Investigating the Interplay between Wilms' Tumour Protein 1 and Oxidised Forms of 5-Methylcytosine in Brain Tumours



University of Nottingham
UK | CHINA | MALAYSIA

Ashley Harry Ramsawhook 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

October 2019

Abstract

5-methylcytosine (5mC) is an epigenetic modification usually associated with transcriptional repression. The Ten–Eleven Translocation proteins (Tet1/2/3) can oxidise 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) in vertebrate DNA. These modifications have not been well characterised but have been detected during embryogenesis, in somatic adult healthy and malignant tissues. Transcription factor and tumour suppressor/ oncogene Wilm's Tumour protein 1 (WT1), though not expressed in healthy brains displays over-expression in high grade neoplasms including Glioblastoma multiforme (GBM) and has been associated with poor patient prognosis. Interestingly, WT1 possesses preferential binding affinity for 5mC and 5caC over unmodified cytosines, 5hmC and 5fC *in vitro.* Thus we speculate that a WT1-5caC relationship may exist within a DNA demethylation associated and possible a transcriptional regulatory context in brain tumours. Utilising a range of techniques including genome editing and Next Generation Sequencing, we demonstrate that 5caC is enriched in brain tumour cell lines, is depleted upon *WT1* mutagenesis and may share genomic occupancy on gene regulatory and intragenic sequences. Additionally, we demonstrate a compromised neurosphere formation phenotype evident in WT1 zinc finger mutant GBM cell lines which correlates with perturbed extracellular matrix and cell adhesion gene expression, and is correlated with reduced tumour formation *in vivo*.

Acknowledgements

I would like to thank my supervisors Dr Alexey Ruzov and Dr Beth Coyle for their constant support and encouragement throughout my project. I would also like to thank the technical and administrative staff of both CBS and the CBTRC who have provided me with invaluable support and advice.

On a more personal note, I would like to first and foremost thank my parents and sister without whom, I would never have had the privilege of studying at this level to forge a career in a field I am so passionate about. I am eternally grateful and indebted to my family for everything and this thesis is dedicated to them.

I would like to thank my dearest friends Dr Lara Lewis, Maria Eleftheriou, Paige Lowe, Zubair Ahmed Nizamudeen, Nigel De Melo, Mohammed Diksin, Dr Elizabeth Scott, Dr Darren Cowley and Dr Sara Pijuan Galito for their continued support, kind words, amazing personalities and encouragement during very stressful times and throughout my tenure at Nottingham.

Special thanks to Dr Alexander Kondroshov and Dr Robert Marcus for lending their superb expertise and experience in CRISPR Cas9 design and confocal microscopy respectively.

Publications

A portion of the results of section 3 detailing the discovery and characterisation of oximC levels in paediatric medulloblastoma and ependymoma cell lines were published in Clinical Epigenetics: Ramsawhook, A., Lewis, L., Coyle, B. and Ruzov, A., 2017. Medulloblastoma and ependymoma cells display increased levels of 5-carboxylcytosine and elevated TET1 expression. *Clinical Epigenetics*, 9(1), p.18.

A portion of the results of section 3 describing the methodology of oximC profiling by immunofluorescence coupled confocal microscopy were published in the Journal of Visualised Experiments: 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)*, (127), p.e56318.

The introduction of section 4 on Wilms' Tumour 1 Biology was published in Frontiers in Cell and Developmental Biology: Ramsawhook, A., Ruzov, A. and Coyle, B., 2018. Wilms' Tumour protein 1 and enzymatic oxidation of 5-methylcytosine in brain tumours: potential perspectives. *Frontiers in cell and developmental biology*, *6*, p.26.

List of Tables

Table 2. 1 Media formulation.....11010

Table 2. 2 List of qPCR Primers 114

Table 2. 3 Standard PCR reaction mix.....11616

Table 2. 4 Thermocycler conditions 116

Table 2. 5 CRISPR Cas9 construction primers.....117

Table 2. 6 Phusion PCR conditions 11717

Table 2. 7 Denaturation / renaturation thermocycler conditions.....118

Table 2. 8 DNA Ligation conditions 120

Table 2. 9 Primary antibody specifications 130

Table 2. 10 Secondary antibody specifications.....1301

Table 2. 11 Antibody specifications.....1344

Table 2. 12 IHC specifications 1355

Table 2. 13 RT-qPCR reagent specifications 1356

Table 2. 14 DNA Immunoprecipitation reagent specifications 1366

Table 2. 15 Chromatin Immunoprecipitation reagents specifications 1366

Table 2. 16 DNA/ RNA purification reagent specifications 1366

Table 2. 17 NGS library preparation reagent specification 1377

Table 2. 18 siRNA reagent specifications 1388

Table 2. 19 CRISPR Cas9 reagent specifications 1388

Table 2. 20 Cloning reagents specifications 1388

Table 2. 21 Cell culture reagents specifications 1399

Table 2. 22 Western blot reagent specifications 1399

Table 2. 23 General reagent specifications 1399

Table 2. 24 Stem cell culture reagent specifications 14040

List of Figures

Figure 1.1 Schematic Representation of DNA Methylation Mechanism..
.....399

Figure 1.2 Multiple factors bind 5mC to influence transcriptional repression.. 432

Figure 1.3 Patterns of DNA methylation erasure and replenishment during human development.. 455

Figure 1.4 Mechanism of iterative oxidation mediated active DNA demethylation. 509

Figure 1.5 models of TET protein mediated 5mC oxidation.. 599

Figure 1.6 Transient accumulation of 5caC as adult mouse neural stem cells differentiate towards glial lineages. 644

Figure 1. 7 Figure 1.7 Physical 5caC occupancy on glial gene promoters correlates with their transcriptional up-regulation.. 644

Figure 1.8 Aberrant DNA methylation landscapes common in cancer.
.....766

Figure 1.9 Glioma stem cells may be tumour initiating cells. 898

Figure 1.10 Staining glioblastoma cell lines LN22, U87MG and LN18 for 5caC presence against HeLa and HEK293T 5caC negative cells.. **Error! Bookmark not defined.**5

Figure 1. 11 2.5D Confocal Microscopy Nuclear Spatial distribution of 5hmC (red) and 5caC (green) signal profiles in LN18 and U87MG glioblastoma cell lines vs 13.5 dpc mouse embryonic brain cells… **Error! Bookmark not defined.**

Figure 1.12. Transcriptional profile of glioblastoma pathway genes expression relative to 5caC negative HeLa cells 105

Figure 3.1 OximC presence in the mammalian brain…….

1444

Figure 3.2 Spectrum of oximC magnitudes between cell lines. 1465

Figure 3.3 Detection of oximCs in paediatric brain tumour cell lines..

.....1498

Figure 3.4 OximC profiles between cells are consistent with TET1 expression. 1511

Figure 3.5 *TET1* siRNA mediated knockdown may reduce 5caC magnitudes via down-regulating TET2 expression. 1543

Figure 3.6 Transcriptional signature of DNA methylation / demethylation across healthy adult somatic – pluripotent and malignant cell lines. 1566

Figure 3.7 Characterisation of GBM cell line DNA demethylation profiles with respect to foetal brain cells.. 1587

Figure 3.8 Examination of heterogeneous oximC staining distributions within cell lines.. 16059

Figure 3.9 Scale of oximC magnitude variation between circled cells with FB83 and GBM cell lines.. 1622
Figure 3.10 Correlation of 5caC signal intensity with neural stem cell marker expression.. 1643
Figure 3.11 Assessment of oximC presence in neural stem cell enriched conditions. 1666
Figure 3.12 Characterisation of DNA demethylation of activity in FB83-GBM cell lines under neurobasal conditions. 16968
Figure 3.13 Comparison of oximC magnitude and distribution characterisations in serum vs neurobasal conditions. 1721
Figure 3.14 Immunofluorescence Analysis of oximC distribution during GBM cell differentiation. 1754
Figure 3.15 Characterisation of DNA demethylation activity during GBM cell line differentiation. 17877
Figure 3.16 Characterisation of WT1 and 5caC enrichment in GBM cell lines. 1800
Figure 3.17 Effect of WT1 knockdown on DNA demethylation profiles in GBM cells. 1822
Figure 3.18 Examination of WT1 expression across GBM cell lines. 1844
Figure 3.19 Construction of WT1 –KTS and +KTS zinc finger plasmids. sequence incorporation. Blue region denotes the 9bp KTS sequence located between exons 9 and 10. 18786
Figure 3.20 Over-expression of WT1 –KTS or +KTS zinc finger domain plasmids in LN229 GBM cells.. 18989

Figure 3.21 Effect of WT1 –KTS and +KTS zinc finger isoform transfection on DNA demethylation in LN229 GBM cells. 1931

Figure 4.1 *WT1* DNA binding domain (exons 7-10) targeting strategy. 20808

Figure 4. 2 Parent plasmids utilised for generating (A) intermediate and (B) Donor HDR plasmids. 21009

Figure 4.3 Schematic of *WT1* zinc finger targeting HDR plasmid construction. 2111

Figure 4.4 UCSC Genome Browser alignment of WT1 gene sequence with genomic features of interest. 2133

Figure 4.5 Amplification of wild type WT1 introns 7 and 9 from U87MG genome. 2154

Figure 4.6 Fusion of *WT1* intron 7 and 9 homologous flanks. 2166

Figure 4.7 Intermediate plasmid. 21717

Figure 4.8 Verification of *WT1* intron 7 – 9 fusion incorporation into pBS vector.. 21818

Figure 4.9 Donor HDR parent pSIN-GFP-IRES-BSD plasmid. 22019

Figure 4.10 Complete WT1 targeting HDR donor plasmid. 2211

Figure 4.11 Verification of pEF-1a-eGFP-IRES-BSD selection cassette incorporation into the intermediate plasmid. 2221

Figure 4.12 small guide RNA (gRNA) cloning vector. 2243

Figure 4.13 Gibson Assembly incorporation of small guide RNAs into vectors.. 2254

Figure 4.14 Verification of sgRNA incorporation into gRNA cloning vector. 2265
Figure 4.15 Blasticidin kill curve and plasmid transfection optimisation in U251.....22928
Figure 4.17 Clone screening for genomic integration of the selection cassette within the WT1 cas9 cleavage locus.....2333
Figure 4.18 Verification of 5′ cassette genomic integration. 2344
Figure 4.19 Verification of 3′ cassette genomic integration. 2355
Figure 4.20 Verification of wildtype U251 WT1 exons 8 and 9 sequences. 23736
Figure 4.21 Screening of U251 C5A WT1 untargeted exon 8 and 9 loci.. 23838
Figure 4.22 Verification of WT1 exons 8 – 9 mutation at the transcription level. 2400
Figure 4.23 Translation of control and mutated WT1 coding sequences. 2422
Figure 4.24 Verification of WT1 protein truncation.. 2433

Figure 5.1 Assessment of WT1 expression in C5A and C7 mutants.. 25757
Figure 5. 2 Characterisation of DNA demethylation activity in wt1 mutants. 26059
Figure 5.3 Effect of wt1 mutagenesis on 5mC and oximC magnitudes. 2643

Figure 5.4 Effect of wt1 mutagenesis on GBM pathway gene transcription.. 2655
Figure 5.5 Profiles of 5caC genomic distribution between cell pluripotent stem cells and GBM cell lines.. 26968
Figure 5. 6 Profiling 5caC genomic distribution in pluripotent and GBM cell lines in gene regulatory and repetitive sequences.. 27069
Figure 5.7 Distribution of 5caC consensus peaks across cell lines. .2711
Figure 5.8 Consensus 5caC positive control peaks in RBLPAT cells. 2733
Figure 5.9 Differential 5caC genomic occupancy within genes in LN18 cells. 2754
Figure 5.10 Differential 5caC genomic occupancy between U251 and LN18 aligns with WT1 enrichment in U251. 2766
Figure 5.11 *In silico* Verification of WT1 ChIP peaks. 27978
Figure 5.12 5caC peak distribution within genes in U251 and LN18 cell lines. 2800
Figure 5.13 Distribution of 5caC and WT1 peaks within U251. 2822
Figure 5.14 Correlation of transcriptional activity with 5caC genomic occupancy. 2843
Figure 5.15 Analysis of WT1 isoform expression between U251 and C5A cell lines. 28686
Figure 5.16 Transcriptome analysis of wt1 mutagenesis in U251 and C5A cell lines under neurobasal conditions. 28989
Figure 5.17 GO analysis of differentially regulated genes in U251 vs C5A under neurobasal conditions. 2910

Figure 5.18 Examination of *in silico* protein interaction network between differentially regulated genes between U251 and C5A.. 2922

Figure 5. 19 Characterisation of the effects of wt1 mutagenesis on DNA demethylation under neural stem cell enrichment conditions.….....2954

Figure 5.20 Effects of wt1 mutagenesis on gliomasphere enrichment. 296

Figure 5.21 Profiling of the effect of wt1 mutagenesis on neurobasal condition-regulated target genes in U251. 299

Figure 5.22 Effects of wt1 mutagenesis on GBM cell line differentiation to neuronal and glial lineages.. 3021

Figure 5.23 Effects of *wt1* mutagenesis on GBM cell line differentiation to neuronal and glial lineages. 30303

Figure 5.24 Effect of wt1 mutagenesis on cell proliferation. 30404

Figure 5.25 Effect of wt1 mutagenesis on in vivo tumour formation. 30606

Figure A.1 Glioblastoma and FB83 cell line differentiation transcriptional profiles.. 35754

Figure B.1 Mouse orthotopic xenografts of U251 control vs C5A mutant derived tumours……...35858

List of Abbreviations

2-bromo-1-(4-dimethylamino-phenyl)-ethanone (BDAPE)

2-oxoglutarate (2OG)

5-carboxylcytosine (5caC)

5'-CG-3' (CpG)

5-formylcytosine (5fC)

5-methylcytosne (5mC)

8-hydroxy-2'-deoxyguanosine (8-oxodG)

Achaete-scute family bHLH transcription factor 1 (ASCL1)

Active removal (AR)

Activation-induced cytidine deaminase (AID)

Actively modified (AM)

Acute myeloid leukaemia (AML)

Adenosine Diphosphate (ADP)

Adeno-associated virus intergration site 1 (AAVS1)

Adenomatous polyposis coli (APC)

Adenosine triphosphate (ATP)

Adult human fibroblasts (GF06)

Alkylglycerol monooxygenase (AGMO)

Alpha actinins (ACTN2)

Annexin VI (ANXA6)

Apurinic/apyrimadinic endonuclease (APE)

ArfGAP with Dual PH Domains 1 (ADAP1)

Aryl hydrocarbon receptor (AHR)

ASH2 Like, Histone Lysine Methyltransferase Complex Subunit (ASH2L)

ATP-dependent, X-linked helicase (ATRX)

Base-excision repair (BER)

Basic helix-loop-helix (bHLH)

BCL2 associated transcription factor 1 (BCLAF1)

Bcl-2-like protein 13 (BCL2L13)

BCL6 Corepressor (BCOR)

Beta-catenin (*CTNNB1*)

Beta Tubulin III (TUBB3)

Blasticidin (BSD)

Bone morphogenetic proteins (BMPs)

Bovine growth hormone (bGH)

Brahma-related gene 1 (*BRG1/SMARCA4*)

Breast Cancer 1 (BRCA1)

C-terminal Cysteine-Histidine (C2-H2)

Cadherin 3 (CDH3)

Cadherin-12 (CDH12)

Carbon dioxide (CO_2)

Caspase 8 (CASP8)

Caveolin 2 (CAV2)

CCAAT enhancer binding protein alpha (C/EBPa)

Cell Division 5-Like (CDC5L)

Central nervous system (CNS)

Chitinase 3-like 1 (CHI3L1)

Chromatin immunoprecipitation (ChIP)

Chronic myelomonocytic leukaemia (CMML)

Chromatin target of PRMT1 (CHTOP)

Claudin 4 (CLDN4)

Closed MAD2 protein (cMAD2)

Cluster of differentiation 15/33/45/133 (CD15/33/45/133)

Clustered regularly interspersed palindromic repeats (CRISPR)

Collagen alpha-2(I) (COL1A2)

Collagen alpha-1(I4) (COL14A1)

Collagen alpha-1(I9) (COL19A1)

CRISPR RNA (crRNA)

CpG island methylator phenotype (CIMP)

CpG islands (CGI)

CUB and Sushi multiple domains 1 (CSMD1)

Cyclin D1 (CCD1)

Cyclin-dependent kinase 4/6 (CDK4/6)

Cyclin-dependent kinases (CDKs)

Cyclin Dependent Kinase Inhibitor 1/2/3A (CDKN1/2/3A)

D-2-hydroxyglutarate (D2HG)

Death-associated protein kinase 1 (DAPK1)

Days post coitum (dpc)

Delta-like protein 3 (DLL3)

Dideoxyribonuclotide triphosphates (ddNTPs)

Differentially methylated regions (DMR)

Digital affinity profiling via proximity ligation (DAPPL)

Diisopropyl ethylenediamine (DIPEA)

Dimethyl sulfoxide (DMSO)

Dissociation constants (K_D)

Dithiothreitol (DTT)

DmX-like protein 1 (DMXL1)

DNA immunoprecipitation (DIP)

DNA immunoprecipitation and subsequent sequencing (DIPSeq)

DNA methyltransferase (DNMT)

Doublecortin (DCX)

Double Homeobox 4 (DUX4)

Downstream double stranded beta helical (DSBH)

Dulbecco's Modified Eagles Medium (DMEM)

Dynein Axonemal Assembly Factor 7 (BLU)

E2F transcription factor 2 (E2F2)

E-Cadherin (CDH1)

Early Growth Response family 1 (EGR1)

Earle's Balanced Salt Soluion (EBSS)

Electro-mobility shift assays (EMSA)

Elongation Factor 1 Alpha (EF-1a)

Endogenous retroviral (ERV)

Enhancer of Zeste Homologue 2 (*EZH2*)

Epidermal growth factor receptor (EGFR)

Epidermal growth factor receptor vIII (EGFRvIII)

Ethylene-diamine tetra-acetic acid (EDTA)

Extracellular matrix (ECM)

Extracellular regulated kinases (ERK)

Family with sequence similarity 230 member C (FAM230C)

Fatty acid binding protein 3 (FABP3)

Fibroblast growth factor 3 (FGF3)

Fluorescence activated cell sorted (FACS)

Fluorescent *in situ* hybridisation (FISH)

Foetal Bovine Serum (FBS)

Fragments per kilobase million (FPKM)

Fluorescent Ubiquitination-based Cell Cycle Indicator (FUCCI)

Gamma-Aminobutyric Acid Type A Receptor Subunit Alpha1 (GABRA1)

Galectin 8 (LGALS8)

Gas chromatography coupled tandem mass spectrometry (GC-MS/MS)

Genetex (GTX)

Glial fibrillary acidic protein (GFAP)

Glioblastoma multiforme (GBM)

Glioblastoma stem cells (GSC)

Glioma CpG island methylator phenotype (G-CIMP)

Gli zinc finger protein 1/2/3 (GLI1/2/3)

Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)

Green Fluorescent Protein (GFP)

Gene Ontology (GO)

Growth Arrest Specific 1 (GAS1)

Hank's buffered saline solution (HBSS)

HEAT Repeat Containing 1 (HEATR1)

HIC ZBTB transcriptional repressor 1 (HIC1)

High-performance liquid chromatography (HPLC)

Hippocalcin Like 1 (HPCAL1)

Histone 3 lysine 4 acetylation (H3K4ac)

Histone 3 lysine 27 acetylation (H3K27Ac)

Histone 3 lysine 4 monomethylation (H3K4me1)

Histone 3 lysine 4 trimethylation (H3K4me3)

Histone 3 lysine 9 trimethylation (H3K9me3)

Histone 3 lysine 27 trimethylation (H3K27me3)

Histone 3 lysine 36 trimethylation (H3K36me3)

Histone deacetylase enzymes (HDACs)

Human induced pluripotent stem cells (hiPSCs)

Human mutL homolog 1 (hMLH1)

Homeobox Containing 1 (HMBOX1)

Homology directed repair (HDR)

Human pluripotent stem cell-derived neural progenitors (hPSC-NPCs)

Hydrochloric acid (HCL)

Hydroxylmethylcytosine (5hmC)

Hydroxymethyl DNA immunoprecipitation (hmeDIP),

Immunocytochemical (ICC)

Immunohistochemical (IHC)

Induced pluripotent stem cells (RBLPAT)

Integrative Genomics Viewer (IGV)

Integrin alpha-3 (ITGA3)

Integrin beta-1/3 (ITGB1/3)

Intercellular Adhesion Molecule 1 (ICAM)

Intracisternal A protein Particle repeats (IAP)

Iron (Fe^{2+})

Isocitrate dehydrogenase genes (*IDH1/2*)

J-Base binding proteins (JBP)

Jagged canonical notch ligand 1 (JAG1)

Janus Kinase 3 (JAK3)

Kilobases (Kb)

L1 Cell Adhesion Molecule (L1CAM)

Lamin B1 (LMNB1)

Left-Right Determination Factor B (LEFTY)

Leukaemia inhibitory factor (LIF)

Leukaemia inhibitory factor receptor (LIFR)

Long Interspersed Nuclear Element 1 (*LINE 1*)

Long terminal repeats (LTRs)

Luria Bertani (LB)

Lysine-threonine-serine ($KTS^{+/-}$)

Lysine Demethylase 2B (KDMB2)

Major late transcription factor (MLTF)

Mass/charge ratio (m/z)

Mass Spectrometry (MS)

Mast/stem cell growth factor receptor (c-KIT)

Matrix assisted laser desorption time of flight (MALDI-TOF)

MAX dimerization factor (MGA)

Meis Homeobox 1/2/3 (MEIS1/2/3)

MER tyrosine-protein kinase (MERTK)

MET hepatocyte growth factor receptor (MET)

Methyl-CpG binding domains (MBDs)

Methyl-CpG binding domain protein (MBD2)

Methyl-CpG binding Protein 2 (MeCP2)

Methyl group ($-CH_3$)

Methylation-Assisted Bisulphite Sequencing (MAB-Seq)

Methylation controlled J-protein (MCJ)

Methylguanine methyltransferase (*MGMT*)

Minimal Eagle's Medium (MEM)

Mixed Lineage Leukaemia (MLL)

Mitochondrial transfer RNA (tRNA)

Mouse embryonic fibroblasts (MEF)

Mouse embryonic stem cells (mESCs)

Myelodysplastic syndrome (MDS)

Nei endonuclease VIII-like 1/3 (NEIL1/3)

Nestin (NES)

Neurofilament Light (NEFL)

Neurofibromatosis type 1 (NF1)

Neuro-progenitor cells (NPCs)

New England Biolabs (NEB)

NFKB repressing factor (NKRF)

Non-essential amino acids (NEAA)

Non-homologous end joining (NHEJ)

NOTCH receptor 1/2/3 (NOTCH1/2/3)

Oligodendrocyte transcription factor 1/2 (OLIG1/2)

Optical density at 600nm (OD_{600})

Ovo Like Zinc Finger 2 (OVOL2)

Oxidised 5mC derivatives (oximCs)

Passively diluted (PD)

Patient derived xenograft (PDX)

Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha (PIK3CA)

Phosphatase and tensin homologue (*PTEN*)

Phosphate buffered saline (PBS)

Phosphorylated retinoblastoma protein (p-pRb)

Platelet Derived Growth Factor Receptor Alpha (PDGFRA)

Platelet factor 4 (PF4)

Pleckstrin homology domain (PHD)

Poly adenylation (polyA)

PolyComb Group (PcG)

Polycomb Repressive Complex (PRC2)

Polytriphosphate (PPPi)

Positional weight matrices (PWMs)

Post coital week (pcw)

Posterior fossa type A (PFA)

Posterior fossa type B (PFB)

Potassium Voltage-Gated Channel Subfamily Q Member 5 (KCNQ5)

PRELI domain-containing protein 2 (PRELID2)

Phosphatidylinositol-3,4,5-trisphosphate (PIP3)-dependent Rac exchanger (PREX2)

Proliferating cell nuclear antigen (PCNA)

Prominin1 (PROM1)

Protein C-ets-1 isoform 2 (ETS1)

Proto-spacer activation motif (PAM)

Quantitative polymerase chain reaction analysis (QPCR)

Retinoic Acid Receptor Beta (RARB)

Radial glial progenitors (RGPs)

Recursive Partition Mixture Model (RPMM)

Reduced representation bisulphite sequencing (RRBS)

RELB proto-oncogene, NF-KB subunit (RELB)

Retinoblastoma protein 1 (RB1)

Retinoblastoma Binding Protein 5 (RBBP5)

Reverse transcriptase PCR (RT-PCR)

Rho Associated Coiled-Coil Containing Protein Kinase 1 (ROCK1A)

Ribosomal RNA (rRNA)

RING1 finger protein 1A/B (RING1A/B)

Ring Finger domains -1 (UHRF1)

RNA induced silencing complex/Argonaute (RISC/AGO)

RNA recognition motif (RRM)

S-adenosyl methionine (SAM/AdoMet)

Search Tool for the Retrieval of Interaction Genes/ Proteins (STRING)

Set and Ring domain (SRA)

Sequence-By-Synthesis (SBS)

Sex determining region Y-box 2 (SOX2)

Small guide RNA (sgRNA)

Short hairpin RNA (shRNA)

Single stranded selected monofunctional uracil glycosylase (SMUG1)

SMAD1 interacting protein 1 (SNIP1)

Small interference RNA (siRNA)

Smoothened (SMO)

Sodium boro-hydride ($NaBH_4$)

Solute Carrier Family 12 Member 5 (SLC12A5)

Sonic hedgehog (Shh)

Specificity protein 1 (SP1)

Stable isotope labelling by amino acids in cell culture (SILAC)

SWItch/ Sucrose Non Fermentable (SWI/SNF)

SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 4 (SMARCA4)

Synaptotagmin-1 (SYT1)

Synuclein Alpha Interacting Protein (SNCAIP)

TATA-Binding Protein (*TBP*)

TATA-Binding Protein or transcription factor IIB (TFIIB).

Ten-Eleven Translocation (TET)

TET-assisted bisulphite sequencing (TAB-Seq)

The cancer genome atlas (TCGA)

Thermus aquaticus (*Taq*)

Thin Layer Chromatography (TLC)

Thymine DNA glycosylase (TDG)

Thymosin Beta 4 (Tβ4)

Thyroid Hormone Receptor Interactor 6 (TRIP6)

Tight Junction Protein 1 (TJP1)

TIMP Metallopeptidase Inhibitor 3 (TIMP3)

TNF Receptor Associated Factor 6 (TRAF6)

Transactivating RNA (tracrRNA)

Transcription activator like effector nucleases (TALENs)

Transcription factor (TF)

Transcription factor 4 (TCF4)

Transcription factor II D (TFIID)

Transcriptional start site (TSS)

Transmembrane Protein with EGF Like And Two Follistatin Like Domains 1 (TMEFF1)

Transforming growth factor beta receptor (TGFβR)

Tumour necrosis factor receptor type 1-associated DEATH domain (TRADD)

Tumour necrosis factor receptor superfamily member 1A (TNFRSF1A)

Tumour Protein 53 (TP53)

Two-dimensional Ultra-Performance Liquid Chromatography coupled tandem Mass Spectrometry (2D-UPLC-MS/MS)

Tyramide signal amplification (TSA)

Untranslated Regions (UTR)

Uracil DNA Glycosylase (UDG)

World Health Organisation (WHO)

Wilms' Tumour Protein 1 (*WT1*)

Zinc fingers (ZNF)

Zinc finger nucleases (ZFNs)

Thesis statement and contributions

The work presented in this thesis is mine (Ashley Harry Ramsawhook) unless otherwise stated. This project was supervised by Dr Alexey Ruzov and Dr Beth Coyle.

Dr Lara Lewis contributed to this work through (1) 5hmC and 5caC immunostaining of RBLPAT cells (fig 3.2); (2) WT1 western blots (fig 4.24B).

Ryszard Olinsky (Nicolaus Copernicus University, Poland) performed mass spectrometric analysis of oximC marks (fig .33B-E).

Maria Eleftheriou performed 5hmC and 5caC immunostaining of LN18, LN229, U251 and U87MG cell lines (fig 3.6A), WT1 and 5caC staining of LN18, LN229, U251 and U87MG (fig 3.16) and WT1 western blots (fig 4.24A)

Vithusha Wijaykumar performed 5hmC, 5caC and stem marker immunostaining on LN18, LN229, U251 and U87MG cell lines (fig 3.10A)

Paige Lowe performed RNA extraction, cDNA synthesis and qPCR analysis on FB83 cell line (Appendix fig A.1).

Next Generation Sequencing (NGS) runs on 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).

Dr Phil Clarke and Alison Ritchie performed mouse orthotopic xenograft injections and brain harvesting and embedding. Sectioning was performed by Dawn Broomfield.

Contents

Abstract 2

Acknowledgements 3

Publications 4

List of Tables 5

List of Figures 6

List of Abbreviations 13

Thesis statement and contributions 25

General Introduction 35

1.1 DNA Methylation 35

1.1.1 *De novo* Methylation 36

1.1.2 Maintenance Methylation 38

1.1.3 Effect of DNA Methylation on Transcription 40

1.1.4 DNA Methylation is Dynamic 43

1.2 DNA Demethylation Mechanisms 47
1.2.1 Passive Demethylation 47
1.2.2 Deamination Pathway 48
1.2.3 Active Iterative Oxidation 48
1.2.4 Mechanism of Oxidation 50
1.2.5 TET Protein Expression and Distribution 52
1.2.6 Distribution of Oxi-methyl-Cytosine Modification 54
1.2.7 Functional significance of oxi-methylcytosine derivatives 58
1.2.8 Readers of Oxi-methylcytosine Derivatives 66
1.3 Epigenetic Aberrations in Brain Tumours 73
1.3.1 Perturbed DNA Methylome Landscapes Are Often Characteristic of Malignancy 74
1.3.2 Multiple Factors May Influence Adult Glioblastoma Multiform Methylome 77
1.3.3 5hmC Genomic Distribution in Glioblastoma Multiforme 82
1.3.4 5mC and 5hmC Distributions within Glioma Stem Cells 88
1.4 Wilm's Tumour 1 Biology 92
1.4.1 Developmental basis for aberrant expression of Wilms' Tumour protein 1 in cancers 92
1.4.2 Impact of ectopic WT1 expression on tumorigenesis 95
1.4.3 WT1 may modulate epigenetic signatures in tumours 98
1.4.4 Hypothesis and Aims 104
Materials & Methods 109
2.1 Cell Culture 109
2.1.1 Cryopreservation of cells 109
2.1.2 Passaging and sub-culturing of cells 110
2.1.3 Preparation of complete media 111

2.1.4 Spheroid Culture 112
2.2 Molecular Biology 112
2.2.1 siRNA Transfection 112
2.2.2 RNA Extraction 113
2.2.3 cDNA Synthesis 113
2.2.4 Quantitative Polymerase Chain Reaction 114
2.2.5 Standard polymerase chain reaction 115
2.2.6 Phusion PCR 117
2.2.7 Gibson Assembly 118
2.2.8 Restriction digests 118
2.2.9 Agarose gel electrophoresis and gel extraction 118
2.2.10 Quantification of nucleic acid yield and purity 119
2.2.11 Ligation 120
2.2.12 Bacterial Transformation 120
2.2.13 Starter culture 121
2.2.14 Overnight culture 121
2.2.15 Plasmid mini-prep 121
2.2.16 Plasmid midi-prep 122
2.2.17 Chromatin Immunoprecipitation (ChIP) 122
2.2.18 Methylated DNA immunoprecipitation (MeDIP) 122
2.2.19 DNA Immunoprecipitation-qPCR 123
2.3.1 Lipofection 124
2.3.2 Electroporation 125
2.3.3 Limiting dilution Assay 126
2.4 Next Generation Sequencing 127
2.4.1 Qubit DNA quantification 127
2.4.2 Adapter ligation (MeDIP-Seq and ChIP-Seq) 127

2.4.3 PCR amplification of MeDIP samples 127
2.4.4 Tapestation analysis of RNA integrity 128
2.4.5 Ribo-depletion 128
2.4.6 Transcriptome library preparation 128
2.4.7 Sequencing 129
2.6 Staining & Imaging 129
2.6.1 Immunofluorescence 129
2.6.2 Fluorescent & confocal microscopy 131
2.7 Functional Studies 131
2.7.1 Cell proliferation 131
2.7.2 Differentiation programmes 131
2.8 Statistical Analysis 132
2.9 Bioinformatics Analysis 133
2.10 Mouse Orthotopic Xenograft 133
2.11 Ethics Statements 133
2.12 List of Reagents 134
Results 141
Characterisation of DNA Demethylation Activity in Brain Tumour Cell Lines 141
3.1 Introduction 141
3.2 Detection of OximC Presence 143
3.2.1 5caC is Detectable in Mammalian Foetal but not Adult Brains 143
3.2.2 Adult Glioblastoma and Foetal Brain Cell lines Display Similar 5caC Signal Magnitudes 145

3.2.3 5caC is Detectable in Paediatric Medulloblastoma and Ependymoma Cell Lines 148

3.2.4 Elevated 5caC Levels Correlate with High *TET1* Expression in Medulloblastoma and Ependymoma Cell Lines 150

3.2.5 TET1 siRNA Knockdown Correlates with Diminished OximC Enrichment in UW228-3 Medulloblastoma Cell Lines 153

3.2.6 Adult Glioblastoma Cell Lines Possess a Unique DNA Demethylation Transcriptional Profile 155

3.2.7 Foetal Brain and Adult GBM Cell Lines show OximC Parity but Transcriptional Discordance 157

3.2.8 Foetal Brain and Adult Glioblastoma Cell Lines Display Heterogeneous Cellular 5hmC and 5caC Distributions 159

3.2.9 Significant Variation in 5hmC Signal Intensity Produce a Perceived Effect of 5caC Cellular Heterogeneity 161

3.2.10 5caC Signal Intensity Does Not Correlate with Neural Stem Cell Markers in Foetal Brain or Glioblastoma Cell Lines 162

3.3 Modulation of OximC Nuclear Distribution 164

3.3.1 Neurosphere Enrichment Conditions Alter OximC Distribution in Adult Glioblastoma Cell Lines 165

3.3.2 Adult Glioblastoma Cell Lines Exhibit Reduced OximC Presence which Correlates with Depleted *TET1* and *TET3* Expression Relative to Foetal Neural Stem Cells 167

3.3.3 Neurosphere Enrichment Significantly Affects OximC Magnitude and Distribution Compared to Serum 169

3.3.4 OximCs Transiently Accumulate during GBM Cell Differentiation 173

3.3.5 Dynamic Patterns of 5caC Abundance Correlate with *WT1* Transcriptional Expression 175

3.4. Investigating the Dynamics of a Possible WT1-5caC Interaction 179

3.4.1 WT1 Expression Correlates with 5caC Presence in all GBM Cell Lines except LN229 179

3.4.2 Transient *WT1* knockdown correlates with reduced global oximC nuclear presence but elevates retro-transposon 5caC enrichment in U87MG cells 181

3.4.3 -/+KTS Splice Variants of *WT1* are Undetected in LN229 .. 183

3.4.4 Verification of –KTS and +KTS Zinc Finger Domain Construct Production 185

3.4.5 Forced Expression of WT1 –/+KTS Zinc Finger Domain in LN229 188

3.4.6 Over-expression of WT1 –KTS and +KTS zinc finger domains significantly alters 5caC magnitudes and induces full-length *WT1* transcription 189

3.5 Technical Discussion 193

Design, Construction and Generation of *wt1* mutant GBM Cell Lines 203

4.1 Introduction 203

4.2 Rationale for Selection of WT1 Target Site 206

4.2.1 WT1 Targeting Strategy 207

4.2.2 Intermediate Plasmid Construction 212

4.2.3 WT1 HDR Donor Plasmid Construction 219

4.2.4 Construction of Guide RNA Plasmids 222

4.3 Transfection Optimisation and Selection 226

4.3.1 Survival Analysis and Transfection Optimisation 226

4.3.2 FAC Sorting of GFP Positive Survivors 229

4.4 Clone Screening and Genotyping 231

4.4.1 Analysis of Cassette Integration 232

4.4.2 Analysis of WT1 Wildtype Zinc Finger Locus 235

4.4.3 CRISPR cas9 U251 and LN18 *wt1* Mutants are Predicted to Produce Truncated WT1 Proteins 239

4.5 Technical Discussion 244

Characterisation *wt1* mutant GBM Cell Lines 253

5.1 Introduction 253

5.2 Analysis of U251 and LN18 *wt1* Mutant DNA Demethylation Profiles 255

5.21 CRISPR cas9 *wt1* Mutant GBM Cell Lines Display Reduced WT1 Gene and Protein Expression 256

5.22 GBM Cell Line *wt1* Mutants Display Significantly Altered 5hmC and 5caC Distributions Correlating with Elevated *TET* Gene Expression 258

5.23 5mC and oximC magnitudes remain consistent in *wt1* mutants relative to their controls 262

5.24 Common GBM Pathology Genes are Up-regulated in *wt1* Mutants 264

5.3 Next Generation Sequencing Characterisation of 5caC Genomic Distribution and Transcriptome Regulation 266

5.31 5caC Enrichment in GBM Cell Lines Occurs at Transcriptional Regulatory Regions, within Gene Bodies and on Repetitive Sequences 266

5.32 5caC Genomic Peaks in U251 and LN18 Predominately Align to Centromeric Repeats 271

5.33 LN18 and U251 Possess Distinct 5caC Enrichment Sites 274

5.34 5caC and WT1 Peaks Align in U251 275

5.35 LN18 and U251 Share Common 5caC Genomic Occupancy 279

5.36 A Subset of Genes Possess WT1 and 5caC Aligned at Identical Genomic Coordinates 280

5.37 5caC Presence Does Not Faithfully Correlate with Transcriptional Up-Regulation of Genes 283

5.4 Transcriptomic Analysis of C5A *wt1* Mutants 285

5.41 C5A *wt1* Mutants Highly Express a Novel *wt1* Isoform 285

5.42 Multiple Genes Demonstrate Strong Differential Expression between C5A *wt1* Mutant and U251 Control Cells 287

5.5 Functional Characterisation of GBM *wt1* Mutants 293

5.51 Mutagenesis of *WT1* May Affect DNA Demethylation Dynamics during Neural Stem Cell Enrichment 293

5.52 C5A and C7 *wt1* Display Diminished Gliomasphere Formation Capacity 295

5.53 Cell Adhesion, Extracellular Matrix and Stemness Genes are Strongly Down-Regulated in C5A *wt1* Mutants 298

5.54 C5A and C7 *wt1* Mutants Exhibit Diminished *TUBB3* and *GFAP* Expression Patterns during Differentiation 300

5.55 C5A and C7 *wt1* Do Not Display Notable Changes in Cellular Proliferation 304

5.56 Reduced Size and Frequency of Tumour Formation in C5A *wt1* Orthotopic Xenograft Mice 305

5.6 Technical Discussion 307

General Discussion 314

6.1 Profiling OximC Dynamics in Brain Tumour Cell Lines 314

6.21 Caveats and Limitations 319

6.3 Functional Impact of WT1 on DNA Demethylation Dynamics 325

6.31 Caveats and Limitations 330

6.4 Significance 331

6.5 Future Work 349
Concluding Remarks 352
Appendix A 354
Glioblastoma and Foetal Brain Cell Line Differentiation 354
Appendix B 358
Mouse Orthotopic Xenografts 358
Bibliography 359

General Introduction

To fully appreciate DNA demethylation mechanisms and their implications on gene expression via effecting transcriptional machinery and chromatin conformation, rudiments of DNA methylation must first be understood.

1.1 DNA Methylation

DNA methylation, a well-documented mechanism associated with transcriptional regulation, entails modification of cytosine residues in a CpG dinucleotide context via the addition of methyl group (-CH_3) to the

5′- carbon atom of the cytosine pyrimidine ring to form 5-methylcytosine (5mC) (Bird, 2002). Approximately 70% of CpGs are methylated which constitutes only 1% of mammalian genomes as they are depleted of this palindromic sequence owing to 5mC mutagenic propensity to spontaneously deaminate to thymine (Youssoufian *et al*, 1985).

The presence of methyl groups on gene promoter sequences shows strong correlation with transcriptional repression in vertebrates (Klose & Bird, 2006). Addition of these methyl groups is catalysed predominantly by highly conserved DNA methyltransferase (DNMT) enzymes DNMT3a, 3b, 3C and 3L, and DNMT1 which modify CpG cytosines in *de novo* and maintenance methylation contexts respectively (Kim *et al*, 2002; Barau *et al*, 2016).

1.1.1 *De novo* Methylation

Methylation modifications on DNA are not inherited from parent to offspring and therefore must be re-established *de novo* following demethylation post fertilization (Santos *et al*, 2002). *De novo* methylation is performed by DNMT3A, B, C and L which catalyse transfer of methyl groups to unmethylated palindromic CpG residues (Santos *et al*, 2002; Barau *et al*, 2016). In mice, DNMT3A/B share structural similarity with another methyltransferase, DNMT2 which is functionally redundant (Reik, Kelsey & Walter, 1999). DNMT3A/B exhibit elevated expression during development in embryonic stem cells and in the epiblast however diminished expression is observed following pluripotent cell lineage commitment to somatic fates during differentiation (Okano, Bell & Haber,

1999). *DNMT3a/b* mutagenesis resulting in functionally null proteins in mouse embryonic stem cells fail to establish methylation profiles commensurate with those departing stemness towards lineage commitment (Okano, Bell & Haber, 1999). This is reflected in *DNMT3a/b*$^{-/-}$ maintenance of *Oct3/4* expression. Additionally, *DNMT3a/b*$^{-/-}$ mouse NIH3T3 fibroblasts cannot methylate viral DNA - an evolutionary hallmark of cellular defence mechanisms against exogenous nucleic acids (Okano, Bell & Haber, 1999). Whilst sharing functional redundancy, DNMT3a and 3b display tissue-specific expression patterns with 3a detected uniformly in mouse embryos but 3b predominantly localised to neuroectoderm and chorionic ectoderm tissues (Okano, Bell & Haber, 1999). Consequent of non-overlapping expression, *DNMT3a*$^{-/-}$ mutants survive 2-3 weeks postnatal whereas *DNMT3b*$^{-/-}$ mutants are embryonically lethal, failing to undergo gastrulation and terminating at embryonic day 9.5 (E9.5) (Li, Bestor & Jaenisch, 1992 & Okano *et al*, 1999). The *de novo* DNA methyltransferase DNMT3C was recently discovered to methylate promoters of transcriptionally active retrotransposons in the genomes of male mice (Barau *et al*, 2016). DNMT3L is an auxiliary factor which complexes with DNMT3a/b to enhance their activity but possesses no methyltransferase activity of its own. However, *DNMT3L* null mutant mice are infertile (Hata *et al*, 2002) as its expression is highest in wildtype primordial germ cells post migration to genital ridges, highlighting the importance of *de novo* methylation in gamete maturation (Bourc'his *et al*, 2001).

1.1.2 Maintenance Methylation

Methylation signatures can be inherited during mitosis from somatic and parent to daughter cells (Eckhardt *et al*, 2006) and in primordial germ cells on repetitive elements prior to meiosis (Lees-Murdock *et al*, 2008). Maintenance methylation involves DNMT1 facilitated modification of CpG cytosine residues existing in hemi-methylated palindromes on double-stranded DNA (Hermann, Goyal & Jeltsch, 2004). DNMT1 can be recruited to DNA at replication forks (Sharif *et al*, 2007) and consequently, genomic methylation levels peak during S phase of the cell cycle (Brown *et al*, 2007). DNMT1 methylates unmodified cytosines thereby demarcating nascent synthesized DNA strands, promoting recognition of self/ non-self genetic material as defence against viruses, X-chromosome inactivation and maintains transcriptional repression profiles (Lin *et al*, 2000). Through recognition of hemi-methylated DNA CpG sequences, DNMT1 maintains established patterns of methylation demarcated by *de novo* methylation e.g. repression of Long Interspersed Nuclear Element 1 (*LINE 1*) retrotransposon promoters to inhibit its potentially carcinogenic propagation (Chalitchagorn *et al*, 2004). Although possessing hemi-methylated DNA preferential binding affinity, DNMT1 can methylate unmethyated CGI sequences in *DNMT3a/b* $^{-/-}$ mutant cells, fulfilling an emergency *de novo* methylation role (Fatemi *et al*, 2002). Thus owing to the semi-conservative nature of DNA replication (Gruenbaum *et al*, 1982), maintenance methylation can faithfully recapitulate methylation signatures from parent to daughter cell (fig 1.1) (Dean, Santos & Reik, 2003).

The source of the methyl group is produced via cyclical methionine biosynthesis. Briefly, methionine adenosyl synthetase facilitated adenosine triphosphate (ATP) hydrolysis generates S-adenosyl methionine (SAM/AdoMet) which reacts with sulfonium lone pair electrons on methionine via nucleophilic substitution. Catalysis of polytriphosphate (PPPi) to adenosine diphosphate (ADP) and inorganic phosphate (Pi) is achieved through innate polyphosphotase activity of methionine adenosyl synthestase. The methyl group attached to methionine sulfonium is cleaved by methyltransferases and acts as a donor for DNA and RNA methylation at CpG and uracil residues respectively (Chiang *et al*, 1996).

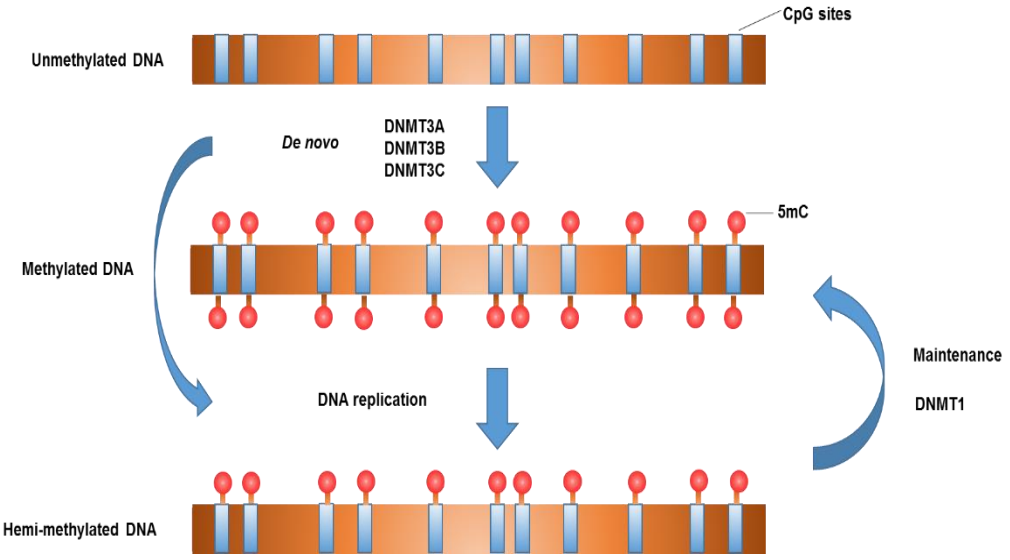


Figure 1.1 Schematic Representation of DNA Methylation Mechanism. DNA methyltransferase (DNMTs) enzymes 3a, 3b and 3c (DNMT3a/b/c) catalyse the transfer of methyl groups (red circles) to unmodified DNA molecules (orange horizontal bar) at CpG sites (blue boxes). Maintenance

methylation enzyme Dnmt1 only transfers methyl groups onto hemi-methylated DNA molecules resulting in double-stranded CpG methylation. (Original figure).

1.1.3 Effect of DNA Methylation on Transcription

To achieve 5-carbon atom methylation of cytosine, DNMTs must access the nucleotide situated within the concave conformation of DNA major grove (Arita *et al*, 2008) Ubiquitin-like containing pleckstrin homology domain (PHD) and Ring Finger domains -1 (UHRF1), recruits DNMT1 to replication forks at hemi-methylated CpG palindromes (Hashimoto *et al*, 2009). The UHRF1 Set and Ring domain (SRA) catalyses the 180° base-flipping and subsequent 65° rotation required to disrupt dipole-dipole base stacking forces (Cheng & Roberts, 2001) and grant DNMT1 access to the cytosine to be modified. Interestingly, the base flipping mechanism is not cytosine-specific but the SAM/AdoMet-dependent methyltransferase reaction is (Hashimoto *et al*, 2009). As posited by Watt & Molloy, 1988, methylation of CpG dinucleotides may attenuate gene transcription by occlusion of basal transcriptional machinery binding to CpG containing transcriptional regulatory sequences (Watt & Molloy, 1988). However evidence revealing unperturbed binding of SP1 transcription factor to high density (10 nucleotide) CG rich consensus sequence with or without methylation suggest the CpG methyl group transcriptional repressive effects may be minimal and or protein specific (Harrington *et al*, 1988). Kass, Pruss & Wolffe suggested an alternate mechanism of transcriptional repression in which methyl-binding domain containing transcription factors recognise and bind to methylated CpG

regulatory sequences thereby contributing to silencing of transcription (Kass, Pruss & Wolffe, 1997).

Indeed, CpG methylation binding proteins can be classified into three families: (1) methyl-CpG binding domain (MBD) family such as Methyl-CpG binding Protein 2 (MeCP2) and MBDs 1-6; (2) Kaiso and Kaiso-like family proteins such as Zinc finger and BTB domain containing (ZBTB) proteins ZBTB4 and ZBTB38 and (3) SET and RING Associated domain (SRA) family proteins such as UHRF1 (Fournier *et al*, 2012).
Proteins such as MeCP2, MBD1, MBD2, MBD3 and MBD4 recognise methylated CpG sequences (Zou *et al*, 2011). The cysteine-rich histidine-rich zinc finger proteins ZBTB4 and ZBTB38 are capable of binding to single methylated CpG containing motifs in a methylation-dependent manner, both with specificity for the methylated *H19/Igf2* locus (Filion *et al*, 2006). Interestingly, MBD3, MBD5 and MBD6 proteins cannot bind to methylated CpG sequences but instead complex with PolyComb Group (PcG) chromatin remodelling complex proteins (Baymaz *et al*, 2014) whilst protein arginine N-methyltransferase PRMT11 protein harbouring MBD7, is expressed in *Arabidopsis thaliana* (Scebba *et al*, 2007).

Methyl CpG binding domains are highly conserved between members of the protein family and consist of loop L1, Beta Sheets B3, B4 and L2 loops within the methyl-CpG binding pocket. Arginine 22 and lysine 46 of the peptide sequence form cationic electrovalent bonds with methylated cytosine whilst hydrogen bonding between opposite cytosine and

guanine and dipole base stacking forces between adjacent nucleotides produce a trifecta of interactions occurring at the DNA-methyl CpG domain interface (Zou *et al*, 2011). These interactions stabilise MBD proteins- DNA binding thus facilitating recruitment of nucleosome remodelling complexes (Hashimshony *et al*, 2003). These complexes such as Nucleosome remodelling and histone Deacetylase (NuRD) and Sin3 consisting scaffolding proteins RbAp46 and RbAp48 mediating ATPase-dependent targeting of histones by histone deacetylase enzymes (HDACs) e.g. HDAC1/2 (Zhang *et al*, 1999). These enzymes catalyse removal of acetyl groups from histone tails at specific residues e.g. Histone 3 lysine 4 acetylation thereby effecting repressive chromatin conformations associated with repressive histone marks such as H3K9me3 (Denslow & Wade, 2007; Nan & Williams Jr, 2019). NuRD and Sin3 complexes may convene at methyl-CpG sequences via interaction with MBD2 and MeCP2 respectively (fig 1.2), thus resulting in locus specific transcriptional silencing through nucleosome remodelling (Ahringer, 2000).

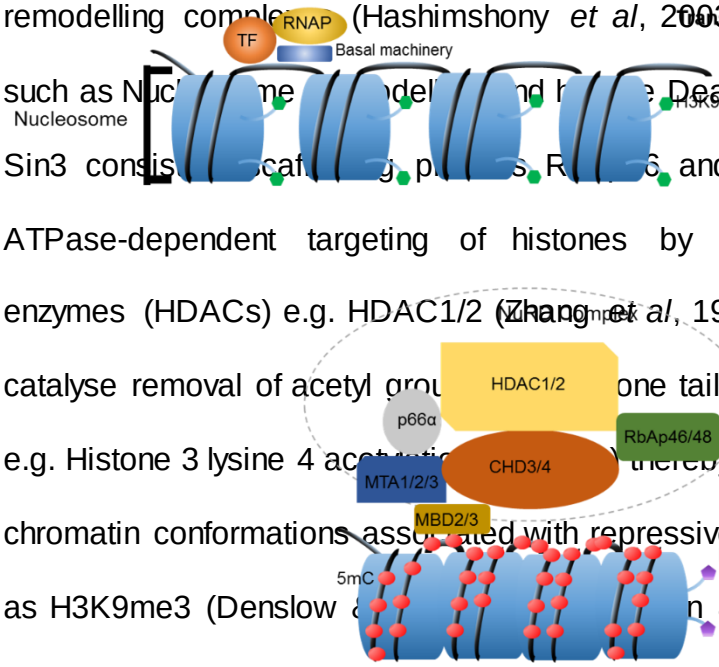


Figure 1.2 Multiple factors bind 5mC to influence transcriptional repression. DNA methylation (red circles) on CpG residues are recognised by methyl-binding domain (MBD) proteins which recruit chromatin remodelling proteins to create a condensed and transcriptionally repressive conformation. Proteins such as MBD2/3 are attracted to open chromatin but transcription factors (TF) and RNA polymerase (RNAP) which bind to transcriptionally permissive chromatin are precluded from binding to repressive chromatin. The repressive chromatin decorated with histone and DNA methylation is associated with NuRD complex proteins such as histone deacetylases 1 & 2 (HDAC1/2), Retinoblastoma associated proteins 46 & 48 (RbAp46/48), chromatin helicase domains 3 & 4 (CHD3/4), GATA zinc finger domain 2 protein A (p66α) and Metastasis Associated protein 1/2/3 (MTA1/2/3). (Original figure).

1.1.4 DNA Methylation is Dynamic

During embryogenesis transcriptionally repressive methylation profiles in gametes undergo erasure by passive and active mechanisms to ensure activation of pluripotency and embryonic development genes whilst silencing terminal differentiation somatic ones (fig 1.3) (Wu & Zhang, 2010). Studies performed by Monk, Boubelik and Lehnert using CpG methylation dependent differential cleavage by restriction enzymes *MspI* and *HpaII* (*HpaII* cannot cleave internal methylated Cytosine in 5'–CCGG-3' sequences to 5'–CC–3' and 5'–GG-3' whereas *MspI* can) followed by α-^{32}P[dCTP] end-labelling, highlighted distinct patterns of methylation and its erasure occurring at landmark stages of embryonic mouse development (Monk, Boubelik & Lehnert, 1987). Analysis of

sperm, ovulated oocyte, 8-cell blastula, blastocyst, E6.5, E7.5 and E16.5 post implantation DNA revealed significant methylation of sperm and blastula genomes compared to embryonic stem cells, E11-13 foetal germ cells and oocytes, the latter of which appeared hypomethylated. Highest methylation densities were observed in E7.5 embryos (Monk, Boubelik & Lehnert, 1987). Owing to high background signal attributable to high density unmethylated CGI sequences, global absolute methylation levels could not be achieved (Monk, Boubelik & Lehnert, 1987). Multiple studies corroborate these findings (Oswald *et al*, 2000; Dean *et al*, 2001; Iqbal *et al*, 2011 Wossidlo *et al*, 2011), facilitating refinement of the embryonic DNA dynamic methylation model. Distinct methylation profiles exist for oocytes and sperm with hypomethylated and hypermethylated genomes observed for the former and latter respectively. Upon pronuclear fusion in the fertilized zygote, paternal sperm derived DNA undergoes rapid global demethylation resulting in significantly reduced methylation levels which diminish throughout embryonic development until blastocyst formation at E3.5 in mice. In contrast, oocyte derived maternal DNA remains hypomethylated during this time and is maintained until E3.5. Global DNA methylation levels are lowest in E3.5 inner cell mass embryonic stem cells and re-methylation commences from E5.5 during uterine wall implantation until E7.5 at which highest global levels are observed (Smith *et al*, 2012). Methylation intensities exhibited during and post-gastrulation signify somatic ground state of DNA methylation which globally remains constant, only fluctuating in discrete cellular conditions (Smith *et al*, 2012).

Methylation erasure at fertilization does not occur uniformly throughout sperm and oocyte genomes, with specific sequences being refractory to demethylation whilst others yield to it (Barton *et al*, 2001 & Santos *et al*, 2002). Interestingly, high resolution 100 base-pair tile methylome profiling of mouse gametes, zygotes and early embryos reveal methylation loss in sperm and oocyte is distinct with hypermethylated *LINE1* retrotransposon elements and to a lesser extent, viral Intracisternal A protein Particle repeats (IAP) in sperm undergoing reduction to intermediate levels in zygotes (Lane *et al*, 2003). In contrast, these *LINE1* sequences displayed low methylation in oocytes and were maintained after transformation to zygotes.

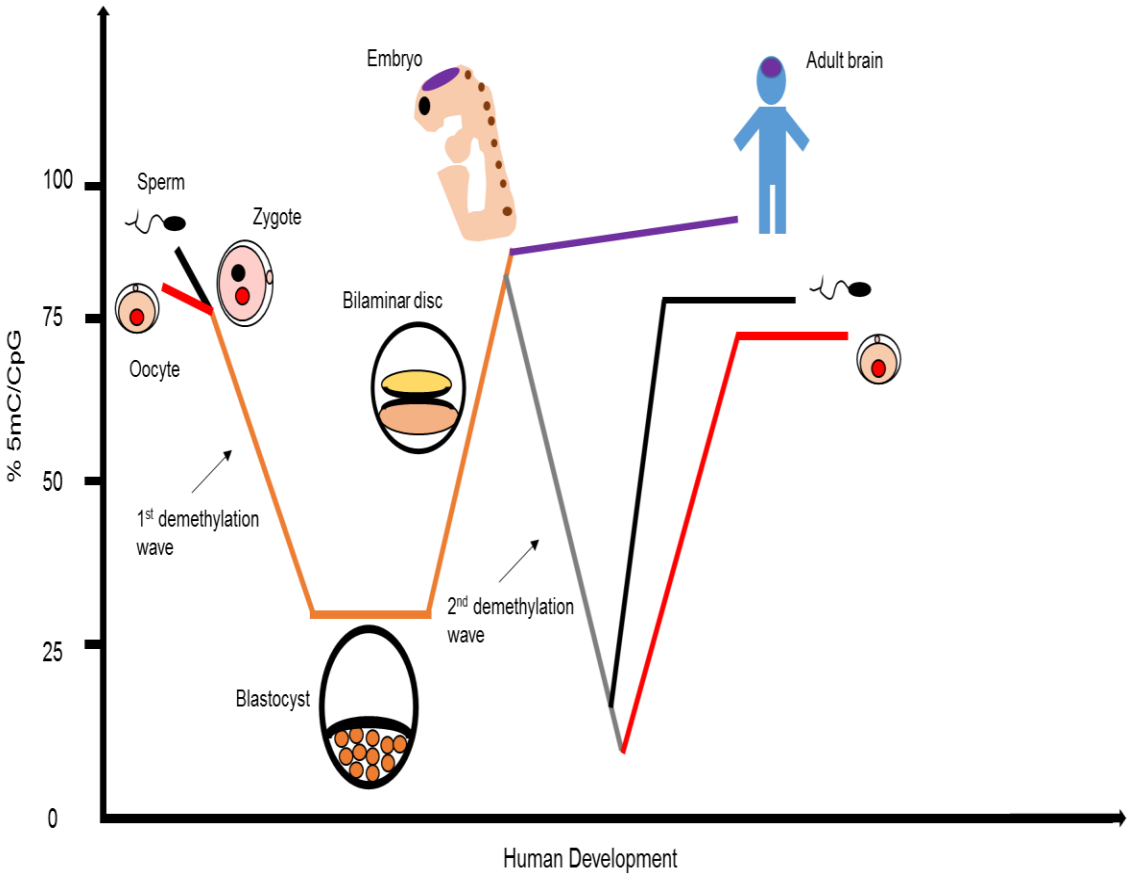


Figure 1.3 Patterns of DNA methylation erasure and replenishment during human development. DNA demethylation occurs bi-phasically during human development, firstly by rapid active and passively in sperm (black line) and oocyte (red line) genomes, ensuing during pre-implantation stages of embryogenesis (orange line). The second wave is (grey line) observed during primordial germ cell migration from the pluripotent epiblast to the genital ridges where gamete methylation is re-established to form sperm and oocyte signatures. Post blastocyst implantation into the uterine wall, foetal DNA methylation signatures are reconfigured during development especially in the brain (purple line), with the somatic genome methylation landscape gradually increasing in scale and magnitude as development ensues towards adulthood. (Original figure).

Intergenic sequences on paternal sperm-derived, and housekeeping gene promoter sequences coinciding with CGIs in oocytes display dramatic fluctuation in methylation intensities throughout embryogenesis (predominantly at pro nuclear fusion- zygote transition stage) and represent differentially methylated regions (DMR). These DMRs are significantly more enriched (10 fold) in sperm than oocytes (Smith *et al*, 2012). Epiblast derived primordial germ cells which number at approximately 35-40 at E7.5 display global 5mC DNA signatures commensurate with the somatic ground state but undergo two waves of demethylation during migration from the epiblast to the genital ridge during E11.5-12.5. Arrival at genital ridge at E13.5 coincides with erasure of random X-chromosome inactivation of maternal or paternal alleles

during migration at E12.5 and establishment of the germline ground state of methylation (Gkountela *et al*, 2015).

While the requirement of DNA demethylation in erasure of repressive transcriptional modifications and thus permitting gene expression is evident, understanding the mechanics and machinery of demethylation is still in its infancy.

1.2 DNA Demethylation Mechanisms

1.2.1 Passive Demethylation

As previously mentioned, erasure of established methylation marks is achieved by passive and active demethylation mechanisms (Wu & Zhang, 2010). At fertilization, oocyte DNA undergoes gradual passive demethylation facilitating removal of 5mC as cell division ensues until formation of embryonic stem cells in the blastocyst inner cell mass (Lane *et al*, 2003). Although genome-wide DNA demethylation occurs, regions of the maternal genome decorated by histone 3 lysine 9 dimethylation (H3K9me2) are protected from 5mC erasure owing to binding by PGC7/Stella to these transcriptionally repressed chromatin regions (Nakamura *et al*, 2012). As DNMT1 nuclear localisation is absent from zygote to blastocyst stages in mammals (Cardoso & Leonhardt, 1999), maintenance methylation on hemi-methylated nascent synthesized DNA stands cannot be performed. Owing to the semi-conservative nature of DNA replication (Gruenbaum *et al*, 1981; Gruenbaum *et al*, 1982), this results in passive dilution of 5mC marks from oocyte DNA (Lane *et al*, 2003). In contrast, 5mC is rapidly lost from paternal sperm derived DNA

at fertilization in the zygote and is heavily depleted by the two-cell stage embryo. This highlights a DNA replication- independent active mechanism of demethylation as one cell division is insufficient for reduction of 5mC levels to this extent. (Wu & Zhang, 2010).

1.2.2 Deamination Pathway

Active DNA demethylation has been documented to occur via an activation induced deamination pathway in which activation-induced cytidine deaminase (AID) catalysed deamination of 5mC to thymine creates thymine: guanine mismatch (Arnott, Bond & Chandrasekaran, 1980; Neuberger & Rada, 2007). Repair of the mismatch, facilitated by thymine DNA glycosylase (TDG) excise single mismatched nucleotides by contorting DNA 45° exposing the deoxyribose-nitrogenous base glycosidic bond for nuleophilic attack. Consequently, an abasic site is produced (Jacobs & Schar, 2012). Apurinic / apyrimadinic endonuclease (APE) family members e.g. APE1 hydrolyse deoxyribose-phophate phospho-diester linkages at the abasic site resulting in a single strand nick. DNA polymerase ß incorporates the correct cytosine base opposite guanine and DNA ligase lll synthesizes the phospho-diester bond thus restoring T:G mismatch to C:G, effectively erasing the local 5mC presence (Ramiro & Barreto, 2015).

