for the subsequent cloning steps. We isolated the product of the first PCR reaction and performed a second PCR incorporating a part of RHA and a part of LHA at the 3' end of LHA and RHA respectively (Fig. 4.5 B); this was an intermediate step in order to proceed with fusion of the two arms of homology using fusion PCR (Fig. 4.5 C). After that, we isolated the PCR product and digested it using the BamHI and XbaI enzymes; the donor plasmid was also digested with the same enzymes in order to create the same 5' and 3' overhangs (Fig. 4.5 D). Finally, we ligated the digested products to achieve incorporation of arms of homology into the donor plasmid. This plasmid was sequenced to ensure the correct ligation reaction (Fig. 4.5 E). The resultant plasmid served as an intermediate plasmid for building of the final TET2 knockout plasmid.

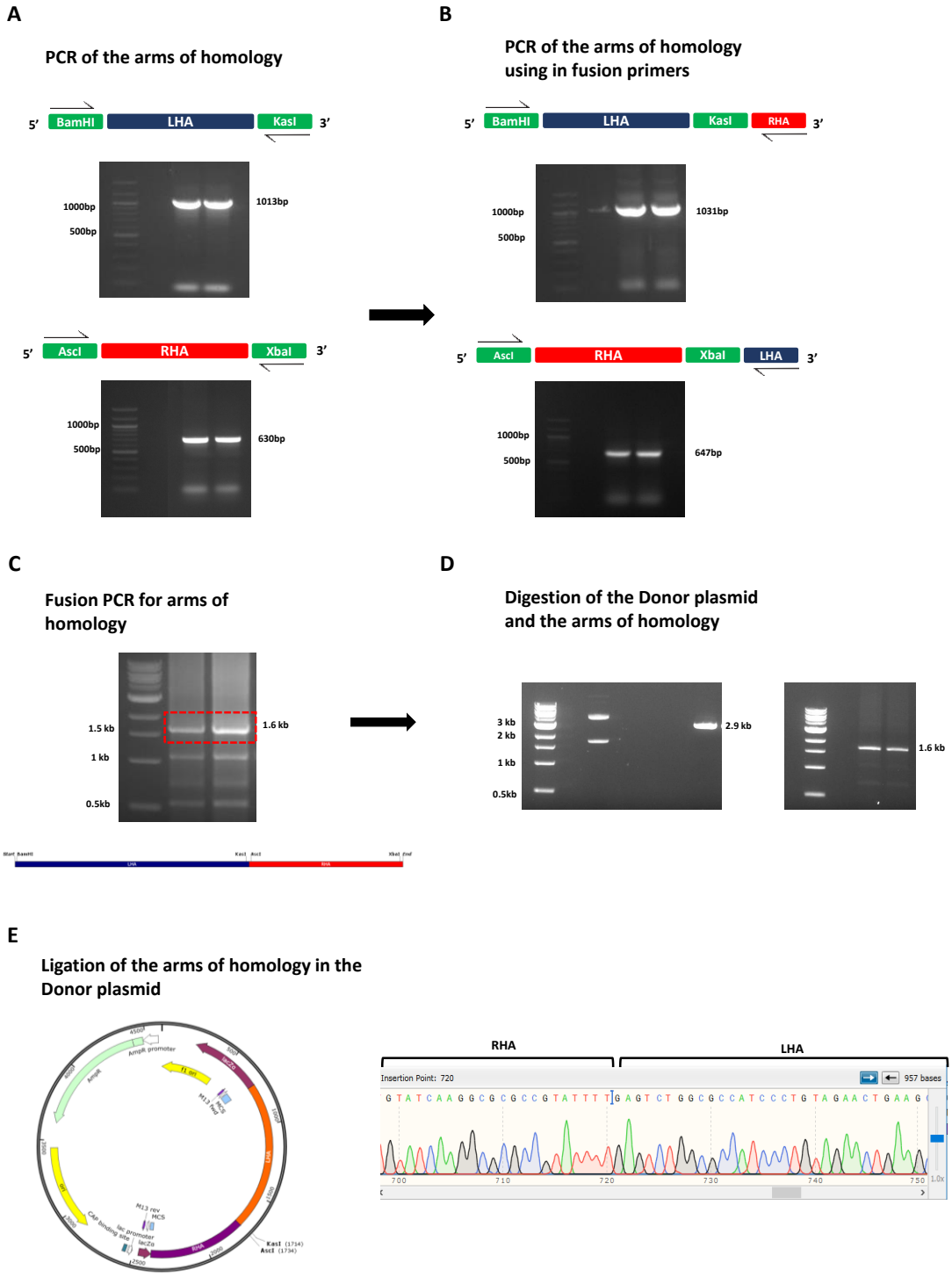


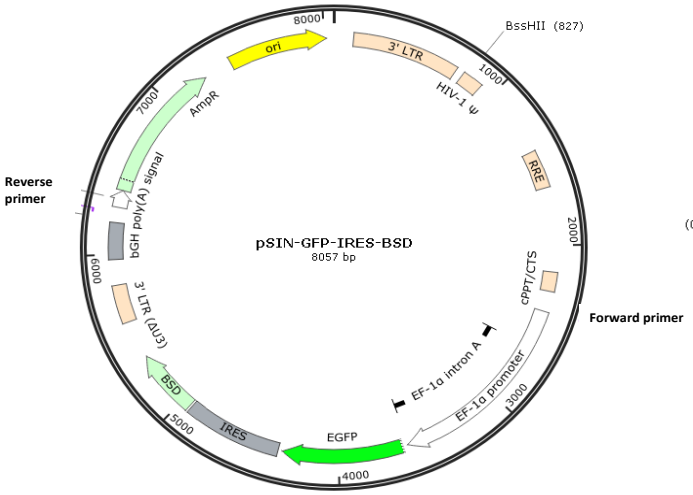
Figure 4.5: Steps for the design of the intermediate plasmid that would harbour the arms of homology.

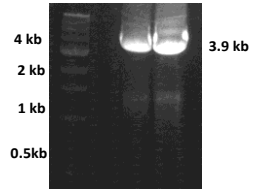
A. PCR for the LHA and RHA incorporating BamHI and KasI restriction sites at 5′ and 3′ end in LHA and AscI and XbaI at 5′ and 3′ end in RHA. B. PCR of the arms of homology incorporating parts of RHA and LHA at the 3′ end in order to allow the fusion of the two products. C. Fusion PCR for the arms of homology resulting in 1.6kb product. D. Digestion of the donor plasmid and the fused arms of homology with the BamHI and XbaI enzymes. E. Schematic illustration of the plasmid after the ligation of the arms of homology and sequencing chromatogram results showing the presence of the arms of homology in the plasmid.

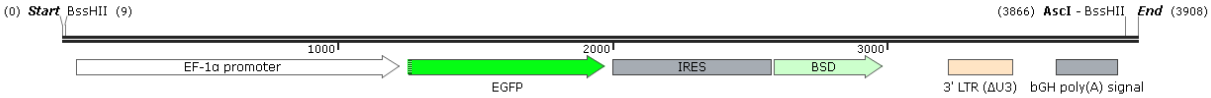
After the generation of the intermediate plasmid, we next sought to incorporate the selection cassette between the arms of homology. Thus, the final donor plasmid would include the arms of homology and the cassette with the two selection marks. We decided to use restriction enzyme based digestion and ligation for this last part of our cloning. The AscI restriction site was already present at one end of the cassette in the pSIN-GFP-IRES-BSD plasmid. Therefore, we first attempted to add the KasI restriction site at the other end of the cassette in order to proceed with restriction enzyme based digestion with AscI and KasI and ligation into the intermediate plasmid. For that reason we designed primers that would have homology with the 5' and 3' end of the cassette as well as the restriction enzymes sequence with the addition of 6 nt at the end. This way we could ensure the correct amplification of the cassette and the addition of the restriction enzymes sites. Nevertheless, the possibilities of amplifying the region of interest are less due to the large size of the primers, their high GC content, the addition of non-complementary sequence and in addition to that the large size of the cassette (3.9 Kb); thus, we were unable to PCR amplify the region of interest. The addition of AscI restriction site at the other end was also unsuccessful for the same reasons. Hence, we decided to add the BssHII restriction site which creates the same overhang ends with AscI but has different recognition site. We successfully PCR amplified the cassette from the plasmid with the addition of BssHII (Fig. 4.6 A). We then isolated the PCR product, digested with AscI and BssHII enzymes (Fig. 4.6 A) and proceeded with digestion of the intermediate plasmid with AscI followed by de-phosphorylation in order to avoid plasmid self-ligation. In the final step, using different ratios of the intermediate plasmid (4 kb) and the cassette (3.9 kb), we achieved the ligation of the cassette into the intermediate plasmid (Fig. 4.6 A). The map of the expected final donor plasmid is displayed in Figure 4.6 B. In order to ensure correct ligation of the two products we performed diagnostic PCRs of the final plasmid (Fig 4.6 B). We used primers to amplify the main junctions in our final donor plasmid, the LHA-EF1α promoter (662 bp) the EF1α-GFP (365 bp) the 3'UTR- RHA (543 bp) and the BSD-RHA (1036 bp) (Fig 4.6 C). Finally, PCR products of these reactions were sequenced; the alignment of the sequencing results to the corresponding sequence led to the confirmation of the final plasmid (Fig. 4.6 D).

A

1. PCR of the cassette and addition of a restriction site

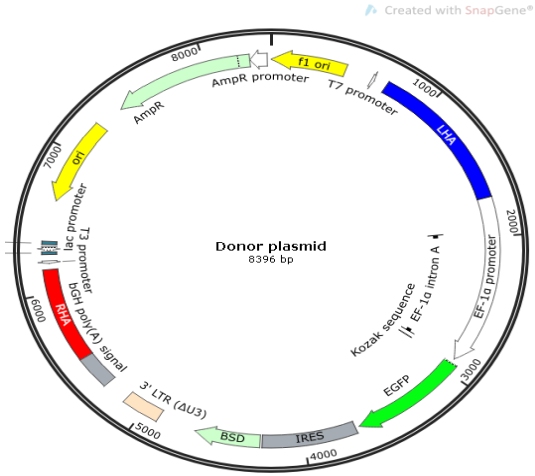




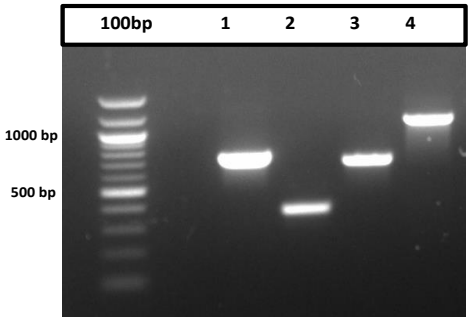


2. Digestion of the cassette with AscI and BssHII ➡ **3. Digestion of the intermediate plasmid with AscI and de-phosphorylation** ➡ **4. Ligation of the cassette into the intermediate plasmid**

B



C



1. LHA- EF1α: 663 bp
2. EF1α- GFP: 365 bp
3. 3'UTR- RHA: 543 bp
4. BSD- RHA: 1036 bp

D

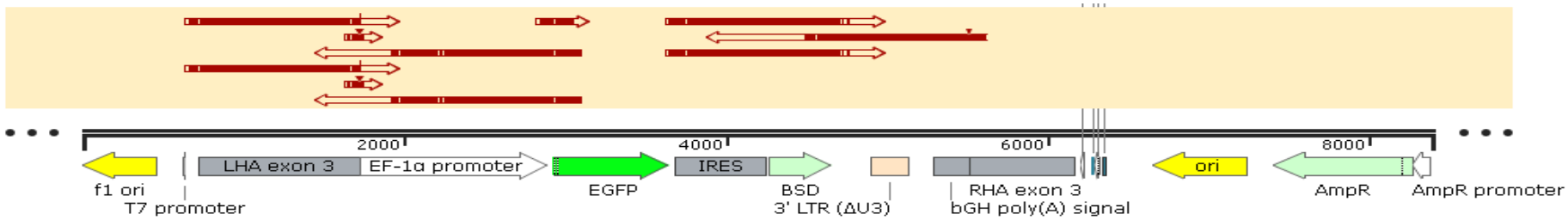


Figure 4.6: Building of the final donor plasmid.

A. Step 1 shows the map of the pSIN-GFP-IRES-BSD plasmid used to isolate the selection cassette; PCR of the cassette was optimised and performed with the addition of a restriction site BssHII at the 3' end. In step 2 the cassette is digested with AscI and BssHII followed by the digestion, the de-phosphorylation of the intermediate plasmid and the ligation of the cassette into the intermediate plasmid. B. The map of the final donor plasmid after ligation. C. Diagnostic PCR to ensure the correct integration of the cassette into the intermediate plasmid. D. Sequencing results of the PCR products shows their alignment to the corresponding areas in the donor plasmid; ensuring the correct integration of the cassette.

4.4.4. Survival analysis and transfection optimisation

In order to determine the minimum antibiotic dosage that kills 100% of untreated control cells, we performed BSD kill curves. Previous publications in GBM cell lines suggested the use of a range between 2.5-10 µg/ml of BSD treatment for 10 days (Carreira *et al.* 2016; Radu *et al.* 2003; Wei *et al.* 2019; Wiedemeyer *et al.* 2010). Therefore, we decided to administer BSD in a range between 2.5-15 µg/ml in LN18 GBM cell line for 14 days. The minimal efficient lethal BSD dosage after 10 days of treatment was 5 µg/ml (Fig. 4.7 A).

We next optimised the transfection conditions, using GFP plasmid as a positive control. We first tried Lipofectamine transfection as it has been previously used for the transfection of siRNAs (Lavanya *et al.* 2016) and plasmid DNA (Würstle *et al.* 2017) into LN18 cell line. We used a range of concentrations between 0.5-5 µg/ml and monitored cell viability and the efficiency based on the presence of GFP (Fig. 4.7 B). Surprisingly, we observed that concentrations as low as 2 µg/ml could cause cell death; therefore we did not proceed with the quantification of the transfection efficiency; although, light microscope images showed that the efficiency was lower than expected. Due to the cytotoxic effects of Lipofectamine in LN18 cell line we decided to change our transfection approach to electroporation. Thus, we used electroporation to transfect LN18 cells with the GFP plasmid. The conditions were optimised to achieve the highest efficiency with minimal cytotoxic effects (Fig. 4.7 C).

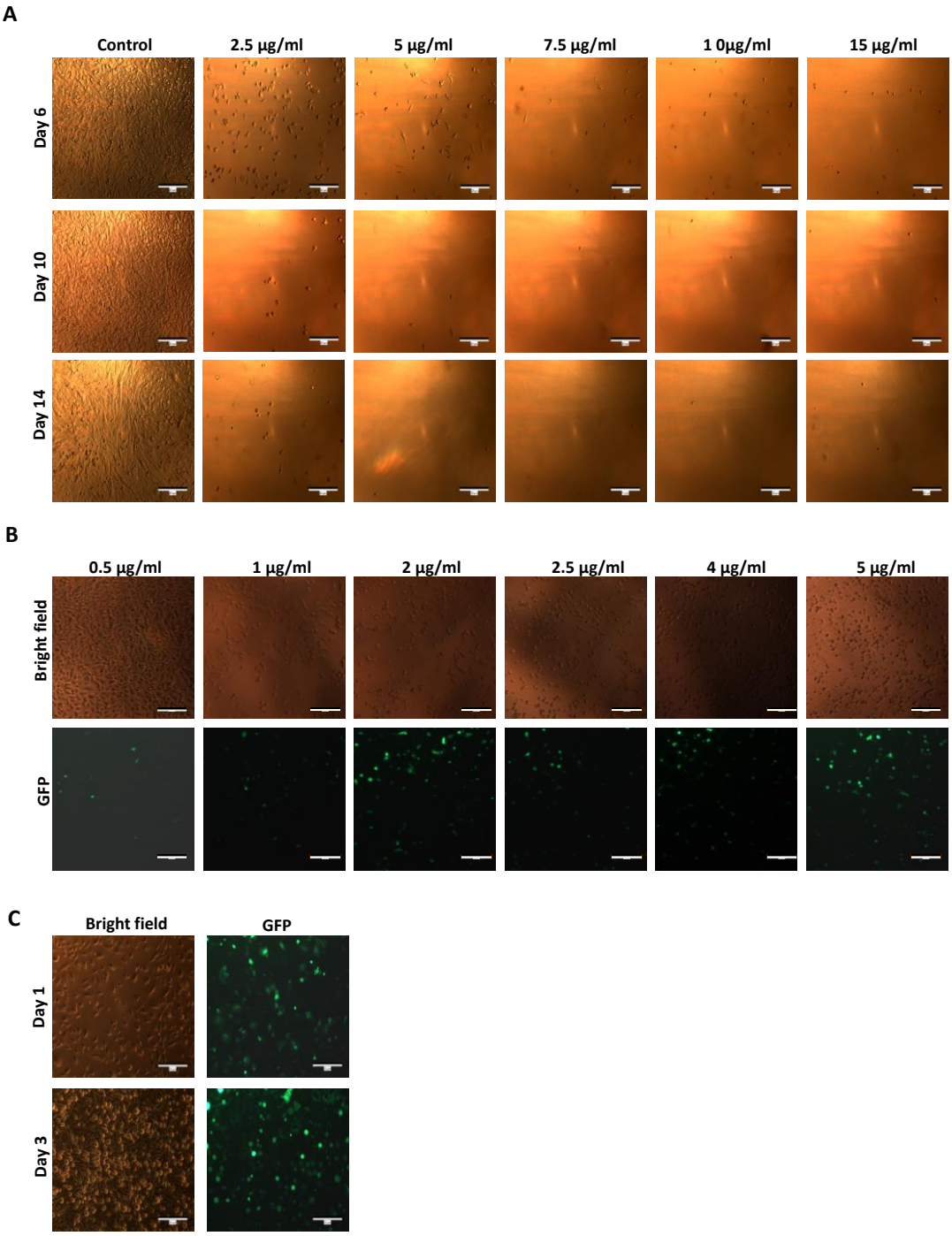


Figure 4.7: BSD kill curve and optimisation of transfection conditions.

A. BSD Kill curve. Light microscope images were taken at 3 different time points to monitor cell death. 5 µg is the minimum concentration that could be used to kill all the cells in culture after 10 days. B. Optimisation of Lipofectamine transfection using GFP plasmid as a control. Light microscope images both in bright field and in the GFP channel were taken to monitor cell survival and efficiency after transfection C. Optimisation of electroporation transfection using a GFP plasmid as control. Light microscope images were taken one day and three days after the transfection to monitor the efficiency and cell survival. Scale bar = 250 µm.

4.4.5. Transfection strategy

The generation of the *TET2* knockout (KO) cell lines included three main steps. The first one was the successful construction of plasmids that directed Cas9 to the *TET2* genomic locus of interest and the donor plasmid that would provide a template for homologous recombination. The second step was the successful transfection of these plasmids into the cell line of interest. The last step was selection of the BSD resistant and GFP positive clones. Considering that our donor plasmid contained two selection markers (GFP and BSD), a two-way screening was performed in order to increase the possibility of selection of successfully targeted clones. Briefly, the plasmids (donor plasmid, gRNAs, and Cas9) were transfected using electroporation the BSD was administered to the cells for 10 days followed by FACS selection of GFP positive cells. Single cells were plated into a 96-well plate in order to generate single clones (Fig. 4.8). Finally, we screened the clones for correct integration of the cassette in the area of interest.

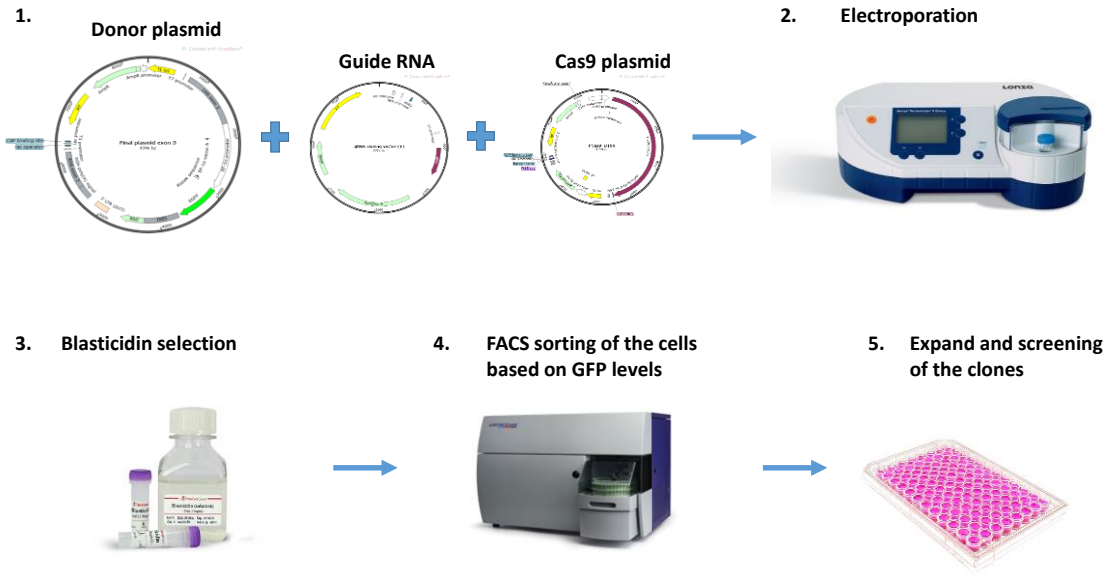


Figure 4.8: Schematic representation of the targeting approach.

The donor, gRNA and Cas9 plasmid were transfected via electroporation into LN18 cell line. Selection of the positive clones was achieved through BSD selection and FACs sorting of the GFP positive cells.

4.4.6. Unsuccessful targeting of TET2 in LN18 cell line using nickase CRISPR/cas9

We first attempted to target LN18 GBM cell line with the nickase CRISPR/Cas9 approach aiming to limit the off-target effects that wild type Cas9 might cause. Therefore, we transfected LN18 cell line with the two gRNA plasmids (Guide A and Guide B), the nickase D10A plasmid and the donor plasmid. A GFP plasmid was transfected separately as a positive control (Fig. 4.9 A). We cultured the cells for 8 days in order to allow them to recover from transfection. We then proceeded with the BSD treatment for an additional 14 days (Fig. 4.9 B). Cells that were transfected with all four plasmids were monitored during BSD treatment and light microscope images were taken every 3-4 days (Fig. 4.9 B). Once the BSD selection was complete, we used FACS to sort GFP positive cells. 80% of the survivors were GFP positive indicating the cassette was integrated into the genome (Fig. 4.9 C).

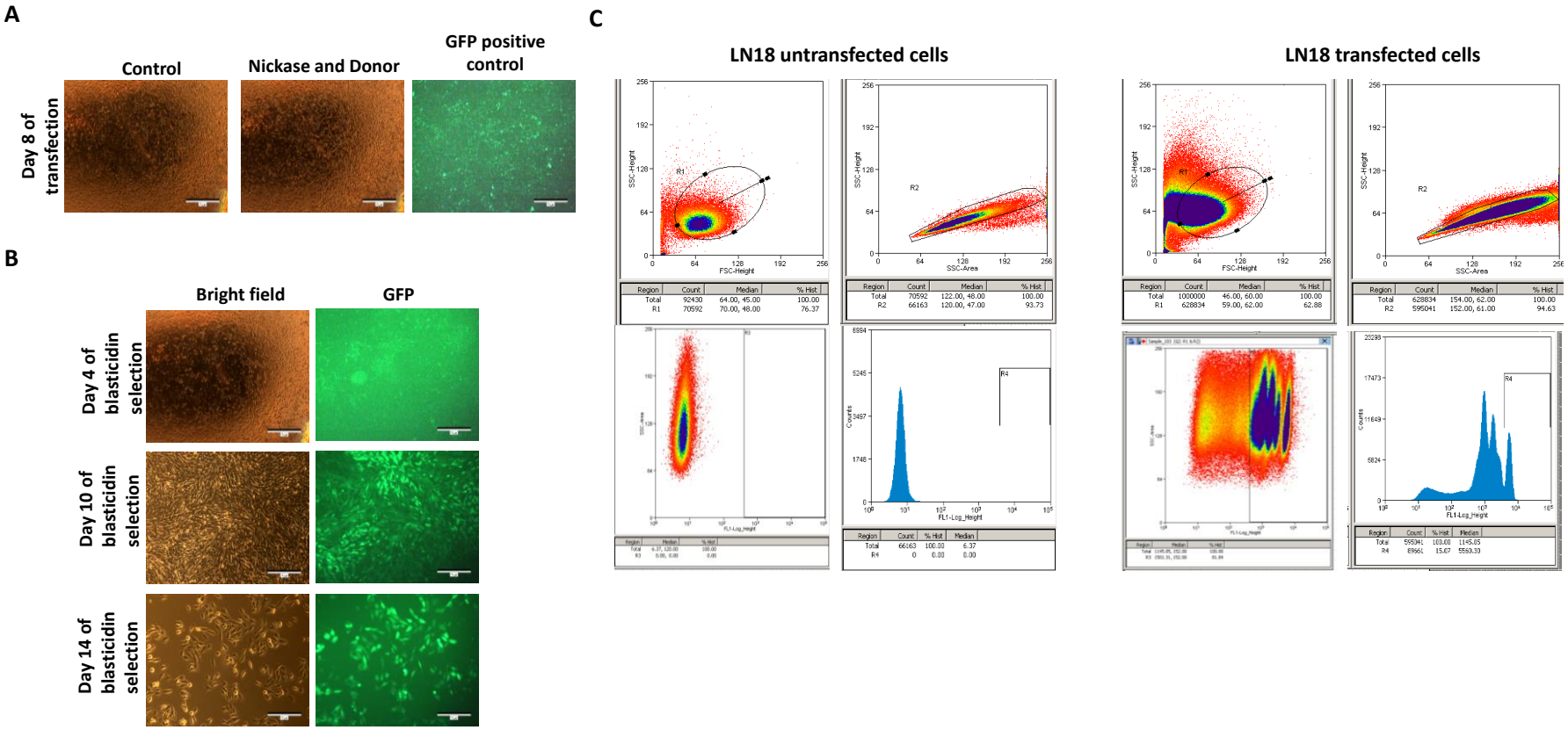


Figure 4.9: Transfection of the plasmids in LN18 cell line. Four plasmids were transfected in total: Donor plasmid, Guide A, Guide B and Nickase plasmid.

A. Light microscope images 8 days after the transfection using electroporation; positive and negative (untransfected cells) controls were used. B. Light microscope images in both bright field and GFP channel at three different time points after the BSD selection. Scale bar = 250 µm. C. FACS of GFP positive cells after the BSD selection. The threshold for FACS selection was set using LN18 untransfected cells and GFP positive cells were selected based on the highest GFP intensity; 80 % of the population of cells were GFP positive.

After FACS of GFP positive cells, single cells were seeded into individual wells of two 96-well plates. The individual clones were then expanded and screened for cassette integration into the locus of interest. We first performed PCRs for 5' and 3' integration of the cassette at exon 3 of *TET2*, analysing 180 clones in total. We designed primers that would recognise the genomic area upstream of LHA into the beginning of the cassette for 5' integration, as shown in the map in Figure 4.10 A; the expected size of the PCR product was 1414 bp. However, none of the clones tested showed any band at the size of interest (Fig. 4.10 A). As for 3' integration, our primers recognised the genomic area downstream of RHA to the end of the cassette with an expected size of 1388 bp. Similar to 5' integration we did not observe any positive clones for 3' integration (Fig. 4.10 B). Our results suggest that the resistance of cells to BSD and the presence of GFP is due to random integration of the donor plasmid into the genome.

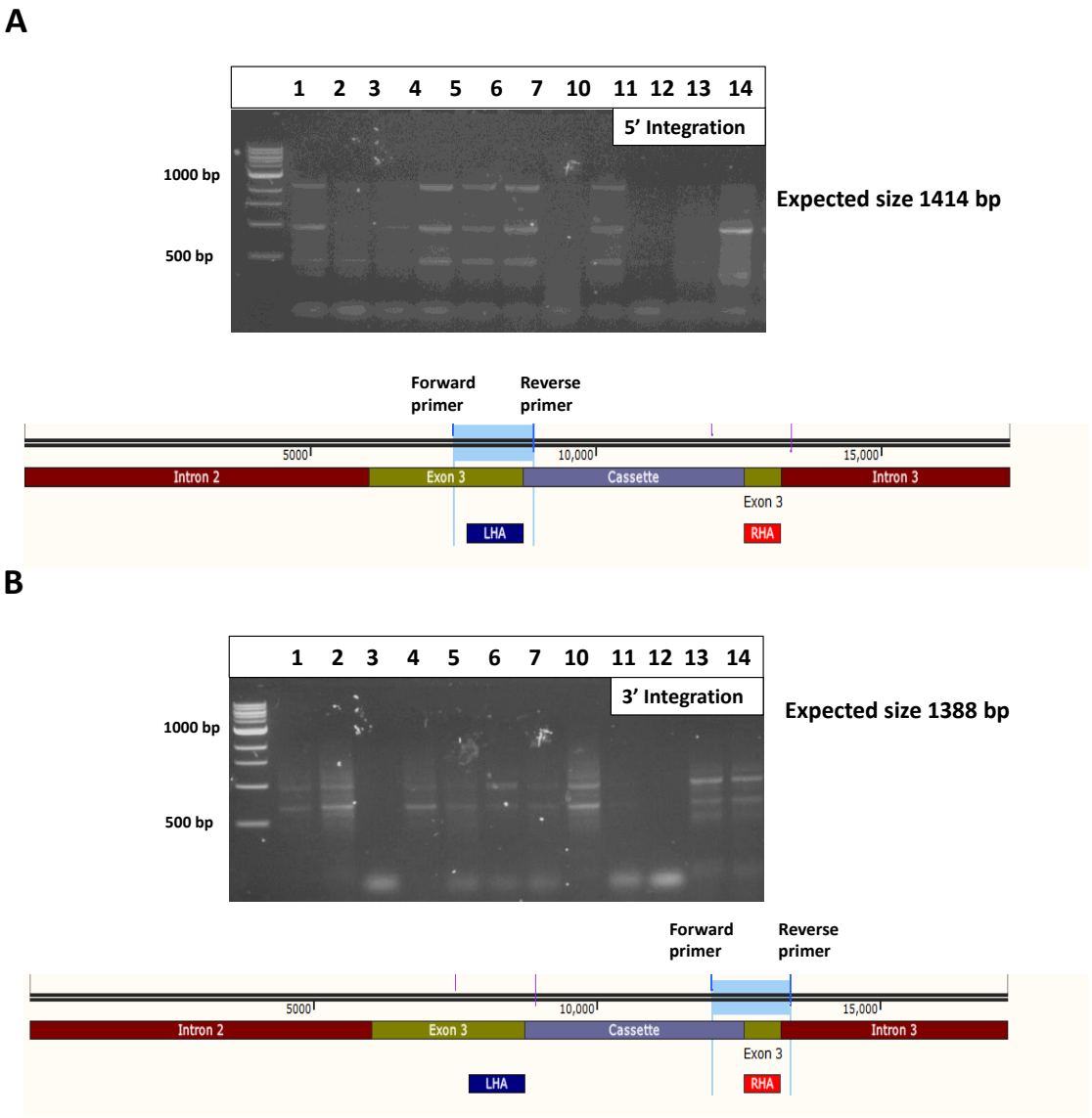


Figure 4.10: Random integration of the cassette into LN18 cells.

Screening of clones for the correct integration of the cassette. 180 clones were analysed in total, with the areas examined shown in the corresponding SnapGene maps. A. Indicative PCR of 14 samples for the 5' integration of the cassette; the expected PCR size was 1414 bp B. Indicative PCR of 14 samples for the 3' integration of the cassette; expected size was 1388 bp.

Our observations led us to hypothesise that random integration was responsible for positive selection of our clones. Therefore, we decided to transfect the cells in the absence of Cas9 plasmid and monitor them under BSD selection. Specifically, we transfected the cells with the donor plasmid only and the donor plus either guide A or B plasmids. A GFP plasmid was transfected as a positive control and light microscope images of both bright field and GFP channels were captured for all conditions (Fig. 4.11 A). We observed that after the end of the BSD treatment the cells that were not transfected with the donor plasmid underwent cell death as expected. Nevertheless, a sub-population of cells that were transfected with the donor plasmid in the absence of Cas9 plasmid, which introduces the DSB, and the presence of gRNAs survived the BSD treatment and were also GFP positive in culture (Fig. 4.11 B). These observations indicate that there is a high risk of random integration of the cassette into the genome, as the cells develop mechanisms to survive BSD selection.

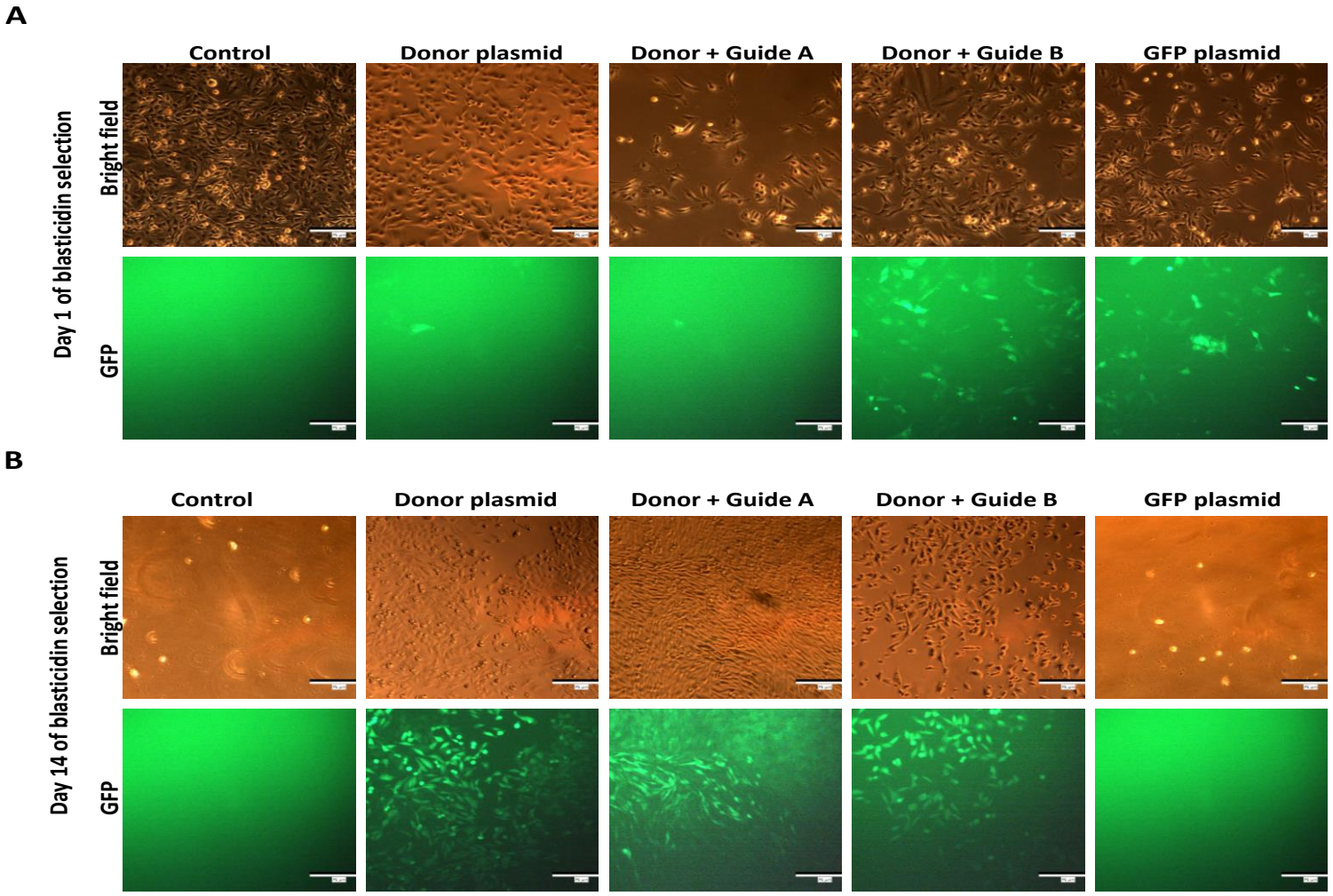


Figure 4.11: Random integration of the donor plasmid occurs even in the absence of Cas9 plasmid.

Light microscope images (bright field and GFP) of negative and positive controls A. Light microscope images of the indicated samples at day 1 of BSD selection. C. Light microscope Images of the indicated samples at the end of BSD selection shows a random integration of the cassette into the genome in the absence of Cas9 plasmid. Scale bar = 250 µm

4.4.7. Successful targeting of TET2 in LN18 cell line using wild type Cas9

After our first unsuccessful attempt of targeting *TET2* in LN18 cell line we decided to target the cells with wild type Cas9 with the two gRNAs individually in order to increase the possibilities of on-target events. Thus, we targeted the cells with donor plasmid, Guide A or Guide B and the Cas9 plasmid (Fig. 4.12 A). This time the BSD selection was applied for 14 days beginning 5 days after transfection of the cells, to allow the cells to recover after the transfection and also to dilute the donor plasmid that has not been integrated into the genome (Fig. 4.12 B and C). After the initial selection with BSD we performed FACS for the selection of the GFP positive cells and sorted single cells into 96-well plates for clonal expansion (Fig. 4.12 D).

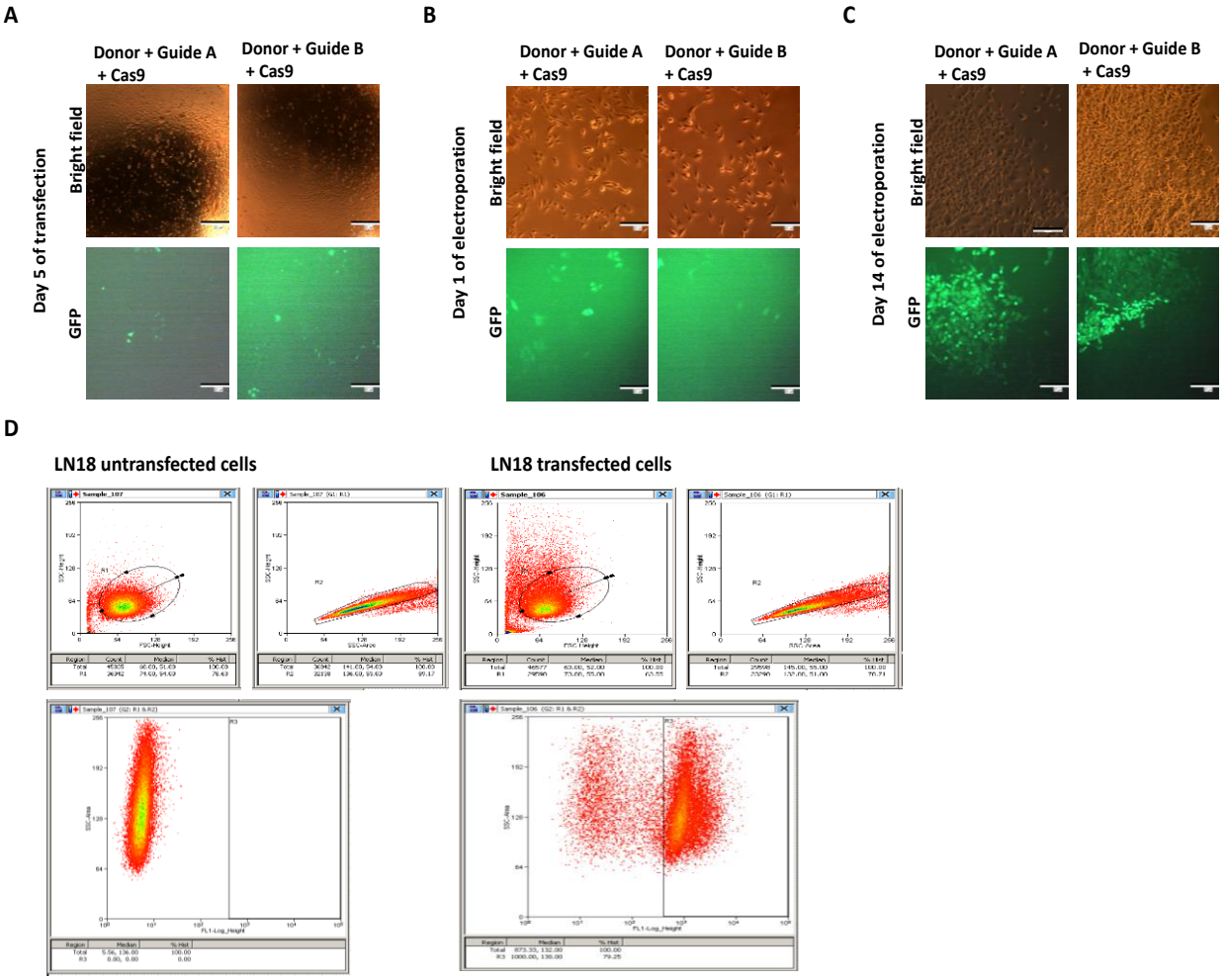


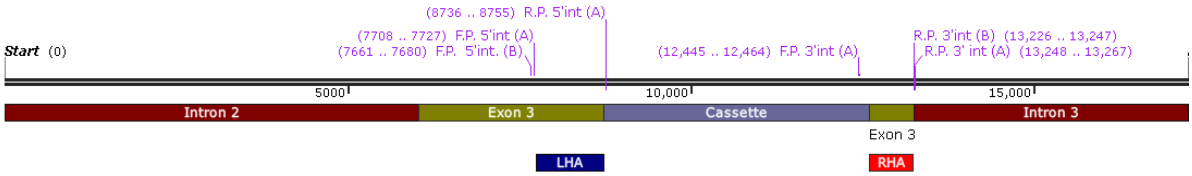
Figure 4.12: Electroporation and selection of the LN18 transfected cells with wild type Cas9 plasmid and two different gRNAs.

A-C. Light microscope images (bright field and GFP) of targeted and non-targeted samples at the indicated time points (5 days post transfection, 1 day and 14 days post BSD treatment) showing the successful integration of the plasmid into most of the cells. Scale bar = 250 µm D. FACS of the CRISPR/CAS9 transfected cells with Guide A. The threshold for FACS selection was set using LN18 untransfected cells and GFP positive cells were selected based on the GFP intensity. 79 % of the cells were GFP positive.

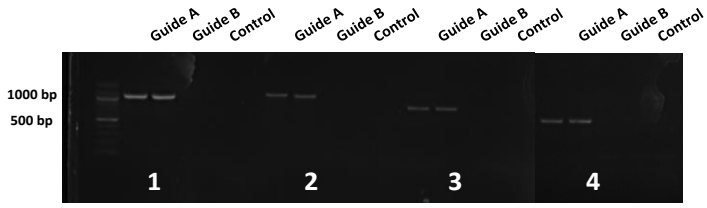
We then proceeded with screening of the pool of GFP positive cells in order to optimise the primers and ensure that correct integration had occurred, before proceeding with clonal expansion of cells. Thus, we used two sets of primers for each 3' and 5' integration PCRs (Fig. 4.13 A). We performed PCR in the population of cells targeted with Guide A, Guide B and in the non-transfected cells as a negative control. We observed that only the cells targeted with Guide A displayed a positive clear band at the correct size in all

primer combinations used (Fig. 4.13 B). We then sent these PCR products for sequencing to ensure that we were amplifying the region of interest (Data not shown).

A



B



		Expected size (bp)
1	5' int (A)	1048
2	5' int (B)	1095
3	3' int (A)	823
4	3' int (A)	803

Figure 4.13: Initial screening of the pool of cells for 5' and 3' integration.

A. SnapGene map showing the areas of interest and the integration primers which are designed to amplify the area outside the arms of homology into the cassette on both sides. B. Gel of PCR products using 2 sets of primers for 5' and 3' integration and the indicated samples; each sample was loaded in duplicates. The table shows the samples and the primers used.

Our initial PCR screening revealed that only clones targeted with Guide A, displayed correct integration of the cassette into the locus of interest. Therefore, we proceeded with clonal expansion and screening of individual clones in these cells. We first screened for 3' and 5' integration using the same primer sets optimised previously, using the pool of cells as a positive control and non-transfected cells as a negative control (Fig. 4.14 A and 4.15 A). We then isolated the PCR products and sequenced them in order to ensure the correct integration of the cassette. The chromatogram of the sequencing results of clone 15 indicates the correct junctions between the genomic sequence and the cassette; Cassette-RHA-Intron 4 for 3' integration (Fig. 4.14 B) and Exon 3- LHA- Cassette for 5' integration (Fig. 4.15 B). Chromatograms also indicate that the results derived from a single clone and there is no contamination in the PCR product (Fig. 4.14 B and 4.15 B).

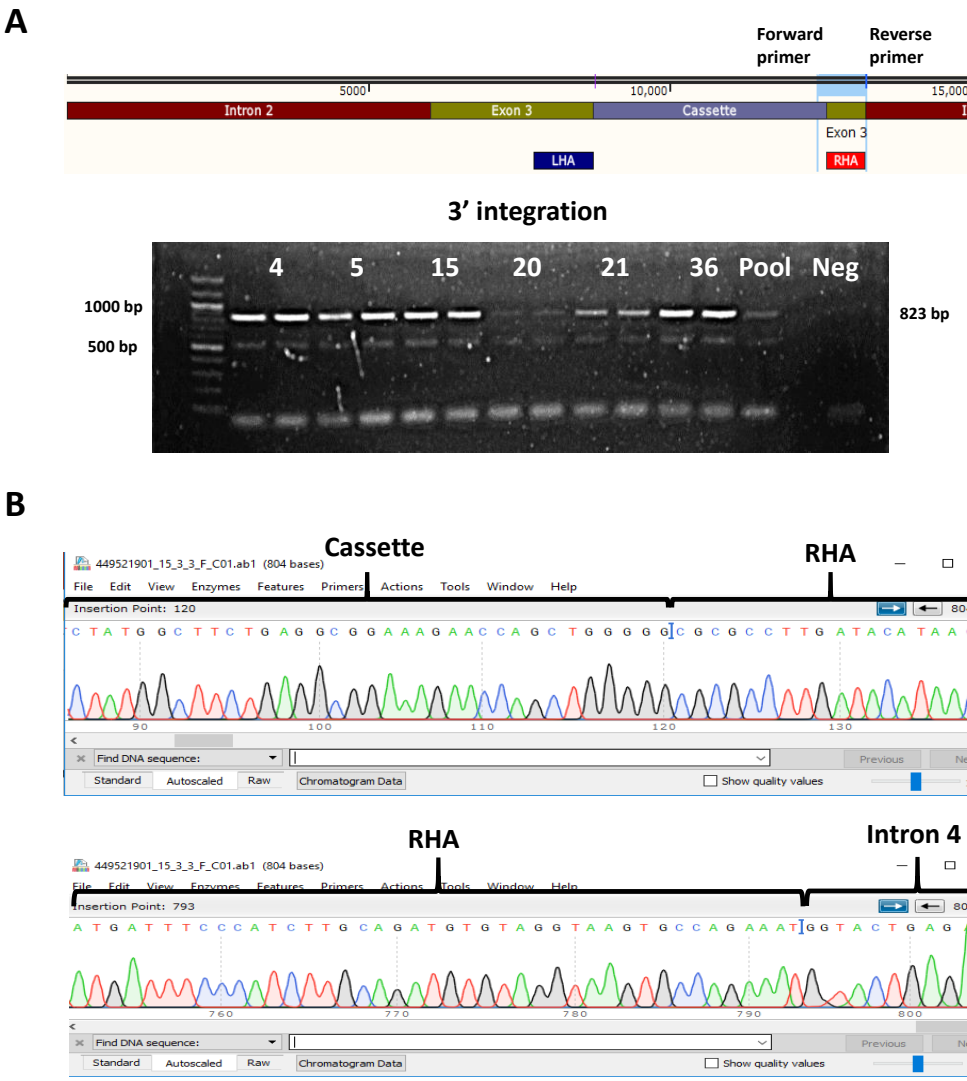


Figure 4.14: Screening of the individual clones for 3' integration.

A. SnapGene map showing the area of the genome that was amplified, a gel with 6 indicative samples screened (loaded in two lines each), all the samples show a positive band at 823 bp. B. Sequencing results for the clone number 15 shows correct 3' integration of the cassette.

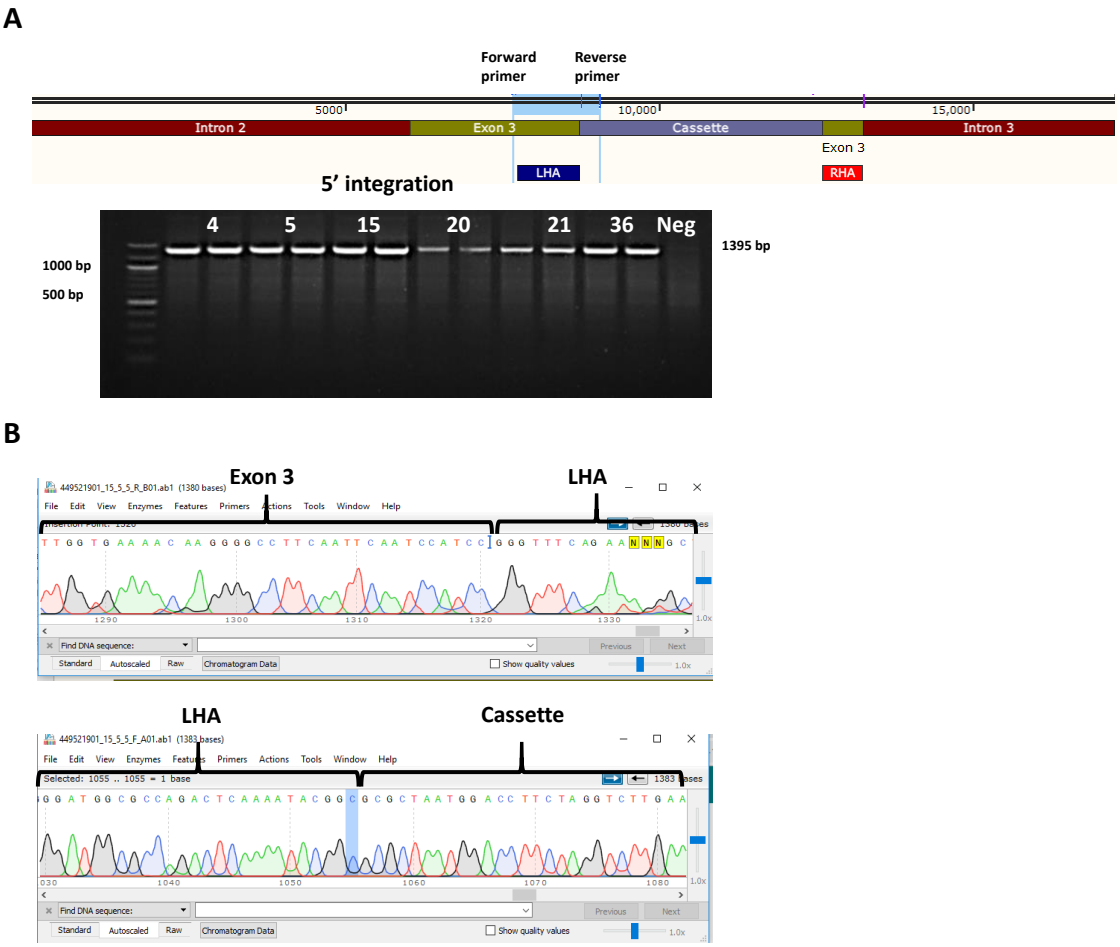


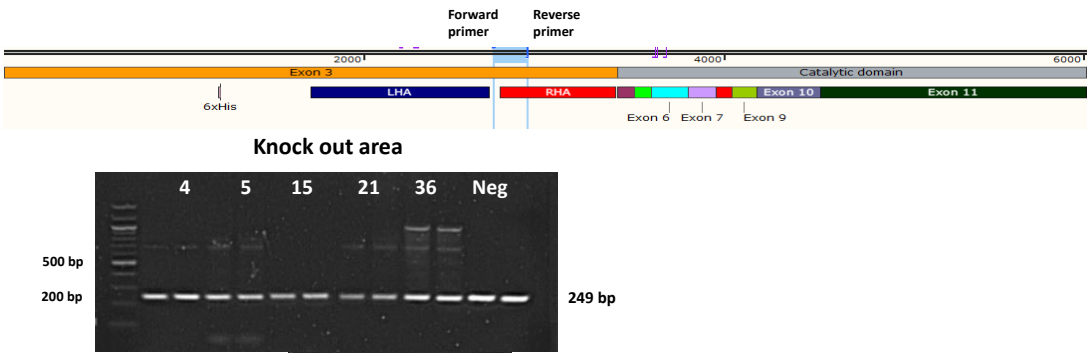
Figure 4.15: Screening of the individual clones for 5' integration.

A. SnapGene map showing the area of the genome that was amplified. A gel with 6 indicative samples screened (loaded in two individual lines each), all the samples show a positive band at 1395 bp. B. Sequencing results for the clone number 15 shows correct 5' integration of the cassette.

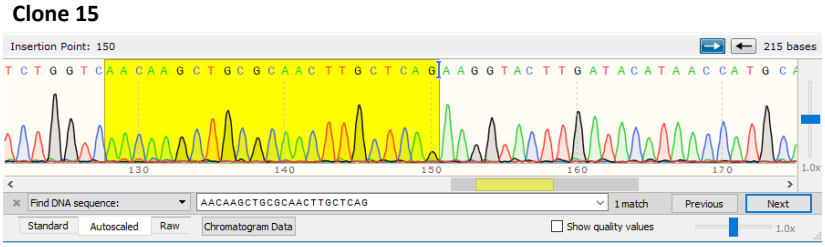
We also aligned the sequencing results with the expected TET2 sequence after the integration of the cassette using the Clustal Omega online tool; which showed 100% alignment for both 3’ and 5’ integrations for clone number 15

In total, we screened 90 clones with 32 % of the clones displaying both 3’ and 5’ integration of the cassette. We then proceeded to further screen the clones that showed positive 5’ and 3’ integrations, in order to examine whether both of the *TET2* alleles were targeted. We performed PCR of the region of deletion; the sgRNA recognition site. Our PCR results revealed that all of the clones examined were heterozygous, as they all displayed a positive band at *TET2* deletion area (Fig. 4.16 A). We then proceeded with the sequencing of the untargeted allele in all the clones to identify any possible deletions or indels in the sequence. Interestingly, alignment of the sequencing results of clone 15 showed the absence of two nucleotides at exon 3 targeted locus of *TET2* gene (Fig. 4.16 B & C).

A



B



C

Knock out area

```
########################################
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 314
# Identity:      95/314 (30.3%)
# Similarity:    95/314 (30.3%)
# Gaps:         217/314 (69.1%)
# Score: 549.5
#
#
########################################

EMBOSS_001         1 ------------------------------------------------      0

EMBOSS_001         1 NNNNNNNNNNNNNNNNNNNNNCCNAAGCAAGANCTTCTTCNCAGGTGCTT     50

EMBOSS_001         1 ------------------------------------------------      0

EMBOSS_001        51 TCAAGAACAGGAGCAGAAGTCACAACAAGCTTCAGTTCTACAGGGATATA    100

EMBOSS_001         1 ------------------TGTCTGGTCAACAAGCTGCGCAACTTGCTCAG     32
                                       ||||||||||||||||||||||||||||||
EMBOSS_001       101 AAAATAGAAACCAAGATATGTCTGGTCAACAAGCTGCGCAACTTGCTCAG    150

EMBOSS_001        33 CAAAGGTACTTGATACATAACCATGCAAATGTTTTTCCTGTGCCTGACCA     82
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       151 CAAGGTACTTGATACATAACCATGCAAATGTTTTTCCTGTGCCTGACCA    198

EMBOSS_001        83 GGGAGGAAGTCACACTCAGACCCCTCCCCAGAAGGACACTCAAAAGCATG    132
                     ||||||||||||||..
EMBOSS_001       199 GGGAGGAAGTCACACNN--------------------------------    215

EMBOSS_001       133 CTGCTCTAAGGTGGCATCTCTTACAGAAGCAAGAACAGCAGCAAACACAG    182

EMBOSS_001       216 ------------------------------------------------    215

EMBOSS_001       183 CAACCCCAAACTGA    196

EMBOSS_001       216 --------------    215
```

Figure 4.16: Screening of the positive clones for the deletion of the genomic area in TET2 gene.

A. SnapGene map shows primers recognition area at *TET2* transcript. The gel shows that all samples examined had a positive band at 249 bp indicating heterozygous disruption of *TET2* gene (each sample was loaded in two individual lanes). B. Chromatogram of the sequencing of the PCR product of clone 15 C. Clustal Omega alignment of clone 15 untargeted allele (bottom strand) with TET2 wild type gene (top strand) showing the absence of two nucleotides in Clone 15 (red rectangle).

4.4.8. Generation of TET2 heterozygous knockout clone that leads to truncated TET2 protein

After the identification of the deletion of two nucleotides in clone 15 we proceeded with a PCR of cDNA of both clone 15 and untargeted cells. We amplified the area from exon 3 to exon 6 (Fig. 4.17 A). Then, the PCR product was sequenced to confirm our previous observations in clone 15. Sequencing results confirm the absence of the two nucleotides at the locus of interest in clone 15 whereas the untargeted wild-type cells did not show any deletion (Fig. 4.17 B and C).

C

Wild type sequence

```
#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1474
# Identity:    1019/1474 (69.1%)
# Similarity:  1019/1474 (69.1%)
# Gaps:         337/1474 (22.9%)
# Score: 5573.5
#
#
#=======================================

EMBOSS_001        1 CCCAAAGCAACAATGATCAGCAAAGAGAAGGATCATTCTTTGGCCA----     46

EMBOSS_001        1 ----------------------------------------------NNNN      4

EMBOSS_001       47 ----------------GACTAAAGTGGAAGAATGTTTTCATGGTGAAAAT     80
                                    .||||...|||||| ||||||||.||||||||||
EMBOSS_001        5 NNNNNNNNNNNNNNNNNACTANNNTGGAAG-ATGTTTTCNTGGTGAAAAT     53

EMBOSS_001       81 CAGTATTCAAAATCAAGCGAGTTCGAGACTCATAATGTCCAAATGGGACT    130
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       54 CAGTATTCAAAATCAAGCGAGTTCGAGACTCATAATGTCCAAATGGGACT    103

EMBOSS_001      131 GGAGGAAGTACAGAATATAAATCGTAGAAATTCCCCTTATAGTCAGACCA    180
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      104 GGAGGAAGTACAGAATATAAATCGTAGAAATTCCCCTTATAGTCAGACCA    153

EMBOSS_001      181 TGAAATCAAGTGCATGCAAAATACAGGTTTCTTGTTCAAACAATACACAC    230
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      154 TGAAATCAAGTGCATGCAAAATACAGGTTTCTTGTTCAAACAATACACAC    203

EMBOSS_001      231 CTAGTTTCAGAGAATAAAGAACAGACTACACATCCTGAACTTTTTGCAGG    280
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      204 CTAGTTTCAGAGAATAAAGAACAGACTACACATCCTGAACTTTTTGCAGG    253

EMBOSS_001      281 AAACAAGACCCAAAACTTGCATCACATGCAATATTTTCCAAATAATGTGA    330
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      254 AAACAAGACCCAAAACTTGCATCACATGCAATATTTTCCAAATAATGTGA    303

EMBOSS_001      331 TCCCAAAGCAAGATCTTCTTCACAGGTGCTTTCAAGAACAGGAGCAGAAG    380
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      304 TCCCAAAGCAAGATCTTCTTCACAGGTGCTTTCAAGAACAGGAGCAGAAG    353

EMBOSS_001      381 TCACAACAAGCTTCAGTTCTACAGGGATATAAAAATAGAAACCAAGATAT    430
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      354 TCACAACAAGCTTCAGTTCTACAGGGATATAAAAATAGAAACCAAGATAT    403
```

```
EMBOSS_001      431 GTCTGGTCAACAAGCTGCGCAACTTGCTCAGCAAAGGTACTTGATACATA    480
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      404 GTCTGGTCAACAAGCTGCGCAACTTGCTCAGCAAAGGTACTTGATACATA    453

EMBOSS_001      481 ACCATGCAAATGTTTTTCCTGTGCCTGACCAGGGAGGAAGTCACACTCAG    530
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      454 ACCATGCAAATGTTTTTCCTGTGCCTGACCAGGGAGGAAGTCACACTCAG    503

EMBOSS_001      531 ACCCCTCCCCAGAAGGACACTCAAAAGCATGCTGCTCTAAGGTGGCATCT    580
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      504 ACCCCTCCCCAGAAGGACACTCAAAAGCATGCTGCTCTAAGGTGGCATCT    553

EMBOSS_001      581 CTTACAGAAGCAAGAACAGCAGCAAACACAGCAACCCCAAACTGAGTCTT    630
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      554 CTTACAGAAGCAAGAACAGCAGCAAACACAGCAACCCCAAACTGAGTCTT    603

EMBOSS_001      631 GCCATAGTCAGATGCACAGGCCAATTAAGGTGGAACCTGGATGCAAGCCA    680
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      604 GCCATAGTCAGATGCACAGGCCAATTAAGGTGGAACCTGGATGCAAGCCA    653

EMBOSS_001      681 CATGCCTGTATGCACACAGCACCACCAGAAAACAAAACATGGAAAAAGGT    730
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      654 CATGCCTGTATGCACACAGCACCACCAGAAAACAAAACATGGAAAAAGGT    703

EMBOSS_001      731 AACTAAGCAAGAGAATCCACCTGCAAGCTGTGATAATGTGCAGCAAAAGA    780
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      704 AACTAAGCAAGAGAATCCACCTGCAAGCTGTGATAATGTGCAGCAAAAGA    753

EMBOSS_001      781 GCATCATTGAGACCATGGAGCAGCATCTGAAGCAGTTTCACGCCAAGTCG    830
                    .||||||||||||||||||||||||||||||||||||||||||||||.|||||
EMBOSS_001      754 NCATCATTGAGACCATGGAGCAGCATCTGAAGCAGTTTCACGCCNAGTCG    803

EMBOSS_001      831 TTATTTGACCATAAGGCTCTTACTCTCAAATCACAGAAGCAAGTAAAAGT    880
                    ||||||||||.|..|||||||||||||||||..|||||||||||||||||||
EMBOSS_001      804 TTATTTGACCATNANNCTCTTACTCTCAAATCACAGANNCAAGTAAAAGT    853

EMBOSS_001      881 TGAAATGTCAGGGCCAGTCACAGTTTTGACTAGACAAACCACTGCTGCAG    930
                    |||||||||..|.|..|.|||||||||||||||||||||||||||||||||||
EMBOSS_001      854 TGAAATGTCNNGNCNNGNCACAGTTTTGACTANACAAACCACTGCTGCAG    903

EMBOSS_001      931 AACTTGATAGCCACACCCCAGCTTTAGAGCAGCAAACAACTT-CTTCAGA    979
                    ||||||||||.|||||||||||||||||||||||||||||.|||| |.|....
EMBOSS_001      904 AACTTGATAGCCNCACCCCAGCTTTAGAGCAGCAAACNACTTNCNTTCNG    953

EMBOSS_001      980 AAAGACACCAACCAAAAGAACAGCTGCTTCTGTTCTCAATAATTTTA-TA   1028
                    |||.||..|.|||||||||||||||||||||||||||||..||||||| ..
EMBOSS_001      954 AAANACNNCNACCAAAAGAACAGCTGCTTCTGTTCTCANNAATTTTANNN   1003

EMBOSS_001     1029 GAGTCACCTT-CCAAA-TTACTAGATACTCCTATAAAAAATTT--ATTGG   1074
                    .||||.|||| |||| |..|||||.|||||.|.||||||.||  .|.||
EMBOSS_001     1004 NAGTCNCCTTNCCAAANTNNCTAGANACTCCNANAAAAAANTTNNNTNGG   1053
```

Clone 15 sequence

```
#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1453
# Identity:    1088/1453 (74.9%)
# Similarity:  1088/1453 (74.9%)
# Gaps:         325/1453 (22.4%)
# Score: 4939.5
#
#
#=======================================

EMBOSS_001        1 CCCAAAGCAACAATGATCAGCAAAGAGAAGGATCAT-------TCTTTGG     43
                                                          |.||||
EMBOSS_001        1 -----------------------------------NNNNNNNTNTTTG-     13

EMBOSS_001       44 CCAGACTAAAGTGGAAGAATGTTTTCATGGTGAAAATCAGTATTCAAAAT     93
                     ...|||||||||| .|||||||||||||||||||||||||||||||||||
EMBOSS_001       14 -NNNACTAAAGTGGA--NATGTTTTCATGGTGAAAATCAGTATTCAAAAT     60

EMBOSS_001       94 CAAGCGAGTTCGAGACTCATAATGTCCAAATGGGACTGGAGGAAGTACAG    143
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       61 CAAGCGAGTTCGAGACTCATAATGTCCAAATGGGACTGGAGGAAGTACAG    110

EMBOSS_001      144 AATATAAATCGTAGAAATTCCCCTTATAGTCAGACCATGAAATCAAGTGC    193
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      111 AATATAAATCGTAGAAATTCCCCTTATAGTCAGACCATGAAATCAAGTGC    160

EMBOSS_001      194 ATGCAAAATACAGGTTTCTTGTTCAAACAATACACACCTAGTTTCAGAGA    243
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      161 ATGCAAAATACAGGTTTCTTGTTCAAACAATACACACCTAGTTTCAGAGA    210

EMBOSS_001      244 ATAAAGAACAGACTACACATCCTGAACTTTTTGCAGGAAACAAGACCCAA    293
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      211 ATAAAGAACAGACTACACATCCTGAACTTTTTGCAGGAAACAAGACCCAA    260

EMBOSS_001      294 AACTTGCATCACATGCAATATTTTCCAAATAATGTGATCCCAAAGCAAGA    343
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      261 AACTTGCATCACATGCAATATTTTCCAAATAATGTGATCCCAAAGCAAGA    310

EMBOSS_001      344 TCTTCTTCACAGGTGCTTTCAAGAACAGGAGCAGAAGTCACAACAAGCTT    393
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      311 TCTTCTTCACAGGTGCTTTCAAGAACAGGAGCAGAAGTCACAACAAGCTT    360

EMBOSS_001      394 CAGTTCTACAGGGATATAAAAATAGAAACCAAGATATGTCTGGTCAACAA    443
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      361 CAGTTCTACAGGGATATAAAAATAGAAACCAAGATATGTCTGGTCAACAA    410
```

```
EMBOSS_001      444 GCTGCGCAACTTGCTCAGCAAAGGTACTTGATACATAACCATGCAAATGT    493
                    ||||||||||||||||||  ||||||||||||||||||||||||||||||||
EMBOSS_001      411 GCTGCGCAACTTGCTCAG--AAGGTACTTGATACATAACCATGCAAATGT    458

EMBOSS_001      494 TTTTCCTGTGCCTGACCAGGGAGGAAGTCACACTCAGACCCCTCCCCAGA    543
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      459 TTTTCCTGTGCCTGACCAGGGAGGAAGTCACACTCAGACCCCTCCCCAGA    508

EMBOSS_001      544 AGGACACTCAAAAGCATGCTGCTCTAAGGTGGCATCTCTTACAGAAGCAA    593
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      509 AGGACACTCAAAAGCATGCTGCTCTAAGGTGGCATCTCTTACAGAAGCAA    558

EMBOSS_001      594 GAACAGCAGCAAACACAGCAACCCCAAACTGAGTCTTGCCATAGTCAGAT    643
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      559 GAACAGCAGCAAACACAGCAACCCCAAACTGAGTCTTGCCATAGTCAGAT    608

EMBOSS_001      644 GCACAGGCCAATTAAGGTGGAACCTGGATGCAAGCCACATGCCTGTATGC    693
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      609 GCACAGGCCAATTAAGGTGGAACCTGGATGCAAGCCACATGCCTGTATGC    658

EMBOSS_001      694 ACACAGCACCACCAGAAAACAAAACATGGAAAAAGGTAACTAAGCAAGAG    743
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      659 ACACAGCACCACCAGAAAACAAAACATGGAAAAAGGTAACTAAGCAAGAG    708

EMBOSS_001      744 AATCCACCTGCAAGCTGTGATAATGTGCAGCAAAAGAGCATCATTGAGAC    793
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      709 AATCCACCTGCAAGCTGTGATAATGTGCAGCAAAAGAGCATCATTGAGAC    758

EMBOSS_001      794 CATGGAGCAGCATCTGAAGCAGTTTCACGCCAAGTCGTTATTTGACCATA    843
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      759 CATGGAGCAGCATCTGAAGCAGTTTCACGCCAAGTCGTTATTTGACCATA    808

EMBOSS_001      844 AGGCTCTTACTCTCAAATCACAGAAGCAAGTAAAAGTTG-AAATGTCA-G    891
                    |||||||||||||||||||||||||||||||||||||||| |||||||| |
EMBOSS_001      809 AGGCTCTTACTCTCAAATCACAGAAGCAAGTAAAAGTTGAAAATGTCAGG    858

EMBOSS_001      892 GGCCAGTCACAG-TTTTGACTAGACAAACCACTGCTGCAG-AACTTG-AT    938
                    |||||||||||| |||||||||||||||||||||||||| |||||| ||
EMBOSS_001      859 GGCCAGTCACAGTTTTTGACTAGACAAACCACTGCTGCAGAAACTTGAAT    908

EMBOSS_001      939 AG-CCACA-CCCCAGCTTT-AGAGCAGC-AAACAACTT--CTTCAGAAAA    982
                    || ||||| |||||||||| |||||||| ||||||| .|||||.||||
EMBOSS_001      909 AGCCCACACCCCCAGCTTTAAGAGCAGCAAAACAACTTTCTTTCAAAAAA    958

EMBOSS_001      983 GACA-CCAACCAAAAGAACAG-CTGC-TTCTG-TTCTCAATAA-TTTTAT   1027
                    |||| |||.|||||.|||||| |||| |||| |||||||||| ||||||
EMBOSS_001      959 GACACCCACCCAAAGGAACAGCCTGCTTTCTGTTTCTCAATAATTTTTAT   1008

EMBOSS_001     1028 AG-AGT-CACCTT--CCAAATTACT-AGATACTCC--TATAAAAAATTT-   1069
                    || ||| |||||| |.|||||||| |||.|.|.|| .||||||||||
EMBOSS_001     1009 AGAAGTCCACCTTTCCAAAATTACTAAGANACNCCCTAATAAAAAATTTA   1058
```

Figure 4.17: RT PCR in the TET2 integration area confirming the absence of two nucleotides in clone 15.

A. SnapGene map shows *TET2* gene and the area of PCR amplification. The gel image shows clone 15 (loaded in two lanes) and the negative control have the same band at 1394 bp confirming that clone 15 is heterozygous. B. The chromatogram of the sequencing results of the PCR bands in the negative control and clone 15 show the absence of two nucleotides in the untargeted allele of clone 15. C. Alignment of the sequencing results of the RT-PCR in wild type and in Clone 15 confirm the absence of two nucleotides in clone 15; bottom strand represents the sequenced PCR product.

Finally, for the last step of clone characterisation, we used the online translation tool (https://web.expasy.org/translate/) to identify whether the deletion can lead to any disruptions in TET2 protein. Indeed, our results show that the deletion of these two nucleotides led to a stop codon at TET2 protein before the start of the catalytic domain (Fig. 4.18). Our sequencing results confirmed that heterozygous TET2 knockout clone 15 leads to the translation of a truncated TET2 protein. In order to confirm our sequencing results, we performed western blot of TET2 protein in clone 15. We used three biological replicates of each wild type and mutant LN18 cells and showed the presence of TET2 protein in the non-transfected wild type LN18 cells only (Fig. 4.18 B).

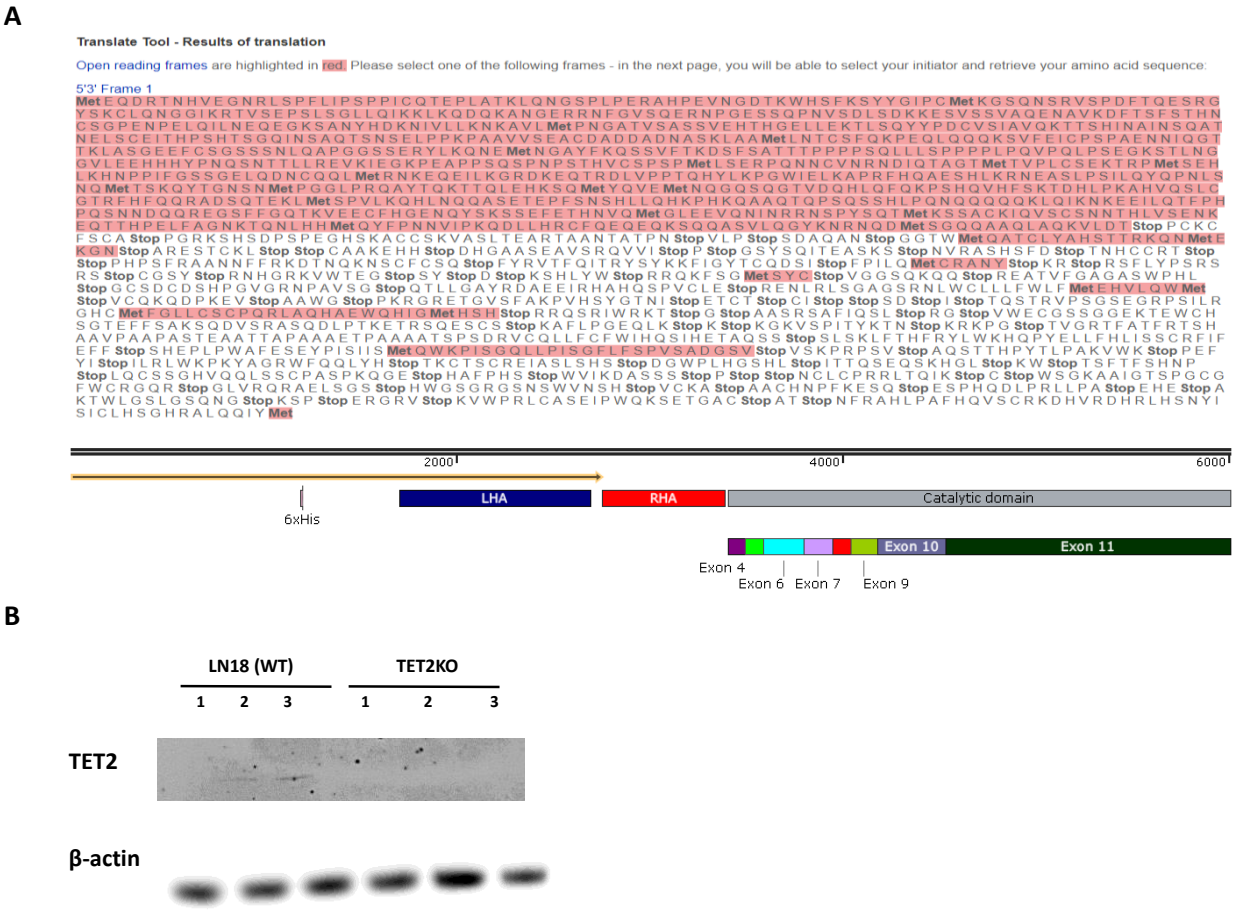


Figure 4.18: Clone 15 untargeted allele leads to truncated TET2 protein.

A. Results from the translation tool used with the sequencing results from the untargeted allele in clone 15. The two nucleotides missing resulted in a stop codon in TET2 protein before the catalytic domain. B. Western blot of TET2 in LN18 wild type cells and LN18 TET2 KO cells (Clone 15) showing the presence of TET2 in the control cells only. β-actin was used as a reference control (n = 3).

4.4.9. Mutant TET2 GBM cell line displays significantly lower levels of oxi-mCs

Once we confirmed that our clone had a missense mutation leading to a stop codon before the start of the catalytic domain we then decided to functionally characterise this clone. We first performed immunocytochemistry for 5hmC/5caC epigenetic marks. This experiment would give us insights into whether TET2 is responsible for the oxidation of 5mC in GBM cell lines. We will be referring to clone 15 as TET2 KO clone for the rest of the study. In this experiment a clone in which only one allele of TET2 was targeted (TET2 -/+) was also used.

We used our sensitive and optimised immunocytochemistry protocol for the identification of 5hmC/5caC marks. However, we had to factor in that our targeted cells were already GFP positive and that the GFP signal could possibly interact with the secondary Alexa fluorophore 488-HRP conjugated antibody we used previously for the detection of 5caC. Therefore, the HCL denaturation step was crucial in this experiment in order to ensure the denaturation of GFP protein. Our experiments confirmed our initial hypothesis about the role of TET2 in active DNA demethylation in GBM cell lines. We showed that, in both TET2-/+ and TET2 KO cells, 5hmC levels were drastically reduced, with the levels being almost undetectable in TET2 KO samples (Fig. 4.19 A). Quantification of 5hmC signal in three individual experiments showed that 5hmC levels were reduced by almost 50% in TET2-/+ ($p < 0.0001$) and were not detectable in TET2 KO ($p < 0.0001$) (Fig. 4.19 B). Regarding 5caC levels we observed similar effects in both samples and signal quantification showed that 5caC levels were significantly reduced in TET2 KO samples ($p < 0.0001$) (Fig. 4.19 A and B).

We next sought to determine whether these changes were due to the inactivation of TET2 protein or if they were impaired by changes in the levels of key genes involved in the active DNA demethylation pathway. Hence, we performed qPCR for the relative quantification of *TET1/TET2/TDG* transcript levels compared to LN18 non-targeted cells. We confirmed that TET2 was undetectable in TET2 KO ($p < 0.001$) and 100 fold lower in TET2 -/+ ($p < 0.0001$) (Fig. 4.19 C). Interestingly, there were no changes in the transcript levels of either *TET1* or *TDG* (Fig. 4.19 C). These results lead to the assumption that TET2 alone is responsible for the oxidation of 5mC in this cell line.

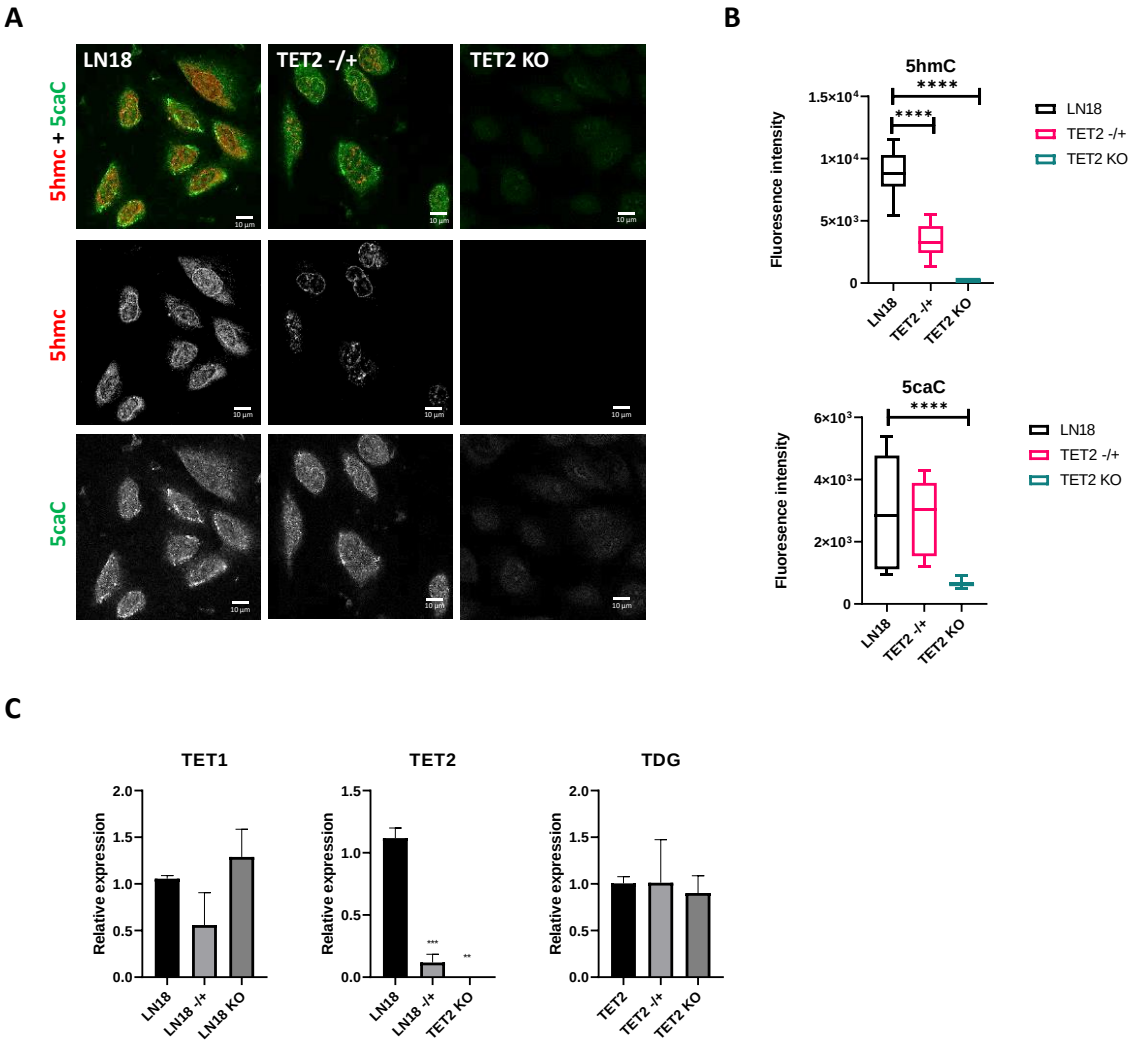


Figure 4.19: TET2 KO leads to almost undetectable immunocytochemically levels of 5hmC/5caC in parallel with no alteration in *TET1* and *TDG* transcripts.