1.2.3 Active Iterative Oxidation

The Ten-Eleven Translocation (TET) proteins belong to a conserved family of dioxygenase proteins capable of iterative oxidation of methyl

groups on CpG residues (Ito *et al*, 2011). These Tet proteins, homologous to J-Base binding proteins (JBP) discovered in *Trapanosome bruceii*, recognise and bind modified DNA bases e.g. 5mC and oxygenate these residues to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine and 5-carboxylcytosine (5caC) (Tahiliani *et al*, 2009). Tet protein facilitated oxidation of 5mC results in stepwise conformation change from methyl to hydroxyl, carbonyl and carboxylate configurations, of which 5fC and 5caC are compatible with TDG recognition and base-excision repair (BER) machinery (He *et al*, 2011). Consequently, TDG/BER mediated oxygenated base removal and polymerase ß – ligase III restoration to unmodified cytosine (figure 1.4) culminates in implicit expulsion of 5mC DNA occupancy (He *et al*, 2011). Interestingly, 5caC is specifically recognised and excised by TDG but 5hmC is not. Likewise, 5hmC can be removed by Uracil DNA Glycosylase (UDG) and single stranded selected monofunctional uracil glycosylase (SMUG1) but not 5caC (He *et al*, 2011).

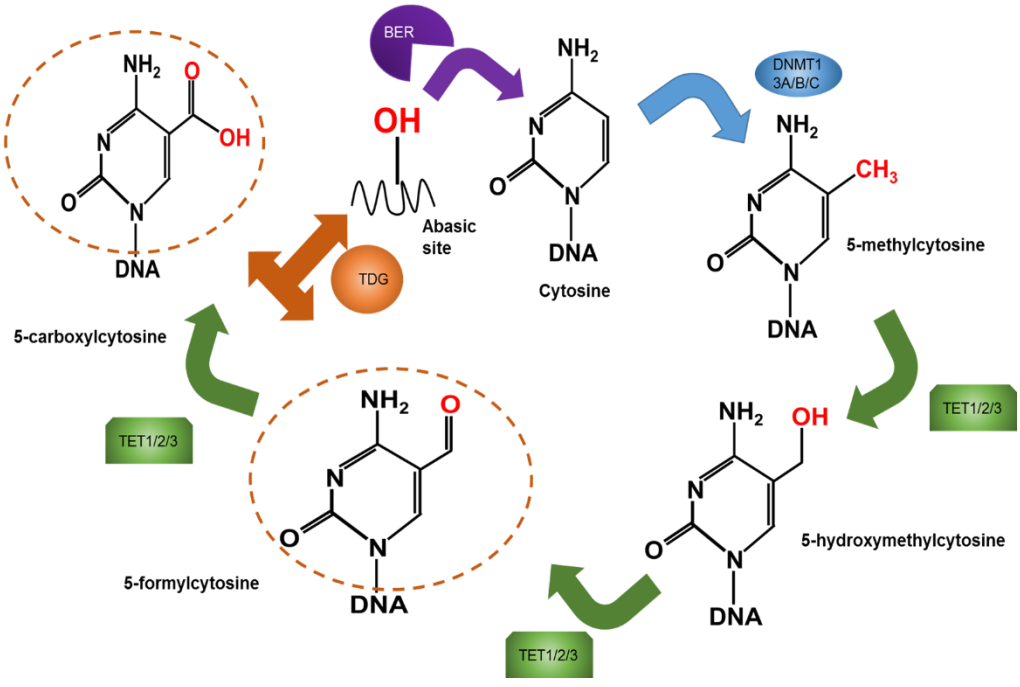


Figure 1.4 Mechanism of iterative oxidation mediated active DNA demethylation. Unmodified cytosine is methylated to 5mC either by *de novo* DNMT3A/B/C or by maintenance methylase DNMT1. 5mC is actively modified via oxidation by TET proteins to stable demethylation intermediate 5hmC which may undergo further oxidation to 5fC and subsequently 5caC (brown dashed circles) and removed by TDG in an active removal step and BER machinery, or passively diluted (PD) via semi-conservative DNA replication. TDG mediated 5fC and or 5caC removal creates an abasic site capable of being restored by an unmodified cytosine residue, thus completing the cycle of methylation erasure. (Original figure).

1.2.4 Mechanism of Oxidation

Evolution of active DNA demethylation machinery may have occurred in response to absence of mammalian enzymes capable of directly cleaving covalent 5'carbon- methyl carbon bond to release the methyl group possibly due to stabilisation of the methyl group by the delocalised electron cloud within the cytosine aromatic ring (Xu *et al*, 2011; Franchini *et al*, 2012). The Tet proteins share strong sequence homology to Mixed Lineage Leukaemia (MLL) fusion protein in the DNA binding domain and to 2-oxoglutarate (2OG) dependent dioxygenase superfamily members (Tahiliani *et al*, 2009). Full-length TET1 & 3 proteins are 2039 and 1668 amino acids in long respectively and both possess cysteine rich CXXC DNA binding domain at their N terminal (Wu & Zhang, 2011). Divergent evolutionary forces in pre-vertebrae ancestors resulted in chromosomal inversion of TET 2 CXXC generating a CXXC4/IDAX domain situated as

an upstream *cis* element transcribed in the reverse orientation to TET2 open reading frame (Ko *et al*, 2013). CXXC and CXXC4 DNA binding domains possess a cysteine rich interface ordered around two zinc ions (Zn^{2+}) in a tetrahedral coordinate configuration enabling binding to high density CpG sequences and discernment of unmethylated/methylated residues (Wu & Zhang, 2011). All three TET proteins share sequence homology of highly conserved cysteine rich regions contiguous with a downstream double stranded beta helical (DSBH, also known as 'jelly-roll motif') motif at the C terminal (Ito *et al*, 2011). These regions comprise the 2-oxoglutarate/ α-ketoglutarate-dependent dioxygenase domain which harbour two non-haem catalytic iron (Fe^{2+}) centres in octahedral coordinate configuration (Schofield & Ratcliffe, 2004). The ferrous ions in contact with two histidine and one aspartate residue create the stabilising dioxygenase interface required for nucleophilic attack of carboxylate group on aspartate towards the co-substrate 2-oxoglutarate 1-carbon carboxylate and 2-carbon ketone groups. Dioxygenase catalysis produces oxidation of CpG 5mC to 5hmC, oxidation of Fe^{2+} to Fe^{3+}, reduction of aspartate carboxylate group and decarboxylation of 2-oxoglutatrate to succinate and liberation of carbon dioxide (CO_2) (Schofield & Ratcliffe, 2004). Interestingly, 2-hydroxyglutarate an onco-metabolite correlated with aberrant DNA methylation in multiple cancers acts as a competitive inhibitor of TET protein activity (Wu & Zhang, 2011). In attempts to characterize TET protein iterative oxidation activity, Tahiliana *et al*, and Ito *et al* employed Thin Layer Chromatography (TLC) and Mass Spectrometry (MS) to observe migration rates of analytes

following 20mer DNA oligonucleotide digestion with *HpaII*, *MspI* and *TaqI* restriction endonucleases capable of cleaving unmodified cytosines; unmodified, 5mC and 5hmC; and all modified methyl-cytosine derivatives respectively. Under combinatorial buffer conditions facilitating separation and migration of C, 5mC, 5hmC, 5fC and 5caC following 20mer CpG oxidation, *TaqI* cleavage, Calf Intestinal Phosphatase 5' phosphate removal and polynucleotide kinase catalysed 5' radioactive [γ^{-32}P]ATP end labelling, analytes migration rates were examined (Tahiliana *et al*, 2009 & Ito *et al*, 2011). Resolution of all oxidised forms was achieved and confirmed by species analysis using mass spectrometry (Tahiliana *et al*, 2009 & Ito *et al*, 2011). *In vitro* analysis of wild type TET 1 vs catalytically attenuated H1671Y and D1673A TET1 mutants revealed increased oxidised methyl cytosine species presence on 2D thin layer chromatography gels correlated with decreased 5mC presence. This was not observed in mutants as 5mC remained constant with undetectable oxidised forms (Tahiliana *et al*, 2009 & Ito *et al*, 2011). Incubation of TET1 with 20mer oligos harbouring 5hmC and 5fC modifications generated migrating spots on 2D TLC identified by MS to be 5caC, implicating 5hmC and 5fC as viable substrates for oxidation in the demethylation process although oxidation efficiency is variable according to substrate (Tahiliana *et al*, 2009 & Ito *et al*, 2011).

1.2.5 TET Protein Expression and Distribution

During development, TET proteins display differential expression patterns as embryogenesis commences. Rapid demethylation of

paternal pronuclear genome at fertilization and prior to pronuclear fusion is attributable to elevated TET3 expression. High TET3 presence correlates with increasing 5hmC concentrations and reduced 5mC levels which are all detectable by immunohistochemistry (Iqbal *et al*, 2011). In contrast to paternal pronuclear 5hmC presence, maternal pronuclear DNA is absent of 5hmC and enriched for 5mC indicating lack of TET protein facilitated active demethylation but consistent with slow replication-dependent passive demethylation mechanisms (Lee *et al*, 2014). Maternal pronuclear DNA achieves immunity to TET3 regulated demethylation via expression of PGC7/Stella complex (Nakamura *et al*, 2007). Interestingly, embryonic requirement for TET3 nuclear expression is curtailed past the 2-cell stage with expression rapidly diminishing. Remarkably, 5hmC concentrations at this point remain elevated suggesting redundancy in active demethylation protein activity and delayed TDG/BER induced removal of oxidised methyl cytosine marks (Iqbal *et al*, 2011). Indeed TET1 expression is significantly up-regulated from the morula to blastocyst stage where it peaks, commensurate with highest observed 5hmC, 5fC, 5caC levels and lowest 5mC levels (Wossidlo *et al*, 2011; Branco *et al*, 2012; Kohli & Zhang, 2013). Suggestive of roles in establishing and maintaining pluripotency, *TET3* and *TET1* conditional knockouts in paternal pronuclei and embryonic stem cells respectively fail to oxidise 5mC present at paternal pronuclear genome *Oct4* and *Nanog* promoters and *LINE1* retrotransposon sequences. Moreover, *TET1* knockout significantly perturbs NANOG expression resulting in departure from pluripotency towards extra-

embryonic lineage differentiation (Gu *et al*, 2011; Dawlaty *et al*, 2013). TET2 expression displays parity with TET1 in mouse embryonic stem cells and *TET1/2* double knockout results in complete abrogation of 5hmC presence culminating in developmental defects but viable offspring. Intriguingly, *TET1/2* double knockout embryonic stem cells remained pluripotent but displayed defective teratoma and chimeric embryo formation capabilities (Dawlaty *et al*, 2013). In primordial germ cells, TET3 has been demonstrated to actively induce erasure of 5mC from randomly imprinted and inactive paternal/maternal X-chromosomes as cells migrate from epiblast to genital ridges (Clark, 2015).

1.2.6 Distribution of Oxi-methyl-Cytosine Modification

A more informative indication of TET protein activity in understanding their regulation is studying the distribution and abundance of oxi-methyl-cytosine marks. Nagae *et al* studied the 5mC distribution at -1.5kb upstream and 1.0kb downstream of promoter sequences across 7 different tissues from primary patient samples. Using 27,578 double stranded cDNA hybridised to 500bp regions in and around promoters, microarrays, methylome profiling and bisulphite sequencing were performed to calculate the coverage of methylated CpGs at these regions. Their results revealed significant 5mC presence at CpG poor promoters in contrast to unmethylated CpG rich regions, the latter being characteristic of CGIs (Herman *et al*, 1996). Across tissues, highest tissue specific methylation was observed in brain, testis and blood whilst testis and oral mucosa exhibited greatest hypomethylation, indicating a

differential methylation pattern occurring at tissue specific promoters (Nagae *et al*, 2011). Owing to the inability of bisulphite sequencing to differentiate between 5mC and 5hmC, Song *et al* implemented β-glucosyltransferase facilitated transfer of glucosyl moieties to 5hmC hydroxyl group generating β-glucosyl-5-hydroxymethylcytosine in the presence of uridine diphosphoglucose .β-glucosyl-5-hydroxymethylcytosine was subsequently tagged with biotin at the β-glucosyl-N3 position. Analysis of chemical addition on 11mer oligo substrates by matrix assisted laser desorption time of flight (MALDI-TOF) mass spectrometry followed by deep sequencing confirmed modification and detection of β-glucosyl-5-hydroxymethylcytosine (Song *et al*, 2011). Dot blot assays employing detection via biotin-conjugated horseradish peroxidase tags exhibited enrichment of 5hmC in male and female postnatal mice cerebellum with highest levels detected at 2.5 and 10 months in males and females (Song *et al*, 2011). 5hmC was readily detectable in mouse embryonic stem cells as expected but unexpectedly present at similar levels in adult mouse neural stem cells at 0.047 and 0.036% of genomic nucleotides respectively (Song *et al*, 2011). HeLa and HEK293 cells exhibited negligible 5hmC levels. Deep sequencing analysis revealed enrichment of 5hmC within gene bodies and within and surrounding (-875 upstream and +100-200 downstream) promoter regions consistent with observations by Nagae *et al* (Song *et al*, 2011). Through utilising sensitive anti-5mC and anti-5hmC antibodies in selective methyl/hydroxymethyl DNA immunoprecipitation (meDIP/hmeDIP), and subsequent high throughput sequencing, Ficz *et*

al demonstrated high 5hmC occupancy at promoters, exons and LINE-1 retrotransposon sequences which correlated with reduced 5mC levels at these locations in mouse embryonic stem cells. Inversely, greatest 5mC enrichment was observed at repetitive satellite sequences where 5hmC presence was limited (Ficz *et al*, 2011).

Studies performed on human frontal lobe brain tissue reveal highly significant 5hmC enrichment, four fold higher than in mouse embryonic stem cells. In concordance with previous observations, high throughput sequencing of frontal lobe tissue illustrated majority (55-59%) of 5hmC signal localised at low density CpG promoter regions, 35-38% within gene bodies and approximately 6% occupancy at intergenic regions. In contrast, 5mC was enriched at intergenic regions (25-26%) and higher within gene bodies (52-5%) but reduced (22-24%) at promoter sequences (Jin *et al*, 2011a). These studies indicate abundance of 5hmC in embryonic stem cells and somatic tissue particularly the brain however investigations on 5fC and 5caC distributions are limited.

Validation of TET2 propensity to synthesize 5caC from 5mC substrates on DNA oligos *in vitro* was confirmed via thin layer chromatography and quantitative mass spectrometry which identified 5caC species with mass/ charge ratio (m/z) of 270.0721 (He *et al*, 2012). In mouse embryonic stem cells, immunohistochemical probing demonstrates presence of 5caC and mass spectrometry analysis quantified its magnitude to $3.0x10^{-6}$, significantly reduced compared to 5hmC ($1.3x10^{-3}$ per 10^6 cytosines). 5fC was quantified slightly higher than 5caC at $2.0x10^{-5}$ per 10^6 cytosines. The markedly diminished presence of 5caC in embryonic stem

cells can be enriched by *TDG* knockout which elevated 5caC concentrations nine fold relative to wild type cells however 5hmC levels remained constant (Ito *et al*, 2012). These observations suggest 5caC depletion may occur as a result of swift TDG facilitated excision of 5caC after synthesis and not solely as a consequence of diminished TET protein catalytic activity (Ito *et al*, 2011). Shen *et al* discovered a similar trend in which 5caC and additionally 5fC concentrations significantly increase (8 fold for 5caC and over 5 fold for 5fC) in TDG knockdown compared to wildtype mouse embryonic stem cells (Shen *et al*, 2013).

Whist 5hmC concentrations are elevated in mouse and human cerebral cortex, brain stem, spinal cord and cerebellum (Globisch *et al*, 2010; Nagae *et al*, 2011) especially Purkinje neurons (Kriaucionis & Heintz, 2009), 5fC and 5caC are undetectable in these tissues. However, recent evidence illustrates strong albeit heterogeneous 5caC presence compared to homogeneously enriched 5hmC staining in 12.5 days post coitum (dpc) embryonic mouse brains. Interestingly, 5hmC signal in adult mouse brain tissue was heterogeneous with high concentrations in positive cells which exhibited co-localisation of 5hmC with post mitotic neuronal marker *NeuN*, suggesting possible lineage or developmental maturity specific role of 5hmC (Abakir *et al*, 2016).

The genomic distributions of 5fC and 5caC in embryonic stem cells are localised at enhancers, promoters, 5' and 3' Untranslated Regions (UTR) but predominate at intragenic exons and introns and, in contrast to 5hmC, cluster at intergenic and repetitive satellite sequences, consistent with observations by Shen *et al*, 2013 (Lu et *al*, 2015). 5fC and 5caC

differ in their proportions of genomic DNA occupancy with greater presence at exons for 5fC and greater presence at introns for 5caC. Both 5fC and 5caC enrichment correlated with poor (15-25%) 5mC and (10-15%)5hmC levels at these regions suggesting terminal oxidation via TET proteins occurs preferentially at low CpG density regions (Lu *et al*, 2015).

1.2.7 Functional significance of oxi-methylcytosine derivatives

The scarcity of 5fC and 5caC marks relative to 5hmC genomic abundance as previously mention initially supported the notion of 5fC and 5caC existing as terminal oxidised marks and simply intermediates in the active DNA demethylation process (Tahaliani *et al*, 2009, Ito *et al*, 2010, He *et al*, 2011, Ito *et al*, 2011). However, examination of the mechanism of mouse TET2-mediated 5mC oxidation suggested a sequential process of oximC derivation of 5fC and 5caC in which single TET2 enzymes may generate these TDG-excisable modifications without releasing 5hmC intermediates (fig 1.5) (Crawford *et al*, 2016). Mass spectrometric examination of oximC abundance at timed intervals of TET2 catalysis of 5mC substrates on 27 nucleotide oligomers revealed no significant difference in proportion of 5hmC, 5fC and 5caC levels generated within the first minutes compared to conclusion of the assay 20 minutes later. These results provide an alternate possibility of oximC generation to the iterative model of 5fC and 5caC production contingent on successive TET enzyme recognition of free available 5hmC substrates within DNA recognition sequences (Crawford *et al*, 2016). This iterative model

implies 5fC and 5caC generation occurs only once stable 5hmC modifications are to be removed and thus 5hmC is enzymatically converted to a higher oxidised state facilitating TDG-mediated excision (Crawford *et al*, 2016). The sequential model, while revealing TET2 mechanisms of oximC generation, also suggests (owing to the propensity for immediate formation of terminal oximC marks) distinct functional roles for 5fC and 5caC which are both chemically stable at steady state conditions (He *et al*, 2011) beyond DNA demethylation intermediates (Crawford *et al*, 2016).

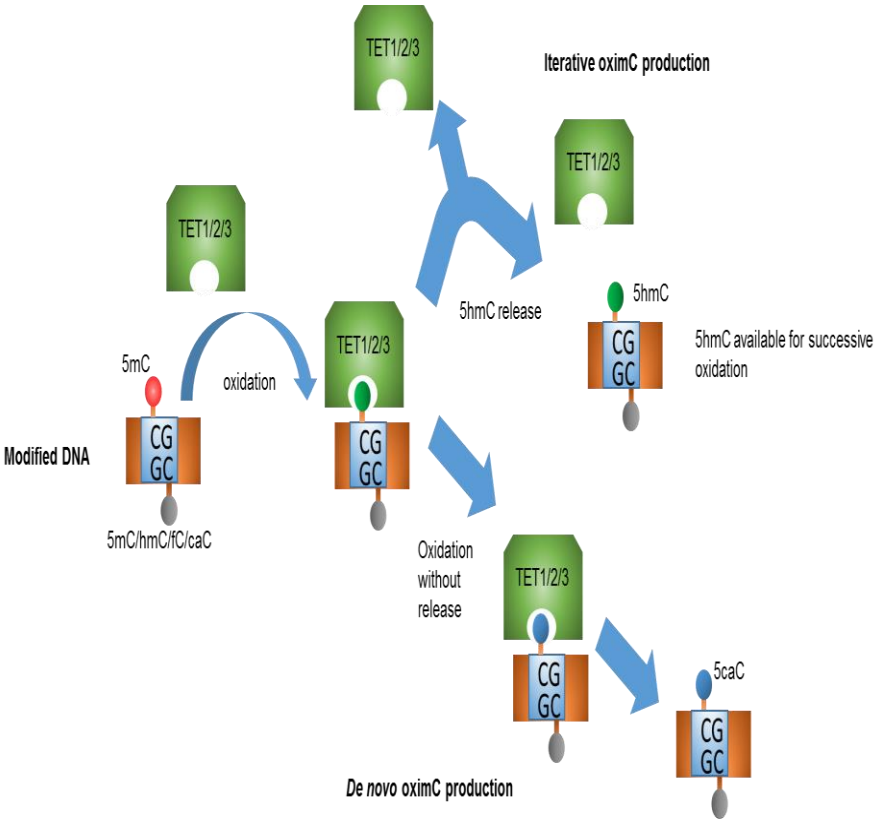


Figure 1.5 models of TET protein mediated 5mC oxidation. Catalysis of oximCs may follow an iterative mechanisms in which TET enzymes release 5hmC residues which are freely exposed for subsequent further catalysis by another TET protein to higher oxidation states. Alternatively, in a sequential *de novo* model, TET enzymes may catalyse the formation of 5fC and 5caC marks

without releasing 5hmC and thus terminal oximC marks from continued catalysis through lower oxidation states. (Adapted from Crawford *et al*, 2016).

Analysis of studies investigating 5hmC, 5fC and 5caC genomic distribution demonstrate overwhelming evidence of occupancy at transcriptionally relevant regions, strongly implicating these modifications in gene expression regulatory roles. Moreover, occupancy of these oxidised marks at these regions correlates with reduced 5mC levels implying erasure of the transcriptionally repressive modification facilitates opportunities for transcriptional activation. Indeed, promoters and exons harbouring 5hmC marks exhibited higher transcription levels than those with unmodified cytosines. Moreover, *TET1/2* knockout resulted in depletion of 5hmC and transcription silencing of pluripotency associated genes, culminating in mouse embryonic stem cell departure from stemness (Ficz *et al*, 2011; Ito *et al*, 2011; Jin *et al*, 2011a; Nagae *et al*, 2011; Song *et al*, 2011; Wossidlo *et al*, 2011; Branco *et al*, 2012; He *et al*, 2012; Ito *et al*, 2012, Dawlaty *et al*, 2013; Lee *et al*, 2013 & Lu *et al*, 2015). In *TDG* null mutant mouse embryonic stem cells, 5caC and 5fC accumulation was mapped by DIP and high throughput sequencing to reveal enrichment at TET1-bound gene sequences of *Sox17*, *Esrrb* and *Tcl1* which was not observed in wild type cells (Shen *et al*, 2013). Additionally, 5fC/5caC peaks were observed at active promoters harbouring RNA polymerase II, distal regulatory regions, active enhancers bound to p300 histone acetyltransferase and transcriptionally permissive histone marks H3K27ac and histone 3 lysine 4

monomethylation (H3K4me1), consistent with observations by Lu *et al*, 2015. Interestingly, this is in contrast to 5hmC patterns which locate at poised enhancers (Wen *et al*, 2014). Curiously, 5caC in *TDG* null mutants is not observed at sequences bound to basal transcriptional machinery such as TATA-Binding Protein or transcription factor IIB (TFIIB). However TDG null mediated 5caC enrichment occurred at Oct4 and Sox2 bound promoters and Cohesin complex bound sequences (Shen *et al*, 2013).

Recent evidence implicates preferential location of 5hmC on sense strands of DNA within gene bodies correlating with higher gene expression compared to transcriptional repression correlating with 5mC locating on antisense strands which may indicate an 'earmarking' mechanism of maintaining active transcription bubbles of the elongation complex (Wen *et al*, 2014). Enrichment of 5hmC at exon-intron junctions may indicate demarcation of exons selected for silencing resulting in variant isoforms of transcripts. 5hmC congregation at exon-intron junction may elicit MeCP2 recruitment to these sites which in turn stimulates summoning of histone deacetylases to remove acetyl groups from H3K27ac histone tails at poised enhances. Consequently this may enable methylation of histone H3K27 by histone methyltransferases facilitating chromatin conformational change over exons to be silenced (Khare *et al*, 2012 & Wen *et al*, 2014).

The significantly high levels of 5hmC in the mammalian brain (10-20% of brain 5mC and 40% of 5mC in Purkinje neurons) predominate in terminally differentiated mature neurons (Globisch *et al*, 2010 & Munzel

et al, 2010). The importance of DNA methylation and its erasure either through mutation of DNA methyltransferases or possibly by demethylation mechanisms can be observed in Rett syndrome mice which exhibit arrhythmic diaphragmatic motions leading to perturbed breathing (Guy *et al*, 2001). Conditional knockout of *DNMT1* has been demonstrated to recapitulate the Rett phenotype (which usually develops due to *MeCP2* deleterious mutations) in neonatal mice which exhibit hypomethylation of *IAP* retro-element sequences which should be transcriptional repressed via 5mC presence (Hutnick *et al*, 2009). Transcriptional induction of these elements correlates with cortical neuron apoptosis and concomitantly, activation of neural stem cell astrogliogenesis, observable through increased early astrocyte marker glial fibrillary acidic protein (GFAP) expression at the cortical plate (Hutnick *et al*, 2009). The 5mC hypomethylation perturbs MeCP2 facilitated transcriptional repression via chromatin remodelling and this effect may be mimicked by 5hmC over enrichment as negative correlations have been observed between 5hmC and MeCP2 (Szulwach *et al*, 2011). Interestingly, abrogation of MeCP2 activity to transcriptionally repress JAK-STAT associated gene and early astrocyte gene e.g. *GFAP* promoters impairs neurogenesis and pushes neural stem cells towards glial lineage differentiation (Martinwich *et al*, 2003 & Fan *et al*, 2005).

During development and differentiation, 5caC has been shown to accumulate as neural stem cells differentiate towards glial lineage but not towards neuronal lineages (Wheldon *et al*, 2014). Wheldon *et al*

examined mouse and human embryonic stem cells, developing mouse embryos and adult mouse brains for presence of 5caC. They discovered enrichment in both mouse and human embryonic stem cells and 5caC presence was observable in 6.5 dpc mouse embryo distal visceral endoderm and multiple issues in sagittal sections of 9.5 dpc embryos (Wheldon *et al*, 2014). Curiously, 5caC was only observed heterogeneously in the sub-ventricular zone of adult brain lateral ventricles and not in the hippocampal dentate gyrus or neocortex. Staining of embryonic mouse forebrains revealed a transient accumulation of 5caC between embryonic days 11.5-13.5 where 5caC absence was observed at 11.5, peaked at 12.5 and diminished by 13.5 dpc (Wheldon *et al*, 2014). This effect was recapitulated by adult neural stem cell induction to glial lineages whereby 5caC absence was observed on day 0, peaked at day 1 and diminished at day 3. However the same pattern was not observable for adult neural stem cells differentiating to neuronal lineages (fig 1.6). Analysis of 5caC physical occupancy at promoters of neuronal genes *Galanin*, *Neuroglobin* and Adenosine Triphosphate Synthase Peripheral stalk unit D (*Atp5h*) by DNA Immunoprecipitation (DIP) revealed 5caC enrichment even though its presence was not notably detectable in adult neural stem to neuronal differentiation by immunohistochemistry (Wheldon *et al*, 2014). Upon *TDG* knockdown by siRNA in adult neural stem cells differentiating to glial lineages, 5caC was 4-5 fold enriched and DIP confirmed this by illustrating 5caC occupancy at promoters of glial markers e.g. *GFAP*, *Olig1*, *Olig2*, *GLAST* and *Msi1,* to name a few (fig 1.7) (Wheldon *et al*,

2014). This data suggests a lineage specifying role of 5caC in differentiation and development by possibly marking promoters of genes perhaps already in a poised transcriptional state in progenitor cells thus enabling rapid expression of lineage specific genes (Wheldon *et al*, 2014).

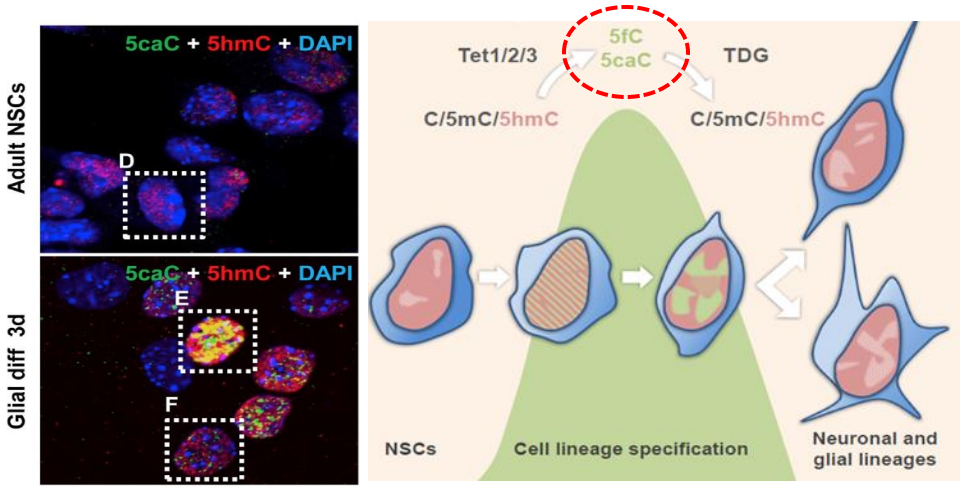


Figure 1.6 Transient accumulation of 5caC as adult mouse neural stem cells differentiate towards glial lineages. 5caC is undetectable in mouse neural stem cells (D) but is enriched heterogeneously (E and F) during glial differentiation. 5fC and 5caC levels peak (red dashed circle) during cell lineage specification. (Adapted from Wheldon *et al*, 2014).

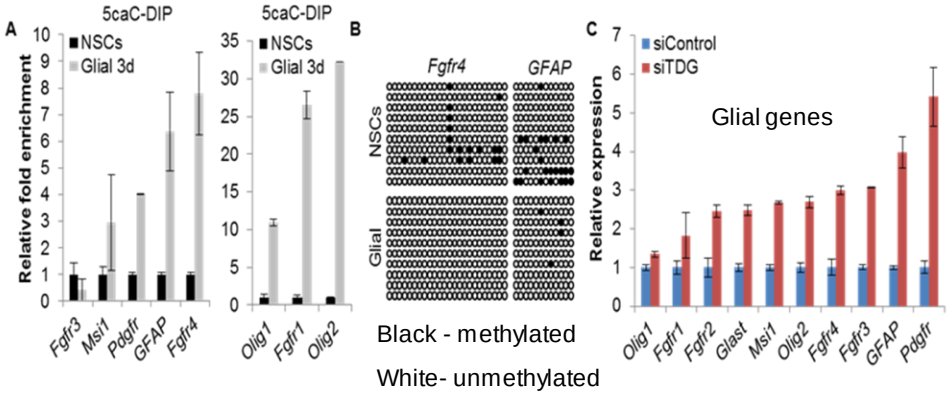


Figure 1.7 Physical 5caC occupancy on glial gene promoters correlates with their transcriptional up-regulation. (A) DNA immunoprecipitation (DIP)

of 5caC on neural stem cells versus differentiating progenitors. (B) Schematic of 5caC presence on glial promoters in adult neural stem cells vs differentiating progenitors. (C) QPCR analysis of glial gene expression (Adapted from Wheldon *et al*, 2014).

Oxidised forms of 5mC may also function through modulation of transcription factor binding to consensus sequences as observed through the interactions of 5hmC, 5fC and 5caC with C/EBPβ basic helix-loop-helix (bHLH) transcription factor. C/EBPβ recognises and binds to the 8 nucleotide consensus sequence 5'-TTGCGTCA-3' but this binding is diminished if the central CG dinucleotide is oxidised from 5mC to 5hmC. However the C/EBPβ-DNA complex interaction is strengthened by the presence of 5fC and 5caC (Sayeed *et al*, 2015). Interestingly, circular dichroism and thermodenaturation studies demonstrate 5caC enhances binding of bHLH transcription factors to DNA most strikingly, the Tcf3/Ascl1 heterodimer binding to 5'-CGCAG GTG-3' E-Box motif was 10-fold higher, an interaction specific to 5caC, but not other oxidised 5mC derivatives (Golla *et al*, 2014). An explanation for these observations may reside in circular dichroism and x-ray crystallography studies which implicate oxidised forms of 5mC i.e. 5fC and 5caC but not 5hmC as producing a B to Z-DNA transition effect (Wang *et al*, 2013; Lercher *et al*, 2014). B-DNA possesses a more relaxed, stable and open conformation whilst Z-DNA has a relatively compressed conformation with backbone phosphate groups in close proximity to another contributing to Z-DNA's energetic instability (Wang *et al*, 2013; Lercher *et al*, 2014). Z-DNA is generally located at high CpG clustered promoter

sequences and is associated with transcriptional activity as RNA polymerase is able to translocate efficiently along Z-DNA (Wang *et al*, 2013; Lercher *et al*, 2014). 5fC and 5caC may preferentially facilitate this B to Z-DNA transition owing to the presence of carbonyl oxygen in their structures which, being highly electronegative, induce intramolecular hydrogen bonding between them and hydrogen atoms of the amine group (Wang *et al*, 2013; Lercher *et al*, 2014). These observations highlight an intriguing dimension of epigenetic modulation of transcription by 5mC derivatives in which a single transcription factor binding sequence can switch accessibility to specific transcription factors by presence or absence of oxi-methylcytosine modifications (Wang *et al*, 2013 & Lercher *et al*, 2014). 5fC and 5caC have also been demonstrated to influence RNA polymerase II activity by reducing the kinetic catalytic rate, increased pausing / stalling during elongation and more frequent RNA transcript cleavage and backtracking (Kellinger *et al*, 2012; Wang *et al*, 2015). Considering 5fC and 5caC can be located at poised and active enhancers, this effect of pausing, slowing and backtracking RNA polymerase II may mediate induction and binding of histone remodelling proteins e.g. p300 and downstream members of demethylation machinery e.g. TDG as well as 5caC specific protein readers (Kellinger *et al*, 2012; Wang *et al*, 2015).

1.2.8 Readers of Oxi-methylcytosine Derivatives

5hmC is the most stable oximC mark and present at magnitudes 10-1000-fold higher than those of 5fC or 5caC modifications, especially in

the mammalian brain (Kriaucionis & Heintz, 2009; Ito *et al*, 2011). Considering the limitations of TET-assisted bisulphite sequencing (TAB-Seq) in discriminating 5fC from 5caC, chemical protection labelling of 5hmC prior to formyl group reduction on 5fC was performed to characterise the distribution of 5fC within mouse ESC genomes (Song *et al*, 2011). 5hmC modifications were protected by reacting with β-glucosyltransferase to conjugate an unmodified glucose moiety to the hydroxyl group of 5hmC. 5fC was reduced via treatment with sodium boro-hydride ($NaBH_4$) to 5hmC which was subsequently reacted with azide-modified glucose group and finally biotin conjugated to facilitate affinity capture pull-down (Song *et al*, 2011). Immunoprecipitation, mass spectrometric validation and sequencing in comparison with previously bisulphite- and TAB-Sequenced genomes at comparable sequencing depth revealed common localisation of 5fC peaks with those of 5hmC. However, multiple 5fC unique peaks were observed at enhancers annotated by bivalent histone marks Histone 3 lysine 4 trimethylation (H3K4me3) and histone 3 lysine 27 trimethylation (H3K27me3), characteristic of poised chromatin (Song *et al*, 2011). Interestingly, the depletion of 5hmC at these 5fC enriched sites coupled with the association of poised chromatin with regulation of cell differentiation (Fisher *et al*, 2011; Xi *et al*, 2011; Kim *et al*, 2018) lends credence to the notion of oximC marks interacting with possible epigenetic reader proteins to facilitate chromatin conformational changes (Song *et al*, 2011). Indeed, the formyl group of 5fC in particular facilitates the interaction between amino acid residue amine groups within proteins to

form Schiff bases which anchor DNA binding proteins to specific DNA motifs (Song *et al*, 2011).

Identification of protein readers capable of recognising and binding to oximC marks in mouse ESCs, neuro-progenitor cells (NPCs) and adult mouse brains was achieved by implementing stable isotope labelling by amino acids in cell culture (SILAC) coupled with affinity capture pull-down and subsequent in-gel trypsinisation and liquid chromatography coupled tandem mass spectrometry (Spruijt *et al*, 2013). Briefly, this process entailed labelling of two identical cell cultures with either heavy or light isotopes of carbon e.g. ^{13}C vs ^{12}C or nitrogen ^{15}N vs ^{14}N prior to incubation with sepharose or agarose bead conjugated to streptavidin. 'Bait' molecules consisting of DNA oligonucleotides possessing either unmodified or modified (5mC/5hmC/5fC/5caC) cytosines conjugated to biotin tags were incubated in heavy or light isotope cell cultures to bind with prey proteins capable of binding to the unmodified or modified 'bait' DNA sequence (Spruijt *et al*, 2013). The bait bound proteins precipitated by pull-down on affinity columns, trypsinised in gels and resolved and subsequently identified by liquid chromatography and mass spectrometry respectively (Spruijt *et al*, 2013). While many protein readers capable of detecting 5mC (Kruppel-Like Factors 2,4 and 5, KLF2, KLF4, KLF5; Methyl CpG binding Protein 2, MECP2; Ubiquitin-like containing PHD and RING Finger domains 1, UHRF1; Methyl-Binding Proteins 1 and 5, MBD1, MBD5; Regulatory Factor X1, RFX1 and Zinc Finger Homeobox 3, ZFHX3) and 5hmC (Thymocyte protein 1, Thy28; WD Repeat containing protein 36, WDR36; Chromosome 3 Open

Reading Frame 37, C3orf37 and MECP2) were identified in mouse ESCs, the screen also revealed dynamic nature of some proteins e.g. KLF4 displaying weak affinity for unmodified cytosine substrates in mouse ESCs but exhibiting high affinity for 5mC marked bait substrates in neuro-progenitors, suggesting context sensitive expression and functional roles in cellular reprogramming (Spruijt *et al*, 2013). Interestingly, interrogation of proteins precipitated and identified in response to 5fC and 5caC modified bait substrates revealed TDG as the only protein common to both higher oxidised 5mC derivatives which was validated by GFP-TDG pull down on 5fC and 5caC modified substrates and by electro-mobility shift assays (EMSA) (Spruijt *et al*, 2013). Proteins discovered to bind uniquely to 5fC modified baits included P53 whilst 5caC modified ones interacted with DNMT1 and SWI/SNF-Related Matrix associated Actin dependent Regulator of Chromatin subfamily C 2 (SMARCC2) exclusively. Differentiation of mouse ESCs to NPCs demonstrated an enrichment of chromatin modifying proteins such as Lysine Demethylase Binding protein 2 (KDMB2), Mixed Lineage Leukemia protein (MLL) and their respective factors BCL6 Corepressor (BCOR) and RING1A/B as well Retinoblastoma Binding Protein 5/ ASH-Like protein (RBBP5/ ASH2L) were found to precipitate with unmodified cytosine bait in NPCs (Spruijt *et al*, 2013). Proteins including Kaiso, UHRF1 and MBD4 displayed affinity for both unmodified cytosine and 5mC whereas DNA glycosylases Nei Endonuclease VIII Like 1 and 3 (NEIL1 and NEIL3) exclusively bound 5hmC in NPCs. Curiously UHRF2 which was absent in mouse ESCs exhibited increasing expression upon

differentiation to NPCs and displayed affinity towards 5hmC modified baits (Spruijt et al, 2013). Adult mouse brain tissue depict the expression of proteins possessing preferential affinity for 5hmC modifications in contract to mouse ESCs and NPCs in which protein binding to unmodified or methylated residues was favoured. Brain enriched 5hmC-binding proteins included but was not limited to Pleiotrophin, Zinc Finger Protein 385B, Zinc finger and BTB domain containing 4, Nuclear Factor Erythroid 2-Like 1, F-Box protein 11, ASH1-Like, Thanatos associated domain-containing Apoptosis-associated Protein 1, Methyl CpG Binding Domain protein 2, Snf2 Related CREBBP Activator Protein, RNA Binding Motif Protein 6, SMAD Nuclear Interacting Protein 1, Phosposerine Phosphatase 1 and Hepatocyte growth factor Regulated tyrosine kinase Substrate (PTN, ZFP385B, ZBTB4, NFE2l1, FBXO11, ASH1l, THAP11, MBD2, SRCAP, RBM6, SNIP1, PSP1 and HRS respectively) (Spruijt *et al*, 2013). Taken together, these results indicate cytosine and cytosine modification specific affinities and binding preferences of numerous proteins. These proteins may exhibit temporally and spatially sensitive dynamics of expression which, depending on the nature of cytosine modification present at cognate DNA binding sequences, may influence distinct functional outputs (Spruijt *et al*, 2013). An unbiased whole transcription factor-modified DNA sequence screen known as digital affinity profiling via proximity ligation (DAPPL) was performed by creating Glutathione-S-transferase fusions of 1235 different transcription factors purified from yeast cells and affixed to glutathione beads (Song *et al*, 2019). The fixed transcription factors were

then fused to unique DNA oligo barcode identifiers via fusion with anchor oligos consisting of a maleimide 'click' chemistry group conjugated to a reverse oligo primer (primer 1) complementary to a primer sequence on the barcode oligo (Song *et al*, 2019). Utilisation of bead conjugated Klenow DNA polymerase facilitated double strand synthesis of the conjugated anchor oligo which, during denaturation and renaturation with the mixed in barcode oligo, enabled hybridisation of the anchor oligo reverse primer sequence to barcode oligo complementary forward primer (primer 1) sequence, thereby creating a joint transcription factor-DNA oligonucleotide complex (Song *et al*, 2019). The inclusion of a *BsaI* restriction site on the 3' terminal of the barcode oligo was crucial for proximity ligation facilitated the joining of the transcription factor-fused barcode to one of $1.0x10^{12}$ randomly symmetrically or hemi-modified 20mer- DNA library sequences flanked a 5' *BsaI* site and a 3' sequencing primer (primer 2) (Song *et al*, 2019). Ligation of *BsaI*-digested libraries and barcode oligos, contingent on library sequence binding to Glutathione bead captured transcription factors facilitated the identification of bound transcription factor to unmodified or modified DNA sequences via powerful next generation sequencing of DNA library-barcode oligo annealed constructs following transcription factor separation (Song *et al*, 2019). Consequently, this technique identified 44 5mC, 97 5hmC, 84 5fC and 107 5caC symmetrical modification binding transcription factors whilst 99 5mC, 103 5hmC, 118 5fC and 139 5caC hemi-modification recognising transcription factors were also discovered (Song *et al*, 2019). Although considerable overlap between transcription

factors recognising multiple distinct oximCs in both symmetrical and hemi-modified configurations, 1 5mC, 27 5hmC, 16 5fC and 42 5caC symmetrical and 7 5mC, 9 5hmC, 17 5fC and 35 5caC hemi-modified uniquely binding transcription factors were identified (Song *et al*, 2019). Depending on the modification, configuration and spacing, 5mC/5hmC/5fC or 5aC presence augmented transcription factor association with a particular DNA template by either increasing, decreasing or modifying transcription factor affinity and or specificity, as discovered through binding kinetics and affinity analysis (Song *et al*, 2019). 5caC DNA occupancy significantly increased Homeobox -contain 1 (HMBOX1) binding affinity to template sequences compared to all other oximCs or unmodified cytosine, a feat observed for and common to all Meis homeobox 1/2/3 (MEIS1/2/3) homologues binding to symmetrical and hemi-modified 5caC configurations. In contrast, 5hmC and 5fC significantly reduced MEIS1/2/3 homologue affinity for the same DNA sequence (Song *et al*, 2019). All cytosine modifications reduced Ovo-Like zinc finger 2 (OVOL2) binding affinity relative to unmodified cytosine on the same DNA sequence. Remarkably, 5caC presence augmented the binding affinities of HMBOX1 and Protein C-ETS1 (ETS1) transcription factors for DNA sequences significantly different from their consensus cognate binding motifs. These findings were corroborated by EMSA and ChIP-Seq experiments (Song *et al*, 2019). Taken together, these results indicated that oximC modifications may increase transcription factor binding kinetics thereby improving rates of gene expression while additionally preferentially specifying activation or

repression of certain target genes (Song *et al*, 2019). These oxidised derivatives may have indeed evolved as method of competitively selecting one transcription factor over another which share cognate sequence motif recognition e.g. Wilms' Tumour protein 1 (WT1) and Early Growth Response protein 1 (EGR1) (Hashimoto *et al*, 2014). Indeed, Hashimoto and colleagues described WT1 as putative oximC 'reader' owing to its preferential binding affinity to 5caC over other oximC derivatives, a phenomenon not shared by EGR1 (Hashimoto *et al*, 2014; Hashimoto *et al*, 2016).

1.3 Epigenetic Aberrations in Brain Tumours

Historically, brain tumour classification, as published by the World Health Organisation (WHO) has been reliant on clinical behaviours such as cellular proliferative ability and invasiveness or circumscription of tumours which indicated the probability of resection (Zulch, 1979). Subsequent updates to the classification occurred in 1993, 2000 and 2007 in which histological examination of tissue to assess and score indices such as cellularity, nuclear atypia, necrosis, micro-vascularisation and endothelial cell presence were included to stratify tumours based on grade (I-IV) of biological severity to predict patient prognosis and clinical outcome (Kleihues, Burger & Scheithauer, 1993; Sallinen *et al*, 2000; Thurnher, 2007). Distinctions in cellular proliferative ability differentiated lower grade I and II tumours such as pilocytic astrocytoma exhibiting low proliferation from high grade III-IV highly proliferative tumours such as glioblastoma (Sallinen *et al*, 2000). The

presence of necrosis and micro-vascularisation categorised tumours into grades III-IV however accurate detection of these morphological features was largely dependent on clinician experience and therefore diagnoses were often subjective (Coons *et al*, 1997; Sallinen *et al*, 2000). Consequently, immunohistochemical detection coupled with *in silico* analysis aimed at standardising measures of necrosis and cellular proliferation detection (Sallinen *et al*, 2000). More recently, and owing to the limitations of microscopic tumour morphology examination, the WHO published new guidelines which included molecular parameters of tumour grading such as genetic mutations and chromosomal abnormalities (Louis *et al*, 2016). These genetic indices have, in the instance of glioblastoma multiforme, enabled distinction between secondary pro-neural and primary classical subtypes which display different mutation profiles e.g. *IDH* gene mutational status (Louis *et al*, 2016). Moreover, epigenetic aberrations have also gained traction as a classifying parameter with DNA methylation landscapes of tumour genomes exhibiting disparate profiles capable of delineating primary from secondary glioblastoma subtypes with the overall aim of refining tumour diagnosis and patient treatment modalities (Capper *et al*, 2018).

1.3.1 Perturbed DNA Methylome Landscapes Are Often Characteristic of Malignancy

While histopathological and molecular signalling pathway analysis have been well established in cancer research (Louis *et al*, 2007), the epigenetic landscape of tumours is only beginning to be unravelled

(Modena *et al*, 2006). The transcriptionally repressive DNA modification 5mC has been well characterized and its presence at gene promoters correlates with H3K27me3 repressive chromatin marks (Fuks, 2005). Methylation patterns must be carefully regulated to prevent the onset of malignancy (fig 1.8) (Fuks, 2005). Typical tumour suppressor genes subject to promoter methylation mediated repression in cancers include Cyclin Dependent kinase Inhibitor 2A (*CDKN2A*), Tumour Protein 53 (*TP53*), Adenomatous Polyposis Coli (*APC*), Breast Cancer Type 1 susceptibility protein (*BRCA1*) and Multi homologue 1 (*hMLH1*) to name a few (Robertson & Jones, 1998; Esteller *et al*, 2000). Multiple neoplasms including breast, prostate, renal, colon, both non-small cell and small cell lung cancer and glioma cell lines display aberrant promoter CpG methylation of tumour suppressor *CDKN2*/p16 (Herman *et al*, 1995; Merlo *et al*, 1995). The p16 protein inhibits cyclin-dependent kinases (CDKs) 4 and 6 from complexing with cyclin D during G1 phase of the cell cycle. Consequently, Retinoblastoma susceptibility protein, pRb remains unphosphorylated, preventing cellular transition from G1 to S phase and ultimately cell cycle arrest at G1/S checkpoint (Herman *et al*, 1995; Merlo *et al*, 1995). Aberrant methylation of *CDKN2* promoter represses *CDKN2* transcriptional activation and hence p16 expression, thus preventing G1/S checkpoint arrest (Herman *et al*, 1995; Merlo *et al*, 1995). Therefore any accumulated deleterious mutation may 'escape' screening prior to irreversible entry into S phase possibly resulting in tumorigenesis (Herman *et al*, 1995; Merlo *et al*, 1995).

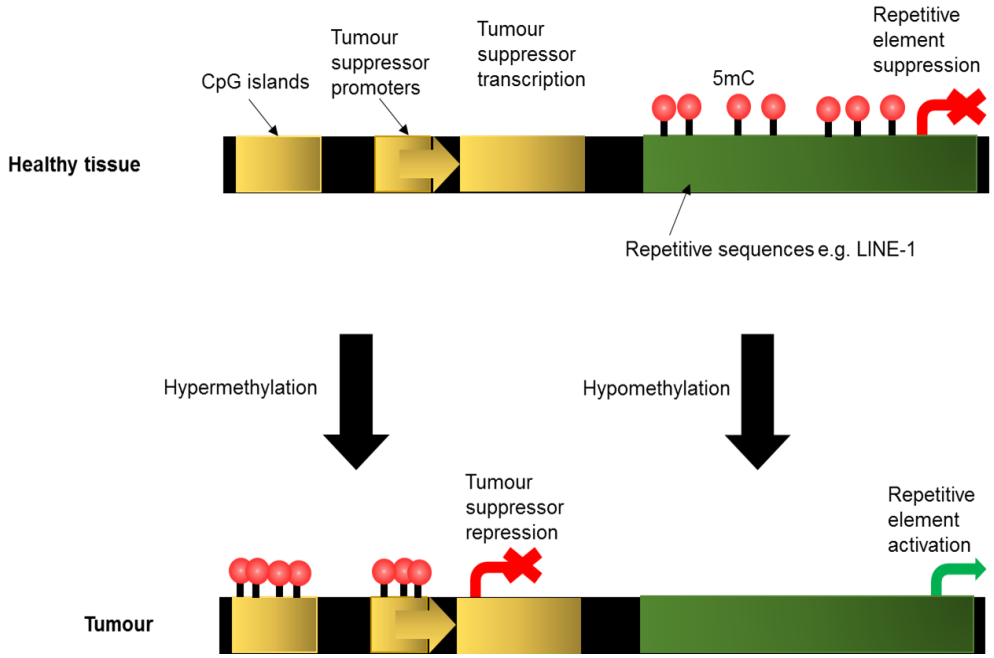


Figure 1.8 Aberrant DNA methylation landscapes common in cancer. Normal healthy tissues possess unmethylated CpG island sequences (yellow boxes) typically coinciding with house-keeping gene or tumour suppressor promoters. Mutagenic repetitive elements (green boxes) such as retrotransposons are hypermethylated (red circles) to prevent their propagation throughout the genome. Malignant tissues possess a reversed epigenetic profile in which CGIs which may coincide with tumour suppressor genes become silenced by hypermethylation and repetitive elements are activated by hypomethylation. (Adapted from Pfeifer, 2018).

1.3.2 Multiple Factors May Influence Adult Glioblastoma Multiform Methylome

Glioblastoma Multiforme (GBM) is the most prevalent and lethal adult brain tumours and associated with poor prognosis (15 months survival following diagnosis) even after surgical intervention, chemo- and radiotherapy and Temozolomide regimen (Parsons *et al,* 2008). Temozolomide is frequently prescribed to patients suffering from GBM tumours possessing methylguanine methyltransferase (*MGMT*) promoter hypermethylation (Martinez & Esteller, 2009). The *MGMT* gene encodes O^6-methylguanine methyltransferase, an enzyme capable of repairing alkylating adduct damage targeted to the O^6 position on guanine residues wrought by alkylating chemo-therapeutic agents (Hegi *et al*, 2005). Hypermethylation of the *MGMT* promoter silences *MGMT* expression, preventing BER-mediated removal of alkylation adducts at O^6 positions on guanine residues, the accumulation of which, results in DNA base mismatch and sensitises GBM cells to undergo apoptosis (Zhang, Stevens & Bradshaw, 2012). GBM tumours are aggressive astrocytomas which in 75% of cases do not exhibit mutations in *IDH1/2* genes or the CIMP prevalent in low grade gliomas (Turcan *et al*, 2012). However, the glioma CIMP (G-CIMP) and *MGMT* hypermethylation phenotypes represent two (pro-neural and classical respectively) of four distinct GBM transcriptional subtypes characterised by comparative genome hybridisation, RNA microarray and RNA-Seq analysis (Verhaak *et al*, 2010). The four subtypes include pro-neural, characterised by over-expression, deletion or mutation of Isocitrate Dehydrogenase 1 (*IDH1*),

Platelet Derived Growth Factor Receptor Alpha (*PDGFRA*), *TP53*, Oligodendrocyte transcription factor 2 (*OLIG2*), *CDKN1A*, Phosphatidylinositol-4,5-bisphosphate 3 Kinase (*PIK3CA*), Sex Determining Region Y Box 2 (*SOX2*), Transcription Factor 4 (*TCF4*), Dublecortin X (*DCX*), Delta-Like Ligand 3 (*DLL3*) and Achaete-Scute homologue 1 (*ASCL1*); neural, characterised predominately by Neurofilament Lght (*NEFL*), Synaptotagmin 1 (*SYT1*), Gamma Aminobutyric Acid Receptor Alpha 1 (*GABRA1*) and Potassium Cotransporter A5 (*SLC12A5*); classical, which possessed over-expression of NESTIN (*NES*), GLI family zinc finger protein 2 (*GLI2*), Smoothened (*SMO*), Growth Arrest-Specific protein 1 (*GAS1*), NOTCH receptor 3 (*NOTCH3*), Jagged Canonical Notch Ligand 1 (*JAG1*) and Epidermal Growth Factor Receptor (*EGFR*) and *EGFR vIII* mutation, wild type *TP53*, homozygous deletion of *CDKN2A* and mutations in Retinoblastoma protein 1 (*RB1*), *CDK4* and Cyclin D1 (*CCND1*); and mesenchymal, which exhibited high expression of Neurofibromatosis 1 (*NF1*), Phosphatase and Tensin homologue 1 (*PTEN*), Chitinase 3-like protein 1 (*CHI3L1*), Cluster of Differentiation 44 (*CD44*), MER proto-oncogene Tyrosine Kinase (*MERTK*), Tumour necrosis factor Receptor type 1 Associated Death Domain (*TRADD)*, RELB Proto-oncogene necrosis factor kappa Beta (*RELB*) and *TNFRSF1A* (Verhaak *et al*, 2010).

The transcriptional profile subtype of glioblastoma contingent on mutational status of specific genes may influence the treatment modality administered which consequently may significantly affect the tumour

epigenetic landscape (Mazor *et al*, 2016). This paradigm is exemplified by studies in which low grade *IDH1* mutant gliomas biopsied from patients were compared with recurrent secondary glioblastoma biopsies from the same patient following resection and temozolomide therapy approximately 11 years later (Johnson *et al*, 2014). Whole exome sequencing of tumour samples from 23 patients revealed a spectrum of genetic divergence between primary glioma and recurrent GBM with 4 patients possessing recurrent tumours sharing 75% of mutations with the initial glioma while 3 patients displayed secondary tumours with less than 25% of mutations in common with the primary low grade glioma (Johnson *et al*, 2014). These cases typified linear and branching models of tumour clonal evolution and the 16 remaining patients within the cohort possessed tumours represented varying degrees of both models at play (Johnson *et al*, 2014). Interestingly, all cohort members possessed primary gliomas with *IDH1* mutations but only 54% of patient recurrent tumours shared mutations (*TP53*, ATRX Chromatin Re-modeller *ATRX* and SWI/SNF Related, Matrix Associated, Actin Dependent Regulator of Chromatin, Subfamily A, Member 4 *SMARCA4*) with the primary glioma. This is suggestive of powerful tumour evolutionary mechanisms contributing to intra-tumour heterogeneity but also may be confounded by sampling errors during biopsies (Johnson *et al*, 2014). Considering the frequency with which secondary recurrent glioblastoma originates from *IDH1* mutant low grade glioma subject to Temozolomide therapy, the alkylating agent itself may be a driving force of tumour evolution in minority of cases (Johnson *et al*, 2014).

In light of the association of *IDH1* mutations and the resulting G-CIMP profile with the pro-neural transcriptional subtype common to secondary glioblastoma (Turcan *et al*, 2012; Mansouri *et al*, 2017), investigations pertaining to the recapitulation of the primary glioma methylome signature in the secondary GBM tumour have been explored (Mazor *et al*, 2016). Methylome analysis utilising Illumina Infinium 450K methylation bead arrays and validated by Methylight Assay profiled 1536 CpG sites representing 807 genes in low grade gliomas, their pair-matched recurrent secondary anaplastic astrocytoma or glioblastoma, *de novo* primary glioblastoma and control healthy region brain tissue (Laffaire *et al*, 2011). Profiling results identified hypermethylation of low grade glioma CpG shores relative to control genomes and hypomethylation of primary GBM CpGs at the same loci relative to healthy brain genomes (Laffaire *et al*, 2011). Hierarchical clustering of array data delineated methylation signatures into three groups with healthy control in one group, *de novo* glioblastoma in the second and low grade glioma, anaplastic astrocytoma and secondary glioblastoma in the third (Laffaire *et al*, 2011). The members of the third group were linked by *IDH1* mutations, *MGMT* promoter methylation and overall hypermethylated phenotype suggesting a shared evolutionary early *IDH1* mutant cell of origin which by linear clonal evolution culminated in a pro-neural glioblastoma transcriptional subtype (Laffaire *et al*, 2011). Analysis of 450K methylation data combined with Affymetrix microarrays aimed at correlating the gene expression levels at differentially methylated targets

between low grade glioma and recurrent GBM revealed that Thyroid hormone Receptor Interactor Protein 6 (*TRIP6*), Left-Right Determination Growth Factor (*LEFTY*) and Glial Fibrillary Acidic Protein (*GFAP*) were hypomethylated in GBM relative to glioma and displayed high expression levels (Laffaire *et al*, 2011). In contrast, hypermethylated and down-regulated targets in GBM relative to glioma included oncogenes Fibroblast Growth Factor 3 (*FGF3*), KIT proto-oncogene receptor tyrosine kinase (*KIT*), Janus Kinase 3 (*JAK3*), MET tyrosine protein kinase (*MET*) and Transforming Growth Factor Beta 2 (*TGFB2*) and tumour suppressors Caveolin 2 (*CAV2*), Fatty Acid Binding Protein 3 (*FABP3*) and Transmembrane Protein with EGF Like And Two Follistatin Like Domains 1 (*TMEFF1*) (Laffaire *et al*, 2011). In contrast to these findings, similar studies investigating the methylome profiles between G-CIMP low grade gliomas and their recurrent secondary glioblastoma discovered an overall CpG hypomethylation profile with 1953 sites hypomethylated, 50 of which existed within a glioma G-CIMP context (Mazor *et al*, 2016). These hypomethylated sites correlated with active chromatin marks H3K4me3 and H3K27ac and correlated with increased gene expression in GBM relative to glioma samples (Mazor *et al*, 2016). Additionally, non G-CIMP contextualised genes exhibiting hypomethylation were over-represented by cell cycle related genes specifically in GBM. Take together these results may be indicative of a different tumour evolution model e.g. neutral evolution in which G-CIMP related mutations and targets may be selected against and hence under-represented in recurrent tumours (Mazor *et al*, 2016).

1.3.3 5hmC Genomic Distribution in Glioblastoma Multiforme

High-performance liquid chromatography (HPLC) analysis of 5mC distribution between different tissues detected highest levels in the brain (Kriaucionis & Heintz, 2009). Isotope labelled liquid chromatography-coupled mass spectrometry studies investigating 5mC levels in the adult human brain depicted significant differences in the distribution of this epigenetic mark (Kraus *et al*, 2012). Frontal and occipital lobes of the cerebral cortex scored highest for 5mC levels whilst frontal and occipital white matter tracts exhibited significantly lower 5mC presence (Kraus *et al*, 2012). Clues as to the nature of this observed disparity in 5mC yield between different organs and even distinct regions of the same tissue may be attributable to the relatively recent discovery of oxidised forms of 5mC and the mechanism governing their generation (Khare *et al*, 2012). In experiments examining the quantity and distribution of 5mC and 5hmC between healthy brain and brain tumour specimens, mass spectrometry-validated immunohistochemical analysis revealed no significant difference in 5mC levels between healthy brains and tissue matched tumours of varying grades and isolated from multiple locations (Kraus *et al*, 2012). Brain tumours representing all World Health Organisation (WHO) classifications (grades I-IV) were interrogated for the presence and magnitude of 5mC and 5hmC marks Kraus *et al*, 2012). The WHO tumour grades increase with disease severity as characterised by tumour cell proliferative index, nuclear abnormality, necrosis, micro-

vascularisation, invasiveness and anti-correlate with patient prognosis and survival rate (Louis *et al*, 2007). Tumours ranging from low to high grade glioma were interrogated for 5hmC quantification (Kraus *et al*, 2012). Overall, immunostaining illustrated heterogeneous 5hmC signatures with greater proportion of positive staining in lower (grade I) tumours (16.7%) relative to higher (grade IV) ones (1.43%). Highest 5hmC signal intensity (5hmC/dG ratio) captured via mass spectrometry was observed also in grade I tumours (0.22%) whereas grade IV tumours exhibited the lowest 5hmC signal intensity (0.078%). Frontal and occipital cortex regions and even their lower white matter tracts scored considerably higher 5hmC levels relative to all brain tumours analysed (Kraus *et al*, 2012). Restriction endonuclease enzyme-facilitated oligonucleotide probe array hybridisation mapping at single base resolution data corroborates earlier evidence for 5hmC distribution and magnitude within healthy human brains relative to organs such as liver, kidney, pancreas, and heart (Khare *et al*, 2012).

Bisulphite and oxi-bisulphite sequencing technologies aim to delineate precise genomic locations of 5mC and oxi-mC marks respectively. Bisulphite and oxidative bisulphite sequencing employing sodium bisulphite and potassium perruthenate chemistries respectively facilitate the discrimination of 5mC from unmodified cytosine and 5fC from 5hmC (Johnson *et al*, 2016). Sequencing of methylated and demethylation intermediate attached bases is possible at single base resolution and has consequently revealed the overall depletion of 5hmC from glioblastoma cells (Johnson *et al*, 2016). Examination of 30 glioblastoma cell lines and

primary tissues exposed distinct 5hmC occupancy at CpG oceans and shelves which are comprised of low density CpG frequency relative to CGI and shores and thus possess intermediate levels of 5mC enrichment. This highlights a DNA demethylation intermediate role for 5hmC as 5mC erasure from sequences augments transcriptional status of genes under regulation (Johnson *et al*, 2016). The CpG shelves which displayed highest 5hmC enrichment (75%) localised within a window of approximately 5 kilobases (Kb) upstream from transcriptional start sites coinciding with enhancer and super-enhancer territory (Johnson *et al*, 2016). These genomic features characterised by association with active chromatin elements such as histone 3 lysine 27 acetylation (H3K27Ac), positively correlated with transcription of pathogenicity related genes in the aggressive glioblastoma cell line U87MG (Johnson *et al*, 2016). In concurrence with previously mentioned studies highlighting the correlation of 5hmC levels depleting as tumour grade increases, Recursive Partition Mixture Model (RPMM) clustering analysis of 5hmC occupancy occurring genome wide in glioblastoma cells linked the cytosine modification to prolonged patient survival, an index associated with pro-neural subtypes of glioblastoma (Verhaak *et al*, 2010; Johnson *et al*, 2016). Despite 5hmC depletion within aggressive brain tumour grades, its presence within genomic features appears critical to glioblastoma pathogenicity with 5hmC marks identified in gene regulatory regions upstream of core glioblastoma signalling genes such as epidermal growth factor receptor vIII (EGFRvIII) and cyclin-dependent kinase 6 (CDK6) (Johnson *et al*, 2016).

More recently, combined analysis using the Illumina Infinium 450K methylation bead array, oxi-bisulphite sequencing and TET-assisted bisulphite sequencing (TAB-Seq) of healthy normal brains versus glioma samples attempted to characterise hydroxymethylation signatures prevalent during malignancy (Fernandez *et al*, 2018). In concordance with findings of other groups (Kraus *et al*, 2012; Turcan *et al*, 2012; Kondo *et al*, 2014; Kraus *et al*, 2015; Johnson *et al*, 2016), 5mC and 5hmC magnitudes were significantly diminished in glioma relative to healthy controls. Interrogation of 5mC and 5hmC levels at *LINE-1* and *AluYb8* retrotransposons and peri-centromeric repeats alpha satellites and NBL-2 sequences demonstrated 5mC and 5hmC presence at greater magnitudes in normal tissue relative to gliomas (Fernandez *et al*, 2018), which was consistent with previous literature (Jin *et al*, 2011). Computational methods utilising the cancer genome atlas (TCGA) array data from different groups combined with Fernandez *et al* 450K and oxi-bisulphite sequencing results together with 5hmC DIP-Seq data generated a map of 5mC and 5hmC distributions between healthy brain and glioma (Fernandez *et al*, 2018). Overall, 7423 CpGs were hypermethylated compared to 44 974 hypohydroxylated CpGs, 4627 of which overlapped with hyper 5mC loci. 5hmC distributions between glioma and healthy control cells anti-correlated in their genomic patters of enrichment. 5hmC reduced loci possessed 5mC hypermethylation in a non-canonical methylation context (Fernandez *et al*, 2018). Furthermore, this non-canonical hypermethylation (occurring at 4627

sites) was unexpectedly commensurate with the localisation of transcriptionally active histone marks such as H3K4me1 and H3K36me3 while anti-correlating with canonical 5mC repressive histone marks H3K27me3 and H3K9me3 (Fernandez *et al*, 2018). Considering the semi-conservative nature of DNA replication and the hemi-methylation mechanism of maintenance methylation by DNMT1, coupled with the increase in 5mC at both canonical and non-canonical loci, the global 5hmC diminishment observed in glioma may not necessarily be attributable to increased cellular proliferation of malignant cells but maybe be due to reduced *TET* gene expression and activity (Fernandez *et al*, 2018).

Indeed, *IDH1* and *IDH2* gene mutations common in low grade glioma and secondary pro-neural glioblastoma have a reduced catalytic ability to generate α-ketoglutarate (Xu *et al*, 2011; Duncan *et al*, 2012). Specifically in the case of *IDH1* R132H substitution mutation of arginine by histidine, confers a gain of function enzymatic activity to isocitrate dehydrogenase to generate 2-hydroxyglutarate instead of TET protein co-factor α-ketoglutarate, thereby reducing TET protein 5mC oxygenase activity resulting in global genomic hypermethylation (Xu *et al*, 2011; Duncan *et al*, 2012). Curiously, the hypermethylation profile of G-CIMP glioblastoma may itself result in transcriptional repression of *TET* gene promoters thereby perpetuating the global 5hmC depletion in *IDH1* mutant gliomas (Carella *et al*, 2019).

However, conflicting reports demonstrate increased 5hmC abundance at CpG loci in *IDH1* mutant glioma compared to *IDH1* wildtype tumours (Glowacka *et al*, 2018). While distribution of 5hmC in *IDH1* mutant gliomas and glioblastoma samples aligned with literature reports of enrichment at CpG shores, promoters and enhancers (Johnson *et al*, 2016; Fernandez *et al*, 2018), the differentially hydroxymethylated probe set newly identified 5hmC enrichment within gene bodies which correlated to their increased gene expression relative to the same genes in *IDH1* wildtype tumours (Glowacka *et al*, 2018). Interestingly, examination of top 1% differentially expressed genes with 5hmC localised to enhancer sequences annotated by H3K27ac presence and validated via H3K27ac ChIP-Seq illustrated unique gene pathways prevalent in *IDH1* mutant and wildtype tumours. *IDH1* mutant glioma 5hmC probes displayed enrichment in genes pertaining to apoptosis and transforming growth factor beta receptor (TGFRβ) signalling pathways whilst *IDH1* wildtype glioblastoma tumours exhibited 5hmC enrichment in pathways relating to WNT, HEDGEHOG, NOTCH, Extracellular Regulated Kinase (ERK) signalling as well as embryonic stem cell network genes (Glowacka *et al*, 2018). This may be indicative of a 5hmC signature overlap between pluripotent stem cells and *IDH1* wildtype grade IV glioblastoma cells (Glowacka *et al*, 2018).

1.3.4 5mC and 5hmC Distributions within Glioma Stem Cells

A hallmark of glioma stem cells is their ability to efflux chemotherapeutic agents during patient therapy following resection which is often consequent in tumour resistance to radio- and chemotherapy treatment modalities culminating in tumour recurrence (fig 1.9) (Lathia *et al*, 2015; Liebelt *et al*, 2016; Sattiraju, Sai & Mintz, 2017). The glioma stem cell hierarchy, in which stem cells reside at the apex, describes the nature of multipotent glioma stem cells to self-renew, recapitulate tumours from clonal density in an irradiated orthotopic xenograft species and generate progenitor cells of restricted plasticity but capable of differentiating towards neuronal, astrocytic and oligodendrocytes lineages (Lathia *et al*, 2015; Liebelt *et al*, 2016; Sattiraju, Sai & Mintz, 2017).

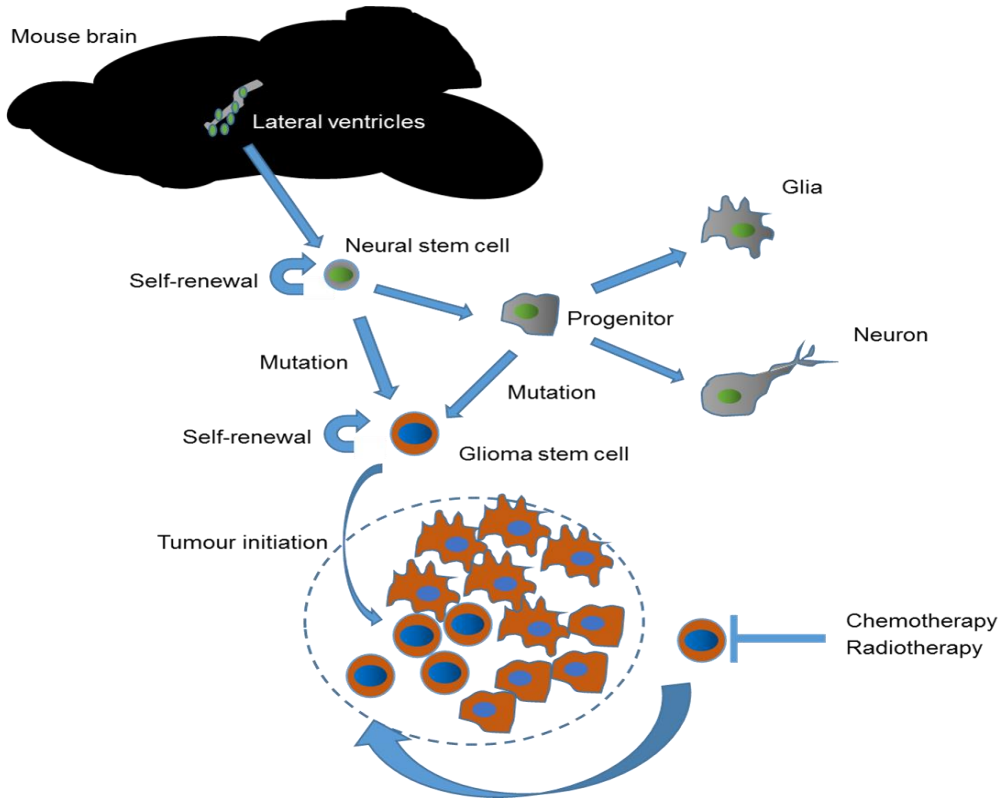


Figure 1.9 Glioma stem cells may be tumour initiating cells. The lateral ventricular zones of the mammalian neocortex represent neural stem cell niches producing multipotent neural stem cells and progenitors possessing multi-lineage differentiation capacity. Mutation in neither neural stem nor progenitor cells may instigate the formation of glioblastoma tumours exhibiting poorly differentiated cells with chemo- and radio-therapeutic resistance, contributing to poor patient prognosis. (Original figure).

Multiple attempts to define glioma stem cells by marker expression have identified discrete stem-like populations staining positively for Cluster of Differentiation 15 (CD15), epidermal growth factor receptor EGFR, CD44, L1 Cell Adhesion Molecule (L1CAM) and Cluster of Differentiation 133 (CD133, Prominin-1) (Hemmati *et al*, 2003; Liu *et al*, 2006; Bao *et al*, 2008; Son *et al*, 2009; Lathia *et al*, 2015). While adult neural stem cells strongly resemble glioma stem cells in multipotency, self-renewal marker expression such as SOX2, EGFR, NESTIN and NOTCH1, they do not exhibit cell surface glycoprotein CD133 (Reya *et al*, 2001; Hemmati *et al*, 2003; Tunici *et al*, 2004). Intriguingly, hierarchical clustering of a 500 gene array panel of 17 glioblastoma cancer stem cell lines delineated populations as being $CD133^{+}$ or $CD133^{-}$ which correlated with a foetal neural stem cell (type I) or adult neural stem cell (type II) profile (Lottaz *et al*, 2010). Type I glioma stem cells possessed properties such as adherent growth in culture, neurosphere formation and unresponsiveness to TGFβ signalling and segregated with a pro-neural glioblastoma transcriptional subtype whereas type II glioma stem

cells exhibited semi-adherent growth, lack of neurosphere formation, heavily express bone morphogenetic proteins (BMPs) and are responsive to TGFβ signalling, segregating with a mesenchymal glioblastoma transcriptional profile (Lottaz *et al*, 2010). Moreover, analysis of transcriptional signatures between type I and type II glioblastoma stem cells revealed no activation in cancer related tumour invasion or survival pathways but did highlight a communal reduction in differentiation pathway signalling and increased proliferative potential (Lottaz *et al*, 2010). This is suggestive of adult stem cells possibly activating a foetal neural stem cell transcriptional profile upon tumorigenesis in the brain stem cell niches (Lottaz *et al*, 2010).

The subventricular-ventricular zone of the lateral ventricles is a major site of mammalian neurogenesis which, together with sub-granular zone of the hippocampal dentate gyrus continue to generate neurons and glia into adulthood (Temple, 2001). During embryonic neurogenesis, ectodermal derived neuroepithelium cells differentiate into radial glial cells in the subventricular zone of the developing cortex (Merkle *et al*, 2004). Excitingly, immunohistochemical detection of 5hmC illustrated enrichment in the vertebrate subventricular zones but weakly in hippocampal dentate gyrus in murine models (Diotel *et al*, 2017). Moreover, *TET1* expression which commences at embryonic day 12 (E12) coincides with neurogenesis at the subventricular zone as do *TET2* and *TET3* (Diotel *et al*, 2017). All *TET*s exhibit diminished expression at E18 however expression resurges during postnatal day 7 (P7),

illustrating a bimodal pattern of *TET* expression (Diotel *et al*, 2017). Although *TET* mRNA and 5hmC presence was weakly detected in the hippocampus using fluorescent *in situ* hybridisation (FISH), both high 5hmC signal and *TET1* expression was detected in neurons in cortical regions at E16.5 however *TET1* and 5hmC signals anti-correlated with intensity of proliferating cell nuclear antigen (PCNA) staining, indicative of 5hmC dilution in actively replicating cells and enrichment in post-mitotic neurons (Diotel *et al*, 2017). Curiously, conditional over-expression of *IDH1* R132H mutant genes in neural stem cell marker positive cells in the adult mouse subventricular zone produced significantly greater numbers of proliferative neural stem cells and GFAP positive progenitors within the subventricular zone as well as increased neuroblast migration and invasion into the surrounding brain parenchyma culminating in increased brain mass in injected mice relative to controls (Bardella *et al*, 2016). Moreover, subventricular zone lesions appear in *IDH1* R132H mutant mice but not in controls. Furthermore, consistent with observations of increased subventricular neural stem cell and progenitors *in vivo*, *IDH1* R132H injected mice neural stem cell culture generated greater and large neurospheres *in vitro* with a greater proliferation rate (Bardella *et al*, 2016). Interestingly, global 5hmC and 5mC levels were decreased and increased respectively which correlated with a tumorigenesis profile and gene set enrichment analysis of subventricular zone cells indicated the activation of glioblastoma signalling pathway genes (Bardella *et al*, 2016). In summary, oxidised 5mC derivatives play a significant role in maintaining the somatic

genome landscape with the adult mammalian brain and aberrations of *TET* gene activity may reconfigure terminally differentiated post mitotic state to a stem cell phenotype, demarcated by aberrant oximC genomic profile (Bardella *et al*, 2016). Commensurate with these findings, IDH1 R132H mutations resulted in pluripotent stem cell gene activation and transcriptional permissive chromatin conformation in glioblastoma cells (Kim *et al*, 2019).

1.4 Wilm's Tumour 1 Biology

1.4.1 Developmental basis for aberrant expression of Wilms' Tumour protein 1 in cancers

Wilms' Tumour protein 1 (WT1) is a Kruppel-like transcription factor important for development of the mammalian embryonic kidney, urogenital system and epicardium (Roberts, 2005) that exhibits developmental and tissue specific expression patterns in mammals (Dressler, 2009).

Onset of WT1 expression in murine embryos commences at 9 days post coitum (dpc) where the protein is detectable in the parietal epithelium lining the coelomic cavity (Armstrong *et al*, 1992). At this stage *WT1* mRNA transcripts (Pritchard-Jones *et al*, 1990) localise to the intermediate mesenchyme/mesoderm (Pelletier *et al*, 1991; Vize *et al*, 1997). At 11 dpc, WT1 expression is detected in the metanephric mesenchyme, presumptive spinal cord motor neurons, gonads and

ureter (Armstrong *et al*, 1992). Significant increase in WT1 expression is observed in the metanephric mesenchyme and nephrogenic condensates at 12.5 dpc, whereas gonadal WT1 expression remains level at this stage (Armstrong *et al*, 1992). Between 13.5 and 15 dpc, developing epicardium, podocytes of the glomerulus, endothelial and stromal cells of the ovaries and uterus, Sertoli cells of the testis, retinal ganglia of the eye and ependymal cells of the fourth ventricle in the brain all possess positive immunostaining for WT1 (Sharma *et al*, 1992).

In a developmental context, during embryonic kidney formation, *WT1* functions as a proto-oncogene, inducing proliferation of metanephric mesenchyme pluripotent progenitor cells required to respond to inductive WNT9b signals from the invading uretic bud to undergo mesenchymal to epithelial transitioning (MET) into glomerular podocytes (Hohenstein & Hastie, 2006).

Approximately 5-20% of Wilms' Tumours cases exhibit somatic *WT1* gene mutations which tend to be sporadic bi-allelic aberrations occurring in blastemal progenitors (Kaneko *et al*, 2015). The inherited nature of mono-allelic germline *WT1* mutations or deletions occurring within chromosomal regions 11p13 predisposes individuals to Wilms' Tumour formation and account for approximately 5% of WT cases (Charlton *et al*, 2017). Nephrogenic rests (precursors to malignant Wilms' tumours) formed as a consequence of mono-allelic inactivation transformed into the eponymous renal malignancy upon secondary *WT1* allele

inactivation, a classic example of Knudson's two hit hypothesis (Ruteshouser *et al*, 2008; Royer-Pokora *et al*, 2010; Kaneko *et al*, 2015). The nephrogenic rests themselves occur more frequently as a consequence of somatic mono-allelic *WT1* mutation in sporadic Wilms' Tumours (90-95%) compared to familial Wilms' tumour germline *WT1* mutations (1-2%) (Cardosa *et al*, 2013).

Nephrogenic rests consist of primitive undifferentiated embryonic blastemal cells which are observed predominately in intralobar nephrogenic rests (ILNR) whereas undifferentiated mesenchymal stromal progenitors populate perilobar nephrogenic rests (PLNR) (Royer-Pokora *et al*, 2010; Charlton *et al*, 2017). *WT1* mutations are associated with ILNR which persist in early kidney development (10 dpc) and is commensurate with induction of WT1 expression in pluripotent mesonephros during this time (Wilm *et al*, 2016). The absence of WT1 mutations in PLNR may be due to a WT1-independent and restricted lineage differentiation potential of metanephric stromal progenitors (Wilm *et al*, 2016). Collectively, this is suggestive of a developmentally early WT1-dependent paradigm of patterning and differentiation of the intermediate mesoderm and metanephric mesenchyme (Pelletier *et al*, 1999; Royer-Pokora *et al*, 2010). However, the increased cellular proliferation rate resulting in the formation of neoplasms with *WT1* mutant genotype evident in 5-20 % of Wilms' Tumour cases implicates a tumour suppressor role for WT1, indicating both proto-oncogene and tumour suppressor functions for this transcription factor (Algar *et al*, 1996;

Yamagami *et al*, 1996; Inoue *et al*, 1998; Menke *et al*, 1998; Tsuboi *et al*, 1999; Loeb & Sukumar, 2002; Li *et al*, 2003; Tatsumi *et al*, 2008).

1.4.2 Impact of ectopic WT1 expression on tumorigenesis

WT1 wildtype mRNA has been detected in a range of malignancies including oesophageal, gastric, colorectal, pancreatic, biliary, lung, prostate, renal, breast, cervical, ovarian, endometrial, soft tissue and brain cancers, indicative of a possible oncogenic contribution of ectopic *WT1* expression in tumorigenesis (Nakatsuka *et al*, 2006). Studies performed on mouse lung cancer cells revealed a WT1-dependent mechanism of oncogenic KRAS induced proliferation (Vincent *et al*, 2010). Human and mouse lung cancer cells with ablated *WT1* reduced proliferation and triggered senescence, highlighting anti-apoptotic consequences of WT1 ectopic overexpression (Vincent *et al*, 2010). These results have been corroborated by immunohistochemistry by both N- and C-terminus-specific WT1 antibodies (Nakatsuka *et al*, 2006). Interestingly, although WT1 is not expressed in astrocytes and is found only in brain endothelium in healthy adult humans (Bourne *et al*, 2010), its elevated expression is common for brain tumours such as pilocytic astrocytoma (grade I), anaplastic astrocytoma (grade III) and glioblastoma multiforme (grade IV) (Izumoto *et al*, 2008; Nakahara *et al*, 2004; Chiba *et al*, 2010; Rauscher *et al*, 2014). Increasing *WT1* levels are commensurate with increasing tumour grade and associated with poor patient prognosis (Izumoto *et al*, 2008; Nakahara *et al*, 2004; Chiba *et al*, 2010; Rauscher *et al*, 2014). Increased incidence of high *WT1*

expression levels have been observed to correlate with severity of tumour grade of paediatric ependymoma with highest levels present in grade III anaplastic ependymoma (Yeung *et al*, 2013). *WT1* overexpression was observed in 98 % of glioblastoma primary cell samples and 83 % of anaplastic astrocytomas compared to 53 % of grade II oligodendroglioma and pilocytic astrocytomas (Schittenhelm *et al*, 2008). Short hairpin RNA (shRNA) molecules targeting *WT1* mRNA transiently silence *WT1* gene expression and reduce glioblastoma cell proliferation, viability and invasion ability suggesting an oncogenic role for WT1 in these malignancies as opposed to the tumour suppressor role this protein plays in Wilms' tumour (Schittenhelm *et al*, 2008; Clark *et al*, 2010; Kijima *et al*, 2014).

Closer inspection of the WT1 structure provides clues to its functional role in tumorigenesis. WT1 is encoded by the corresponding gene located on chromosome 11p13 that is composed of 10 exons with exons 5 and 9 undergoing alternate splicing to form functionally different isoforms (Call *et al*, 1990). The WT1 gene harbours two ATG start codons, with one at the conventional +1 site and the second residing between exons 1 and 2, translation from which, synthesizes a curtailed protein variant at the N terminal (Dallosso *et al*, 2004). Full length WT1 protein consists of an N terminal RNA recognition motif (RRM), dimerization domain, activation domain, alternatively spliced 17 amino acid auxiliary interaction domain encoded by exon 5 and 4 C-terminal Cysteine-Histidine (C2-H2) zinc finger domains (Bickmore *et al*, 1992).

Zinc fingers 1-4 are encoded by exons 7, 8, 9 and 10 respectively (Ladomery & Dellaire, 2002). Alternate splicing between zinc fingers 3 and 4 generates two isoforms either possessing or absent for lysine-threonine-serine ($KTS^{+/-}$) amino acid triplet sequence located between zinc fingers 3 and 4 (Clark *et al*, 2007). WT1 Zinc fingers 1 & 2 can interact with p53 facilitating its sequestration and stabilisation for subsequent ubiquitination and proteolysis (Maheshwaran *et al*, 1995). Stabilisation of p53 by WT1 prevents induction of pro-apoptotic pathways (Maheshwaran *et al*, 1995) Full length WT1 DNA binding domain is required for its physical interaction with p53 and as such, only +KTS and not -KTS isoforms can bind to wildtype p53 to stabilise the protein (Maheswara *et al*, 1995). Surprisingly, studies in osteosarcoma models indicated that the pro-apoptotic effect of p53 expression can be inhibited by WT1 +KTS isoforms but not –KTS isoforms (Maheswara *et al*, 1995). One explanation for this observation may include the WT1 facilitated inhibition of E6/E6AP mediated p53 ubiquitination and consequent proteolysis. Upon complexing with p53, WT1, may sterically prevent proteolytic degradation of p53 by occluding its ubiquitination sites from targeting by E6 ubiquitin ligase (Yamanouchi *et al*, 2014). In a murine model of ovarian cancer, transient overexpression of WT1 -17aa/-KTS isoform was linked to significant increases in cellular proliferation, migration and angiogenesis leading to a significant reduction of mouse survival time (Yamanouchi *et al*, 2014). These results were in stark contrast to over-expression of +KTS isoforms which did not significantly affect these parameters and the +17aa/-KTS splice variant which

induced apoptosis via suppression of *EGFR* transcription (Yamanouchi *et al*, 2014).

1.4.3 WT1 may modulate epigenetic signatures in tumours

Whilst *WT1* over-expression and null mutations have both been documented in cancer, implicating its dual properties as a proto-oncogene and tumour suppressor (Algar *et al*, 1996; Yamagami *et al*, 1996; Inoue *et al*, 1998; Menke *et al*, 1998; Tsuboi *et al*, 1999; Loeb & Sukumar, 2002; Li *et al*, 2003; Tatsumi *et al*, 2008), recent evidence has begun to demystify potential epigenetic regulatory functions of this protein in malignant neoplasms (Akpa *et al*, 2015). Thus, according to several studies, WT1 is able to prime metanephric mesenchyme progenitors for differentiation triggered by inductive WNT9b signals from the uretic bud via the repression of Enhancer of Zeste Homologue 2 (*EZH2*) transcription (Akpa *et al*, 2015). The consequent reduction of H3K27me3 enables transcription of beta-catenin (*CTNNB1*) in these progenitors and their subsequent nephrogenic differentiation (Akpa *et al*, 2015). In *Wt1*$^{+/-}$ mutant cells, progenitors cannot respond to WNT9b induction and thus proliferate into nephrogenic rests that transform into malignant tumours upon *wt1* ablation (Dressler, 2009; Akpa *et al*, 2015). According to a growing body of experimental evidence, oxidised forms of 5mC may play specific roles in epigenetic regulation of gene expression (Wu & Zhang, 2010; Seisenberger *et al*, 2012; Hackett *et al*, 2013; Smith & Meissner, 2013; Guo *et al*, 2014; Hu *et al*, 2014; Lewis *et al*, 2017) and serve as intermediates in the DNA demethylation process (Jones &

Liang, 2009; Tahiliani *et al*, 2009; He *et al*, 2011; Shen *et al*, 2013; Dawlaty *et al*, 2014) and, therefore, may be functionally involved in tumorigenesis (Dawson *et al*, 2012; Ehrlich & Lacey, 2013; Eleftheriou *et al*, 2015; Tian *et al*, 2016). Correspondingly, TET proteins may display tumour suppressor activity and their mutations, generating defective or non-functional proteins may result in tumorigenesis (Ko *et al*, 2010). Thus, TET2 protein has been implicated in the development of haematological malignancies such as Acute Myeloid Leukaemia (AML), Chronic Myelomonocytic Leukaemia (CMML) and Myelodysplastic Syndrome (MDS) (Tefferi *et al*, 2009). Interestingly, WT1 has been demonstrated to bind directly to TET2 and TET3 and recruit them to their potential target sequences in acute myeloid leukaemia models where mutation of WT1 resulting in its inactivation is accompanied with significant locus specific diminishing of 5hmC levels, a phenotype mimicked by mutations in the isocitrate dehydrogenase genes (*IDH1/2*) (Verhaak *et al*, 2010; Rampal *et al*, 2014; Kelly *et al*, 2017). In AML and pro-neural glioblastoma, a G-CIMP is common (Noushmehr *et al*, 2010; Turcan *et al*, 2012) and hypermethylation of non-housekeeping genes has been observed (Sturm *et al*, 2012; Hughes *et al*, 2013; Weisenberger, 2014). Substitution mutations such as arginine 132 replaced by histidine (R132H) mutate *IDH1,* erroneously converting isocitrate to 2-hydroxyglutarate instead of alpha-ketoglutarate, the canonical co-factor required by TET enzymes for oxidation of 5mC (Rampal *et al*, 2014). The consequent inhibition of TET enzyme activity correlates with significant 5mC hypermethylation and reduction in 5hmC

observed in Acute Myeloid Leukaemia cells (Rampal *et al*, 2014). Similar effects of hypomethylation and 5hmC reduction are characteristic of *Tet2* missense mutations that either result in its attenuated or abrogated catalytic ability (Konstandin *et al*, 2014).

Interestingly, the aberrant methylation signatures observed in AML can be recapitulated in cells with wildtype *IDH1&2* and *Tet2* but mutant *WT1* (Rampal *et al*, 2014). Correspondingly, a negative correlation of *IDH1&2* and *Tet2* mutations with mutations of *WT1* was reported for a cohort of 398 AML patients (Rampal *et al*, 2014). Moreover, a significant reduction in 5hmC levels was demonstrated for these *WT1* mutants compared to *IDH1 & 2*, *TET2* and *WT1* wildtype AML background controls (Rampal *et al*, 2014). Furthermore, next generation sequencing-based analysis revealed the existence of differentially methylated and hydroxymethylated regions between the wildtype AML controls and *WT1* mutants with the majority of 5hmC peaks localising to enhancers and distal regulatory regions whereas differential 5mC peaks between wildtype and mutants clustered around transcriptional start sites (Rampal *et al*, 2014). The 5hmC levels at differentially hydroxymethylated distal enhancers were lower in *WT1* AML mutants compared to controls and differentially methylated regions at transcriptional start sites were more greatly enriched for 5mC in *WT1* mutants (Rampal *et al*, 2014). This may suggest a transcriptional priming role for 5hmC in haematopoiesis which when perturbed by *WT1* mutations, may contribute to tumorigenesis (Ko *et al*, 2010; Yang *et al*, 2013). In glioblastoma multiforme, WT1 over-

expression may facilitate competition or co-operative binding with other transcription factors capable of recognising 5hmC occupancy at promoter and enhancer sequences (Takai *et al*, 2014). Intriguingly, DNA immunoprecipitation and subsequent sequencing (DIPSeq) investigations into the physical location of oxi-mC marks demonstrate possible interactions between cytosine modifications and protein complexes capable of 'reading' them and altering the transcriptional status of associated genes (Takai *et al*, 2014). Core GBM signalling pathway genes e.g. *BRAF*, *AKT*, *EGFR* & *CDK6* possessed 5hmC enrichment at their promoters and intergenic regions which positively correlated with Chromatin Target of PRMT1 (CHTOP) promoter presence and binding and consequent transcriptional activation of these genes (Takai *et al*, 2014). Excitingly, WT1 and CHTOP share binding affinity for the same promoter sequence, implicating WT1 as a strong contender for influencing sustained pathogenic transcriptional signalling in glioblastoma (Hashimoto *et al*, 2015).

In addition to its interaction with TET proteins, recent studies suggest that WT1 can also specifically bind certain oxi-mCs (Hashimoto *et al*, 2014; Hashimoto *et al*, 2016). WT1, a member of the Early Growth Response family (EGR) of transcription factors shares DNA binding consensus sequence 5'-GCG(T/G)GGGCG-3' (EGR-1 consensus) with its fellow family members EGR1 and Zif268 (Stoll *et al*, 2007). WT1 possesses oximC vs mC discrimination capabilities in contrast to 5mC vs unmodified cytosine recognition as demonstrated by EGR1 and Zif268 (Hashimoto

et al, 2014). Amongst oximC derivatives, WT1 displays preferential binding affinity for 5caC over 5hmC and 5fC. Additionally, preferential binding affinity for 5caC is enhanced by presence of asymmetric methylation on the same transcription factor "recognition strand" and by occupancy of 5mC at CpG sites on the complementary strand (Hashimoto *et al*, 2016). This is indicative of specific epigenetic configurations on DNA sequences which can be "read" by transcription factors such as WT1 (Hashimoto *et al*, 2016). Moreover, the markedly reduced binding affinity for 5caC and 5mC displayed by WT1 +KTS isoform may demonstrate a splice variant-specific requirement of certain gene promoters to undergo transcriptional activation (Hashimoto *et al*, 2016). A significant discrepancy is observed between zinc finger binding affinities to 5caC in WT1: zinc finger 2 possesses higher affinity whilst zinc finger 4 possesses lower, depicted by their dissociation constants (K_D) of approximately 50nm and greater than 12µm respectively (Hashimoto *et al*, 2014). Mapping of residues within WT1 protein structure revealed that arginine 366 (R366) with zinc finger 2 is able to contact with 3' guanine in the consensus sequence whilst Glutamine 369 (Q369) amine group hydrogen bonds with carboxylate groups on 5caC at the 5' CG dinucleotide. The carboxylate within 5caC intrabase hydrogen bonds with its own amine group and Q369 rotates approximately 70° from the position it occupied for bonding with 5mC. This rotation of Q369 together with stabilisation by S365 and water molecules gives WT1 specificity to interact with 5mC, 5hmC, 5fC and 5caC. In the +KTS isoform which possesses 5caC binding affinity, Q369

is involved in hydrogen bonding but the –KTS isoform has perturbed zinc finger 4 binding and thus possesses a reduced dissociation constant (Hashimoto *et al*, 2014).

WT1 is strongly implicated in neural progenitor differentiation to glial lineage. Considering Wheldon *et al*, 2014 findings on 5caC accumulation as adult neural stem cells differentiate towards glial lineages, we speculate that WT1 may be involved mechanistically in this accumulation, possibly by modulating TET function or by stabilising 5caC at promoter sequences. Ergo, this may suggest that oximC presence or absence on promoters and enhancers may perturb or enhance WT1-contingent gene expression, thus contributing WT1-implicated pathologies (Wheldon *et al*, 2014; Hashmoto *et al*, 2016). Considering the observations across the board that 5hmC levels are significantly reduced in malignancies compared to their healthy matched tissue (Figueroa *et al*, 2010; Ko *et al*, 2010; Jin *et al*, 2011), the presence of detectable 5caC in the brain tumours may seem rather unexpected and may depend on certain specific features of their methylation/demethylation machinery influenced by overexpression of *WT1* and or *TET2* / *IDH1 & 2* mutations in these cancers.

Collectively, these studies imply two potential modes of WT1 interplay with TET-dependent 5mC oxidation that may influence tumorigenesis. Thus, WT1 can either (1) serve as a binding partner of TET proteins participating in their recruitment to target sequences (Wang *et al*, 2015)

or (2) act as a specific “reader”- of certain oxi-mCs (most notably 5caC) (Hashimoto *et al*, 2014; Hashimoto *et al*, 2016). Presently available experimental evidence suggests that both these scenarios may take place during the pathogenesis of brain tumours, particularly glioblastoma multiforme (Schittenhelm *et al*, 2008; Eleftheriou *et al*, 2015). Considering the high mortality rate of this tumour and an immediate need for an efficient strategy for its therapy (Smoll *et al*, 2013; Thakkar *et al*, 2014; Ostrom *et al*, 2015), elucidating potential roles of WT1 and its potential interaction with TETs/oxi-mCs in glioblastomas should represent an important direction of the future research.

1.4.4 Hypothesis and Aims

Studies performed by Eleftheriou *et al*, 2015 examined low grade and high grade gliomas for the presence of 5caC as an indicator of elevated TET protein expression. Unexpectedly they discovered significant enrichment of 5caC in both low grade and high grade gliomas (Eleftheriou *et al*, 2015). Considering the observations across the board that 5hmC levels are significantly reduced in malignancies compared to their healthy matched tissue, 5caC presence is most unexpected (Jin *et al*, 2011).

Following unexpected discovery of 5caC enrichment in low and high grade glioma primary tumours (Eleftheriou *et al*, 2015), analysis of 5caC presence was conducted in high grade glioblastoma cell lines LN18,

LN229 and U87MG (fig 1.10). HeLa and HEK293 cells display absence and negligible 5caC levels respectively whilst all three glioblastoma cells lines demonstrate heterogeneous enrichment of 5caC and uniform presence of 5hmC. Interestingly, confocal microscopy signal intensity analysis revealed strikingly similar nuclear distribution of 5hmC and 5caC signatures to 13.5 dpc embryonic brain (fig 1.11) in concordance with studies performed by Abakir *et al*. Moreover, Quantitative PCR analysis of genes exhibiting 5hmC and more markedly, 5caC enrichment at their promoters were heavily up-regulated in U87MG cells compared to HeLa (fig 1.12). Notably, *WT1* expression was greater than ten- fold higher in U87MG cells relative to HeLa. This observation, coupled with Wheldon *et al* reports on 5caC enrichment observed at pro-glial lineage promoters in *TDG* knockdown mouse neural stem cells suggested a WT1- 5caC interaction influencing transcription of pro-glial lineage genes and a DNA demethylation dependent mechanism regulating neural stem cell differentiation.

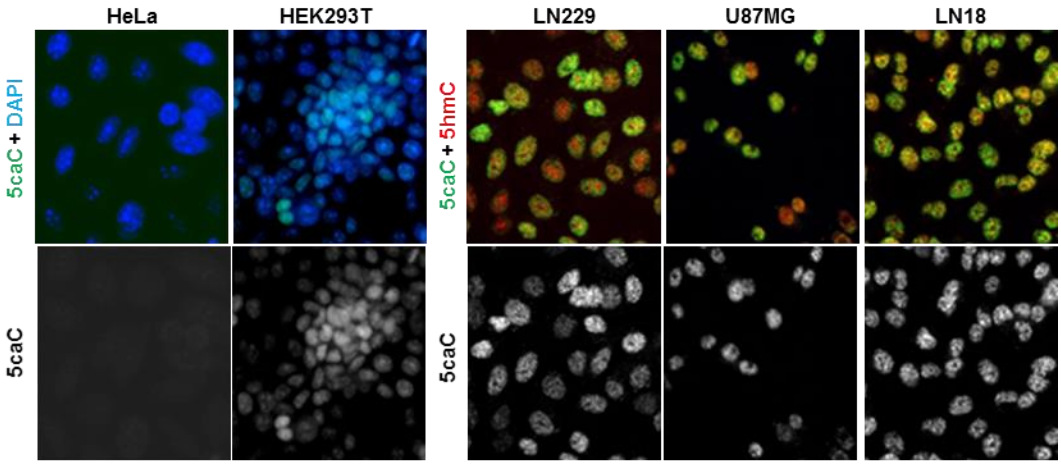


Figure 1.10 Staining glioblastoma cell lines LN22, U87MG and LN18 for 5caC presence against HeLa and HEK293T 5caC negative cells. Cells are stained for 5hmC (red) and 5caC (green). (Ruzov lab unpublished data).

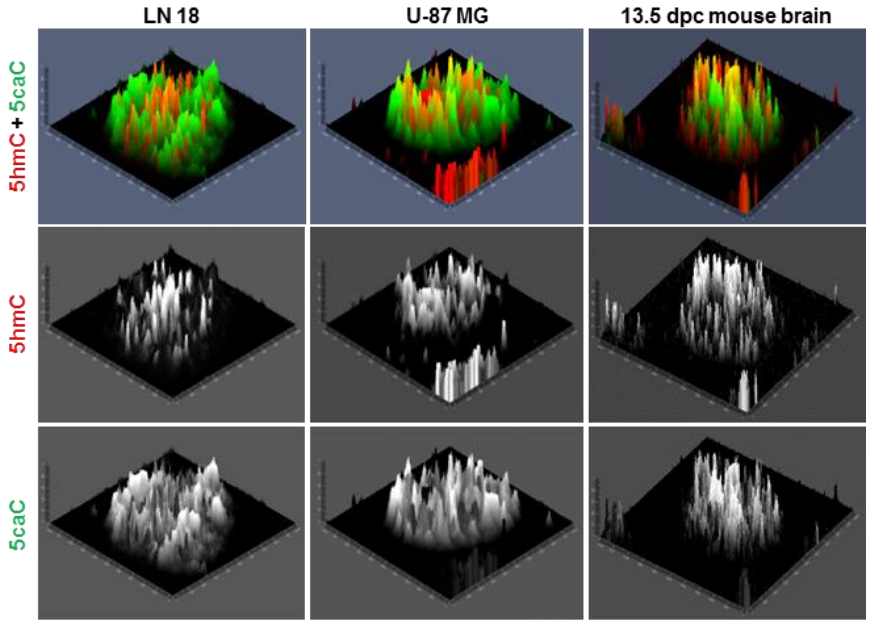


Figure 1.11 2.5D Confocal Microscopy Nuclear Spatial distribution of 5hmC (red) and 5caC (green) signal profiles in LN18 and U87MG glioblastoma cell lines vs 13.5 dpc mouse embryonic brain cells. (Ruzov lab unpublished data).

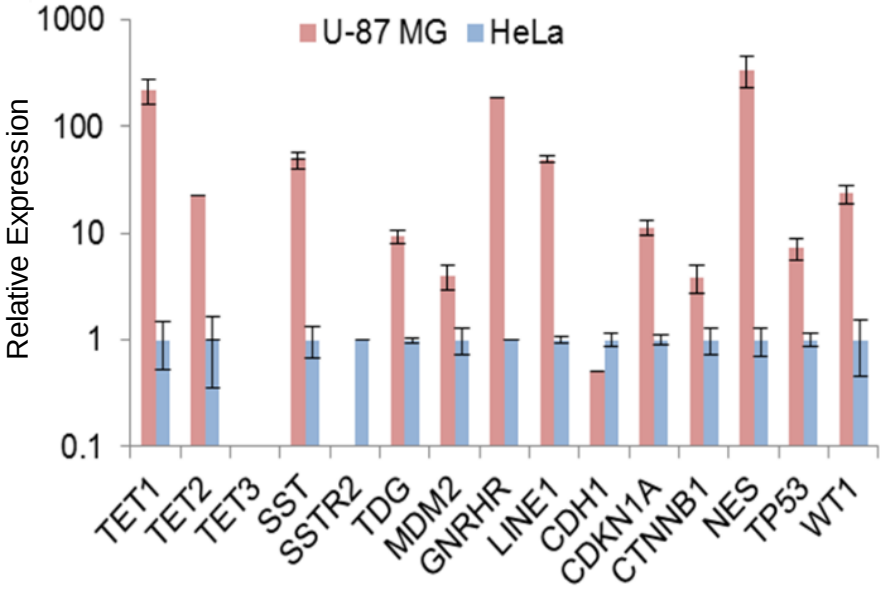


Figure 1.12 Transcriptional profile of glioblastoma pathway genes expression relative to 5caC negative HeLa cells. Mean ± SD, N =1, (Ruzov lab unpublished data).

Considering absence *WT1* of expression and 5caC presence in healthy adult mouse and human terminal differentiated glial cells but their significant up-regulation in embryonic brain and glioblastoma cells (which purportedly possess glial progenitor origins), we postulate a WT1-dependent 5caC modulation of glial progenitor persistence and proliferation in glioblastoma. Consequently, we speculate that transcription of glioblastoma pathogenic genes may be dependent on WT1-5caC mediation and conversely, hypothesize depletion of WT1 from glioblastoma cells may significantly diminish if not abrogate transcription of these genes. If correct, elevated *WT1* expression in adult human glia may signify a potential early prognostic marker and novel therapeutic drug target for the treatment of gliomas and glioblastoma.

As such we endeavour to verify 5caC presence in glioblastoma cell lines and examine WT1 expression levels in these cells. We subsequently aim to assess WT1 influence on transcription of core glioblastoma pathway genes (previously illustrated to possess 5caC occupancy) by small interference RNA (siRNA) specific to *WT1*. Following this, we intend to create *wt1* null mutants by Clustered Recognition of Interspersed Short Palindromic Repeats, CRISPR cas9 technology. This should facilitate examination of effects of permanent WT1 depletion on 5caC gene sequence occupancy assessed by DNA immunoprecipitation, quantitative mass spectrometry and Next Generation Sequencing. Chromatin immunoprecipitation will be performed to investigate possible WT1-5caC interactions at glioblastoma genes of interest assessed by

quantitative PCR. Finally functional studies will be performed upon successful selection and isolation of *WT1* null mutant GBM cells to assess WT1 depletion on cell proliferation, tumorigenicty, stemness and differentiation potential.

Materials & Methods

All reagents used in this thesis can be found by referring to tables 2.11 – 2.24.

2.1 Cell Culture

2.1.1 Cryopreservation of cells

$1x10^6$ cells were pelleted by centrifugation at 200g for 5 minutes. Supernatant was aspirated and pellets were resuspended in 1ml Foetal bovine serum (FBS) (Hyclone) and 5% dimethyl sulfoxide (DMSO) and promptly placed in a Mr Frosty freezing container to enable constant rate of cooling at -1°C/minute. Cells were immediately transferred to the -80°C freezer for 24 hour incubation. Cells were then removed from Mr Frosty container and transferred to liquid nitrogen for long term storage. A summary of media formulations is compiled in table 2.1.

Table 2. 1 Media formulation

Cell Line (serum culture)	**Media components**
U87MG, U251, LN229 & LN18	DMEM + 10% FBS
FB83	DMEM + 15% FBS
UW228-3	DMEM F12 + 10%FBS + 2 mL-Glutamine, 100 g/mL streptomycin and 100 U/mL penicillin
DAOY	MEM/EBSS + 10% FBS + 2ml Glutamine + 1% sodium pyruvate + non-essential amino acids + 100 g/mL streptomycin and 100 U/mL penicillin
DKFZ, BXD1425-EP1N	DMEM + 15% FBS + 100 g/mL streptomycin and 100 U/mL penicillin
Cell Line (serum-free culture)	**Neuro-basal media**
U251, FB83, LN18, LN229	DMEM F12 + 100µl of 100x N2 + 200µl of 50x B27 + 100 g/mL streptomycin and 100 U/mL penicillin + 20ng/ml bFGF + 20ng/ml EGF (Sigma, Poole, UK)
U251, FB83, LN18, LN229	**Neurogenic media**
	DMEM F12 + 100ul of 100x N2 + 200ul of 50x B27 + 100 g/mL streptomycin and 100 U/mL penicillin + 0.5% FCS, 1µM retinoic acid (Sigma), 10 ng/ml BDNF).
U251, FB83, LN18, LN229, U87	**Gliagenic media**
	DMEM F12 + 100ul of 100x N2 + 200ul of 50x B27 + 100 g/mL streptomycin and 100 U/mL penicillin + 5% FBS

2.1.2 Passaging and sub-culturing of cells

1 ml of $1x10^6$ cells were rapidly thawed from liquid nitrogen for 2 minutes in a 37°C water bath and seeded into a T25 (NUNC) pre-warmed (37°C) flask containing 4ml of appropriate media (table 2.1). After 24 hours, fresh media was replenished to remove any residual DMSO from the growth conditions. Media was replenished every 2-3 days and cells were passaged upon reaching 70-80% confluence by washing with 10ml of Hank's buffered saline solution (HBSS) (cat# 14170138) to remove residual media. Aspiration of HBSS and subsequent addition of 2ml of 1X 0.25% trypsin and 0.53mM ethylene-diamine tetra-acetic acid (EDTA) (cat# 93283) dissolved in HBSS was performed and flasks were

incubated for 2-3 minutes in a humidified incubator at 95% humidity, 37°C and 5% CO_2. Following incubation, the flasks were tapped gently to agitate and dislodge adherent cells from the tissue culture plastic surface. Cell trypsinisation was neutralised using 8ml of complete (serum-containing) media. Cancer cells were seeded into new flasks at a ratio of 1:10 whilst foetal neural stem cells were split at a 1:3 ratio in their appropriate media.

2.1.3 Preparation of complete media

50ml of pre-warmed (37°C) 10% Foetal Bovine Serum (FBS) was added to pre-warmed 500ml of Dulbecco's Modified Eagles Medium (DMEM) 1X (Gibco, Life Technologies). Solution was mixed, ready for use for U87MG, LN229, LN18 and U251 glioblastoma cell lines.

BXD-1425EPN and DKFZ-EP1NS cell lines were cultured in DMEM 1X supplemented with 10% Foetal Bovine Serum and 1% Penicillin/Streptomycin (100 g/mL streptomycin and 100 U/mL penicillin). DAOY cells were cultured in Minimal Eagles Medium/ Earle's Balanced Salt Solution (MEM/ EBSS) supplemented with 10% heat-activated foetal bovine serum, sodium pyruvate, non-essential amino acids (NEAA), 2mL Glutamine, 100 g/mL streptomycin and 100 U/mL penicillin.
UW228 cells were cultured in DMEM/F12 (cat# 21331020) supplemented with 10% heat-activated foetal bovine serum, 2 mL-Glutamine, 100 g/mL streptomycin and 100 U/mL penicillin.

2.1.4 Spheroid Culture

U251, LN18, U87, LN229 and FB83 cells were passaged as previously described. Trypsinised cells were pelleted and resuspended in neuro-basal media (DMEM F12, 1:1 B27, N2, 1% L-glutamatine, 20ng/ml bFGF and 20ng/ml EGF). Cells were seeded at a density of 500 -2000 cells in 200ul per well of a 96 round bottom well ultra-low attachment plate (NUNC). Cells were incubated at 37°C, 5% CO_2 in a humidified incubator. Media was changed every 2-3 days, avoiding agitation of growing spheroid. Spheroids were grown for 10 days until diameters of 100-300µm were achieved.

2.2 Molecular Biology

2.2.1 siRNA Transfection

siRNA transfection master mixes were formulated by dissolving 3.3nmol lyophilized WT1 or control scramble siRNA (Santa Cruz Biotech) in 330µl siRNA buffer to make a 10µM stock. Stock solutions were diluted 0.5µl in 190µl serum-free Opti-MEM® (Life Technologies) and mixed with 0.2µl DharmaFECT™ (Dharmacon GE Life Science) transfection reagent diluted in 198µl serum-free Opti-MEM®. This mixture was incubated for 5 minutes at room temperature, dilutes in 1600µl complete media, incubated at room temperature and subsequently loaded onto U87MG passage 6 (p6) cells seeded in 24 well plates and 8 well chamber slides (NUNC). siRNA transfection master mix was added to 500µl on cells at seeding density 3.0 x 10^4 cells/well to make final siRNA concentration of 25nM and 50nM per well. U87MG cells in 24 well plates and chamber

slides were incubated at 37°C and 5% CO_2 for 24 hours after which the procedure was performed again and incubated for a further 24 hours. U-87MG transfected cells were harvested for RNA extraction and reverse transcriptase PCR at 48 and 72 hours. U-87MG cells in chamber slides were harvested for immunohistochemical analysis at 72 hours only.

2.2.2 RNA Extraction

RNA extraction was performed using QIAGEN RNeasy RNA extraction kit as described by manufacturer. Briefly, cell pellets were lysed and homogenised by resuspension in 350µl of RLT buffer containing β-mercaptoethanol as instructed in the protocol. Samples were precipitated with 100% ethanol and on-column DNase digestion was performed to remove genomic DNA. Samples were washed with RW1 buffer, precipitated with RPE buffer and eluted in 30µl RNase-free water.

2.2.3 cDNA Synthesis

2µg of RNA of each sample was added to a 20µl reaction mix containing 10mM dNTPs (1µl) (Invitrogen), 10X random hexamers (Invitrogen) (2µl). Reaction mix was incubated at 65°C for 5 minutes followed by immediate ice incubation for 5 minutes. 5X first strand buffer (4µl), Dithiothreitol (DTT) 0.1M (1µl) and Superscript III Reverse Transcriptase (1µl) (both Invitrogen) were added to the reaction mix, vortexed, incubated at room temperature for 5 minutes followed by incubation at 50°C for 1 hour. Enzyme inactivation step occurred subsequently at 70°C for 15 minutes.

Reaction mix was diluted in sterile distilled water to make final volume of 200µl.

2.2.4 Quantitative Polymerase Chain Reaction

Candidate genes (table 5.2) expression were measured via RT-qPCR performed on Applied Biosystems FAST 7500 RT-qPCR machine and software. 10µl reaction volumes contained 5µl 2X SYBR Green (Promega), 0.25µl forward and reverse primers (5µM) (Sigma Aldrich) (which were designed using www.Primer3.ut.ee and BLASTED for specificity using www.ncbi.nlm.nih.gov BLAST tools) and 2µl sample cDNA. Gene expression was measured using comparative Ct standard run at 40 cycles with denaturation at 95°C, annealing at 60°C and 72°C extension. Target gene expression was normalised to Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) or TATA-Binding Protein (*TBP*) expression.

Table 2. 2 List of qPCR Primers

Forward Primer	Sequence (5'-3')	Reverse Primer	Sequence (5'-3')
TUBB3-F	CCTCAAGATGTCCTCCACCTTCAT	TUBB3-R	GTTCATGTTGCTCTCGGCCTC
CD15-F	GACGACTTCCCAAGTGCCTC	CD15-R	CAAGTTCCGTATGCTCTTGGGC
CD133-F	GATGCCTCTGGTGGGGTATTTC	CD133-R	CGACTCCTTTTGATCCGGGTTC
CDH1-F	CTGCTCTTGCTGTTTCTTC	CDH1-R	AGCTGGCTCAAGTCAAAG
CDKN1A-F	CAAGCTCTACCTTCCCAC	CDKN1A-R	GGGTACAAGACAGTGACA
CTNNB1-F	CCATCTGTGCTCTTCGTC	CTNNB1-R	GGACAAAGGGCAAGATTTC
EGFR-F	CATTTAGGGGTGACTCCTTCACAC	EGFR-R	GACGACTGCAAGAGAAAACTGACC
FABP7-F	CCAGCTGGGAGAAGAGTTTG	FABP7-R	CTCATAGTGGCGAACAGCAA
GAPDH-F	GATGCTGGCGCTGAGTACG	GAPDH-R	GCAGAGATGATGACCCTTTTGG

GFAP-F	GCCAGCTACATCGAGAAGGTTC	GFAP-R	CTGTGCCAGATTGTCCCTCTC
GLAST-F	GTGAATGGAGTCAATGCCCTG G	GLAST-R	CTCCACAATCTTCCCAGCAATC AG
GNRHR-F	CCTCTCCTGAACAGAATC	GNRHR-R	GGAAGAAAGTAACCGTCA
HES1-F	GTCTGAGCCAGCTGAAAACACT G	HES1-R	CCAAGTGTGCTGGGGAAGTAC
LINE-1-F	GAGAACAAAGACACCACATACC	LINE-1-R	GGCATTTAGTGCTATAAATTTC CC
MDM2-F	CCATGATCTACAGGAACTTG	MDM2-R	CTCACTCACAGATGTACC
MS1-F	GGTGGAATGTAAGAAAGCTCAG CC	MS1-R	GAACTGGTAGGTGTAGCCAGG G
NANOG-F	CAGCCTTTACTCTTCCTACCAC CAG	NANOG-R	GTTATAGAAGGGACTGTTCCAG GCC
NESTIN-F	GGAAGAGGAAGAGAACCTGGG AAA	NESTIN-R	CTCAGATTCAGCTCTGCCTCAT C
NF1-F	GCCCTAAAGAAGGTTGCGC	NF1-R	GTGGCCAAACTGCTGCTTTAC
NOTCH1-F	CAACTGCGAGATCAACCTGGAT G	NOTCH1-R	GTGAAGCCATTGATGCCGTCC
POU5F1-F	CAGGCCCGAAAGAGAAAGCG	POU5F1-R	GTTGTGCATAGTCGCTGCTTGA TC
OLIG1-F	ACTTCCTCCTCCTCCACGA	OLIG1-R	CTGAGTAGGGCAGGATGACC
OLIG2-F	GATGACCTTTTTCTGCCGGCC	OLIG2-R	GTGCTGGACGAGGATGACTTG
PDGFRA-F	GAGATGCTTTGGGGAGAGTGA AG	PDGFRA-R	GCCTGCCTTCAAGCTCATTCTC
RB1-F	GGAAGCAACCCTCCTAAACC	RB1-R	TTTCTGCTTTTGCATTCGTG
S100B-F	ATTCTGGAAGGGAGGGAGAC	S100B-R	CGTGGCAGGCAGTAGTAACC
SOX2-F	CACATGAACGGCTGGAGCAAC	SOX2-R	GTAGGACATGCTGTAGGTGGG
SOMATOSTATIN-F	CTGGCCAAGTACTTCTTG	SOMATOSTATIN-R	GACAGATCTTCAGGTTCC
SOMATOSTATIN R2-F	GGCTATCCATTCCATTTG	SOMATOSTATIN R2-R	GCTTGTCAGGTCATAGTA
TBP-F	TATAATCCCAAGCGGTTTGC	TBP-R	GCTGGAAAACCCAACTTCTG
TDG-F	CAGCTATTCCCTTCAGCA	TDG-R	GGAACTTCTTCTGGCATTTG
TET1-F	CTTGGTATGAGTGGGAGTG	TET1-R	GAGCATTAAAGGTAGCAATTG
TET2-F	GCAAGATCTTCTTCACAG	TET2-R	GCATGGTTATGTATCAAGTA
TET3-F	CTCTGAAGTCAGAGGAGAA	TET3-R	GTCCAGGAAGTTGTGTTC
TP53-F	GCCATCTACAAGCAGTCA	TP53-R	CCATCGCTATCTGAGCAG
WT1-F	CCTTCATCAAACAGGAGCCG	WT1-R	GTAGGGCGCGTTAGGAAAC

2.2.5 Standard polymerase chain reaction

Standard PCR reactions were performed in 50µl reaction volumes in as detailed in table 2.3. All primers used were designed in-house, *in silico* and ordered commercially. Standard amplifications were performed using 10x thermostable *Thermus aquaticus* (*Taq*) polymerase buffer

(New England Biolabs (NEB), 10mM dNTPs (NEB) and *Taq* DNA polymerase (NEB). Stock 100µM primers were diluted to working concentrations of 10µM.

Table 2. 3 Standard PCR reaction mix

Component	Volume (µl)
Template DNA	50ng (genomic) 1ng (plasmid)
Thermo *Taq* pol buffer 10x	5
dNTP mix 10mM	1
Forward primer 10µM	2
Reverse primer 10µM	2
Taq DNA polymerase	0.5
Nuclease – free H_2O	Up to 50

Primer annealing temperature were calculated from individual primer sequences utilising thermodynamic equations described by Breslauer *et al*, 1986; Sugimoto *et al*, 1986 and Xia *et al*, 1998, as employed by Oligocalc oligonucleotide properties calculator (https//:biotools.nubic.northwestern.edu) and nearest neighbour temperatures were selected for each primer pair. Polymerase extension phase time at 72°C was estimated as 1 minute per kilobase to facilitate adequate amplicon polymerisation. PCR reaction conditions are summarised in Table 2.4 and were performed on Tecne TC312 thermocycler.

Table 2. 4 Thermocycler conditions

Condition	Temperature °C	Minutes/ seconds
Initial denaturation	94	5 minutes
Denaturation	94	30 sec
Annealing	variable	30sec
Extension	72	variable
Final extension	72	5 minutes
Hold	4	indefinitely

Table 2. 5 CRISPR Cas9 construction primers (designed in-house)

Primer	Sequence
WT1ex6F2	CACTTACTAGTATGAACCACAGCACAGGGTACGAGA
WT1ex10R	TATCGGAATTCTCAAAGCGCCAGCTGG
KTS R	GGGCTTTTCACTTGTTTTACCTGT
7a3F2	GGATCTCATTAAAGGCAACCTCTCC
7a3R	GGCGCGCCGCATTAAGGCGCCGGATCTCATTAAAGGCAACCTCTCC
9a3F2	GGCGCCTTAATGCGGCGCGCCGAGAGAAGGAATAGTGCGTGG
9a3R	TCTAGAAGCTTCACCAGCCTTGCAGAAAGC
FLcas9F	ACGCACTATTCCTTCTCTCGGGCAGCTAATGGACCTTCT
FLcas9R	GGTTGCCTTTAATGAGATCCGCCATAGAGCCCACCGCA
sgRNA 8AF	TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGACCAGCT CAAAAGACACCAA
sgRNA 8AR	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTTGGTGTCT TTTGAGCTGGTC
7aRHAR	AGTATGTTGGGAGGCGTAGG
WT1RintgF	TGGGAAGACAATAGCAGGCA
WT1LintgR	CTTCTCTAGGCACCGGTTCA
RHRF	CAGTCTCCTCATCTGTAAGATGGG
In7ajncF	CCTAGTAGGAGAGGTTGCCTT
9ajncR	GCCACGCACTATTCCTTCTCT
WT1cDNAex4F	TCAGATGAACTTAGGAGCCAC
WT1ex10R	GCCAGCTGGAGTTTGGTCA

Table 2. 6 Phusion PCR conditions

Component	Volume per reaction µl
Nuclease-free water	2
5x Phusion GC Buffer	3
10mM dNTPs	-
Fragment 1	x
Fragment 2	x
Phusion DNA polymerase	
Total	15

2.2.6 Phusion PCR

Phusion PCR oligonucleotides were designed with 20bp overlaps to ensure complementary sequence hybridisation the during denaturation/re-naturation step. Equimolar (100-1000ng) DNA template fragments with complementary 20bp overlaps were prepared in the

phusion PCR reaction mix (table 2.6) and incubated in the thermocycler under the following conditions (table 2.7):

Table 2. 7 Denaturation / renaturation thermocycler conditions

Step	Temperature °C	Time seconds/minutes
Initial Denaturation	94	5 minutes
Annealing Cycle	94	30 sec
	75	5 minutes
	52	5 minutes
	45	5 minutes
Hold	4	-

2.2.7 Gibson Assembly

Gibson Assemblies were performed using the NEB Gibson assembly master mix and protocol.

2.2.8 Restriction digests

Restriction digests were performed on 2µg of plasmid DNA to ensure that sufficient quantity of DNA would remain following gel extraction and purification for subsequent subcloning. Digests were performed in 50µl reaction volumes in appropriate reaction buffers (NEB). DNA was digested with 1 unit of restriction enzyme per µg of DNA in the reaction mix incubated at 37°C in a water bath for 4 hours. Overnight incubation was performed when high-fidelity NEB enzymes were used to ensure complete digestion. Following incubation, restriction enzymes were deactivated by incubation of reaction mixtures at 70°C for 15 minutes.

2.2.9 Agarose gel electrophoresis and gel extraction

Detection of DNA of required size in base pairs was resolved by agarose gel electrophoresis. Gels of either 1% or 1.3% were cast to separate DNA

band sizes of 500 – 8000bp and 100 – 500bp respectively. Molecular grade agarose (Sigma) was dissolved in 1X TAE buffer by heating at maximum power for 3-4 minutes in a microwave. Molten gel was cooled prior to addition of ethidium bromide (Invitrogen) at a final concentration of 250ug/ml was poured into electrophoresis trays with well combs in place and allowed to set for 15 minutes. DNA samples were mixed with 6x loading dye (NEB) at a sample to dye ration of 6:1. For verification of band presence or absence, 5µl of sample/dye mixture was loaded into set electrophoresis gels submerged in 1X TAE running buffer whereas 50µl of sample/dye mix was loaded for bands requiring subsequent gel extraction. 200ng of DNA molecular weight quick-load hyper ladder markers of either 1kb or 100bp (NEB) were run in parallel with samples of appropriate band size. Gels were run at 80-90 volts, 400mA for 1 hour and visualised on Fujifilm LAS 4000 UV transilluminator. Pictures were captured using LAS 4000 image acquisition software version 2.1. Gel bands were extracted using clean scalpel and forceps and purified using QIAGEN gel DNA purification kit. DNA was eluted in 30ul and quantified by nanodrop spectrophotometer.

2.2.10 Quantification of nucleic acid yield and purity

Quantification of DNA and RNA samples was performed by measuring UV absorbance at 260nm using Nanodrop 1000 spectrophotometer. Absorbance ratios of DNA and RNA were calculated as a ratio of nucleic acid: protein at 260/280nm. Protein contamination was measured as an absorbance ratio of 260/230nm.

2.2.11 Ligation

Ligations were performed in 10µl reaction volumes consisting of vector and insert DNA at a ratio of 1:3 (10-100ng) respectively. Phosphodiester bond linkage between insert and vector backbones was catalysed by T4 DNA ligase (NEB) as summarized in table 5.8. Ligations were performed overnight at room temperature.

Table 2. 8 DNA Ligation conditions

Reagent	Volume µl
Vector DNA	x
Insert DNA	3x
T4 DNA ligase Buffer	1
T4 DNA ligase	1
Nuclease-free water	Up to 10

2.2.12 Bacterial Transformation

Transformations were performed using One-shot Top 10 chemically competent DH5α *Erechia coli* bacteria (Invitrogen). Bacterial cells (50µl) were thawed on ice for 10 minutes before addition of 5µl of ligation mix or 1ul of plasmid (50ng). Cells were gently mixed and incubated on ice for 30 minutes. Cells were heat-shocked at 42°C for 30 seconds and immediately placed on ice for 2 minutes. Cells were then resuspended in 250µl of pre-warmed (37°C) SOC media (Invitrogen) and incubated for 1 hour at 37°C under shaking conditions at 225RPM. Cells in suspension were plated onto agar plates inoculated with the appropriate antibiotic selection (amplicilin 100ug/ml or kanamycin 50ug/ml). Plates were incubated overnight for 16-18 hours. Bacterial colonies were scored and harvested for starter culture and subsequent plasmid mini-prep extraction.

2.2.13 Starter culture

Individual transformed bacterial colonies were picked from agar selection plates using a sterile P10 pipette tip and used to inoculate 5ml of Luria Bertani (LB) broth with antibiotic selection in a 15ml falcon tube. Each colony picked was placed in a separate tube for incubation overnight for 12-16 hours at 37°C until an optical density at 600nm (OD_{600}) was achieved, measured by spectrometer. Starter cultures were centrifuged at 8000rpm for 5 minutes and supernatants were discarded. Pellets were either frozen for mini-prep plasmid extraction or resuspended in 100ml LB broth for midi-prep plasmid extraction.

2.2.14 Overnight culture

Bacterial cell pellets were resuspended in 100ml of antibiotic selection LB broth and incubated in an orbital shaker at 225 revolutions per minute (rpm) overnight at 37°C. Cell suspensions were centrifuged to pellet cells from which plasmid DNA was extracted using midi-prep kits.

2.2.15 Plasmid mini-prep

Alkaline lysis plasmid mini-prep was performed by using Promega Wizard Plus SV Miniprep DNA Purification System according to the manufacturer's protocol.

2.2.16 Plasmid midi-prep

Plasmid midi prep DNA extraction was performed using Macherey Nagel Nucleobond Xtra Midi-prep kit according to the manufacturer's protocol.