A. Co-detection of 5caC/5hmC in LN18 untargeted cells, LN18 -/+ targeted at only one allele and in TET2 KO cells. Merged views and individual channels are shown. The cell cultures were stained in parallel at the same experimental conditions and were imaged at identical settings. Scale bar = 10 µm. B. Quantification of 5hmC and 5caC signals in the indicated samples. Mean values of the average intensities of each individual cell in each cell line are shown in three individual experiments (n = 3). Error bars are shown as ± SD. C. Relative quantification of *TET1/2 TDG* transcripts in the indicated samples showing that TET2 KO does not affect transcript levels of *TET1* and *TDG*. Mean values of three individual experiments are shown (n = 3). Error bars are shown as ± SD. Statistical analysis was performed using un-paired t test; p ≤ 0.05; * p < 0.05; ** p < 0.01; *** p < 0.001 and **** p < 0.0001.

In order to validate our immunocytochemistry results we next proceeded with the quantification of all oxi-mC derivatives using 2D-UPLC-MS/MS. Mass spectrometry analysis would give us information on the actual number of modified cytosines in the samples. Our results showed that there are no changes in the levels of 5mC and 5hmU in all the samples tested irrespective of TET2 enzymatic activity (Fig. 4.20 A and C). Nevertheless, 5hmC levels are drastically reduced from 0.02 cytosine modifications per 1000 DNA bases to 0.003 in TET2 -/+ ($p < 0.0001$) and 0.001 modifications ($p < 0.0001$) in TET2 KO (Fig. 4.21 B). Moreover, 5fC levels were also significantly reduced in TET2 KO samples; from 0.3 DNA modifications per $1x10^6$ DNA bases to 0.09 ($p = 0.04$) (Fig. 4.20 D). Surprisingly, in contrast to our immunocytochemistry experiments, 5caC levels were slightly decreased in TET2 KO from 22 cytosine modifications per $1x10^9$ DNA bases to 17, however this reduction was not statistically significant ($p = 0.33$) (Fig. 4.20 E). Finally, 8-oxodG levels were comparable across the samples indicating that levels of oxidative damage were similar in all the cell lines tested (Fig. 4.20 F). Our results show that inactivation of TET2 protein leads to drastically reduced levels of active DNA demethylation marks on DNA.

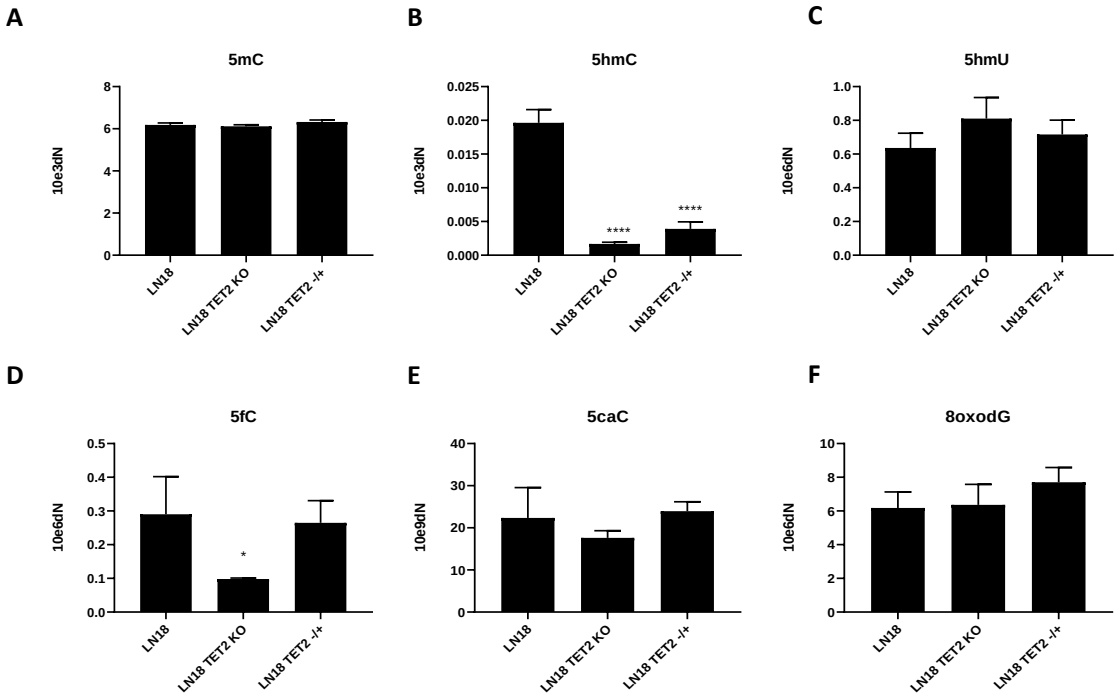


Figure 4.20: Mass spectrometry determined levels of DNA methylation marks in TET2 KO and TET2-/+ samples.

The results of 2D-UPLC–MS/MS quantification of the indicated forms of modified cytosine. A. 5mC and C. 5hmU levels remained stable after the depletion of TET2. B. 5hmC levels are drastically reduced in both LN18 KO sample and TET2-/+. D. 5fC levels are reduced in TET2 KO samples. E. 5caC levels showed a slight decrease in TET2 KO. F. The levels of oxidative DNA damage (8-oxodG) are comparable between the samples. Mean values of three individual experiments are shown (n = 3). Experimental error is shown as ± SD. Statistical analysis was performed using un-paired t test; $p \leq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ and **** $p < 0.0001$.

4.4.10. Non-functional TET2 protein does not affect cell proliferation in different culture conditions

We next examined the growth rate of the TET2 mutant cells in comparison with the control, non-transfected cells. We monitored cell growth for 8 days in TET2 KO compared to LN18 (Fig. 4.21). We did not observe any significant difference in cell number at any of the time points analysed. Moreover, the doubling time of the cells was also calculated based on the cell numbers obtained; for LN18 doubling time was 31.43 hours while for TET2 KO samples was 31.6 hours.

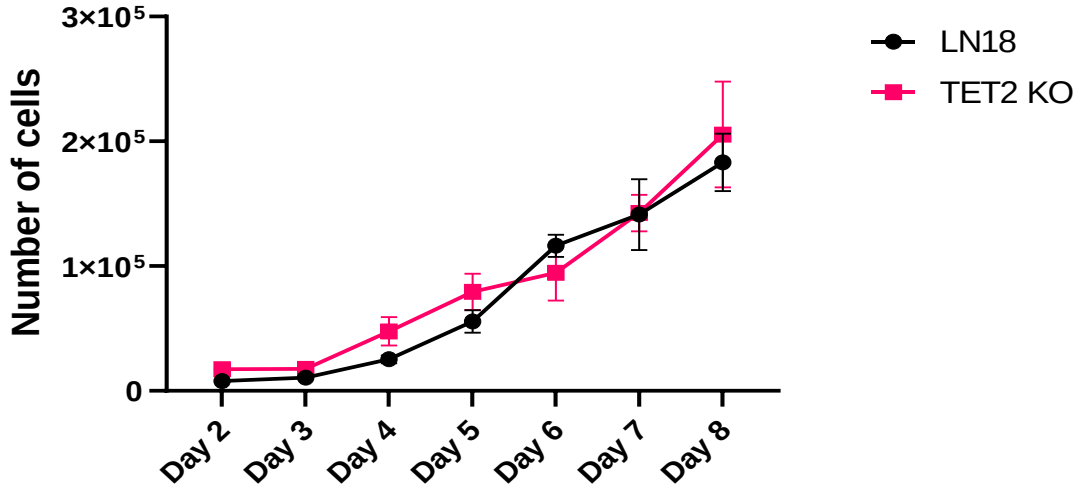


Figure 4.21: TET2 KO does not impair cell proliferation in adherent cell culture conditions.

Growth curve shows cell proliferation rates in LN18 and TET2 KO mutants. Mean values of eight individual experiments are shown (n = 8).Experimental error is shown as mean ± SD. Statistical analysis was performed using un-paired two tailed Student's t-tests; LN18 and TET2 KO growth differences were not significant NS $p \geq 0.05$.

The growth rate of cells was also calculated when cells were cultured in neurobasal medium for cancer stem cells enrichment. Similar to normal adherent cell culture conditions, we did not observe any significant differences in the growth rate of the two populations (Fig. 4.22). Doubling

times for LN18 grown in neurobasal medium was 48 hours and for TET2 KO was 45 hours; this difference was not statistically significant (p = 0.6). Our results show that TET2 enzymatic inactivity is unlikely to cause any disruption in cell proliferation *in vitro*.

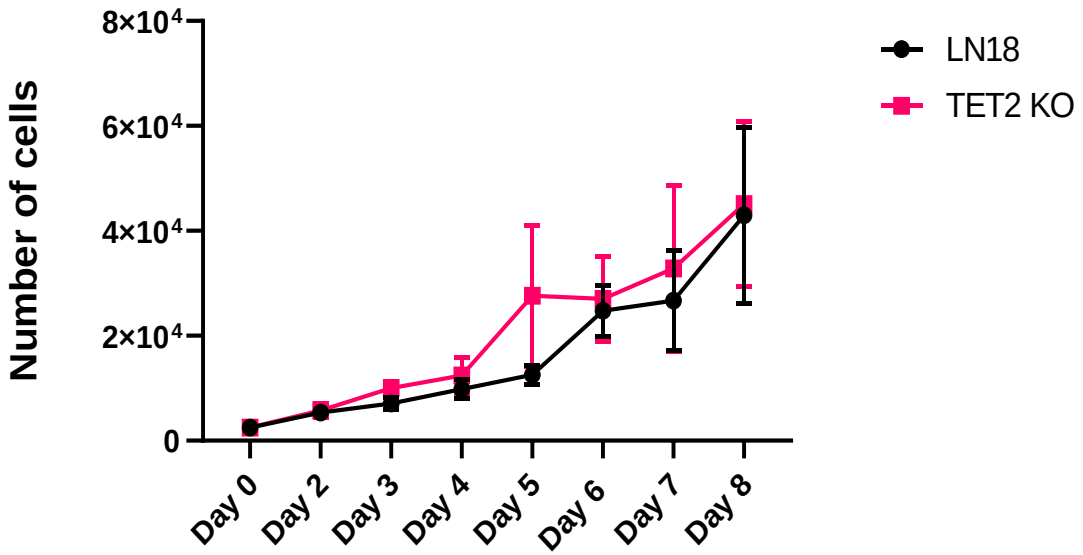


Figure 4.22: TET2 KO does not impair cell proliferation in neurobasal cell culture conditions.

Growth curve illustrates cell proliferation rates in LN18 and TET2 KO mutants. Mean values of four individual experiments are shown (n = 4). Experimental error is shown as mean ± SD. Statistical analysis was performed using un-paired two tailed Student's t-tests; LN18 and TET2 KO growth differences were not statistically significant NS $p \geq 0.05$.

4.4.11. TET2 KO mutants display unique expression of key cancer stem cell markers and reduced tumorigenic ability in culture

We next examined the transcript levels of key genes involved in the pathogenesis of GBM and in the maintenance of cancer stem cells. We observed that some genes including *EGFR, GFAP, CD133,* Notch homolog 1, translocation-associated (*NOTCH1) and* β-tubulin III did not show any alteration in their relative expression levels in TET2 mutant cells compared to LN18 non-targeted cells (Fig. 4.23). However, significantly lower levels (10 fold decreased) were found in *PDGF, SOX2, NANOG* and Glutamate Aspartate Transporter 1 (*GLAST)* ($p < 0.0001$, $p = 0.006$, $p = 0.02$, $p = 0.001$ respectively).

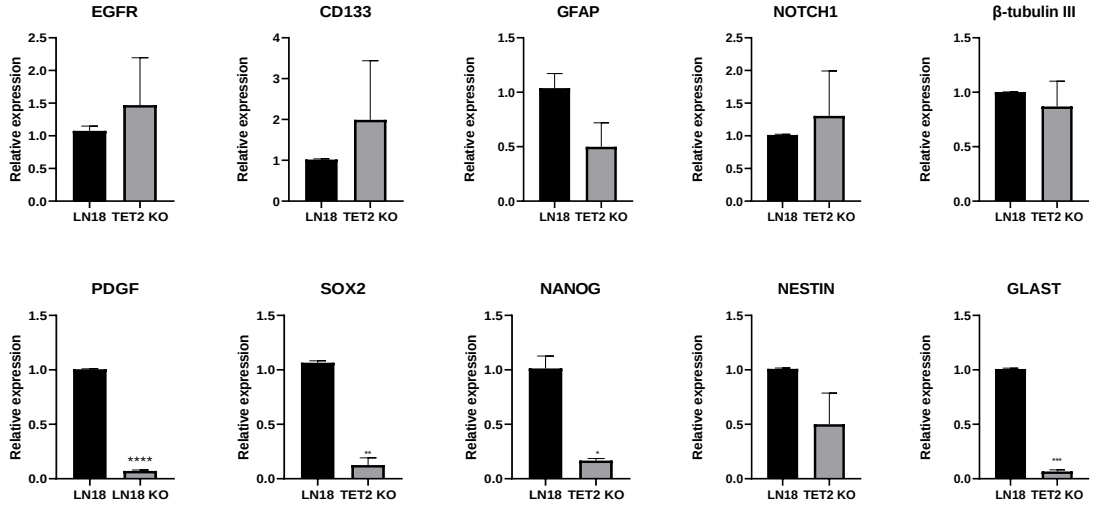


Figure 4.23: TET2 KO leads to reduced transcript levels of key glioma stem cell markers.

Expression levels of key transcripts involved in the maintenance and proliferation of glioma stem cells in TET2 KO samples compared to LN18 non-targeted cells. Mean values of three independent experiments are shown (n = 3). Experimental error is shown as ± SD. Statistical analysis was performed using un-paired t test; $p \leq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ and **** $p < 0.0001$.

Our findings prompted us to examine the tumorigenic capacity of the cells. We decided to perform extreme limited dilution assay (ELDA) and assess the development of gliomaspheres in culture with TET2 KO cells compared to the control. Cells were seeded in 96-well plates in a serial dilution starting from 250 cells down to 5 cells per well. Seeding of 1 cell per well was not feasible because of technical problems. Cells started forming spheres at day 8 in culture and reached 100 μm in size on day 14. We observed that after 14 days in culture TET2 KO cells were not producing gliomaspheres at the same rate as LN18 control cells (Fig. 4.24 A). We then analysed our data by measuring the average size of spheres in culture at different seeding conditions. We found that the TET2 KO gliomasphere average size was lower compared to the control; however those results were not statistically significant (Fig. 4.24 B). Nevertheless, the most commonly used way to examine the tumorigenic capacity of cells in culture is to count the number of gliomaspheres formed during a certain period of time. Therefore, we counted the number of gliomaspheres larger than 100 μm using Image J software. Interestingly, we found that TET2 KO samples formed significantly lower number of gliomaspheres in culture compared to the control in all seeding densities examined (Fig. 4.24 C). In most of the cases, these numbers were up to 10 fold lower compared to the control; with all the findings being statistically significant in a total number of 6 independent biological replicates including 6 technical replicates in each repeat (Fig. 4.24 C).

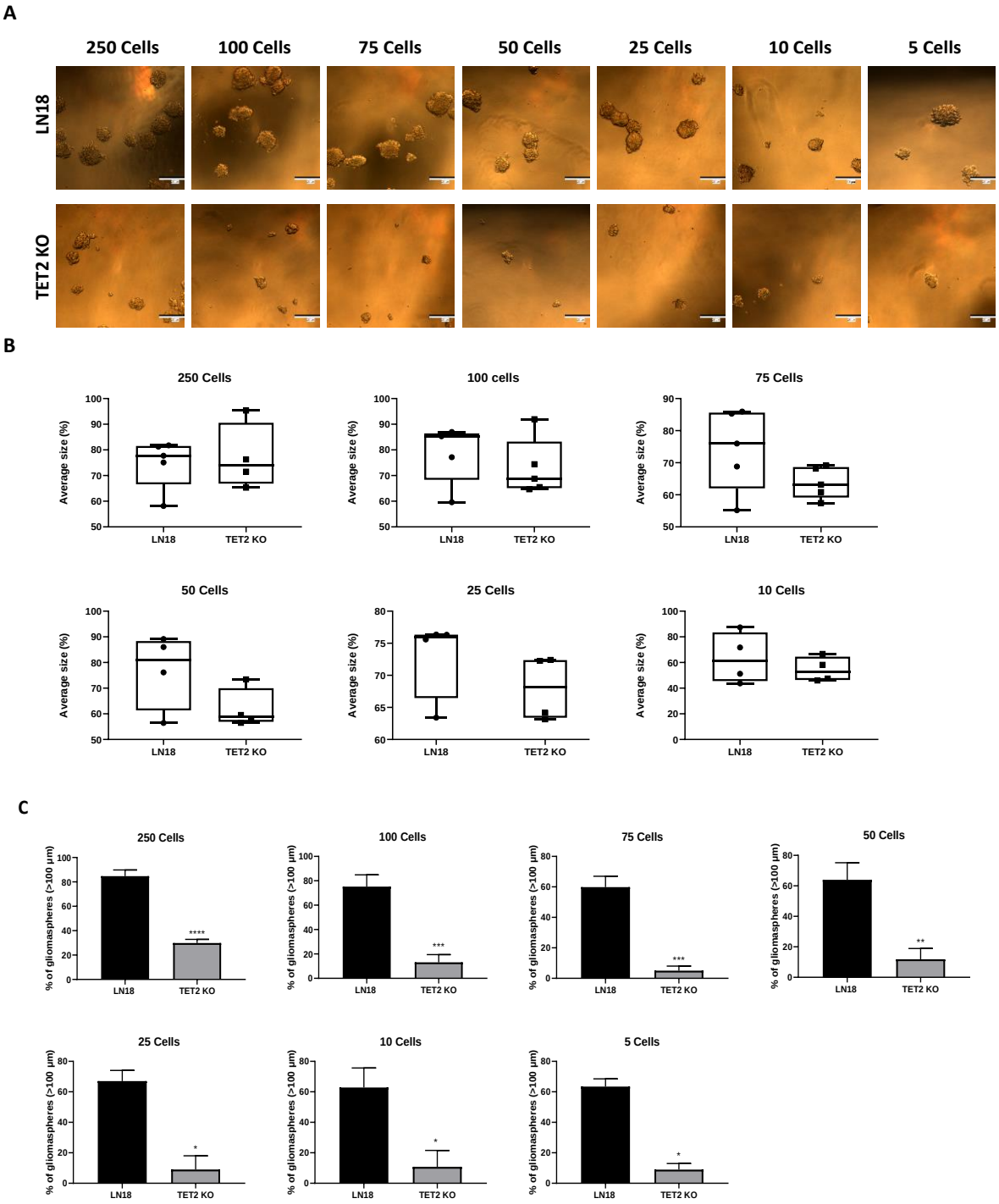


Figure 4.24: TET2 KO mutants display significantly lower tumorigenic capacity in culture.

A. Light microscope images of cells seeded in serial dilution in gliomasphere formation cell culture conditions. Scale bar = 250 µm. B. Average size of gliomaspheres (displayed in percentage) in the indicated samples showing no significant difference between LN18 non-targeted cells and TET2 KO in six individual experiments (n = 6). C. Percentage of gliomaspheres formed (> 100 µm) in the indicated samples showing a significant difference between the two samples (n = 6). Error bars shown as ± SD. Statistical analysis was performed using un-paired t test; $p \leq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ and **** $p < 0.0001$.

We then dissociated and re-seeded the cells in new 96-well plates in order to examine if they will be able to re-form gliomaspheres. Our results show that, TET2 KO cells were not able to re-form the spheres in culture indicating their inability to initiate tumours *in vitro* (Fig. 4.25 A). Quantification of the gliomapsheres formed at the end of the experiment showed that TET2 KO cells did not form any spheres larger than 100 μm ($p < 0.0001$) (Fig. 4.25 B). Therefore, our results show that TET2 protein inactivation has significant effects in the stemness of GBM cell lines, likely by blocking the tumour initiator cells forming the tumour.

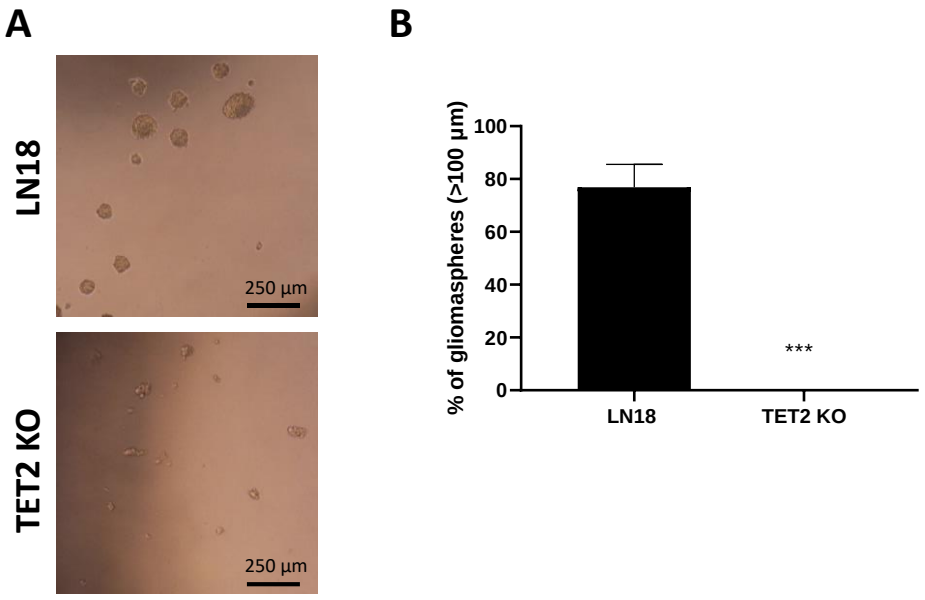


Figure 4.25: TET2 KO cells did not form gliomaspheres after dissociation and re- seeding.

A. Light microscope images of the gliomaspheres after re-seeding. Scale bar = 250 μm. B. Percentage of gliomaspheres formed in culture in the indicated samples showing that in TET2 KO samples there is no formation of gliomaspheres larger than 100 μm in three individual experiments (n = 3). Error bars shown as ± SD. Statistical analysis was performed using un-paired t test; $p \leq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ and **** $p < 0.0001$.

4.4.12. TET2 KO mutants fail to respond to differentiation cues

Given the evidence we have of an impact of TET2 inactivity on LN18 cancer stem cells characteristics and tumorigenic capacity; we next attempted to differentiate the cells into neuronal and glial lineages. In brief, we conditioned the cells in neurobasal medium containing FGF and EGF-2 growth factors for 4 days in order to enrich for the population of cancer stem cells (Fig. 4.26). We next withdrew the growth factors from the media and added FBS and BMP4 for glial differentiation and BDNF and RA for neural differentiation for 6 days (Fig. 4.26). The differentiation capacity of cells at each stage was examined using immunocytochemistry and qPCR.

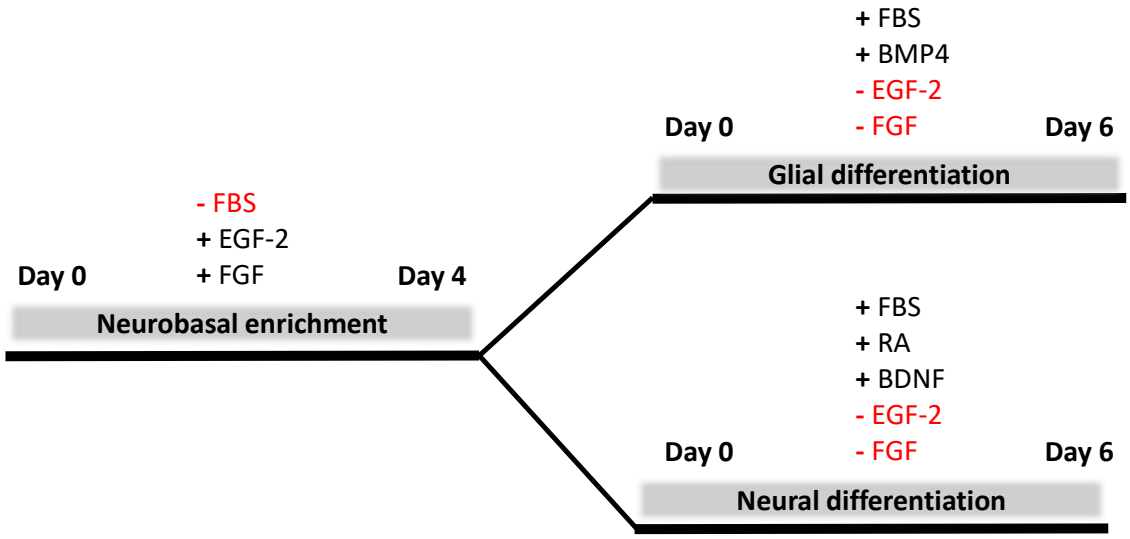


Figure 4.26: Summary of the differentiation approach; including the growth factors used at each differentiation stage.

Hence, we first conditioned the two cell sample populations in neurobasal conditions and after 4 days in culture we collected the samples for immunocytochemistry and qPCR analysis. We performed immunocytochemistry for NESTIN as a key cancer stem cell marker, for GFAP as a marker for glial differentiation and β-tubulin III as a marker for neuronal differentiation (Fig. 4.27 A). Due to the fact that the cells are grown in non-adherent conditions at this stage, with the removal of FBS from the culture, fixation and staining was extremely challenging and in most cases cells were

detached from the slides. We observed that TET2 KO cells displayed lower protein levels of NESTIN, however levels of GFAP did not display any significant differences β-tubulin III levels were upregulated in TET2 KO however, we could not perform any statistical analysis of this sample due to technical issues of the staining experiment (Fig. 4.27 A and B). Therefore, due to the difficulties explained, we decided to perform qPCR to examine the transcript levels of these marks. *NESTIN* and *SOX2* transcript levels were significantly downregulated by 3-fold and 2- fold respectively (Fig. 4.27 C). These two genes are involved in cancer stem cell self-renewal, indicating TET2 KO effects the stemness of LN18 cell line. The transcript levels of β-tubulin III however remained unaltered (Fig. 4.27 C). Interestingly, although GFAP protein was detectable immunohistochemically we were unable to detect *GFAP* transcripts in our qPCR experiments.

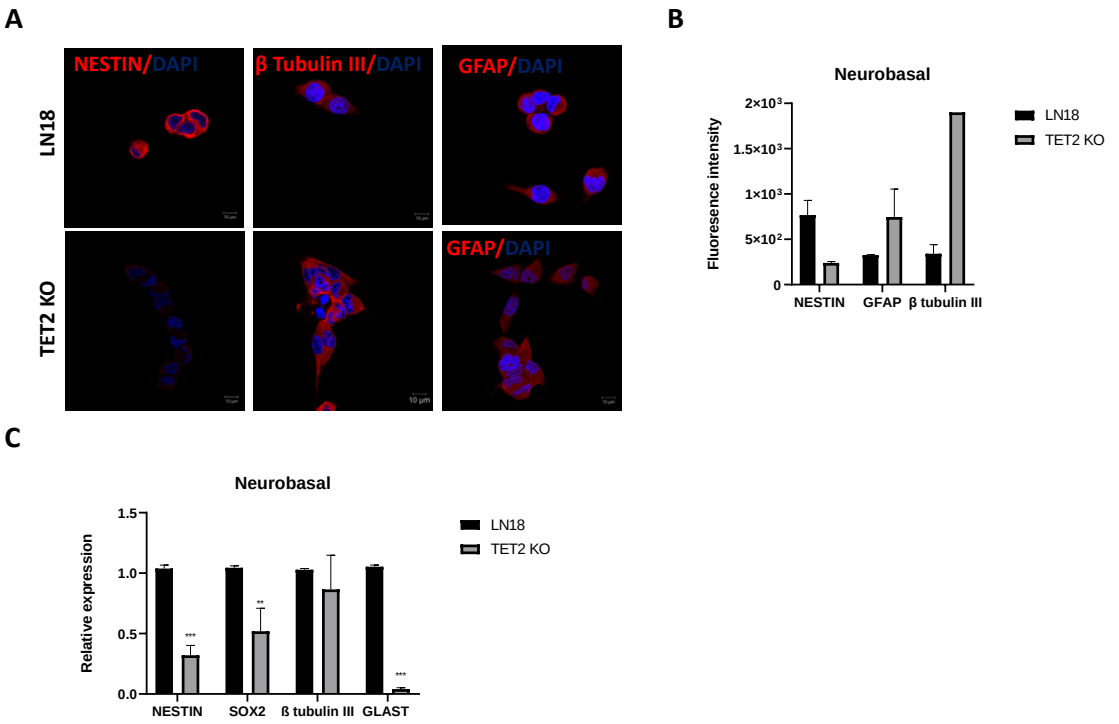


Figure 4.27: TET2 mutants display reduced expression of key stem cell self-renewal markers in neurobasal medium.

A. Immunocytochemistry of LN18 and TET2 KO samples of NESTIN, β-tubulin III and GFAP co-stained with DAPI. The cell cultures were stained in parallel at the same experimental conditions and were imaged at identical settings. Scale bar = 10 μm. B. Quantification of NESTIN, GFAP and β-tubulin III cytoplasmic signal in the indicated samples. Mean values of the average intensities of each individual cell in each cell line are shown in three independent experiments. Error bars are shown as ± SEM (n = 3). C. Relative quantification *of NESTIN, SOX2,* β-tubulin III and *GLAST* transcript levels in LN18 and TET2 KO samples. Mean values of the average expression levels of each sample is shown in six individual experiments. Error bars are shown as ± SD (n = 6), un-paired t test was performed $p \leq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ and **** $p < 0.0001$.

We performed neural and glial differentiation in parallel for both LN18 control cells and TET2 KO cells. Cells were monitored and light microscope images were taken across the whole experiment, samples for qPCR and immunocytochemistry were also obtained at regular time points. Our results show that there are no detectable morphological changes in both samples during neuronal or glial differentiation indicating that cells might have failed to undergo differentiation (Fig. 4.28 A and B). Surprisingly, *NESTIN* and *SOX2* levels increased in response to differentiation cues in both lineages in TET2 KO cells (Fig. 4.28 C and D). Considering that both genes are responsible for cancer stem cells self-renewal, the increase in their levels in TET2 KO cell line could indicate the de-differentiation of the cells instead of differentiation. Taking into account that β-tubulin III marker for neural differentiation did not show any difference in expression levels throughout the experiment and *GFAP* and *GLAST* markers for glial differentiation were significantly down regulated during glial differentiation (Fig. 4.28 C and D), it is more apparent that TET2 mutant GBM cell line cannot differentiate under our experimental conditions. Our immunocytochemistry experiments also showed similar results. We observed that NESTIN protein levels increased in response to differentiation to neuronal and glial lineages (Fig. 4.29). However, we did not observe any significant changes in the levels of β-tubulin III and GFAP (Fig. 4.29). Our results indicate that under our experimental conditions LN18 and TET2 KO cell lines fail to undergo differentiation into neuronal or glial lineages.

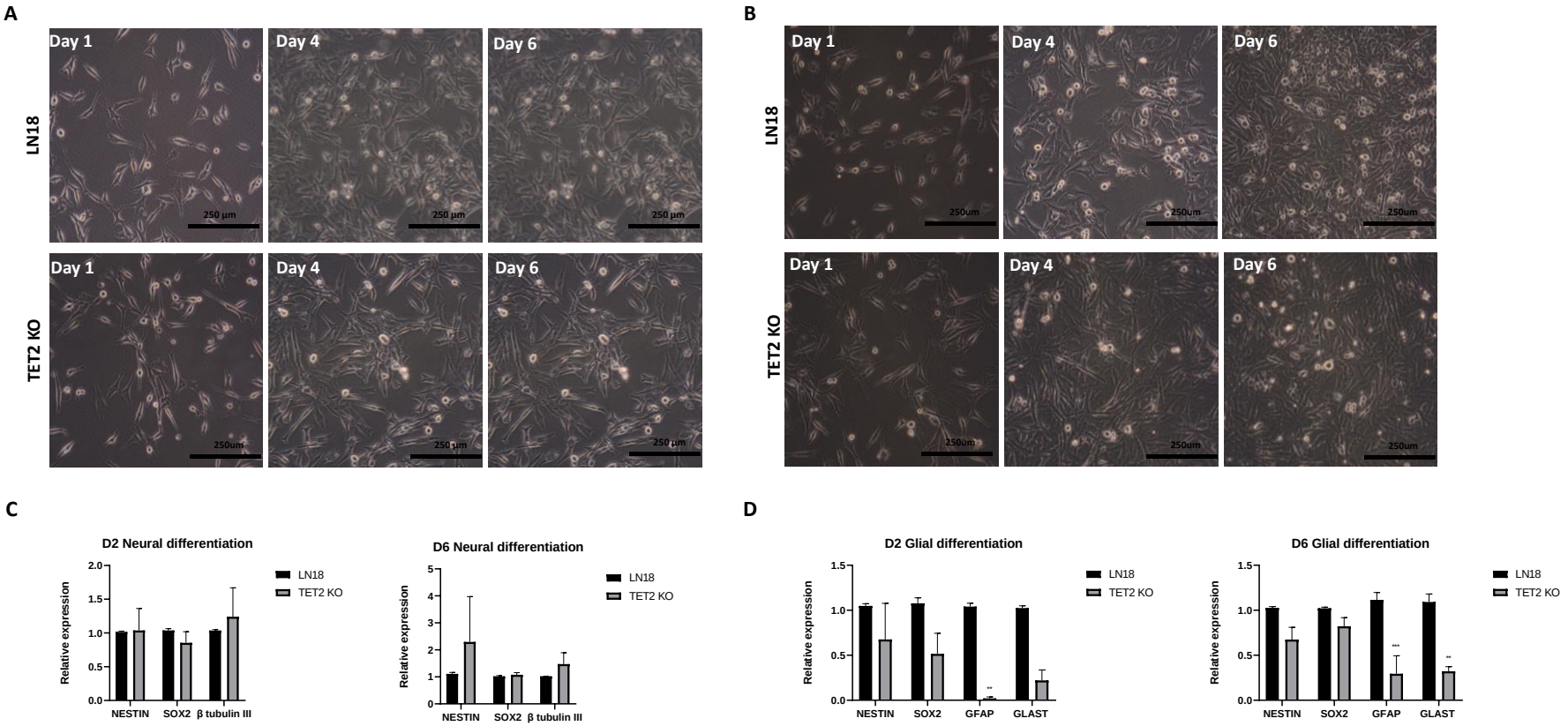
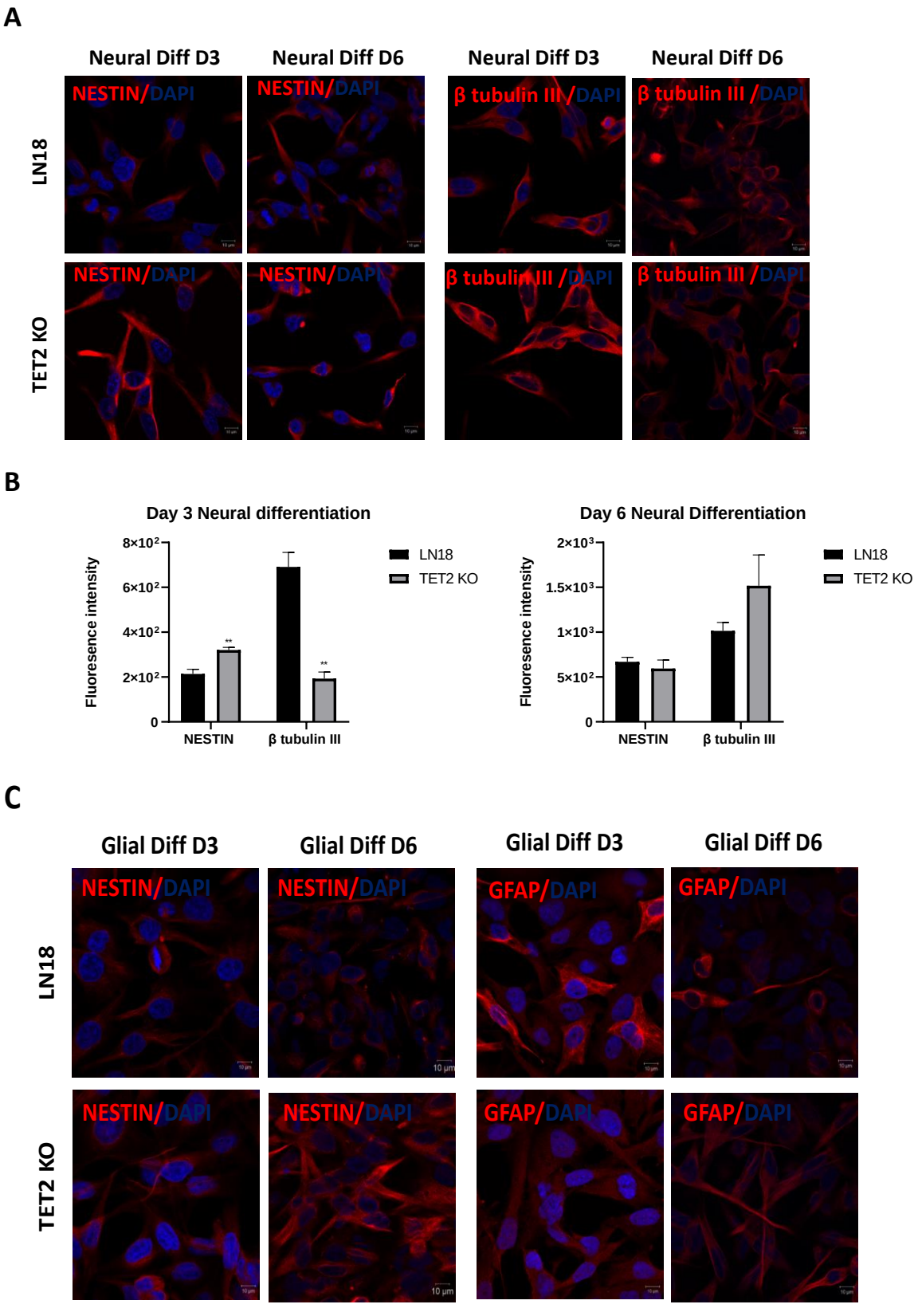


Figure 4.28: TET2 KO cell lines failed to respond to differentiation cues.

A and B. Light microscope images of A. neuronal differentiation and B. glial differentiation at the indicated time points do not show any morphological changes for differentiation. Scale bar = 250 µm. C. Expression of *NESTIN Sox2* and β-tubulin III in TET2 KO cells relative to LN18 cell line at the indicated time points during neural differentiation. D. Expression of *NESTIN, SOX2, GFAP* and *GLAST* in TET2 KO cells compared to LN18 during glial differentiation at the indicated time points. Mean values of four individual experiments (n = 4) are shown and experimental error is shown as ± SD (n = 4). Statistical analysis was performed using un-paired t test; $p \leq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ and **** $p < 0.0001$.


A
Neural Diff D3
Neural Diff D6
Neural Diff D3
Neural Diff D6
LN18
TET2 KO
NESTIN/DAPI
β tubulin III /DAPI
B
Day 3 Neural differentiation
Day 6 Neural Differentiation
Fluorescence intensity
LN18
TET2 KO
NESTIN
β tubulin III
C
Glial Diff D3
Glial Diff D6
Glial Diff D3
Glial Diff D6
GFAP/DAPI

D

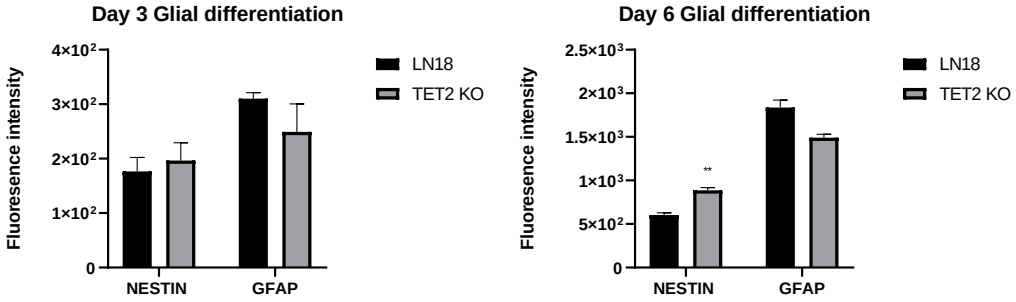


Figure 4.29: NESTIN protein levels were increased upon differentiation to neuronal and glial lineages in TET2 KO cells.

A. Co-immunostaining of NESTIN/DAPI and β-tubulin III / DAPI during neural differentiation in LN18 and TET2 KO cells. Scale bar = 10 μm. B. Quantification of NESTIN and β-tubulin III cytoplasmic signal in the indicated samples at day 3 and 6 of neural differentiation. C. Co-immunostaining of NESTIN/DAPI and GFAP/DAPI during glial differentiation in LN18 and TET2 KO cells. Scale bar = 10 μm. D. Quantification of NESTIN and GFAP cytoplasmic signal at day 3 and 6 of glial differentiation in the indicated samples. Mean values of the average intensities of each individual cell in each condition are shown and error bars are shown as ± SEM in three individual experiments (n = 3). The cell cultures were stained in parallel at the same experimental conditions and were imaged at identical settings. Un-paired t test was performed $p \leq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ and **** $p < 0.0001$.

4.5. Technical discussion

The aim of this chapter was to successfully design a CRISPR/Cas9 system to achieve the complete knockout of *TET2* gene in LN18 GBM cell line and proceed with functional characterisation of the cell line created. Our aim was to disrupt TET2 function at the end of exon 3 before the start of the catalytic domain of the protein; in order to ensure that the protein is no longer active. Plasmid constructs for sgRNAs and donor plasmid were successfully synthesised and transfected in LN18 cell line. The transfected cells were then collected and screened for the presence of the selection cassette at the locus of interest.

The minimal catalytically active fragment in TET2 enzyme is at positions 1129-1936 containing two DSBH highly conserved regions (Hu *et al.* 2013a). It has been previously shown that multiple mutations of TET2 gene are present in patients with leukaemia. The majority of them are missense mutations that impair the enzymatic activity of TET2 leading to reduced 5hmC levels (Konstandin *et al.* 2011). These mutations tend to be in the two DSBH highly conserved regions of the catalytic domain of TET2 protein (Fig. 4.1). Moreover, another study showed that H1302Y and D1304A substitutions in the catalytic domain of TET2 protein leads to diminished levels of 5hmC in the nuclei of HEK293T cells. Other mutations in the catalytic domain of TET2 have a similar effect indicating that a single mutation at the catalytic domain will significantly affect its enzymatic activity (Ko *et al.* 2010). These indications led us to target the TET2 catalytic domain to achieve our goal for complete protein inactivation.

We experienced multiple difficulties during the process of construction of the donor plasmid; therefore we had to change the building strategy multiple times. The main problem we faced during the building of the donor plasmid was the PCR of the selection cassette from pSIN plasmid. The fact that there were no unique restriction enzymes at both sides of the cassette made the restriction enzyme based cloning challenging. Therefore, we attempted to switch to an alternative cloning method, Gibson Assembly, for the integration of the cassette into the intermediate plasmid. The two fragments, selection

cassette and intermediate plasmid, to be annealed had to be extended by PCR so they overlap. However, both fragments were approximately 4 kb long and GC-rich, making the PCR challenging. Thus, we decided to introduce restriction enzymes at each end of the cassette and attempt restriction enzyme based cloning. This method was also challenging due to difficulties in PCR amplification of the fragment due to the high GC content of the primers resulting from the addition of the restriction enzymes. After several attempts using different restriction sites we managed to amplify the fragment of interest and ligate it into the intermediate plasmid for the generation of the final donor plasmid.

Once the plasmids were built and sequenced, we next proceeded with their transfection into the cells. Transfection using lipofection based delivery methods is considered to be one of the "gold standard" techniques. This is due to the fact that it has been shown to efficiently introduce both RNA and dsDNA plasmids in a broad range of cell lines using simple protocols with high reproducibility (Karra *et al.* 2010). Moreover, previous studies in the LN18 cell line successfully used Lipofectamine as a method for transfection of siRNA (Lavanya *et al.* 2016), miRNA (Li *et al.* 2016b) and dsDNA (Liao *et al.* 2018; Würstle *et al.* 2017). Therefore, we first choose the conventional lipid-mediated gene delivery using Lipofectamine.

This method is based on cationic lipid molecules which use small liposomes that interact with negatively charged nucleic acids allowing their fusion with the plasma membrane (Karra *et al.* 2010). One of the main advantages of the use of Lipofectamine is that it is technically simple, does not require any specialised equipment and shows high reproducibility (Karra *et al.* 2010). Nevertheless, we had to optimise multiple parameters in order to achieve the highest desired transfection efficiency. Those parameters included cell density, cargo concentration and transfection reagent concentration. Our attempts to optimise all the conditions did not lead to the desirable efficiency (Fig. 4.7 B), mainly due to the toxicity that the transfection reagent had on the cells leading to cell death. It is important to note that we performed our optimisation using a single GFP plasmid, however in our experiments our plan was to use 4 different plasmids which could further increase toxicity. Nevertheless, we attempted to transfect our plasmids using Lipofectamine in

low concentrations to avoid as much toxicity as possible; however transfection efficiency was extremely low and cells did not survive BSD selection (Data not shown).

Therefore, we decided to use electroporation as an alternative. As it has been previously used in GBM cell lines for the transfection of shRNA (Liao *et al.* 2018). During electroporation, cells are exposed to a voltage pulse, temporarily altering the plasma membrane allowing extracellular material to enter the cell (Washbourne & McAllister 2002). Although electroporation offers relatively high transfection efficiency with little optimisation required (Karra *et al.* 2010), it requires relatively expensive equipment and reagents; for that reason electroporation was not our first choice to perform our experiments.

Transfected cells were selected through BSD selection and FACS cell sorting, based on their GFP expression, followed by clonal expansion and screening. We initially transfected LN18 cell line with four individual plasmids using the nickase D10A plasmid. Our aim was to use nickase Cas9 with a pair of gRNAs that would introduce DNA breaks in both DNA strands, reducing off-target events. However, although antibiotic-resistant and GFP positive cells were obtained, we could not identify any mono-allelic nor bi-allelic mutants after screening the clones. This could be due to two main reasons: a reduced specificity of D10A Cas9 plasmid and a non- functional gRNA plasmid which could be a result of a poor isolation of the plasmid. Therefore, in order to test this hypothesis, wild-type Cas9 was used with guide A and guide B in separate transfections. Interestingly, it was only in guide A transfected cells that we managed to identify mono-allelic mutants (Fig. 4.10). We then reasoned that the inability of guide B to direct SSB at DNA might have been caused by the significantly reduced activity of D10A Cas9.

Traditionally, in order to express Cas9 and sgRNAs in cells in culture, dsDNA plasmids which encode them are constructed and transfected via lipofection or electroporation. Therefore, we decided to follow this route for our transfection. Unfortunately though, this method is risky as the use of plasmids often favours the random integration of all or part of the plasmid DNA into the genome (Kim *et al.* 2014). These random integrated sequences

are often responsible for host immunity responses (Wagner 2001). It has been previously shown that transfection of human embryonic kidney cells (HEK293FT), and human fibroblasts with Cas9 protein, instead of dsDNA plasmid increases the specificity of targeting (Kim *et al.* 2014; Liang *et al.* 2015). However, there is a disagreement in the observed transfection efficiency as the study in HEK293FT cells shows that efficiency does not change with either mRNA Cas9, protein Cas9 or dsDNA (Liang *et al.* 2015); whereas the study in fibroblasts shows that transfection efficiency with protein Cas9 is at least two fold higher compared to plasmid transfection (Kim *et al.* 2014).

Indeed, in our experiments we observed random integration of the donor plasmid in a sub-population of cells and as a result this population was BSD resistant and GFP positive making the selection of the correct clones much more challenging. It has been shown that prolonged expression of Cas9 and sgRNAs, which can last several days post transfection, can magnify off-target effects (Gaj *et al.* 2012). Interestingly, Cas9 protein accumulated over time in cells transfected with plasmid DNA whereas in mRNA transfected cells Cas9 protein peaked 4 hours post transfection and remained stable for 48 hours before diminishing (Liang *et al.* 2015). On the other hand in cells that were transfected with Cas9 protein, Cas9 peaked early after transfection and rapidly decreased (Liang *et al.* 2015). Kim et al., showed similar results and demonstrated that chromosomal DNA is cleaved almost immediately after delivery and Cas9 protein degrades rapidly; therefore reducing the risk of off-target events (Kim *et al.* 2014). It is possible that prolonged presence of the plasmids in our cells led to the random integration we observed. Therefore, in future experiments it may be more efficient to use Cas9 protein for our transfections, reducing the possibilities of off-target events.

The next aim of this chapter was to functionally characterise the cell line developed and to try and understand the role of TET2 in the active DNA demethylation pathway in GBM as well as its role in tumorigenesis. There have been numerous publications showing the correlation of 5hmC levels and TET2 enzymatic activity in malignancies and other diseases. More specifically it has been shown that TET2 is significantly downregulated in multiple sclerosis (MS) and this is associated with aberrant methylation in its promoter

corresponding to lower 5hmC levels (Calabrese *et al.* 2014). Therefore, lower 5hmC levels in peripheral blood of MS patients is likely to depend on TET2 down-regulation and variation in DNA demethylation marks in peripheral blood might be used as an indicator of ongoing alterations in the CNS (Calabrese *et al.* 2014). Moreover, 5hmC levels are positively associated with TET2 expression in breast precancerous lesions (Zhang *et al.* 2019). It has been also shown that reduced levels of TET2 enzyme in human melanoma is responsible for 5hmC loss, as re-introduction of active TET2 in melanoma cell lines results in the global increase of 5hmC levels (Lian *et al.* 2012). Our immunocytochemistry and mass spectrometry results, similar to previous publications, revealed that TET2 is responsible for the oxidation of 5mC in LN18 GBM cell line. Specifically both qualitative and quantitate analysis we performed via immunocytochemistry and mass spectrometry showed that 5hmC levels were significantly reduced in both TET2-/+ and TET2 KO samples.

Nevertheless, it is striking that although 5caC levels were diminished in TET2 KO cells in our immunocytochemistry experiments, the mass spectrometry analysis showed that there is no significant difference between TET2 KO cell line and the non-targeted LN18 cells. Our previous results based on meDIP followed by NGS showed that the levels of 5caC in LN18 are extremely low compared to hiPSCs and UW228-3 cell line. This might indicate that due to the extremely low levels of 5caC, the threshold set for mass spectrometry detection was low and resulted in false positive results in all the samples. Current literature about 5caC detection especially in cancer tissues is limited and is mainly based on antibody based detection methods (Chou *et al.* 2016; Eleftheriou *et al.* 2015; Ramsawhook *et al.* 2017; Storebjerg *et al.* 2018). Therefore, it might well be that there is a limitation in mass spectrometry detection of 5caC in our experiments. However, we cannot exclude the possibility that 5caC is still present in LN18 cell line after TET2 inactivation and is possibly oxidised by TET1 enzyme, the levels of which remain stable after the knockout of TET2.

Functional characterisation of the TET2 KO cell line showed that there is a correlation between TET2 and stemness. We first used qPCR screening to examine the transcript levels of key genes involved in GBM pathogenesis and

glioma stem cell maintenance in TET2 KO cells. *PDGF* and *EFGR* were among those genes, as it is well known that overexpression of PDGFR and EGFR can contribute to the malignant transformation of human GBM (Craig Clark *et al.* 1992). In TET2 KO cell line *PDGF* levels were approximately 14 fold lower (Fig. 4.23) compared to the LN18 cell line, indicating that *PDGF* might represent a downstream target of *TET2*. PDGF overexpression is associated with tumour development in a wide range of tumours including brain, liver, lung and haematopoietic malignancies (Kim *et al.* 2012). Moreover, *in vitro* studies in mouse models implicated PDGF pathway in tumour angiogenesis and cellular invasion (Shih & Holland 2006). Inhibition of PDGF signalling pathway induces apoptosis in GBM stem cells (Cenciarelli *et al.* 2014; Kim *et al.* 2012) and PDGFR depletion in glioma cancer stem cells leads to attenuation of stemness-related and glial markers expression (Cenciarelli *et al.* 2016). We also observed similar patterns in the expression levels of *SOX2* and *NANOG*. SOX2 activity is associated with the maintenance of cancer stem cells in GBM by maintaining tissue homeostasis and regeneration (Sarkar & Hochedlinger 2013). Moreover, high *NANOG* levels have been associated with tumour aggressiveness and worse prognosis in GBM; regulating glioma stem cells (Ben-Porath *et al.* 2008; Zbinden *et al.* 2010). Taken into account that 5hmC levels are extremely low in TET2 KO cells it is possible that *PDGF, SOX2* and *NANOG* promoters remained silent due to the inability of TET2 to induce further oxidation; explaining the extremely low levels we observed in our clone. Our results indicate that there might be a direct correlation between the maintenance of glioma stem cells and active DNA demethylation. Therefore, these observations coupled with the current knowledge led us to direct our research into glioma stem cells and the tumorigenic capacity of TET2 KO cells.

One way to examine tumorigenic capacity of cells *in vitro* is to examine the ability to form gliomaspheres in culture. We cultured the cells in a medium containing EGF-2 and FGF growth factors in the absence of serum in order to efficiently promote enrichment of cancer stem cells as has previously been described (Hong *et al.* 2012). This method has been widely used for the study of glioma stem cells (Garros-Regulez *et al.* 2016; Hardee *et al.* 2012; Richichi *et al.* 2013). We observed that TET2 KO cells formed a significantly smaller number of gliomaspheres *in vitro* compared to LN18 cell lines (Fig. 4.24);

with diminished ability to form secondary gliomaspheres in culture (Fig. 4.25). Our results confirm our initial indications that impaired active DNA demethylation pathway is associated with stemness in GBM LN18 cell line, inhibiting the ability of cells to form gliomaspheres in culture. However, these observations do not correlate with any changes in the proliferation rate of the cells indicating that TET2 enzymatic inactivity affects the stemness of the tumour only, without affecting the cell cycle.

We then examined the ability of cells to differentiate into neuronal and glial lineages *in vitro*. Terminally differentiated cells have a limited lifespan; therefore differentiation therapy has the potential to eradicate proliferating tumour cells (Sell 2004). Our aim was to examine the difference, if any, in the differentiation capacity of LN18 GBM cell line and TET2 KO cells. The withdrawal of growth factors from culture usually results in spontaneous multi-lineage differentiation of glioma stem cells where a subpopulation of cells undergoes cell cycle exit (Pollard *et al.* 2009). It has been shown that astrocytic differentiation can be achieved following 7 days of BMP4 treatment coupled with the withdrawal of growth factors (Carén *et al.* 2015); therefore we used this method in order to achieve astrocytic differentiation. However, we did not observe any significant changes in the morphology of the cells that would indicate their differentiation, with qPCR and immunocytochemistry experiments also showing the same results (Fig. 4.28 and 4.29). We also attempted neuronal differentiation with similar results. Our results indicate that LN18 GBM cell line and TET2 KO generated cell line fail to undergo differentiation under our experimental conditions. Although we did not manage to differentiate the cells into the two lineages, we surprisingly observed that in TET2 KO samples, glial stem cell markers (*SOX2* and *NESTIN*) levels increased during glial and neuronal differentiation. However, further experiments need to be performed in order to examine why this occurred.

In conclusion, in this chapter we managed to generate a CRISPR/Cas9 system for TET2 KO. We targeted LN18 GBM cell line and were able to create a TET2 KO cell line with TET2 catalytically inactive protein. Finally, we showed that TET2 has a crucial role in active DNA demethylation pathway in GBM as well as in the stemness of the tumour.

Chapter 5. 5caC potential readers in paediatric and adult brain tumours cell lines

5.1. Introduction

The addition of a methyl group to DNA and histone modifications are of crucial importance in many biological processes. The key players underlying these modifications have been designated as epigenetic modifiers (Biswas & Rao 2018). They are many enzymes that catalyse these modifications, regulating the diversity of the epigenetic state allowing further levels of complexity in gene expression (Biswas & Rao 2018). These epigenetic modifiers can be classified into three categories: writers, readers and erasers. Writers are responsible for the establishment of the epigenetic modifications in the genome. Erasers are the enzymes responsible for the removal of these marks and readers are capable of recognizing specific epigenetic marks in a region (Feinberg *et al.* 2016; Hashimoto *et al.* 2010; Torres & Fujimori 2015). In this chapter we will focus on 5caC readers; their role in the regulation and maintenance of DNA methylation and oxidative derivative marks.

Several proteins have been identified to be specific readers of 5mC. Histone deacetylase complex (MECP1) was the first protein complex identified as a 5hmC reader (Meehan *et al.* 1989); this family shares a common methyl CpG binding domain (MBD) (Du *et al.* 2015). MBD2, which is a component of this complex, was confirmed to have a binding specificity to 5mC (Zhang *et al.* 1999). Other members of this complex, MBD1, MBD3, MBD4 and MECP2 share the same binding specificity for 5mC (Du *et al.* 2015) and their recruitment to methylated DNA imposes an inactive chromatin formation. Interestingly enough, the MBD domains of these proteins are unable to bind to sequences with 5hmC modification (Jin *et al.* 2010); suggesting that their function as readers is blocked once 5mC is oxidised to 5hmC. Although earlier studies have suggested that MBD3 (Yildirim *et al.* 2011), MECP2 (Mellen *et al.* 2012) might specifically recognize 5hmC; follow- up studies failed to

support these findings (Hashimoto *et al.* 2012b; Khrapunov *et al.* 2014). Other groups of proteins identified as 5mC readers include transcription factors, such as kruppel-like factor 4 (KLF4) and specific subset of zinc finger proteins including zinc fingers ZNF114 and ZNF416 (Buck-Koehntop & Defossez 2013; Hu *et al.* 2013b; Spruijt & Vermeulen 2014).

In contrast to 5mC specific readers, 5hmC readers seem to be very limited in number. Experiments based on mass spectrometry have revealed a small number of proteins that preferentially interact with 5hmC (Iurlaro *et al.* 2013; Spruijt *et al.* 2013). It has been shown that the SRA domain of ubiquitin-like, containing PHD and RING finger domains 2 (UHRF2) specifically recognizes and interacts with 5hmC (Spruijt *et al.* 2013). This finding was further supported by another structural and biochemical study, which showed that the SRA domain of UHRF2 preferentially binds to fully hydroxymethylated DNA 3.2 fold and hemi-hydroxymethylated DNA 1.5 fold stronger than it does to hemi-methylated DNA (Zhou *et al.* 2014).

Although 5caC is in low abundance in the genome, it is believed that this modification has significant impact on gene regulation and genome stability (Hashimoto *et al.* 2014a; Shen *et al.* 2013; Wang *et al.* 2015a). A number of studies have shown that proteins involved in transcription and genomic maintenance, specifically recognize and bind at sequences with 5caC modification (Hashimoto *et al.* 2014a; Jin *et al.* 2016; Spruijt *et al.* 2013; Wang *et al.* 2015a; Zhang *et al.* 2012). Specifically, the CXXC domain of Tet3 was found to be a specific reader of 5caC (Jin *et al.* 2016). Briefly, using electrophoretic mobility shift assay (EMSA) confirmed that the CXXC domain of *Tet3* binds at unmodified cytosine while further studies showed that mouse *Tet3*-CXXC domain has even stronger binding affinity for 5caC containing DNA (Jin *et al.* 2016). Genomic mapping of TET3 in neuronal and mESCs showed that it is recruited to the TSS of a subset of genes involved in RNA splicing, DNA base excision repair and lysosomal activities (Jin *et al.* 2016). This suggests that the recognition of 5caC by TET3 has a critical role in maintenance of the methylation-free state of TET3 target promoters.

Moreover, 5caC might regulate gene transcription by modulating the recognition of transcriptional proteins of target DNA (Hashimoto *et al.* 2014a;

Wang *et al.* 2015a)**.** One example is WT1, which is a transcription factor that binds to DNA with a higher affinity to 5caC containing sequences (Hashimoto *et al.* 2014a).

Although there is an increased body of experimental evidence about 5mC readers, the readers of 5mC oxidative derivatives at a genome wide level are poorly studied. Iurlaro et al., established a proteomics screen for C, 5mC, 5hmC and 5fC binding proteins using promoter sequences as baits, then comparing protein binding patterns to unmodified or modified cytosines using DNA derived from mESCs. The promoter regions of the Paired Box 6 (*Pax6)* and Fibroblast growth factor 15 (*Fgf15)* genes were used to make PCR probes (Iurlaro et al. 2013). Both of the regions were enriched for 5hmC in mESCs and TET1 siRNA mediated knockdown led to decreased 5hmC levels on their promoters with corresponding changes in gene expression (Ficz et al. 2011). Modified cytosines were incorporated during PCR and then incubated with nuclear protein extracts from mESCs, proteins were then eluted and identified by mass spectrometry (Iurlaro et al. 2013). Only a few proteins showed preference for 5hmC including RBM14, PRP8, MSH6 and PNKP (Iurlaro et al. 2013). However there were a lot of proteins showing preference for 5fC including transcriptional regulators, DNA repair factors and chromatin regulators (Iurlaro et al. 2013). These results were validated through Enzyme-Linked Immunosorbent Assay (ELISA) with purified recombinant proteins and differentially modified probes (Iurlaro et al. 2013).

In a similar study, published by Spruijt et al., they used mass spectrometry based proteomics to identify the readers of 5mC, 5hmC, 5fC and 5caC in mESCs, NPC and adult brain tissue (Spruijt *et al.* 2013). Specifically, they developed a DNA pull-down approach followed by quantitative mass spectrometry. Briefly, nuclear extracts from mESCs grown in "light" or "heavy" SILAC medium were incubated with double stranded DNA sequence probed with a non-modified C or mC, hmC. Bound proteins were then in-gel digested with trypsin and analysed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) (Spruijt *et al.* 2013). They first identified 19 proteins enriched for mC compared to C in mESCs nuclear extracts, including MeCP2, MDB1 and MDB4; confirming previous studies (Du et al. 2015; Jin et al. 2010). 5hmC readers showed partial overlap with mC readers; only 3

proteins interacted with both mC and hmC: MeCP2, UHRF1, and Thymocyte Nuclear Protein 1 *(*THYN1) although MECP2 bound to 5mC with much higher affinity; confirming previous studies (Mellen et al. 2012). Surprisingly, 5caC and 5fC recruited many more proteins than 5hmC with very little overlap. TDG represents one of the proteins that binds to caC and fC but not hmC (Spruijt *et al.* 2013). Furthermore, P53 a well-known protein involved in DNA damage response (Williams & Schumacher 2016), is a specific reader for fC while DNMT1 specifically interacts with caC only (Spruijt *et al.* 2013). Interestingly, 5fC specific protein readers are related to DNA repair according to the GO term enrichment for biological process analysis whereas 5caC readers are not enriched for any biological process (Spruijt *et al.* 2013). These findings indicate a major role for active DNA demethylation in both transcription and DNA repair.

5.2. Aims

Readers of 5caC in mESCs include a number of histone modifying enzymes (e.g. Euchromatic Histone Lysine Methyltransferase 1 (EHMT1)), members of chromatin remodelling complexes (e.g. SWI/SNF Related, Matrix Associated, Actin Dependent Regulator of Chromatin Subfamily C member 2 (SMARCC2)) and transcription factors (e.g. WIZ Zinc Finger (WIZ)) (Spruijt *et al.* 2013). However, the biological roles of these proteins in interpreting 5caC are essentially unknown. Our aim in this chapter is to functionally characterise several candidate proteins which are likely involved in 5caC interpreting active DNA demethylation in cancer cell lines. We used real-time PCR experiments for the initial screening of the gene expression of 5caC readers in adult and paediatric brain tumours, immunocytochemistry for the examination of the levels of 5caC/5hmC and loss of function experiments. The characterisation of the biological roles of these proteins will add to the growing body of information regarding the specific role of the 5caC epigenetic modification in adult and paediatric brain tumours and lead to better understanding of 5caC role in the development of these tumours.

5.3. Results

5.3.1. Investigation of the 5caC potential readers expression levels in paediatric brain tumour cell lines

We selected 28 genes from the specific readers of 5caC in mESCs identified by quantitative mass-spectrometry (Spruijt *et al.* 2013). Including, transcription factors (e.g. RCOR2, WIZ), histone modifying enzymes (e.g. KDM1A, EHMT1) and chromatin remodelling factors (e.g. ZNF462, SMARCC2). We first examined their transcription levels in EPN, MB and GBM cell lines using real time PCR (qPCR) relative to GAPDH as a reference gene. This initial screening was performed in order to identify the trends in the expression levels of 5caC potential readers in the cell lines with elevated 5caC levels compared to the HeLa cell line, which according to our previous experiments did not exhibit immunochemically detectable levels of 5caC (Ramsawhook *et al.* 2017).

DKFZ-EP1NS and BXD-1425 EPN cell lines were used as representative cell lines for EPN paediatric brain tumours. We identified a number of 5caC candidate readers that had higher relative expression; including the transcription repressor Rest Corepressor 2 (*RCOR2)* that was present at 60 fold higher levels in BXD-1425 cell line compared to HeLa (Fig. 5.1). Other transcripts with higher expression levels were transcription factors such as *SMARCC2*, chromatin remodelling factors including multiple myeloma 1 (*MUM1)* and Ankyrin Repeat Domain 17 *(ANKRD17)* a cell cycle regulator (Fig. 5.1).

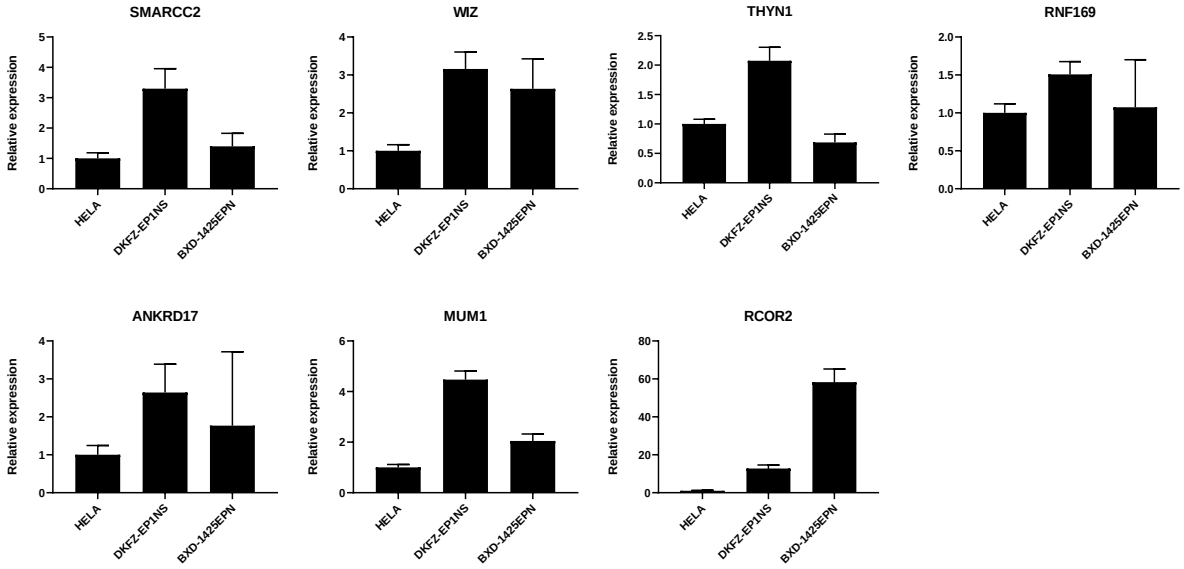


Figure 5.1: Potential 5caC readers that are upregulated in EPN cell lines compared to HeLa.

Graphs show highly expressed transcripts compared to HeLa cell line in which 5caC levels are undetectable. Mean values of three replicates are shown (n = 3). Error bars shown as ± SD.

Nevertheless, the majority of the genes (12 out of 28; 42%) examined, were downregulated compared to HeLa (Fig. 5.2). The genes that displayed the lowest levels of transcription were genes responsible for the regulation of gene expression including Scm Like With Four Mbt Domains 2 *(SFMBT2),* FLT3 Interacting Zinc Finger 1 *(FIZ1) and* Zinc Finger and BTB Domain Containing 7A *(ZBTB7A)* (Fig. 5.2).

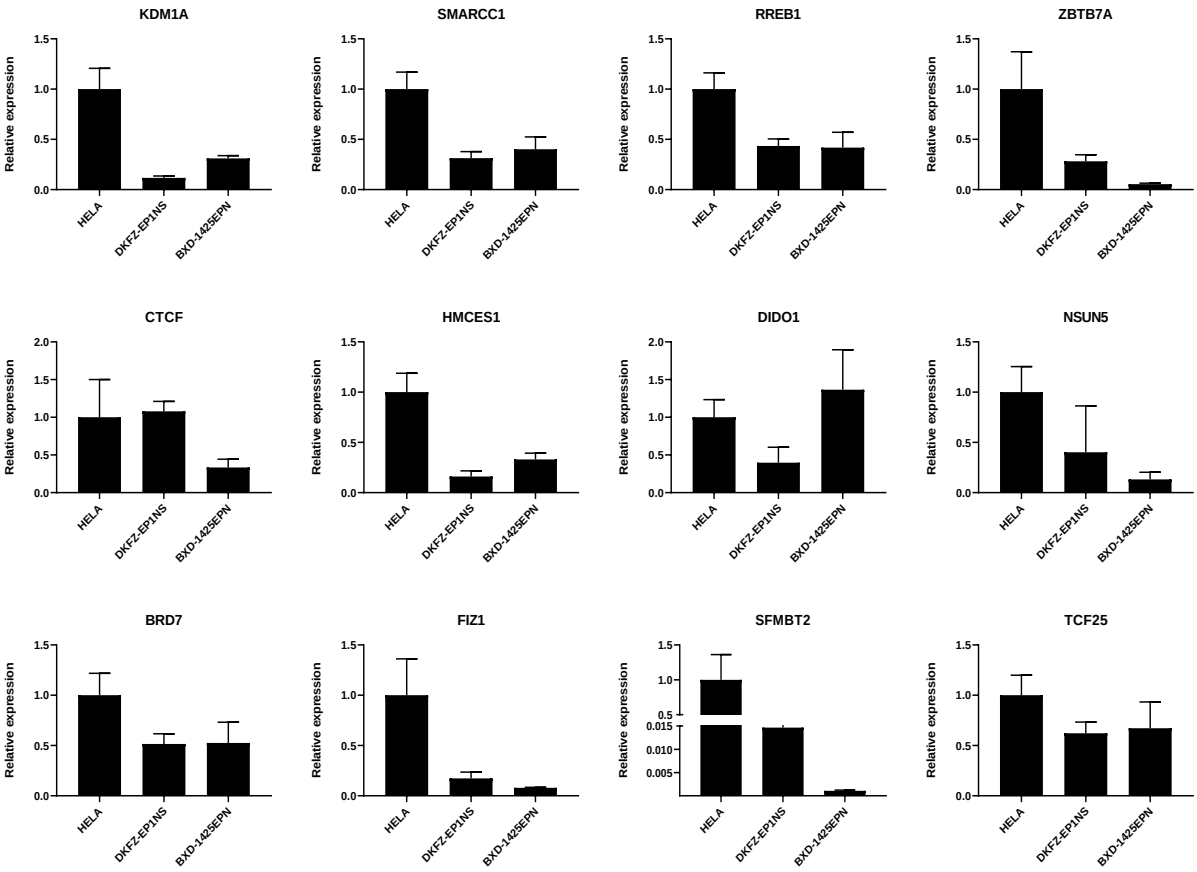


Figure 5.2: Potential 5caC readers that are downregulated in EPN cell lines compared to HeLa.

Graphs show downregulated transcripts compared to HeLa cell line where 5caC is absent. Mean values of three replicates are shown (n = 3). Error bars shown as ± SD.

We also observed a number of genes with significant differences in their relative expression, when comparing the two EPN cell lines, not showing any specific pattern of expression, indicating the differences that exist between cell lines derived from the same type of tumour (Fig. 5.3). These genes include, the DNA methyltransferase *DNMT1*, DNA glycosylase *NEIL1,* the histone methyltransferase *EHMT1* and transcription factors such as nuclear respiratory factor 1 (*NRF1)* and Zinc Finger and SCAN Domain Containing 26 *(ZSCAN26)* (Fig. 5.3).

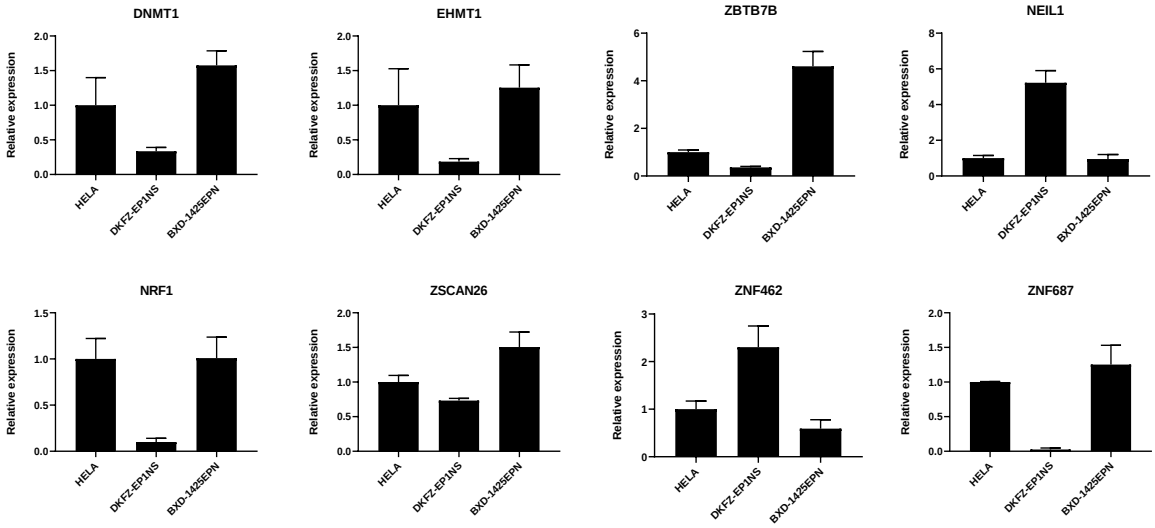


Figure 5.3: A number of 5caC potential readers that do not show any specific pattern of expression in EPN cell lines.

Graphs present transcripts that do not show specific patter of expression between the two EPN cell lines compared to HeLa cell line where 5caC is absent. Mean values of three replicates are shown (n = 3). Error bars shown as ± SD.

We next examined the relative expression levels of 5caC potential readers in DAOY and UW228-3 SHH MB cell lines. We observed that most of the genes shared similar patterns of expression with EPN; *RCOR2, SMARCC2, ANKRD17* and *WIZ* present at higher levels in both paediatric brain tumour cell lines compared to HeLa (Fig. 5.1 and 5.4). Specifically, *SMARCC2* exhibits 10 fold higher levels in the MB cell lines examined compared to HeLa (Fig. 5.4). Moreover, *RCOR2* and *ANKRD17* have approximately 5 fold higher levels in both MB cell lines (Fig. 5.4). Similarly, EPN and MB present the same pattern of expression in the down regulated genes, including *lysine demethylase 1A* (*KDM1A)*, *SMARCC1*, *ZBTB7B* a*n*d ras responsive element binding protein 1 (*RREB1)* (Fig. 5.2 and 5.5).

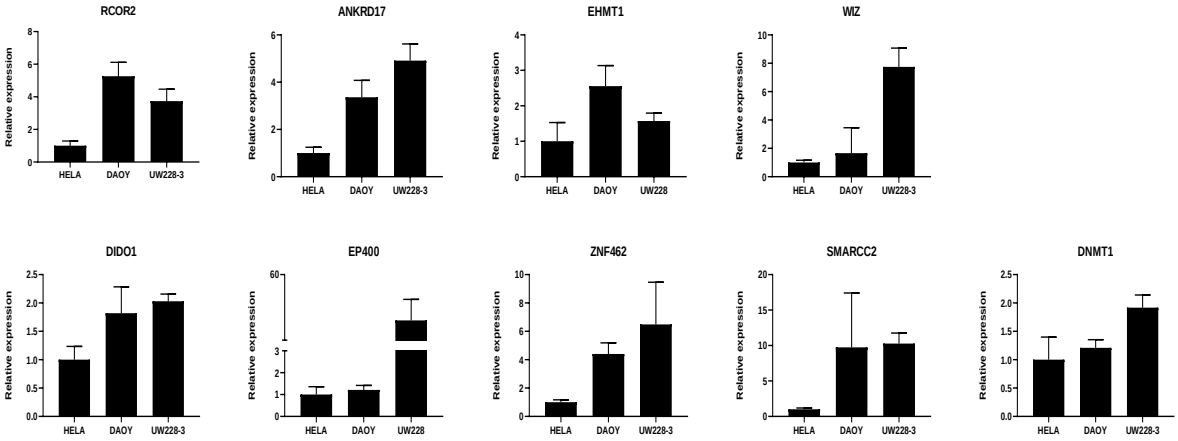


Figure 5.4: Upregulated 5caC potential readers in MB cell lines compared to HeLa.

Graphs present upregulated transcripts compared to HeLa cell line in which 5caC levels are undetectable. Mean values of three replicates are shown (n = 3). Error bars shown as ± SD.

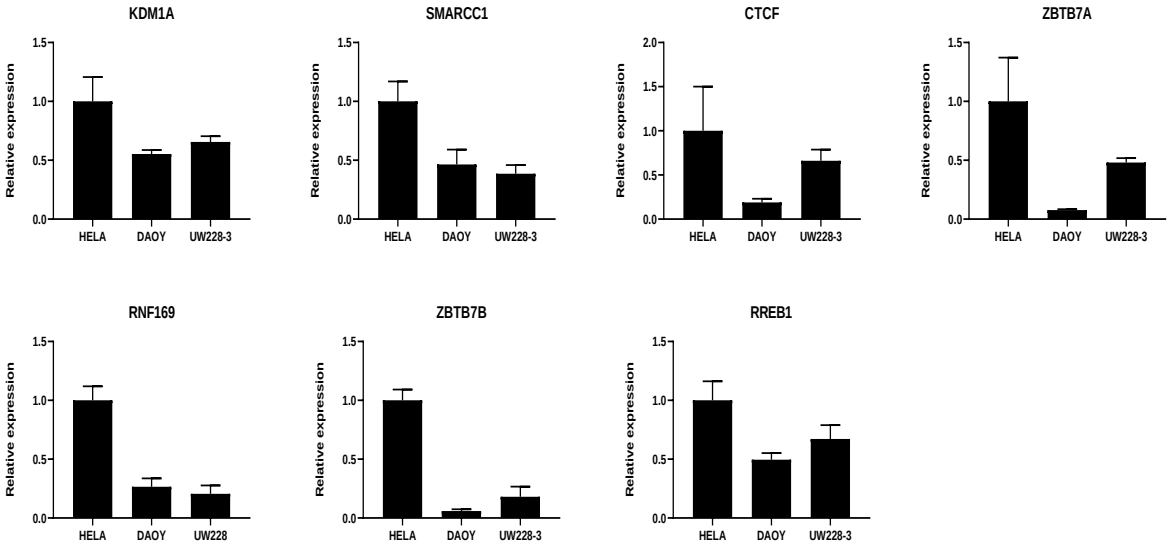


Figure 5.5: Downregulated 5caC potential readers in MB cell lines compared to HeLa.

Graphs present downregulated transcripts compared to HeLa cell line in which 5caC levels are undetectable. Mean values of three replicates are shown (n = 3). Error bars shown as ± SD.

Although DAOY and UW228-3 belong to the same subgroup of MB (SHH) we observed a few genes with no distinct pattern of expression between the two cell lines; including *NEIL1, (THYN1)* and *MUM1* (Fig. 5.6) pointing at the intra heterogeneity that exists between two cell lines in the same subgroup of MB.