2.2.17 Chromatin Immunoprecipitation (ChIP)

ChIP experiments were optimised on RBLPAT induced pluripotent stem cell WT1 control cell line. Two 80% confluent T75 flasks of U251 cells per biological repeat were used as staring material for ChIP.Chromatin Immunoprecipitation was performed according to Magna ChIP kit manufacturer's instructions. DNA was made up to 130µl with nuclease-free water and sonicated using Covaris S220 ultrasonicator using the Chromatin Shear programme to generate 200-1000 base pair fragments. Sonicated DNA was verified on Agilent Tapestation D5000 screen tapes. We adapted the protocol by using half-volumes of all reagents to save on cost. 2.75µg of GTX Rabbit monoclonal anti-Wilms' Tumour 1 antibody was used for precipitation. For preparation of Next Generation Sequencing libraries, adaptor ligation was performed following column purification of DNA.

2.2.18 Methylated DNA immunoprecipitation (MeDIP)

MeDIP experiments were optimised on RBLPAT induced pluripotent stem cell 5caC control cell line. Modification of the MeDIP protocol devised by Mohn *et al*, 2009 enabled isolation of high yield methylated DNA fragments for Next Generation Sequencing and PCR analysis. Brief modifications include: cell pellets were resuspended in 1000µl of lysis

buffer. 37.5µl of RNase A at 100mg/ml was added to lysis buffer. One volume of phenol/chloroform (Sigma) was mixed with the lysate to generate organic-interphase and aqueous phase separation. Aqueous and interphases were taken and placed in fresh tubes. DNA was precipitated with 1/10 by volume of 5M NaCl mixed in the aqueous phase and 1 volume of isopropanol. DNA pellets were resuspended in 50µl of nuclease-free water. DNA was made up to 130µl with nuclease-free water and sonicated using Covaris S220 ultrasonicator at a frequency of 550KHz using programme 350bp_60s to generate 350 base pair fragments. Adaptor ligation was performed at this stage prior to immunoprecipitation. Modifications to the immunoprecipitation protocol included using 2.5µl of Active Motif Rabbit anti-5caC antibody. Sheep anti-rabbit magnetic Dynabeads M-280 (Invitrogen) were used to precipitate 5caC modified DNA fragments. 3.75µl of Proteinase K at 20mg/ml in 250µl Proteinase K buffer enabled separation of Dynabeads from antibody bound modified DNA fragments prior to overnight incubation at 55°C. DNA was purified using QIAGEN QIAquick column purification kit and eluted in 25µl of nuclease-free water.

2.2.19 DNA Immunoprecipitation-qPCR

DNA isolation from cell pellets and sonication was performed as described in section 2.2.18. Sonicated DNA fragments were immunoprecipitated without prior adaptor ligation and purified in nuclease-free water as described in 2.2.18. Purified DNA was quantified using nanodrop 2000 and Qubit spectrometers. DNA was

diluted to 10ng/µl using sterile distilled water and 2ul of this DNA fragment mixture was used as template DNA for qPCR analysis of 5caC enrichment at gene promoters / enhancers. QPCR primers were designed *in-silico* using http://primer3.ut.ee/ website to amplify 200-500bp regions situated 1kb upstream of transcriptional start sites of genes of interest. Primers thermodynamic specifications were verified *in-silico* using http://biotools.nubic.northwestern.edu/OligoCalc.html primer design tools. Primer sequences were designed to fall within 50-60% GC content, be 22-25 nucleotides in length and possess ≤1°C difference in annealing temperature between forward and reverse primer pairs. Designed primer pairs were NCBI BLAST and University of California Santa Cruz (UCSC) genome browser https://genome.ucsc.edu/cgi-bin/hgPcr verified to bind to their specific target sequence. QPCR was performed was described in section 2.2.4 using sonicated DNA instead of cDNA.

2.3 Transfection and Selection of Cells

2.3.1 Lipofection

Cells were seeded in 6 well plates at a density of $2.5x10^4$ cells per well in complete media. 24 hours post seeding cells were examined and once reached 50% confluence, were incubated with transfection complex medium. Transfection complex medium consisted of 12µl of Lipofectamine 2000 (Life Technologies) dissolved in 150µl of Optimem

(Gibco). Plasmid DNA was dissolved in 150ul of Optimem and allowed to equilibrate for 5 minutes before lipofectamine-Optimem and DNA-Optimem mixtures were combined at a 1:1 volume ratio and incubated at room temperature for 5 minutes to facilitate DNA-lipid vesicle complex formation. Complete media was aspirated off cells and 700µl of Optimem was added to lipofectamine-DNA complex to constitute a 1ml volume which was administered per well to cells. Cells were incubated for 24 hours at 95% humidity, 37°c, 5% CO_2 in humidified incubator and transfection media was replaced by complete media the next day. Transfection efficiency was calculated by scoring green fluorescent protein positive cells against total number of cells per field of vision under fluorescent microscope examination.

2.3.2 Electroporation

Electroporation of cells was performed using the Nucleofector II AMAXA (Lonza) with the VCA-1005 kit (Lonza). Briefly, for CRISPR-cas9 facilitated gene targeting of WT1 in U251 and LN18 glioblastoma cell lines, plasmid DNA was aliquoted into sterile Eppendorf tubes at the following quantities: Nickase D10A plasmid – 1µg; guide RNA 8A plasmid – 500ng; guide RNA 8B plasmid – 500ng and homology directed repair donor template plasmid – 3µg. GFP positive control plasmid was used at 3µg. 500 000 cells per transfection were harvested from T75 flasks, pelleted and resuspended gently in 100µl of nucleofection buffer containing added nucleofection supplement per transfection. Resuspended cells were aliquotted in 100µl volumes to individual

plasmid containing Eppendorf's for respective treatments. Cells were mixed with plasmid DNA and transferred to cuvettes and placed in nucleofector II AMAXA electroporator. Cells were electroporated using programmes X-001 U87MG (U251) and U-030 C6 glioma high efficiency pulses (LN18). Cuvettes were then incubated at 37°C, 5% CO_2 and 95% humidity for 10 minutes after which 900µl of warm complete media was added to cells in cuvettes and resuspended cells were seeded into 6-well plates containing pre-warmed 2ml of complete media. Cells were incubated at 37°C, 5% CO_2 and 95% humidity overnight. Cells were analysed the next day to assess transfection efficiency and cell viability.

2.3.3 Limiting dilution Assay

Glioblastoma Cells to be clonally distributed were harvested from T75 flasks by spent media aspiration, HBSS wash and aspiration followed by trypsinisation, agitation and collection as previously described. Cells were counted using a haemocytometer. Average cell number present calculated by averaging the total of cell counts in 4 quadrants of the haemocytometer loaded with 10ul of cell suspension in technical duplicate. The average cell count per µl was calculated and divided by a volume to generate 48 cells/µl. 1ul of this suspension was resuspended in 19.2ml of 1X DMEM + 10% FBS and mixed. 200µl of this cell suspension was aliquoted into each well of a fresh 96 well plate to achieve a theoretical density of 1 cell per 2 wells. Cells were grown until clonal density was achieved.

2.4 Next Generation Sequencing

2.4.1 Qubit DNA quantification

Quantification of DNA samples was measured using Qubit dsDNA broad range and or high sensitivity kits on the Qubit fluorometer according to the manufacturer's instructions.

2.4.2 Adapter ligation (MeDIP-Seq and ChIP-Seq)

Adapter ligation was carried out according to the NEBNext Ultra II DNA library Prep Kit for Illumina instruction manual. Input and MeDIP DNA samples were used at quantities of 1µg and 5µg respectively. Input DNA was not diluted nor was ligation enhancer and ligation mastermix mixed ahead of time as stated by the protocol. Clean up of adaptor ligated DNA was favoured over size selection owing to our purchase of AMPure XP beads. Following ethanol washing of beads and air-drying, DNA was eluted from magnetic beads in 20µl of 0.1X TE for input, and 50µl for MeDIP samples. Libraries were verified for DNA size distribution using Agilent Tapestation D1000 chips.

2.4.3 PCR amplification of MeDIP samples

PCR amplification of MeDIP samples was carried out according to the NEBNext Ultra II DNA library Prep Kit for Illumina instruction manual. PCR amplification cycles were determined based on yield of adaptor ligated DNA fragments following 5caC immunoprecipitation and AMPure

XP bead selection clean up. PCR cycles were selected to generate 100ng of DNA for sequencing.

2.4.4 Tapestation analysis of RNA integrity

RNA quantity and integrity was measured on the Agilent RNA ScreenTape System according to the manufacturer's instructions. Following RNA extraction, broad range RNA Qubit kits and Agilent RNA screentapes used on the Agilent Tapestation were performed to quantify RNA levels based on the manufacturer's instructions. Only samples possessing RNA integrity numbers (RIN) greater than 7 were used for ribo-depletion.

2.4.5 Ribo-depletion

Double ribo-depletion (back to back ribo-depletions) were performed using Illumina Ribo-Zero ribosomal RNA depletion kits according to the manufacturer's instructions. Qiagen RNeasy Min-Elute Clean-up kits were used to purify ribo-depleted RNA. 4ug of RNA per sample were used for ribo-depletion and high sensitive RNA screentapes were used on the Agilent Tapestation to verify ribosomal RNA peak depletion, RNA yield and integrity following elution.

2.4.6 Transcriptome library preparation

RNA-Seq libraries were prepared using the New England Biolabs NEBNext Ultra II Directional RNA Library Prep Kit for Illumina. RNA library adaptors were diluted 25 fold prior to library ligation and PCR

cycles were selected based on ribo-depleted RNA input yield. RNA libraries were verified for suitability for sequencing using the Agilent D1000 chip to check for DNA library size distribution.

2.4.7 Sequencing

Sequencing was performed at Queen's Medical Centre DEEPSEQ facility on the Illumina NextSeq500 short read platform. ChIP-Seq and RNA-Seq libraries were run using 75bp pair-ended reads while 5caC DIP-Seq libraries were run using 150bp single-ended reads. 45 million reads were generated per biological repeat.

2.6 Staining & Imaging

2.6.1 Immunofluorescence

Immunostaining on fixed cells was performed according to protocol detailed by Abakir, Wheldon & Ruzov, 2015.

Briefly, cells seeded at a density of $3.0x10^4$ per well in 8 well chamber slide were fixed in 4% paraformaldehyde (PFA), washed twice in xylene and rehydrated in 95, 75 and 50% ethanol, all at room temperature for approximately 10 minutes. Slides were subsequently fixed, washed and permeabilised in PFA (15 minutes), phosphate buffered saline (PBS) (5 minutes), and PBX (30 minutes) respectively. Slides were quickly washed in PBT. Samples were depurinated in 2 or 4 N hydrochloric acid (HCL) for 1 hour and neutralized in 10mM Tris-HCL for 30 minutes. Both steps performed at room temperature. 10 minute wash was performed at

PBS/0.1& Tween-20 and wells were subsequently demarcated using hydrophobic barrier pen. Owing to possible denaturation of native protein antigen epitopes by HCL treatment, depurination was omitted for the interrogation of protein markers e.g. nestin and without oxi-mC colocalisation. We incubated blocked (10% bovine serum albumin, BSA in PBS) samples with 100µl anti – 5caC rabbit polyclonal (1:500), anti – 5fC rabbit polyclonal (1:500) and anti – 5hmC mouse monoclonal (1:500) primary antibodies (table 2.9) (Active Motif) for 60 minutes in humid chamber (table 2.9). Following PBT wash, goat anti – rabbit horseradish peroxidase (HRP) conjugated (1:400) and donkey anti – mouse Alexa 555 fluor conjugated (1:400) secondary antibodies (table 5.10) (Molecular Probe) were added to samples and incubated for 60 minutes in humid chamber. Tyramide signal amplification (TSA) (Perkin Elmer) was added (1:200) to samples after 3 PBT washes, excess removed and DAPI inoculated mounting medium nuclear stain was added drop-wise to samples on slide before coverslip was placed and sealed with nail polish (Abakir, Wheldon & Ruzov, 2015).

Table 2. 9 Primary antibody specifications

Antibody	Host	dilution	Clonality
Anti-WT1 (GTX)	Rabbit	1:100	Polyclonal
Anti-WT1 (SC)	Mouse	1:200	Polyclonal
Anti-WT1 (INV)	Mouse	1:200	Monoclonal
Anti-5hmC	Mouse	1:500	Monoclonal
Anti-5caC	Rabbit	1:500	Polyclonal
Anti-SOX2	Mouse	1:200	Monoclonal
Anti-NESTIN	Mouse	1:200	Monoclonal
Anti-NOTCH1	Mouse	1:200	Monoclonal
Anti-Ki67	Mouse	1:200	Monoclonal
Anti-TUBB3	Mouse	1:200	Monoclonal
Anti-GFAP	Chicken	1:200	Polyclonal

Table 2. 10 Secondary antibody specifications

Antibody	Host	dilution	Fluorophore wavelength(nm)
Anti-rabbit	Donkey	1:400	488
Anti-mouse	Goat	1:400	555
Anti-mouse	Goat	1:400	633

2.6.2 Fluorescent & confocal microscopy

Fluorescence images were captured using Nikon i90 fluorescent and Zeiss 780 LSM microscope and Zen software. All images were captured at identical exposure settings and acquisition protocols for red, blue and green channels. Zen software was used to quantify fluorescence from images capture on Zeiss 780 LSM confocal laser scanning microscope. Full details of confocal image signal quantification and analysis are reported in *Immunostaining for DNA Modifications: Computational Analysis of Confocal Images*; Ramsawhook *et al*, 2018.

2.7 Functional Studies

2.7.1 Cell proliferation

Cells were seeded in 6-well plates at a density of $1.0x10^5$ per well and monitored over the course of 10 days. After 24 hours post seeding and subsequent harvesting, cells were counted in 10ul duplicates on a haemocytometer. The averages of 4 technical replicates were recorded each day and the assay was performed in biological triplicate.

2.7.2 Differentiation programmes

For FB83 and GBM cell lines grown in 10% FBS, cells were conditioned for 4 days in neurobasal media on adherent 6-well plates. Following

conditioning, neurobasal media was aspirated off and substituted by pro-neuronal (DMEM F12 1:1 N2, B27, 1um retinoic acid, 10ng/ml BDNF and 0.5% FBS) or pro-glial (DMEM F12 1:1 N2, B27 and 10ng/ml BMP4) media. Media was changed every 48 hours and cells were harvested 6 days post pro-neuronal or pro-glial media induction.

2.8 Statistical Analysis

All experiments unless otherwise stated were performed in biological triplicate (n = 3), the mean and standard deviation (mean ± SD) of which were reported and plotted in figures. Statistical analysis was performed using GraphPad Prism 6 software. All data was checked for normal distributions using D'Agostino-Pearson Omnibus and Shapiro-Wilk normality tests. Unpaired two tailed parametric Student's t-tests, one-way and two-way ANOVA tests were performed where appropriate on normally distributed data whereas non-parametric Mann-Whitney, one-tailed t-tests and Kruskal-Wallis tests were performed on non-Gaussian data sets. All statistical tests were conducted using a 95% confidence interval and therefore not significant, significant, highly significant, very highly significant and extreme significantly different results were denoted by NS $p \geq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ and **** $p < 0.0001$ respectively.

2.9 Bioinformatics Analysis

Bioinformatics analysis of Next Generation Sequencing was performed at the Advanced Data Analysis Centre, ADAC at the University of Nottingham, Sutton-Bonington by Dr Tom Giles.

2.10 Mouse Orthotopic Xenograft

All procedures described here were performed by qualified and trained technicians at the Queen's Medical Centre animal house facility.

$1.0x10^6$ U251 control or C5A mutant cells were injected in a volume of 5µl of saline into the cerebrum of 8 female immunocompromised CD-1 Nu/Nu mice at coordinates: 7mm posterior to bregma, 1mm right of the midline, 3mm deep with adjustments for brain size using stereotactic manipulators. Mice were monitored daily for 40 days for signs of weight loss, diarrhoea, subdued behaviour or ocular and or nasal discharge. Mice exhibiting sudden weight loss greater than or equal to 10% of total body weight were culled by cervical dislocation. Brains were harvested and fixed in formalin for sectioning. Sections were wax embedded and sliced to 4µm thin sections using Leica RM2235 microtome.

2.11 Ethics Statements

These studies, and the experimental protocols required, were reviewed and approved by the National Research Ethics Service Committee East Midlands – Nottingham 2 and have therefore been performed in accordance with the ethical standards laid down in an appropriate

version of the 1975 Declaration of Helsinki, as revised in 1983. For all patients, informed consent was obtained from the patient, or a parent and/or legal guardian where the patient was under 18 years of age, prior to their inclusion in the study. All animal experiments were performed in accordance with the United Kingdom Animals (Scientific Procedures) Act 1986, under the UK Home Office project license authority PPL40/3559. Ethical approval was granted by the University of Nottingham Animal Welfare and Ethical Review Board.

2.12 List of Reagents

Table 2. 11 Antibody specifications

Product Name	Host	1°/2°	Concentration	Company	Cat#
5-Carboxylcytidine/5-Carboxylcytosine (5-caC)	Rabbit pAb	Primary		Active Motif	61225
5-Carboxylcytosine (5-caC) antibody (pAb)	Rabbit pAb	Primary		Active Motif	61229
5-Carboxylcytosine / 5-caC antibody	Rabbit pAb	Primary		GeneTex	GTX60801
5-Hydroxymethylcytosine (5-hmC) antibody (mAb)	Mouse	Primary		Active Motif	39999
5-Hydroxymethylcytosine (5-hmC) antibody (pAb)	Rabbit pAb	Primary		Active Motif	39791
Anti-Nestin Antibody, clone 10C2	Mouse mAb	Primary		Millipore	MAB5326
Anti-Olig2 Antibody, clone 211F1.1	Mouse mAb	Primary		Millipore	MABN50
EGFR Monoclonal Antibody (H11)	Mouse mAb	Primary		Invitrogen	MA5-13070
GFAP Polyclonal Antibody	Chicken pAb	Primary		ThermoFischer	PA1-10004
NOTCH1 Monoclonal Antibody (A6)	Mouse mAb	Primary		Invitrogen	MA5-11961
SOX2 Monoclonal Antibody (20G5)	Mouse mAb	Primary		Invitrogen	MA1-014
Wilms Tumour 1 Antibody	Rabbit pAb	Primary		GeneTex	GTX15249
Wilms Tumour 1 Antibody	Mouse mAb	Primary		Invitrogen	MA1-46028
WT1 antibody (H-1)	Mouse mAb	Primary		Santa Cruz Biotechnology	sc-393498

Alexa Fluor(TM) 633 Goat Anti-Mouse IgG (H+L) highly cross-adsorbed	x	Secondary	2mg/ml	Invitrogen/Life Technologies	A21052
Alexa Fluor(TM) 633 Goat Anti-Mouse IgG (H+L) highly cross-adsorbed Invitrogen	x	Secondary		Invitrogen	A-21052
AlexaFluor 555 Donkey Anti-Mouse	x	Secondary		Invitrogen	A31570
Donkey Anti-Goat IgG H&L (HRP)	x	Secondary	500ug	Abcam	ab205723
Donkey anti-Rabbit IgG (H+L) Secondary Antibody, HRP	x	Secondary		Invitrogen	A16035
Goat anti-Chicken IgY (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 633	x	Secondary	2mg/ml	Invitrogen	A-21103
Goat Anti-Rabbit IgG	x	Secondary		Invitrogen	T30954
Goat Anti-Rabbit IgG (H+L), HRP conjugate	x	Secondary	1mg/ml	Invitrogen	65-6120
Goat anti-Rat IgG-heavy and light chain Antibody DyLight® 594 Conjugated	x	Secondary		Bethyl Lab/Cambridge Bioscience Ltd	A110-105D4

Table 2.12 IHC specifications

Product Name	Details	Company	Cat#
CFM-3 moutning medium	(without DAPI)	Citifluor	CFM3
Chamber Slide, Lab-Tek, Glass, 8-well	16pk	Thermo Scientific Nunc	177402
Chamber Slide, Lab-Tek, Permanox, 8-well	16pk	Thermo Scientific Nunc	177445
FluoroshieldTM with DAPI, histology mounting medium	20ml (with DAPI)	SIGMA	F6057
Mounting Medium for Fluoresence	with DAPI	Vector-Vectashield	H-1200
SLS Coverslips No 0 22x64mm	100pk	Scientific Laboratory Supplies	MIC3208
TSA Kit 41* 50-150	(frozen buffer)	Molecular Probes	T30954
TSA Plus Fluorescein System	(preferred)	Perkin Elmer	NEL741001KT
TSA Plus Fluorescein System	(preferred)	Perkin Elmer	NEL741001KT
Xylene	2.5L	Fischer Scientific	x/0200/17

Table 2.13 RT-qPCR reagent specifications

Product Name	Details	Company	Cat#
Molecular Grade Water Biology Reagent	1L	SIGMA	W4502
Deoxynucleotide (dNTP) Solution Mix	8uml/40umol - 10mM each nt	NEB	N0447S/N0447L

Deoxynucleotide (dNTP) Solution Set	25um of each, 100mM	New England Biolabs	N0446S
GoTaq - qPCR master mix		Promega	A6001/A6002
MicroAMP Fast Optical 96-well Reaction plate	with Barcode (0.1ml)	Applied Biosystems (Life Tech)	4346906
MicroAMP Fast Optical 96-well Reaction plate	(0.1ml)	Applied Biosystems (Life Tech)	4346907
Optical Adhesive Covers	PCR compatible	Applied Biosystems (Life Tech)	4360954/4311971
PB Buffer	500ml bottle	QIAgen	19066
Random Hexamers	50uM	Thermo Fischer	N8080127
SuperScript® III Reverse Transcriptase	10,000 units/4x10,000 units	Invitrogen	180800(44/85)

Table 2. 12 DNA Immunoprecipitation reagent specifications

Product Name	Details	Company	Cat#
E. coli DNA ligase	200 units	NEB	M0205S
Klenow Fragment (3'→5' exo-)	200 units	NEB	M0212S
Proteinase K	100mg	Promega	V3021
Proteinase K (PK) Solution (20mg/ml)	4ml	Promega	MC5005
Rnase A (100mg/ml) 2.5ml	17500 U	QIAgen	19101

Table 2. 13 Chromatin Immunoprecipitation reagents specifications

Product Name	Details	Company	Cat#
20-163 | SDS Lysis Buffer - for use in ChIP Assay	Nuclear Lysis Buffer in #17-610 - 10ml	Millipore	20-163
EZ-Magna ChIP A	Chromatin Immunoprecipitation Kit (22)	Millipore	17-408
Immunoprecipitation kit		Abcam	ab206996
Magna ChIP A	Chromatin Immunoprecipitation Kit (22)	Millipore	17-610
Protease Inhibitor Cocktail 200X in DMSO	1ml	SIGMA	P1860
Protein G Magnetic Beads		NEB	S1430S
Sheep anti-mouse IgG	Dynabeads M-280	Invitrogen	11201D
Sheep anti-rabbit IgG	Dynabeads M-280	Invitrogen	11203D

Table 2. 14 DNA/ RNA purification reagent specifications

Product Name	Details	Company	Cat#

Phenol:Chloroform:Isoamyl Alcohol 25:24:1	Saturated with 10 mM Tris, pH 8.0, 1 mM EDTA.	SIGMA	P2069
RNase A (17,500 U)	2.5ml (100mg/ml)	QIAgen	19101
DNeasy Blood and tissue kit	50	Qiagen	69504
RNase-Free DNase Set (50)		QIAgen	79254
Glycogen, RNA grade, 20mg/mL in water	2x0.1ml	Thermo Scientific	11873933
Ambion; Acid-Phenol:Chloroform; pH 4.5; with IAA, 125:24:1;Liquid; Molecular Biology Grade	100ml	Ambion/invitrogen	10095904/AM9720
NucleoBond Xtra Midi EF (Macherey Nagel) 740420.10	10/50 samples	Fischer Scientific	12773550 (10)/12798412 (50)
QIAquick PCR Purification Kit	50/250	QIAgen	28104/28106
DirectPCR Lysis Reagent (Cell) 100ml #302-C	100ml	bioquote	302-C
NucleoSpin® RNA	50/250 samples	Macherey-Nagel	740955-50/250
RNeasy mini kit (50/250)	50 RNeasy Mini Spin Columns, Collection Tubes (1.5 ml and 2 ml), RNase-free Reagents and Buffers	QIAgen	74104/74106
DNase I Reaction Buffer		NEB	B0303S
Taq DNA Polymerase with ThermoPol Buffer	2000 units	NEB	M0267L
DNase I (RNase-free)		NEB	M0303S
Q5® High-Fidelity DNA Polymerase	100/500 units	NEB	M0491S/L
Promega™ GoTaq™ Long PCR Master Mix	100 units	Promega	M4021
RQ1 RNase-Free Dnase	1,000u	Promega	M6101
Quick-Load 100bp Ladder	1.25ml	NEB	N0467S
1Kb ladder -	100ug	NEB	N3232S/L
50bp DNA Ladder	0.1ml	NEB	N3236S
Phenol:Chloroform:Isoamyl Alcohol 25:24:1 Saturated with 10 mM Tris, pH 8.0, 1 mM EDTA	100/400ml	SIGMA	P2069 100ml/400ml

Table 2. 15 NGS library preparation reagent specification

Product Name	**Details**	**Company**	**Cat#**
Trizma® hydrochloride solution BioPerformance Certified, pH 7.5, 2 M, suitable for cell culture	100ml	SIGMA	T2944
Tris-EDTA buffer solution BioUltra, for molecular biology, pH 8.0	100ml	SIGMA	93283
NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina®	24 reactions	NEB	E7760S
NEBNext® Ultra™ II DNA Library Prep Kit for Illumina®	24 reactions	NEB	E7645S
NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1)	24 reactions	NEB	E7335S
NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 2)	24 reactions	NEB	E7500S
NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 4)	24 reactions	NEB	E7730S
Ribo-Zero Gold rRNA Removal Kit (H/M/R)	6 reaction	illumina	MRZG126
Ribo-Zero Gold rRNA Removal Kit (H/M/R)	24 reactions	illumina	MRZG12324
RNeasy MinElute Cleanup Kit (50)	50 reactions	Qiagen	74204
NEBNext rRNA Depletion Kit (Human Mouse Rat)	6 reactions	NEB	E6310S

Table 2. 16 siRNA reagent specifications

Product Name	Details	Company	Cat#
Individual: ON-TARGETplus TET1 siRNA Targeted Region:ORF (human)	10nmol	GE Healthcare	J-014635-23-0010
DharmaFECT 1	0.2ml/0.75 ml	Dharmacon	T-2001-01/02
Tet2 siRNA (human)	10uM	Santa Cruz Biotech	sc-88934
WT1 siRNA (human)	10 µM	Santa Cruz Biotech	sc-36846

Table 2. 17 CRISPR Cas9 reagent specifications

Product Name	Details	Company	Cat#
Phusion HF Buffer Pack	6ml	NEB	B0518S
A1460 Wizard Plus SV Minipreps DNA Purification System, 250 preps	250	Promega	A1460
Alkaline Protease Solution	3ml	Promega	A1441
Blasticidin	50mg (5x1ml)	Invivogen	ant-bl-05
Cell Line Nucleofector® Kit L	25 RCT	Lonza	VCA-1005
Lipofectamine® 2000 Transfection Reagent	0.3ml	Invitrogen	11668030
NEB Buffer 2.1	5 ml 10X concentration	NEB	B7202S
NEB Gibson Assembly® Master Mix	10 reactions	NEB	E2611S
One Shot® TOP10 Chemically Competent E. coli	20 reactions	Invitrogen	C404003
Phusion High-Fidelity DNA Polymerase	100units	NEB	M0530S/L
QIAquick Gel Extraction Kit (50)	50	QIAgen	28704
QIAquick Spin Columns (100)	100	QIAgen	28115
Quick-Load® Purple 1 kb DNA Ladder 50ug/ml	1.25ml	NEB	N0552S
Shrimp Alkaline Phosphatase (rSAP)	1000 units/ml	NEB	M0371S (500units)/M0371L (2500units)

Table 2. 18 Cloning reagents specifications

Product Name	Details	Company	Cat#
NEB KasI restriction endonuclease	250 Units 5000U/ml	NEB	R0544S
BsmI	500 units	NEB	R0134S
BsrGI-HF	1000 units (20 000 units/ml)	NEB	R3575S
BsrGI-HF	1000 units (20,000 units/ml)	NEB	R3575S
BssHII	500 units (5000units/ml)	NEB	R0199S
BstX1	1000 units (10,000 units/ml)	NEB	R0113S
DNA Polymerase I (E. coli)	500 units (10000units/ml)	NEB	M0209S
EcoRI-HF	10,000 units	NEB	R3101S
HindIII-HF	10,000 units	NEB	R3104S

Kas	250 units (5000units/ml)	NEB	R0544S
KfII	20 reactions	Thermo Scientific	FD2164
NEB AscIrestriction endonuclease	500 Units 10000U/ml	NEB	R0558S
SpeI-HF	500 units (20 000 units/ml)	NEB	R3133S
XbaI	3000 units	NEB	R0145S

Table 2. 19 Cell culture reagents specifications

Product Name	Details	Company	Cat#
DMEM Dulbecco>=s Modified Eagle Medium (1X) liquid (Low Glucose) [with L-Glutamine 1000 mg/L D-Glucose Sodium Pyruvate] with L-glutamine, D-glucose, sodium pyruvate Invitrogen Gibco	500 ml bottle	Fischer Scientific	31885023
10453302 - Serum, Growth; HyClone; U.S. origin; For cell culture; FBS Replacement; Bovine; Bovine Calf Serum; Supplemented w/chemically defined components to stimulate cell growth and proliferation; 500mL	500 ml bottle	Fischer Scientific	10453302 (SH30541.03)
Blasticidin (solution) ant-bl-05	50mg (5x1ml)	invivogen	ant-bl-05
DMEM/F-12 (1:1) (1X) liquid [without L-Glutamine] without L-glutamine Invitrogen Gibco	500 ml bottle	Fischer Scientific	21331020
HBSS, no calcium, no magnesium	10x500ml	Fischer Scientific	14170138

Table 2. 20 Western blot reagent specifications

Product Name	Details	Company	Cat#
RIPA lysis buffer	50ml	Sigma	R0278-50ml
Amersham ECL Prime Western Blotting Detection Reagent	1000cm3 membrane	GE Healthcare	RPN2232
Blue Loading Buffer Pack	8ml	NEB	B7703S
Bradford reagent	500ml	Sigma	B6916-500ML
BSA Standards 2mg/ml 10x	10x	Sigma	P0834
Nitrocellulose Transfer Membrane- 0.45 um	10x	Abcam	ab133412
Nitrocellulose/Filter Paper Sandwich, 0.45 µm pore size, 8.5 cm x 13.5 cm	16/pack	Thermofisher	LC2006
NuPAGE MES SDS Running Buffer	20x	Thermofisher	NP0002
NuPAGE Transfer Buffer (20X)	20x	Thermofisher	NP0006
NuPAGE™ 4-12% Bis-Tris Protein Gels, 1.5 mm, 10-well	10 gels (1 box)	Invitrogen	NP0335BOX
Protease Inhibitor Cocktail	1ml	Sigma	P8340-1ML
TBS,Tris Buffered Saline 10x	10x	Invitrogen	BP24711

Table 2. 21 General reagent specifications

Product Name	Details	Company	Cat#
10316380 - Hydrochloric Acid, 37%, Certified AR for Analysis, d=1.18; CAS No: 7647-01-0; HDPE plastic bottle; 2.5L	2.5L	Thermo Fisher	10316380

PBS tablets	1 per 1L (10 tablets)	Calbiochem	524650-IEA
PBS tablets each tablet makes 500mL liquid (consists of NaCl, PO4 Buffer, KCl)	100 tablets	Invitrogen Gibco	11510546
Storage Box f. 100 Micro tubes #95.64.997	100 tubes	Sarstedt	95.64.997
Triton™ X-100 for molecular biology	100ml	Sigma	T8787-100ML
Tween-20	100ml	SIGMA	P1379
Ultraclear 1.7ml tubes	500 tubes	Axygen	MCT-175-C

Table 2. 22 Stem cell culture reagent specifications

Product Name	Details	Company	Cat#
2-Mercaptoethanol 50 mM	20ml	Gibco	31350-010
Accutase	100ml	BioLegend	423201
Accutase Solution	100ml	Sigma	A6964
B27 Supplement 50x	10ml	LifeTech/Gibco	17504-044
BSA Soln 7.5%	100ml	Gibco	15260-037
D-(+)-Glucose solution 100 g/L in H2O	100ml	Sigma	G8644
DMEM/HAMS-F12 which comes with Hepes, Sodium Bicarbonate and L-Glutamine	500ml	Sigma	D8437
Laminin	1MG	Sigma	L2020-1MG
Laminin - Cultrex Laminin	5mls at 6mg/ml	R&D Systems (BioTechne)	3446-005-01
MEM NEAA 100x	100ml	Gibco	11140-035
N2 Supplement 100x	5ml	LifeTech/Gibco	17502-048
Pen-Strep	100ml	Gibco	15140-122
Retinoic acid >=98% (HPLC), powder	50mg	SIGMA	R2625
rhFGFbasic	500ug	Peprotech	100-18b
rmEGF	500ug	Peprotech	315-09

Results

Characterisation of DNA Demethylation Activity in Brain Tumour Cell Lines

3.1 Introduction

Techniques such as chromatography, mass spectrometry, methylation-sensitive restriction enzyme digestion and sequencing exhibit high sensitivity and specificity detecting DNA modifications however their universal limitation of aggregating raw biological data and producing an overall readout is attributable to the homogenisation of biological material e.g. cell pellets or tissue during sample preparation (Hotchkiss, 1948;

Wyatt, 1950; Razin & Riggs, 1980; Gruenbaum, Cedar & Razin, 1981; Ehrlich *et al*, 1982; McClelland & Nelson, 1985; Fisher & Giese, 1988; Frommer *et al*, 1992; Song *et al*, 2012; Bassil *et al*, 2013; Yin *et al*, 2013; Tang *et al*, 2015; Guo *et al*, 2017).

Consequently, information such as potential heterogeneity of signal distribution within populations of cells is lost. Immunohistochemistry and immunofluorescence based techniques, albeit not as sensitive as gas or liquid chromatograph or mass spectrometry, possess the advantage of highlighting the signal distribution of DNA modifications in cells when coupled to microscopic examination and analysis (Zluvova *et al*, 2001; Gu *et al*, 2011; Ito *et al*, 2012; Abakir *et al*, 2016; amouroux *et al*, 2016 Lewis *et al*, 2017). These conjugated antibodies can therefore facilitate illustration of any heterogeneity of their presence in samples containing multi-lineage and polyclonal cell types such as cancer cell lines (Nagai *et al*, 2005; Fillmore and Kuperwasser, 2008; Haffner *et al*, 2011; Lawrence *et al*, 2013; Poloni *et al*, 2013; Yang *et al*, 2013; Ramsawhook *et al*, 2017).

Multiple groups have focused heavily on the characterisation of oximC distributions at very early stages of development in mice and humans (Koh *et al,* 2011; Branco *et al*, 2012; Gao *et al*, 2013; Dawlaty *et al*, 2014; Amouroux *et al*, 2016) while others have assessed oximC pattern dynamics during reprogramming from adult somatic to embryonic-like states (Le *et al*, 2011; Gao *et al*, 2013). We sought to address where brain tumour cell line oximC presence falls on the spectrum between healthy early embryonic and adult somatic cells from a human

development context. We employed immunofluorescent detection and confocal microscopy methods to discern subtle oximC distribution patterns not perceivable via conventional DNA modification detection methods such as mass spectrometry, chromatography and sequencing. Due to unavailability and limited access to healthy control adult human brain cells, we employ HeLa cells as a 5caC negative control owing to their lack of 5caC possession and as a cancer cell line control. We also utilise human foetal neural stem cells and RBLPAT human iPSC cell lines as 5hmC and 5caC positive controls as is conventionally performed by other groups (Le *et al*, 2011; Gao *et al*, 2013; Yoo & Bieda, 2014; Ye *et al*, 2016).

3.2 Detection of OximC Presence

To determine whether active DNA demethylation was observable, we implemented immunofluorescence based interrogation of cell lines and tissues coupled with confocal microscopic analysis to examine the distribution, profile and magnitude of oximC presence as an initial parameter for characterisation of DNA demethylation activity.

3.2.1 5caC is Detectable in Mammalian Foetal but not Adult Brains

OximCs demonstrate a significant reduction in magnitude going from abundant enrichment in pluripotent stem cells during embryogenesis

(Hemberger, Dean & Reik, 2009; Seisenberger *et al*, 2013; Messerschmidt, Knowles & Solter, 2014) to diminished but stable 5hmC in adult somatic tissues (Munzel *et al*, 2010; Swagierczak *et al*, 2010; Jin *et al*, 2011; Branco, Ficz & Reik, 2012). To contextualise observations of 5hmC and 5caC presence from a developmental standpoint and establish positive and negative controls respectively, healthy mammalian brains, 22 post coital week (pcw) old human foetal brain and 6 week old adult mouse brain sagittal sections were immuno-stained and imaged using confocal microscopy (fig 3.1A-B).

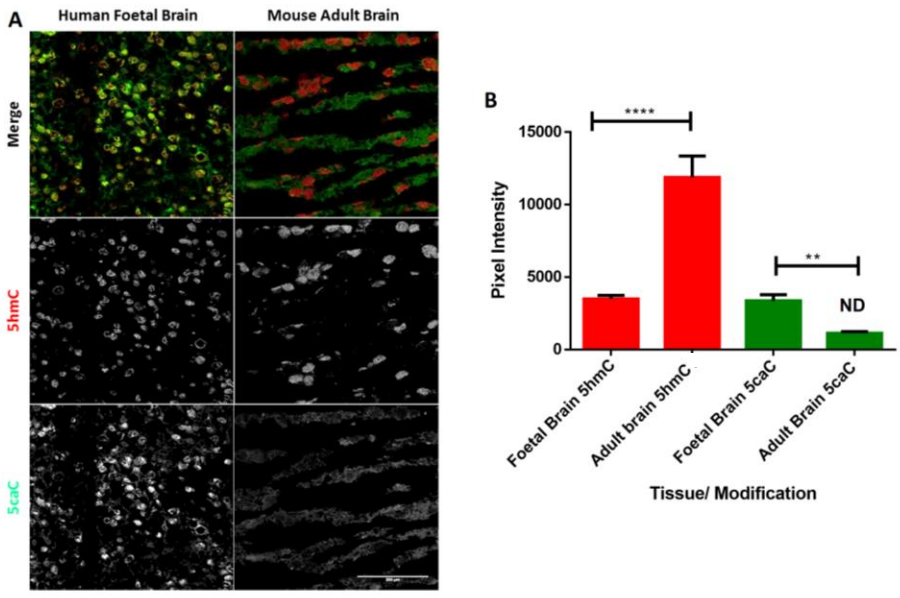
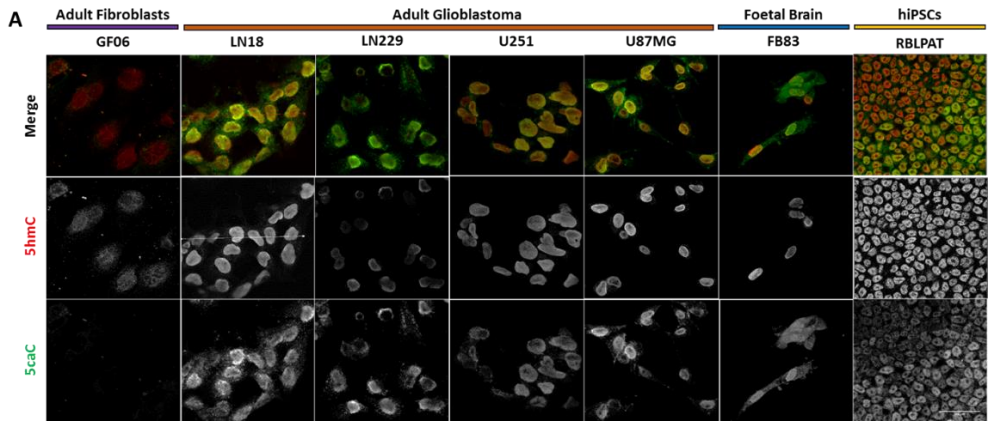


Figure 3.1 OximC presence in the mammalian brain. (A) Immunofluorescent detection and (B) quantification of 5hmC (red) and 5caC (green) signal intensities in foetal and adult mammalian brains. Scale bar represents 200µm and all staining and images acquired at identical settings. Statistical analysis was performed using unpaired, two tailed Student's t-test, mean ± SD. ND denotes not-detected, ** $p < 0.001$; **** $p < 0.0001$.

Interrogation of human foetal and mouse cerebrums revealed the presence of 5hmC and 5caC in human foetal sections. In the adult mouse, only 5hmC was detectable above negligible magnitudes (>1000 pixels) as determined by secondary only antibody probing (appendix 1) (fig 3.1A). The 5hmC signal intensity was significantly higher in adult compared to the foetal brain (11880.4 versus 3491.45, $p < 0.0001$) whilst only 5caC was detectable (3363.8) in the foetal brain, displaying an opposite profile to 5hmC distributions (fig 3.1B). The detectability of 5hmC and 5caC in these tissues indicated the existence of TET protein mediated DNA demethylation active at these developmental time points.

3.2.2 Adult Glioblastoma and Foetal Brain Cell lines Display Similar 5caC Signal Magnitudes

To understand where glioblastoma cells fall on the spectrum of 5caC presence (fig 3.2A-C), adult human fibroblasts (GF06) and induced pluripotent stem cells (RBLPAT) were used as negative and positive controls respectively.



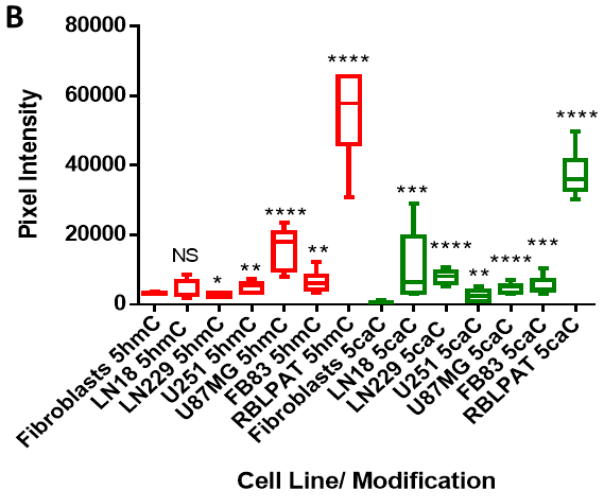


Figure 3.2 Spectrum of oximC magnitudes between cell lines. (A) Immunostaining of 5hmC and 5caC signal presence ranging across healthy adult somatic fibroblasts, adult glioblastoma and foetal and pluripotent cell lines. Scale bar represents 200µm, staining and images acquired at identical settings. (B) Quantification of 5hmC (red) and 5caC (green) signal intensities across cell lines. Unpaired two tailed Kruskal-Wallis test performed relative to GF06 fibroblasts. Mean ± SD, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$, n = 3.

Cell lines were used as oppose to tissues sections to account for discrepancies between immunohistochemical (IHC) and immuno-cytochemical (ICC) staining methods owing to IHC antigen retrieval techniques and artificial *in vitro* culture environments (fig 3.2A). Glioblastoma LN229, U251 and U87MG, foetal brain FB83 and iPSC RBLPAT lines all displayed significantly higher 5hmC signal intensity relative to fibroblasts ($p = 0.049$, $p = 0.006$, $p < 0.0001$, $p = 0.0047$ and $p < 0.0001$ respectively) however no significance was observed in the LN18 cell line ($p = 0.337$). All cells showed very significantly reduced 5hmC levels relative to RBLPAT (all $p < 0.0001$) with smallest disparity existing between RBLPAT and U87MG ($p = 0.042$). All cell lines

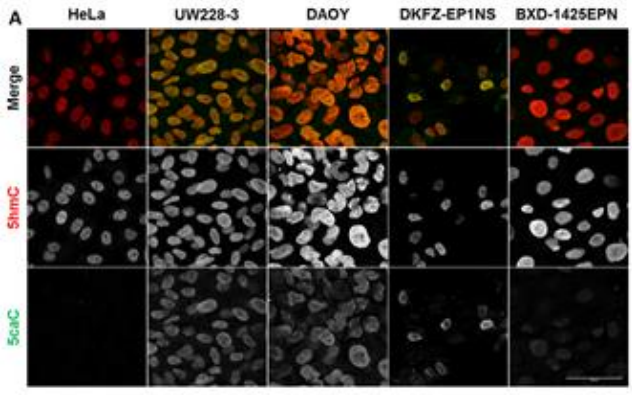
exhibited significantly higher 5caC signal intensity relative to the negligible signal from adult fibroblasts with LN18 possessing the highest pixel intensity of the GBM cell lines (102089.3, $p = 0.0006$). RBLPAT iPSC illustrated dramatic and highly significant ($p < 0.0001$) 5hmC and 5caC presence relative to all compared cell lines with 3 fold higher 5hmC signal than U87MG and 16 fold higher than GF06. Similarly, 5caC was detected at 3.6 fold higher magnitudes in RBLPAT than LN18 and 54 fold greater than in GF06 adult fibroblasts (fig 3.2B).

Further analysis of the 5hmC and 5caC staining distribution revealed no significant population variance in 5hmC signal between GF06 and any GBM cell line nor FB83 ($p > 0.05$) however very high significance was detected between GF06 and RBLPAT ($p < 0.0001$). No significance in 5hmC signal variance was detected when comparing FB83 to GBM cell lines ($p > 0.05$) with the exception of U87MG ($p = 0.00012$). RBLPAT displayed very high significance compared to all analysed cell lines ($p < 0.0001$). Interestingly, GBM cell lines showed similarly non-significant population variance between themselves except for U87MG with which LN18 and LN229 showed significant population variance ($p = 0.012$ and $p < 0.0001$ respectively). Fibroblasts displayed significant 5caC population variance with LN18, LN229 and RBLPAT ($p = 0.0006$, $p = 0.0006$ and $p < 0.0001$ respectively). No significant population variation was observed between GBM cell lines and FB83 ($p = 0.061$) or between GBM cell lines themselves ($p = 0.057$). Additionally, only a small overlap between GF06 fibroblasts and U251 ($U = 2.00$) whereas large overlaps were observed between all GBM cell lines LN18 ($U = 31.00$), LN229 (U

= 13.00), U251 (U = 10.00) and U87MG (U = 33.00) and FB83 (fig 3.2B). No overlaps were observed between either GBM or FB83 cell lines with RBLPAT (U = 0). These results taken together illustrate unique 5hmC and 5caC staining patterns for GBM cell lines which do not statistically overlap with 5hmC and 5caC signal profiles of foetal brain, fibroblast or pluripotent cell lines.

3.2.3 5caC is Detectable in Paediatric Medulloblastoma and Ependymoma Cell Lines

In light of the discovery of similar 5caC signatures in adult glioblastoma cell lines and foetal brain cells, we questioned whether such findings would be consistent with tumours of the developing human. To this end paediatric medulloblastoma DOAY and UW228-3 and ependymoma BXD1425EPN and DKFZ-EP1NS cell lines were probed for 5caC presence (fig 3.3A-E).



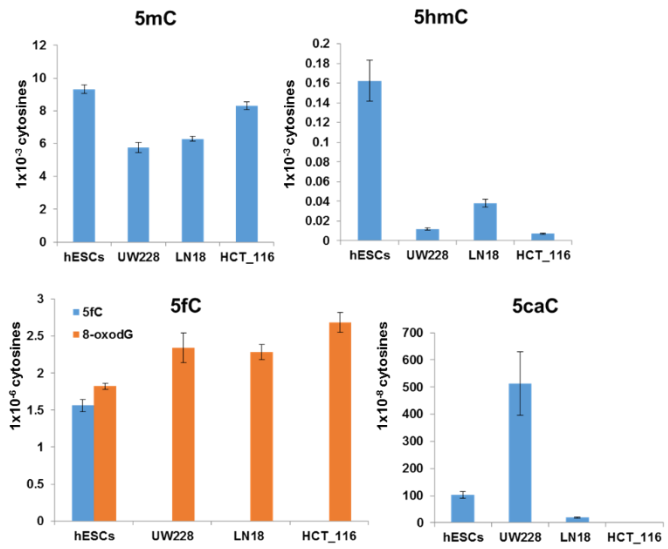


Figure 3.3 Detection of oximCs in paediatric brain tumour cell lines. (A) Immunostaining of 5hmC, 5fC and 5caC in medulloblastoma and ependymoma cell lines. Scale bar represents 200µm, staining and images acquired at identical settings. (B-E) Mass spectrometric quantification of oximC magnitudes across cell lines. Mean ± SD, N =2.

Cervical cancer cell line HeLa was used as a control for non-brain lineage neoplasms. All medulloblastoma and ependymoma cell lines displayed enrichment for both 5hmC and 5caC but the latter was not detectable in HeLa. While the 5hmC staining distribution appeared homogeneously across nuclei, 5caC exhibited a speckled, punctate pattern, which bordered the nuclear periphery in DKFZ-EP1NS cells. To validate these findings, pellets of glioblastoma LN18, medulloblastoma UW228-3, human embryonic stem cells Hues7 and human colorectal cancer HCT_116 cell lines were analysed by mass spectrometry for oximC species (fig 3.3B-E). Comparable levels of 5mC exist in LN18, UW228-3, HCT_116 and hESCs with the latter possessing the greatest magnitudes. 5hmC levels were again highest in hESCs and levels were

comparable amongst UW228-3, LN18 and HCT_116. Enrichment of 5fC was observed in hESCs together with 8-hydroxy-2'-deoxyguanosine (8-oxodG), an indirect co-factor influencing TET1 mediated 5mC oxidation and marker of oxidative stress (Allgayer *et al*, 2016). Abundant levels of 8-oxodG comparable to those in hESCs were observed in UW228-3, LN18 and HCT_116 but no 5fC was detectable in these lines. Unexpectedly, UW228-3 demonstrated 5caC magnitudes 5-fold and 28.5-fold higher than hESCs and LN18 respectively. No 5caC was detectable in HCT_116 cell lines. Thus mass spectrometry results verified immunohistochemical staining in all cases with the exception of UW228-3 which demonstrated an anomalous signature of 5caC, compounded by lower 5mC and 5hmC species abundance than both hESCs and LN18 cell lines.

3.2.4 Elevated 5caC Levels Correlate with High *TET1* Expression in Medulloblastoma and Ependymoma Cell Lines

Quantification and further characterisation of 5hmC and 5caC distributions in paediatric cell lines revealed similar overlapping trends in 5hmC profile across nuclei amongst the different cell lines (fig 3.4A).

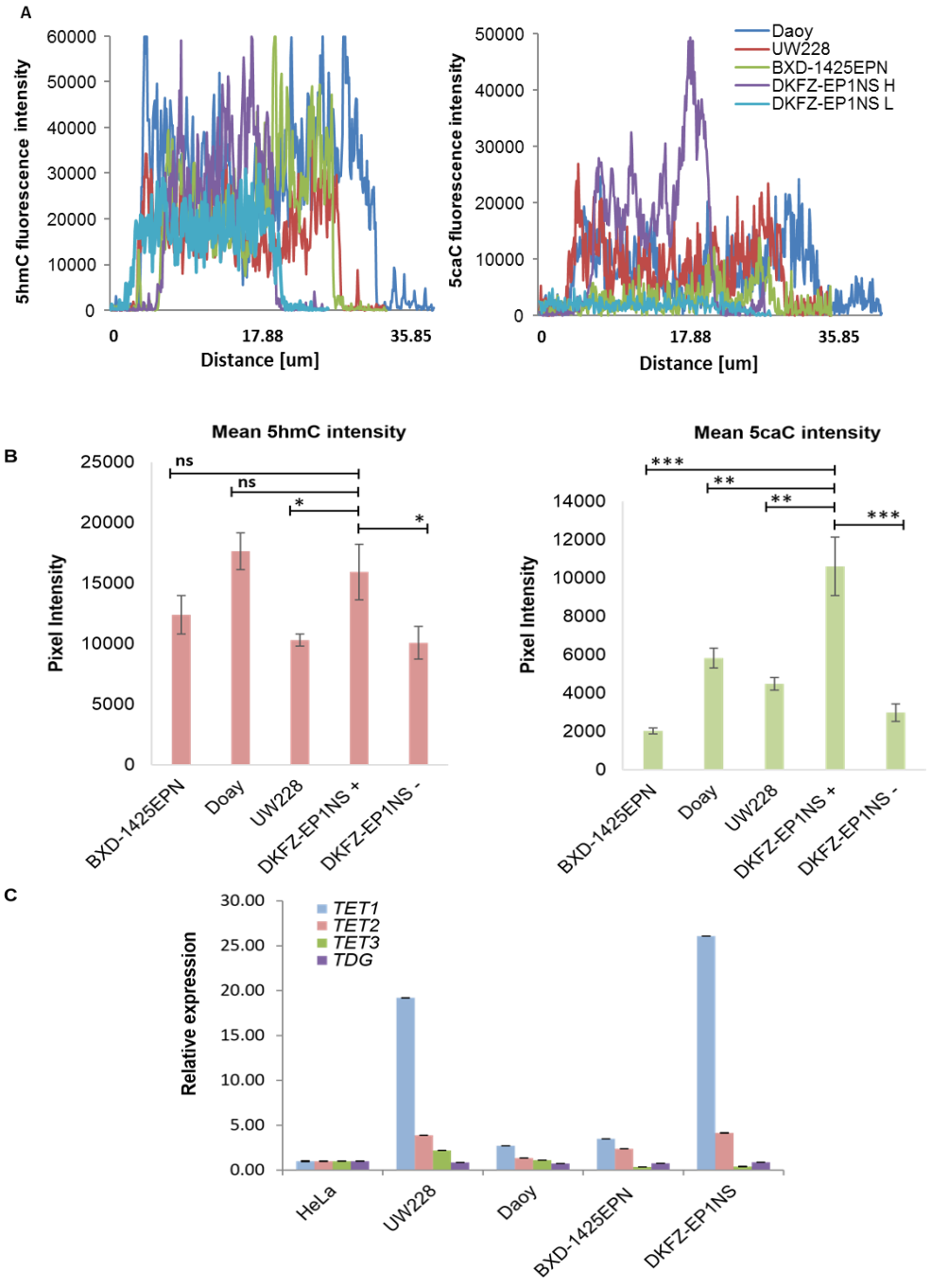


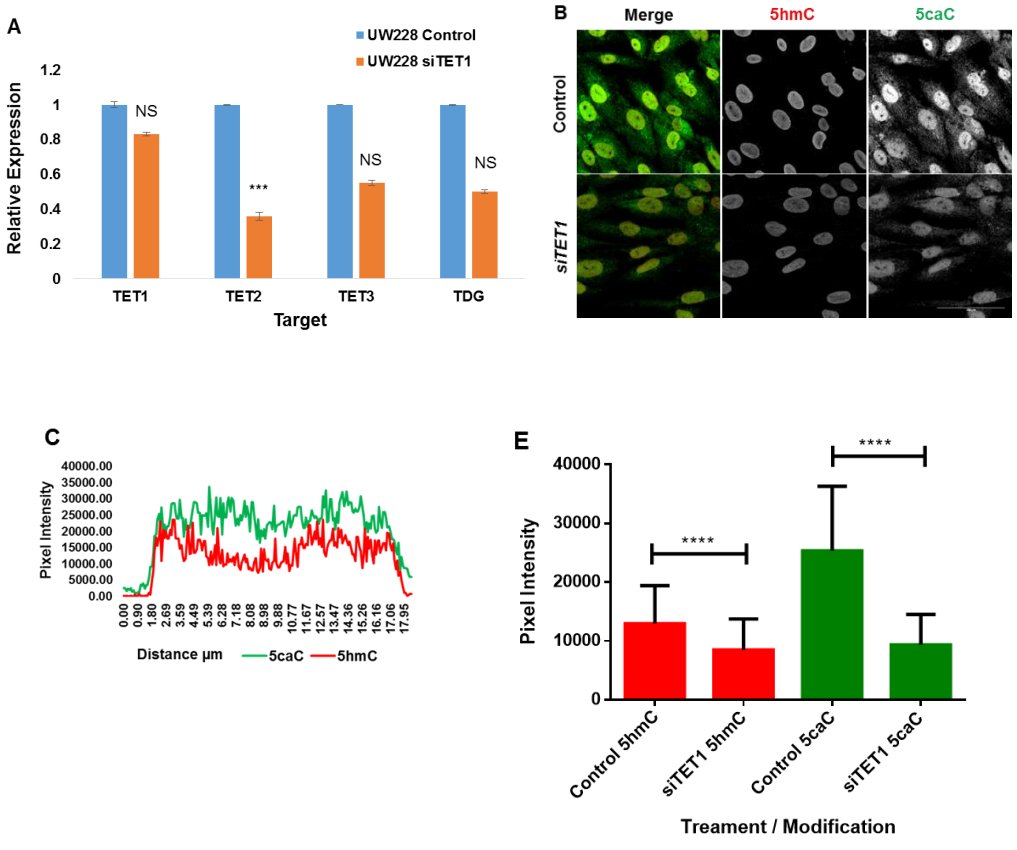
Figure 3.4 OximC profiles between cells are consistent with TET1 expression. (A) Confocal profile analysis of 5hmC and 5caC signal magnitudes and distributions. (B) Quantification of 5hmC and 5caC signal magnitudes relative to DKFZ-EP1NS 5caC positive cells. Unpaired two tailed Student's t-test, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$, n = 3. (C) QPCR analysis of DNA demethylation machinery transcriptional levels in paediatric brain tumour cell lines relative to HeLa cells, Mean ± SD, n =2.

Analysis of 5caC distribution profiles across cell lines illustrated similar trends in spatial pattern and magnitude however DKFZ-EP1NS cells could be segregated into 5caC 'high' or 5caC 'low' presenting cells on the basis of superficial image analysis. This entailed selecting cells in which 5caC staining was detectable and measuring the 5caC and 5hmC signal magnitude separately to 5caC undetectable cells. Profile analysis concordantly depicted distinct 5caC distribution signatures for DKFZ-EP1NS high and low cells which did not overlap in spatial pattern. Quantification of 5hmC and 5caC signal intensities (fig 3.4B) demonstrated comparable magnitudes amongst medulloblastoma and ependymoma lines with no significance observed between BXD-1425EPN, DOAY and DKFZ-EP1NS+ ('high') cells ($p < 0.05$). UW228-3 and DKFZ-EP1NS- ('low') cells possessed significantly lower 5hmC magnitudes relative to DKFZ-EP1NS+ cells. All samples displayed highly significantly different 5caC magnitudes compared to DFKZ-EP1NS+ cells ($p < 0.01$ – $p < 0.001$). Notably, similar 5caC magnitudes were observed between BXD-1425EPN and DKFZ-EP1NS- lines, both of which exhibit moderate 5caC levels relative to DKFZ-EP1NS. Therefore the 5caC was not depleted from DKFZ-EP1NS- cells as the '-' symbol may have suggested on the basis of microscopic examination but rather 'masked' by high 5hmC levels in these cells. To verify whether the enriched 5hmC and 5caC presence observable in paediatric brain tumours was consistent with DNA demethylation activity, gene expression analysis was performed (fig 3.4C). High *TET1* and *TET2*

expression in all medulloblastoma and ependymoma cell lines relative to HeLa were predominately observed with UW228-3 and DKFZ-EP1NS in particular possessing 19- and 26-fold higher *TET1* expression levels than HeLa. *TET3* and *TDG* did not exhibit significantly altered expression levels to HeLa, suggesting a substantial involvement of TET1 in oximC presence in these brain tumour lines.

3.2.5 TET1 siRNA Knockdown Correlates with Diminished OximC Enrichment in UW228-3 Medulloblastoma Cell Lines

We next examined the extent to which 5hmC and 5caC presence in UW228-3 was dependent on TET1 activity owing to its high expression in this cell line (fig 3.5A-D).



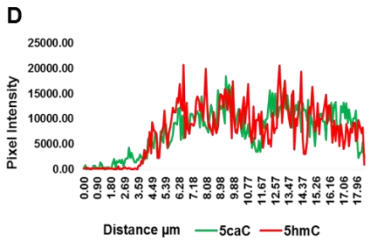


Figure 3.5 TET1 siRNA mediated knockdown may reduce 5caC magnitudes via down-regulating *TET2* expression. (A) QPCR analysis of DNA demethylation machinery gene expression following *TET1* knockdown. Unpaired two tailed Mann-Whitney tests were performed, NS $p \geq 0.05$, * $p < 0.001$. (B) Immunostaining of 5hmC and 5caC signal intensities on control vs si*TET1* cells, scale bar represents 200µm, staining and images acquired at identical settings. (C-D) 5hmC and 5caC signal distribution profiles of control (C) vs si*TET1* (D) cells. (E) 5hmC and 5caC signal quantification, unpaired two tailed Student's t-test was performed, mean ± SD, **** $p < 0.0001$, $n = 3$.

A modest reduction in *TET1* expression (17%) was achieved via si*TET1* mediated *TET1* knockdown resulting in decreased fold change in *TET2*, *TET3* and *TDG* gene expression by 64%, 45% and 50% respectively. Interestingly, only *TET2* reduction was significant ($p = 0.0009$) while *TET1*, *TET3* and *TDG* gene expression reduction was trivial ($p = 0.292$; $p = 0.118$ and $p = 0.096$ respectively). Analysis of *siTET1* transfected cells revealed significant reductions in both 5hmC ($p < 0.001$) and 5caC ($p < 0.05$) signal intensities relative to control scramble transfected cells. Consequently, the signal distributions of 5hmC and 5caC which, in control cells, exhibited distinct peaks and troughs, produce overlapping 5hmC and 5caC signal profiles both in magnitude and spatial distribution.

3.2.6 Adult Glioblastoma Cell Lines Possess a Unique DNA Demethylation Transcriptional Profile

Owing to DNMT-generated 5mC functioning the substrate for TET protein-mediated production of oximCs, We next asked whether the 5caC enrichment observable in brain tumour cell lines can be attributable to elevated *TET* gene expression coupled with reduced *TDG* gene expression or increased activity of DNA methyltransferase enzymes producing an abundance of 5mC substrates for TET protein mediated oxidation. Transcriptional analysis (fig 3.6A-B) investigating DNA methylation, demethylation and transcription factor expression relative to adult fibroblasts was performed and demonstrated DNA methyltransferase genes expression reduced by one to two orders of magnitude (e.g. *DNMT1* = 0.08, *DNMT3A* = 0.01 and *DNMT3B* = 0.1) in LN18, LN229, U251 and U87MG, comparable to levels recorded in RBLPAT. *TET1* expression levels were also reduced by at least one order of magnitude relative to fibroblasts in all GBM cell lines except U251 which possessed 3.04 fold higher expression. RBLPAT pluripotent stem cells possessed 31-fold higher *TET1* expression than fibroblasts and two orders of magnitude higher than GBM cells. Interestingly, *TET2* and *TET3* were the predominant *TET* genes expressed in GBM cells with highest levels recorded in U251 at 6.32- and 23.63-fold higher levels than fibroblasts.