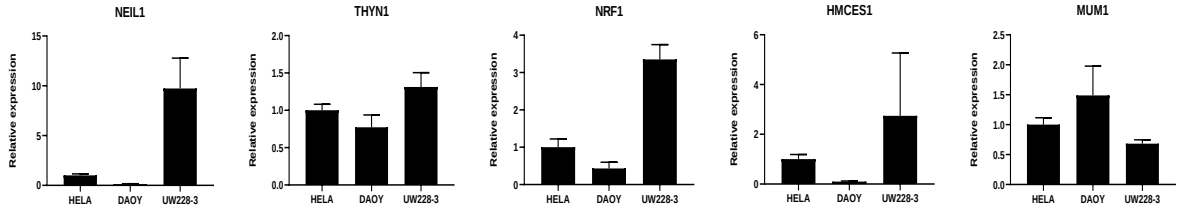


Figure 5.6: A number of 5caC readers do not show any specific pattern of expression in MB cell lines.

Graphs show expression of transcripts between the two MB cell lines compared to HeLa cell line that do not show any specific pattern of expression. Mean values of three replicates are shown (n = 3). Error bars shown as ± SD.

We next sought to plot the gene expression levels of both MB and EPN cell lines examined in a heat map in order to visualise the similarities in their transcripts levels compared to HeLa (Fig. 5.7). The similar patterns in the expression levels of 5caC potential readers in both MB and EPN cell lines may be explained by the fact that they are both paediatric brain tumours; although arising from different parts of the developing brain from different cell lineages as MB is neuronal whereas EPN is derived from glial cells.

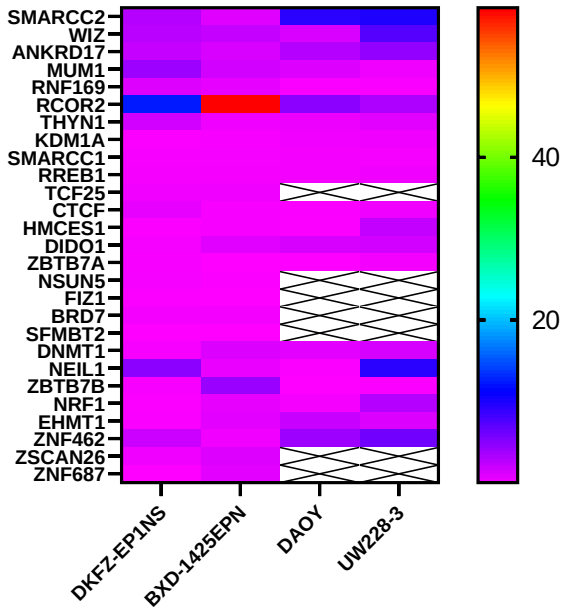


Figure 5.7: The expression patterns of 5caC readers is similar in MB and EPN cell lines

The heat map based comparison of relative expression of 5caC readers in EPN and MB cell lines.

5.3.2. Investigation of the 5caC potential readers expression in GBM cell lines

Finally, we examined the relative expression levels of 5caC potential readers in LN18, LN229 and U87-MG GBM cell lines. In contrast with paediatric brain tumours, in GBM we observed more differences in the patterns of expression of the potential readers between the different cell lines (Fig. 5.8). LN18 and LN229 cell lines share more common expression patterns of the transcripts examined probably because they both belong in the same GBM subgroup characterised by *P53* mutation and *PTEN* wild type. Most of the genes examined do not show any specific pattern of expression between the GBM cell lines (Fig. 5.9). While we only identified a small number of readers that had relatively low expression in all the cell lines compared to HeLa, including *RCOR2* gene which was highly expressed in the paediatric brain tumour cell lines examined (Fig. 5.10).

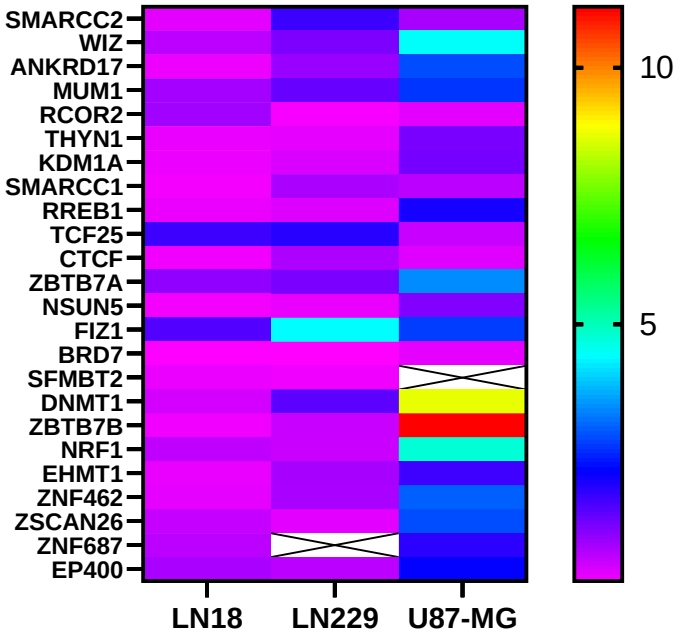


Figure 5.8: The patterns of 5caC reader's expression are not the same in GBM cell lines

The heat map based comparison of relative expression of 5caC readers in three GBM (LN18, LN229, and U87-MG) cell lines.

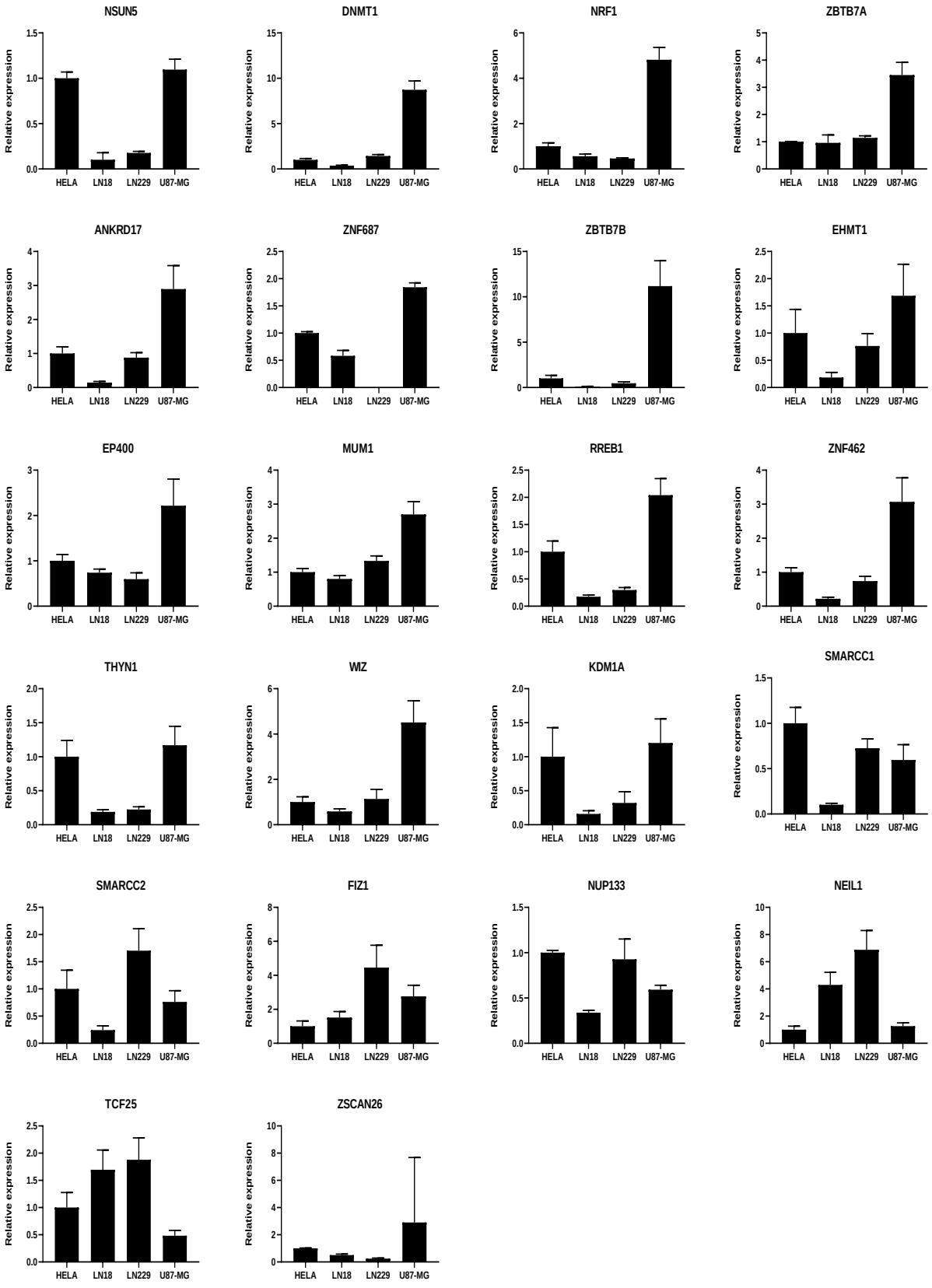


Figure 5.9: Most 5caC potential readers do not exhibit any specific expression pattern in GBM cell lines.

Graphs show expression of transcripts between the GBM cell lines compared to HeLa cell line that do not show any specific pattern of expression. Mean values of three replicates are shown (n = 3). Error bars shown as ± SD.

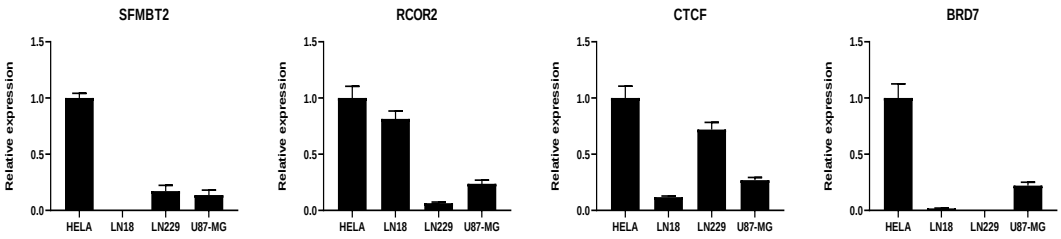


Figure 5.10: Downregulated 5caC potential readers in GBM cell lines compared to HeLa.

Transcripts expressed at low levels in GBM compared to HeLa cell line that 5caC is not detectable. Mean values of three replicates are shown (n = 3). Error bars shown as ± SD.

5.3.3. RCOR2 knockdown leads to reduced immunocytochemically detectable levels of 5caC

We next decided to perform siRNA mediated knockdown of two of the most highly expressed genes in both EPN and MB cell lines, relative to their expression in HeLa cells. We chose *RCOR2* which is a transcriptional repressor of a number of neuronal-specific genes in NSCs (Huang & Bao 2012) while its deletion in mice, leads to defects in brain development and neurogenesis; promoting neural proliferation (Monaghan *et al.* 2017). *SMARCC2* which is involved in transcriptional regulation through chromatin remodelling was the second gene we chose. It has been implicated in the regulation of cortical neurogenesis in mice (Tuoc *et al.* 2013) and in neural differentiation of mESCs (Lessard *et al.* 2007). The aim was to determine whether the knockdown of these two genes in MB can influence active DNA demethylation and 5caC levels as well as the transcription levels of key genes involved in active DNA demethylation and in the development of MB.

As such, we first performed siRNA mediated transient knockdown of *RCOR2* in UW228-3 MB cell line, using a chemical transfection reagent. We first optimised the conditions in order to achieve the highest possible efficiency of RCOR2 knockdown without affecting cell integrity and viability (Fig. 5.11 A). The levels of *RCOR2* were then assessed 48 hours after the treatment using qPCR compared against the non-siRNA treated cells; after the optimisation

we achieved approximately 95% downregulation of *RCOR2* transcripts (Fig. 5. 11 B).

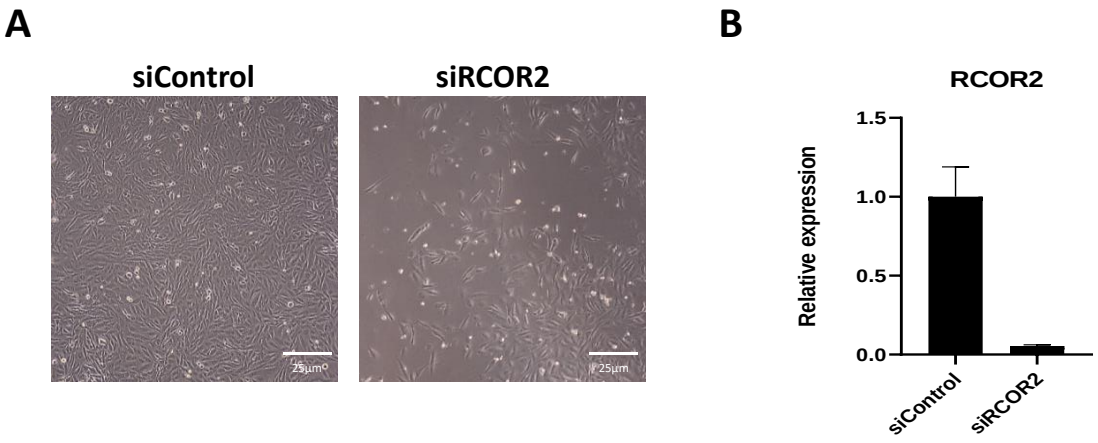


Figure 5.11: Successful siRNA mediated knockdown of RCOR2 in the MB cell line UW228-3.

A. Light microscopy images show that cells' morphology and viability do not change 48 hours upon siRNA treatment. Scale bar = 25 µm. B. Relative expression of *RCOR2* transcripts at the siRNA treated sample compare with the control (50 nM siRCOR2 and 2 µl Dharmafect); indicating 95% knockdown of *RCOR2.* Mean values of three individual samples are shown (n = 3). Error bars are shown as ± SD.

After optimisation of the siRNA treatment and the verification of the knockdown, we performed qPCR experiments to first examine the levels of the key genes involved in active DNA demethylation. *RCOR2* knockdown did not cause any effects in the transcript levels of *TETs* transcripts nor *TDG* transcript (Fig. 5.12 A). We also examined the transcript levels of key genes involved in pluripotency, that are related with the cancer stem cells subpopulation of MB including *NANOG, NESTIN* and *SOX2* (Huang *et al.* 2016a). Interestingly, we observed the upregulation of key cancer stem cell markers including *NOTCH1* and *NESTIN* (Fig. 5.12 C) indicating a potential interplay of RCOR2 and the cancer stem cells compartment of the tumour. Moreover, genes involved in SHH driven MB, Beta-arrestin-2 *(ARRB2),* Hedgehog-Interacting Protein *(HHIP),* and *GLI3* were significantly upregulated in the siRNA treated samples (Fig. 5.12 D).

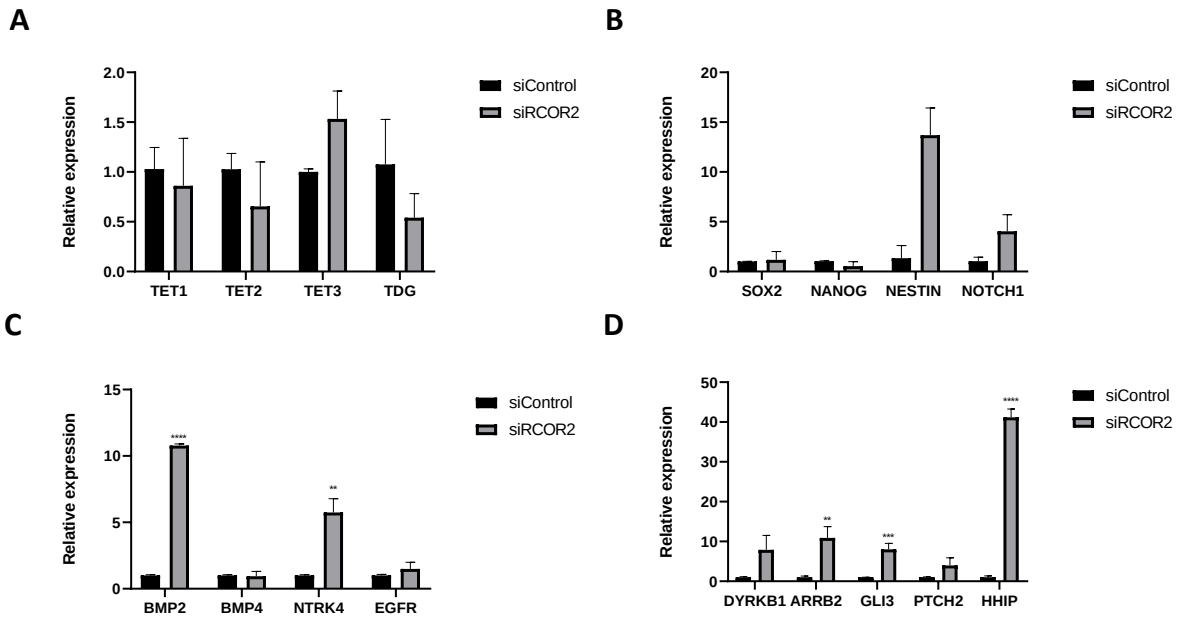


Figure 5.12: Expression levels of DNA methylation, SHH pathway and cancer stem cell related genes upon RCOR2 knockdown.

A. Transcripts involved in active DNA demethylation. B, C. Transcripts important for cancer stem cells proliferation and maintenance. D. Transcripts related to SHH signalling pathway. Mean values of three individual experiments (n = 3) are shown. Error bars are shown as ± SD. Un-paired t-test was performed: *:p < 0.05, **:p < 0.01, ***:p < 0.01, ****:p < 0.001.

Next, we performed our sensitive immunocytochemistry protocol to examine the levels of 5caC/5hmC upon *RCOR2* transient knockdown. We used high resolution confocal microscopy imaging in order to obtain the absolute intensity values of 5hmC and 5caC (Fig. 5.13 A). Signal intensity levels per cell in each condition were calculated using image J and statistical analysis of the results was performed. We observed significantly lower signal intensities ($p < 0.001$) for 5caC in RCOR2 siRNA treated samples whereas 5hmC levels remain unaltered (Fig. 5.13 B). These observations do not correlate with our qPCR findings where *TET* levels were not altered in the *RCOR2* knockdown samples. Our results therefore, indicate that RCOR2 might be recruiting or stabilising 5caC in MB in a TET-independent manner.

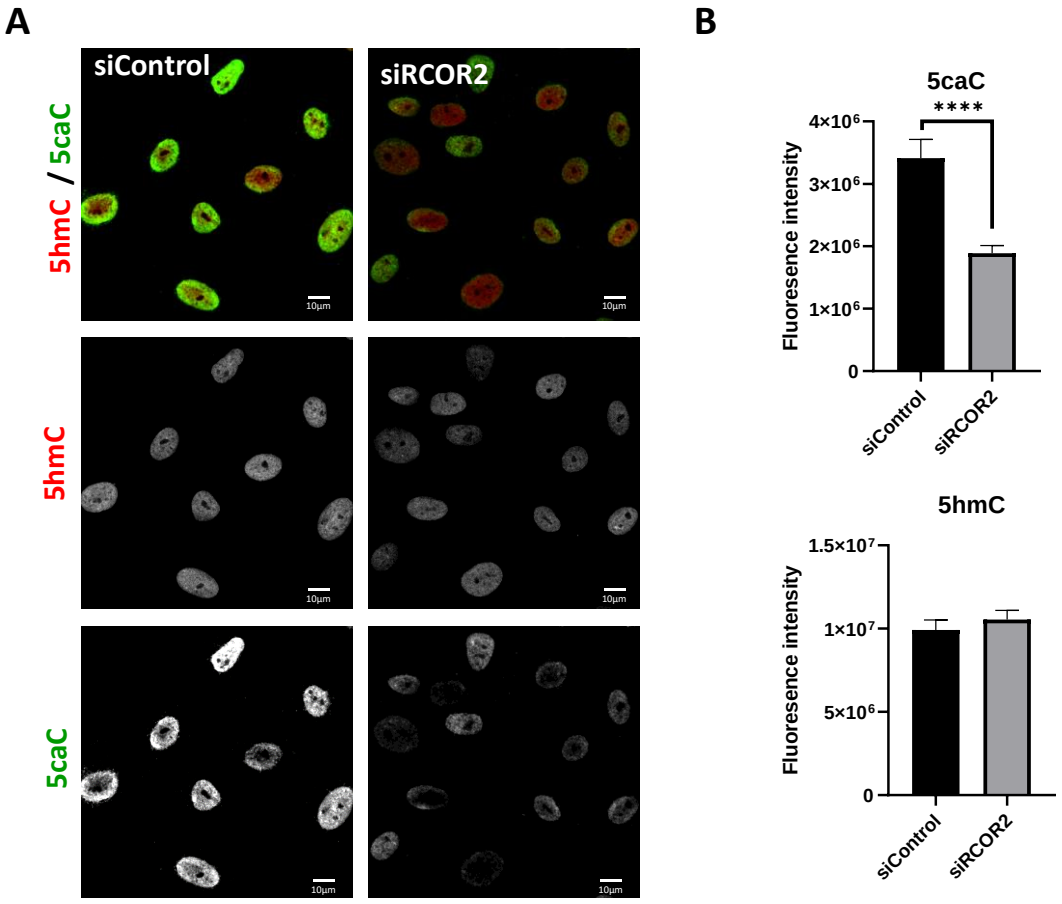


Figure 5.13: RCOR2 knockdown leads to reduced levels of immunocytochemically detectable 5caC.

A. Co-detection of 5caC/5hmC in siRNA control and siRNA treated cells. Merged views and individual channels are shown. The cell cultures were stained in parallel at the same experimental conditions and were imaged at identical settings. Scale bar = 10 µm. C. Quantification of 5hmC and 5caC signals. Mean values of the average intensities of each individual cell for each cell population are shown in two independent experiments (n = 2). Error bars are shown as ± SEM. Un-Paired t test has been performed. *:p < 0.05, **:p < 0.01, ***:p < 0.01, ****:p < 0.001.

5.3.4. SMARCC2 siRNA mediated knockdown leads to elevated levels of 5hmC/5caC in MB

We next performed siRNA mediated knockdown of *SMARCC2,* which was highly expressed in both MB and EPN cell lines. First, the conditions for *SMARCC2* siRNA knockdown were optimised using different concentrations of both the transfection reagent and the siRNA reagent (Fig. 5.14 A and B). After

optimisation of the conditions, we assessed the efficiency of the siRNA knockdown using qPCR 48 hours after the siRNA treatment. We achieved a high efficiency of 95% *SMARCC2* knockdown in the siRNA treated samples (Fig. 5.14 C).

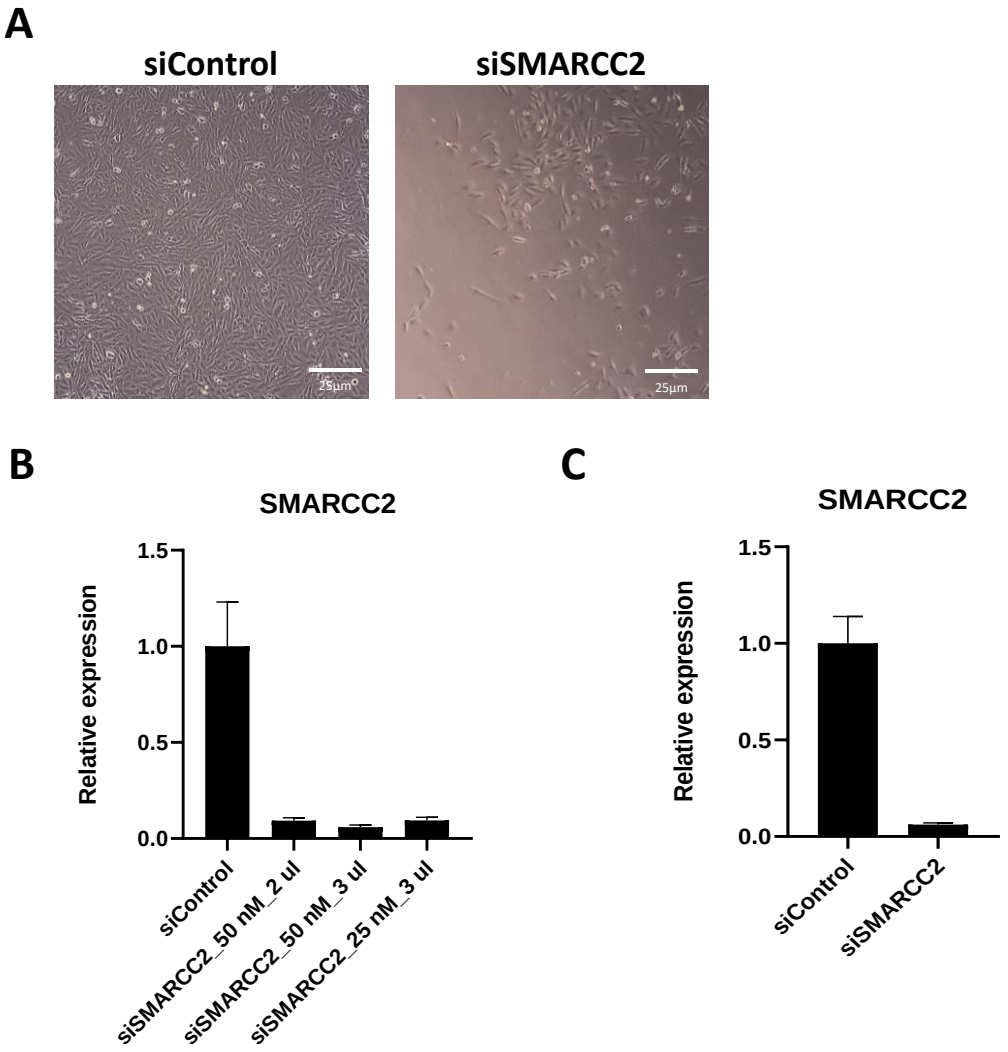


Figure 5.14: siRNA mediated knockdown of SMARCC2 in the MB cell line UW228-3.

A. Light microscopy images show that cells' morphology and viability do not change 48 hours upon siRNA treatment. Scale bar = 25 µm. B. Optimisation of the siRNA mediated knockdown of *SMARCC2* in UW228-3 cell line. Relative expression of *SMARCC2* in 2 different concentrations of the siRNA and the transfection reagent; 3 different conditions in total. 50 nM of siRNA and 3 µl of Dharmafect lead to 95% knockdown of SMARCC2. C. Relative expression of *SMARCC2* transcripts at the siRNA treated sample compare with the control; indicating 95% knockdown of *SMARCC2*. Mean values of three individual experiments (n = 3) are shown. Error bars are shown as ± SD.

Similar to our previous experiments, we then performed qPCR experiment to measure the relative expression levels of genes involved in active DNA demethylation as well as MB progression. We did not observed any alterations at the expression levels of *TET1/2/3* and *TDG* transcripts in SMARCC2 knockdown samples (Fig. 5.15 A). Nevertheless, we observed significantly lower levels of *SOX2* and *NOTCH1,* key regulators in the maintenance of cancer stem cells population in MB (Fig. 5.15 B). Key genes involved in the SHH pathway including *GLI3* and Patched 2 *(PTCH2)* were upregulated upon *SMARCC2* knockdown (Fig. 5.15 D).

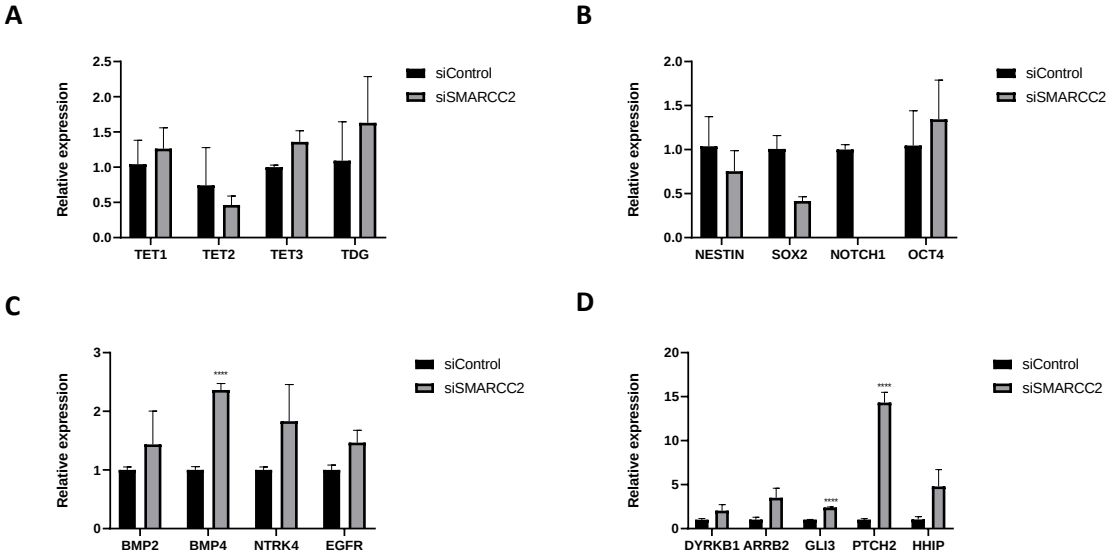


Figure 5.15: Alterations in gene expression levels of DNA methylation and cancer related markers changes upon SMARCC2 knockdown.

A. Transcripts involved in active DNA demethylation. B, C. Transcripts important for cancer stem cells proliferation and maintenance. D. Transcripts related to the SHH signalling pathway. Mean values of three individual experiments (n = 3) are shown. Error bars are shown as ± SD. Un-paired t-test was performed: *:p < 0.05, **:p < 0.01, ***:p < 0.01, ****:p < 0.001.

We next examined the levels of both 5hmC and 5caC using immunohistochemistry followed by confocal microscopy imaging. In contrast to our observations after RCOR2 depletion, SMARCC2 knockdown led to elevated levels of both 5hmC and 5caC ($p < 0.001$) (Fig. 5. 16). The elevated levels of these epigenetic marks do not correlate with TET levels which remain stable after the knockdown (Fig. 5.15 A). Our results indicate that SMARCC2 might be a negative regulator in the maintenance of 5hmC and 5caC in the DNA of MB independent from TET activity.

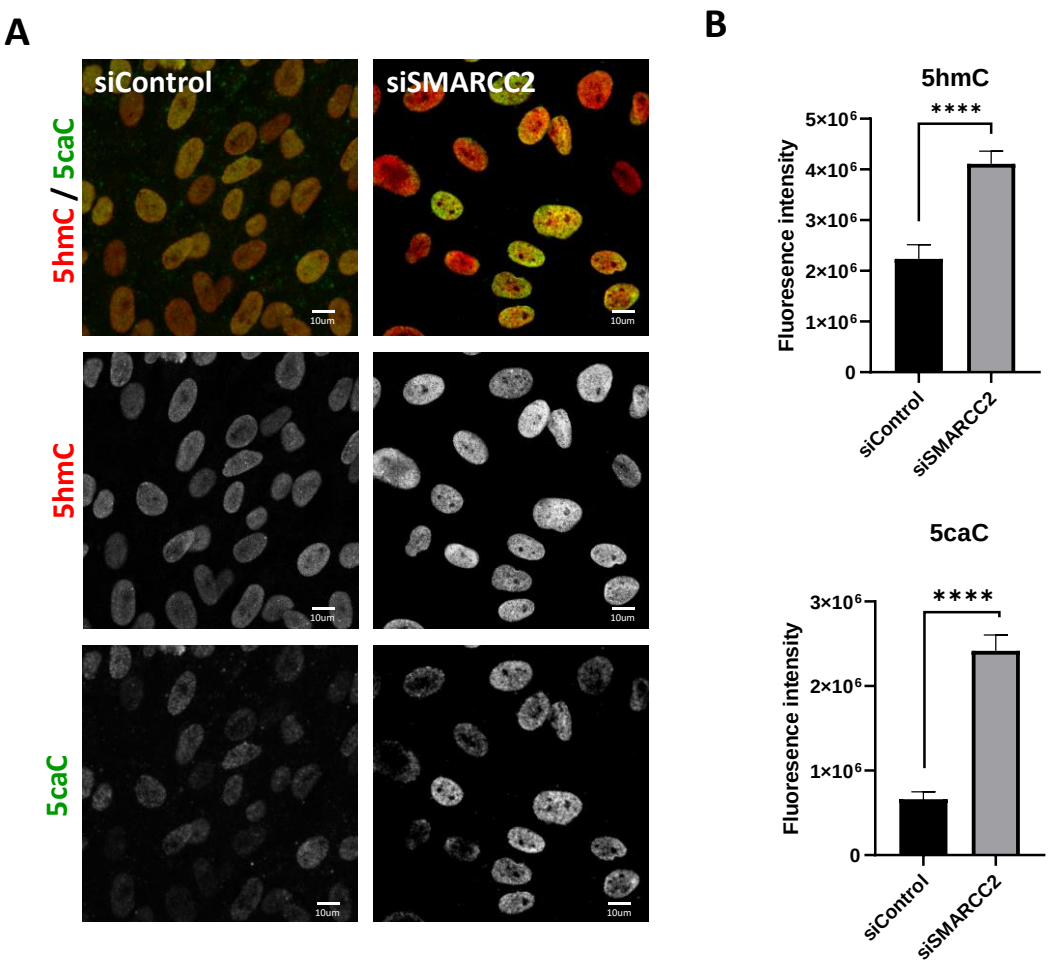


Figure 5.16: SMARCC2 knockdown leads at increased immunocytochemically detectable levels of 5caC/5hmC.

A. Co-detection of 5caC/5hmC in siRNA control and siRNA treated cells. Merged views and individual channels are shown. The cell cultures were stained in parallel at the same experimental conditions and were imaged at identical settings. Scale bar = 10 µm. C. Quantification of 5hmC and 5caC signals. Mean values of the average intensities of each individual cell for each cell population in two individual experiments (n = 2) are shown. Error bars are shown as ± SEM. Un-Paired t test has been performed. *:p < 0.05, **:p < 0.01, ***:p < 0.01, ****:p < 0.001.

5.4. Technical discussion

Following up on the previous publications from our group in which we showed the presence of 5caC in brain and breast cancer tissues (Eleftheriou *et al.* 2015) and in paediatric brain tumour cell lines (Ramsawhook *et al.* 2017) as well as our findings in the first chapters of this thesis. We next sought to examine the biological role of 5caC presence in both adult and paediatric brain tumours. Thus, we studied the readers of 5caC as identified by a study in mESCs using mass spectrometry approach (Spruijt *et al.* 2013). Our initial screening of 5caC potential readers revealed that RCOR2 and SMARCC2 are highly expressed in both MB and EPN cell lines characterised by high levels of 5caC. Therefore, we proceeded with loss of function experiments using siRNA mediated knockdown approach. Our results showed that, RCOR2 knockdown leads to significantly lower levels of 5caC in the UW228-3 MB cell line while SMARCC2 leads to elevated levels of both 5hmC and 5caC in the same cell line. Our observations therefore gave us a good indication that RCOR2 and SMARCC2 might act as 5caC readers in MB.

We performed our initial screen of gene expression of selected 5caC potential readers in our cell lines compared to HeLa. The rationale for this comparison comes from previous observations in our lab that HeLa cells do not exhibit immunocytochemically detectable levels of 5caC whereas there is a subpopulation of cells in MB and EPN with increased levels of 5caC (Ramsawhook *et al.*, 2017). Therefore we made the assumption that the differences we observed in the gene expression levels between HeLa and our cell lines examined could possibly be explained by the presence of 5caC in our system.

Nevertheless our approach has several limitations. The first one, is that the HeLa cell line was first established 70 years ago and it has been extensively used for cervical cancer research and drug screening; however it has been argued that HeLa genome has been transformed during the years in culture and is no longer a human line (Landry *et al.* 2013) making HeLa cells a less favourable model to use nowadays. Thus, the differences we observed in the relative expression of 5caC potential readers in our system may be due to the

transformation of the HeLa cell line if it is indeed derived from completely different tissue with different gene expression profiling. Ideally, we could use human astrocytes or human NSCs as a better comparison control however, we did not have access to these samples at that time. Another limitation is the variability that has been observed in gene expression profiling of different HeLa clones (Frattini *et al.* 2015); to overcome this we used the same clone of HeLa cells throughout our experiments in order to limit variability. Thus, we proceeded with using HeLa cells for the initial screening of the readers.

MB and EPN represent the two most common forms of malignant paediatric brain tumours while GBM represents the most aggressive adult brain tumour. They have all been categorised into molecular subgroups based on their transcriptome and methylome characteristics (Johnson *et al.* 2010; Taylor *et al.* 2012; Verhaak *et al.* 2010) supporting the hypothesis that DNA methylation and demethylation plays a key role in the development and pathogenesis of brain cancer. We used BXD-1425 EPN and DKFZEP1NS EPN cell lines, that had been isolated from patients with recurrent supratentoria anaplastic EPNs (Milde *et al.* 2011; Yu *et al.* 2010) which is the most aggressive subgroup of EPN. Both of these cell lines harbour a C11orf95-RELA fusion protein which has the ability to translocate to the nucleus and activate the NF-κB pathway transforming NSCs; which are considered to be the cell of origin of EPN (Parker *et al.* 2014). We also used UW228-3 and DAOY MB cell lines which harbour TP53 mutations and therefore represent a very high risk subgroup of SHH driven MB (Kunkele *et al.* 2012; Ramaswamy *et al.* 2016). For GBM, LN229 and LN18 were used, both of which harbour TP53 mutations, deletions in p16 /p14ARF tumour suppressor genes and are PTEN wild type and also U87-MG, which is PTEN mutated and contains p16/p14ARF deletion (Lines *et al.* 1999).

Our results from the initial screening of the cell lines examined showed that EPN and MB display similar gene expression patterns with some genes not having a distinct pattern of expression among the cell lines (Fig. 5.7). Both paediatric brain tumours are characterised by deregulation of signalling pathways important for embryonic brain development including SHH, WNT and Notch (de Bont *et al.* 2008). Moreover, both EPN and MB have been found

to contain a population of cancer stem- like cells (Milde *et al.* 2012; Vanner *et al.* 2014).

Our screening of the 5caC potential readers in GBM revealed further cell line heterogeneity with U87-MG presenting a more unique pattern of expression compared to LN18 and LN229 (Fig. 5.8). Differences in gene expression levels are mainly because they are derived from different subtypes of GBM. A recent study based on genotyping of short tandem repeats (STR) has shown that the DNA of U87-MG cell line differs from that of the original cell line derived from GBM patient almost 50 years ago; suggesting that it is still a glioma cell line but with different characteristics than the original (Allen *et al.*, 2016). Other studies also showed that different culture conditions can alter the transcriptome of GBM cell lines (Lee *et al.*, 2006). Cell lines grown in serum containing medium in adherent conditions deviate from the tumour of origin whereas cell lines maintained in serum-free neural stem cell medium maintain their tumour specific characteristics (Pollard *et al.*, 2009). The differences between the cell lines as well as the culture conditions used might explain the differences in the gene expression profiles observed.

Our screening showed that all 5caC readers examined were not expressed or expressed at low levels in GBM cell lines thus we decided to proceed with our experiments focusing on paediatric brain tumours. We found that SMARCC2 and RCOR2 are highly expressed in all the paediatric brain tumours cell lines examined. Briefly, SMARCC2 is a member of the SWI/SNF family of ATP-dependent chromatin remodelling complexes and regulates transcription through chromatin remodelling (Kadam *et al.* 2000). It is also required for forebrain development in the regulation of proliferation, differentiation and cell survival of neural progenitor cells (Narayanan *et al.* 2015). *RCOR2* gene encodes repressor element 1 (RE1) silencing transcription factor (REST) corepressor 2 protein, a transcriptional repressor of a number of neuronal-specific genes in NSCs and is critical for the maintenance of NSCs (Huang & Bao 2012). It has been shown that REST can promote the development of MB arising from NSCs or NPCs (Fuller *et al.* 2005). The cross talk between REST and the SHH signalling pathway or WNT signalling pathway during early embryonic development might also have an important role during MB formation (Singh *et al.* 2008).