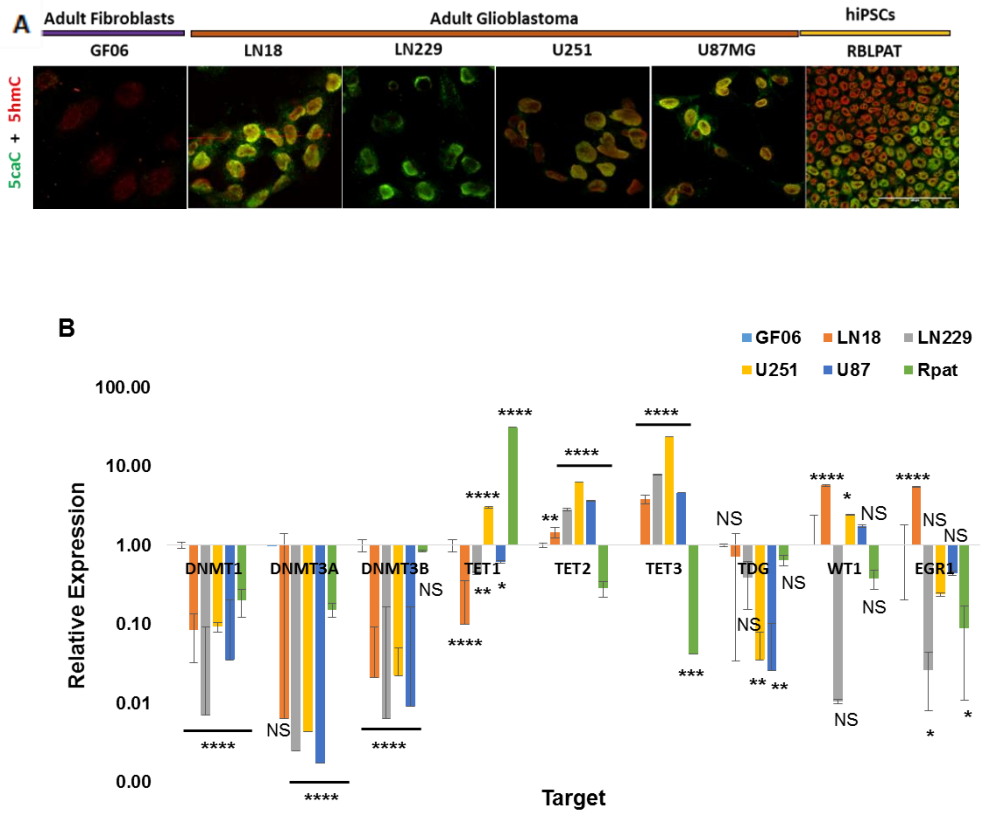


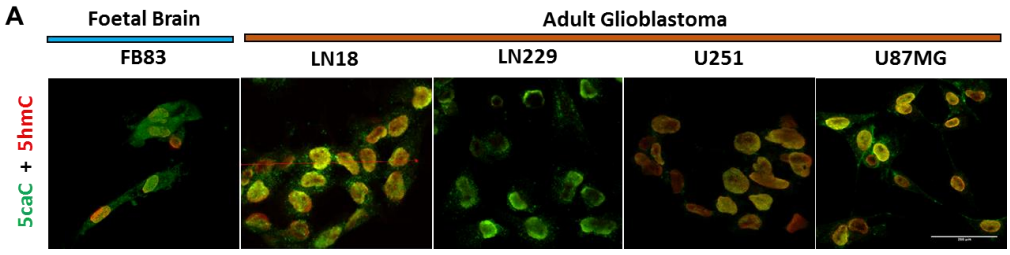
Figure 3.6 Transcriptional signature of DNA methylation / demethylation across healthy adult somatic – pluripotent and malignant cell lines. (A) OximC staining across cell lines, scale bar represents 200µm, staining and images acquired at identical settings. (B) QPCR analysis of DNA methylation, demethylation, *WT1* and *EGR1* target genes across cell lines relative to 5caC negative GF06 fibroblasts. Unpaired two tailed Kruskal-Wallis test, mean ± SD * $p < 0.05$; ** $p < 0.01$; *** $P < 0.001$, **** $p < 0.0001$, $n = 3$.

All GBM and RBLPAT lines had substantial reduced *TDG* levels, the highest among GBM lines pertaining to U251 (0.04-fold lower) and comparable to RBLPAT (0.03-fold lower). Transcription factor and

putative TET2 interacting partner *WT1* was elevated in all GBM cell lines except LN229, the highest expression occurring in LN18 (5.7-fold) and the lowest in U87MG (1.7-fold). Early Growth Response gene 1 *EGR1* which shares consensus DNA binding motifs with WT1 displayed similar expression levels in LN18, LN229 and RBLPAT to *WT1* in these cell lines but opposite profiles in U251 and U87MG. Overall these findings describe a molecular landscape in GBM cell lines of *TET2 & 3* gene transcription higher than adult somatic fibroblasts cells and lower than that of fibroblast-derived iPSCs while simultaneously possessing DNA methyltransferase transcription far below fibroblasts levels.

3.2.7 Foetal Brain and Adult GBM Cell Lines show OximC Parity but Transcriptional Discordance

Considering the similarity in 5hmC and 5caC magnitude between FB83 and GBM cell lines (fig 3.7A-B), transcriptional analysis was performed to determine the level of contribution of DNA demethylation gene expression to this pattern (fig 3.7C).



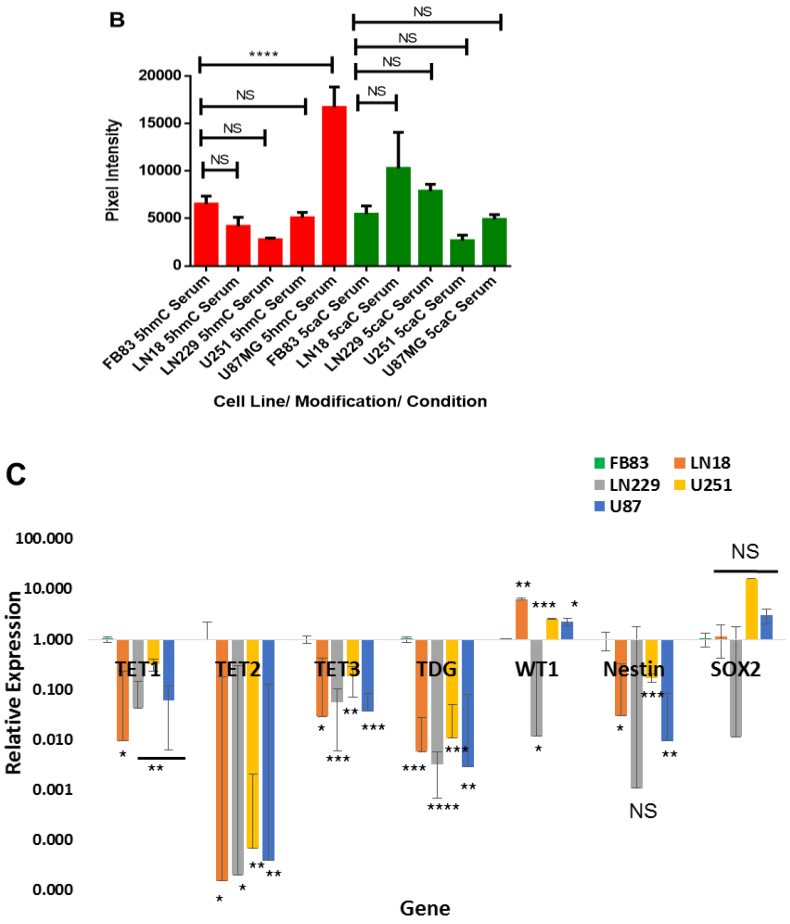


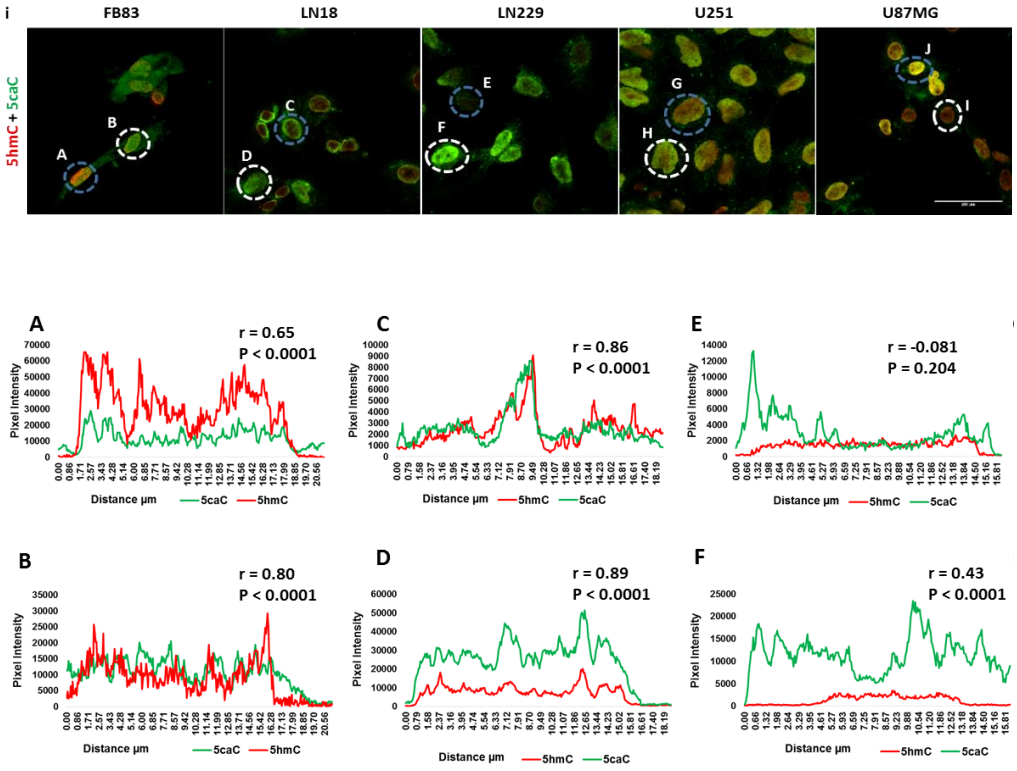
Figure 3.7 Characterisation of GBM cell line DNA demethylation profiles with respect to foetal brain cells. (A) Immunofluorescent detection of 5hmC and 5caC signals in FB83 vs GBM cell lines. Scale bar represents 200µm, staining and images acquired at identical settings. (B) OximC staining intensities of GBM cell lines relative to FB83 foetal brain cells, unpaired two tailed Kruskal-Wallis test. (C) QPCR analysis of DNA demethylation machinery, *WT1* and stem cell target genes, unpaired, two tailed Kruskal-Wallis test, mean ± SD, NS $p \geq 0.05$; * $p < 0.05$; ** $p < 0.01$; * $p < 0.001$; **** $p < 0.0001$, n = 3.

DNA demethylation machinery associated genes *TET1*, *TET2*, *TET3* and *TDG* all showed dramatic reduction in GBM cell lines relative to FB83 with fold reductions ranging from 0.31-fold for *TET1* in U251 to 0.0001-fold in LN18. *WT1* was expressed at 6.36-fold, 2.57-fold and 2.23-fold higher levels than FB83 but LN229 exhibited marked reduction at 0.01-fold less. The significantly reduced ($p < 0.05$ – $p < 0.001$) *TET* gene

expression in GBM cell lines relative to FB83 may have contributed to a diminished 5hmC and 5caC presence in the lines.

3.2.8 Foetal Brain and Adult Glioblastoma Cell Lines Display Heterogeneous Cellular 5hmC and 5caC Distributions

Examination of 5hmC and 5caC immunostaining patterns across FB83 and all GBM cell lines highlighted apparent cellular heterogeneity with regards to 5caC levels (fig 3.8(i)A-J).



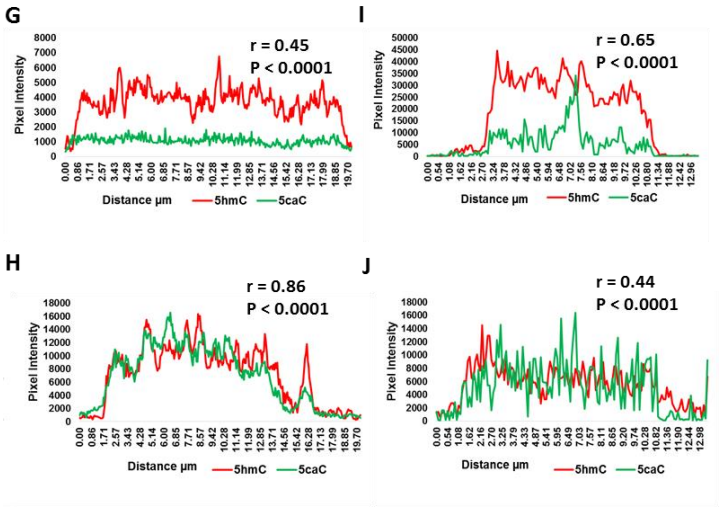


Figure 3.8 Examination of heterogeneous oximC staining distributions within cell lines. (i) 5hmC (red) and 5caC (green) immunofluorescent staining in FB83 and all GBM cell lines. Cells encircled by blue and white dashed circles indicate cells interrogated by profile analysis (A-J). Circled lettered cells correspond to lettered profiles of 5hmC and 5caC signal scale and distribution across the length of each circled cell. Scale bar represents 200µm, staining and images acquired at identical settings. Spearman's correlation performed on all cells, n =3.

All analysed cell lines displayed varying degrees of 5caC enrichment while some cells appear 5caC depleted (fig 3.8i). Disparate 5caC distribution between cells for each cell line was highlighted (encircled cells) and their 5hmC and 5caC spatial distribution and magnitude profiles were plotted. The profiles illustrate distinct 5hmC and 5caC signatures within and between the cell lines with all cell lines except LN229 demonstrating highly significant ($p < 0.0001$) positive correlations between 5hmC and 5caC distribution (Pearson's correlation, r). A negative and non-significant correlation ($r = -0.081$, $p = 0.204$) was observed in low 5caC staining LN229 cells compared to highly significant positively correlating ($r = 0.43$, $p < 0.0001$) high 5caC staining cells (fig

3.8E & F respectively). The strongest coincidence of 5hmC and 5caC spatial distribution and magnitudes occurred in FB83 (B), LN18 (C) and U251 (H) cells ($r = 0.8$, $r = 0.86$, $r = 0.86$, $p < 0.0001$ for all values respectively). Curiously, LN18 (D) cells reported the highest 5hmC and 5caC correlation yet the signal distributions do not overlap but do exhibit the same trends.

3.2.9 Significant Variation in 5hmC Signal Intensity Produce a Perceived Effect of 5caC Cellular Heterogeneity

Further analysis to understand the extent of 5hmC and 5caC signal variation between 'high' and low' 5caC enriched cells (lettered cells) within each cell line (fig 3.9A-E) revealed highly significant ($p < 0.0001$) differences in the mean 5hmC signal intensities and also the variation of signal intensity distributions. Comparisons within cell lines LN18, LN229 and U251 demonstrated that actual discrepancies in 5caC magnitudes contributed to observed degrees in 5caC presence. While LN18 and U251 'high' and 'low' cells within each line possess significantly distinct mean 5caC signal intensities and additionally for LN229, the skewed distribution in 5caC 'high' cells compounded a perceived heterogeneity ($p < 0.0001$). Significantly high variation and mean signal intensity of 5hmC masks 5caC signal in U87MG cell lines and erroneously depicts FB83 cells as 'low' in 5caC although it displays significantly higher signal intensity that the perceived 'high' cells.

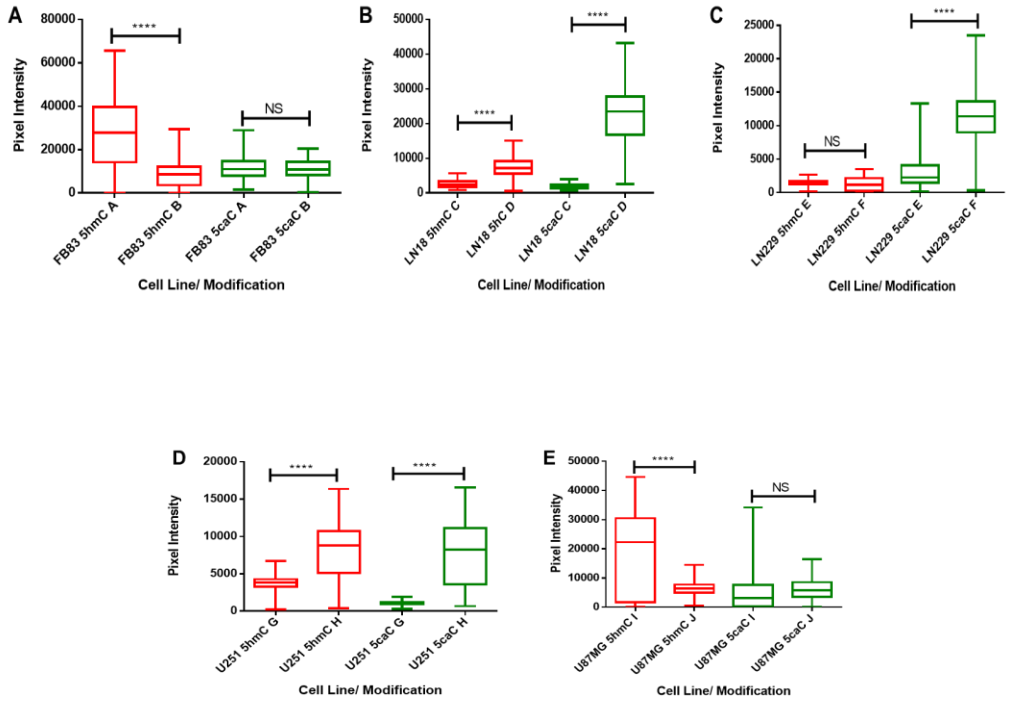
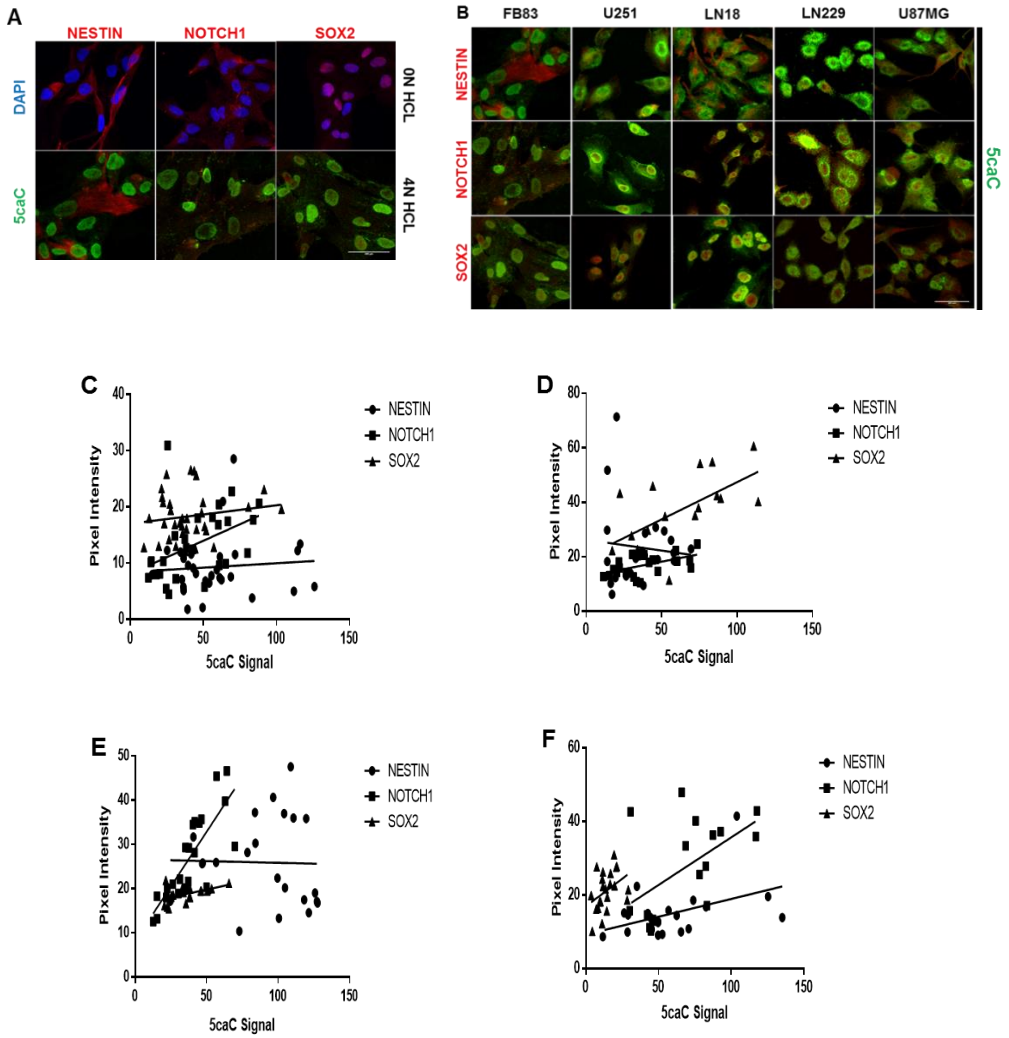


Figure 3.9 Scale of oximC magnitude variation between circled cells with FB83 and GBM cell lines. (A-E) Box plots of 5hmC and 5caC signal distribution within individual cells showing mean, interquartile range, minimum and maximum. Mann-Whitney tests, NS $p \geq 0.05$; **** $p < 0.0001$, n = 3.

3.2.10 5caC Signal Intensity Does Not Correlate with Neural Stem Cell Markers in Foetal Brain or Glioblastoma Cell Lines

To understand the variation of 5hmC and 5caC staining intensities, co-staining of 5caC with neural stem cell markers was performed to determine whether 5caC intensity segregated with intrinsic cellular properties such as stemness (fig 3.10A-G). Significant positive correlation of 5caC signal intensity with Nestin ($r = 0.41$, $p < 0.0001$), SOX2 ($r = 0.33$, $P < 0.0001$) but negatively correlated with Notch1

intensity (r = -0.19, p <0.0001) was observed in human foetal neural stem cell line FB83. The GBM cell lines produced significant profiles of correlation with only one or two any of the stem markers but not all three i.e. LN18 correlated significantly with NOTCH1 (p = 0.013) and SOX2 (p = 0.0093), LN229 with NOTCH (p < 0.0001) and SOX2 (p = 0.014), U251 with NESTIN (p = 0.045) and NOTCH1 (p = 0.014) and U87MG with NOTCH1 (p = 0.017). The profile established by FB83 represents the pattern of 5caC distribution within a stem cell line despite not strongly correlating with all stem markers.



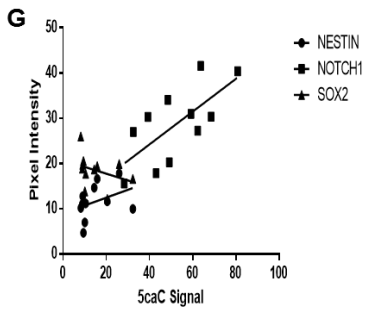


Figure 3.10 Correlation of 5caC signal intensity with neural stem cell marker expression. (A) FB83 immunostaining at 0N HCL and 4N HCL conditions. (B) 5caC and neural stem cell markers immunofluorescent staining between FB83 and GBM cell lines. Images acquired at identical settings, scale bar represents 200µm. (C-G) Linear regression analysis scatterplots of 5caC and neural stem cell marker signal correlation per cell lines FB83, LN18, LN229, U251 and U87MG respectively, n = 3.

3.3 Modulation of OximC Nuclear Distribution

Genome-wide bisulfite and methylated DNA immunoprecipitation sequencing methods have facilitated the discovery of 5mC-rich loci (Li & Tollefsbol, 2011). The preponderance of 5mC enrichment at promoters and enhancers of silenced genes has implicated this mark as transcriptional repressive (Weber *et al*, 2007). Considering 5mC is a substrate for TET-protein mediated oxidation and generation of oximCs, the presence of which have been associated with transcriptional activation. We postulated that the magnitude and distribution of oximCs could be influenced by forcibly modulating cellular transcriptional

programmes allowing transience or permanence of these marks to be determined.

3.3.1 Neurosphere Enrichment Conditions Alter OximC Distribution in Adult Glioblastoma Cell Lines

To address the possibility of plasticity of 5hmC and 5caC presence in foetal brain and GBM cells, cell lines were cultured in neurobasal conditions to promote neural stem cell proliferation and hence enrich for stemness in the population (fig 3.11A-B). Neurobasal conditions promoted neurosphere formation in FB83 and all GBM cell lines on adherent surface. All cell lines formed spheroids under non-adherent surfaces in neurobasal conditions. Morphological indicators of stemness were corroborated through detection of neural stem cell marker (NESTIN, NOTCH1 and SOX2) expression. Sporadic and variable 5hmC and 5caC staining was observed in LN18 and LN229 neurospheres in contrast to U251 and U87MG which demonstrated more homogeneous staining. FB83 did not form spheroids on adherent surfaces and possessed uniform 5hmC and 5caC nuclear staining.

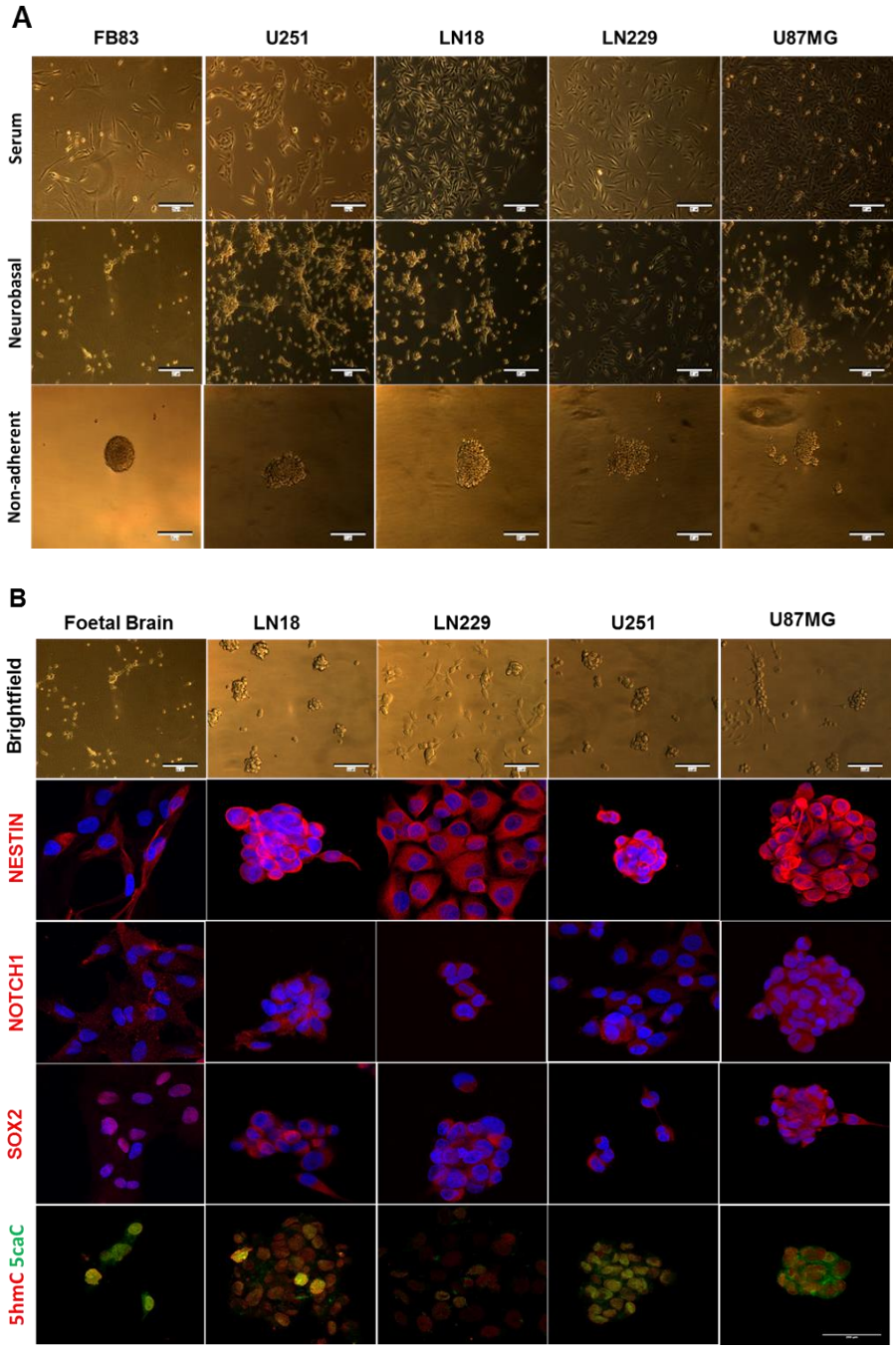
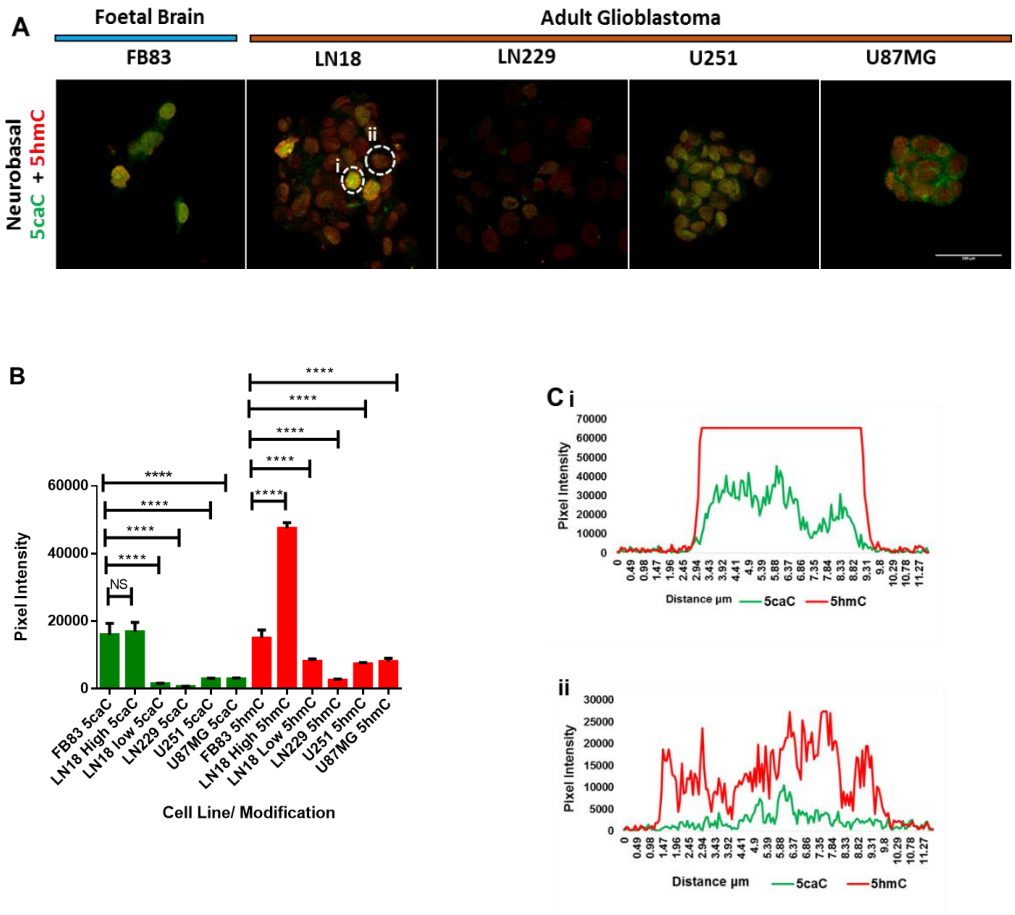


Figure 3.11 Assessment of oximC presence in neural stem cell enriched conditions. (A) Brightfield images of FB83 and GBM cell lines cultured in serum vs adherent and non-adherent surface neurobasal conditions for assessment of neurosphere / gliomasphere formation. Scale bar represents 25µm. (B) Brightfield images and Immunostaining of FB83 and GBM cells for neural stem cell markers and oximC presence in neurobasal conditions. Scale bar represents 25µm (Brightfield) and 200µm (immunofluorescence), n = 3.

3.3.2 Adult Glioblastoma Cell Lines Exhibit Reduced OximC Presence which Correlates with Depleted *TET1* and *TET3* Expression Relative to Foetal Neural Stem Cells

Quantification of 5hmC and 5caC signal intensities observed under neurobasal conditions (fig 3.12A-B) revealed that both 5hmC and 5caC levels were significantly reduced in all GBM cell lines relative to FB83 neural stem cells. Lowest 5hmC magnitudes belonged to LN229 (82.7% lower, $p < 0.0001$) and LN18 exhibited highest 5hmC (45.7% lower, $p = 0.0011$). U251 showed highest 5caC levels (81.6% lower, $p = 0.0047$). Lowest 5caC quantification was reported in LN229 cells (95.5% lower, $p < 0.0001$). No significant difference was observed between 5hmC or 5caC magnitudes between GBM cell lines LN18, U251 or U87MG however LN229 5hmC and 5caC magnitudes were highly significant compared to all other GBM cell lines ($p < 0.0001$). Comparison of highly enriched 5hmC and 5caC cells (encircled by dotted lines) to low abundance cells In LN18 (fig 3.12Ci-ii) highlighted their distinct nuclear 5hmC and 5caC signal profiles. Magnitudes of 5hmC are present at greater than or equal to the confocal laser detection limit and thus appear saturated, producing a plateaued profile spanning the nucleus of the highly enriched cell. The 5caC profile is very significantly lower than 5hmC in this cell ($p < 0.0001$) and displayed a variable spatial distribution in its abundance across the nucleus. The less abundant cell exhibits a varied spatial distribution and scale of 5hmC and 5caC level with 5hmC signal intensities significantly higher than 5caC ($p < 0.0001$). The

reductions in 5hmC and 5caC staining intensities in GBM cell lines relative to FB83 neural stem cells correlated with two to four orders of magnitude lower *TET1* and *TET3* and *TDG* gene expression levels (fig 3.12D). All substantial *TET2* gene expression was recorded in all GBM cell lines relative to FB83, most notably in LN18 and LN229 with levels over 100-fold higher, U87MG showed only marginal increase in *TET2* expression. Interestingly, *WT1* expression was consistent with the profile of the DNA demethylation genes. Curiously, LN18, U251 and U87MG depicted *SOX2* expression levels 5-10 fold higher than foetal neural stem cells, with U251 exhibiting greater than 10-fold elevation.



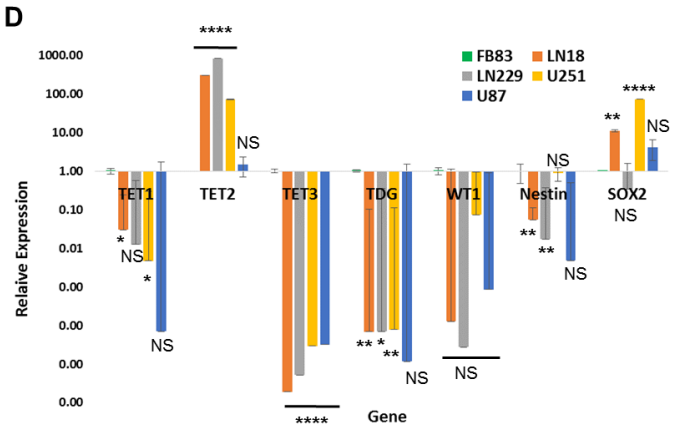


Figure 3.12 Characterisation of DNA demethylation activity in FB83-GBM cell lines under neurobasal conditions. (A) OximC distribution within neurospheres / gliomaspheres. LN18 cells marked (i) and (ii) are indicative of distinct variation in oximC magnitudes within cell lines. Scale bar represents 200µm. (B) Quantification of neurobasal condition oximC levels between and within (LN18) cell lines. LN18 'high' and 'low' pertain to cells (i) and (ii) respectively. Unpaired two tailed Kruskal-Wallis test, mean ± SD, NS $p > 0.05$; **** $p < 0.0001$. (Ci-ii) Confocal oximC profiles of LN18 'high' (i) and 'low' (ii) cells. (D) QPCR analysis of DNA demethylation machinery genes, *WT1* and neural stem cell markers, mean ± SD, unpaired, two tailed Kruskal-Wallis test, * $p < 0.05$; ** $p < 0.01$, *** $p < 0.001$, **n = 3.**

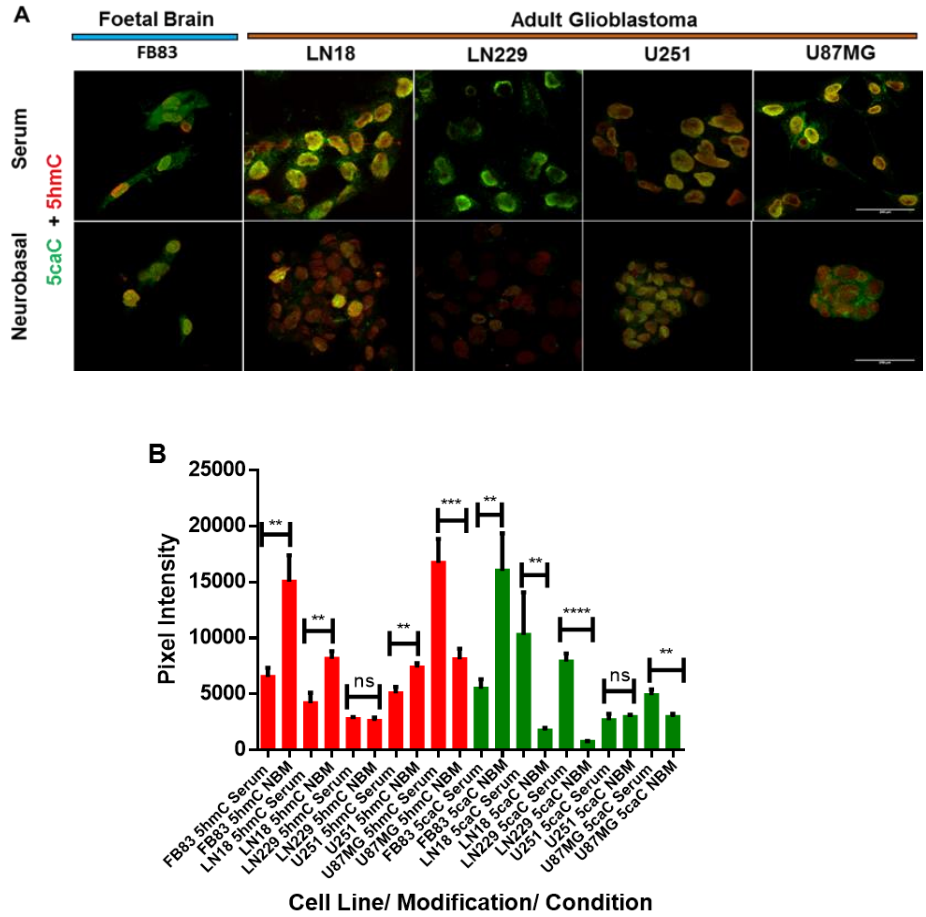
3.3.3 Neurosphere Enrichment Significantly Affects OximC Magnitude and Distribution Compared to Serum

Comparing 5hmC and 5caC signal intensities in neurobasal and serum culture conditions (fig 3.13A-B) demonstrated that higher 5hmC signals were observed in neurobasal media for FB83 ($p = 0.0015$), LN18 ($p = 0.0028$) and U251 ($p = 0.0044$). Greater 5hmC signal intensity was observed in U87MG ($p = 0.0006$) and no significance was observed between neurobasal and serum conditions in LN229 ($p = 0.804$). An

opposite trend of 5caC signal intensities in neurobasal versus serum conditions was observed in GBM cell lines LN18, LN229 and U87MG 5caC levels lower in neurobasal compared to serum in contrast to FB83 which experienced a very significant ($p = 0.0028$) increase. Interestingly, unpaired, two tailed, two-way ANOVA analysis of cell line vs media condition revealed greater 5hmC and 5caC signal intensity variation attributable to differences between cell lines compared to media condition (5hmC:36.28%, $p < 0.0001$ vs 2.61% $p = 0.09$ and 5caC: 17.92%, $p = 0.0012$ vs 0.63%, $p = 0.37$). Normalising gene expression signatures of cell lines to serum as the conventional culture medium (fig 3.13C) we observed elevation of DNA demethylation genes *TET1*, *TET3* and *TDG* transcription in FB83 cells in neurobasal conditions relative to serum. Together with expected stem markers *NESTIN*, *NOTCH1* and *SOX2*, these genes increased expression 8.8-fold to 60-fold. Unexpectedly, *WT1* expression shot up 620-fold and *TET2* was down-regulated to negligible levels (0.0000001-fold). All GBM cell lines experienced reduced *TET3* expression and all three *TET* genes were reduced by 1-2 orders of magnitude relative to serum levels. *WT1* expression was most dramatically altered in LN18 (0.01-fold reduction) and U251 (17.75-fold increase) and LN229 showed a marginal increase relative to serum (1.46-fold).

To determine whether the 5caC modifications were physically present on promoters of stemness associated genes, anti-5caC DNA immunoprecipitation coupled qPCR (5caC-DIP-qPCR) was performed in

U251 (fig 3.13D). Negligible differences in 5caC fold enrichment were observed between mock IgG non-5caC-specific antibody controls and targeted 5caC probing under serum conditions. In contrast, *NESTIN*, *NOTCH1*, *OLIG2*, *POU5F1* (*OCT4*), *NANOG* and *EGFR* promoters possessed greater 5caC enrichment ranging from 3.07-fold (*POU5F1*) to 19.04-fold (*NOTCH1*) relative to their counterparts in serum cultured cells. No change in 5caC enrichment between treatments was observed on *SOX2* and *CD133* promoters.



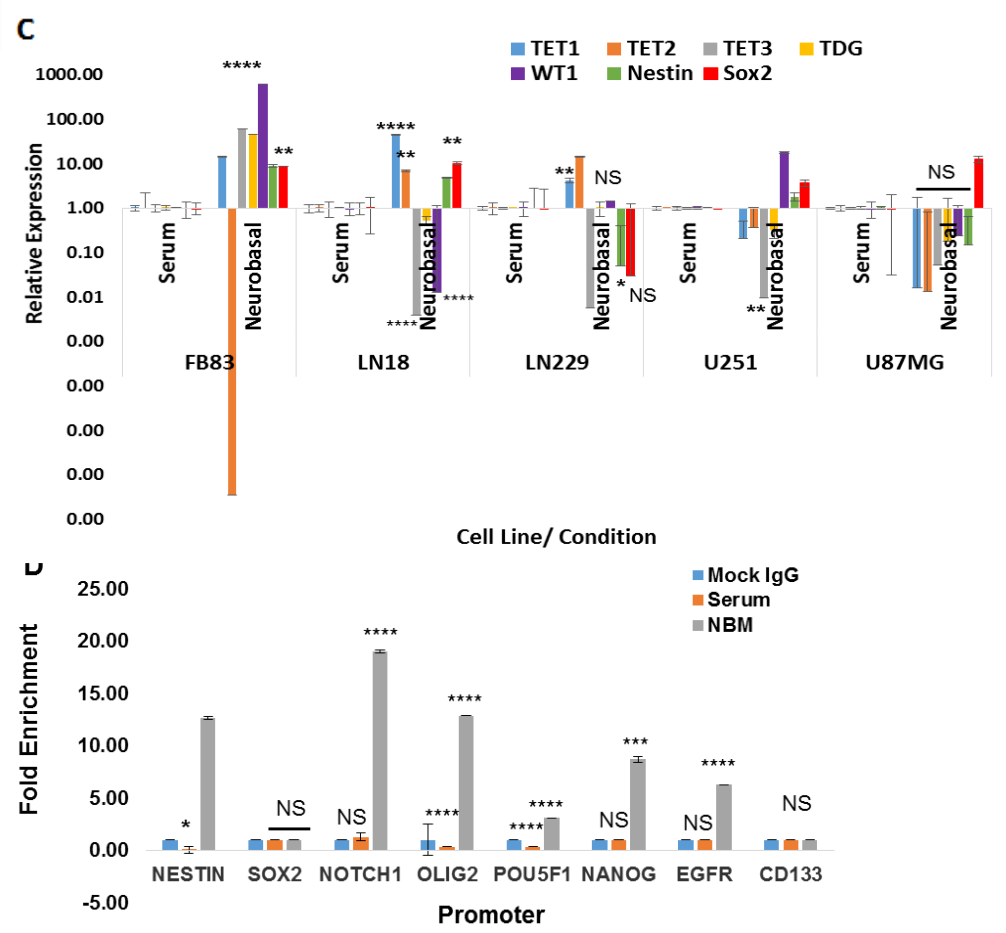
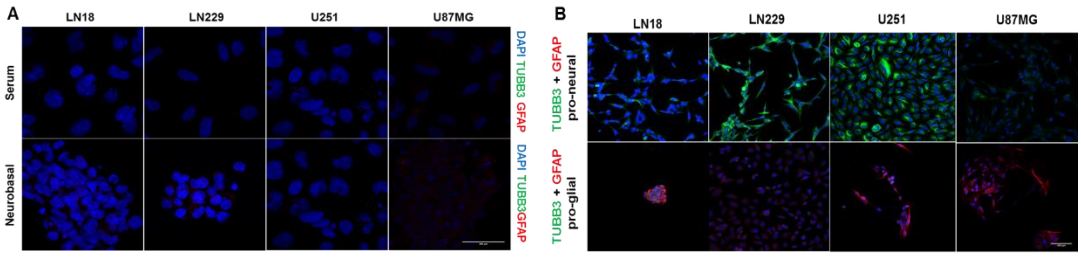


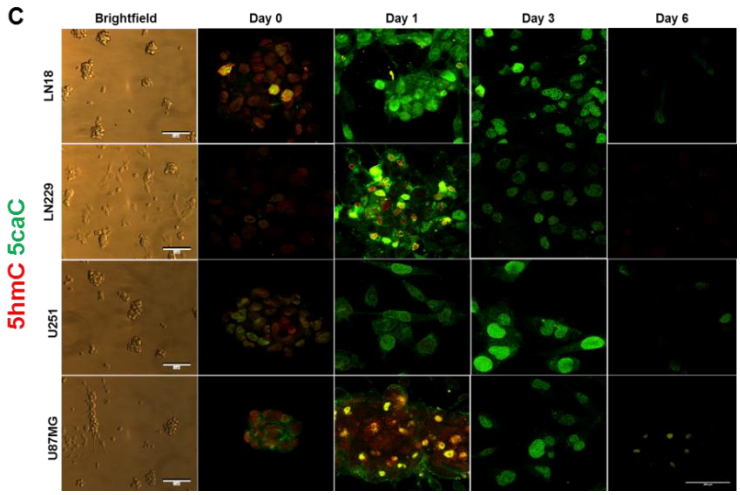
Figure 3.13 Comparison of oximC magnitude and distribution characterisations in serum vs neurobasal conditions. (A) 5hmC and 5caC immunostaining in serum vs neurobasal conditions, staining and images acquired at identical settings. Scale bar represents 200µm. (B) Quantification of oximC signal intensities in serum vs neurobasal conditions per cell line, unpaired two tailed Mann-Whitney tests, mean± SD. (C) QPCR analysis of DNA demethylation machinery genes, *WT1* and neural stem cell marker genes in serum vs neurobasal conditions, mean ± SD, unpaired two tailed Mann-Whitney test (serum vs neurobasal media). (D) 5caC DIP-qPCR analysis of 5caC enrichment on neural stem cell marker gene promoters in serum vs neurobasal media relative to mock IgG controls, unpaired, two tailed Kruskal Wallis test, mean ± SD, NS $p \geq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; $p < 0.0001$, n = 3.

3.3.4 OximCs Transiently Accumulate during GBM Cell Differentiation

Expanding the concept of forcible modulation of epigenetic reconfiguration of cells further, differentiation of GBM cell lines to either glial or neuronal lineages was performed (fig 3.14A-D). As cell lines were conventionally cultured under serum conditions which, as an undefined media, may contain a complex cocktail of conflicting paracrine signals, cell lines were co-stained for glial and neuronal markers Glial Fibrillary Acidic Protein (GFAP) and Beta Tubulin III (TUBB3) respectively. Staining was performed to determine if GBM cells maintained in serum were responsive and inducted into any neuronal or glial differentiation programme. No substantial TUBB3 or GFAP expression was detected in serum. Cancer cells do not undergo lineage commitment and terminal differentiation, retaining their capacity to enter the cell cycle even after specification and determination (Caren *et al*, 2015). As it was unknown if serum cultured cells were partially specified towards glial or neuronal lineages, neurobasal conditions were administered to encourage neural stem cell proliferation and enrich the pool of progenitors capable of undergoing differentiation. Weak GFAP staining but no TUBB3 presence was observed in neurobasal conditioned GBM cell lines. Following subjection to 4 day neurobasal conditions and either 6 day neuronal or glial differentiation programmes, GBM lines showed positive expression of neuronal TUBB3 or glial GFAP markers. Patterns of rapid 5hmC and 5caC signal intensity increase and subsequent diminishment were

observed in all GBM cell lines subjected to both neuronal and glial differentiation programmes.





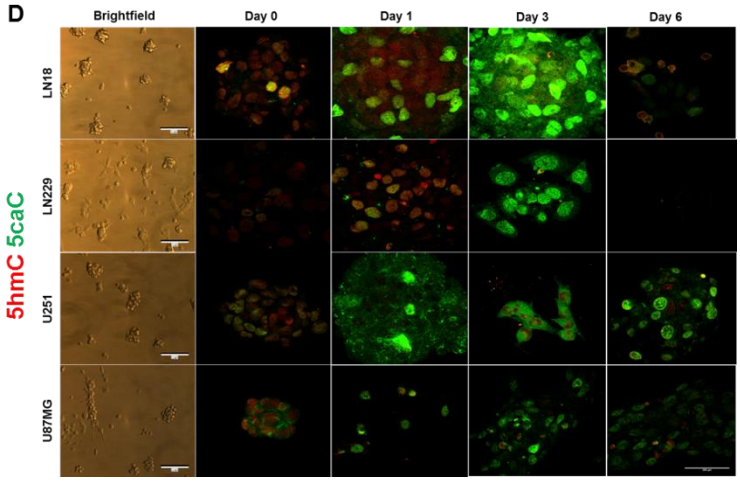


Figure 3.14 Immunofluorescence analysis of oximC distribution during GBM cell differentiation. (A) Immunostaining of neuronal beta III tubulin (TUBB3) and glial fibrillary acid protein (GFAP) signal intensities in GBM cell lines under serum conditions. (B) Immunostaining of TUBB3 and GFAP on day 6 of neuronal and glial differentiation. (C-D) 5hmC and 5caC staining during the 6 day neuronal (C) or glial (D) differentiation programme. Scale bar represents 200µm, n = 3.

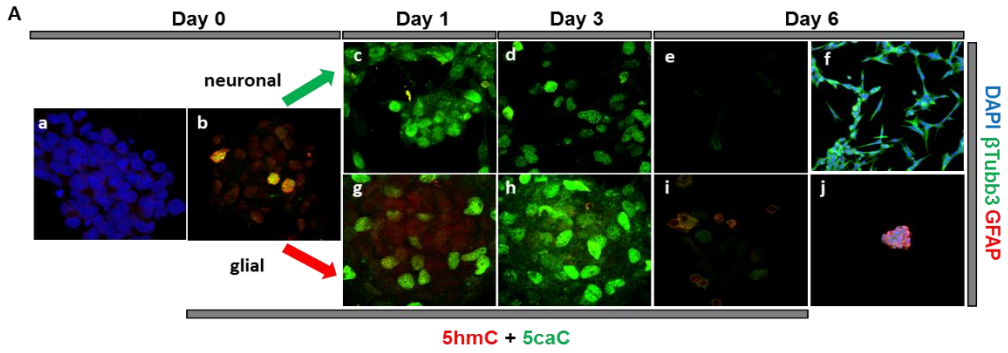
3.3.5 Dynamic Patterns of 5caC Abundance Correlate with *WT1* Transcriptional Expression

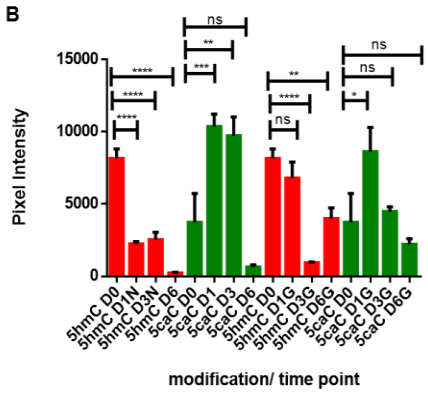
Quantification of 5hmC and 5caC levels in LN18 and the other GBM cell lines (appendix) produced similar profiles of oximC magnitudes over time (fig 3.15A-B). Highest 5hmC presence was detectable on day 0 (4^{th} and final day of neurobasal conditioning) which subsequently significantly decreased by 72.6% during days 1 and 3 neural and by 88.5% for glial differentiation (D1N, D3N, D1G, D3G respectively). Levels then depleted at day 6 for neuronal (D6N) but recovering to 48.6% of initial day 0 5hmC levels on day 6 of glial (D6G) differentiation. In contrast to 5hmC, 5caC levels increased by 2.77-fold on days 1 and 3 but dropped to 16% of day 0 levels on day 6 of the neuronal programme. Similarly, 5caC levels peaked on day 1 (2.31-fold increase) and declined to 1.20-fold higher followed by a 0.49-fold lower than day 0 5caC levels on days 3 and 6 of the glial programme respectively. The effect of the differentiation programmes duration on oximC signal intensities was diminished relative to oximC immunofluorescence signal variation attributable to cell line differences. Unpaired two tailed two-way ANOVA demonstrated that for

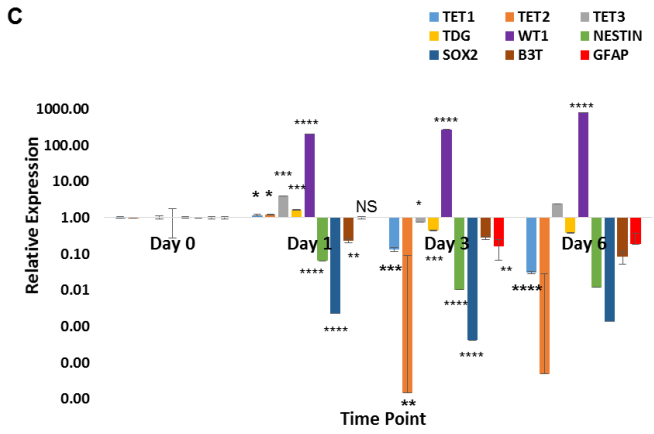
neuronal differentiation, analysis of cell line versus time point produced 5hmC percentage variation: 18.39%, $p < 0.0001$ vs 10.57%, $p < 0.0001$ and 5caC percentage variation: 29.71%, $p < 0.0001$ vs 3.30%, $p = 0.0002$. OximC signal intensity variation occurring during glial differentiation entailed 5hmC percentage variation (cell line vs time point): 7.56%, $p < 0.0001$ vs 4.44%, $p = 0.0035$ whereas 5caC percentage variation was 15.17%, $p< 0.0001$ vs 9.65%, $p < 0.0001$.

Curiously, *TET* gene expression over the 6 day differentiation regimen was markedly decreased relative to day 0 conditioning (fig 3.15C-D). The highest observed *TET* gene, *TET3* expression varied between days 0-6 however did not surpass a 10-fold increase. *TET1*, *TET2* and *TDG* displayed 1-3 orders of magnitude decrease over the 6 day programmes with *TET2* showing the greatest down-regulation. Interestingly, *WT1* expression dramatically increased relative to day 0 on days 1, 3 and 6 of neuronal differentiation at 203.51-fold, 269.85-fold and 792.08-fold respectively and at 122.50-fold, 12.19-fold and 3.82-fold expression at days 1, 3 and 6 of glial differentiation respectively. Profile analysis of 5hmC and 5caC spatial distribution and magnitudes across nuclei during the differentiation regime (fig 3.15E) illustrates the dynamic patterns of oximC presence in response to media induced paracrine signalling over time. 5hmC and 5caC profiles follow similar trends albeit separate in magnitude however the most obvious profile discrepancies were observed at day 0, illustrating the heterogeneity of 5caC signal under neurobasal conditions, and day 3 of glial differentiation as 5caC displays signal saturation across the nucleus. These findings may point to oximC

involvement in a reconfiguration of the transcriptional paradigm occurring during differentiation.







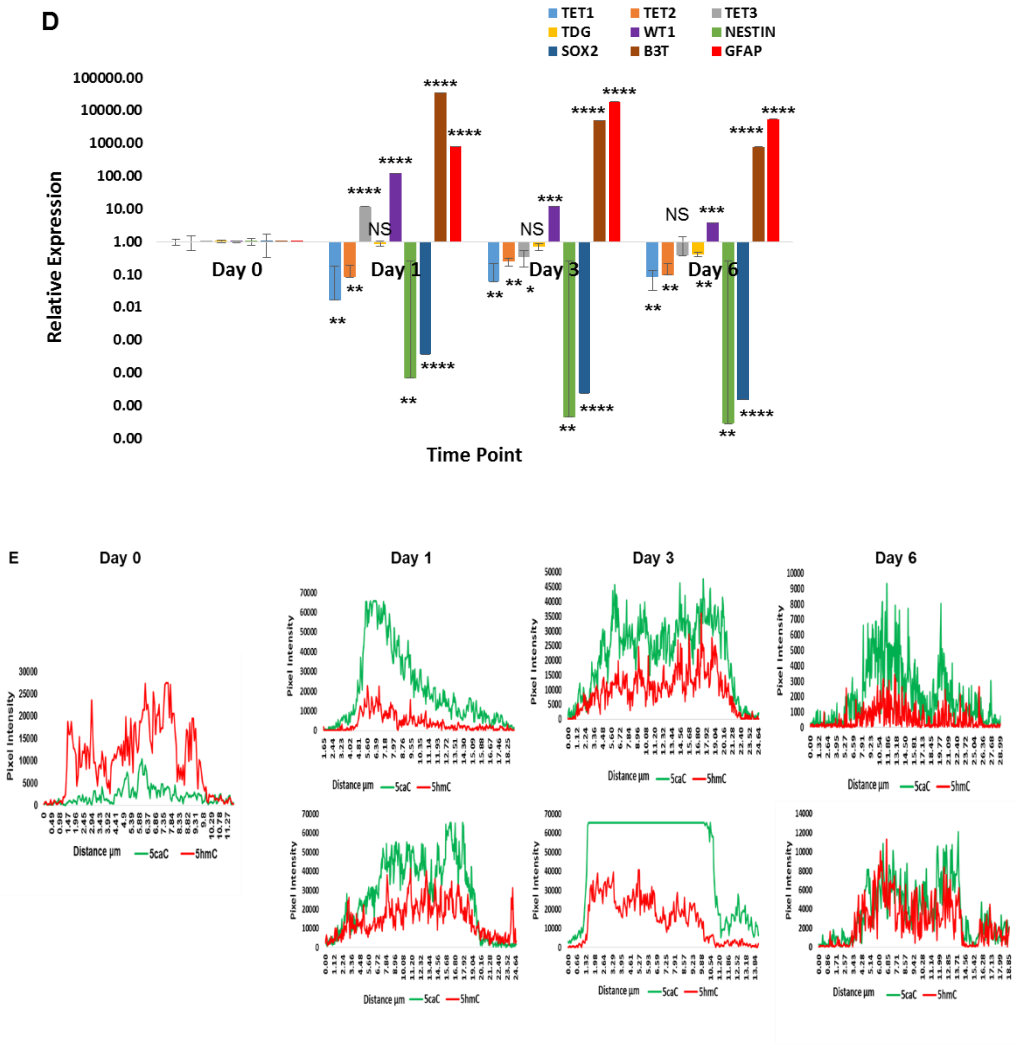


Figure 3.15 Characterisation of DNA demethylation activity during GBM cell line differentiation. (A) Immunostaining of oximC presence during neuronal and glial differentiation. Day 0 cells (a-b) were stained for oximC presence as well as neuronal and glial markers. OximC presence was examined during stages of neuronal (c-e) and glial (g-i) differentiation. Scale bar represents 200µm. (B) Quantification of oximC magnitudes during neuronal and glial differentiation in LN18. (C-D) QPCR transcriptional analysis of DNA demethylation, WT1 and neuronal (C) and glial (D) specific marker gene expression during differentiation. (E) Profile analysis of oximC cellular distribution during differentiation. Unpaired two tailed Kruskal-Wallis test, mean ± SD, NS $p \geq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$, n = 3.

3.4. Investigating the Dynamics of a Possible WT1-5caC Interaction

Our data thus far has alluded to the over-expression of WT1 in GBM cell lines in conditions of serum culture, and how through enforcing stem-like conditions and transcriptional profile modulation via cell differentiation, WT1 expression patterns correlate with 5caC presence in most GBM cell lines. Deeper probing into this perceived correlation required functional analysis specifically targeting WT1 expression. As such proof-of principle experiments were devised and employed to characterise a possible WT1-5caC engagement. No mutations or copy number alterations of WT1 were present in any of the U87MG, U251, LN18 or LN229 GBM cell lines according to the Catalogue of Somatic Mutations in Cancer (COSMIC).

3.4.1 WT1 Expression Correlates with 5caC Presence in all GBM Cell Lines except LN229

Interrogation of WT1 and 5caC presence by immunostaining (fig 3.16A-C) revealed the highest WT1 signal intensity occurring in U251 and U87, the former of which was significantly higher ($p = 0.0028$) than LN18 and the latter showed no significant difference ($p = 0.16$). Although both U251 and U87MG exhibited higher WT1 expression, LN18 demonstrated the highest WT1:5caC correlation ($r = 0.84$, $P < 0.0001$). LN229 showed negligible WT1 expression compared to secondary antibody-only

controls (not shown) and hence did not produce a significant correlation with 5caC (r = -0.076, p = 0.146).

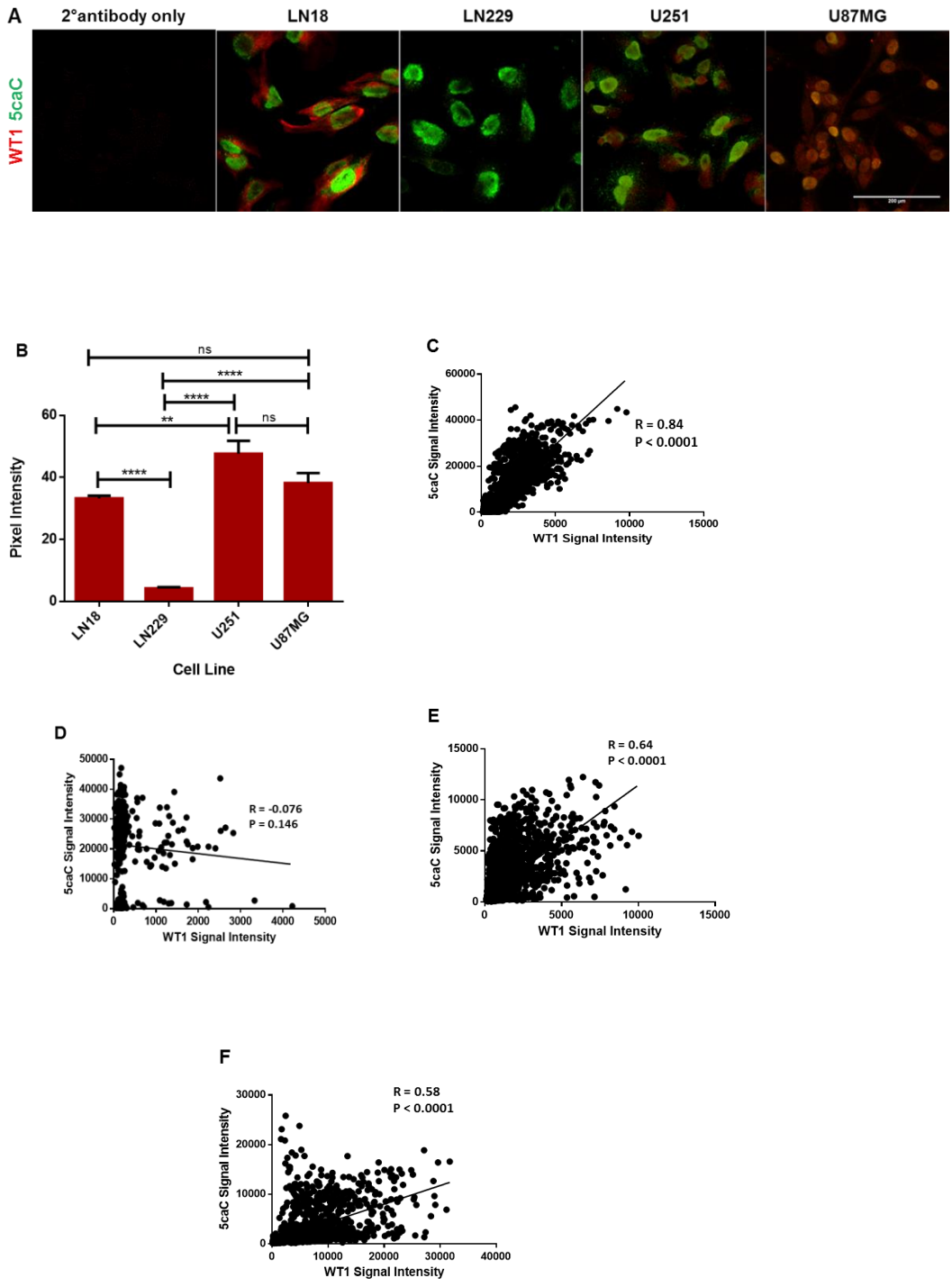


Figure 3.16 Characterisation of WT1 and 5caC enrichment in GBM cell lines. (A) Immunofluorescent detection of WT1 and 5caC across GBM cell lines, staining and images acquired at identical settings, scale bar represents 200µm. (B) Quantification of WT1 fluorescence signal intensity, mean ± SD,

unpaired two tailed ANOVA. (C-F) Scatterplots depicting correlation between WT1 and 5caC signal intensities within each cell line C-F: LN18, LN229, U251 and U87MG respectively, linear regression, two tailed Spearman correlation, NS $p \geq 0.05$; ** $p < 0.01$; **** $p < 0.0001$, n =3.