We showed that, after depletion of either RCOR2 or SMARCC2, the levels of members of the active DNA demethylation pathway remain unaltered (Fig 5.12 and 5.15). Rather unexpectedly, 5caC levels were significantly lower in RCOR2 knockdown whereas they were significantly higher in SMARCC2 knockdown. Our findings indicate that RCOR2 and SMARCC2 might act as readers of 5caC in MB in a TET-independent manner. One possible explanation is that RCOR2 might interact with other complexes in order to stabilise the 5caC mark while SMARCC2 is a negative regulator of both 5caC/5hmC. We also found that depletion of either RCOR2 or SMARCC2 leads to the an upregulation of SHH pathway related genes (*DYRKB1, GLI3*, and ARRB2) which were previously found to be upregulated in this type of tumour (Northcott *et al.* 2011), indicating that the knockdown of RCOR2 and SMARCC2 might lead to further activation of the SHH pathway in MB. Moreover, there was an upregulation of certain cancer stem cell markers including *NESTIN* in RCOR2 treated cells. NESTIN can specifically bind to GLI3 promoting the development of SHH MBs enhancing tumour growth (Li *et al.* 2016a). These results indicate that the depletion of RCOR2 and SMARCC2 in MB can result in alterations at 5hmC/5caC levels. However, it can also result in the up regulation of both SHH related genes and cancer stem cell markers. Collectively our results indicate that SMARCC2 and RCOR2 are 5caC readers in MB; however further experiments need to be done to support these findings.

Chapter 6. Main discussion

6.1. Spatial and genomic distribution of 5caC in GBM and MB cell lines

6.1.1. GBM and MB cell lines are characterised by unique methylation marks

The presence of 5caC in cancer tissues has not yet been systematically studied due to the extremely low levels of this modification in the genome as it is up to 100 times fold lower than 5hmC (Ito *et al.* 2011). However, several studies have suggested its distinct role in gene expression regulation; indicating that 5fC and 5caC presence within gene bodies reduce the elongation rate of RNA polymerase II (Kellinger *et al.* 2012; Wang *et al.* 2015a). Therefore, the aim of the first chapter of this thesis was to examine the presence, as well as the spatial and genomic distribution of 5hmC/5caC in brain tumour's cell lines in an attempt to examine its biological role in the development of brain tumours.

The first report studying the levels of 5caC in cancer tissues was published by Chowdhury et al., in 2014 (Chowdhury *et al.* 2014). In this study they designed a biotin-avidin mediated enzyme-based immunoassay (EIA) in order to evaluate the levels of 5hmC/5caC in the blood of patients diagnosed with metastatic lung, pancreatic and bladder cancer versus healthy controls (Chowdhury *et al.* 2014). They showed that global 5hmC levels in the blood of patients with metastatic lung cancer were significantly lower (0.013 ± 0.0003%) compared to healthy controls (1.025 ± 0081%) (Chowdhury *et al.* 2014). Indicating that global 5hmC levels can serve as a biomarker of metastatic lung cancer. Interestingly, 5caC was detectable in the blood samples of both healthy and cancer patients with no significant difference in the levels (Chowdhury *et al.* 2014). In contrast to this study, our studies, using immunohistochemistry showed that the levels of 5caC are elevated in a proportion of breast cancer tissues, independent of 5hmC levels, while 5caC

was not detectable in healthy breast tissues (Eleftheriou *et al.* 2015). These data indicated that there might be preferential oxidation of TET proteins from 5mC to 5caC in a proportion of cancer tissues (Eleftheriou *et al.* 2015). The presence of 5caC in cancer tissues but not in the healthy tissues adjacent indicates a potential role of 5caC and active DNA demethylation in the development of tumours (Eleftheriou *et al.* 2015). In a subsequent study in breast cancer tissues a method for the accurate simultaneous quantification of 5mC and its oxidative derivatives have been developed (Guo *et al.* 2017). This method is based on chemical derivatization coupled with UPLC-electrospray quadrupole time of flight mass spectrometry analysis (LC-HRMS/MS); derivatization was performed to improve separation and enhance sensitivity (Guo *et al.* 2017). Similar to our study (Eleftheriou *et al.* 2015), elevated 5fC and 5caC levels were found in breast cancer tissues compared to healthy controls (Guo *et al.* 2017). Elevated levels of 5caC were also found in prostate cancer tissue samples compared to healthy controls and were associated with favourable prognosis (Storebjerg *et al.* 2018). Quantitative analysis with 2D-UPLC-MS/MS showed that 5hmC levels were 5 fold lower in colorectal carcinoma relative to normal tissues adjacent to tumour (Gackowski *et al.* 2016). Nevertheless, the levels of 5caC and 5fC were also detectable in the healthy tissues adjacent to cancer as well as in colorectal carcinoma displaying although at lower levels (Gackowski *et al.* 2016; Tang *et al.* 2015); in contrast to our previous findings in breast cancer tissues.

Correspondingly, here we show for the first time that GBM cell lines display immunocytochemically and mass spectrometry detectable levels of 5caC (Fig. 3.1 and 3.9). 5caC is present in a sub-population of cells in the GBM cell lines examined, while 5hmC is present in a homogeneous spatial distribution within the nuclei of GBM cell lines examined. However, direct comparison of the absolute immunocytochemically detectable levels of 5hmC and 5caC in GBM cell lines was not possible due to the incorporation of tyramide amplification step in the detection of 5caC. The presence of 5caC in GBM cell lines is accompanied by high TET2 transcript and protein levels. Our siRNA mediated knockdown of TET2 leads to diminished 5hmC/5caC in LN18 GBM cell line but not in UW228-3 MB cell line. These findings indicate that TET2 is responsible for the oxidation of 5mC into 5hmC and 5caC in LN18 GBM cell lines. However,

in MB it is possible that the high levels of TET1 (Ramsawhook *et al.* 2017) are responsible for the establishment of 5hmC and 5cac.

Interestingly, primary GBM cell lines also displayed immunocytochemically detectable levels of 5hmC/5caC (Fig. 3.5). Nevertheless, 5caC spatial distribution was significantly different compared with continuous GBM cell lines indicating that cell culture conditions affect the levels and the distribution of 5caC. Indeed previous studies showed that 5hmC levels decrease rapidly as cells adapt to cell culture; with the concomitant decrease of TET transcript levels (Nestor *et al.* 2012). Moreover a study examining the global levels of 5hmC and 5caC using 2D-UPLC-MS/MS showed a global erasure of 5hmC/5caC levels in the DNA of cultured cells lines compared with DNA from primary malignant tissue (Foksinski *et al.* 2017). Global levels of 5hmC in colorectal cancer tissues were 15-fold lower than the levels in cultured colon cancer cell line (HCT116); while 5caC levels were 14-fold lower (Foksinski *et al.* 2017). Therefore, the differences observed between primary and continuous cell lines could be attributable to cell culture conditions.

Single-base resolution analysis reveals that 5hmC is enriched at CpG islands associated with transcriptional regulators at distal-regulatory regions and LINE elements (Booth *et al.* 2012; Yu *et al.* 2012). 5hmC is strongly enriched at exons near transcriptional start sites of genes marked with H3K27me3 and H3K4me3 indicating its role in transcriptional regulation (Pastor *et al.* 2011). Unlike 5hmC levels, 5fC and 5caC levels are significantly lower in all mammalian tissues and cells (Ito *et al.* 2011). 5fC is enriched in CpG islands of promoters and at poised enhancers (Iurlaro *et al.* 2016; Song *et al.* 2013a); its presence in mESCs is often associated with elevated H3K4me3 and active transcription (Raiber *et al.* 2012);. Moreover, 5fC and 5caC are enriched at active gene promoters of highly expressed genes and enhancers in ESCs (Neri *et al.* 2015). Here, we examined the genomic distribution of 5caC in hiPSCs cell lines in LN18, U251 GBM cell lines and in UW228-3 MB cell lines. In agreement with our mass spectrometry results, 5caC presence in cancer cell lines is significantly lower than hiPSCs. In common with previous findings (Lewis *et al.* 2017; Neri *et al.* 2015; Shen *et al.* 2013; Song *et al.* 2013a; Wheldon *et al.* 2014), 5caC was enriched at protein coding transcripts in all the cell lines examined, indicating its potential role in the regulation of

gene transcription. Nevertheless the extremely low abundance of 5caC in GBM cell lines revealed by 5caC-DIP followed by NGS did not allow us to make any conclusions based on the genomic distribution of this mark in GBM and its role in transcriptional regulation. Therefore, subsequent experiments examining 5caC presence at single base resolution using caMAB-seq, based on bisulfite conversion, should be performed in order to validate our meDIP results.

6.1.2. Different MB subgroups are characterised by distinct 5hmC levels and detectable levels of 5fC/5caC

Deregulation of signalling pathways important for normal brain development such as SHH and WNT is a hallmark of MB pathogenesis (Northcott *et al.* 2011; Taylor *et al.* 2012). Epigenetic modifications and DNA methylation play key roles during normal cerebellar development and MB progression (Roussel & Stripay 2018). Currently, MB can be classified into four distinct subgroups based on their genetic epigenetic and transcriptomic characteristics (Northcott *et al.* 2011, 2017; Taylor *et al.* 2012). The role of DNA methylation is well known during MB pathogenesis; however less is known about the role and the distribution of the oxidative derivatives of 5mC in different MB subgroups. The present study is the first study to examine the absolute levels of 5mC oxidative derivatives (5hmC, 5caC, 5fC) in three of the subgroups of MB through a very sensitive mass spectrometry method.

According to a number of studies, oxidised forms of 5mC are of high importance for tumour pathogenesis (Ficz & Gribben 2014). Levels of 5hmC are significantly reduced in many cancers including MB compared with its levels in normal tissues (Jin *et al.* 2011); this reduction correlates with tumour progression and poor prognosis (Ficz & Gribben 2014). Interestingly our results show a precise correlation of distinct 5hmC levels with different subtypes of MB where group 4 has the highest levels of 5hmC followed by group 3 and SHH (Fig. 3.10). In our experiments, the levels of 5hmC do not correlate with poor prognosis across the MB subgroups. It is also important to note that our results are in contrast with a recent study examining the levels of 5hmC in different primary MB samples using dot-blot assays, that

did not find any correlation between 5hmC levels and MB subtypes (Bezerra Salomão *et al.* 2018). It is likely that these differences are attributable either to different sensitivity of mass spectrometry compared with antibody-based techniques or due to difference between primary and established cell lines.

In agreement with previous publications from our group that showed elevated levels of 5caC in cancer tissues such as breast cancer, gliomas, MB and EPN using immunocytochemistry (Eleftheriou *et al.* 2015; Ramsawhook *et al.* 2017); here, we confirm that 5caC and 5fC are detectable in the majority of the cell lines tested; however the levels of these modifications do not show any correlation with different MB molecular subgroups. Given the embryonal nature of MB (Leto *et al.* 2016), it is possible that the presence of 5caC and 5fC in these tumours may be linked with previously reported accumulation of these marks during the differentiation of neural and glial cells in normal brain development (Wheldon *et al.* 2014). Moreover, the detection of comparatively high levels of 5caC that do not correlate with 5fC or 5hmC content in our experiments implies independent generation of these modifications in these tumours; which might reflect to different cell of origin (discussed in main introduction) in different MB subgroups. This may also indicate potential distinct roles of these oxidative forms of 5mC in MB pathogenesis.

In this context, according to a number of recent studies, in addition to their roles as intermediates in the process of active demethylation, both 5fC and 5caC are likely to play specific roles in the regulation of gene expression (Bachman *et al.* 2015; Wheldon *et al.* 2014). Correspondingly, these modifications have been shown to interact with various transcription factors and proteins involved in chromatin remodelling and DNA repair (Spruijt *et al.* 2013). Furthermore, 5caC may also directly retard Pol II elongation on gene bodies (Wang *et al.* 2015a). Relatively high content of this modification in MB suggest that these mechanisms may also be involved in the pathogenesis of these tumours.

MB represents a heterogeneous disease and current approaches of treatment with standard chemotherapy are clearly outmoded (Crawford *et al.* 2007; Martin *et al.* 2014). The diversity of genetic and epigenetic events within MB (Cavalli *et al.* 2017; Northcott *et al.* 2017) warrants the need for more

personalised therapies. Thus, conventional chemotherapy drugs used for MB treatment generally represent cytotoxic agents that interfere with DNA replication and non-specifically target proliferating cells leading to numerous side effects including increased risk of infection, nausea fatigue and alopecia (Crawford *et al.* 2007; Martin *et al.* 2014). Apart from the short term side effects, survivors of MB often experience an onset of neurological conditions post treatment such as hearing loss, seizures, impairments in balance and coordination (King *et al.* 2017). The diversity of genetic and epigenetic events within different molecular subgroups of MB (Cavalli *et al.* 2017; Northcott *et al.* 2017) suggest that epigenetic approaches for therapy may be beneficial.

At present, several drugs targeting different subgroups of MB are undergoing clinical and pre-clinical trials (Coluccia *et al.* 2016). Thus, Vismodegib and Sonidegib are approved by the FDA for the treatment of basal cell carcinoma (Migden *et al.* 2015; Sekulic *et al.* 2012) and have been shown to inhibit Smoothened (SMO) in SHH-driven MB (Kool *et al.* 2014). However, these drugs are effective only in a subset of tumours that harbour mutations in the genes acting upstream of SMO in SHH pathway (Kool *et al.* 2014). Bromodomain and extra-terminal motif (BET) inhibitors have also been shown to decrease tumour cell viability in group 3 MB in preclinical studies (Bandopadhayay *et al.* 2014; Bolin *et al.* 2018). Moreover, high throughput drug screening for MYC driven MB identified HDAC inhibitors as potential therapeutic agents, as their administration combined with treatment with phosphoinositide 3-kinase (PI3K) inhibitors leads to tumour growth inhibition and prolongs survival of MYC driven MB bearing mice (Pei et al. 2016). Thus, novel markers that can be used for classification of MBs into different molecular subgroups are of particular interest due to the need for personalised therapies for treatment of these highly heterogeneous tumours.

Recent studies suggest that liquid biopsy may represent a quick non-invasive approach for early diagnosis of cancer progression (Saenz-Antoñanzas *et al.* 2019; Zeng *et al.* 2019). Tissue biopsy is invasive and, especially in the case of brain tumours, clinically risky as it requires surgical resection of the tumour that is linked with the risks of damaging the normal brain tissue potentially leading to catastrophic outcome (Ilié & Hofman 2016; Saenz-Antoñanzas *et al.* 2019). Moreover in addition to the risks of surgery, the information

obtained using this approach is often limited as the biopsies are usually acquired from a single-region within the tumour and, thus, the intra-tumour heterogeneity is not measured (Sabaawy 2012). In contrast to tissue biopsy, liquid biopsy offers a less invasive tool for cancer diagnostics based on the analysis of circulating materials such as circulating free DNA (cfDNA) a fragmented cellular DNA that is released into the bloodstream by cells undergoing apoptosis and necrosis (Jahr *et al.* 2001; Wan *et al.* 2017). In healthy people, it is derived from apoptotic hematopoietic cells (Lehmann-Werman *et al.* 2016), however in cancer patients, it can be of tumour tissue and microenvironment origin (Wyatt *et al.* 2017). There is an increasing number of clinical trials based on the 5hmC profiling of cfDNA in tumour patients-derived samples (Zeng *et al.* 2019). Importantly, studies performed on colorectal and gastric cancers, indicate that cfDNA can provide details of 5hmC content that can potentially be used to identify various cancer types and track tumour stages (Li *et al.* 2017; Song *et al.* 2017). In this context, our results indicate that the assessment of global 5hmC levels in cfDNA may allow early, quick, non-invasive stratification of patients into different MB subtypes as well as contributing to the development of novel targets for specific treatment of these tumours.

6.2. Catalytically inactive TET2 protein leads to diminished 5hmC levels in LN18 cells resulting in impaired tumorigenic capacity

It is well established that disrupted DNA methylation contributes to initiation and progression of tumorigenesis (Baylin and Jones 2011; Esteller, 2007). Previous studies in our group demonstrated transient accumulation of 5caC/5fC mediated by TET enzymes during differentiation into neuronal and hepatic lineages, while levels decreased drastically in terminally differentiated cells (Lewis *et al.* 2017; Wheldon *et al.* 2014). Therefore, we hypothesised that 5caC presence in GBM cell lines might indicate glioma stem cells' differentiation state; and active DNA demethylation might be operative in this system contributing to tumorigenesis. Our previous findings based on transient knockdown of TET2 indicate that TET2 might be responsible for the

oxidation of 5mC into 5hmC and 5caC in LN18 GBM cell line. Therefore, we generated a TET2 KO cell line in order to further examine our hypothesis on the role of active DNA demethylation via TET2 mediated oxidation in GBM tumorigenesis.

Reduced TET2 enzymatic activity in malignancies and diseases has been correlated with reduced 5hmC levels (Calabrese *et al.* 2014; Lian *et al.* 2012; Puig *et al.* 2018; Tong *et al.* 2019; Zhang *et al.* 2019). Nevertheless, according to Kraus et al., alterations at TET2 genes in gliomas are not responsible for the reduced 5hmC levels observed in these tumours (Kraus *et al.* 2015); suggesting that nuclear exclusion of TET proteins in gliomas might be responsible for the reduced oxidation (Müller *et al.* 2012). In this study we showed that the introduction of a stop codon before the catalytic domain of TET2 protein leads to catalytically inactive enzyme; resulting in diminished 5hmC levels in the GBM cell line examined while the levels of TET1 gene remain unaltered (Fig. 4.19 and 4.20). Additionally, heterozygous inactivation of TET2 also leads to 6 fold lower 5hmC levels, suggesting that disruption of only one allele of TET2 can significantly affect the oxidation of 5mC into 5hmC (Fig. 4.19 and 4.20). Moreover, in agreement with previous studies in glioma stem cells (Zhou *et al.* 2018a), TET2 KO in our system led to significantly lower levels of 5fC indicating its role in establishing 5fC in GBM. Therefore, suggesting that TET2 is responsible for the active DNA demethylation occurring in LN18 GBM cell line.

Although there has been huge progress in the study of 5mC oxidative derivatives on DNA with methods including immunohistochemistry, meDIP, bisulfite sequencing and mass spectrometry there is still a diversity in their use. Therefore, a future standardisation would be useful in order to establish a cutoff to identify 5caC positive versus negative DNA bases. The discrepancy we observed at the levels of 5caC between our immunocytochemistry experiment and the mass spectrometry experiment is striking. Due to the extremely low levels of 5caC in the genome, our immunocytochemistry experiment involves a tyramide amplification step which allows the amplification of the signal increasing the detection ability (Abakir *et al.* 2016). On the other hand the 2D-UPLC-MS/MS method applied has the ability to detect extremely low levels for 5caC. Therefore, it is possible that the levels

of 5caC detected by mass spectrometry might be a result of false positive signal. Moreover, it could be possible that 5caC is present on the RNA in GBM cell lines and our observations in mass spectrometry experiments could be due to RNA contamination in the samples. Emerging evidence shows that TET enzymes have the ability to catalyse the formation of 5hmC (Fu *et al.* 2014), 5caC and 5fC in RNA *in vitro* (Basanta-sanchez *et al.* 2017) while endogenous 5hmC, 5fC and 5caC was found in the RNA in mammals (Huang *et al.* 2016b).

A growing body of evidence suggests TET2 has a role as a tumour suppressor in haematological malignancies (Ko *et al.* 2010, 2015; Li *et al.* 2011; Quivoron *et al.* 2011; Wang *et al.* 2015b; Weissmann *et al.* 2012). Indicating that inactive TET2 leading to reduced 5hmC levels contributes to the development of haematological malignancies. Moreover, according to a more recent study, TET2 inactivation in leukaemia progression leads to aberrant self-renewal while TET2 restoration of activity with Vitamin C induction reverses the phenotype (Cimmino *et al.* 2017). In contrast to these findings, our results revealed that TET2 KO does not impair either the proliferation of GBM cells or the proliferation of GBM cells enriched for glioma stem cells population (Fig. 4.21 and 4.22). Indicating a possible distinct role of TET2 in GBM compared to haematological malignancies.

It has previously been shown that the majority of the tumours are characterised by a hierarchically organized system that includes cancer stem cells with quiescence and self-renewal capacity within the tumours. These cells are responsible for tumour maintenance and recurrence (Behnan *et al.* 2017; Chen *et al.* 2012; Singh *et al.* 2004). GBM contains distinct cellular subpopulations that express neural stem and progenitor cell markers including NESTIN, SOX2 and OLIG2 (Singh *et al.* 2004; Suvà *et al.* 2014). Glioma stem cells identified within the tumour bulk share biological similarities with NSCs however they also have distinct alterations favouring their malignant growth (Zhou *et al.* 2018a).

Previous studies showed increased TET2 levels in glioma stem cells of primary GBM tumours correlating with stem markers, including SOX2 (Zhou *et al.* 2018a). Our data showed that key markers associated with glioma stem cells maintenance and proliferation are downregulated in the TET2 KO clone (Fig.

4.23). Interestingly, TET2 KO cells cultured in conditions enabling the proliferation of glioma stem cells showed limited tumorigenic ability *in vitro* (Fig. 4.24 and 4.25). A recently published study showed that TET2 enzymatic activity has the ability to control glioma cancer cells survival and recurrence (Puig *et al.* 2018). It has been shown that, TET2 was highly expressed in the subpopulation of cancer stem cells of different tumour types including colon cancer, myeloma and GBM (Puig *et al.* 2018). TET2 knockout in cancer stem cells in colon cancer resulted in increased apoptosis *in vitro* and delayed subcutaneous tumour regrowth *in vivo,* increasing progression-free survival (Puig *et al.* 2018). Moreover, high TET2 levels in glioma stem cells contributes to their chemo-resistance suggesting TET2 potential association with DNA damage (Zhou *et al.* 2018a). Our results, similar to Puig et al., showed that TET2 knockout could inhibit growth of cancer stem cells *in vitro* in GBM cell lines. Moreover, current evidence indicates that TET2 expression levels correlates with a short relapse time in recurrent GBM suggesting that inhibition of TET2 could be used as a therapeutic approach for recurrent GBM (Briand *et al.* 2018). Similarly, patients with colorectal carcinoma that showed enrichment of TET2 levels and treated with chemotherapy, relapsed significantly earlier (Puig *et al.* 2018).

Another study, however, showed that ectopic expression of TET2 in U87-MG and U251 resulted in the inhibition of tumour invasiveness and cell proliferation *in vitro* and growth *in vivo* (Chen *et al.* 2017). However, in this study, cells were grown in adherent conditions and there is no indication of the phenotype of TET2 overexpression when cells are cultured in glioma stem cell enrichment medium. Although we observed reduced tumorigenic capacity of TET2 KO cells that does not mean that TET2 could act as a tumour suppressor in the bulk tumour. Another study suggests that in colon cancer TET2 could inhibit proliferation only in the population of slow-cycling cancer stem cells (Puig *et al.* 2018). It is well known that GBM is characterised by intratoumoral heterogeneity, our results showed that, TET2 blocks the formation of neurospheres *in vitro*; however it does not show any effect when cells are not enriched in cancer stem cells conditions.

6.3. 5caC potential readers in adult and paediatric brain tumours cell lines. The implication of RCOR2 and SMARCC2 in active DNA demethylation in MB development.

The aim of the last chapter of this thesis was to study the 5caC potential readers in adult and paediatric brain tumour cell lines. A study based on a mass spectrometry approach identified the readers of 5mC oxidative derivatives in mESCs (Spruijt *et al.* 2013). However, although there are several studies identifying 5caC readers in mESCs (Shin *et al.* 2016; Song *et al.* 2016; Spruijt *et al.* 2013); their role in interpreting 5caC has not been systematically assessed in brain tumours. We therefore, selected 5caC unique readers and studied their expression in brain tumours cell lines with regards to their contribution to active DNA demethylation pathway.

Previous studies in our group showed that EPN and MB paediatric tumour cell lines are characterised by high levels of 5caC (Ramsawhook *et al.* 2017). In contrast, the HeLa cell line does not have any immunocytochemically detectable 5caC levels (Ramsawhook *et al.* 2017). Thus, we proceeded with the initial screening of the expression levels of 5caC readers in brain tumour cell lines relative to HeLa using qPCR. Our results showed that RCOR2 and SMARCC2 levels are elevated in MB and EPN cell lines (Fig. 5.1 and 5.4). Hence, loss of function experiments using a siRNA mediated knockdown approach were performed for *RCOR2* and *SMARCC2* genes. Our results indicate that knockdown of both RCOR2 and SMARRC2 impairs 5caC levels in UW228-3 MB cell line (Fig. 5.13 and 5.16). These findings suggest that RCOR2 and SMARCC2 expression interprets 5caC oxidation in UW228-3 MB cell line.

SMARCC2 belongs to the SWI/SNF complex, an ATP-dependent chromatin remodelling complex which regulates chromatin structure acting either as a transcription repressor or as an activator (Hargreaves & Crabtree 2011; Wu 2012). Inactivation of the SWI/SNF complex impairs the appropriate cell response to signalling preventing cell differentiation and sustaining cell

growth (Romero & Sanchez-Cespedes 2014). It has been shown that conditional knockdown of SMARCC2 in mice represses cortical neurogenesis resulting in an enlarged cortex while overexpression of SMARCC2 rescues that phenotype (Tuoc *et al.* 2013). Although SMARCC2 is not expressed in ES cells it has been shown that differentiation into neural progenitor cells leads to its activation (Lessard *et al.* 2007). *SMARCC2* mutations have also recently been related with neurodevelopmental abnormalities; while individuals with variable neurodevelopmental delay and growth retardation have mutations in the *SMARCC2* gene (Machol *et al.* 2019).

These findings indicate that SMARCC2 might have a distinct role in the regulation of chromatin during neuronal differentiation. Considering that MB develops from the deregulation of critical pathways involved in neuronal differentiation (Kool *et al.* 2012; Northcott *et al.* 2012b, 2017), SMARCC2 knockdown in MB cell lines would also give an indication of the phenotype in MB development. The UW228-3 MB cell line is derived from a SHH MB therefore it has an activated SHH pathway. Thus, following siRNA mediated knockdown of SMARCC2 we examined the levels of key genes found to be upregulated in SHH derived MB (Northcott *et al.* 2011). Our results show a further upregulation of *ARRB2, GLI3* and *PTCH2* (Fig. 5. 15). Indicating that SMARCC2 downregulation might lead to a further activation of the SHH pathway in UW228-3 cell lines. Moreover, in parallel with this, we observed elevated 5caC levels in SMARCC2 knockdown, which suggests that SMARCC2 might be a negative regulator of 5caC. It could be possible therefore that SMARCC2 acts as a 5caC reader in SHH MB interpreting active DNA demethylation as well as having an effect in SHH pathway driven MB.

RCOR2 is a repressor of the REST/neuron-restrictive silencer factor (REST/NRSF) which is a critical regulator of neuronal differentiation, acting by repressing the expression of neuronal genes in non-neuronal cells and neural progenitors (Chong *et al.* 1995; Negrini *et al.* 2013; Schoenherr & Anderson 1995). It has been suggested that REST can act as a tumour suppressor in non-neural tumours and as an oncogene in neural tumours (Negrini *et al.* 2013). Overexpression of REST and MYC in NSCs blocks neuronal differentiation and leads to MB development *in vivo* (Su *et al.* 2006). Interestingly, MB is characterised by high levels of REST/NRSF compared with

neuronal progenitor cells (Lawinger *et al.* 2000). The replacement of the NRSF repressor domain with the activation domain of viral protein 16 (VP16) and its transient expression in MB cells leads to stimulation of neuronal promoters as well as inhibited proliferation and growth of MB cell lines (Lawinger *et al.* 2000). Intratoumoral injection of the REST-VP16 complex in nude mice harbouring MB tumours led to inhibition of their growth (Lawinger *et al.* 2000).

According to a recent study RCOR2 and RCOR1 form a complex with the TF insulinoma-associated 1 (INSM1) that induces cell cycle arrest in the embryonic brain (Monaghan *et al.* 2017). Deletion of both RCOR1 and RCOR2 in mice leads to embryonic lethality and defects in brain development with more neural progenitors but fewer terminally differentiated neurons and oligodendrocytes (Monaghan *et al.* 2017). REST transcripts are upregulated in the absence of RCOR2 and reducing REST transcripts partially rescues the phenotype (Monaghan *et al.* 2017). Suggesting that elimination of RCOR1 and RCOR2 promotes neural proliferation over differentiation (Monaghan *et al.* 2017). In agreement with these observations, our results show upregulation of *NESTIN* and *NOTCH1* genes (Fig. 5.12) which are involved in the proliferation of NPCs (Bernal & Arranz 2018). Suggesting a possible role of RCOR2 in the regulation of neural proliferation in MB development.

Moreover, another study shows the interplay between RCOR2 and SHH pathway in brain development. RCOR2 forms a complex with lysine specific histone demethylase (LSD1) targeting SHH pathway and brain development (Wang *et al.* 2016). Whereas, *in vitro* RCOR2 knockdown leads to decreased neurosphere size indicating that it significantly affects the stemness (Wang *et al.* 2016). Mouse embryos with RCOR2 KO display reduced brain sizes and structural abnormalities of the cortex affecting neocortical development. As a consequence of alterations in chromatin structure through nucleosomal demethylase activities of LSD1 complex at the promoter regions of key genes involved in SHH (Wang *et al.* 2016). Transcriptome analysis of RCOR2 KO cells reveals upregulation of genes involved in neuronal differentiation as well as genes involved in SHH pathway (Wang *et al.* 2016). In agreement with these studies, our results show that RCOR2 knockdown in SHH MB cell line leads to the upregulation of genes involved in SHH pathway (Fig. 5.12)

indicating a possible role of RCOR2 role in the activation of SHH pathway in MB development.

siRNA mediated knockdown of RCOR2 not only led to the upregulation of key genes involved in NPC proliferation and in the SHH pathway but also led to significantly lower levels of 5caC. Our findings show that 5caC levels upon RCOR2 depletion were significantly lower (Fig. 5.13). Surprisingly no significant changes were observed in the levels of TET transcripts (Fig. 5.12); indicating that the reduction of 5caC levels we observed was independent of TETs enzymatic activity. Our observations in the last chapter of this thesis provide a good indication that RCOR2 and SMARCC2 might act as 5caC readers in MB. However, further evidence is needed to support these findings.

6.4. Concluding remarks and future work

In conclusion, the data generated in this thesis reveal the unique methylomes that exists in adult and paediatric brain tumours. The implementation of immunostaining, confocal microscopy, mass spectrometry and meDIP followed by NGS has revealed the presence of 5caC in brain tumour cell lines. While distinct levels of 5hmC in different subgroups of MB indicate that 5hmC can be used as a future biomarker for the classification of MB. Moreover, we suggested the potential role of RCOR2 and SMARCC2 as 5caC epigenetic readers in MB development and pathogenesis.

We also showed that, TET2 is responsible for the oxidation of 5mC into 5hmC/5caC/5fC in LN18 cell lines. TET2 CRISPR/Cas9 mediated knockout leads to reduced levels of oxi-mCs derivatives. Moreover, although TET2 enzymatically inactive protein does not impair the proliferation of GBM cells it does affect the stemness of the tumours. We showed that, several key glioma stem cell markers were downregulated in the TET2 KO cells while the cells had significantly lower tumorigenic ability *in vitro*. Indicating that, unlike haematological malignancies, TET2 might act as a tumour oncogene in glioma stem cells; proposing a new novel target for GBM therapy.

Nevertheless, in order to complete this work and consolidate our hypothesis, several experiments need to be conducted. Owing to the differences in the oxi-mCs abundance and distribution in primary and continuous cell lines (Nestor *et al.* 2012), cell lines might not be the ideal model to examine 5caC genomic distribution in GBM as it is possible that the extremely low 5caC levels we observed in GBM cell lines in our meDIP analysis were due to cell culture conditions. Therefore, 5caC and 5hmC meDIP coupled with transcriptomic analysis could be performed in primary GBM cell lines in order to gain a better understanding on the distribution of those marks in GBM as well as their role in transcriptional regulation. Moreover, meDIP of 5hmC in both TET2 KO cell line and LN18 cell line accompanied with transcriptomic analysis would help us identify the genomic distribution of 5hmC prior and after TET2 KO as well as its role in the transcriptional regulation; verifying TET2 potential role in the regulation of transcription of key glioma stem cells markers. Finally, *in vivo* experiments would lead to a better understanding on the effects of TET2 KO in GBM development; injection of TET2 KO cell line into nude mice would confirm our *in vitro* observations.

References

Aapola, U., Shibuya, K., Scott, H. S., … Peterson, P. (2000). Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. *Genomics*, **65**(3), 293–298.

Abakir, A., Giles, T. C., Cristini, A., … Ruzov, A. (2020). N 6-methyladenosine regulates the stability of RNA:DNA hybrids in human cells. *Nature Genetics*, **52**(1), 48–55.

Abakir, A., Wheldon, L., Johnson, A. D., Laurent, P., & Ruzov, A. (2016). Detection of modified forms of cytosine using sensitive immunohistochemistry. *Journal of Visualized Experiments*, **2016**(114), 3–6.

Achour, M., Jacq, X., Rondé, P., … Bronner, C. (2008). The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression. *Oncogene*, **27**(15), 2187–2197.

Agathocleous, M., Meacham, C. E., Burgess, R. J., … Morrison, S. J. (2017). Ascorbate regulates haematopoietic stem cell function and leukaemogenesis. *Nature*, **549**(7673), 476–481.

Alcantara Llaguno, S., Chen, J., Kwon, C. H., … Parada, L. F. (2009). Malignant Astrocytomas Originate from Neural Stem/Progenitor Cells in a Somatic Tumor Suppressor Mouse Model. *Cancer Cell*, **15**(1), 45–56.

Alcantara Llaguno, S. R., Wang, Z., Sun, D., … Parada, L. F. (2015). Adult Lineage-Restricted CNS Progenitors Specify Distinct Glioblastoma Subtypes. *Cancer Cell*, **28**(4), 429–440.

Bachman, M., Uribe-lewis, S., Yang, X., … Balasubramanian, S. (2015). 5-Formylcytosine can be a stable DNA modification in mammals. *Nature Chemical Biology*, **11**(august), 3–6.

Bandopadhayay, P., Bergthold, G., Nguyen, B., … Cho, Y. J. (2014). BET bromodomain inhibition of MYC-amplified medulloblastoma. *Clinical Cancer Research*, **20**(4), 912–925.

Barrangou, R., Fremaux, C., Deveau, H., … Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*, **315**(March), 1709–1712.

Bartoccetti, M., van der Veer, B. K., Luo, X., … Koh, K. P. (2020). Regulatory Dynamics of Tet1 and Oct4 Resolve Stages of Global DNA Demethylation and Transcriptomic Changes in Reprogramming. *Cell Reports*, **30**(7), 2150-2169.e9.

Basanta-sanchez, M., Wang, R., Liu, Z., Ye, X., & Li, M. (2017). TET1-Mediated Oxidation of 5-Formylcytosine (5fC) to 5-Carboxycytosine (5caC) in RNA. *ChemBioChem*, 72–76.

Behnan, J., Finocchiaro, G., & Hanna, G. (2019). The landscape of the mesenchymal signature in brain tumours. *Brain*, **142**(4), 847–866.

Behnan, J., Stangeland, B., Hosainey, S. A. M., … Brinchmann, J. E. (2017). Differential propagation of stroma and cancer stem cells dictates tumorigenesis and multipotency. *Oncogene*, **36**(4), 570–584.

Beier, D., Röhrl, S., Pillai, D. R., … Beier, C. P. (2008). Temozolomide preferentially depletes cancer stem cells in glioblastoma. *Cancer Research*, **68**(14), 5706–5715.

Ben-Hattar, J., & Jiricny, J. (1988). Effect of cytosine methylation on the cleavage of oligonucleotide duplexes with restriction endonucleases HpaII and MspI. *Nucleic Acids Research*, **16**(9), 11063.

Ben-Porath, I., Thomson, M. W., Carey, V. J., … Weinberg, R. A. (2008). An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nature Genetics*, **40**(5), 499–507.

Bernal, A., & Arranz, L. (2018). Nestin-expressing progenitor cells: function, identity and therapeutic implications. *Cellular and Molecular Life Sciences*, **75**(12), 2177–2195.

Bestor, T. H. (2000). The DNA methyltransferases of mammals. *Human Molecular Genetics*, **9**(5), 2395–2402.

Bezerra Salomão, K., Cruzeiro, G. A. V., Bonfim-Silva, R., … Gonzaga Tone, L. (2018). Reduced hydroxymethylation characterizes medulloblastoma while TET and IDH genes are differentially expressed within molecular subgroups. *Journal of Neuro-Oncology*, **139**(1), 33–42.

Bhattacharyya, S., Yu, Y., Suzuki, M., … Verma, A. (2013). Genome-wide hydroxymethylation tested using the HELP-GT assay shows redistribution in cancer. *Nucleic Acids Research*, **41**(16), 1–10.

Bickle, T. A., & Kruger, D. H. (1993). Biology of DNA restriction. *Microbiological Reviews*, **57**(2), 434–450.

Bird, A., Taggart, M., Frommer, M., … Road, M. (1985). A Fraction of the Mouse Genome That Is Derived from Islands of Nonmethylated, CpG-Rich DNA. *Cell*, **40**(January), 91–99.

Biswas, S., & Rao, C. M. (2018). Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. *European Journal of Pharmacology*, **837**(August), 8–24.

Blaschke, K., Ebata, K. T., Karimi, M. M., … Ramalho-Santos, M. (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature*, **500**(7461), 222–226.

Boch, J., & Bonas, U. (2010). Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function . *Annual Review of Phytopathology*, **48**(1), 419–436.

Boch, J., Scholze, H., Schornack, S., … Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**(5959), 1509–1512.

Bolin, S., Borgenvik, A., Persson, C. U., … Swartling, F. J. (2018). Combined BET bromodomain and CDK2 inhibition in MYC-driven medulloblastoma. *Oncogene*, **37**(21), 2850–2862.

Booth, M. J., Branco, M. R., Ficz, G., Oxley, D., & Krueger, F. (2012). Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. *Science*, **336**(May), 934–938.

Bostick, M., Kim Kyong, J., Pierre-Olivier, Estève Amander, C., Pradhan, S., & Jacobsen, S. E. (2007). UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science*, **317**(9), 1760–1764.

Bourc'his, D., & Bestor, T. H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, **431**(7004), 96–99.

Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., & Bestor, T. H. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science*, **294**(5551), 2536–2539.

Branco, M. R., Ficz, G., & Reik, W. (2012). Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nature Reviews Genetics*, **13**(1), 7–13.

Brennan, C. W., Verhaak, R. G. W., McKenna, A., … McLendon, R. (2013). The somatic genomic landscape of glioblastoma. *Cell*, **155**(2), 462–

477.

Briand, J., Nadaradjane, A., Vallette, F. M., & Cartron, P. F. (2018). The TET2 Expression Level Correlates with a Short Relapse Time in Glioblastoma Multiforme. *Journal of Clinical Epigenetics*, **04**(02), 1–5.

Buck-Koehntop, B. A., & Defossez, P. A. (2013). On how mammalian transcription factors recognize methylated DNA. *Epigenetics*, **8**(2), 131–137.

Calabrese, R., Valentini, E., Ciccarone, F., … Caiafa, P. (2014). TET2 gene expression and 5-hydroxymethylcytosine level in multiple sclerosis peripheral blood cells. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, **1842**(7), 1130–1136.

Carén, H., Stricker, S. H., Bulstrode, H., … Pollard, S. M. (2015). Glioblastoma stem cells respond to differentiation cues but fail to undergo commitment and terminal cell-cycle arrest. *Stem Cell Reports*, **5**(5), 829–842.

Carreira, P. E., Ewing, A. D., Li, G., … Faulkner, G. J. (2016). Evidence for L1-associated DNA rearrangements and negligible L1 retrotransposition in glioblastoma multiforme. *Mobile DNA*, **7**(1), 1–14.

Cavalli, F. M. G., Remke, M., Rampasek, L., … Taylor, M. D. (2017). Intertumoral Heterogeneity within Medulloblastoma Subgroups. *Cancer Cell*, **31**(6), 737-754.e6.

Cedar, H., Solage, A., & Hebrew, T. (1979). Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI. *Nucleic Acids Research*, **6**(6), 2125–2132.

Cenciarelli, C., Marei, H. E., Felsani, A., … Mangiola, A. (2016). PDGFRα depletion attenuates glioblastoma stem cells features by modulation of STAT3, RB1 and multiple oncogenic signals. *Oncotarget*, **7**(33), 53047–53063.

Cenciarelli, C., Marei, H. E. S., Zonfrillo, M., … Mangiola, A. (2014). PDGF receptor alpha inhibition induces apoptosis in glioblastoma cancer stem cells refractory to anti-Notch and anti-EGFR treatment. *Molecular Cancer*, **13**(1), 1–15.

Chang, V. Y., Federman, N., Martinez-Agosto, J., Tatishchev, S. F., Nelson, S. F., & 1. (2013). Whole Exome Sequencing of Pediatric Gastric Adenocarcinoma Reveals an Atypical Presentation of Li-Fraumeni Syndrome. *Pediatric Blood & Cancer*, **60**, 57–574.

Charlet, J., Duymich, C. E., Lay, F. D., … Jones, P. A. (2016). Bivalent Regions of Cytosine Methylation and H3K27 Acetylation Suggest an Active Role for DNA Methylation at Enhancers. *Molecular Cell*, **62**(3), 422–431.

Chen, B., Lei, Y., Wang, H., … Liu, L. (2017). Repression of the expression of TET2 by ZEB1 contributes to invasion and growth in glioma cells. *Molecular Medicine Reports*, **15**(5), 2625–2632.

Chen, J., & Li, Q. (2011). Life and death of transcriptional co-activator p300. *Epigenetics*, **6**(8), 957–961.

Chen, J., Li, Y., Yu, T.-S., … Parada, L. F. (2012). A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature*, **488**(7412), 522–526.

Chen, Q., Chen, Y., Bian, C., Fujiki, R., & Yu, X. (2013). TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature*, **493**(7433), 561–564.

Chen, T., Ueda, Y., Dodge, J. E., Wang, Z., & Li, E. (2003). Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b. *Molecular and Cellular Biology*, **23**(16), 5594–5605.

Cho, S. W., Kim, S., Kim, J. M., & Kim, J. S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature Biotechnology*, **31**(3), 230–232.

Chong, J. A., Tapia-Ramirez, J., Kim, S., … Mandel, G. (1995). REST: A mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell*, **80**(6), 949–957.

Chou, N. H., Tsai, C. Y., Tu, Y. T., … Tsai, K. W. (2016). Isocitrate dehydrogenase 2 dysfunction contributes to 5-hydroxymethylcytosine depletion in gastric cancer cells. *Anticancer Research*, **36**(8), 3983–3990.

Chowdhury, B., Cho, I. H., Hahn, N., & Irudayaraj, J. (2014). Quantification of 5-methylcytosine, 5-hydroxymethylcytosine and 5-carboxylcytosine from the blood of cancer patients by an enzyme-based immunoassay. *Analytica Chimica Acta*, **852**, 212–217.

Christian, M., Cermak, T., Doyle, E. L., … Voytas, D. F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, **186**(2), 756–761.

Chylinski, K., Makarova, K. S., Charpentier, E., & Koonin, E. V. (2014). Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Research*, **42**(10), 6091–6105.

Cimmino, L., Dawlaty, M. M., Ndiaye-Lobry, D., … Aifantis, I. (2015). TET1 is a tumor suppressor of hematopoietic malignancy. *Nature Immunology*, **16**(6), 653–662.

Cimmino, L., Dolgalev, I., Wang, Y., … Aifantis, I. (2017). Restoration of TET2 Function Blocks Aberrant Self-Renewal and Leukemia Progression. *Cell*, **170**(6), 1079-1095.e20.

Clifford, S. C., Lusher, M. E., Lindsey, J. C., … Ellison, D. W. (2006). Wnt/Wingless pathway activation and chromosome 6 loss characterize a distinct molecular sub-group of medulloblastomas associated with a favorable prognosis. *Cell Cycle (Georgetown, Tex.)*, **5**(22), 2666–2670.

Coluccia, D., Figuereido, C., Isik, S., Smith, C., & Rutka, J. T. (2016). Medulloblastoma: Tumor biology and relevance to treatment and prognosis paradigm. *Current Neurology and Neuroscience Reports*, **16**(5). doi:10.1007/s11910-016-0644-7

Compere, S. J., & Palmiter, R. D. (1981). DNA methylation controls the inducibility of the mouse metallothionein-I gene in lymphoid cells. *Cell*, **25**(1), 233–240.

Cong, L., Ran, F. A., Cox, D., … Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, **339**(February), 819–824.

Costa, Y., Ding, J., Theunissen, T. W., … Wang, J. (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature*, **495**(7441), 370–374.

Craig Clark, W., Robertson, J. T., & Unnisa Ali, I. (1992). Amplification and/or Overexpression of Platelet-derived Growth Factor Receptors and Epidermal Growth Factor Receptor in Human Glial Tumors. *Cancer Research*, **52**(16), 4550–4553.

Crawford, J. R., MacDonald, T. J., & Packer, R. J. (2007). Medulloblastoma in childhood: new biological advances. *Lancet Neurology*, **6**(12), 1073–1085.

Czapiewski, P., Gorczynski, A., Radecka, K., … Biernat, W. (2016). Expression of SOX11, PAX5, TTF-1 and ISL-1 in medulloblastoma. *Pathology Research and Practice*, **212**(11), 965–971.

Dang, L., Jin, S., & Su, S. M. (2010). IDH mutations in glioma and acute myeloid leukemia. *Trends in Molecular Medicine*, **16**(9), 387–397.

Dawlaty, M. M., Breiling, A., Le, T., Barrasa, M. I., & Gao, Q. (2014). Loss of Tet Enzymes Compromises Proper Differentiation of Embryonic Stem Cells. *Developmental Cell*, **20**, 102–111.

Dawlaty, M. M., Breiling, A., Le, T., … Jaenisch, R. (2013). Combined Deficiency of Tet1 and Tet2 Causes Epigenetic Abnormalities but Is Compatible with Postnatal Development. *Developmental Cell*, **3**, 310–323.

Dawlaty, M. M., Ganz, K., Powell, B. E., … Jaenisch, R. (2011). Tet1 Is Dispensable for Maintaining Pluripotency and Its Loss Is Compatible with Embryonic and Postnatal Development. *Stem Cell*, **9**(2), 166–175.

de Bont, J. M., Packer, R. J., Michiels, E. M., Boer, M. L. den, & Pieters, R. (2008). Biological background of pediatric medulloblastoma and ependymoma: A review from a translational research perspective. *Neuro-Oncology*, **10**(6), 1040–1060.

de Conti, A., Kobets, T., Escudero-Lourdes, C., … Pogribny, I. P. (2014). Dose-and time-dependent epigenetic changes in the livers of fisher 344 rats exposed to furan. *Toxicological Sciences*, **139**(2), 371–380.

De Souza-Pinto, N. C., Eide, L., Hogue, B. A., … Bohr, V. A. (2001). Repair of 8-oxodeoxyguanosine lesions in mitochondrial DNA depends on the oxoguanine DNA glycosylase (OGG1) gene and 8-oxoguanine accumulates in the mitochondrial DNA of OGG1-defective mice. *Cancer Research*, **61**(14), 5378–5381.

Delhommeau, F., Dupont, S., Valle, V. Della, … Bernard, O. A. (2009). Mutation in TET2 in Myeloid Cancers. *The NEW ENGLAND JOURNAL of MEDICINE*, **360**, 2289–2301.

Deplus, R., Delatte, B., Schwinn, M. K., … Fuks, F. (2013). TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO Journal*, **32**(5), 645–655.

Dhliwayo, N., Sarras, M. P., Luczkowski, E., Mason, S. M., & Intine, R. V. (2014). Parp inhibition prevents ten-eleven translocase enzyme activation and hyperglycemia-induced DNA demethylation. *Diabetes*, **63**(9), 3069–3076.

Dickson, G. J., Lappin, T. R., & Thompson, A. (2009). Complete Array of HOX Gene Expression by RQ-PCR. *Leukemia, Methods in Molecular*

Biology, **538**(1), 369–393.

Du, Q., Luu, P. L., Stirzaker, C., & Clark, S. J. (2015). Methyl-CpG-binding domain proteins: Readers of the epigenome. *Epigenomics*, **7**(6), 1051–1073.

Ehrlich, M. (2002). DNA methylation in cancer: too much, but also too little. *Oncogene*, **21**(35), 5400–5413.

Eleftheriou, M., Pascual, A. J., Wheldon, L. M., … Ruzov, A. (2015). 5-Carboxylcytosine levels are elevated in human breast cancers and gliomas. *Clinical Epigenetics*, 1–6.

Elliott, E. N., Sheaffer, K. L., & Kaestner, K. H. (2016). The 'de novo' DNA methyltransferase Dnmt3b compensates the Dnmt1-deficient intestinal epithelium. *ELife*, **5**, 1–15.

Farzadfard, F., Perli, S. D., & Lu, T. K. (2013). Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS Synthetic Biology*, **2**(10), 604–613.

Feinberg, A. P., Koldobskiy, M. A., & Göndör, A. (2016). Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nature Reviews Genetics*, **17**(5), 284–299.

Ficz, G., Branco, M. R., Seisenberger, S., … Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*, **473**(7347), 398–404.

Ficz, G., & Gribben, J. G. (2014). Loss of 5-hydroxymethylcytosine in cancer: Cause or consequence? *Genomics*, **104**(5), 352–357.

Foksinski, M., Zarakowska, E., Gackowski, D., … Olinski, R. (2017). Profiles of a broad spectrum of epigenetic DNA modifications in normal and malignant human cell lines: Proliferation rate is not the major factor responsible for the 5-hydroxymethyl-2'-deoxycytidine level in cultured cancerous cell lines. *PLoS ONE*, **12**(11), 1–13.

Frattini, A., Fabbri, M., Valli, R., … Maserati, E. (2015). High variability of genomic instability and gene expression profiling in different HeLa clones. *Scientific Reports*, **5**, 1–9.

Freudenberg, J. M., Ghosh, S., Lackford, B. L., … Jothi, R. (2012). Acute depletion of Tet1-dependent 5-hydroxymethylcytosine levels impairs LIF/Stat3 signaling and results in loss of embryonic stem cell identity. *Nucleic Acids Research*, **40**(8), 3364–3377.

Friedmann-Morvinski, D., Bushong, E. A., Ke, E., … Verma, I. M. (2012).

Dedifferentiation of Neurons and Astrocytes by Oncogenes Can Induce Gliomas in Mice. *Science*, **338**(November), 1080–1085.

Frommer, M., McDonald, L. E., Millar, D. S., … Paul, C. L. (1992). A genomic sequencing protocol that yields a positive display of 5- methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America*, **89**(5), 1827–1831.

Fu, L., Guerrero, C. R., Zhong, N., … Wang, Y. (2014). Tet-Mediated Formation of 5 - Hydroxymethylcytosine in RNA, 1–4.

Fu, Y., Foden, J. A., Khayter, C., … Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, **31**(9). doi:10.1038/nbt.2623

Fuller, G. N., Su, X., Price, R. E., … Majumder, S. (2005). Many human medulloblastoma tumors overexpress repressor element-1 silencing transcription (REST)/neuron-restrictive silencer factor, which can be functionally countered by REST-VP16. *Molecular Cancer Therapeutics*, **4**(3), 343–9.

Gackowski, D., Starczak, M., Zarakowska, E., … Olinski, R. (2016). Accurate, direct, and high-throughput analyses of a broad spectrum of endogenously generated DNA base modifications with isotope-dilution two-dimensional ultraperformance liquid chromatography with tandem mass spectrometry: Possible clinical implication. *Analytical Chemistry*, **88**(24), 12128–12136.

Gaj, T., Guo, J., Kato, Y., Sirk, S. J., & Barbas, C. F. (2012). Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nature Methods*, **9**(8), 805–807.

Gajjar, A., Bowers, D. C., Karajannis, M. A., Leary, S., Witt, H., & Gottardo, N. G. (2015). Pediatric brain tumors: Innovative genomic information is transforming the diagnostic and clinical landscape. *Journal of Clinical Oncology*, **33**(27), 2986–2998.

Gajjar, A., Chintagumpala, M., Ashley, D., … Gilbertson, R. J. (2006). Risk-adapted craniospinal radiotherapy followed by high-dose chemotherapy and stem-cell rescue in children with newly diagnosed medulloblastoma (St Jude Medulloblastoma-96): long-term results from a prospective, multicentre trial. *Lancet Oncology*, **7**(10), 813–820.

Gao, Y., Chen, J., Li, K., … Gao, S. (2013). Replacement of Oct4 by Tet1

during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell*, **12**(4), 453–469.

García, M. G., Carella, A., Urdinguio, R. G., … Fraga, M. F. (2018). Epigenetic dysregulation of TET2 in human glioblastoma. *Oncotarget*, **9**(40), 25922–25934.

Garros-Regulez, L., Aldaz, P., Arrizabalaga, O., … Matheu, A. (2016). MTOR inhibition decreases SOX2-SOX9 mediated glioma stem cell activity and temozolomide resistance. *Expert Opinion on Therapeutic Targets*, **20**(4), 393–405.

Gaudet, F. (2003). Induction of Tumors in Mice by Genomic Hypomethylation. *Science*, **300**(5618), 489–492.

Genereux, D. P., Johnson, W. C., Burden, A. F., Stöger, R., & Laird, C. D. (2008). Errors in the bisulfite conversion of DNA: Modulating inappropriate- and failed-conversion frequencies. *Nucleic Acids Research*, **36**(22), 1–19.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, **6**(5), 343–345.

Gibson, P., Tong, Y., Robinson, G., … Gilbertson, R. J. (2010). Subtypes of medulloblastoma have distinct developmental origins. *Nature*, **468**(7327), 1095–1099.

Gilbert, L. A., Larson, M. H., Morsut, L., … Qi, L. S. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, **154**(2), 442.

Globisch, D., Münzel, M., Müller, M., … Bru, T. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE*, **5**(12), 1–9.

Goldberg, A. D., Allis, C. D., & Bernstein, E. (2007). Epigenetics: A Landscape Takes Shape. *Cell*, **128**(4), 635–638.

Goll, M. G., & Bestor, T. H. (2005). Eukaryotic Cytosine Methyltransferases. *Annual Review of Biochemistry*, **74**(1), 481–514.

Grunau, C. (2001). Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Research*, **29**(13), 65e – 65.

Gu, T.-P., Guo, F., Yang, H., … Xu, G.-L. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*,

477(7366), 606–610.

Guibert, S., Forne, T., Weber, M., Forné, T., & Weber, M. (2012). Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Research*, **22**(4), 633–641.

Guo, F., Li, X., Liang, D., … Xu, G. L. (2014). Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell*, **15**(4), 447–459.

Guo, M., Li, X., Zhang, L., … Tang, D. (2017). Accurate quantification of 5-Methylcytosine, 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine in genomic DNA from breast cancer by chemical derivatization coupled with ultra performance liquid chromatography-electrospray quadrupole time . *Oncotarget*, **8**(53), 91248–91257.

Hackenberg, M., Barturen, G., & L., J. (2012). DNA Methylation Profiling from High-Throughput Sequencing Data. *DNA Methylation - From Genomics to Technology*. doi:10.5772/34825

Hahn, M. A., Qiu, R., Wu, X., … Lu, Q. (2013). Dynamics of 5-Hydroxymethylcytosine and Chromatin Marks in Mammalian Neurogenesis. *Cell Reports*, **3**(2), 291–300.

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell*, **144**(5), 646–674.

Hardee, M. E., Marciscano, A. E., Medina-Ramirez, C. M., … Barcellos-Hoff, M. H. (2012). Resistance of glioblastoma initiating cells to radiation mediated by the tumor microenvironment can be abolished by inhibiting transforming growth factor-β (TGFβ). *Cancer Research*, **72**(16), 4119–4129.

Hargreaves, D. C., & Crabtree, G. R. (2011). ATP-dependent chromatin remodeling: Genetics, genomics and mechanisms. *Cell Research*, **21**(3), 396–420.

Hashimoto, H., Hong, S., Bhagwat, A. S., Zhang, X., & Cheng, X. (2012a). Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: Its structural basis and implications for active DNA demethylation. *Nucleic Acids Research*, **40**(20), 10203–10214.

Hashimoto, H., Liu, Y., Upadhyay, A. K., … Cheng, X. (2012b). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Research*, **40**(11), 4841–4849.

Hashimoto, H., Olanrewaju, Y. O., Zheng, Y., Wilson, G. G., & Zhang, X. (2014a). Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. *Genes & Development*, **4**, 2304–2313.

Hashimoto, H., Pais, J. E., Zhang, X., … Cheng, X. (2014b). Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature*, **506**(7488), 391–395.

Hashimoto, H., Vertino, P. M., & Cheng, X. (2010). Molecular coupling of DNA methylation and histone methylation. *Epigenomics*, **2**(5), 657–669.

Hata, K., Okano, M., Lei, H., & Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development*, **129**(8), 1983–1993.

Hatten, M. F. R. and M. E. (2011). Cerebellum: Development and Medulloblastoma.

He, Y.-F., Li, B.-Z., Li, Z., … Xu, G.-L. (2011). TET-Mediated Formation of 5-Carboxylcytosine and its excision by TDG in Mammalian DNA. *Science*, **333**(6047), 1303–1307.

Hermann, A., Goyal, R., & Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *Journal of Biological Chemistry*, **279**(46), 48350–48359.

Holkers, M., Maggio, I., Liu, J., … Gonçalves, M. A. F. V. (2013). Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Research*, **41**(5). doi:10.1093/nar/gks1446

Holliday, R., & Pugh, J. (1975). DNA modification mechanisms and gene activity during development. *Science*, **187**(4173), 226–232.

Hong, X., Chedid, K., & Kalkanis, S. N. (2012). Glioblastoma cell line-derived spheres in serum-containing medium versus serum-free medium: A comparison of cancer stem cell properties. *International Journal of Oncology*, **41**(5), 1693–1700.

Hotchkiss, D. (1948). THE QUANTITATIVE SEPARATION OF PURINES , AND NUCLEOSIDES BY PAPER CHROMATOGRAPHY. *Journal of Biological Chemistry*, (i).

Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, **157**(6), 1262–1278.

Hsu, P. D., Scott, D. A., Weinstein, J. A., … Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, **31**(9). doi:10.1038/nbt.2647

Hu, L., Li, Z., Cheng, J., … Xu, Y. (2013a). Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell*, **155**(7), 1545–1555.

Hu, L., Lu, J., Cheng, J., … Xu, Y. (2015). Structural insight into substrate preference for TET-mediated oxidation. *Nature*, **527**(7576), 118–122.

Hu, S., Wan, J., Su, Y., … Zhu, H. (2013b). DNA methylation presents distinct binding sites for human transcription factors. *ELife*, **2013**(2), 1–16.

Hu, X., Zhang, L., Mao, S. Q., … Xu, G. L. (2014). Tet and TDG mediate DNA demethylation essential for mesenchymal-to- epithelial transition in somatic cell reprogramming. *Cell Stem Cell*, **14**(4), 512–522.

Huang, G. H., Xu, Q. F., Cui, Y. H., Li, N., Bian, X. W., & Lv, S. Q. (2016a). Medulloblastoma stem cells: Promising targets in medulloblastoma therapy. *Cancer Science*, **107**(5), 583–589.

Huang, W., Lan, M.-D., Qi, C.-B., … Feng, Y.-Q. (2016b). Formation and determination of the oxidation products of 5-methylcytosine in RNA. *Chemical Science*, **7**, 5495 – 5502.

Huang, Y., Chavez, L., Chang, X., … Rao, A. (2014). Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, **111**(4), 1361–1366.

Huang, Z., & Bao, S. (2012). Ubiquitination and deubiquitination of REST and its roles in cancers. *FEBS Letters*, **586**(11), 1602–1605.

Ignatova, T. N., Kukekov, V. G., Laywell, E. D., Suslov, O. N., Vrionis, F. D., & Steindler, D. A. (2002). Human cortical glial tumors contain neural stem-like cells expressing astroglial and neuronal markers in vitro. *Glia*, **39**(3), 193–206.

Ilié, M., & Hofman, P. (2016). Pros: Can tissue biopsy be replaced by liquid biopsy? *Translational Lung Cancer Research*, **5**(4), 420–423.

Inda, M. del M., Bonavia, R., & Seoane, J. (2014). Glioblastoma multiforme:A look inside its heterogeneous nature. *Cancers*, **6**(1), 226–239.

Inoue, A., Shen, L., Dai, Q., He, C., & Zhang, Y. (2011). Generation and

replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Research*, **21**(12), 1670–1676.

Iqbal, K., Jin, S.-G., Pfeifer, G. P., & Szabo, P. E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proceedings of the National Academy of Sciences*, **108**(9), 3642–3647.

Ito, S., Dalessio, A. C., Taranova, O. V., … Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, **466**(7310), 1129–1133.

Ito, S., Shen, L., Dai, Q., … Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science*, **333**(6047), 1300–1303.

Iurlaro, M., Ficz, G., Oxley, D., … Reik, W. (2013). A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biology*, **14**(10), 1–11.

Iurlaro, M., Mcinroy, G. R., Burgess, H. E., … Reik, W. (2016). In vivo genome-wide profiling reveals a tissue-specific role for 5-formylcytosine. *Genome Biology*, 1–9.

Iyer, L. M., Tahiliani, M., Rao, A., & Aravind, L. (2009). Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle*, **8**(11), 1698–1710.

Jackson-Grusby, L., Beard, C., Possemato, R., … Jaenisch, R. (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nature Genetics*, **27**(1), 31–39.

Jahr, S., Hentze, H., Englisch, S., … Knippers, R. (2001). DNA fragments in the blood plasma of cancer patients: Quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Research*, **61**(4), 1659–1665.

Jiang, H., Lei, R., Ding, S. W., & Zhu, S. (2014). Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*, **15**(1), 1–12.

Jin, S. G., Jiang, Y., Qiu, R., … Pfeifer, G. P. (2011). 5-hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Research*, **71**(24), 7360–7365.

Jin, S. G., Kadam, S., & Pfeifer, G. P. (2010). Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Research*, **38**(11), 1–7.

Jin, S. G., Zhang, Z. M., Dunwell, T. L., … Pfeifer, G. P. (2016). Tet3 Reads 5-Carboxylcytosine through Its CXXC Domain and Is a Potential Guardian against Neurodegeneration. *Cell Reports*, **14**(3), 493–505.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, **337**(August), 816–822.

Johnson, K. C., Houseman, E. A., King, J. E., Von Herrmann, K. M., Fadul, C. E., & Christensen, B. C. (2016). 5-Hydroxymethylcytosine localizes to enhancer elements and is associated with survival in glioblastoma patients. *Nature Communications*, **7**, 1–11.

Johnson, R. A., Wright, K. D., Poppleton, H., … Gilbertson, R. J. (2010). Cross-species genomics matches driver mutations and cell compartments to model ependymoma. *Nature*, **466**(7306), 632–636.

Johnson, T. B., & Coghill, R. D. (1925). Researches on pyrimidines. C111. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus. *Journal of the American Chemical Society*, **47**(11), 2838–2844.

Jones, D. T. W., Jäger, N., Kool, M., … Lichter, P. (2012). Dissecting the genomic complexity underlying medulloblastoma. *Nature*, **488**(7409), 100–105.

Jones, P. A. (2012). Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, **13**(7), 484–492.

Jones, P. a, & Baylin, S. B. (2002). The fundamental role of epigenetic events in cancer. *Nature Reviews. Genetics*, **3**(6), 415–28.

Kadam, S., Mcalpine, G. S., Phelan, M. L., Kingston, R. E., Jones, K. A., & Emerson, B. M. (2000). Functional selectivity of recombinant mammalian SWI / SNF subunits. *In Vitro*, 2441–2451.

Kaelin, W. G., & McKnight, S. L. (2013). Influence of metabolism on epigenetics and disease. *Cell*, **153**(1), 56–69.

Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M., & Saitou, M. (2013). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *EMBO Journal*, **32**(3), 340–353.

Kaneda, M., Okano, M., Hata, K., … Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature*, **429**(6994), 900–903.

Kang, K. A., Piao, M. J., Kim, K. C., … Hyun, J. W. (2014). Epigenetic modification of Nrf2 in 5-fluorouracil-resistant colon cancer cells: Involvement of TET-dependent DNA demethylation. *Cell Death and Disease*, **5**(4), 1–13.

Karajannis, M., Allen, J. C., & Newcomb, E. W. (2009). NIH Public Access, **217**(3), 584–589.

Karra, D., Dahm, R., Hahn, S., … Zeitelhofer-, M. (2010). Transfection Techniques for Neuronal Cells. *The Journal of Neuroscience*, **30**(18), 6171–6177.

Kawauchi, D., Robinson, G., Uziel, T., … Roussel, M. F. (2012). A Mouse Model of the Most Aggressive Subgroup of Human Medulloblastoma. *Cancer Cell*, **21**(2), 168–180.

Kellinger, M. W., Song, C., Chong, J., Lu, X., He, C., & Wang, D. (2012). 5-formylcytosine and 5-carboxylcytosine reduce rate and substrate specificity of RNA polymerase II transcription. *Nature Structural & Molecular Biology*, **19**(8), 831–834.

Khan, S. H. (2019). Genome-Editing Technologies : Concept , Pros , and Cons of Various Genome-Editing Techniques and Bioethical Concerns for Clinical Application. *Nucleic Acids Research*, **16**(June), 326–334.

Khrapunov, S., Warren, C., Cheng, H., Berko, E. R., Greally, J. M., & Brenowitz, M. (2014). Unusual characteristics of the DNA binding domain of epigenetic regulatory protein MeCP2 determine its binding specificity. *Biochemistry*, **53**(21), 3379–3391.

Kilday, J.-P., Rahman, R., Dyer, S., … Grundy, R. (2009). Pediatric Ependymoma: Biological Perspectives. *Molecular Cancer Research*, **7**(6), 765–786.

Killian, J. K., Kim, S. Y., Miettinen, M., … Meltzer, P. S. (2013). Succinate dehydrogenase mutation underlies global epigenomic divergence in gastrointestinal stromal tumor. *Cancer Discovery*, **3**(6), 648–657.

Kim, H., & Kim, J. (2014). A guide to genome engineering with programmable nucleases. *Nature Publishing Group*, **15**(May). doi:10.1038/nrg3686

Kim, J. K., Samaranayake, M., & Pradhan, S. (2009). Epigenetic

mechanisms in mammals. *Cellular and Molecular Life Sciences*, **66**(4), 596–612.

Kim, S., Kim, D., Cho, S. W., Kim, J., & Kim, J. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins, 1012–1019.

Kim, Y. G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, **93**(3), 1156–1160.

Kim, Y. H., Pierscianek, D., Mittelbronn, M., … Ohgaki, H. (2011). TET2 promoter methylation in low-grade diffuse gliomas lacking IDH1/2 mutations. *Journal of Clinical Pathology*, **64**(10), 850–852.

Kim, Y., Kim, E., Wu, Q., … Rich, J. N. (2012). Platelet-derived growth factor receptors differentially inform intertumoral and intratumoral heterogeneity. *Genes and Development*, **26**(11), 1247–1262.

King, A. A., Seidel, K., Di, C., … Packer, R. J. (2017). Long-term neurologic health and psychosocial function of adult survivors of childhood medulloblastoma/PNET: A report from the Childhood Cancer Survivor Study. *Neuro-Oncology*, **19**(5), 689–698.

Kishikawa, S., Murata, T., Ugai, H., Yamazaki, T., & Yokoyama, K. K. (2003). Control elements of Dnmt1 gene are regulated in cell-cycle dependent manner. *Nucleic Acids Research. Supplement (2001)*, (3), 307–308.

Klug, A. (2010). The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation. *Annual Review of Biochemistry*, **79**(1), 213–231.

Ko, M., An, J., Bandukwala, H. S., … Rao, A. (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature*, **497**(7447), 122–126.

Ko, M., An, J., Pastor, W. A., Koralov, S. B., Rajewsky, K., & Rao, A. (2015). TET proteins and 5-methylcytosine oxidation in hematological cancers. *Immunological Reviews*, **263**(1), 6–21.

Ko, M., Bandukwala, H. S., An, J., … Rao, A. (2011). Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *PNAS*, **108**(35), 14566–14571.

Ko, M., Huang, Y., Jankowska, A. M., … Rao, A. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*, **468**(7325), 839–843.

Koh, K. P., Yabuuchi, A., Rao, S., … Daley, G. Q. (2012). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell*, **8**(2), 200–213.

Kohli, R. M., & Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*, **502**(7472), 472–479.

Konstandin, N., Bultmann, S., Szwagierczak, A., … Bohlander, S. K. (2011). Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. *Leukemia*, **25**(10), 1649–1652.

Kool, M., Jones, D. T. W., Jäger, N., … Pfister, S. M. (2014). Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothened inhibition. *Cancer Cell*, **25**(3), 393–405.

Kool, M., Korshunov, A., Remke, M., … Pfister, S. M. (2012). Molecular subgroups of medulloblastoma: An international meta-analysis of transcriptome, genetic aberrations, and clinical data of WNT, SHH, Group 3, and Group 4 medulloblastomas. *Acta Neuropathologica*, **123**(4), 473–484.

Kozmik, Z., Sure, U., Rüedi, D., Busslinger, M., & Aguzzi, A. (1995). Deregulated expression of PAX5 in medulloblastoma. *Proceedings of the National Academy of Sciences of the United States of America*, **92**(12), 5709–5713.

Kraus, T. F. J., Globisch, D., Wagner, M., … Kretzschmar, H. A. (2012). Low values of 5-hydroxymethylcytosine (5hmC), the "sixth base," are associated with anaplasia in human brain tumors. *International Journal of Cancer*, **131**(7), 1577–1590.

Kraus, T. F. J. J., Greiner, A., Steinmaurer, M., Dietinger, V., Guibourt, V., & Kretzschmar, H. A. (2015). Genetic characterization of ten-eleven-translocation methylcytosine dioxygenase alterations in human glioma. *Journal of Cancer*, **6**(9), 832–842.

Kriaucionis, S., & Heintz, N. (2009). The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science*, **324**(May), 929–931.

Kumar, S., Chinnusamy, V., & Mohapatra, T. (2018). Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Frontiers in Genetics*, **9**(December), 1–14.

Kunkele, A., Preter, K. De, Heukamp, L., … Germany, A. K. (2012). Pharmacological activation of the p53 pathway by nutlin-3 exerts anti-tumoral effects in medulloblastomas. *Neuro*, **14**(7), 859–869.

Kuo, K., A-McCune, R., & Charles, W. G. (1980). Quantitative reversed-phase high performance liquid chromatographic determination of major and modified deoxvribonudeosides in DNA. *Nucleic Acids Research*, **8**(20).

Kurdyukov, S., & Bullock, M. (2016). DNA methylation analysis: Choosing the right method. *Biology*, **5**(1), 1–21.

Landry, J. J. M., Pyl, P. T., Rausch, T., … Steinmetz, L. M. (2013). The genomic and transcriptomic landscape of a hela cell line. *G3: Genes, Genomes, Genetics*, **3**(8), 1213–1224.

Lathia, J. D., Mack, S. C., Mulkearns-Hubert, E. E., Valentim, C. L. L., & Rich, J. N. (2015). Cancer stem cells in glioblastoma. *Genes and Development*, **29**, 1203–1217.

Laurent, L., Wong, E., Li, G., … Wei, C. L. (2010). Dynamic changes in the human methylome during differentiation. *Genome Research*, **20**(3), 320–331.

Lavanya, C., Sibin, M. K., Srinivas Bharath, M. M., … Chetan, G. K. (2016). RNA interference mediated downregulation of human telomerase reverse transcriptase (hTERT) in LN18 cells. *Cytotechnology*, **68**(6), 2311–2321.

Lawinger, P., Venugopal, R., Guo, Z. S., … Majumder, S. (2000). The neuronal repressor REST/NRSF is an essential regulator in medulloblastoma cells. *Nature Medicine*, **6**(7), 826–831.

Lee, H. J., Hore, T. A., & Reik, W. (2014). Reprogramming the methylome: Erasing memory and creating diversity. *Cell Stem Cell*, **14**(6), 710–719.

Lehmann-Werman, R., Neiman, D., Zemmour, H., … Dor, Y. (2016). Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **113**(13), E1826–E1834.

Lentini, A., Lagerwall, C., Vikingsson, S., … Nestor, C. E. (2018). A reassessment of DNA-immunoprecipitation-based genomic profiling.

Nature Methods, **15**(7), 499–504.

Leonhardt, H., Page, A. W., Weier, H. U., & Bestor, T. H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell*, **71**(5), 865–873.

Lessard, J., Wu, J. I., Ranish, J. A., … Crabtree, G. R. (2007). An Essential Switch in Subunit Composition of a Chromatin Remodeling Complex during Neural Development. *Neuron*, **55**(2), 201–215.

Leto, K., Arancillo, M., Becker, E. B. E., … Hawkes, R. (2016). Consensus Paper: Cerebellar Development. *Cerebellum*, **15**(6), 789–828.

Lewis, L. C., Cho, P., Lo, K., … Aithal, G. P. (2017). Dynamics of 5-carboxylcytosine during hepatic differentiation : Potential general role for active demethylation by DNA repair in lineage specification. *Epigenetics*, **12**(4), 277–286.

Li, E., Bestor, T. H., & Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**(6), 915–926.

Li, E., & Zhang, Y. (2014). DNA methylation in mammals. *Cold Spring Harbor Perspectives in Biology*, **6**(5). doi:10.1101/cshperspect.a019133

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**(14), 1754–1760.

Li, H., Handsaker, B., Wysoker, A., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**(16), 2078–2079.

Li, P., Lee, E. H., Du, F., … Yang, Z. J. (2016a). Nestin mediates hedgehog pathway tumorigenesis. *Cancer Research*, **76**(18), 5573–5583.

Li, T., Yang, D., Li, J., Tang, Y., Yang, J., & Le, W. (2014). Critical Role of Tet3 in Neural Progenitor Cell Maintenance and Terminal Differentiation. *Molecular Neurobiology*, **51**(1), 142–154.

Li, W., Li, X., Xu, S., Ma, X., & Zhang, Q. (2016b). Expression of Tim4 in glioma and its regulatory role in LN-18 glioma cells. *Medical Science Monitor*, **22**, 77–82.

Li, W., Zhang, X., Lu, X., … Liu, J. (2017). 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers. *Cell Research*, **27**(10), 1243–1257.

Li, Z., Cai, X., Cai, C. L., … Xu, M. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of

myeloid malignancies. *The American Society of Hematology*, **118**(17), 4509–4518.

Lian, C. G., Xu, Y., Ceol, C., … Shi, Y. G. (2012). Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of Melanoma. *Cell*, **150**(6), 1135–1146.

Liang, X., Potter, J., Kumar, S., … Chesnut, J. D. (2015). Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *Journal of Biotechnology*, **208**, 44–53.

Liao, J., Karnik, R., Gu, H., … Meissner, A. (2015). Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nature Genetics*, **47**(5), 469–478.

Liao, W. C., Liao, C. K., Tsai, Y. H., … Liu, C. H. (2018). DSE promotes aggressive glioma cell phenotypes by enhancing HB-EGF/ERBB signaling. *PLoS ONE*, **13**(6), 1–16.

Lines, G. C., Ishii, N., Maier, D., … Meir, E. G. Van. (1999). p14 ARF , PTEN Tumor Suppressor Genes in Human. *Brain Pathology*, **479**, 469–479.

Lister, R., Mukamel, E. A., Nery, J. R., … Ecker, J. R. (2013). Global epigenomic reconfiguration during mammalian brain development. *Science*, **341**(6146). doi:10.1126/science.1237905

Lister, R., Pelizzola, M., Dowen, R. H., … Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**(7271), 315–322.

Liu, J., Jiang, J., Mo, J., … Wang, H. (2019). Global DNA 5-Hydroxymethylcytosine and 5-Formylcytosine Contents Are Decreased in the Early Stage of Hepatocellular Carcinoma. *Hepatology*, **69**(1), 196–208.

Liu, K.-W., Pajtler, K. W., Worst, B. C., Pfister, S. M., & Wechsler-Reya, R. J. (2017). Molecular mechanisms and therapeutic targets in pediatric brain tumors, **7593**(March).

Liu, S., Wang, J., Su, Y., … Wang, Y. (2013a). Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA. *Nucleic Acids Research*, **41**(13), 6421–6429.

Liu, X., Gao, Q., Li, P., … Wong, J. (2013b). UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nature Communications*, **4**. doi:10.1038/ncomms2562

Losman, J. A., Looper, R. E., Koivunen, P., … Kaelin, W. G. (2013). (R)-2-hydroxyglutarate is sufficient to promote leukemogenesis and its effects are reversible. *Science*, **340**(6127), 1621–1625.

Louis, D. N., Ohgaki, H., Wiestler, O. D., … Kleihues, P. (2007). The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathologica*, **114**(2), 97–109.

Louis, D. N., Perry, A., Reifenberger, G., … Ellison, D. W. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathologica*, **131**(6), 803–820.

Lu, X., Song, C. X., Szulwach, K., … He, C. (2013). Chemical modification-assisted bisulfite sequencing (CAB-seq) for 5-carboxylcytosine detection in DNA. *Journal of the American Chemical Society*, **135**(25), 9315–9317.

Luo, S., & Preuss, D. (2003). Strand-biased DNA methylation associated with centromeric regions in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(19), 11133–11138.

Ma, Y., Zhang, L., & Huang, X. (2014). Genome modification by CRISPR / Cas9. *The FEBS Journal*, **281**, 5186–5193.

Mabbott, D. J., Penkman, L., Witol, A., Strother, D., & Bouffet, E. (2008). Core neurocognitive functions in children treated for posterior fossa tumors. *Neuropsychology*, **22**(2), 159–168.

Machol, K., Rousseau, J., Ehresmann, S., … Campeau, P. M. (2019). Expanding the Spectrum of BAF-Related Disorders: De Novo Variants in SMARCC2 Cause a Syndrome with Intellectual Disability and Developmental Delay. *American Journal of Human Genetics*, **104**(1), 164–178.

Maeder, M. L., Linder, S. J., Cascio, V. M., Fu, Y., Ho, Q. H., & Joung, J. K. (2013). CRISPR RNA-guided activation of endogenous human genes. *Nature Methods*, **10**(10), 977–979.

Maiti, A., & Drohat, A. C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for active demethylation of CpG sites. *Journal of Biological Chemistry*, **286**(41), 35334–35338.

Mali, P., Aach, J., Stranges, P. B., … Church, G. M. (2013). CAS9

transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*, **31**(9), 833–838.

Marinov, G. K., Kundaje, A., Park, P. J., & Wold, B. J. (2014). Large-scale quality analysis of published ChIP-seq data. *G3: Genes, Genomes, Genetics*, **4**(2), 209–223.

Martin, A. M., Raabe, E., Eberhart, C., & Cohen, K. J. (2014). Management of Pediatric and Adult Patients with Medulloblastoma. *Current Treatment Options in Oncology*, **15**(4), 581–594.

Martinez-Lage, M., Lynch, T. M., Bi, Y., … Dahmane, N. (2019). Immune landscapes associated with different glioblastoma molecular subtypes. *Acta Neuropathologica Communications*, **7**(1), 1–12.

Martins-Taylor, K., Schroeder, D. I., Lasalle, J. M., Lalande, M., & Xu, R. H. (2012). Role of DNMT3B in the regulation of early neural and neural crest specifiers. *Epigenetics*, **7**(1), 71–82.

Marusyk, A., & Polyak, K. (2010). Tumor heterogeneity: causes and consequences Andriy. *Biochim Biophys Acta.*, **105**(1), 1–7.

Maupetit-Méhouas, S., Montibus, B., Nury, D., … Arnaud, P. (2016). Imprinting control regions (ICRs) are marked by mono-allelic bivalent chromatin when transcriptionally inactive. *Nucleic Acids Research*, **44**(2), 621–635.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., & Haaf, T. (2000). Demethylation of the zygotic paternal genome. *Nature*, **403**(6769), 501–502.

Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L., & Bird, A. P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell*, **58**(3), 499–507.

Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S., & Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell*, **151**(7), 1417–1430.

Messerschmidt, D. M., Knowles, B. B., & Solter, D. (2014). DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes and Development*, **28**(8), 812–828.

Migden, M. R., Guminski, A., Gutzmer, R., … Dummer, R. (2015). Treatment with two different doses of sonidegib in patients with locally advanced or metastatic basal cell carcinoma (BOLT): A multicentre,

randomised, double-blind phase 2 trial. *The Lancet Oncology*, **16**(6), 716–728.

Milde, T., Hielscher, T., Witt, H., … Korshunov, A. (2012). Nestin expression identifies ependymoma patients with poor outcome. *Brain Pathology*, **22**(6), 848–860.

Milde, T., Kleber, S., Korshunov, A., … Witt, O. (2011). A novel human high-risk ependymoma stem cell model reveals the differentiation-inducing potential of the histone deacetylase inhibitor Vorinostat. *Acta Neuropathologica*, **122**(5), 637–650.

Miller, J. C., Tan, S., Qiao, G., … Rebar, E. J. (2011). A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology*, **29**(2), 143–150.

Minor, E. A., Court, B. L., Young, J. I., & Wang, G. (2013). Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine. *Journal of Biological Chemistry*, **288**(19), 13669–13674.

Moarefi, A. H., & Chédin, F. (2011). ICF syndrome mutations cause a broad spectrum of biochemical defects in DNMT3B-mediated de novo DNA methylation. *Journal of Molecular Biology*, **409**(5), 758–772.

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, **155**(3), 733–740.