3.4.2 Transient *WT1* knockdown correlates with reduced global oximC nuclear presence but elevates retro-transposon 5caC enrichment in U87MG cells

Transient knockdown of *WT1* gene transcription via siRNA mediated transfection (fig 3.17A-B) resulted in significantly decreased 5hmC and 5caC immunostaining signal intensities compared to control transfections ($p < 0.0001$ and $p = 0.0049$ respectively). WT1 and 5caC co-staining signal intensity decreased highly significantly compared to transfection controls ($p < 0.0001$, $p < 0.0001$ respectively). The transcriptional effect of 99.98% *WT1* knockdown was observed in *TET1*, *TET2*, *TET3*, *TDG* and *LINE1* element up-regulation, most prominently by *TET1* (2.07-fold increase) (fig 3.17C). *LINE1* transcriptional retrotransposon up-regulation was associated with a significant 5caC depletion on its promoter region ($p < 0.0001$) suggesting a transcriptionally repressive role of 5caC (fig 3.17D).

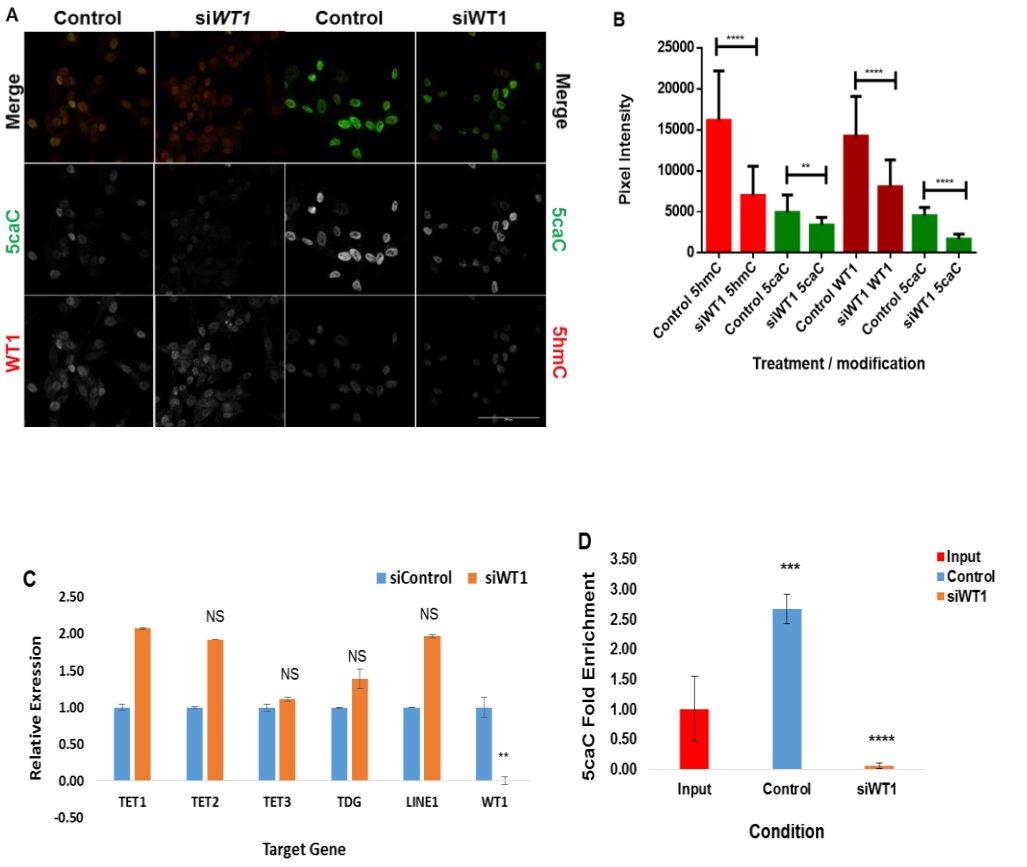
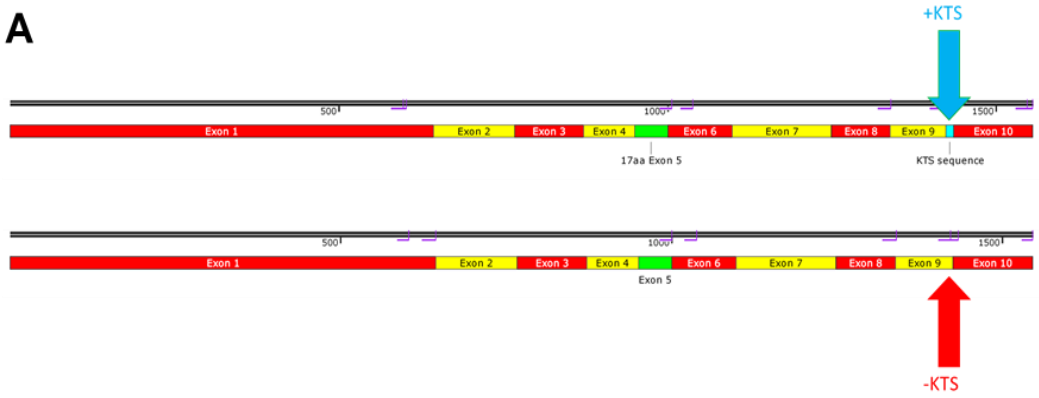


Figure 3.17 Effect of WT1 knockdown on DNA demethylation profiles in GBM cells. (A) Immunostaining analysis of siRNA mediated *WT1* knockdown on 5hmC and 5caC presence in U87MG cells. Scale bar represents 200µm, staining and images acquired at identical settings. (B) Quantification of WT1, 5hmC and 5caC signal intensities in control vs si*WT1* knockdown cells. (C) QPCR analysis of DNA demethylation machinery, *WT1* and *LINE1* target genes normalised against input (1% of total sonicated DNA ng/µl). (D) 5caC DIP-qPCR of 5caC genomic occupancy on the *LINE1* promoter in U87MG cells normalised against input (1% of total sonicated DNA ng/µl). (B-D) mean ± SD, unpaired two tailed Mann-Whitney tests, NS $p \geq 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$, n = 3.

3.4.3 -/+KTS Splice Variants of *WT1* are Undetected in LN229

Closer examination of WT1 absence in LN229 compared to LN18, U251 and U87MG GBM cell lines (fig 3.18A-C) revealed LN18 possessed the highest *WT1* transcriptional expression (3772-fold increase, $p < 0.0001$) with U251 and U87MG both demonstrating high expression but to a lesser degree (679-fold, $p < 0.0001$ and 431-fold, $p < 0.0001$ respectively). LN18, U251 and U87MG cell lines expressed both – and + KTS WT1 isoforms with LN18 and U251 possessing stronger –KTS bands than +KTS while U87MG showed stronger +KTS expression than –KTS. LN229 did not express either – or +KTS isoform (red and blue dashed boxes, fig 3.19B).



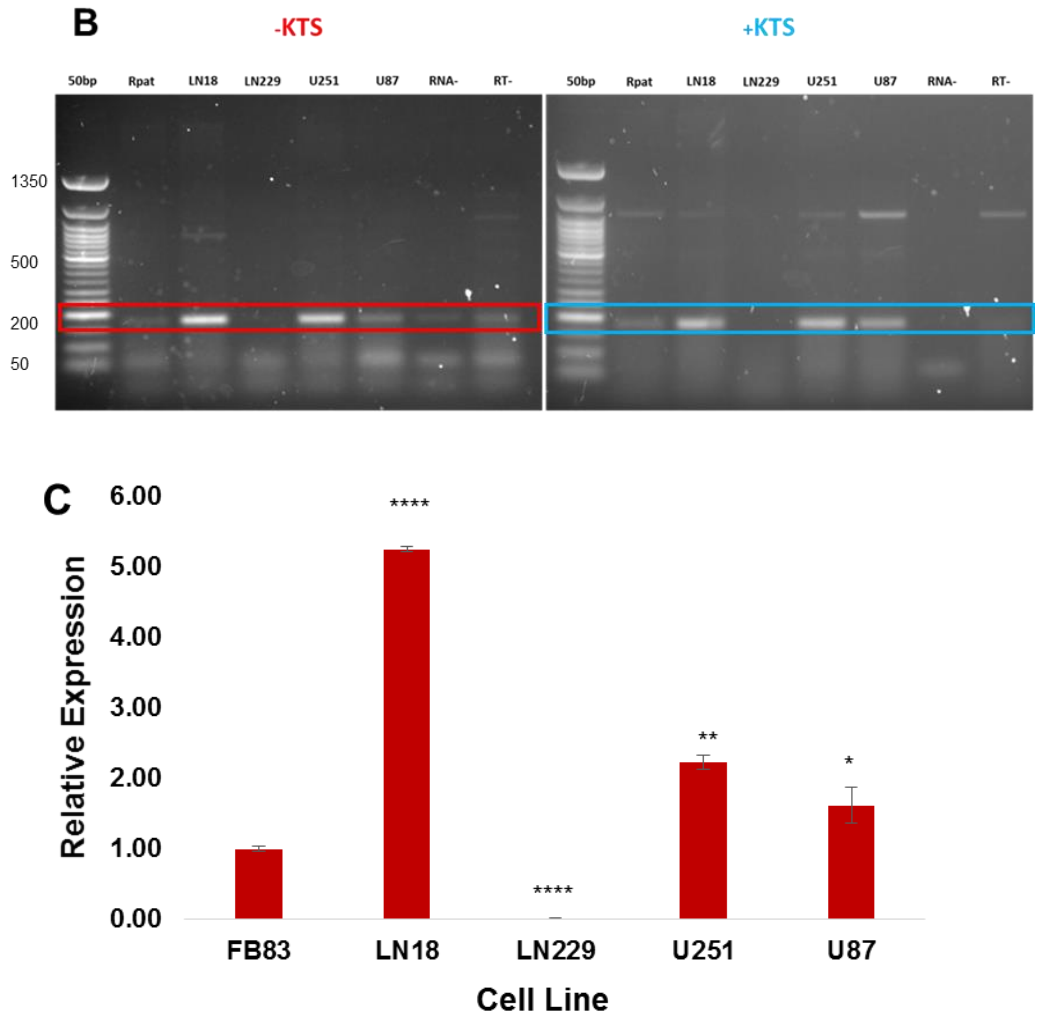
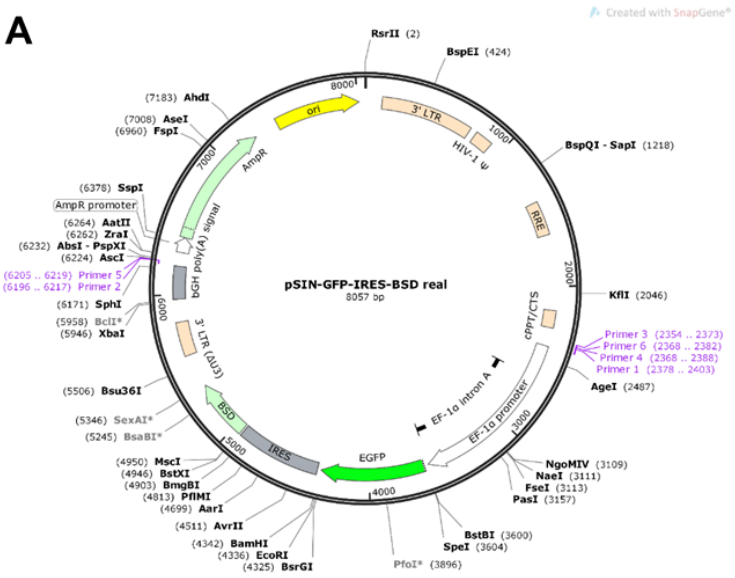


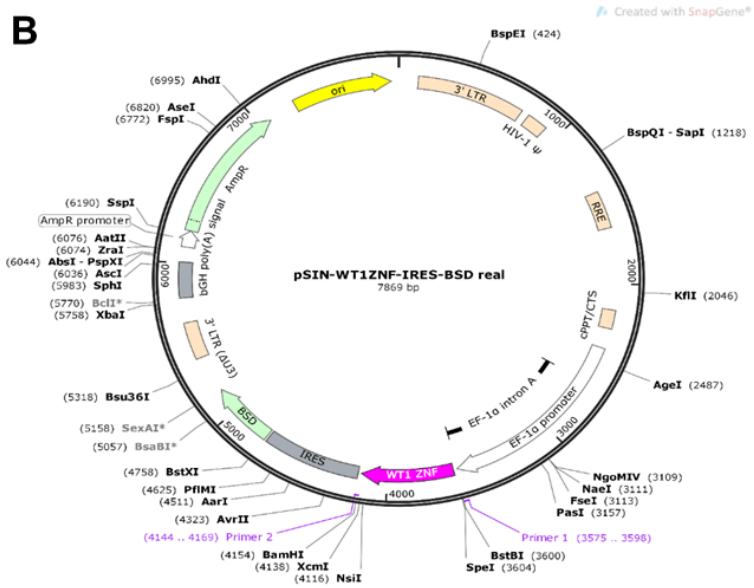
Figure 3.18 Examination of *WT1* expression across GBM cell lines. (A) Schematic representations of *WT1* –KTS (red arrow) and +KTS (blue arrow) sequence locations within alternative splicing isoforms. (B) Agarose gel of –KTS and + KTS *WT1* isoform expression across GBM cell lines relative to RPBLAT (RPAT) negative controls. (C) QPCR analysis of *WT1* expression across GBM cell lines relative to foetal brain FB83 cells. Mean ± SD, unpaired two tailed Mann-Whitney test, * $p < 0.05$; ** $p < 0.01$; **** $p < 0.0001$, n =3.

3.4.4 Verification of –KTS and +KTS Zinc Finger Domain Construct Production

The combined findings of immunostaining, qPCR and RT-PCR-based detection methods imply WT1 is not expressed by LN229 under serum conditions. LN229 cells display a pronounced 5caC heterogeneity independent of WT1 expression. To determine whether WT1 may influence 5caC distribution and presence via a protein-oximC binding interaction as opposed to trans-activating transcription factor recruitment of DNA demethylation machinery, only exons 6-10 (exons 7-10 encode *WT1* DNA binding zinc finger domains) were cloned into an expression vector for transfection (fig 3.19 A-F). Exon 6 was included as a linker region to ensure that zinc finger domain folding would not be perturbed by the addition of a *SpeI* endonuclease site into the cloning construct. The pSIN-GFP-IRES-BSD plasmid was used as the parent vector for *WT1* zinc-finger domain subcloning. The parent plasmid was digested with *SpeI* and *EcoRI* restriction enzymes to release the Green Fluorescent Protein (GFP) coding fragment. The *WT1* zinc-finger domains were PCR amplified from LN18 cDNA using exon 6 forward (WT1ex6F2: 5'-CACTTACTAGTATGAACCACAGCACAGGGTACGAGA-3') and exon 10 reverse primers (WT1ex10R: 5'-TATCGGAATTCTCAAAGCGCCAGCTGG-3') with *SpeI* and *EcoRI* restriction sites incorporated into them respectively. PCR product bands were gel purified, ligated into the digested parent vector, cloned in TOP10 DH5alpha *E.coli*, miniprepped and sequenced for incorporation

of the correct WT1 fragment. The successfully cloned *WT1* -/+KTS vectors were designated pSIN-WT1ZNF-IRES-BSD real. Successful subcloning of –KTS and +KTS fragments into pSIN-WT1ZNF-IRES-BSD was verified by Sanger sequencing.





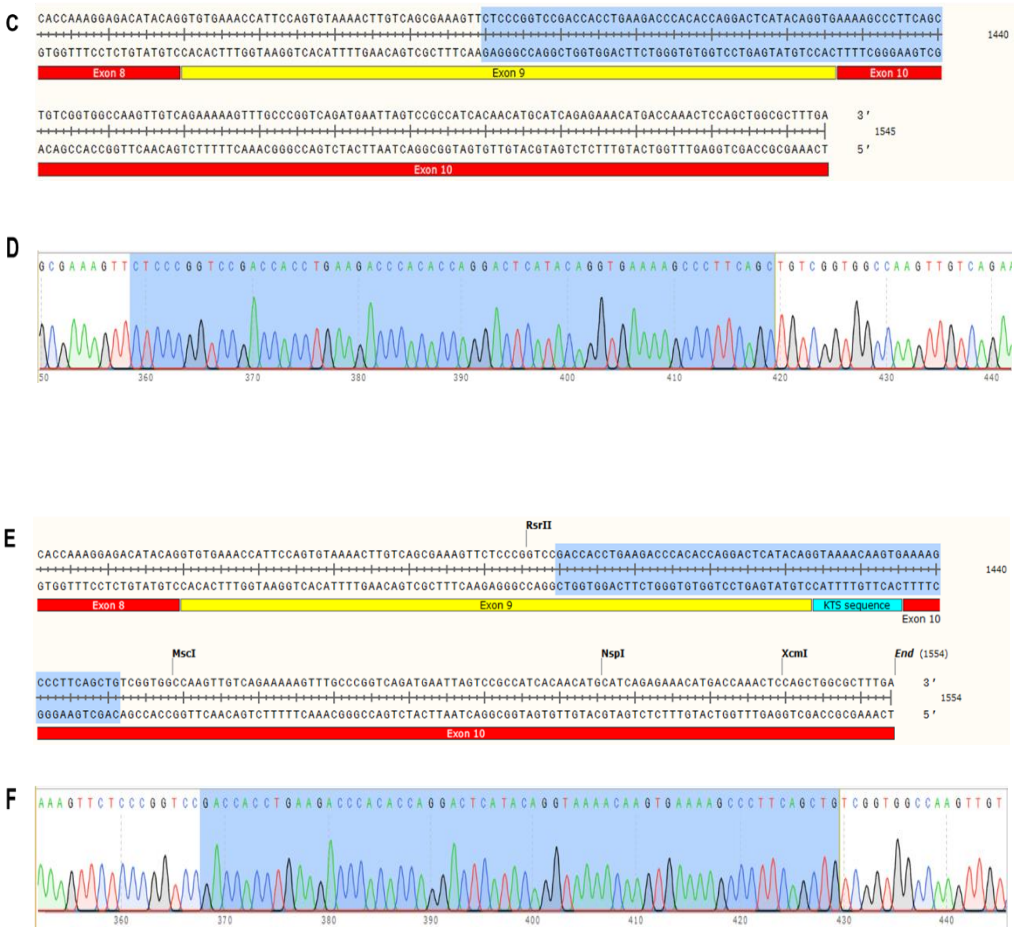


Figure 3.19 Construction of WT1 –KTS and +KTS zinc finger plasmids. (A) Parent pSIN-GFP-IRES-BSD plasmid. (B) Parent pSIN-GFP-IRES-BSD plasmid was digested to remove the e-GFP coding sequence which was subsequently substituted by either –KTS or +KTS WT1 zinc finger domain coding sequence (pink). (C-D) Verification of –KTS WT1 isoform zinc finger domain coding sequence incorporation. (E-F) Verification of +KTS WT1 isoform zinger finger domain coding sequence incorporation. Blue region denotes the 9bp KTS sequence located between exons 9 and 10.

3.4.5 Forced Expression of WT1 –/+KTS Zinc Finger Domain in LN229

LN229 cells were transfected with 3µg of either *WT1*-/+KTS plasmid or the unmodified parent vector with *GFP* intact (fig 3.20). Transfection of the parent vector served not only as a control for the effect of vector delivery but also facilitated the tracking of GFP expression over time and thus WT1 zinc finger expression as the Elongation Factor 1 Alpha (EF-1a) promoter drives transcription of both fragments. Untransfected cells were subjected to electroporation only. GFP expression was tracked over 24-48 hours post transfection to assess the level of promoter usage to determine at which stage to harvest cells for qPCR and immunostaining analysis. The 24-48 hour temporal window was selected as we anticipated that our truncated WT1 protein may undergo proteolytic degradation after 48 hours which may severely reduce its detection. A highly dense population of cells was observable for untransfected and parent plasmid only (pSIN) transfected cells 24 hours post transfection and both became confluent at 48 hours. A greater proportion of cells expressing a GFP signal higher at 48 than 24 hours which informed us of the appropriate cell harvest time. Although all treatment conditions were performed at the same cell density with the same concentration of vector DNA under the same transfection parameters, fewer surviving cells were visible in the –KTS and +KTS transfected wells compared to untransfected and pSIN only transfected ones. Fewer surviving cells were visible in –KTS transfected cells than +KTS transfected ones. Immunofluorescent interrogation of transfected cells revealed

detectability of WT1 in –KTS and +KTS transfected but not in pSIN or untransfected cells.

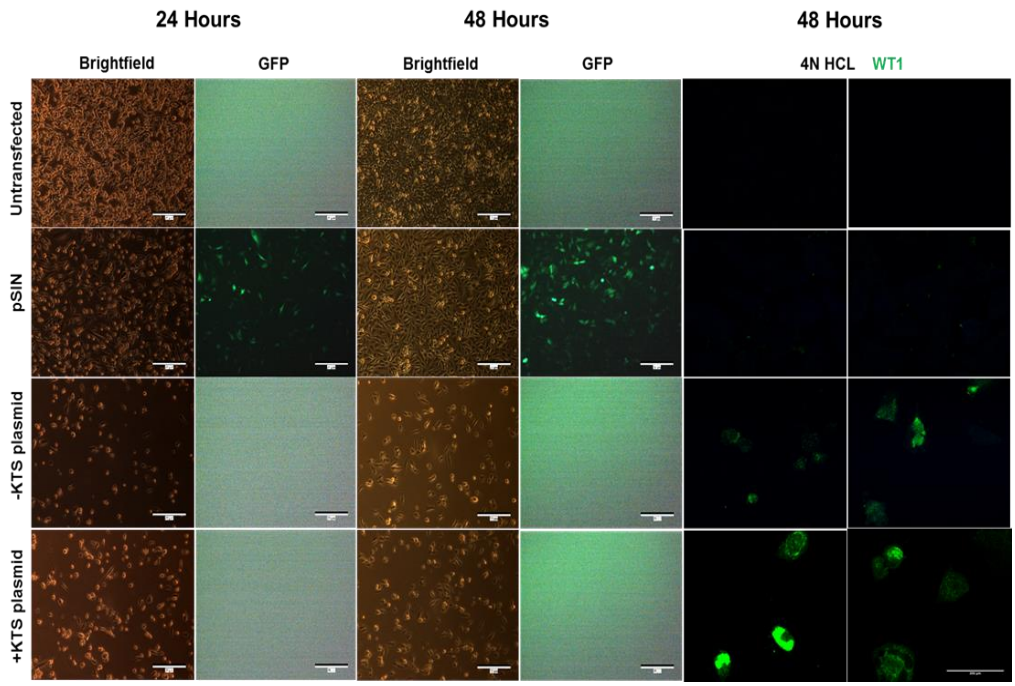


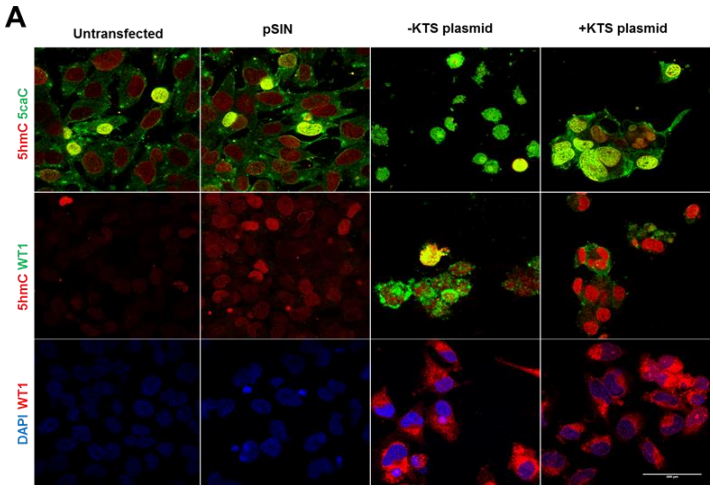
Figure 3.20 Over-expression of WT1 –KTS or +KTS zinc finger domain plasmids in LN229 GBM cells. Brightfield and GFP images scale bars represent 25µm. WT1 Staining and immunofluorescence images acquired at identical settings, scale bar represents 200µm.

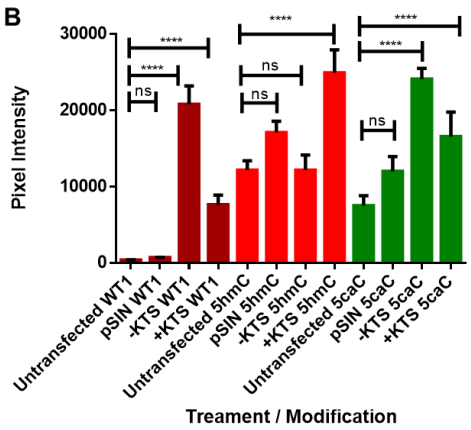
3.4.6 Over-expression of WT1 –KTS and +KTS zinc finger domains significantly alters 5caC magnitudes and induces full-length *WT1* transcription

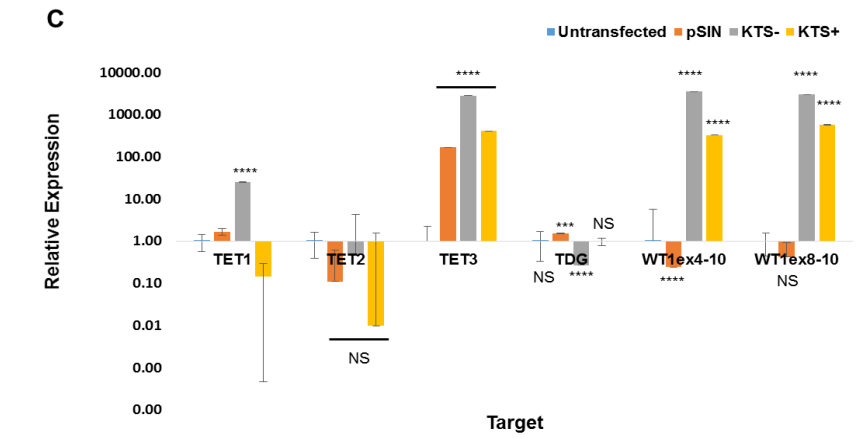
Immunostaining and confocal microscopic quantification of WT1 presence in transfected LN229 cells (fig 3.21A-B) resulted in highly significantly elevated WT1 signal intensity in –KTS and +KTS transfected cells compared to pSIN parent plasmid or untransfected cells, in which

WT1 was undetectable ($p < 0.0001$, $p < 0.0001$ respectively). The –KTS zinc finger domain transfected cells possessed 2.71-fold higher WT1 signal intensity than +KTS transfected cells ($p = 0.0026$) which were 48.57-fold and 27.45-fold higher than untransfected and 17.90-fold and 10.12-fold higher than pSIN transfected cells respectively. Analysis of 5hmC levels showed no significant difference between cell treatments except WT1+KTS transfection possessed very highly significant levels ($p < 0.0001$) and no significant difference in 5caC signal magnitude was observed between untransfected and pSIN conditions ($p > 0.05$). Interestingly, -KTS zinc finger domain transfected cells possessed highly significantly elevated 5caC signal intensities ($p < 0.0001$) 3.59-fold, 2.00-fold and 1.45-fold higher than untransfected, pSIN and +KTS zinc finger transfected cells respectively. The +KTS transfected cells also depicted elevated 5caC levels relative to untransfected ($p < 0.0001$) and 2.19-fold and 1.38 fold higher than untransfected and pSIN controls respectively. Transcriptional profile qPCR analysis of DNA demethylation machinery genes (fig 3.21C) revealed inconsequential changes in *TET1*, *TET2* and *TDG* expression between untransfected and pSIN treated cells. *TET1* and *TET2* were down-regulated in WT1+KTS zinc finger domain transfected cells compared to control treatments (0.15-fold and 0.01-fold decreases respectively). Surprisingly, *TET3* expression was 165-fold, 404-fold and 2813-fold elevated in pSIN, +KTS and –KTS treated cells respectively. WT1 exons 8-10 present in the zinc finger construct were expressed 576-fold and 2955-fold higher in +KTS and –KTS treatments respectively but transcripts including exon 4 (not part of the zinc-finger

construct) were detected at 3514-fold and 313-fold levels of expression relative to untransfected cells. This suggests that transfection of the zinc-finger domain alone is sufficient to stimulate expression of full length *WT1*. Additionally, WT1/5hmC and 5hmC/5caC signal profile analysis revealed the strongest signal correlations between WT1/ 5hmC and 5hmC/5caC occurred in –KTS transfected LN229 cells (fig 6.21D profiles e & f). This may suggest the differential WT1 zinc finger – oximC interactions contingent on distinct WT1 isoform expression.






C
Untransfected
pSIN
KTS-
KTS+
Relative Expression
10000.00
1000.00
100.00
10.00
1.00
0.10
0.01
0.00
0.00
TET1
TET2
TET3
TDG
WT1ex4-10
WT1ex8-10
****
***
NS
Target

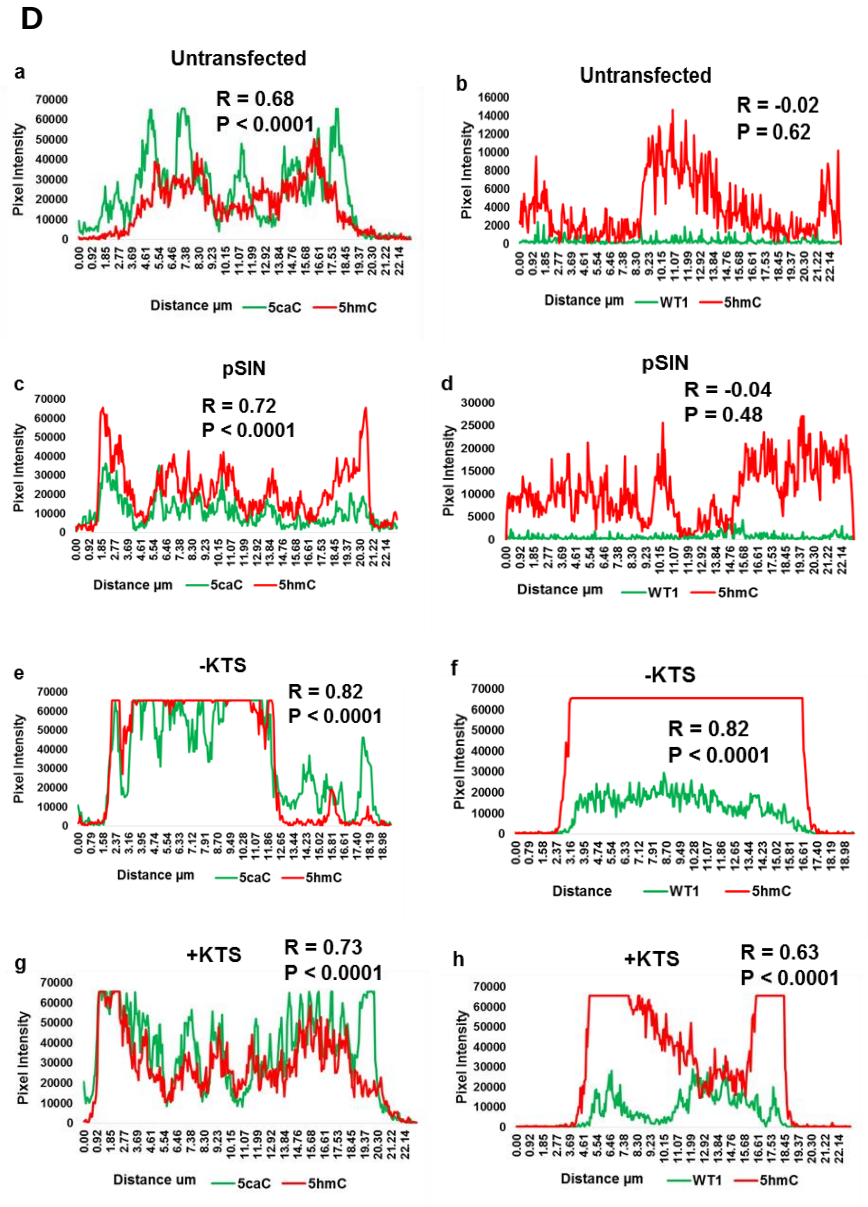

D
a
Untransfected
R = 0.68
P < 0.0001
b
Untransfected
R = -0.02
P = 0.62
c
pSIN
R = 0.72
P < 0.0001
d
pSIN
R = -0.04
P = 0.48
e
-KTS
R = 0.82
P < 0.0001
f
-KTS
R = 0.82
P < 0.0001
g
+KTS
R = 0.73
P < 0.0001
h
+KTS
R = 0.63
P < 0.0001
Pixel Intensity
Distance µm
5caC
5hmC
WT1
Distance

Figure 3.21 Effect of WT1 –KTS and +KTS zinc finger isoform transfection on DNA demethylation in LN229 GBM cells. (A) WT1, 5hmC and 5caC immunostaining reflecting the effects of transfection. Staining and images acquired at identical settings, scale bar represents 200µm. (B) Quantification of WT1, 5hmC and 5caC signal intensity following WT1 –KTS or +KTS zinger finger transfection. (C) QPCR analysis of DNA demethylation genes and WT1 amplicon transcriptional levels following transfection. Mean ± SD, unpaired two tailed Kruskal-Wallis tests. (D) Confocal profile analysis of WT1/5caC and 5hmC/5caC signal correlations at different transfection conditions. Spearman correlation, NS $p \geq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$, n =3.

3.5 Technical Discussion

The aim of this chapter was to detect and characterise oximC presence in adult glioblastoma cell lines as an initial attempt at understanding the scale and significance of active demethylation mechanisms. Moreover, we sought to accrue empirical evidence endeavouring to unveil what relationship if any, WT1 bears to oximC distribution and abundance within nuclei. Interrogation of 5hmC and 5caC presence in foetal human and adult mouse brain sections (fig 3.1A-B) demonstrated differential 5hmC magnitudes between the two samples and most notably, 5caC enrichment at levels comparable to 5hmC in foetal but inconsequential amounts in the adult mammalian brain.

Although multi-methodological analysis is desirable, even pertinent to validate experimental results, the gold standard of HPLC-MS methods would be improper and impractical for oximC magnitude verification as homogenisation of brain tissue regions at quantities sufficient for detection would be impossible given our sample availability. Moreover, LC-MS quantification of 5hmC and 5caC abundances derived from tissue homogenisation would not reflect potential cellular heterogeneity of the oximC signals present in different anatomical structures. Nevertheless, anti-5hmC and anti-5caC primary antibodies implemented in this work have been LC-MS validated.

Owing to sample availability, whole foetal mouse brains sagittal sections were compared to coronal mouse brain sections. The coordinates of mouse brain cerebrums had to be approximated as the 4N HCL conditions necessary for oximC antigen retrieval disrupt hydrogen bonding between bases within the DNA helix. Consequently, the utility of DNA intercalating dyes such as DAPI, which could provide indications of specific tissue landmarks within the brain, was severely limiting. Therefore, accurate navigation of the brain tissue proved problematic and microscope objectives were positioned in approximation of the cerebrum based on mouse brain and developing human brain atlases (www. atlas.brain-map.org).

A more comprehensive and accurate analysis of 5hmC and 5caC staining distribution could have been achieved if sections of multiple brain regions with distinct anatomical structures e.g. cerebellum and

hippocampus in the same plane e.g. sagittal were available to us however this was not an option. Unfortunately human adult temporal lobe sections did not survive the antigen retrieval process and resulted in non-specific background staining and therefore prevented us from conducting intra-species analysis.

Our observations of the adult mammalian brain possessing greater 5hmC signal than foetal brains is consistent with literature evidence in which multiple independent groups employed LC-MS, antibody-based dot-blot assays and oxi-bisulfite sequencing techniques to detect and quantify 5hmC magnitudes in these tissues (Munzel *et al*, 2010; Song *et al*, 2011; Szulwach *et al*, 2011; Wang *et al*, 2012; Spiers *et al*, 2017). The detection of 5caC in adult glioblastoma (fig 3.2A-B) and paediatric brain tumour cell lines but not HeLa by confocal microscopy, corroborated by mass spectrometry analysis which generated a similar trend of 5hmC and 5caC abundance substantiates the presence of these epigenetic marks. The mass spectrometry abundance for 5hmC and 5caC correlate with immunostaining signal intensities trends for pluripotent stem cells and LN18 glioblastoma, however UW228-3 demonstrated anomalous 5caC magnitudes 5-fold greater than human embryonic stem cells (fig 3.3A-E) which did not directly correlate with confocal staining levels (fig3.4B). The 5hmC and 5caC signal intensities observable in DKFZ-EP1NS^{+} (high 5caC) cells did correlate with high *TET1* expression however high and low 5caC signal cells would require sorting to produce

separate populations to ensure qPCR analysis discriminates DNA demethylation gene transcriptional profiles of distinct subclones.

High *TET1* expression in UW228-3 correlated with 5hmC and 5caC enrichment however the significant drops in 5hmC and 5caC signal intensities may be attributable to the impact of *siTET1* transfection on *TET2*, *TET3* and *TDG* machinery together as they were down-regulated as well as *TET1* (fig 3.5A-D). The poor *TET1* reduction achieved may have occurred owing to low transfection efficiency or low potency of the siTET siRNA molecule itself. Proper optimisation of the transfection conditions may have resulted in stronger *TET1* knockdown and perhaps more convincing 5hmC and 5caC reduction. However significant *TET1* suppression towards depletion may actually maintain 5hmC and 5caC abundance via compensatory expression of *TET2* and *TET3* (Dawlaty *et al*, 2011). Interestingly, the high *TET1* expression and enriched 5caC levels detected by mass spectrometry may suggest UW228-3 exhibits an epigenetic profile and transcriptional profile closer to pluripotent stem cells than neural stem cells. Analysis of 5hmC and 5caC immunostaining intensities (fig 3.2A-B), abundance via mass spectrometry (fig 6.3A-E) and DNA methylation and demethylation machinery gene expression (figs 3.6A-8 & 3.7A-C) would suggest that adult glioblastoma cell lines possess a unique epigenetic environment located in between adult somatic tissue, the developing foetal brain and reprogrammed somatic tissue in the form of pluripotent stem cells.

Considering the levels of DNA demethylation machinery gene expression are so dramatically reduced in GBM cell lines relative to FB83 even though oximC levels are comparable (fig3.7A-C) between cell lines, is indicative that qPCR data which are relative and not absolute, are not solely reliable estimations of gene expression or protein activity.

The 5hmC and 5caC magnitudes detectable by immunostaining and confocal microscopy only reflect a semi-quantitative assessment of proportions of one oximC signal in one cell relative to another. Owing to the differential sensitivities of 5hmC and 5caC –specific primary antibodies, compounded by fluorophore conjugated secondary antibodies and systems such as Tyramide which produce non-linear signal amplification, the signal intensities of 5hmC and 5caC channels are directly incomparable. Consequently, while it is improper to compare 5hmC signal intensities to those of 5caC as a measure of differential oximC abundance, correlations of signal spatial distribution can be performed as the point-by-point illumination technology of laser scanning confocal microscopy analyses signal intensities of separate channels at the same coordinates. However, the 5hmC and 5caC signal intensity values recorded per pixel across the genome do not accurately reflect real dimensions of oximC presence within the nuclei.

The ratio of antibody to epitope molecule recognition is unknown but the disparity between protein antibody and oximC size suggests steric hindrance would preclude two primary antibody molecules from binding to oximC residues at adjacent CpG sites. Moreover, administration and

binding of fluorophore conjugated secondary antibodies would compound issues of steric hindrance at the sites. Furthermore, all immunostaining for 5hmC and 5caC levels we conducted were performed at x63 magnification and at 2048 x 2048 pixels resolution, which produced 5hmC and 5caC signal intensity profiles with a 65nm resolution limit between adjacent pixels. Considering the 0.34nm distance between adjacent bases on DNA and the supercoiled nature of DNA package within chromatin, adjacent pixels represent approximately 190bp of DNA at this magnification and resolution. Taken together, owing to the non-linearity of scale in signal intensity capture and absolute oximC presence attributable to physical disparities in antibody-epitope dimensions, coupled with the resolution limit of confocal microscopy, 5hmC and 5caC staining magnitudes and distributions only reflect an overall pattern of oximC representation within nuclei. As such, we implemented these techniques to gauge an initial indication of 5hmC and 5caC levels in these cells relative to healthy developmental controls which display patterns within the realm of current literature.

The nature of our antigen retrieval protocol necessary to visualise oximC epitopes requires formaldehyde-fixed and depurinated nuclei which precludes flow cytometry or FACS-based sorting of these cells by their degree of 5hmC or 5caC staining intensity. Co-staining of 5caC with NESTIN, NOTCH1 or SOX2 neural stem cell markers was performed to circumnavigate this technical limitation (fig. 3.10A-F). The immunostaining results did not visually depict 5caC segregation with any

neural stem cell marker in GBM cell lines and showed poor statistical correlation within FB83 neural stem cells themselves. Additionally, NESTIN, NOTCH1 and SOX2 did not mark discrete populations of cells within GBM cell lines thereby enabling visual identification of stem cells. The expression of all stem markers in FB83 may have produced false positive staining effects when applied to GBM cell lines, thereby masking non-specificity of primary and or secondary antibodies for these stem markers. Thus this approach may be too rudimentary for investigating differential oximC staining patterns and could be rectified by, although a controversial, live cell sorting with CD15 and CD133 glioma stem cell markers to produce a side population of cells.

Combination staining of 5caC with NESTIN, NOTCH1 and SOX2 could be performed on sorted cells and analysed to investigate any correlation. Modulation of foetal neural stem cell and GBM transcriptional profiles via stem cell enrichment produced the morphological responses consistent with a neural stem cell phenotype (fig 3.11A-B) and pronounced NESTIN, NOTCH1 and SOX2 staining. However, while stem marker expression appeared uniform within gliomaspheres in GBM cell lines, 5hmC and 5caC staining patterns appeared heterogeneous and, with respect to LN18, distinct high magnitude 5hmC and 5caC cells were observable which bore resemblance to FB83 5caC levels (fig 3.12 A-B).

Stemness enrichment successfully provoked transcriptional profile modulation in foetal neural stem cells and GBM cells which correlated

with increased 5hmC magnitudes but decreased 5caC, consistent with a down-regulation of *TET3* and *WT1* gene expression. Although 5caC staining intensities were not substantially changed between serum and neurobasal conditions in U251, the transcriptional up-regulation of *NESTIN* and *SOX2* neural stemness markers was consistent with increased 5caC enrichment at *NESTIN* promoters however no fold-change in 5caC occupancy was observed on the *SOX2* promoter. This may be indicative of (i) 5caC-independent transcriptional modulation of *SOX2* (ii) *SOX2* promoter 5caC levels were below detection level thresholds required for immunoprecipitation of those sequences or (iii) 5caC modifications were not present within the *SOX2* amplicon but were present further upstream or downstream. These technical considerations could be partly addressed via implementation of more sophisticated 5caC-DIP sequencing approaches which utilise adaptor sequences to unbiasedly ensure sonicated and precipitated genomic DNA fragments are uniformly amplified and sequenced.

Glioblastoma cell lines demonstrated responsiveness to differentiation cues both morphologically and at the transcriptional level with *GFAP* up-regulation over time during glial differentiation (fig 3.14A-C). Beta tubulin III did not show transcriptional up-regulation over time relative to day 0 even though no TUBB3 marker expression was detected at day 0 but strongly pronounced by day 6 (fig 3.15A-D) Interestingly, in contrast to DNA demethylation machinery genes *TET1*, *TET2*, *TET3* and *TDG* show marked reduction as differentiation ensues, the dynamic pattern of 5hmC

and 5caC enrichment and dissipation over time coincides with WT1 up-regulation which remains high after oximC diminishment. The accumulation of oximC presence, particularly 5caC is concentrated centrally within nuclei (fig 3.16), possibly reflective of occupancy within euchromatic DNA (Sadoni *et al*, 2001; Bolzer *et al*, 2005). Markers depicting the expression of genes commensurate with mature neuronal and astrocyte phenotypes e.g. NeuN and S100β (Sarnat *et al*, 1998; Raponi *et al*, 2007) respectively could be chosen to determine the differentiation efficiency of cells and their ability to maintain specialisation following protracted maturation programmes however this was not our intention. Exposure of adult glioblastoma cells to differentiation via paracrine signalling factors was deliberately intended to characterise the active DNA demethylation signatures occurring in cancer cells and determine the level of rigidity or plasticity conferred by the epigenetic configuration on gene expression.

The absence of WT1 expression at the transcriptional or protein level in LN229 highlights the WT1-independent nature of 5caC presence and may be reflecting the same relationship in WT1 positive LN18, U251 and U87MG cells (fig 3.17A-C). Testing this idea via targeted *WT1* knockdown in U87MG and transient WT1 zinc finger domain over-expression in LN229 significantly altered 5hmC and 5caC levels in the respective cells and correlated with modest up-regulation of *TET1* and *TET2*, *TET3* and *TDG* in the U87MG knockdown and marked up-regulation of *TET1*, *TET3* but *TDG* down-regulation in the over-

expression studies, well above the control treatments, indicating some degree of interaction with DNA demethylation machinery. Interestingly, there seems to be some differential affinity for 5caC exhibited between –KTS and +KTS WT1 isoforms with the latter being more potent as a lower abundance of the +KTS domain indicated at the transcriptional and protein level, correlated with proportionally higher 5caC presence. Furthermore, transient *WT1* depletion in U87MG correlated with increased physical occupancy on LINE-1 element retrotransposons which was consistent with elevated LINE-1 transcriptional output. These findings, although correlations, strengthens the evidence for WT1 association with active DNA demethylation processes in adult glioblastoma cell lines.

Design, Construction and Generation of *wt1* mutant GBM Cell Lines

4.1 Introduction

Clustered Regularly Interspersed Palindromic Repeats (CRISPR) Cas9 technology has revolutionised the efficiency of genetic manipulation in multiple mammalian cells types (Ann Ran *et al*, 2013; Wang *et al*, 2013; Dominguez, Lim & Qi, 2016; Tsai & Joung, 2016; Jiang & Doudna 2017). Previously used technologies such as small interference RNA (siRNA) facilitated gene silencing via RNA induced silencing complex/ Argonaute (RISC/AGO) mediated mRNA degradation (Tatiparti *et al*, 2017) can produce significant transcript knockdown but is limited by its transient expression, low efficiency and possible off-target effects which reduce siRNA specificity (Jackson *et al*, 2003; Mohr *et al*, 2014). Zinc finger nucleases (ZFNs) and transcription activator like effector nucleases (TALENs) employ protein DNA-binding domains to bind and cleave DNA in a site specific manner based utilising bespoke subunit combinations targeting genes of interest (Gaj, Gersbach & Barbas III, 2013). However, the construction of bulky 30 amino acid or greater modules in custom combinations and their fusion to FOK-1 endonuclease which require

dimerization is a significant time and labour intensive limitation which is compounded by poor *in vitro* plasmid and protein delivery methods (Gaj, Gersbach & Barbas III, 2013; Mahfouz, Piatek & Stewart Jr, 2014).

CRISPR Cas9 derived from *Streptoccocus pyogenes* functions as an adaptive immune bacterial defence mechanism capable of disrupting foreign infectious DNA elements from attacking host cells (Ann Ran *et al*, 2013; Wang *et al*, 2013; Dominguez, Lim & Qi, 2016; Tsai & Joung, 2016; Jiang & Doudna 2017). Following bacteriophage infection in *S. pyogenes*, fragments of Cas9-cleaved bacteriophage DNA are incorporated into the CRISPR locus enabling synthesis of short CRISPR RNA (crRNA) oligonucleotides specific to the infectious bacteriophage genomic sequence (Gaj, Gersbach & Barbas III, 2013; Mahfouz, Piatek & Stewart Jr, 2014). Transactivating RNA (tracrRNA) which displays a stem-loop secondary RNA structure, functions as a scaffold capable of fusing to crRNA sequences (also termed small guide RNAs in mammalian cells, sgRNA) and subsequently binds to Cas9 endonuclease thereby facilitating the localisation of DNA cleavage machinery to a genomic locus specified by the crRNA (sgRNA) sequence (Ann Ran *et al*, 2013; Gaj, Gersbach & Barbas III, 2013 Wang *et al*, 2013; Mahfouz, Piatek & Stewart Jr, 2014 Dominguez, Lim & Qi, 2016; Tsai & Joung, 2016; Jiang & Doudna 2017). The Proto-spacer activation motif (PAM), a 3 nucleotide sequence typically '5-NGG-3' where 'N' denotes any base, is located directly adjacent and upstream to the crRNA hybridising region on DNA target genome. The PAM sequence,

recognised by Cas9 endonuclease designates the cleavage site as between the 3rd and 4th base downstream from the 'N' base of the PAM sequence (Ann Ran *et al*, 2013; Gaj, Gersbach & Barbas lll, 2013 Wang *et al*, 2013; Mahfouz, Piatek & Stewart Jr, 2014 Dominguez, Lim & Qi, 2016; Tsai & Joung, 2016; Jiang & Doudna 2017).

Additionally, while wildtype Cas9 activity involves symmetric double-strand endonuclease cleavage dependent on HNH and RuvC domains, single strand nick or nickase variants have been generated using mutants of either HNH or RuvC domains ($HNH^+/RuvC^-$ or $HNH^-/RuvC^+$) (Hsu, Lander & Zhang, 2014). Alternatively, a D10A RuvC mutant capable of nicking DNA targets as designated by sgRNAs can, in concert with inverted D10A nickase endonuclease bound to a separate sgRNA species with specificity to an additional sequence within the target region, create a nick on the opposite DNA strand. However, both D10A nickase – sgRNA complexes must be within 20 nucleotides of one another (Hsu, Lander & Zhang, 2014). The localisation of CRISPR/D10A nickase Cas9 complexes in reverse orientations to one another at DNA target sites increases the specificity of cleavage as effective double strand cleavage can only be accomplished if both complexes are present and within the required distance from each other, and greatly increases the cleavage site, creating 5' overhangs and 3' underhangs, which may be repaired by non-homologous end joining or homology directed repair (Ann Ran *et al*, 2013; Gaj, Gersbach & Barbas lll, 2013 Wang *et al*, 2013; Hsu, Lander & Zhang, 2014; Mahfouz, Piatek & Stewart Jr, 2014 Dominguez, Lim &

Qi, 2016; Tsai & Joung, 2016; Jiang & Doudna 2017). We utilised CRISPR-Cas9 technology to create *WT1* mutant LN18 and U251 glioblastoma cell lines to examine WT1-5caC 'reader' interactions and relationship with DNA demethylation.

4.2 Rationale for Selection of WT1 Target Site

Our siRNA mediated *WT1* knockdown and *WT1* zinc finger domain over-expression studies demonstrated a correlation between *WT1* gene expression and 5caC immunodetection, promoter DNA occupancy and *TET* gene transcription profiles, suggesting an interplay between *WT1* and DNA demethylation machinery. Owing to the heterogeneous nature of siRNA and non-integrating plasmid delivery and their transient expression and subsequent degradation, the protracted effects of *WT1* modulation could not examined. Thus *WT1* CRISPR cas9 knockout constructs were designed to facilitate the generation of *wt1* null mutant cell lines possessing perturbed DNA binding ability. The clonal genotype of these lines would ameliorate confounding issues of population heterogeneity obfuscating the interpretation of sporadic 5caC distribution and DNA demethylation machinery transcriptional profiles. Multiple studies have demonstrated that the zinc finger domains confer DNA binding ability to WT1 (Little *et al*, 1992; Josina *et al*, 1995; Stoll *et al*, 2007) with zinc fingers 2, 3 and 4 responsible for enabling sequence specific localisation while zinc finger 1 contributes a DNA anchorage function to the protein (Laity *et al*, 2000; Nurmemmedov *et al*, 2009). Alternative splice variants of the Lysine-Serine-Threonine (KTS) amino

acid triplet sequence can occur between zinc fingers 3 and 4. The KTS insertion (+KTS) results in significant distortions in the linker sequence between the two terminal fingers which consequently perturbs zinc 4 interaction with the DNA major grove, thus lowering WT1 DNA binding affinity relative to the –KTS isoform. Therefore the –KTS variant possesses higher DNA binding affinity and sequence specificity compared to the +KTS isoform which has been suggested to bequeath a splicing regulatory function to WT1 via putative RNA binding affinity afforded by the three amino acid insertion (Larsson *et al*, 1995; Barbaux *et al*, 1997; Laity *et al*, 2000). Moreover, within the DNA binding zinc finger domain, recognition and binding of oximC DNA modifications can be attributable to conserved residues glutamine 369 (Q369) and glutamic acid 427 (E427) residing in zinc fingers 2 and 4 respectively (Hashimoto *et al*, 2014; Hashimoto *et al*, 2016).

4.2.1 WT1 Targeting Strategy

To generate clonal lines *WT1* exons 8 and 9 encoding zinc fingers 2 and 3 respectively were selected for targeting to disrupt both the transcription factor's oximC and DNA binding ability (Fig 4.1A-C). The resulting mutant truncated protein would possess only zinc finger 1 of its zinc finger motif as the cassette incorporation via homologous recombination would delete exons 8 and 9 and disrupt the reading frame of exon 10. Moreover, the termination and bovine growth hormone (bGH) poly adenylation (polyA) signal of the cassette upstream of intron 9 would induce premature transcriptional termination prior to RNA polymerase reading

of the exon 10 sequence. Successful generation of mutant clonal lines would also enable analysis of oximC patterns during functional assays such as stem enrichment and differentiation as well as tumour formation *in vivo*.

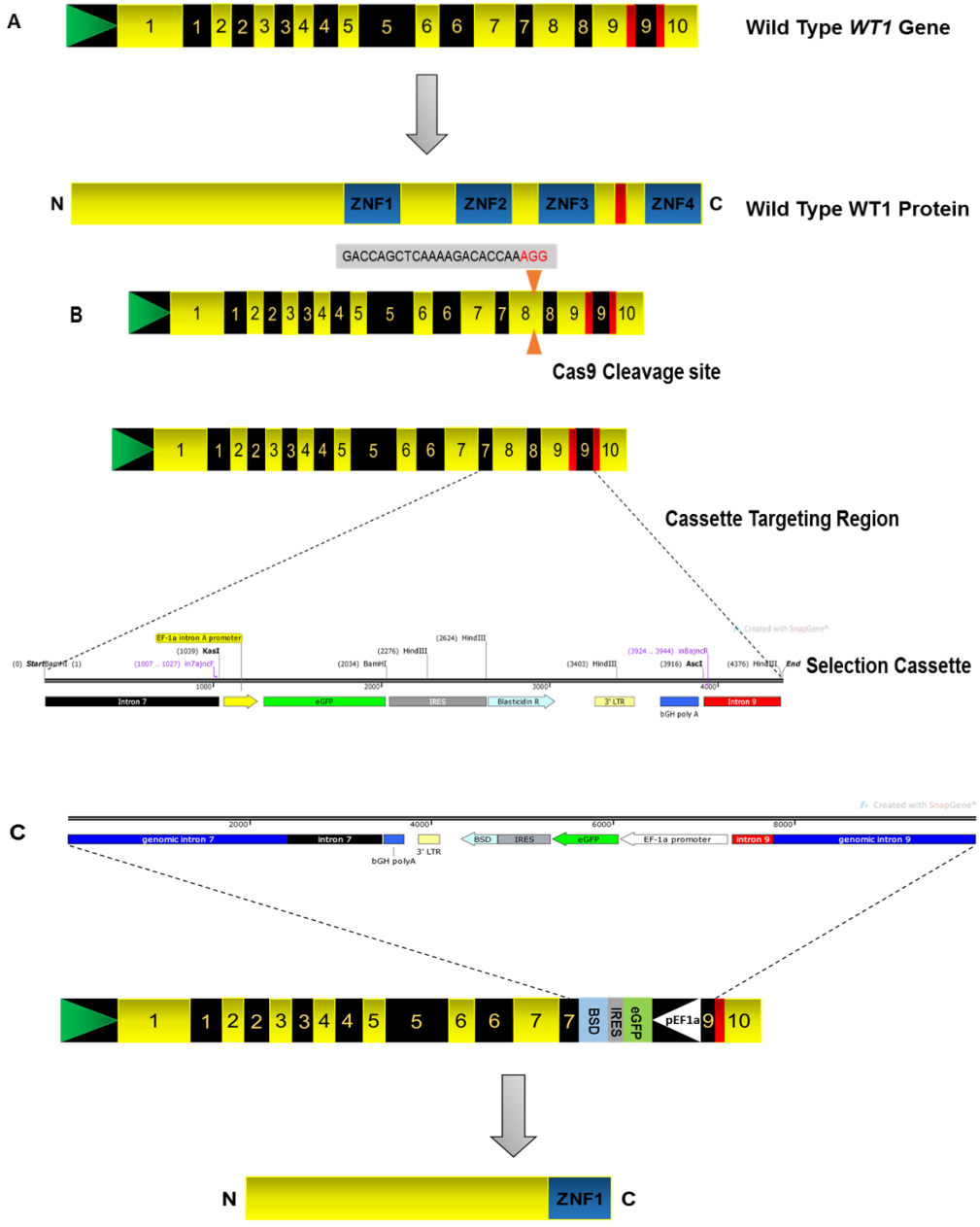
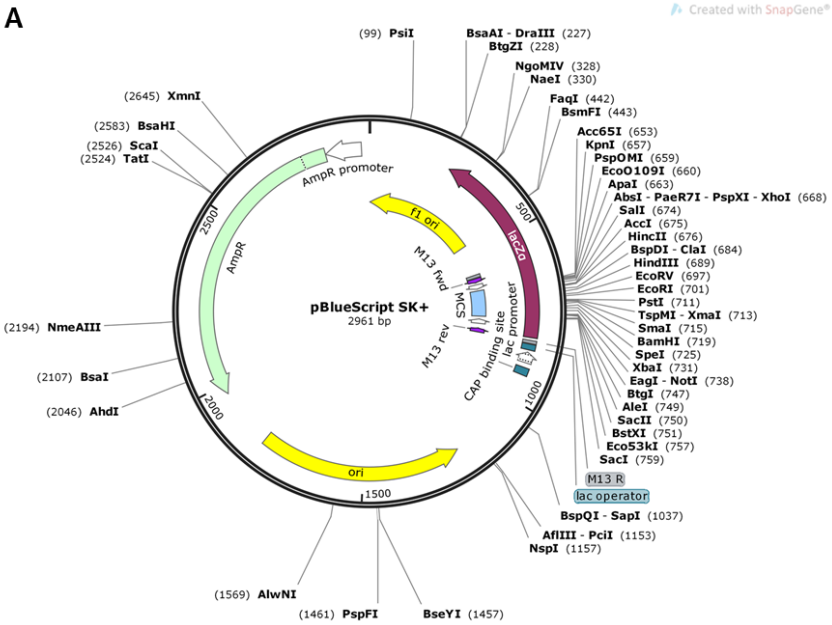


Figure 4.1 *WT1* DNA binding domain (exons 7-10) targeting strategy. (A) The wild type *WT1* gene encodes a transcription factor with four zinc fingers

(ZNF) in its DNA binding C terminal domain. The 3 amino acid KTS sequence (red) exists between zinc fingers 3 and 4. (B) The small guide RNA (grey) in relation to the genomic PAM sequence (red) designates the site of CRISPR cas9 mediated cleavage of *WT1* within exon 8 which facilitates incorporation of the selection cassette into this cleavage point by homology directed repair. (C) Incorporation of the selection cassette into the targeted *WT1* locus would disrupt the coding sequence of the DNA binding domain by deletion of exons 8 and 9 encoding zinc fingers 2 and 3 respectively, yielding a truncated, DNA binding null mutant WT1 protein.

The generation of an exons 8-9 targeting donor plasmid was contingent on the generous donation of parental vectors pSIN-GFP-IRES-Blasticidin (hereafter referred to as pSIN) and pBluescript (hereafter referred to as pBS) by Dr Alexander Kondroshov (fig 4.2A-B).



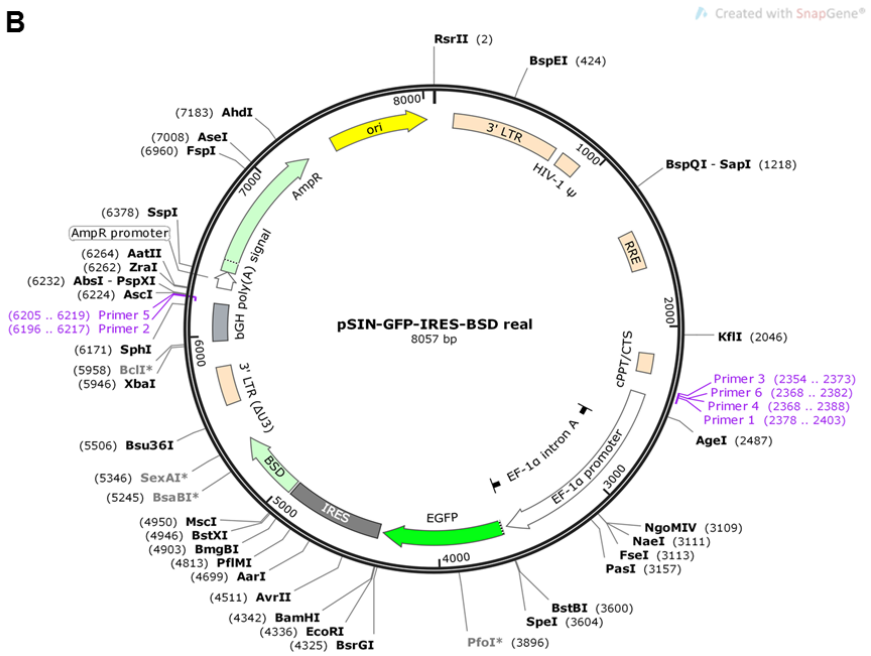


Figure 4. 2 Parent plasmids utilised for generating (A) intermediate and (B) Donor HDR plasmids.

The targeting strategy (fig 4.3) for synthesis of the donor plasmid entailed subcloning the GFP-IRES-Blasticidin-bGH polyA sequence from pSIN vector into a pBS intermediate plasmid containing left and right homologous arms with sequence identity to *WT1* introns 7 and 9 respectively. Subcloning of the GFP-IRES-BSD-bGH polyA fragment into the pBS vector derived intermediate plasmid was devised to limit the size of the resulting plasmid to enable greater efficiency of transfection. Additionally, lack of unique restriction sites upstream of the EF-1 alpha promoter reduced the flexibility of directional cloning options for the incorporation of WT1 homology arms.

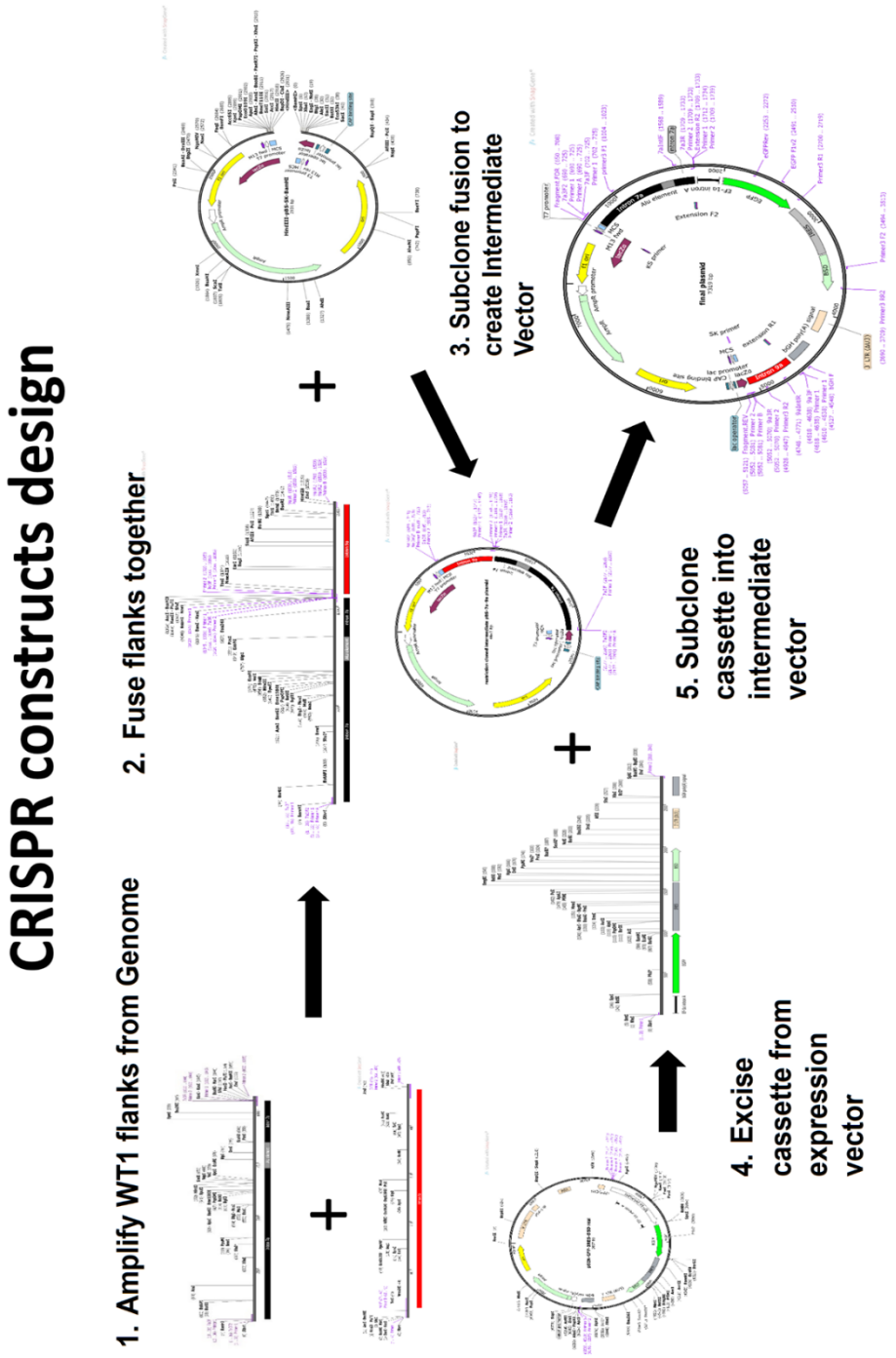


Figure 4.3 Schematic of *WT1* zinc finger targeting HDR plasmid construction.

4.2.2 Intermediate Plasmid Construction

An 'intermediate plasmid' consisting of *WT1* intron 7 left and intron 9 right homology arms was designed to facilitated efficient restriction cloning of the pSIN GFP-IRES-BSD-bGH polyA cassette into the pBS vector such that introns 7and 9 would flank the cassette upstream and downstream respectively. The intermediate plasmid, once harbouring the pSIN cassette would become the homologous donor plasmid. Examination of pBS and pSIN vectors revealed *HindIII* and *BamHI* unique restriction sites on pBS and *KasI* and *AscI* sites on pSIN which could serve as cleavage points for directional cloning of *WT1* intron 7 and 9 arms and GFP-IRES-BSD-bGH polyA cassette fragment into pBS respectively. Analysis of the *WT1* genomic zinc finger encoding region via the University of California Santa Cruz (UCSC) Genome Browser (fig 4.4) (https://genome.ucsc.edu/) revealed *Alu* Short Interspersed Nuclear Element (SINE) repetitive sequences (fig 4.4 red dashed boxes) present in introns 7 (304bp) and 9 (563bp) which, we predicted may interfere with homologous recombination. Thus intron 7 and intron 9 homology arms were shortened to 1065bp and 485bp to avoid these repeat sequences.

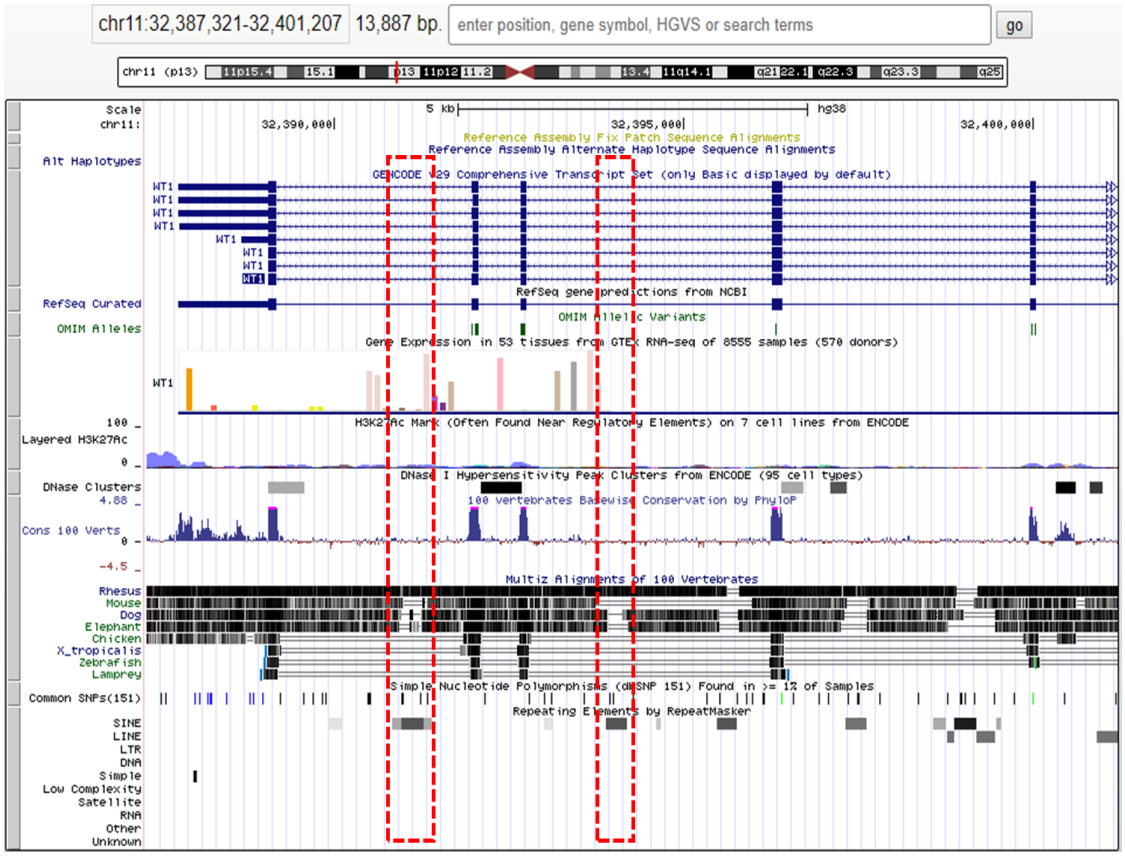
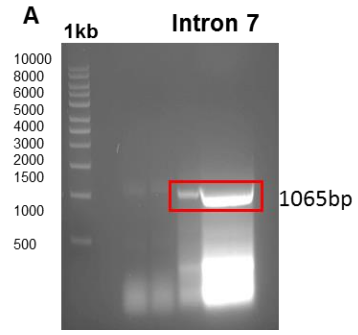
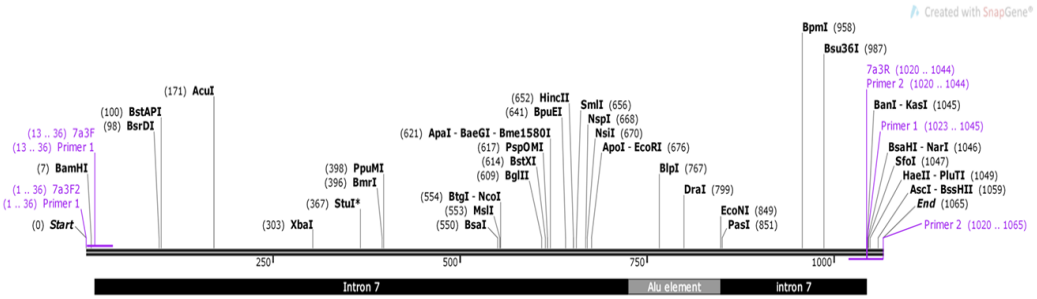


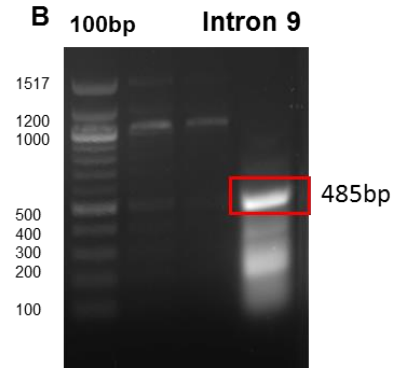
Figure 4.4 UCSC Genome Browser alignment of WT1 gene sequence with genomic features of interest. Red dashed boxes demarcate regions harbouring SINE repetitive elements within introns 7 and 9.

Forward and reverse primers 7a3F2: 5'-GGATCTCATTAAAGGCAACCTCTCC-3' and 7a3R: 5'-GGCGCGCCGCATTAAGGCGCCGGATCTCATTAAAGGCAACCTCTCC -3' and 9a3F2: 5'-GGCGCCTTAATGCGGCGCGCCGAGAGAAGGAATAGTGCGTGG-3' and 9a3R: 5'- TCTAGAAGCTTCACCAGCCTTGCAGAAAGC -3' respectively were used to amplify intron 7 and 9 fragments from the U251 genomic *WT1* sequence (fig 4.5A-B). Primers 7a3F2 and 9a3R possessed *BamHI* and *HindIII* sites within their 5' sequences

respectively whilst 7a3R and 9a3F2 possessed *KasI* and *AscI* sites to facilitate the restriction digest-based subcloning of the GFP-IRES-BSD-bGH polyA fragment into an intermediate vector containing the fused intron 7 and 9 flanks.







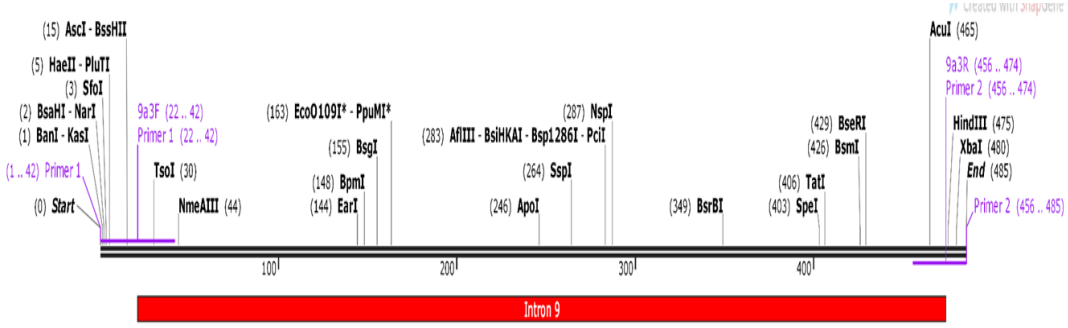


Figure 4.5 Amplification of wild type WT1 introns 7 and 9 from U87MG genome. (A) PCR amplification of 1065bp intron 7 (red box). Black schematic represents intron 7 amplicon with 300bp *Alu* element depicted in grey. (B) PCR amplification of 485bp intron 9 (red box). Red schematic represents intron 9 amplicon.

As both intron 7a3R and 9a3F2 primers contain overlapping sequences harbouring *KasI* and *AscI* restriction sites, the resulting fusion consequent of *KasI* digestion of both intron 7 and 9 flanks and their subsequent ligation (fig 4.6A-B) produced a linear fragment of introns 7-9 with BamHI and and HindIII sites at the intron 7 and intron 9 terminal ends and *KasI* and *AscI* sites existing at the fusion junction. Fusion of the two flanks was confirmed by sequencing, with the fusion junction depicted in the schematic and chromatogram (fig 4.6B-C).

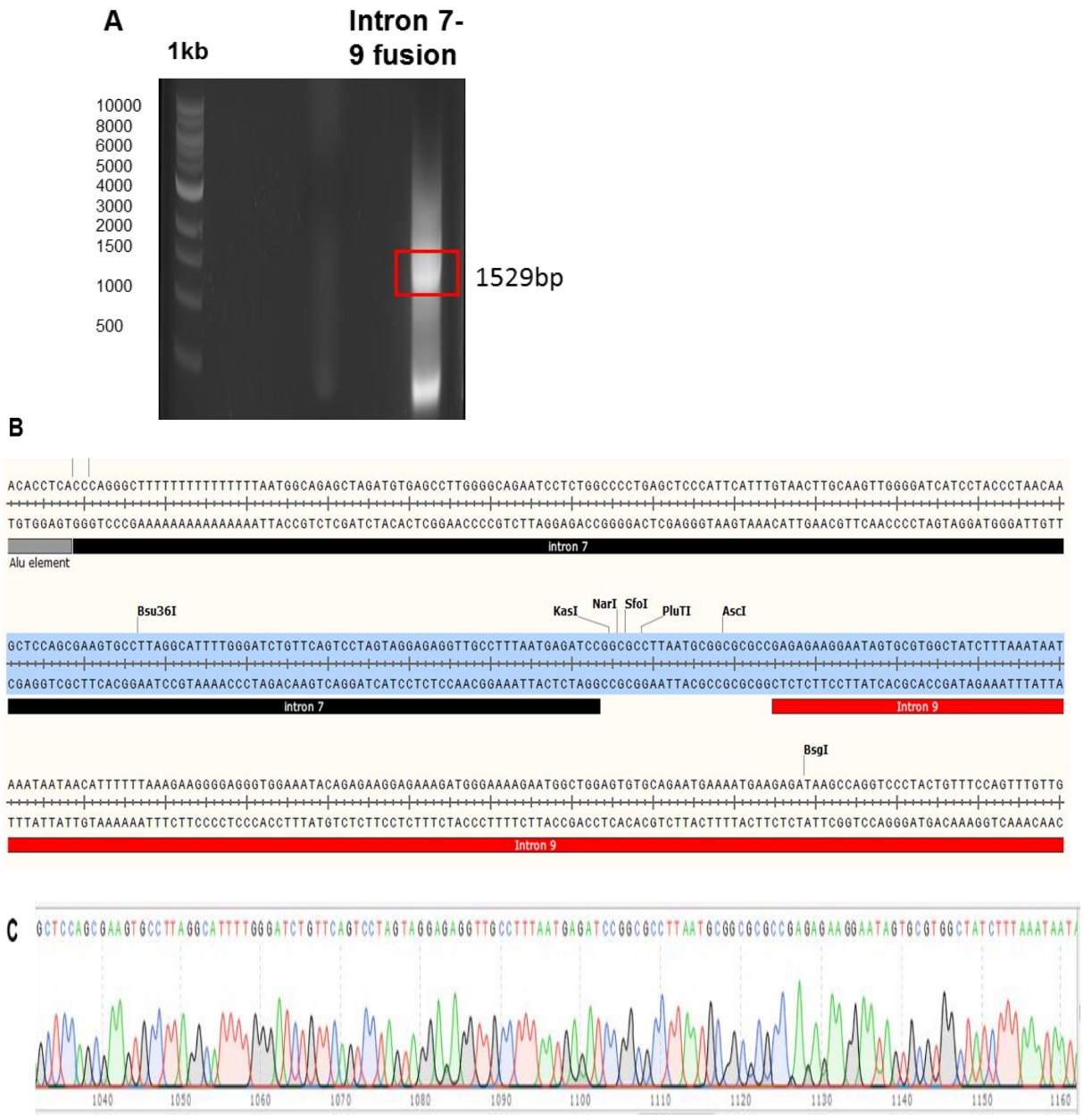


Figure 4.6 Fusion of *WT1* intron 7 and 9 homologous flanks. (A) Agarose gel electrophoresis of ligated 1529bp intron 7-9 fusion (red box). (B) Schematic of intron 7- 9 fusion. Blue highlighted region represents the sequence depicted in chromatogram (C) verifying the fusion junction between introns 7-9.

Incorporation of the *WT1* intron 7-9 fusion fragment was achieved by *BamHI* and *HindIII* double digestion of both the fusion fragment and the

pBS vector and subsequent ligation to form the intermediate plasmid (fig 4.7)

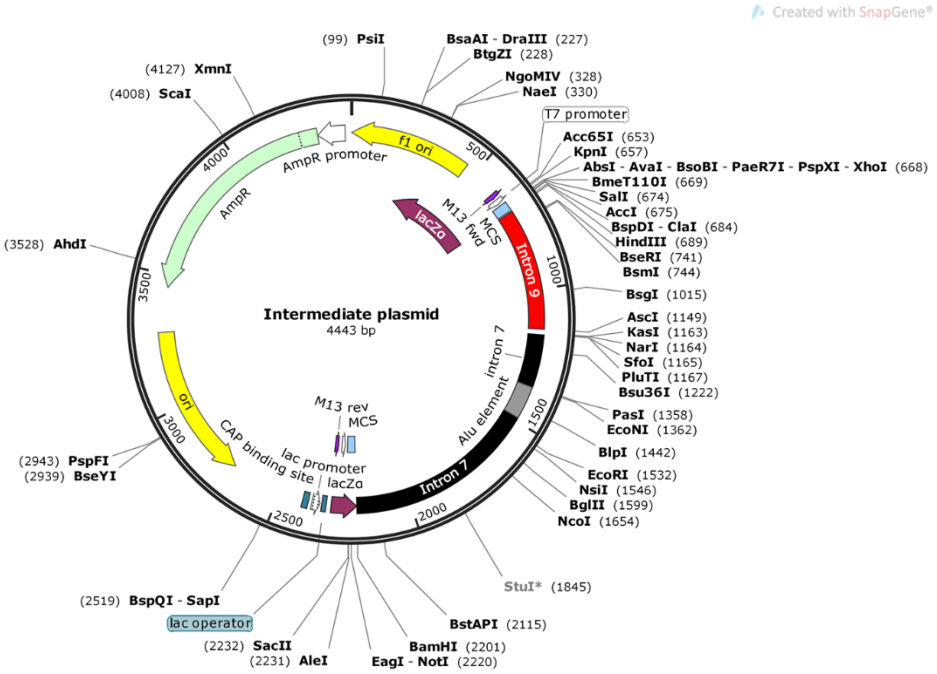


Figure 4.7 Intermediate plasmid. Schematic representation of the Intermediate plasmid containing fused *WT1* intron 7 (black) and 9 (red) homologous flanks containing *AscI* and *KasI* restriction sites located at the intron 7 – 9 junction. The intron 7 – 9 fused construct was subcloned into the pBS vector.

Sequencing of the pBS-intron 7 arm and pBS-intron 9 arm junctions was performed to confirm plasmid structure (fig 4.8A-B)

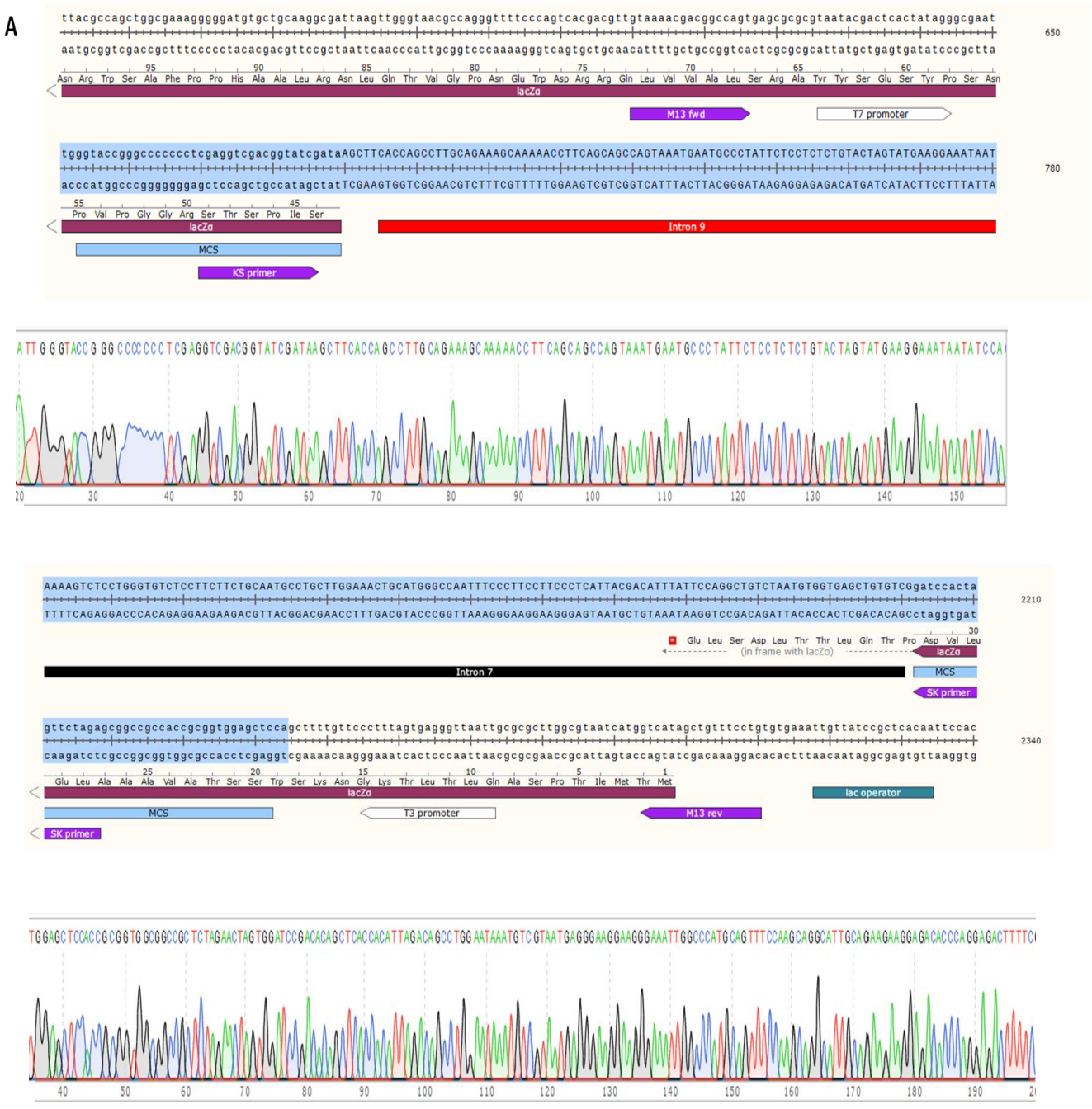
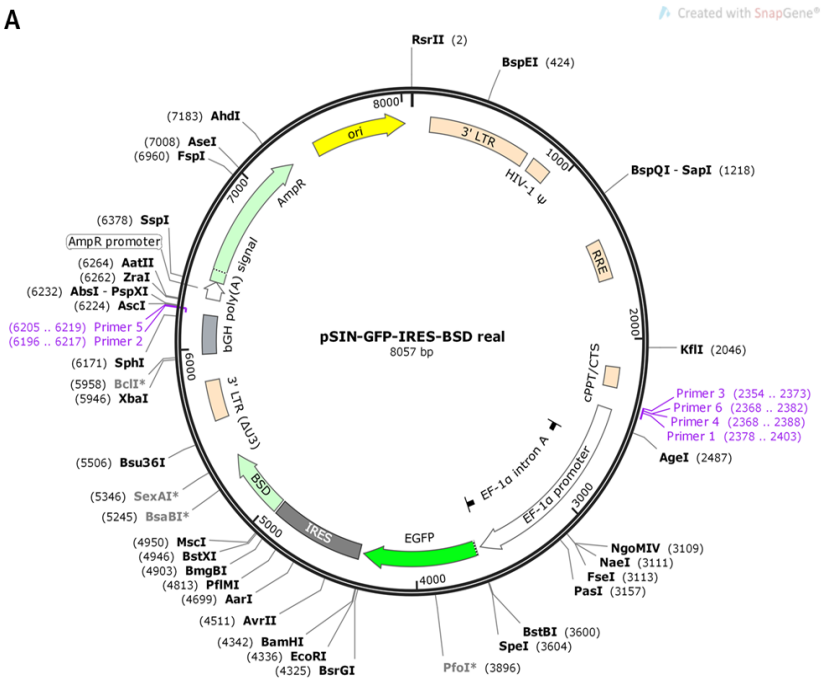


Figure 4.8 Verification of *WT1* intron 7 – 9 fusion incorporation into pBS vector. The highlighted blue regions in schematics (A) & (B) illustrate the pBS - intron 9 and intron 7 - pBS junctions respectively, the sequences of which, depicted in their respective chromatograms.

4.2.3 WT1 HDR Donor Plasmid Construction

Owing to lack of restriction digests sites flanking the full length pEF-1a-GFP-IRES-BSD-bGH polyA cassette (incorrect annotation of pSIN plasmid map only depicted a minimal pEF-1a promoter with a *KasI* site upstream and *AscI* downstream of the bGH polyA terminator respectively), fusion primers FLcas9F: 5'-ACGCACTATTCCTTCTCTCGGGCAGCTAATGGACCTTCT-3' and FLcas9R: 5'-GGTTGCCTTTAATGAGATCCGCCATAGAGCCCACCGCA-3' which partially bind to pEF-1a and bGH polyA terminator respectively were utilised to facilitate subcloning of the amplified cassette fragment (fig4.9A-B) into *AscI* digested intermediate plasmid.



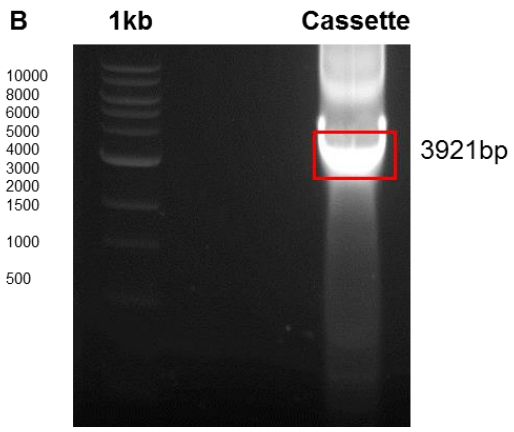


Figure 4.9 Donor HDR parent pSIN-GFP-IRES-BSD plasmid. (A) Schematic of the parent plasmid containing the pEF-1a-eGFP-IRES-BSD selection cassette required for donor plasmid generation. (B) PCR amplification of the selection cassette (red box) from the intact pSIN-GFP-IRES-BSD plasmid.

Extraction and purification of the amplified cassette was performed and subcloned into the digested intermediate plasmid via Gibson Assembly to produce the WT1 homology directed repair (HDR) donor cassette (fig 4.10) which was used for targeting *WT1* exons 8 and 9.

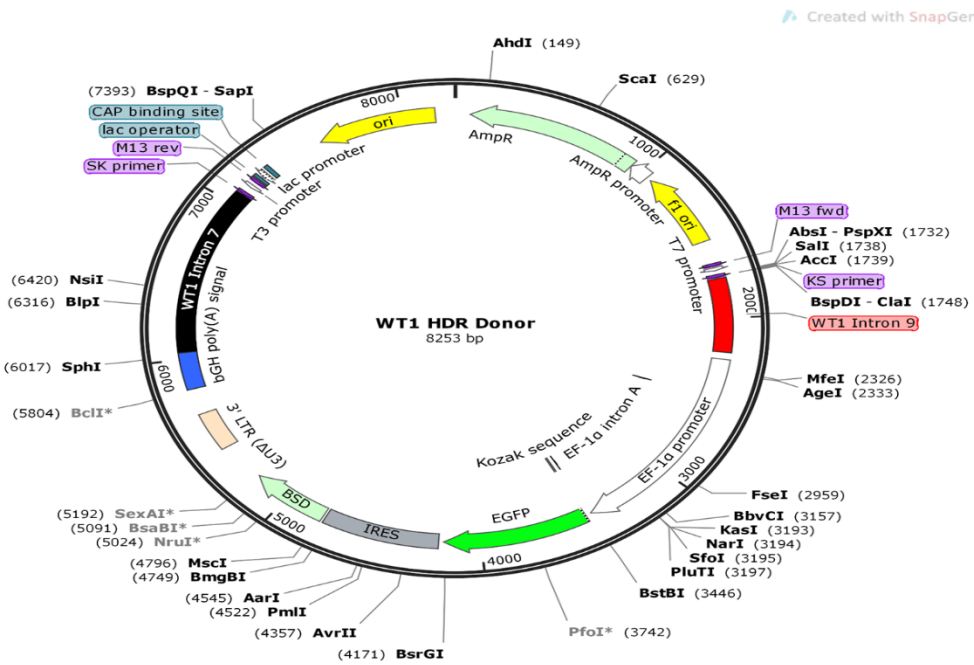


Figure 4.10 Complete WT1 targeting HDR donor plasmid. The plasmid contains introns 7 (black) and 9 (red) homologous arms flanking the incorporated pEF-1a-eGFP-IRES-BSD selection cassette which had been subcloned into the intermediate plasmid.

Sequencing was performed on donor plasmids to verify the *WT1* Intron 7-bGH polyA and *WT1* intron 9-pEF-1a junctions to verify incorporation of the cassette between the WT1 homology arms (fig 4.11A-B).



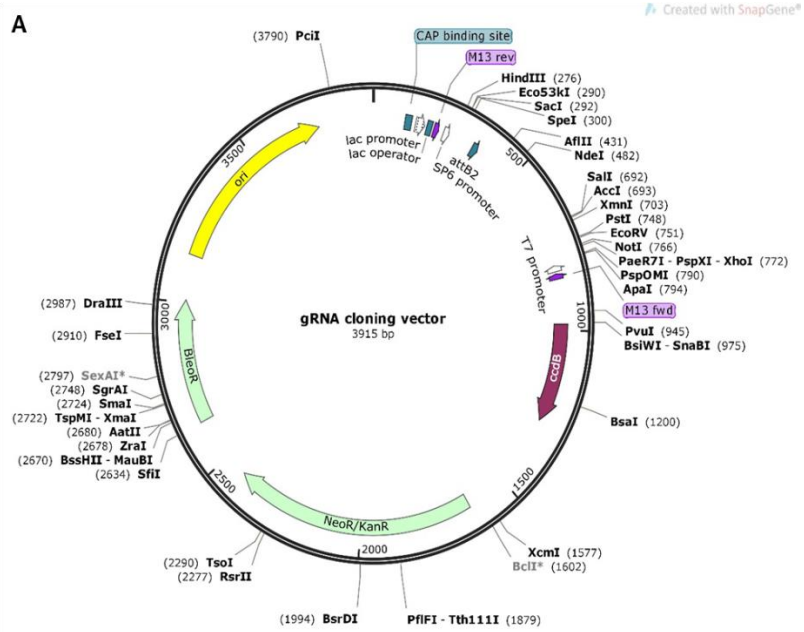


Figure 4.11 Verification of pEF-1a-eGFP-IRES-BSD selection cassette incorporation into the intermediate plasmid. (A-B) Schematics of intron 9 flank - p-EF-1a and bGH polyA terminator- intron 7 flank junctions respectively. The blue highlighted regions illustrate the sequences depicted in each of the associated chromatograms.

4.2.4 Construction of Guide RNA Plasmids

Guide RNAs specific to exon 8 were generated using Feng Zhang's lab online CRISPR guide RNA tool (https://zlab.bio/guide-design-resources). The two highest ranked guides based on high binding specificity and fewest number of mismatched off-targets were selected which possessed sequences 5'-GACCAGCTCAAAAGACACCAA-3' (guide 8A) and 5'- CACAGTCCTTGAAGTCACAC-3' (guide 8B) and were ranked 94% and 91% respectively. To enable guide RNA transfection and expression within cells, 61bp primers were designed to include 40 bases of RNA pol III U6 promoter upstream of the 21 nucleotide guide RNA sequence in forward primers and 40 bases of small nuclear RNA scaffold upstream of 21 nucleotides of the guide RNA reverse complement i.e.

Guide 8A forward primer: 5'-TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGACCAGCTCAAAAGACACCAA-3' and Guide 8A reverse primer: 5'-GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTTGGTGTCTTTTGAGCTGGTC-3'. Linearisation of guide RNA cloning vector (fig 4.12A-B) through digestion of *AflII* restriction site located between a U6 promoter and guide RNA scaffold sequence facilitated the subcloning of annealed forward and reverse guide 8A cloning primers.



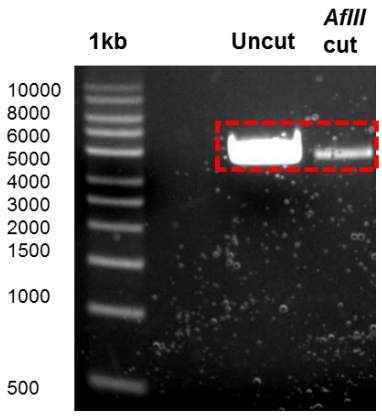
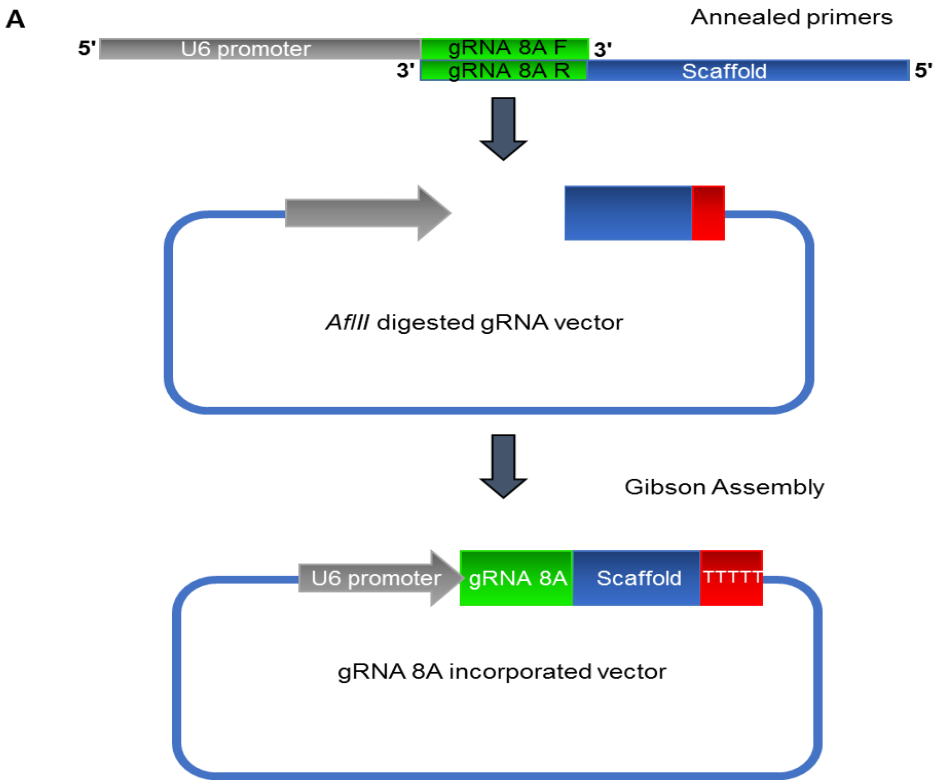


Figure 4.12 small guide RNA (gRNA) cloning vector. (A) Schematic of the gRNA cloning vector. (B) Digestion of gRNA cloning vector with *AflII*.

Primers annealed at their complementary sequences to produce heterodimers capable of hybridising to the linearised guide RNA cloning vector (fig 4.13A-B) via Gibson Assembly.



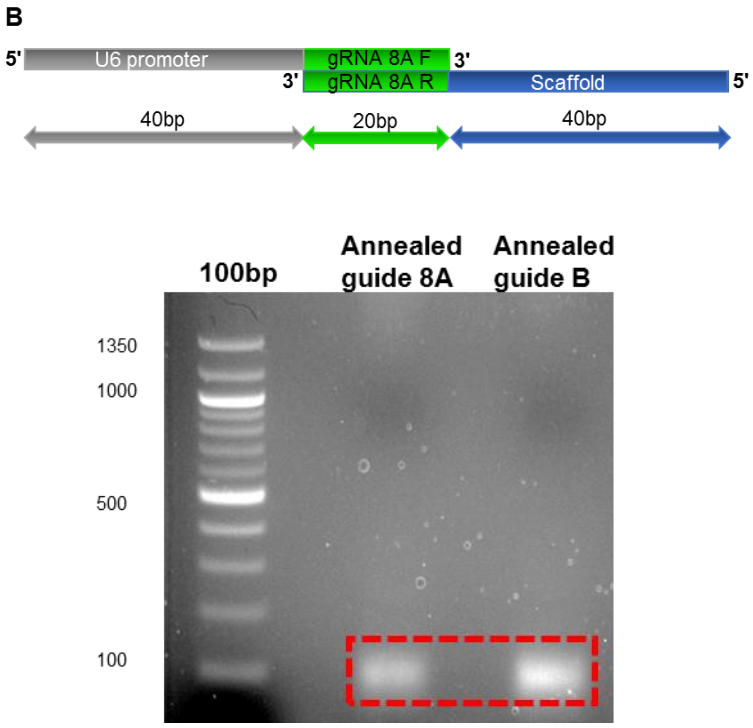
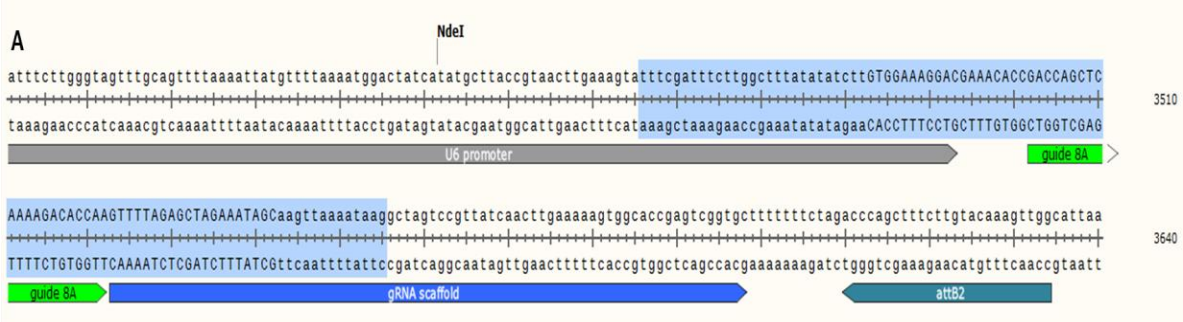


Figure 4.13 Gibson Assembly incorporation of small guide RNAs into vectors. (A) Forward and reverse sgRNA primers are annealed together and incorporated into *AflII* digested gRNA cloning vector by Gibson Assembly. (B) Agarose gel confirming the presence of annealed sgRNAa 8A and 8B (red dashed box, 100bp).

Guide 8A RNA cloning vectors were sequenced to verify guide location and fidelity in preparation for transfection (fig 4.14).



B

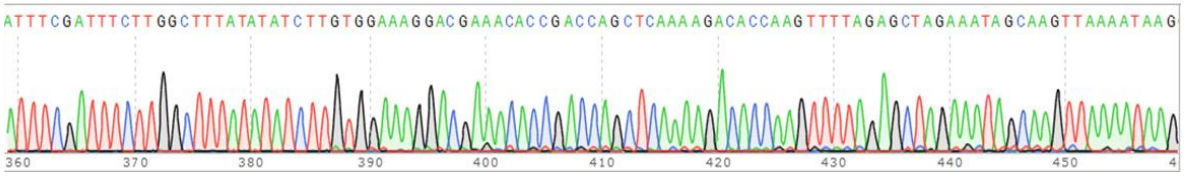


Figure 4.14 Verification of sgRNA incorporation into gRNA cloning vector. (A-B) Schematic of sgRNA insertion between U6 promoter and scaffold sequences, highlighted by blue region and illustrated by chromatogram (B).

4.3 Transfection Optimisation and Selection

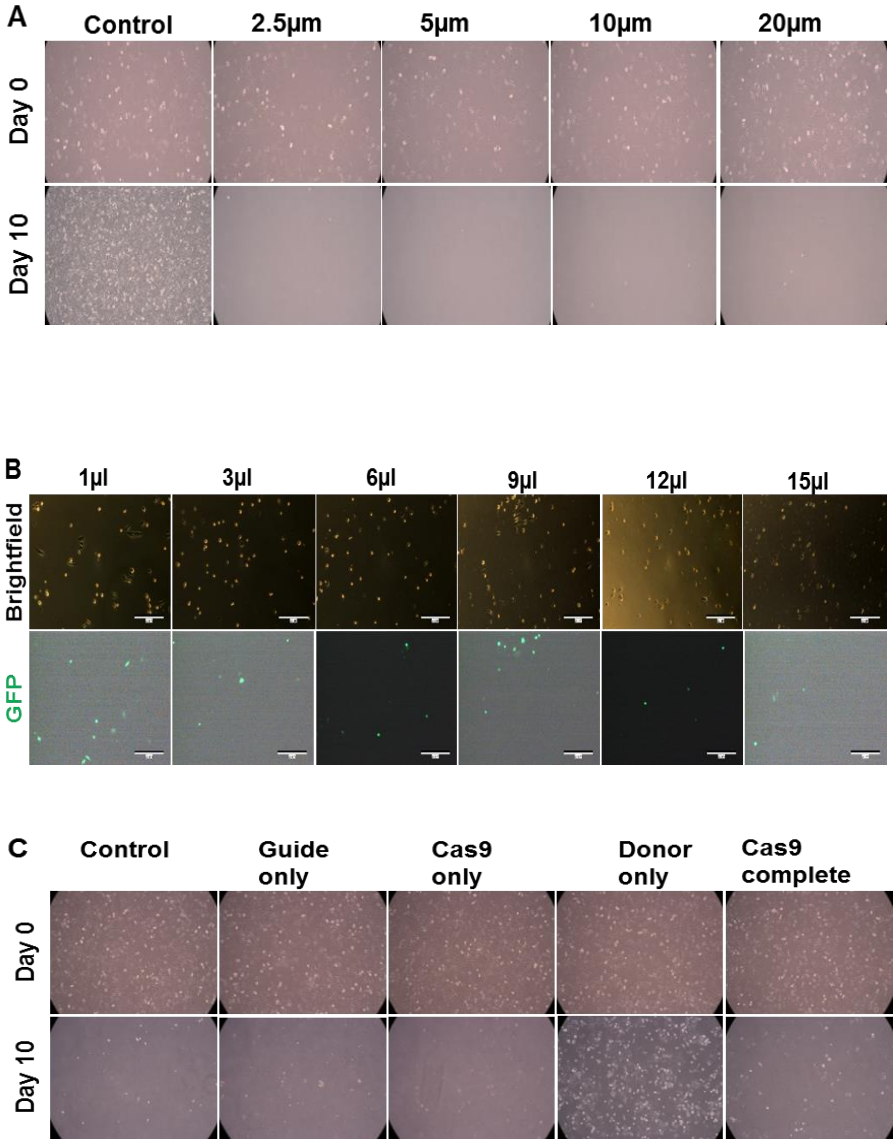
Successful construction of plasmids for cas9 and gRNA delivery and homology directed repair contributed to the first step of a tripartite stage process involving generation of CRISPR targeted mutant *wt1* GBM cell lines, the second and third stages entailing transfection and selection and clone screening respectively. Considering the GFP and blasticidin (BSD) selection markers present in the cassette, two-way screening would be performed to isolate positively transfected cells to increase the probability of successfully targeted mutants. Owing to their high WT1 expression, LN18 and U251 cell lines were chosen for targeting.

4.3.1 Survival Analysis and Transfection Optimisation

Blasticidin kill curves were conducted to determine the minimum antibiotic dosage to kill 50% of untransfected cells (LD_{50}). Literature examination recommended 2.5-10µm of Blasticidin for U251 and LN18 cell lines (Radu *et al*, 2003; Levitt *et al*, 2005; Zhang *et al*, 2005; Wang *et al*, 2017; Zhang *et al*, 2017) so blasticidin was administered at ranges

of 2.5-20µm (fig4.15A) to be sure if our cell line response and blasticidin potency was consistent with the literature. The minimal efficient lethal blasticidin dosage was 2.5µm for both LN18 and U251 which killed all cells after 10 days of treatment. Lipofectamine produced low efficiency of transfection success, evident from GFP plasmid positive control signals (fig 4.15B). Transfection efficiencies at 1-15ul of reagent were 26.2%, 16.6%, 17.1%, 34.4%, 10%, 11% respectively. Considering the increased cytotoxicity of lipofectamine at elevated concentrations and concordant drop in transfection efficiency, 2ul of lipofectamine was selected as optimal for transfection. To confirm that blasticidin resistance could only be attributable to possession of the BSD gene present on the donor plasmid, individual vectors were administered under BSD selection conditions (fig 4.15C). As expected, donor only and Cas9 complete (Cas9 plasmid + gRNA plasmid + donor plasmid) conditions exhibited surviving cells after the treatment duration. Dramatic disparity in surviving cells was observed between donor only and cas9 complete treatments during BSD selection. Transfections were repeated under either BSD or no BSD conditions to understand whether cell death observed in Cas9 complete treatments was attributable to BSD effects (fig 4.15D). No observable difference was present between no BSD and BSD treated donor only or Cas9 complete cells. We reasoned that transfection of only one plasmid in the donor only treatments conferred greater survival to cells relative to Cas9 complete in which three different plasmids may compete for cellular entry. Moreover, only BSD resistance conferred by the BSD resistance cassette present on the donor plasmid would enable

survival of cells and thus Cas9 complete treated cells have reduced probability of survival. To overcome this limitation, cells were transfected by 'nucleofection' electroporation (fig 4.15E) which produced a greater transfection efficiency. Multiple cells possessing bright GFP signal in Cas9 complete treatment were observed growing as large colonies within a background monolayer of GFP positive cells. Sorting of surviving cells GFP positive cells would be performed to establish the clonal density required for genotypic screening.



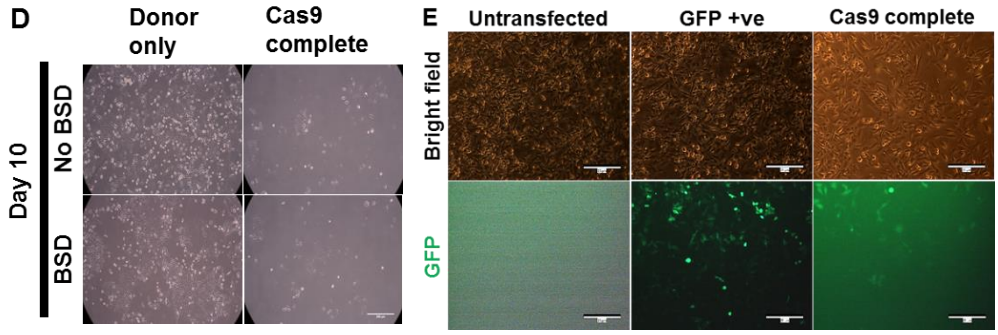
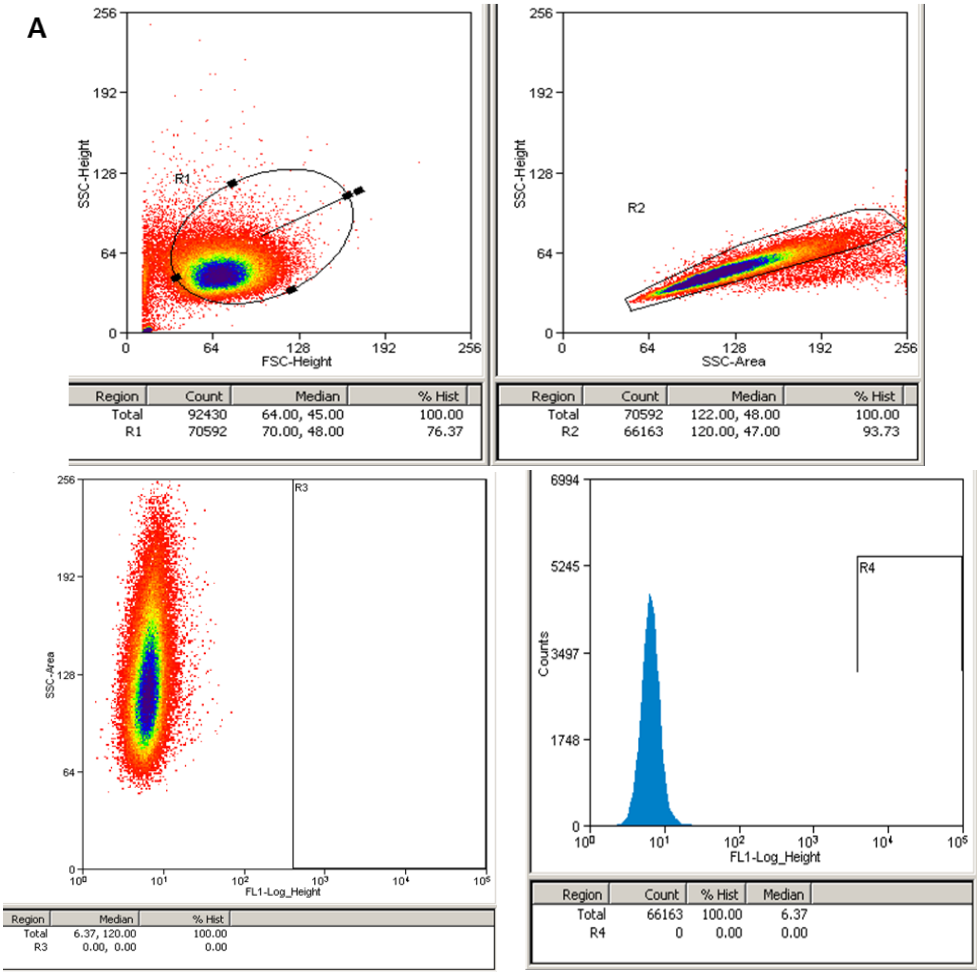


Figure 4.15 Blasticidin kill curve and plasmid transfection optimisation in U251. (A) Blasticidin kill curve. (B) GFP plasmid transfection optimisation using Lipofectamine 2000. (C) Confirmation of Donor plasmid blasticidin resistance. (D) Determination of lethality of multi-plasmid transfection. (E) U251 transfection via Nucleofection.

4.3.2 FAC Sorting of GFP Positive Survivors

GFP U251 and LN18 surviving cells were analysed for GFP signal against untransfected control cells for each cell line (fig 4.16A). Initially gating analysis on U251 using forward scatter versus side scatter revealed 76% GFP positive cells (fig 4.16A, R1) which were sorted to produce a 93% enriched population (R2). Side scatter height versus side scatter area gating was performed on 93% GFP positive population and doublets were excluded. Sorting produced a 100% GFP positive population which was subsequently gated, analysed, enriched and plated to yield the highest GFP signal cells which were sorted into 96 well plates with one cell per well to achieve clonal lines. The same pipeline was used for sorting on LN18 transfected cells (fig 4.16B).

A



Region	Count	Median	% Hist
Total	92430	64.00, 45.00	100.00
R1	70592	70.00, 48.00	76.37

Region	Count	Median	% Hist
Total	70592	122.00, 48.00	100.00
R2	66163	120.00, 47.00	93.73

Region	Median	% Hist
Total	6.37, 120.00	100.00
R3	0.00, 0.00	0.00

Region	Count	% Hist	Median
Total	66163	100.00	6.37
R4	0	0.00	0.00

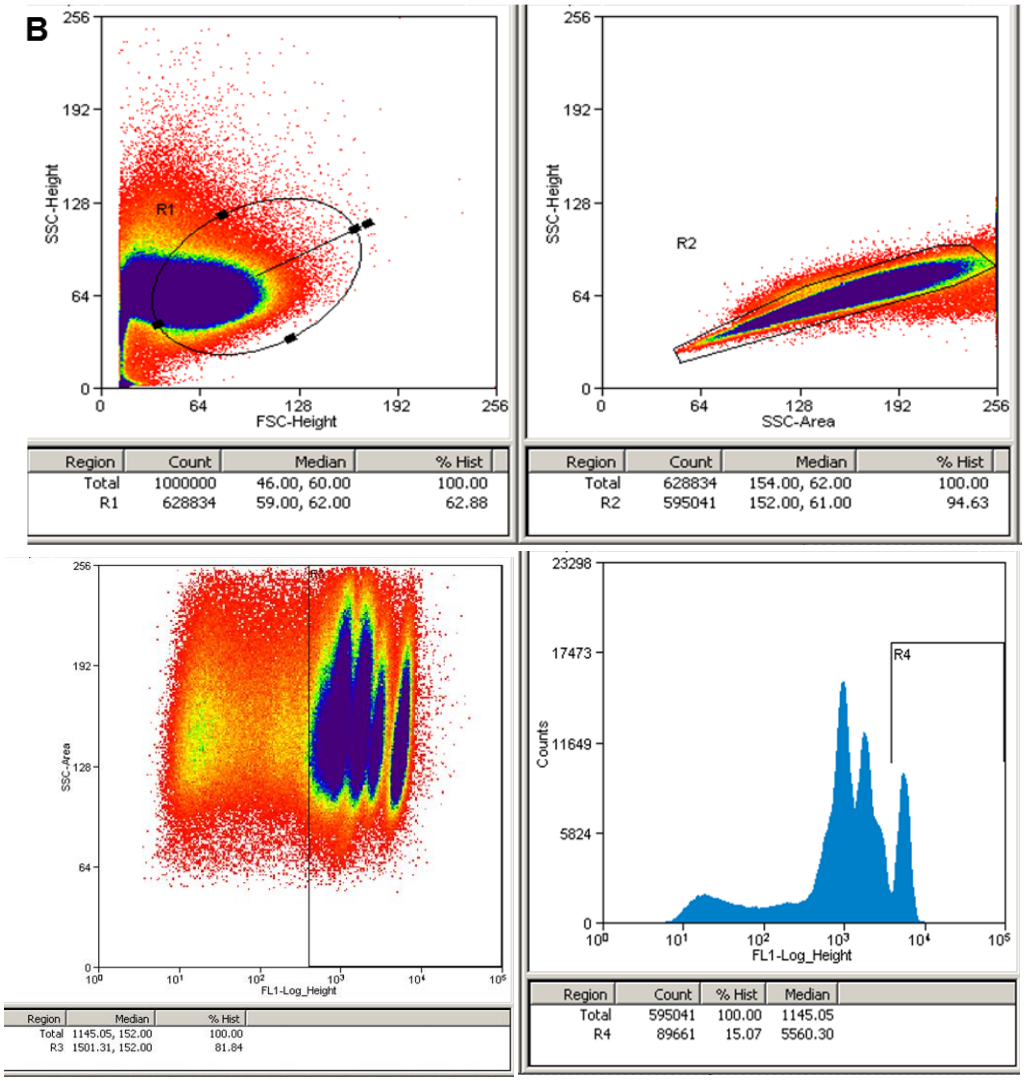


Figure 4. 16 FAC sorting of U251 CRISPR cas9 transfected cells for eGFP signal. (A) U251 eGFP positive cells (R1) were enriched (R2) and plated at one cell per well. (B) Sorting of LN18 cell lines.

4.4 Clone Screening and Genotyping

Clone screening and genotyping by PCR amplification of cassette integration ensures not only that cas9 targeting has occurred in the correct location but also that, in addition to the target locus, off target sites have not been affected. Genotyping for homozygous or

heterozygous insertions would be performed to determine if one or both alleles have been targeted.

4.4.1 Analysis of Cassette Integration

Cassette integration was analysed by PCR amplification of 5' and 3' regions within introns 7 and 9 respectively of the *WT1* genomic zinc finger locus (fig 4.17). Integration of the left homologous arm (5' region) was detected using 7aRHAR: 5'-AGTATGTTGGGAGGCGTAGG-3' and WT1RintgF: 5'-TGGGAAGACAATAGCAGGCA-3' primers whilst WT1LintgR: 5'-CTTCTCTAGGCACCGGTTCA-3' and RHRF: 5'-CAGTCTCCTCATCTGTAAGATGGG-3' primers detected right arm (3' region) integration. No PCR bands were generated at the correct sizes in control untargeted cells. Multiple clones for both U251 (C5A, C5B, D4, F6 and F8) and LN18 (C7 and C9) were positive for cassette integration which was absent in control untargeted cells which was corroborated via sequencing of 5' region (fig 4.18) and 3' region (fig 4.19) integration junctions.

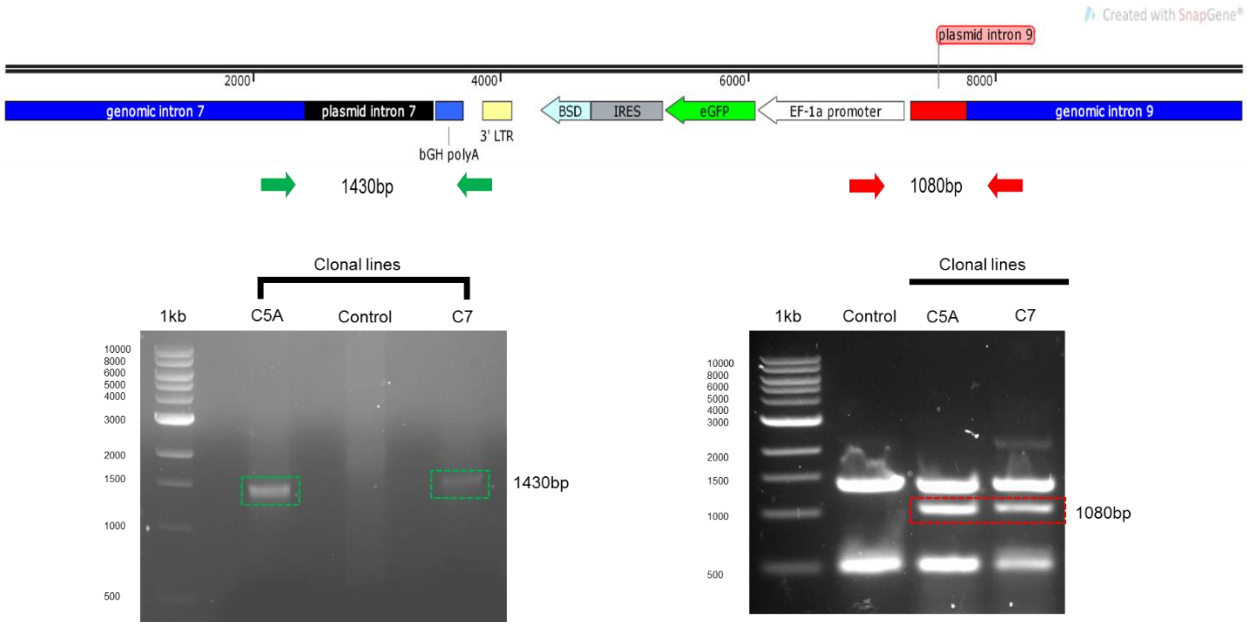


Figure 4.17 Clone screening for genomic integration of the selection cassette within the WT1 cas9 cleavage locus. Green and red arrows align to internal genomic and cassette sequences which produce amplicon products contingent on successful and site-specific genome integration. Green and red dashed boxes on agarose gels correspond to amplicons denoted by green and red arrows respectively.



Figure 4.18 Verification of 5′ cassette genomic integration. (A-B) 5′ forward and reverse reads of PCR amplified DNA from U251 clones positive for 1430bp bands representing 5′ intron 7 – bGH poly A terminator integration into WT1 cas9 target locus. Chromatograms correspond to blue highlighted sequences in the respective schematics

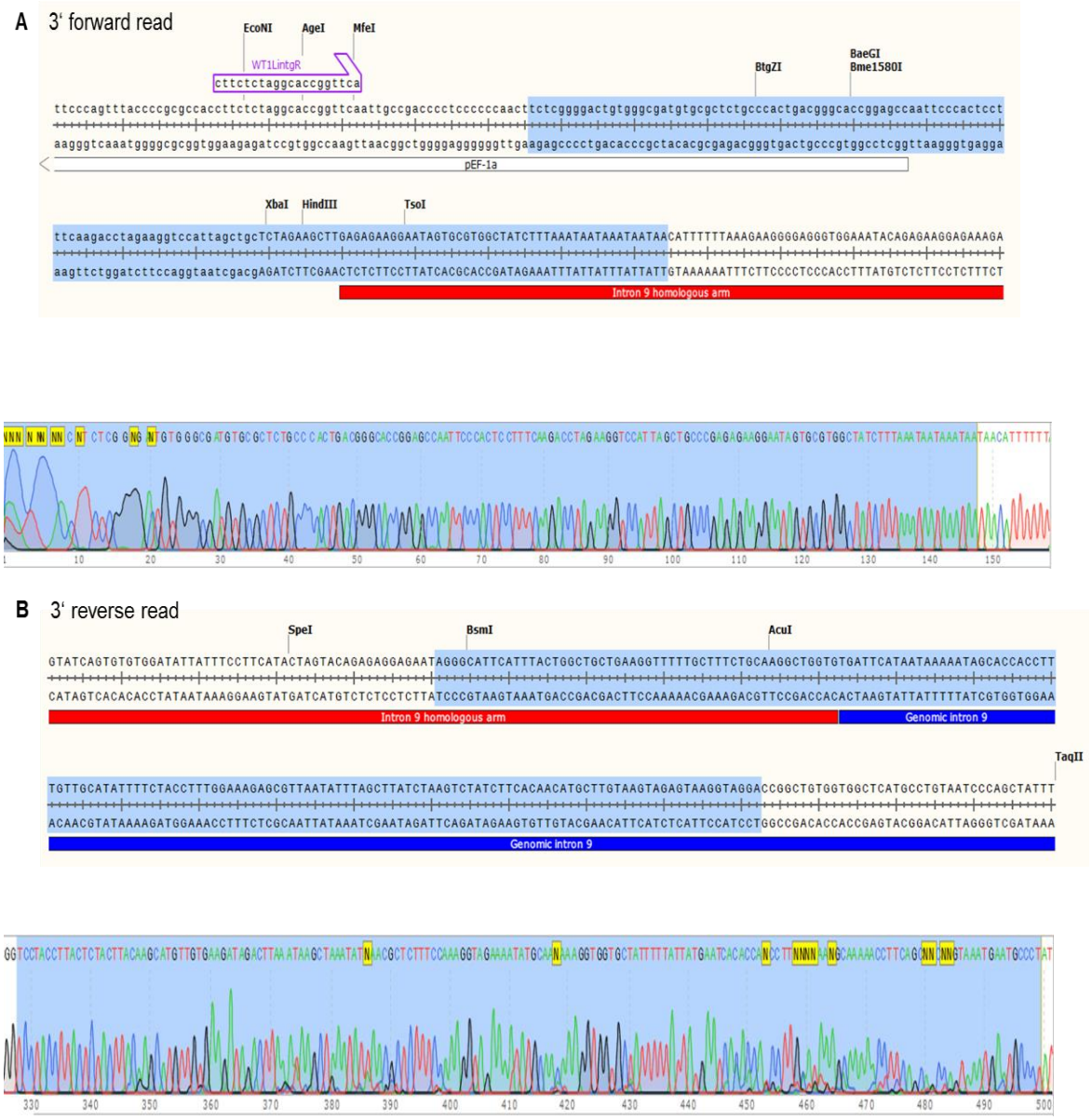
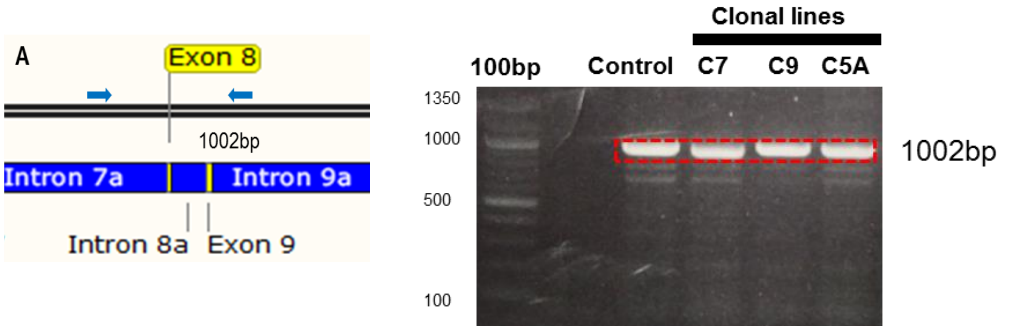


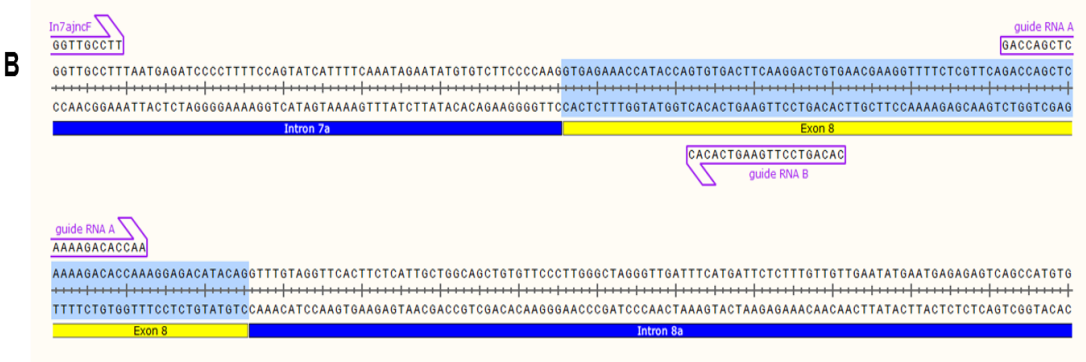
Figure 4.19 Verification of 3′ cassette genomic integration. (A-B) 3′ forward and reverse reads of PCR amplified DNA from U251 clones positive for 1080bp bands representing 3′ pEF-1a – intron 9 integration into WT1 cas9 target locus. Chromatograms correspond to blue highlighted sequences in the respective schematics.