Monaghan, C. E., Nechiporuk, T., Jeng, S., … Mandel, G. (2017). REST corepressors RCOR1 and RCOR2 and the repressor INSM1 regulate the proliferation- differentiation balance in the developing brain. *Proceedings of the National Academy of Sciences of the United States of America*, **114**(3), E406–E415.

Moore, L. D., Le, T., & Fan, G. (2012). DNA Methylation and Its Basic Function. *Neuropsychopharmacology*, **38**(1), 23–38.

Moran-Crusio, K., Reavie, L., Shih, A., … Levine, R. L. (2011). Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation. *Cancer Cell*, **20**(1), 11–24.

Moscou, M. J., & Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**(5959), 1501.

Müller, T., Gessi, M., Waha, A., … Pietsch, T. (2012). Nuclear exclusion of TET1 is associated with loss of 5- hydroxymethylcytosine in IDH1 wild-

type gliomas. *American Journal of Pathology*, **181**(2), 675–683.

Narayanan, R., Pirouz, M., Kerimoglu, C., … Tuoc, T. (2015). Loss of BAF (mSWI/SNF) Complexes Causes Global Transcriptional and Chromatin State Changes in Forebrain Development. *Cell Reports*, **13**(9), 1842–1854.

Negrini, S., Prada, I., D'Alessandro, R., & Meldolesi, J. (2013). REST: An oncogene or a tumor suppressor? *Trends in Cell Biology*, **23**(6), 289–295.

Neri, F., Incarnato, D., Krepelova, A., … Oliviero, S. (2015). Single-Base resolution analysis of 5-formyl and 5-carboxyl cytosine reveals promoter DNA Methylation Dynamics. *Cell Reports*, **10**(5), 674–683.

Neri, F., Krepelova, A., Incarnato, D., … Oliviero, S. (2013). Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. *Cell*, **155**(1), 121.

Nestor, C. E., Ottaviano, R., Reddington, J., … Meehan, R. R. (2012). Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Research*, **22**(3), 467–477.

Nishimasu, H., Ran, F. A., Hsu, P. D., … Nureki, O. (2014). Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell*, **156**(5), 935–949.

Northcott, P. A., Buchhalter, I., Morrissy, A. S., … Lichter, P. (2017). The whole-genome landscape of medulloblastoma subtypes. *Nature*, **547**(7663), 311–317.

Northcott, P. A., Jones, D. T. W., Kool, M., … Pfister, S. M. (2012a). Medulloblastomics: the end of the beginning. *Nature Reviews Cancer*, **12**(12), 818–834.

Northcott, P. A., Korshunov, A., Witt, H., … Taylor, M. D. (2011). Medulloblastoma comprises four distinct molecular variants. *Journal of Clinical Oncology*, **29**(11), 1408–1414.

Northcott, P. A., Shih, D. J. H., Peacock, J., … Taylor, M. D. (2012b). Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature*, **487**(7409), 49–56.

Nowell, P. C. (1976). The clonal evolution of tumor cell populations. *Science*, **194**(4260), 23–28.

O. Abdel-Wahab, A. Mullally, C. H. et al. (2009). Genetic char- acterization of TET1, TET2, and TET3 alterations in myeloid malignancies,. *Blood,*

Vol. 114, No. 1, Pp. 144–147, **114**(1), 144–148.

Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**(3), 247–257.

Olinski, R., Starczak, M., & Gackowski, D. (2016). Enigmatic 5-hydroxymethyluracil: Oxidatively modified base, epigenetic mark or both? *Mutation Research - Reviews in Mutation Research*, **767**, 59–66.

Ono, R., Taki, T., Taketani, T., Taniwaki, M., Kobayashi, H., & Hayashi, Y. (2002). LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Research*, **62**(14), 4075–4080.

Ooi, S. K. T., Qiu, C., Bernstein, E., … Bestor, T. H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, **448**(7154), 714–717.

Pajtler, K. W., Witt, H., Sill, M., … Pfister, S. M. (2015). Molecular Classification of Ependymal Tumors across All CNS Compartments, Histopathological Grades, and Age Groups. *Cancer Cell*, **27**(5), 728–743.

Parker, M., Mohankumar, K. M., Punchihewa, C., … Vadodaria, B. (2014). C11orf95–RELA fusions drive oncogenic NF-kB signalling in ependymoma. *Nature*, **506**, 451–554.

Pastor, W. A., Aravind, L., & Rao, A. (2013). TETonic shift : biological roles of TET proteins in DNA demethylation and transcription. *Molecular Cell Biology*, **14**(June). doi:10.1038/nrm3589

Pastor, W. A., Pape, U. J., Huang, Y., … Rao, A. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature*, **473**(7347), 394–397.

Patel, A. P., Tirosh, I., Trombetta, J. J., … Suvà, M. L. (2014). NIH Public Access. *Science*, **344**(6190), 1396–1401.

Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., & Liu, D. R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology*, **31**(9). doi:10.1038/nbt.2673

Paul, C. L., & Clark, S. J. (1996). Cytosine methylation: Quantitation by automated genomic sequencing and GENESCAN(TM) analysis. *BioTechniques*, **21**(1), 126–133.

Pei, Y., Liu, K.-W., Wang, J., … Wechsler-Reya, R. J. (2016). HDAC and PI3K Antagonists Cooperate to Inhibit Growth of MYC-driven Medulloblastoma. *Cancer Cell*, **29**(3), 311–323.

Peng, D. F., Kanai, Y., Sawada, M., … Hirohashi, S. (2006). DNA methylation of multiple tumor-related genes in association with overexpression of DNA methyltransferase 1 (DNMT1) during multistage carcinogenesis of the pancreas. *Carcinogenesis*, **27**(6), 1160–1168.

Perez-Pinera, P., Ousterout, D. G., Brunger, J. M., … Gersbach, C. A. (2013). Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nature Methods*, **10**(3), 239–242.

Persson, A. I., Petritsch, C., Swartling, F. J., … Weiss, W. A. (2010). Non-Stem Cell Origin for Oligodendroglioma. *Cancer Cell*, **18**(6), 669–682.

Pfaffeneder, T., Hackner, B., Truß, M., … Carell, T. (2011). The discovery of 5-formylcytosine in embryonic stem cell DNA. *Bioorganic Chemistry*, **50**(31), 7008–7012.

Pfaffeneder, T., Spada, F., Wagner, M., … Carell, T. (2014). Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nature Chemical Biology*, **10**(7), 574–581.

Phillips, H. S., Kharbanda, S., Chen, R., … Aldape, K. (2006). Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell*, **9**(3), 157–173.

Piccolo, F. M., Bagci, H., Brown, K. E., … Fisher, A. G. (2013). Different Roles for Tet1 and Tet2 Proteins in Reprogramming-Mediated Erasure of Imprints Induced by EGC Fusion. *Molecular Cell*, **49**(6), 1023–1033.

Pollard, S. M., Yoshikawa, K., Clarke, I. D., … Dirks, P. (2009). Glioma Stem Cell Lines Expanded in Adherent Culture Have Tumor-Specific Phenotypes and Are Suitable for Chemical and Genetic Screens. *Cell Stem Cell*, **4**(6), 568–580.

Portela, A., & Esteller, M. (2010). Epigenetic modifications and human disease. *Nature Biotechnology*, **28**(10), 1057–1068.

Pradhan, S., Bacolla, A., Larson, J. E., Roberts, R. J., & Wells, R. D. (1999). Recombinant Human DNA (Cytosine-5) Methyltransferase. *Journal of Biological Chemistry*, **274**(12), 33002–33010.

Puig, I., Tenbaum, S. P., Chicote, I., … Palmer, H. G. (2018). TET2 controls chemoresistant slow-cycling cancer cell survival and tumor recurrence.

Journal of Clinical Investigation, **128**(9), 3887–3905.

Qi, L. S., Larson, M. H., Gilbert, L. A., … Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**(5), 1173–1183.

Qiao, Y., Wang, X., Wang, R., … Jing, N. (2015). AF9 promotes hESC neural differentiation through recruiting TET2 to neurodevelopmental gene loci for methylcytosine hydroxylation. *Cell Discovery*, **1**, 1–19.

Quivoron, C., Couronné, L., Della Valle, V., … Bernard, O. A. (2011). TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis. *Cancer Cell*, **20**(1), 25–38.

Radu, A., Neubauer, V., Akagi, T., Hanafusa, H., & Georgescu, M.-M. (2003). PTEN Induces Cell Cycle Arrest by Decreasing the Level and Nuclear Localization of Cyclin D1. *Molecular and Cellular Biology*, **23**(17), 6139–6149.

Raiber, E. A., Beraldi, D., Ficz, G., … Balasubramanian, S. (2012). Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biology*, **13**(8), R69.

Raimondi, A. J., & Tomita, T. (1983). Brain tumors during the first year of life. *Child's Brain*, **10**(3), 193–207.

Raizis, A. M., Schmitt, F., & Jost, J.-P. (1995). A Bisulfite Method of 5-Methylcytosine Mapping That Minimizes Template Degradation. *Analytical Biochemistry*, **226**, 161–166.

Rakyan, V. K., Hildmann, T., Novik, K. L., … Beck, S. (2004). DNA methylation profiling of the human major histocompatibility complex: A pilot study for the Human Epigenome Project. *PLoS Biology*, **2**(12). doi:10.1371/journal.pbio.0020405

Ramaswamy, V., Remke, M., Bouffet, E., … Pomeroy, S. L. (2016). Risk stratification of childhood medulloblastoma in the molecular era: the current consensus. *Acta Neuropathologica*, **131**(6), 821–831.

Rampal, R., Alkalin, A., Madzo, J., … Levine, R. L. (2014). DNA Hydroxymethylation Profiling Reveals that WT1 Mutations Result in Loss of TET2 Function in Acute Myeloid Leukemia. *Cell Reports*, **9**(5), 1841–1855.

Ramsahoye, B. H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A. P., &

Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences of the United States of America*, **97**(10), 5237–5242.

Ramsawhook, A., Lewis, L., Coyle, B., & Ruzov, A. (2017). Medulloblastoma and ependymoma cells display increased levels of 5- carboxylcytosine and elevated TET1 expression. *Clinical Epigenetics*, 1–9.

Ran, F. A., Hsu, P. D., Lin, C., & Gootenberg, J. S. (2013). Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *CELL*, **154**(6), 1380–1389.

Raybaud, C., Ramaswamy, V., Taylor, M. D., & Laughlin, S. (2015). Posterior fossa tumors in children: developmental anatomy and diagnostic imaging. *Child's Nervous System*, **31**(10), 1661–1676.

Rein, T., Zorbas, H., & DePamphilis, M. L. (1997). Active mammalian replication origins are associated with a high-density cluster of mCpG dinucleotides. *Molecular and Cellular Biology*, **17**(1), 416–426.

Renfree, M. B., Suzuki, S., & Kaneko-Ishino, T. (2013). The origin and evolution of genomic imprinting and viviparity in mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **368**(1609). doi:10.1098/rstb.2012.0151

Rhee, I., Bachman, K. E., Park, B. H., … Vogelstein, B. (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature*, **416**(6880), 552–556.

Rhee, I., Jalr, K. W., Yen, R. W. C., … Schuebel, K. E. (2000). CpG methylation is maintained in human cancer cells lacking DNMT1. *Nature*, **404**(6781), 1003–1007.

Richichi, C., Brescia, P., Alberizzi, V., Fornasari, L., & Pelicci, G. (2013). Marker-independent method for isolating slow-dividing cancer stem cells in human glioblastoma. *Neoplasia (United States)*, **15**(7), 840–847.

Robinson, G., Parker, M., Kranenburg, T. A., … Gilbertson, R. J. (2012). Novel mutations target distinct subgroups of medulloblastoma. *Nature*, **488**(7409), 43–48.

Romero, O. A., & Sanchez-Cespedes, M. (2014). The SWI/SNF genetic blockade: Effects in cell differentiation, cancer and developmental diseases. *Oncogene*, **33**(21), 2681–2689.

Roussel, M. F., & Stripay, J. L. (2018). Epigenetic Drivers in Pediatric Medulloblastoma. *Cerebellum*, **17**(1), 28–36.

Ruzov, A., Tsenkina, Y., Serio, A., … De Sousa, P. (2011). Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. *Cell Research*, **21**(9), 1332–1342.

Sabaawy, H. E. (2012). Genetic Heterogeneity and Clonal Evolution of Tumor Cells and their Impact on Precision Cancer Medicine. *J Leuk (Los Angel)*, **1**(4), 1–7.

Saenz-Antoñanzas, A., Auzmendi-Iriarte, J., Carrasco-Garcia, E., … Matheu, A. (2019). Liquid biopsy in glioblastoma: Opportunities, applications and challenges. *Cancers*, **11**(7), 1–20.

Saitou, M., Kagiwada, S., & Kurimoto, K. (2012). Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development*, **139**(1), 15–31.

Sander, J. D., & Joung, J. K. (2014). CRISPR-Cas systems for editing , regulating and targeting genomes. *Nature Biotechnology*, **32**(4), 347–355.

Santiago, M., Antunes, C., Guedes, M., Sousa, N., & Marques, C. J. (2014). TET enzymes and DNA hydroxymethylation in neural development and function - How critical are they? *Genomics*, **104**(5), 334–340.

Santos, F., & Dean, W. (2005). *Using immunofluorescence to observe methylation changes in mammalian preimplantation embryos. Nuclear reprogramming*. (S. Pells, Ed.)*Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki*, Vol. 325, Humana press. doi:10.1385/1597450057

Sarkar, A., & Hochedlinger, K. (2013). The Sox family of transcription factors: Versatile regulators of stem and progenitor cell fate. *Cancer Stem Cell*, **12**(1), 15–30.

Schoenherr, C. J., & Anderson, D. J. (1995). The neuron-restrictive silencer factor (NRSF): A coordinate repressor of multiple neuron-specific genes. *Science*, **267**(5202), 1360–1363.

Schomacher, L., Han, D., Musheev, M. U., … Niehrs, C. (2016). Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. *Nature Structural and Molecular Biology*, **23**(2), 116–124.

Schwartz, S., Meshorer, E., & Ast, G. (2009). Chromatin organization marks

exon-intron structure. *Nature Structural and Molecular Biology*, **16**(9), 990–995.

Seisenberger, S., Andrews, S., Krueger, F., … Reik, W. (2012). The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. *Molecular Cell*, **48**(6), 849–862.

Sekulic, A., Migde, M. R., Oro, A. E., … Hauschild, A. (2012). Efficacy and Safety of Vismodegib in Advanced Basal-Cell Carcinoma. *New England Journal of Medicine*, **366**(23), 2171–2179.

Sell, S. (2004). Stem cell origin of cancer and differentiation therapy. *Critical Reviews in Oncology/Hematology*, **51**(1), 1–28.

Sharif, J., Muto, M., Takebayashi, S. I., … Koseki, H. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, **450**(7171), 908–912.

Shen, B., Zhang, W., Zhang, J., … Skarnes, W. C. (2014a). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nature*, **11**(4). doi:10.1038/nmeth.2857

Shen, L., Inoue, A., He, J., Liu, Y., Lu, F., & Zhang, Y. (2014b). Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes. *Cell Stem Cell*, **15**(4), 459–471.

Shen, L., Wu, H., Diep, D., … Zhang, K. (2013). Genome-wide Analysis Reveals TET- and TDG-Dependent 5-Methylcytosine Oxidation Dynamics. *Cell*, **153**(3), 692–706.

Shibata, M., & Shen, M. M. (2013). The roots of cancer: Stem cells and the basis for tumor heterogeneity. *Bioassays*, **35**(3), 253–260.

Shih, A. H., & Holland, E. C. (2006). Platelet-derived growth factor (PDGF) and glial tumorigenesis. *Cancer Letters*, **232**(2), 139–147.

Shin, J. H., Xu, L., & Wang, D. (2016). RNA polymerase II acts as a selective sensor for DNA lesions and endogenous DNA modifications. *Transcription*, **7**(3), 57–62.

Singh, S. K., Clarke, I. D., Terasaki, M., … Dirks, P. B. (2003). Identification of a cancer stem cell in human brain tumors. *Cancer Research*, **63**(18), 5821–5828.

Singh, S. K., Hawkins, C., Clarke, I. D., … Dirks, P. B. (2004). Identification of human brain tumour initiating cells. *Nature*, **432**(7015), 396–401.

Singh, S. K., Kagalwala, M. N., Parker-Thornburg, J., Adams, H., & Majumder, S. (2008). REST maintains self-renewal and pluripotency of

embryonic stem cells. *Nature*, **453**(7192), 223–7.

Smallwood, S. A., & Kelsey, G. (2012). De novo DNA methylation: A germ cell perspective. *Trends in Genetics*, **28**(1), 33–42.

Smith, J. (2000). Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Research*, **28**(17), 3361–3369.

Smith, Z. D., Chan, M. M., Humm, K. C., … Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. *Nature*, **511**(7511), 611–615.

Smith, Z. D., Chan, M. M., Mikkelsen, T. S., … Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*, **484**(7394), 339–344.

Smith, Z. D., & Meissner, A. (2013). DNA methylation: Roles in mammalian development. *Nature Reviews Genetics*, **14**(3), 204–220.

Song, C. X., Szulwach, K. E., Dai, Q., … He, C. (2013a). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell*, **153**(3), 678–691.

Song, C. X., Yin, S., Ma, L., … Quake, S. R. (2017). 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Research*, **27**(10), 1231–1242.

Song, J., Pfeifer, G. P., & Rapids, G. (2016). Are there specific readers of oxidized 5-methylcytosine bases? *Bioessays*, **38**(10), 1038–1047.

Song, L., James, S. R., Kazim, L., & Karpf, A. R. (2005). Specific method for the determination of genomic DNA methylation by liquid chromatography-electrospray ionization tandem mass spectrometry. *Analytical Chemistry*, **77**(2), 504–510.

Song, S. J., Ito, K., Ala, U., … Pandolfi, P. P. (2013b). The oncogenic microRNA miR-22 targets the TET2 tumor suppressor to promote hematopoietic stem cell self-renewal and transformation. *Cell Stem Cell*, **13**(1), 87–101.

Song, S. J., Poliseno, L., Song, M. S., … Pandolfi, P. P. (2013c). MicroRNA-antagonism regulates breast cancer stemness and metastasis via TET-family-dependent chromatin remodeling. *Cell*, **154**(2), 311.

Sorek, R., Lawrence, C. M., & Wiedenheft, B. (2013). CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea. *Annual Review of Biochemistry*, **82**(1), 237–266.

Sowar, K., Straessle, J., Donson, A. M., Handler, M., & Foreman, N. K. (2006). Predicting which children are at risk for ependymoma relapse. *Journal of Neuro-Oncology*, **78**(1), 41–46.

Spans, L., Van den Broeck, T., Smeets, E., … Claessens, F. (2016). Genomic and epigenomic analysis of high-risk prostate cancer reveals changes in hydroxymethylation and TET1. *Oncotarget*, **7**(17), 24326–24338.

Spiegler, B. J., Bouffet, E., Greenberg, M. L., Rutka, J. T., & Mabbott, D. J. (2004). Change in neurocognitive functioning after treatment with cranial radiation in childhood. *Journal of Clinical Oncology*, **22**(4), 706–713.

Spruijt, C. G., Gnerlich, F., Smits, A. H., … Vermeulen, M. (2013). Dynamic readers for 5-(Hydroxy)methylcytosine and its oxidized derivatives. *Cell*, **152**(5), 1146–1159.

Spruijt, C. G., & Vermeulen, M. (2014). DNA methylation: Old dog, new tricks? *Nature Structural and Molecular Biology*, **21**(11), 949–954.

Stark, R and Brown, G. (2016). DiffBind: differential binding analysis of ChIP-Seq peak datatle, 1–29.

Storebjerg, T. M., Strand, S. H., Høyer, S., … Sørensen, K. D. (2018). Dysregulation and prognostic potential of 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) levels in prostate cancer. *Clinical Epigenetics*, **10**(1), 1–16.

Sturm, D., Witt, H., Hovestadt, V., … Pfister, S. M. (2012). Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma. *Cancer Cell*, **22**(4), 425–437.

Su, X., Gopalakrishnan, V., Stearns, D., … Majumder, S. (2006). Abnormal Expression of REST/NRSF and Myc in Neural Stem/Progenitor Cells Causes Cerebellar Tumors by Blocking Neuronal Differentiation. *Molecular and Cellular Biology*, **26**(5), 1666–1678.

Suvà, M. L., Rheinbay, E., Gillespie, S. M., … Bernstein, B. E. (2014). Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell*, **157**(3), 580–594.

Szopa, W., Burley, T. A., Kramer-marek, G., & Kaspera, W. (2017). Diagnostic and Therapeutic Biomarkers in Glioblastoma : Current Status and Future Perspectives, **2017**. doi:10.1155/2017/8013575

Szulwach, K. E., Li, X., Li, Y., … Jin, P. (2011a). 5-hmC-mediated epigenetic

dynamics during postnatal neurodevelopment and aging. *Nature Neuroscience*, **14**(12), 1607–1616.

Szulwach, K. E., Li, X., Li, Y., … Jin, P. (2011b). Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genetics*, **7**(6). doi:10.1371/journal.pgen.1002154

Szwagierczak, A., Bultmann, S., Schmidt, C. S., Spada, F., & Leonhardt, H. (2010). Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Research*, **38**(19). doi:10.1093/nar/gkq684

Tahiliani, M., Koh, K. P., Shen, Y., … Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**(5929), 930–935.

Takai, D., & Jones, P. A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proceedings of the National Academy of Sciences of the United States of America*, **99**(6), 3740–3745.

Takai, H., Masuda, K., Sato, T., … Akiyama, T. (2014). 5-Hydroxymethylcytosine plays a critical role in glioblastomagenesis by recruiting the CHTOP-Methylosome complex. *Cell Reports*, **9**(1), 48–60.

Takamiya, T., Hosobuchi, S., Asai, K., … Okuizumi, H. (2006). Restriction landmark genome scanning method using isoschizomers (MspI/HpaII) for DNA methylation analysis. *Electrophoresis*, **27**(14), 2846–2856.

Tan, L., Xiong, L., Xu, W., … Shi, Y. G. (2013). Genome-wide comparison of DNA hydroxymethylation in mouse embryonic stem cells and neural progenitor cells by a new comparative hMeDIP-seq method. *Nucleic Acids Research*, **41**(7), 1–12.

Tanaka, K., & Okamoto, A. (2007). Degradation of DNA by bisulfite treatment. *Bioorganic and Medicinal Chemistry Letters*, **17**(7), 1912–1915.

Tang, Y., Zheng, S. J., Qi, C. B., Feng, Y. Q., & Yuan, B. F. (2015). Sensitive and Simultaneous Determination of 5-Methylcytosine and Its Oxidation Products in Genomic DNA by Chemical Derivatization Coupled with Liquid Chromatography-Tandem Mass Spectrometry Analysis. *Analytical Chemistry*, **87**(6), 3445–3452.

Taylor, M. D., Northcott, P. A., Korshunov, A., … Pfister, S. M. (2012). Molecular subgroups of medulloblastoma: The current consensus. *Acta*

Neuropathologica, **123**(4), 465–472.

Thompson, M. C., Fuller, C., Hogg, T. L., … Gilbertson, R. J. (2006). Genomics identifies medulloblastoma subgroups that are enriched for specific genetic alterations. *Journal of Clinical Oncology*, **24**(12), 1924–1931.

Thomson, J. P., Hunter, J. M., Nestor, C. E., … Meehan, R. R. (2013). Comparative analysis of affinity-based 5-hydroxymethylation enrichment techniques. *Nucleic Acids Research*, **41**(22), 1–15.

Thomson, J. P., & Meehan, R. R. (2017). The application of genome-wide 5-hydroxymethylcytosine studies in cancer research. *Epigenomics*, **9**(1), 77–91.

Tong, M., Gao, S., Qi, W., … Che, Y. (2019). 5-Hydroxymethylcytosine As a Potential Epigenetic Biomarker in Papillary Thyroid Carcinoma. *Oncology Letters*, **18**(3), 2304–2309.

Torres, I. O., & Fujimori, D. G. (2015). Functional coupling between writers,erasers and readers of histone and DNA methylation. *Current Opin Struct Biol.*, **35**(1), 68–75.

Tost, J., & Gut, I. G. (2007). *Analysis of Gene-Specific DNA Methylation Patterns by Pyrosequencing® Technology*. *Methods Molecular Biology*. doi:10.1385/1597453773

Trevino, A. E., & Zhang, F. (2014). *Genome Editing Using Cas9 Nickases*. *Methods in Enzymology*, 1st edn, Vol. 546, Elsevier Inc. doi:10.1016/B978-0-12-801185-0.00008-8

Tsai, Y., Chen, H., Chen, S., Cheng, W., Wang, H., & Shen, Z. (2014). TET1 regulates hypoxia-induced epithelial-mesenchymal Transition By Acting As a Co-Activator. *Genome Biology*, 1–13.

Tuoc, T. C., Boretius, S., Sansom, S. N., … Stoykova, A. (2013). Chromatin Regulation by BAF170 Controls Cerebral Cortical Size and Thickness. *Developmental Cell*, **25**(3), 256–269.

Umer, M., & Herceg, Z. (2013). Deciphering the epigenetic code: An overview of DNA methylation analysis methods. *Antioxidants and Redox Signaling*, **18**(15), 1972–1986.

Vanner, R. J., Remke, M., Gallo, M., … Dirks, P. B. (2014). Quiescent Sox2+ Cells Drive Hierarchical Growth and Relapse in Sonic Hedgehog Subgroup Medulloblastoma. *Cancer Cell*, **26**(1), 33–47.

Vella, P., Scelfo, A., Jammula, S. G., … Pasini, D. (2013). Tet Proteins

Connect the O-Linked N-acetylglucosamine Transferase Ogt to Chromatin in Embryonic Stem Cells. *Molecular Cell*, **49**(4), 645–656.

Verhaak, R. G. W., Hoadley, K. A., Purdom, E., … Hayes, D. N. (2010). Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*, **17**(1), 98–110.

Waddington, C. H. (2012). The epigenotype. 1942. *International Journal of Epidemiology*, **41**(1), 10–13.

Wagner, H. (2001). Toll Meets Bacterial CpG-DNA Minireview. *Cell*, **14**, 499–502.

Wan, J. C. M., Massie, C., Garcia-Corbacho, J., … Rosenfeld, N. (2017). Liquid biopsies come of age: Towards implementation of circulating tumour DNA. *Nature Reviews Cancer*, **17**(4), 223–238.

Wang, H., Yang, H., Shivalila, C. S., … Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/cas-mediated genome engineering. *Cell*, **153**(4), 910–918.

Wang, J., Garancher, A., Ramaswamy, V., & Wechsler-Reya, R. J. (2018). Medulloblastoma: From Molecular Subgroups to Molecular Targeted Therapies. *Annual Review of Neuroscience*, **41**(1), 207–232.

Wang, L., Zhou, Y., Xu, L., … He, C. (2015a). Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature*, **523**(621). doi:10.1038/nature14482

Wang, Q., Hu, B., Hu, X., … Verhaak, R. G. W. (2017). Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell*, **32**(1), 42-56.e6.

Wang, Y., Wu, Q., Yang, P., … Wang, X. (2016). LSD1 co-repressor Rcor2 orchestrates neurogenesis in the developing mouse brain. *Nature Communications*, **7**. doi:10.1038/ncomms10481

Wang, Y., Xiao, M., Chen, X., … Ma, S. (2015b). WT1 Recruits TET2 to Regulate Its Target Gene Expression and Suppress Leukemia Cell Article WT1 Recruits TET2 to Regulate Its Target Gene Expression and Suppress Leukemia Cell Proliferation. *Molecular Cell*, **57**(4), 662–673.

Ward, E., Desantis, C., Robbins, A., Kohler, B., & Jemal, A. (2014). Childhood and Adolescent Cancer Statistics , 2014. *Ca Cancer J Clin*, **64**(2), 83–103.

Warnecke, P. M., Stirzaker, C., Song, J., Grunau, C., Melki, J. R., & Clark, S. J. (2002). Identification and resolution of artifacts in bisulfite sequencing. *Methods*, **27**(2), 101–107.

Washbourne, P., & McAllister, A. K. (2002). Techniques for gene transfer into neurons. *Current Opinion in Neurobiology*, **12**(5), 566–573.

Weber, A. R., Krawczyk, C., Robertson, A. B., … Schar, P. (2016). Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nature Communications*, **7**(0372), 10806.

Weber, M., Davies, J. J., Wittig, D., … Schübeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genetics*, **37**(8), 853–862.

Wei, L., Li, P., Zhao, C., Wang, N., & Wei, N. (2019). Upregulation of microRNA-1270 suppressed human glioblastoma cancer cell proliferation migration and tumorigenesis by acting through WT1. *OncoTargets and Therapy*, **12**, 4839–4848.

Weissmann, S., Alpermann, T., Grossmann, V., … Kohlmann, A. (2012). Landscape of TET2 mutations in acute myeloid leukemia. *Leukemia*, **26**(5), 934–942.

Wesseling, P., van den Bent, M., & Perry, A. (2015). Oligodendroglioma: pathology, molecular mechanisms and markers. *Acta Neuropathologica*, **129**(6), 809–827.

Wheldon, L. M., Abakir, A., Ferjentsik, Z., … Ruzov, A. (2014). Transient accumulation of 5-carboxylcytosine indicates involvement of active demethylation in lineage specification of neural stem cells. *Cell Reports*, **7**(5), 1353–1361.

Wiedemeyer, W. R., Dunn, I. F., Quayle, S. N., … Chin, L. (2010). Pattern of retinoblastoma pathway inactivation dictates response to CDK4/6 inhibition in GBM. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(25), 11501–11506.

Wiedenheft, B., Sternberg, S. H., & Doudna, J. A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, **482**(7385), 331–338.

Williams, A. B., & Schumacher, B. (2016). p53 in the DNA-damage-repair process. *Cold Spring Harbor Perspectives in Medicine*, **6**(5), 1–16.

Williams, K., Christensen, J., & Helin, K. (2012). DNA methylation: TET proteins—guardians of CpG islands? *Nature Publishing Group*, **13**(1), 28–35.

Williams, K., Christensen, J., Pedersen, M. T., … Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature*, **473**(7347), 343–349.

Williams, S. R., Joos, B. W., Parker, J. C., & Parker, J. R. (2008). Papillary glioneuronal tumor: a case report and review of the literature. *Ann Clin Lab Sci*, **38**(3), 287–292.

Wojdacz, T. K., & Dobrovic, A. (2007). Methylation-sensitive high resolution melting (MS-HRM): A new approach for sensitive and high-throughput assessment of methylation. *Nucleic Acids Research*, **35**(6). doi:10.1093/nar/gkm013

Worm, J., Aggerholm, A., & Guldberg, P. (2001). In-tube DNA methylation profiling by fluorescence melting curve analysis. *Clinical Chemistry*, **47**(7), 1183–1189.

Wossidlo, M., Nakamura, T., Lepikhov, K., … Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nature Communications*, **2**(1). doi:10.1038/ncomms1240

Wright, J. H. (1910). Neurocytoma or Neuroblastoma, a Kind of Tumor Not Generally Recognized. *The Journal of Experimental Medicine*, **12**(4), 556–61.

Wu, Y.C., and Ling, Z. . (2014). The role of TET family proteins and 5-hydroxymethylcytosine in human tumors.

Wu, H., Alessio, A. C. D., Ito, S., Wang, Z., Cui, K., & Zhao, K. (2011a). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Cells and Development*, **25**, 679–684.

Wu, H., D'Alessio, A. C., Ito, S., … Zhang, Y. (2011b). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature*, **473**(7347), 389–394.

Wu, H., Wu, X., Shen, L., & Zhang, Y. (2014). Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nat Biotechnoilogy*, **32**(12), 1231–1240.

Wu, H., Wu, X., & Zhang, Y. (2016). Base-resolution profiling of active DNA

demethylation using MAB-seq and caMAB-seq. *Nature Protocols*, **11**(6), 1081–1100.

Wu, H., & Zhang, Y. (2011). Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes & Development*, **25**, 2436–2452.

Wu, H., & Zhang, Y. (2014). Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions. *Cell*, **156**(6), 45–58.

Wu, J. I. (2012). Diverse functions of ATP-dependent chromatin remodeling complexes in development and cancer Epigenetic Regulation of Transcription During Development Brg1 / Brm-Associated Factors (BAF), A Mammalian SWI / SNF-like ATP-dependent. *Breast*, 54–69.

Wu, X., & Zhang, Y. (2017). TET-mediated active DNA demethylation: Mechanism, function and beyond. *Nature Reviews Genetics*, **18**(9), 517–534.

Würstle, S., Schneider, F., Ringel, F., … Schlegel, J. (2017). Temozolomide induces autophagy in primary and established glioblastoma cells in an EGFR independent manner. *Oncology Letters*, **14**(1), 322–328.

Wyatt, A. W., Annala, M., Aggarwal, R., … Gleave, M. E. (2017). Concordance of Circulating Tumor DNA and Matched Metastatic Tissue Biopsy in Prostate Cancer. *Journal of the National Cancer Institute*, **109**(12), 1–9.

Wyatt, C. R., & Cohen, S. S. (1952). A new pyrimidine Base from Bacteriophage Nucleic Acids. *Nature*, **4338**(1073), 1072–1073.

Xie, S., Wang, Z., Okano, M., … Li, E. (1999). Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene*, **236**(1), 87–95.

Xu, J., Margol, A., Asgharzadeh, S., & Erdreich-Epstein, A. (2015). Pediatric brain tumor cell lines. *Journal of Cellular Biochemistry*, **116**(2), 218–224.

Xu, Y., Wu, F., Tan, L., … Shi, Y. G. (2011). Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. *Molecular Cell*, **42**(4), 451–464.

Xu, Y., Xu, C., Kato, A., … Shi, Y. G. (2012). Tet3 CXXC Domain and Dioxygenase Activity Cooperatively Regulate Key Genes for Xenopus Eye and Neural Development. *Cell*, **151**(4), 1200–1213.

Yamaguchi, S., Hong, K., Liu, R., … Zhang, Y. (2012). Tet1 controls meiosis

by regulating meiotic gene expression. *Nature*, **492**(7429), 443–447.

Yang, H., Lin, H., Xu, H., … Ye, D. (2014). TET-catalyzed 5-methylcytosine hydroxylation is dynamically regulated by metabolites. *Cell Research*, **24**(8), 1017–1020.

Ye, D., Ma, S., Xiong, Y., & Guan, K. L. (2013). R-2-hydroxyglutarate as the key effector of IDH mutations promoting oncogenesis. *Cancer Cell*, **23**(3), 274–276.

Yildirim, O., Li, R., Hung, J. H., … Fazzio, T. G. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell*, **147**(7), 1498–1510.

Yin, R., Mao, S. Q., Zhao, B., … Wang, H. (2013). Ascorbic acid enhances tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. *Journal of the American Chemical Society*, **135**(28), 10396–10403.

Yu, L., Baxter, P. A., Voicu, H., … Li, X. N. (2010). A clinically relevant orthotopic xenograft model of ependymoma that maintains the genomic signature of the primary tumor and preserves cancer stem cells in vivo. *Neuro-Oncology*, **12**(6), 580–594.

Yu, M., Hon, G. C., Szulwach, K. E., … He, C. (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*, **149**(6), 1368–1380.

Zalatan, J. G., Lee, M. E., Almeida, R., … Lim, W. A. (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell*, **160**(1–2), 339–350.

Zbinden, M., Duquet, A., Lorente-Trigos, A., Ngwabyt, S. N., Borges, I., & Ruiz I. Altaba, A. (2010). NANOG regulates glioma stem cells and is essential in vivo acting in a cross-functional network with GLI1 and p53. *EMBO Journal*, **29**(15), 2659–2674.

Zeng, C., Stroup, E. K., Zhang, Z., Chiu, B. C. H., & Zhang, W. (2019). Towards precision medicine: Advances in 5-hydroxymethylcytosine cancer biomarker discovery in liquid biopsy. *Cancer Communications*, **39**(1), 1–9.

Zeng, Y., & Chen, T. (2019). DNA methylation reprogramming during mammalian development. *Genes*, **10**(4). doi:10.3390/genes10040257

Zhang, L., Lu, X., Lu, J., … He, C. (2012). Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nature*

Chemical Biology, **8**(4), 328–330.

Zhang, Q., Zhao, K., Shen, Q., … Cao, X. (2015). Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. *Nature*, **525**(7569), 389–393.

Zhang, R., Cui, Q., Murai, K., … Xu, G. (2013). Short Article Tet1 Regulates Adult Hippocampal Neurogenesis and Cognition. *Stem Cell*, **13**(2), 237–245.

Zhang, W., & Xu, J. (2017). DNA methyltransferases and their roles in tumorigenesis. *Biomarker Research*, **5**(1), 1–8.

Zhang, Y., Liu, T., Meyer, C. A., … Shirley, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, **9**(9). doi:10.1186/gb-2008-9-9-r137

Zhang, Y., Ng, H. H., Erdjument-Bromage, H., Tempst, P., Bird, A., & Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes and Development*, **13**(15), 1924–1935.

Zhang, Z., Jin, Y., Zhang, W., … Zou, Q. (2019). Values of 5mC, 5hmC, and TET2 for identifying the presence and progression of breast precancerous lesion. *Journal of Clinical Laboratory Analysis*, (September), 1–8.

Zhao, S. L., Zhu, S. T., Hao, X., Li, P., & Zhang, S. T. (2011). Effects of DNA methyltransferase 1 inhibition on esophageal squamous cell carcinoma. *Diseases of the Esophagus*, **24**(8), 601–610.

Zhou, D., Alver, B. M., Li, S., … Robertson, K. D. (2018a). Distinctive epigenomes characterize glioma stem cells and their response to differentiation cues. *Genome Biology*, **19**(1), 1–25.

Zhou, J., Chen, Q., Zou, Y., Chen, H., Qi, L., & Chen, Y. (2019). Stem cells and cellular origins of breast cancer: Updates in the rationale, controversies, and therapeutic implications. *Frontiers in Oncology*, **9**(AUG), 1–12.

Zhou, T., Xiong, J., Wang, M., … Xu, R. M. (2014). Structural Basis for Hydroxymethylcytosine Recognition by the SRA Domain of UHRF2. *Molecular Cell*, **54**(5), 879–886.

Zhou, Y., Xia, L., Wang, H., … Cao, D. (2018b). Cancer stem cells in progression of colorectal cancer. *Oncotarget*, **9**(70), 33403–33415.

Zoghbi, H. Y., & Arthur L. Beaudet. (2004). Epigenetics and Human

Disease. *Annual Review of Genomics and Human Genetics*, **5**(1). doi:10.1146/annurev.genom.5.061903.180014

.

Appendix I - Publications

Eleftheriou, M, Pascual AJ, Wheldon LM, Perry C, Abakir A, Arora A, Johnson AD, Auer DT et al (2015) 5-Carboxylcytosine levels are elevated in human breast cancers and gliomas. Clinical Epigenetics, 7:88 (Designated "Highly accessed" by the journal)

Ramsawhook, A.H, Lewis, L.C, **Eleftheriou, M,** Abakir, A, Durczak, P, Markus, R, Rajani, S., Hannan, N.R, Coyle, B, and Ruzov, A (2017). Immunostaining for DNA modifications: computational analysis of confocal images. *JoVE (Journal of Visualized Experiments)*

Ehren R.V. Moler, Abdulkadir Abakir, **Maria Eleftheriou**, Jeremy S. Johnson, Konstantin V. Krutovsky, Lara C. Lewis, Alexey Ruzov, Amy V. Whipple, and Om P. Rajora. (2018). Population Epigenomics: Advancing Understanding of Phenotypic Plasticity, Acclimation, Adaptation and Diseases, in Om P. Rajora (ed.), Population Genomics: Concepts, Approaches and Applications, © Springer International Publishing AG, part of Springer Nature, pp. 179-260.2

Abakir, A, Giles, T, Cristini, A, Foster, J, Dai, N, Starczak, M, Rubio-Roldan, A, Li, M, **Eleftheriou, M**, Crutchley, J, Flatt, L, Young, L, Gaffney, D, Denning, C, Dalhus, B, Emes, R, Gackowski, D, Corrêa, I, Garcia-Perez, J, Klungland, A, Gromak, N, Ruzov, A (2020) N6-methyladenosine regulates the stability of RNA:DNA hybrids in human cells. Nat Genet 52, 48–55.

Eleftheriou M, *Ruzov A* Modified forms of cytosine in eukaryotes: DNA (de)methylation and beyond. In: Ruzov A. Gering M. *DNA Modifications: Methods and Protocols. Methods in Molecular Biology,* vol 2198, Springer US, DOI: 10.1007/978-1-0716-0876-0; ISBN 978-1-07-160875-3.