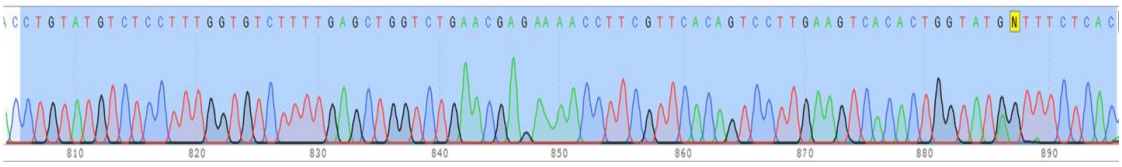
4.4.2 Analysis of WT1 Wildtype Zinc Finger Locus

To determine if presence of cassette integration of was homozygous or heterozygous, PCR amplification across exons 8-9 was performed (fig

4.20) using primers In7ajncF: 5'-CCTAGTAGGAGAGGTTGCCTT-3' and 9ajncR: 5'-GCCACGCACTATTCCTTCTCT-3'. Although PCR amplification detected presence of an amplified product commensurate with the wildtype amplicon size, sequencing revealed the presence of multiple substitutions (red boxes) and deletions (red lines) within exon 8 of mutants with the 5'- ATACAGGT-3' sequence at the exon 8-intron 9 junction representing a significant deletion near the guide 8A binding region (figs 4.20 & 4.21). Multiple deletions and substitutions were also present in intron 8 however no substitution or indels were observed in exon 9. Therefore cas9 mutant clones were heterozygous *wt1* mutants.







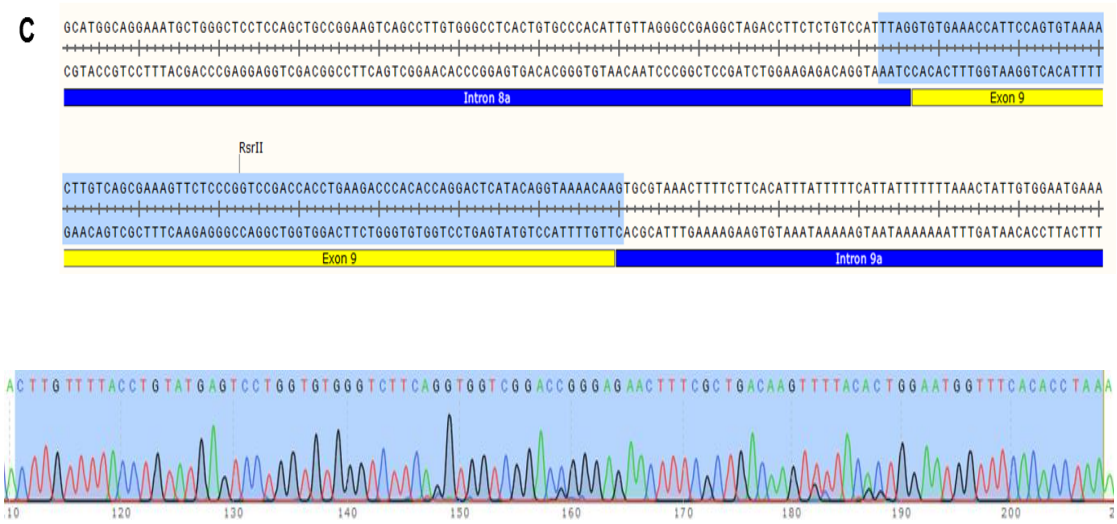


Figure 4.20 Verification of wildtype U251 WT1 exons 8 and 9 sequences.

(A) PCR screening of wildtype *WT1* exons 8 and 9 using intron 7 and 9 primers targeting the region illustrated in schematic (blue arrows). (B –C) Reverse sequencing reads verifying the wild type *WT1* exon 8 and 9 sequences in untransfected control U251 cells.

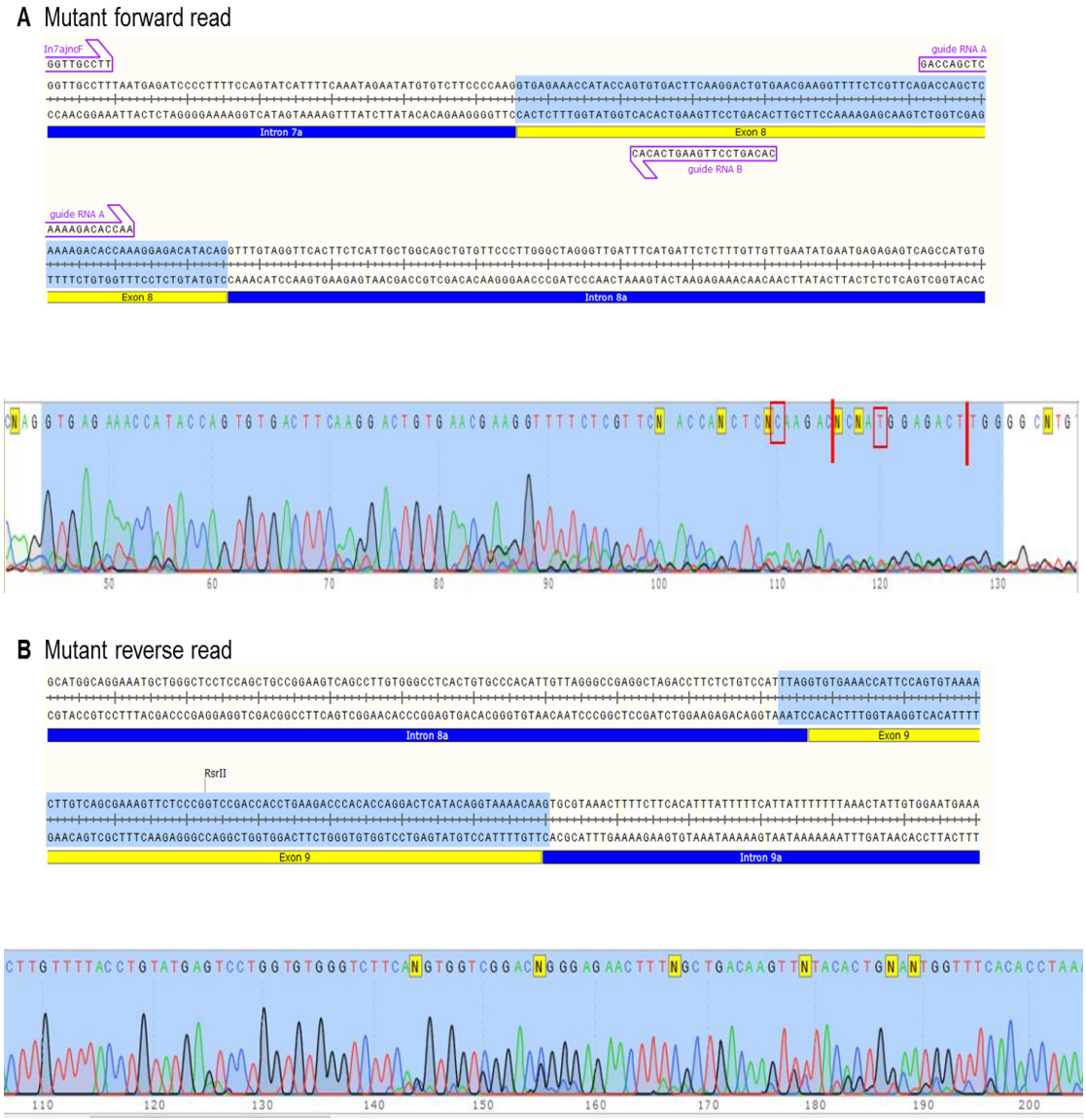


Figure 4.21 Screening of U251 C5A WT1 untargeted exon 8 and 9 loci. (A) Exon 8 forward sequencing read corresponding to blue highlighted region in schematic. Red boxes and red vertical lines denote base substitutions and deletions respectively. (B) Exon 9 reverse sequencing read. Chromatogram corresponds to blue highlighted region in the schematic.

4.4.3 CRISPR cas9 U251 and LN18 *wt1* Mutants are Predicted to Produce Truncated WT1 Proteins

To verify the genotyping of U251 and LN18 mutants, RT-PCR sequencing was performed to determine if mutations in the genomic sequence affected WT1 transcripts (fig 4.22). RT-PCR amplification of cDNA from region of WT1 exon 4 –exon 10 was performed using primers WT1cDNAex4F: 5'-TCAGATGAACTTAGGAGCCAC-3' and WT1ex10R: 5'-GCCAGCTGGAGTTTGGTCA-3'. Very faint bands were observed in all mutants compared to the untargeted control. Sequencing revealed perfect mapping of sequence traces to the wildtype reference in control samples however mutants showed multiple indels and substitutions throughout the transcript but the regions flanking indels were *WT1* specific.

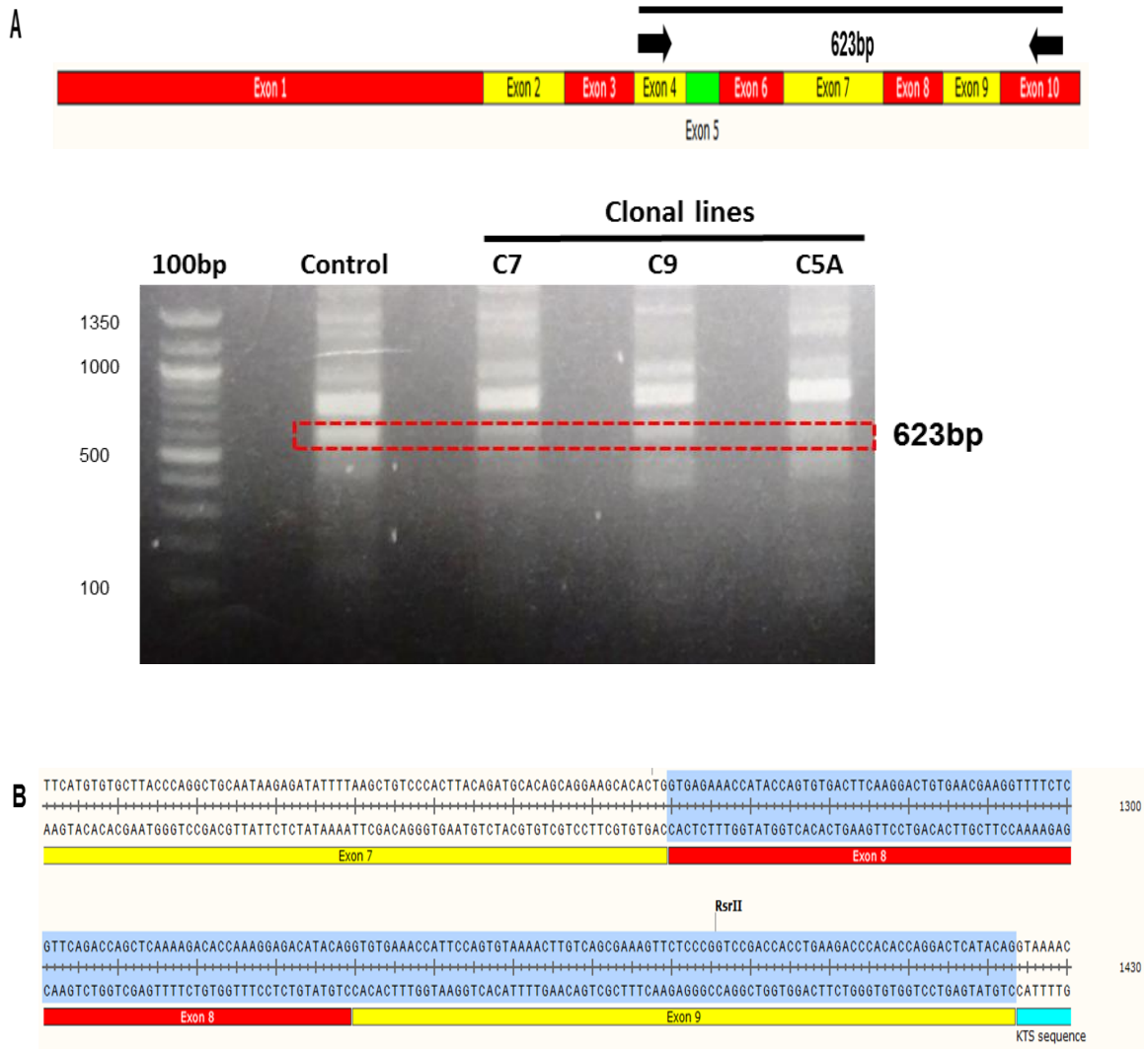


Figure 4.22 Verification of *WT1* exons 8 – 9 mutation at the transcription level. (A) PCR amplification of WT1 cDNA transcripts from exon 4 – 10 as illustrated by forward and reverse primers (black arrows). The 623bp band present in control U251 was absent or diminished in clonal lines (red dashed

box). (B) Sequencing verification of U251 *WT1* amplicons at exons 8-9 versus C5A mutants. Blue highlighted sequence in schematic corresponds to blue chromatogram sequence. Multiple substitutions (red boxes), deletions (red vertical line) and insertions (black box) were observed in the C5A mutant but not controls.

Translation of the *wt1* mutant transcripts did not match any identifiable protein sequence. Mutations present with exon 8 based on sequence traces were mapped and translated as a putative WT1 cas9 mutant sequence. Expasy translation predicted indels in exon 8 result in a truncated protein exhibiting deletion of zinc fingers 3 and 4 with a partial deletion of zinc finger 2 as amino acids 435 to 517 appear to be absent (fig 4.23A-B).

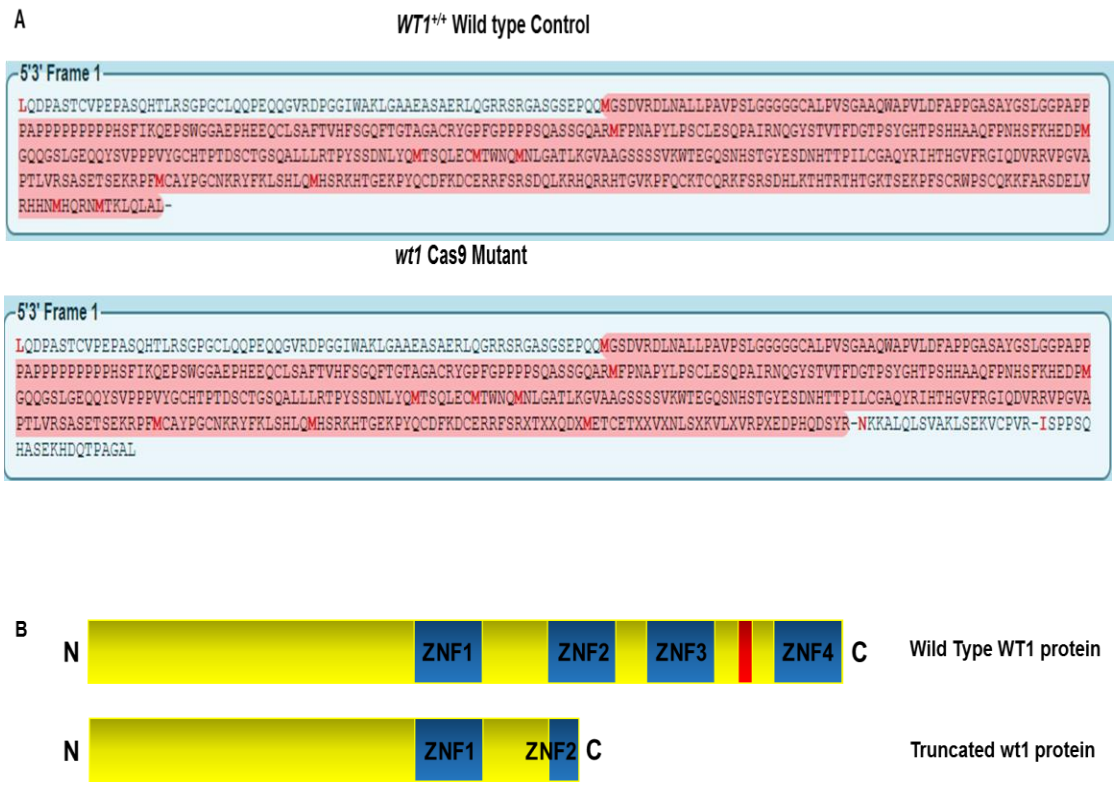


Figure 4.23 Translation of control and mutated WT1 coding sequences. (A) Translated U251 WT1 wildtype versus C5A mutant full length coding sequences. Expasy predicted premature termination occurring within exon 8 encoding zinc finger 2. (B) Schematic of the predicted WT1 protein truncation in C5A relative to U251 wild type control (red box represents KTS sequence).

Unfortunately, assessment of CRISPR-cas9 mutation on DNA binding domain zinc-fingers could not be efficiently determined owing to complications experienced during western blots (fig 4.24).

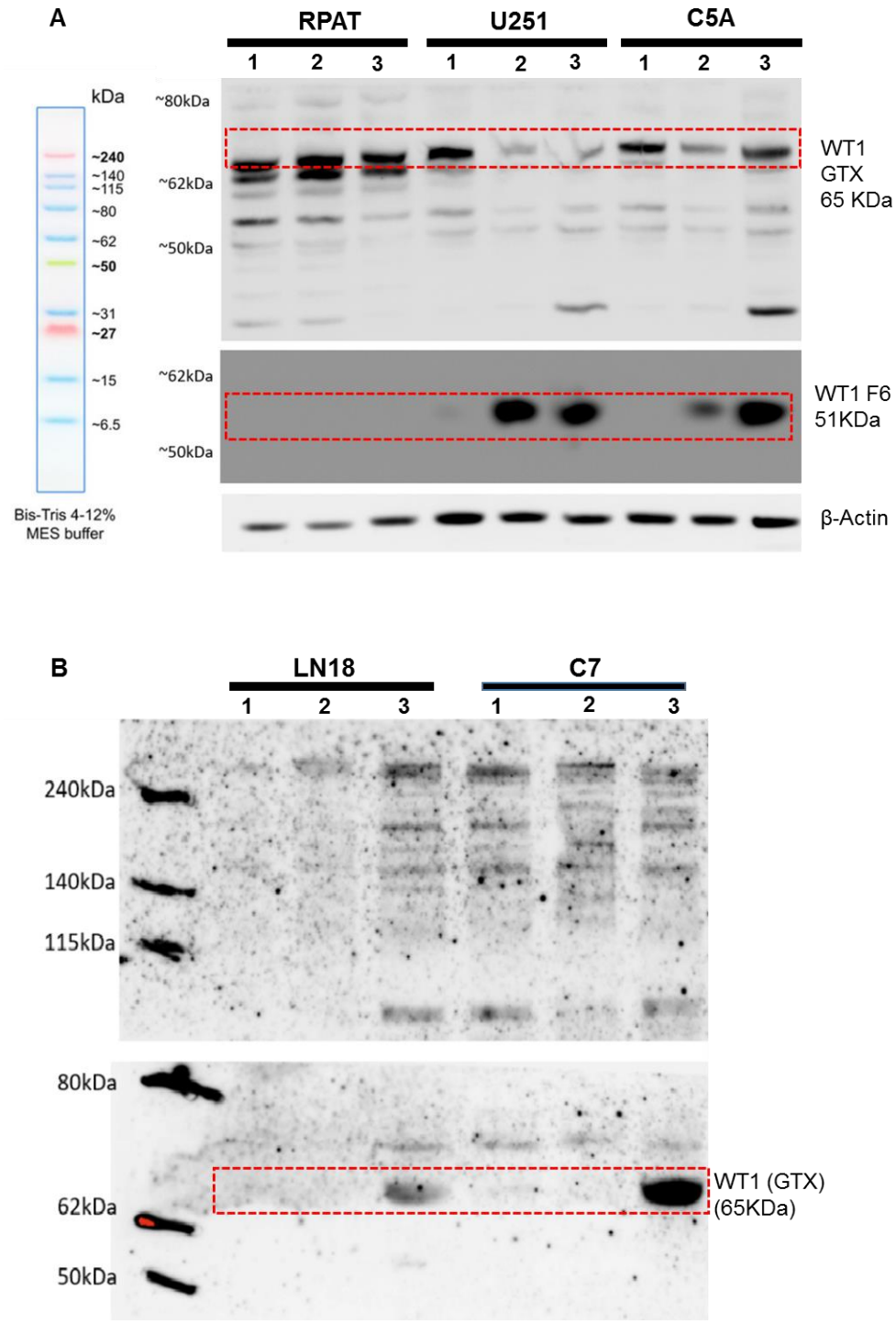


Figure 4.24 Verification of WT1 protein truncation. (A) U251 control and C5A mutant cell lines were probed for WT1 presence vs RBLPAT (RPAT) negative control using GTX C terminal and F6 N terminal WT1 primary antibodies. Beta actin served as internal housekeeping control. (B) LN18 and

C7 mutant detection of WT1 presence using GTX antibody. Red dashed boxes depict bands corresponding to their appropriate size estimations (A-B).

Inconsistency between biological repeats for all cell lines U251, C5A, LN18, C7 and RBLPAT controls and inaccurate band sizes of detected proteins precluded conclusive analysis of WT1 protein band sizes. Moreover, the appearance of multiple bands in close proximity reduced the confidence in which WT1 specific band recognition could be estimated. This effect was observed using both N terminal Santa Cruz F6 recognising 51KDa WT1 and Genetex (GTX) C terminal antibody recognising 65KDa WT1. Furthermore, the epidermoid carcinoma A-431 WCL positive control cell line expressing WT1 sent to us directly from Santa Cruz Biotech did not produce any WT1 signal using their F6 antibody. Therefore CRISPR-cas9 mediated WT1 DNA binding domain mutation in C5A and C7 GBM mutant lines was inconclusive.

4.5 Technical Discussion

This chapter entailed the design and construction of CRISPR cas9 *WT1* zinc finger knockout constructs aimed at generating DNA binding null mutants to facilitate the examination of a possible WT1-oximC interaction. Constructs were successfully synthesised, verified by sequencing and transfected into U251 and LN18 glioblastoma cell lines and clonal *WT1* zinc finger mutants were identified and isolated.

Multiple point mutations within the *WT1* zinc finger encoding region have been demonstrate to perturb DNA binding ability (Little *et al*, 1995). Missense substitution mutations such as R394W and D396G/N affect zinc finger 3 and present clinically as Denys-Drash Syndrome whilst R362X substitution in zinc finger 2 introduces a stop codon, preventing transcription and hence translation of zinc fingers 3 and 4, completely abolishes DNA binding ability of WT1 (Little *et al*, 1995). Considering biases of codon usage prevalent in malignant cells (Wu & Li, 2016; Son *et al*, 2017), introduction of point mutations into oncogene *WT1* in U251 and LN18 GBM cells may substitute canonical amino acids for unpredicted residues which either do not change DNA binding ability or the base substitution on DNA may be silent and not result in a corresponding amino acid variation. Therefore complete deletion of exons 8 and 9 encoding zinc fingers 2 and 3 respectively was considered the more favourable *WT1* targeting strategy.

Multiple obstacles were experienced during the process of generating *WT1* zinc finger null mutants. Our initial targeting strategy design implemented the D10A nickase mutant variant of the Cas9 enzyme which would utilise two *WT1* exon 8 guides (gRNA 8A and 8B with sequences 5'- GACCAGCTCAAAAGACACCAA-3' and 5'- CACAGTCCTTGAAGTCACAC-3' respectively) to ensure greater specificity of cleavage with no off-targets. Guide RNAs 8A and 8B were subcloned into two separate plasmids of the same parental backbone such that complete nickase machinery transfection would involve the

transfection of four plasmids: D10A nickase, guide 8A plasmid, guide 8B plasmid and HDR donor plasmid. U251 and LN18 cells had been transfected using a liposomal transfection strategy (lipofectamine 2000) to facilitate cytosolic entry of gene editing constructs. Whilst multiple publications detail highly efficient delivery of CRISPR machinery and successful editing of cells (Liang *et al*, 2015; Yu *et al*, 2016; Hegge *et al*, 2018 & Shen *et al*, 2018), these studies entailed incorporation of Cas9 proteins or mRNA transcripts together with one guide RNA plasmid or full U6 promoter-guide-scaffold transcript into the liposomal package delivery system i.e. a maximum of two plasmids is simultaneously transfected into cells. Contrarily, the utilisation of mutant D10A nickase cas9 for precise genome editing with minimal off-targets chosen for WT1 editing required double guide RNA cut site designation, facilitating single strand 'nicking' by the D10A cas9 mutant enzyme.

To improve the chances of homology-directed repair following effective double strand cleavage, presence of a homologous template is paramount to facilitate incorporation of an identifiable and selectable marker signalling successful WT1 target locus editing. Unfortunately, implementation of this strategy did not produced desired results.

Parameters effecting transfection efficiency of liposomal based delivery include: cell density, cargo concentration and transfection reagent concentration. These parameters were manipulated to generate the optimal transfection conditions however multiple caveats exist which

confound both liposomal based cellular delivery and the transfection optimisation process itself: 1. optimal transfection was measured by metrics including transfection efficiency and cell viability. Transfection efficiency pertained to the number of transfected cells as a fraction of the total number of cells in a population. To gauge transfection efficiency, cells were transfected with a constant green fluorescent protein (GFP) plasmid concentration with constant cell density at varying lipofectamine concentrations (fig 4.15B). The concentration exhibiting highest GFP signal with lowest cell death when examined by fluorescent microscopy was selected as optimal. However, gene editing transfections entail delivery of four different plasmids (nickase, guide A, guide B and homology donor template plasmid) all significantly larger than the positive control GFP plasmid. Desired *WT1* targeted biallelic homologous recombination was contingent on the presence of all four plasmids per cell. Consequently, by analysis of GFP signal alone successful CRISPR transfection could not be determined. 2. Lipofectamine is cytotoxic at increasing concentrations (Zhong *et al*, 2008) and thus compromises between cell viability and transfection efficiency were optimised however the assays were swayed in favour of transfection efficiency owing to the excesses of plasmid cargo impinging on optimal transfection. 3. Lipofectamine 2000 is designed for the delivery of RNA and linear DNA constructs into cells, not intact plasmids. Large plasmid DNA (>2kb) transfection induces genotoxic stress on cells which reduces cell viability and transgene expression efficiency. With the added cytotoxicity of the lipofectamine 2000 reagent alone, liposomal based delivery methods of

plasmid transfection retrospectively appeared to be unsuitable. Moreover, the conductance of multiple transfection controls with individual elements of the CRISPR cas9 machine alone and in combinations, with and without antibiotic selection revealed the presence of transfection induced stress on cells. However, no significant stress and loss in viability was observed between large single donor plasmid transfection versus CRISPR cas9 machinery plasmids without donor plasmid (data not shown). Simply put, this indicated that no significant genotoxic stress was induced by delivery of nickase cas9 components compared to the large 8.0 kb donor plasmid. This is highly irregular considering nickase creates large indels within the guide RNA target site and would cause significant mutation even if repair by the non-homologous end joining (NHEJ) DNA repair pathway.

To address the issue of cargo delivery, nucleosomal based transfection methods were employed using the Lonza Nucleofectotor II electroporation system. Optimisation of this transfection method produced significantly higher transfection efficiency and increased cell viability. Transfected cells were screened by blasticidin selection and fluorescence activated cell sorted (FACS) on the basis of GFP expression and seeded into 7 96 well plates at a density of one cell per well. Following clonal expansion, analysis of surviving GFP positive clones did not yield any *WT1* edited cells as screened by PCR verification of exogenous cassette integration into the WT1 zinc finger domain. No mono-allelic nor bi-allelic mutants were found nor NHEJ mutants. This

may suggest that, considering the requirement of two guide RNAs for efficient nickase mediated double strand cleavage, the improper functioning of one or both guide RNAs may render cleavage improbable. Hence to test this scenario, wild type CRISPR cas9 was utilised with guides A and B in separate transfections. Transfection of wildtype cas9 with separate guide 8A and 8B yielded vastly different results with no edited clones observed for U251 or LN18 cells transfected with guide B. In contrast multiple clones were identified in both cell lines transfected with guide 8A (fig 4.17). We surmised that reduced or absent binding ability of guide 8B may have significantly reduced D10A nickase mediated cleavage which is contingent on efficient binding of both guides. Insights into this limitation could have been achieved prior to transfection (thereby mitigating the labour and time intensive endeavour) by *in vitro* assessment of guide function.

Recombinant wildtype cas9 could have been incubated in buffered solution with sgRNA 8A or B and a 500bp DNA template containing exon 8 and 9. Incubation of the reaction at 37°C followed by agarose gel analysis for restriction fragment lengths could be performed to determine guide cutting efficiency (Jinek *et al*, 2012; Larson *et al*, 2013; Mali *et al*, 2013). Our initial generation of cas9 edited clones involved manual limiting dilution isolation of individual cells from a pool of GFP positive blasticidin survivor cells. After 10 days of blasticidin treatment, surviving cells were serially diluted at ratios of 1:100 and reseeded in fresh plates in an attempt to generate smaller colonies of cells, thus facilitating more

efficient segregation of stably incorporated cassette clones. The failure of this procedure may be attributable to a passive blasticidin resistance conferred from cassette harbouring cells to untransfected cells, perhaps via exosomal delivery (Lv *et al*, 2014). Indeed, FACs analysis which eventually resulted in isolation of successfully edited clones revealed the persistence of a GFP negative blasticidin surviving population (24.6% of analysed cells, fig 7.16A) which may have been obscured from microscopic detection owing to the over representation of GFP positive cells coupled with low resolution of the fluorescent microscope. Additionally, random plasmid integration exhibited by cancer cells possessing Phosphatase and tensin homologue (*PTEN*) mutations and consequently compromised non-homologous end joining mechanisms contribute to problems of distinguishing transient versus stably transfected cells (Murnane, Yezzi & Young, 1990; Santivasi & Xia, 2013; Annovazzi *et al*, 2015; Kang *et al*, 2017).

To mitigate these issues, serial passaging of cells could be performed shortly (24 hours) after recovery from transfection. Passaging of cells on a 48 hour basis in the absence of blasticidin section pressure could force dilution of non-integrated plasmid during mitosis, thus limiting the quantity of plasmid capable of randomly integrating within the genome. Antibiotic mediated selection could be performed following 2 weeks of such passaging to prevent possible excision or silencing of incorporated cassettes due to excess metabolic expenditure in steady state conditions (Pikaart, Recillas-Targa & Felsenfeld, 1998; &Mutskov & Felsenfeld,

2004; Riu *et al*, 2007). Only heterozygous CRISPR cas9 mediated *wt1* mutants were obtained even when gating for the highest GFP expressing cells which were reasoned to potentially contain two targeted *wt1* alleles each substituted by a cassette and hence exhibit the brightest signal. Increasing the concentration of donor plasmid concentration from 3ug to 10ug (by controlled titration) may have yielded homozygous mutants even though concentrations of 3mg have been effective (Chu *et al*, 2015). Although only one *WT1* allele was targeted in our *wt1* mutants, multiple indels appeared in the non-targeted allele which may have arisen consequent of NHEJ mediated via cas9 cleavage as these mutations appear around the guide 8A cleavage site and at the exon 8-intron 9 boundary, characteristic of wildtype cas9 activity (Canver *et al*, 2014; Hsu *et al*, 2014; Inui *et al*, 2014).

Owing to the initial design strategy of utilising the D10A 'nickase' variant of Cas9, no guide RNA off-targets were predicted for WT1 guide 8A and 8B combination. Considering the duration of time troubleshooting D10A nickase transfections and subsequent clone screening, prioritisation of time and resources was skewed towards clone genotyping. Consequently, separate off-targets for WT1 guides 8A and 8B were investigated and primers design however execution off-target screening (due to time constraints) was not performed. As positive clones were only obtained using guide 8A with wildtype cas9, top 3 off-targets for guide 8A predicted at co-ordinates chr5:91,158,448-91,158,470, chrX:25,082,860-25,082,882 and chr4:128,628,327-128,628,349 with

scores of 1.8 (2mismatches), 1.5 (3 mismatches) and 1.4 (4 mismatches) respectively would need to be screened as a priority before subsequent experiments with this dataset can be performed.

The lack of conclusive western blot results following our probing for WT1 protein size in U251 vs C5A and LN18 vs C7 cell lines may be due to multiple factors such as non-specificity of the WT1 antibodies. The GTX C terminal specific antibody, although utilised in immunofluorescence and ChIP, may not have been suitable for western blot. The Santa Cruz F6 antibody, while detecting bands roughly approximating 51KDa in U251 and C5A, did not produce reproducible results between biological replicates and did not produce detectable bands in the positive control. Greater reliability can be attributable to the F6 antibody compared to the GTX as RBLPAT cells which do not express *WT1* at detectable transcriptional levels, display a large increase in WT1 expression at 24 hours post induction during hepatic lineage differentiation, detectable with the F6 antibody (data not shown). Therefore, to mitigate these complications, additional specialised WT1 specific antibodies could be implemented to detect WT1 expression in control and mutant cell line and protein mass spectrometry could be performed to identify mutant WT1 species in these cells.

Characterisation *wt1* mutant GBM Cell Lines

5.1 Introduction

Since the advent of the first generation DNA sequencing technologies by Frederick Sanger which employed target-specific primer PCR-dependent amplification of sequences of interest via incorporation of radio-labelled, and later on, fluorescence labelled dideoxyribonuclotide triphosphates (ddNTPs), substantial progress in depth and scale of genome sequencing has been achieved (Smith *et al*, 1985; Sanger, 1988; Mitra *et al*, 2003; Zhou *et al*, 2010). Multiple Next Generation Sequencing technologies are now available which utilise techniques such as bridge or emulsion PCR or *in situ* polonies in which DNA molecules affixed to solid surfaces such as flow cells facilitate massively paralleled sequencing enabling significantly increased depth and scale of genome coverage as oppose to the limitations attributable to gel-slab and capillary PCR of the 1st sequencing generation (Smith *et al*, 1985; Sanger, 1988; Mitra *et al*, 2003; Zhou *et al*, 2010). Although multiple

technologies employing variants in the amplification and base calling methods of sequencing exist on different platforms offered by competing technologies such as but not exclusive to Illumina, Oxford Nanopore, Ion Torrent and ABI SOLiD, (Zhou *et al*, 2010; Mardis, 2013; Levy & Myers, 2016), only Illumina Sequence-By-Synthesis (SBS) methods will be described here since this technology was employed in this thesis.

In contrast to 1st generation sequencing technologies such as Sanger sequencing which required target-specific primers to isolate sequences of interest, NGS utilises generic forward and reverse adaptor primers which can be ligated to genomic fragment terminals created via shearing whole genomic DNA through sonication or enzymatic digestion (Zhou *et al*, 2010; Mardis, 2013; Levy & Myers, 2016). This DNA library preparation process possessed multiple quality control stages in which fragmented genomic DNA ends are repaired to blunt end any 5' and 3' overhangs, adaptor ligate, purify, PCR amplify and verify library size distribution (Zhou *et al*, 2010; Mardis, 2013; Levy & Myers, 2016). DNA libraries possessing forward and reverse adaptor sequences are denatured and hybridise *in situ* to complementary adaptor primers covalently affixed to a solid phase within an Illumina microfluidic channel flow cell (Zhou *et al*, 2010; Mardis, 2013; Levy & Myers, 2016). Conduction of bridge PCR using Phusion DNA polymerase over approximately 35 cycles generates millions clonal clusters of DNA fragments stemming from a single template molecule. Chain elongation is terminated by incorporation of ddNTPS. SBS may commence as

fluorescently labelled ddNTPS are incorporated into clonal clustered fragments according to Watson and Crick base-pair rules simultaneously and following washing, are detected and read at each cluster positioned at unique coordinates within the flow cell (Zhou *et al*, 2010; Mardis, 2013; Levy & Myers, 2016). The massively paralleled SBS approach generates millions of short reads which may vary dependent on platform specifications but is ideal for whole transcriptome analysis by RNA-Seq or Genomic oximC enrichment profiling and transcription factor DNA occupancy interrogation (Zhou *et al*, 2010; Mardis, 2013; Levy & Myers, 2016).

Hence, we employed Illumina SBS methodologies to understand genomic 5caC and WT1 enrichment in LN18 and U251 GBM cell lines and the effect of their presence (or absence) on GBM transcriptional profiles.

5.2 Analysis of U251 and LN18 *wt1* Mutant DNA Demethylation Profiles

The results generated so far have correlated *WT1* expression with the presence of 5caC in GBM cell lines relative to healthy foetal neural stem cells and human induced pluripotent stem cells (hiPSCs) and fibroblast control cell lines. Moreover, the timing of *WT1* transcriptional up-regulation coincides with transient periods of 5caC enrichment during

both neural and glial differentiation in LN18 and U251 cell lines. Furthermore, *WT1* siRNA mediated silencing and plasmid facilitated over-expression elicit significant 5caC signal magnitude reduction and elevation respectively, thereby implicating a non-trivial level of involvement of this transcription factor in DNA demethylation. Therefore, utilisation of the U251 and LN18 CRISPR cas9 generated *wt1* zinc finger mutants C5A and C7 respectively, enabled the investigation of WT1 function in DNA demethylation dynamics and what possible contribution (if any) this interplay may influence GBM pathology.

5.21 CRISPR cas9 *wt1* Mutant GBM Cell Lines Display Reduced WT1 Gene and Protein Expression

To determine the effect of CRISPR cas9 mediated *WT1* zinc finger mutation on *WT1* expression levels, Immunofluorescence staining and qPCR analysis of WT1 protein and gene expression levels respectively was performed on C5A and C7 mutants (fig 5.1 A-C). Clusters of brightly staining *puncta* were observed surrounding nuclei periphery (white arrows) in both mutants but not control cell lines (fig 5.1A). Quantification of WT1 levels via immunofluorescence revealed highly significantly ($p < 0.0001$) reduced WT1 protein levels in C5A and C7 mutants relative to their U251 and LN18 cell line controls respectively. WT1 signal intensities in C5A and C7 mutants were detected at 60 and 70% lower magnitudes respectively compared to their controls (fig 5.1B).

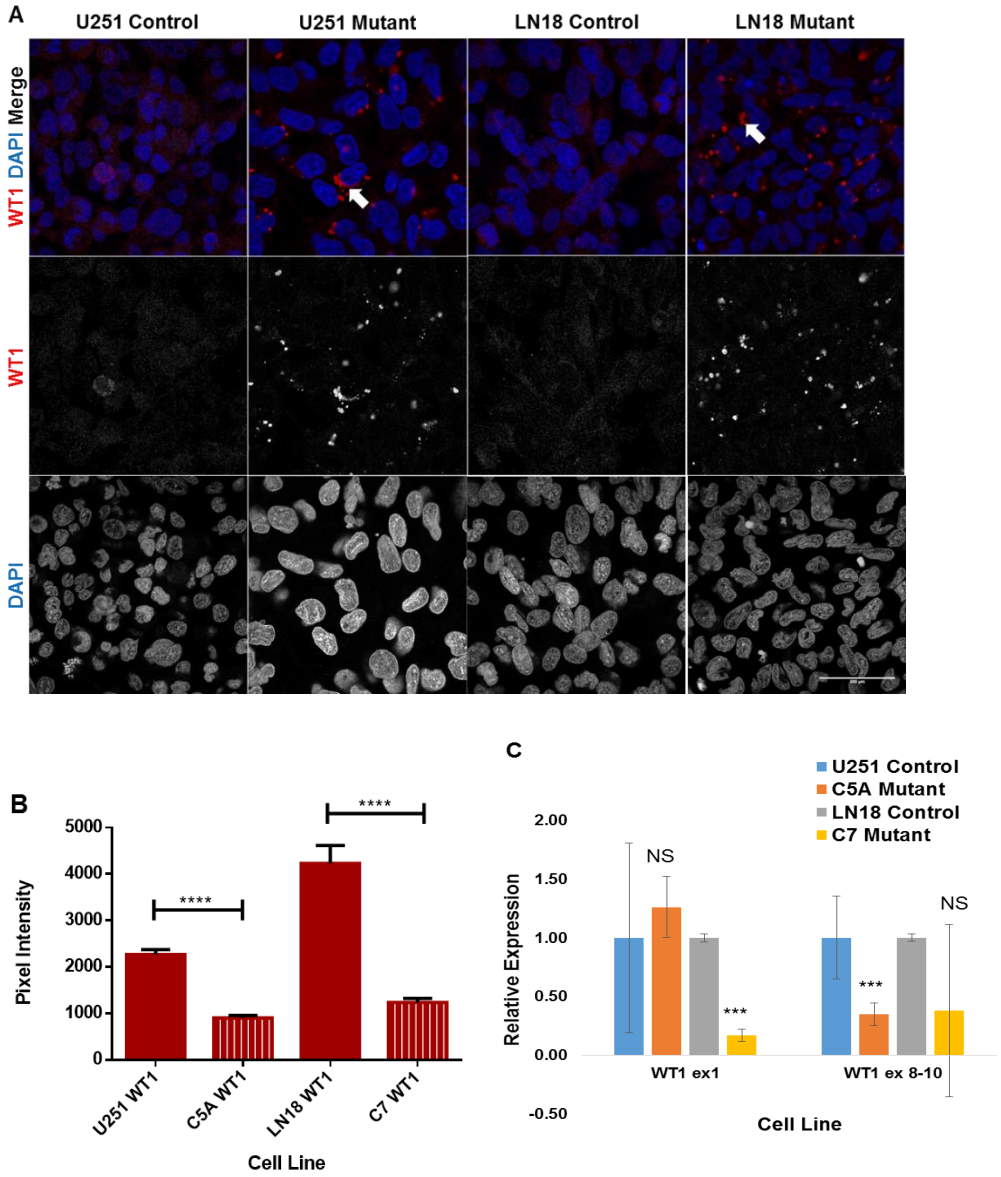


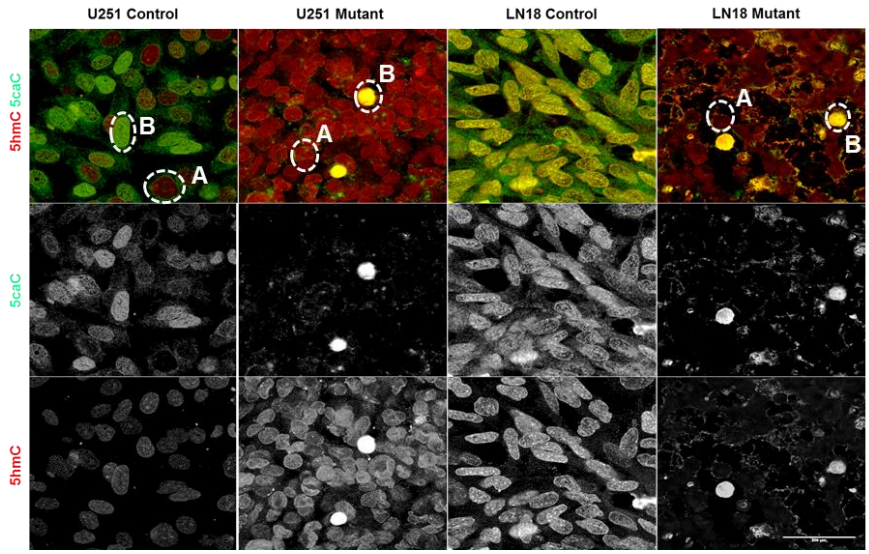
Figure 5.1 Assessment of WT1 expression in C5A and C7 mutants. (A) Immunostaining of WT1 (red) in C5A and C7 wt1 mutants vs their controls. Staining and images captured at identical settings, scale bar represents 200µm. (B) Quantification of WT1 immunostaining signal intensity. Mean ± SD, unpaired two tailed Student's t-test, **** $p < 0.0001$. (C) QPCR analysis of *WT1* transcript levels in C5A and C7 mutants vs their respective controls. Mean ± SD, unpaired two tailed Mann-Whitney test, NS $p \geq 0.05$; *** $p < 0.001$, n =3.

Analysis of *WT1* gene expression depicted similar trend of *WT1* transcriptional levels to the WT1 immunofluorescence signal intensities (fig 5.1C). Considering the CRISPR cas9 mutagenesis of WT1 targeted zinc fingers 2 and 3 encoded by exons 8 and 9 respectively, interrogation of exons 8-10 was performed as well as exon 1 to determine if expression of upstream regions of the targeting site was occurring. No significant difference in exon1 expression was observed between U251 control and C5A mutants ($p = 0.206$) however a 65% reduction in exon 8-10 zinc finger domain expression in C5A was detected ($p = 0.0071$). In contrast, 83% exon 1 expression reduction occurred in the C7 mutant relative to LN18 ($p = 0.0012$) whilst a 62% reduction in exons 8-10 expression was observed in C7 relative to control LN18 ($p = 0.285$).

5.22 GBM Cell Line *wt1* Mutants Display Significantly Altered 5hmC and 5caC Distributions Correlating with Elevated *TET* Gene Expression

Immunostaining and qPCR analysis was performed to evaluate the effect of *WT1* zinc finger mutation on the magnitude and distribution of 5hmC and 5caC modifications as an indicator of DNA demethylation activity in the cell lines (fig 5.2A-E). Immunofluorescence interrogation of samples with Alexa 488 and 633nm fluorophore conjugated secondary antibodies delineated the 5hmC and 5caC signal heterogeneity (fig 5.2A). Cells with higher 5hmC signal than 5caC (overall red) and those with higher 5caC than 5hmC (overall green) were designated as 'A' and 'B' cells (white

dashed circles) respectively. Type 'A' cells exhibited greater 5hmC signal than type 'B' cells within the U251 control cell line whilst 5caC signal intensities in 'B' cells was greater than that of type A 5caC ($p < 0.0001$). Type 'A' cells 5hmC was greater than type 'B' cell 5caC indicating a strong 5hmC presence and not simply lower 5caC levels in these cells than type 'B' ones ($p < 0.0001$). The stark heterogeneity in 5hmC and 5caC signal of U251 control cells was not observed in LN18 controls, therefore no type 'A' or 'B' cells were demarcated. Remarkably, both C5A and C7 *wt1* mutant lines displayed significant increase in 5hmC signal intensity compared to respective control lines ($p < 0.0001$). Again, type 'A' and 'B' cells were designated in C5A and C7 mutants with C5A type 'A' cells exhibiting 1.5 times and 2.2 times greater 5hmC signal intensities than



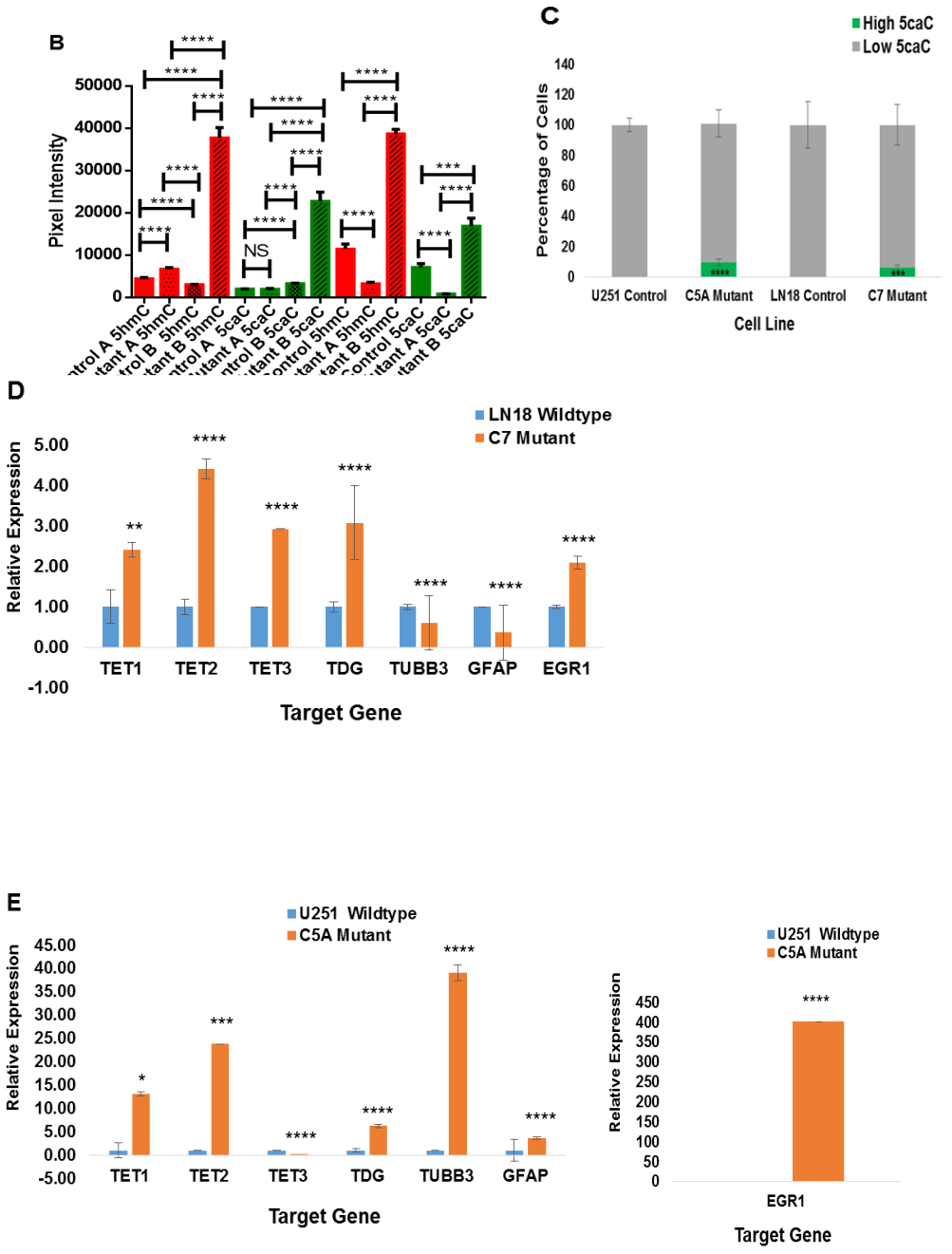


Figure 5. 2 Characterisation of DNA demethylation activity in wt1 mutants.

(A) 5hmC and 5caC immunostaining in C5A and C7 mutants relative to their U251 and LN18 respective controls. White dashed circles demarcate cells displaying disparate staining intensities within each population which are represented in the histogram (B) Immunostaining and images captured at identical settings, scale bar represents 200µm. (B) Quantification of 5hmC and

5caC signal intensities. Encircled cells 'A' and 'B' for each cell line are represented and quantified relative to the appropriate control. Mean ± SD, unpaired, two tailed Mann-Whitney tests. (C) Proportion of high 5caC signal possessing cells in C5A and C7 mutants relative to their controls. Mean ± SD, unpaired two tailed Z-test of proportions. (D-E) QPCR transcriptional analysis of DNA demethylation machinery, lineage specific genes and EGR1 expression in C5A and LN18 mutants relative to their controls. Mean ± SD, unpaired, two tailed Kruskal-Wallis tests, NS $p \geq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$, $n = 3$.

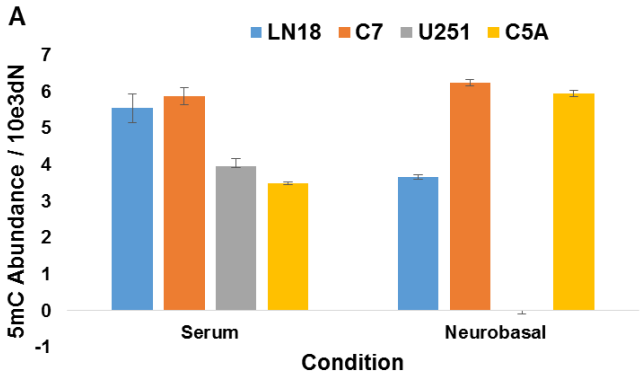
U251 type 'A' and 'B' cells respectively ($p < 0.0001$). No significant difference was observed between C5A type A 5caC signal levels compared to U251 controls ($p = 0.329$) however C5A type A 5caC levels were 1.7 fold reduced relative to U251 control type B cells ($p < 0.0001$). C5A type 'B' cells possessed very highly enriched 5hmC and 5caC signal intensities 5.7 fold and 12.0 fold higher respectively than their own type B cells ($p < 0.0001$) and very highly significantly greater than U251 type A and B cell 5hmC and 5caC signal intensities ($p < 0.0001$). Similar 5hmC and 5caC staining distributions were observed in C7 mutants however type A C7 cells exhibited 3.6 times lower 5hmC than control LN18 ($p < 0.0001$) and 9.25 times lower 5caC (and at negligible levels) than LN18 control cells ($p < 0.0001$). C7 type B cells displayed 12.2 fold and 3.4 fold greater 5hmC than C7 type A and LN18 control cells respectively whilst C7 type B 5caC levels were detected at 22 times and 2.4 times greater magnitudes than C7 type A and LN18 control 5caC levels respectively ($p < 0.0001$). Type B cells possessing 5hmC and

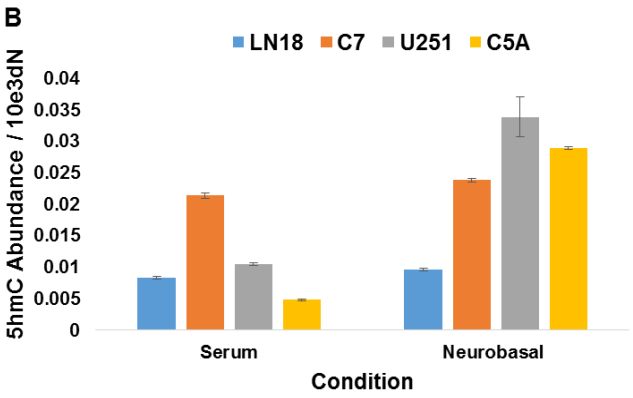
5caC magnitudes exemplified by C5A and C7 cell lines were not observed in control U251 and LN18 cell lines. The mutant type B cells only represented 9.4% ($z = 4.529$, $p < 0.0001$) and 6% ($z = 3.558$, $p = 0.00038$) of the population (fig 5.2C). Significantly Increased gene expression of *TET1*, *TET2*, *TDG* and *EGR1* was observed in both C5A and C7 mutants relative to their respective controls (fig 5.2 D-E). Of the *TET* protein genes, *TET2* displayed the greatest expression at 23.78 fold ($p = 0.002$) and 4.41 ($p < 0.0001$) fold levels higher than in U251 and LN18 controls respectively. Contrarily, *TET3* which was down-regulated by four orders of magnitude in C5A ($p < 0.0001$), was up-regulated by 2.93 fold in C7 mutant cells ($p < 0.0001$). Notably, *TUBB3* and *EGR1* were very significantly up-regulated by 39 and 401-fold respectively in C5A mutants ($p < 0.0001$). B cells in C5A and C7 mutants may actually be dead cells which could be verified by using viability stains such as Crystal Violet.

5.23 5mC and oximC magnitudes remain consistent in *wt1* mutants relative to their controls

Mass spectrometric analysis of C5A and C7 cell lines relative to their controls revealed similar absolute 5mC, 5hmC and 5caC magnitudes in wt1 mutants relative to U251 and LN18 (fig 5.3). Control and mutant cells displayed similar levels of 5mC within either serum but larger discrepancies were observed under neurobasal conditions e.g. LN18 vs C7 in serum (5.54 vs 5.87) compared to LN18 vs C7 in neurobasal (3.66 vs 6.24). Comparison of LN18 and C7 cell lines demonstrated an

increase in 5mC, 5hmC and 5caC levels in C7 mutants which was in each case more disparate than differences in magnitude of 5mC, 5hmC or 5caC observed between U251 and C5A mutants. Indeed, C5A mutants possessed lower serum levels of 5mC and 5hmC relative to U251 and similar serum levels of 5caC. Neurobasal 5hmC levels were also decreased however differential 5mC and 5caC levels under these conditions could not be established owing to no detectable levels produced for those samples.





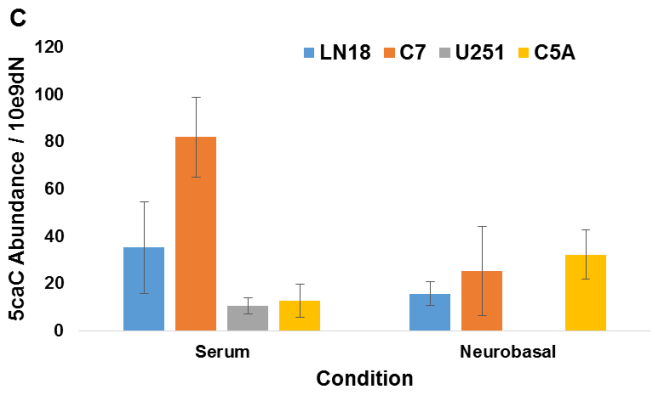


Figure 5.3 Effect of wt1 mutagenesis on 5mC and oximC magnitudes. (A-C) Mass spectrometric quantification of 5mC, 5hmC and 5caC levels in *wt1* mutants vs their controls in serum and neurobasal conditions. Mean ± SD, n = 1.

5.24 Common GBM Pathology Genes are Up-regulated in *wt1* Mutants

Analysis of commonly mutated and disease-relevant genes *EGFR*, *RB1*, *NF1*, *TP53*, *CDKN2A* and *PDGFRA* revealed elevated expression of all targets in both C5A and C7 mutants (fig 5.4A-B). Targets *RB1*, *NF1*, *TP53* and *CDKN2A* were up-regulated by 30.9 ($p = 0.0277$), 7.6 ($p = 0.0284$), 10.5 ($p = 0.0121$) and 6.1-fold ($p = 0.128$) respectively while *PDGFRA* was very highly up-regulated by $1.73x10^7$ –fold ($p = 0.0039$) in C5A mutants relative to U251 control cells (fig 5.3A).

No significant difference in gene expression was observed for *EGFR* between C5A and U251 ($p = 0.100$). Similar trends were observed in C7 mutants however *CDKN2A*, *TP53* and *PDGFRA* were the highest up-regulated at 9.4 ($p < 0.0001$), 19.8 ($p < 0.0001$) and $1.5x10^5$-fold ($p <$

0.0001) expression relative to LN18. In contrast to C5A, NF1 was not significantly up-regulated in C7 cells (p = 0.250).

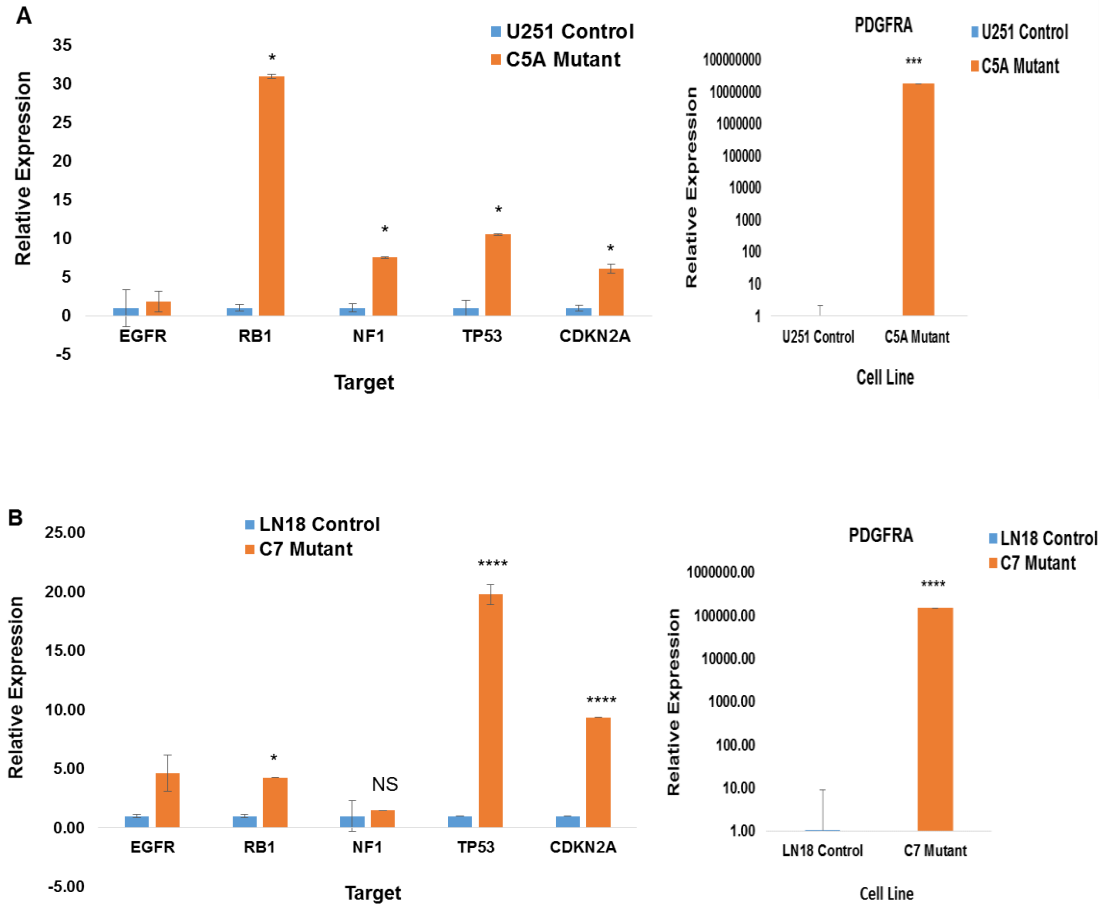


Figure 5.4 Effect of wt1 mutagenesis on GBM pathway gene transcription. (A-B) QPCR transcriptional analysis of C5A and C7 wt1 mutants relative to their U251 and LN18 respective controls. Mean ± SD, unpaired two tailed Kruskal-Wallis test, NS $p \geq 0.05$; * $p < 0.05$; *** $p < 0.001$; **** $p < 0.0001$, n = 3.

5.3 Next Generation Sequencing Characterisation of 5caC Genomic Distribution and Transcriptome Regulation

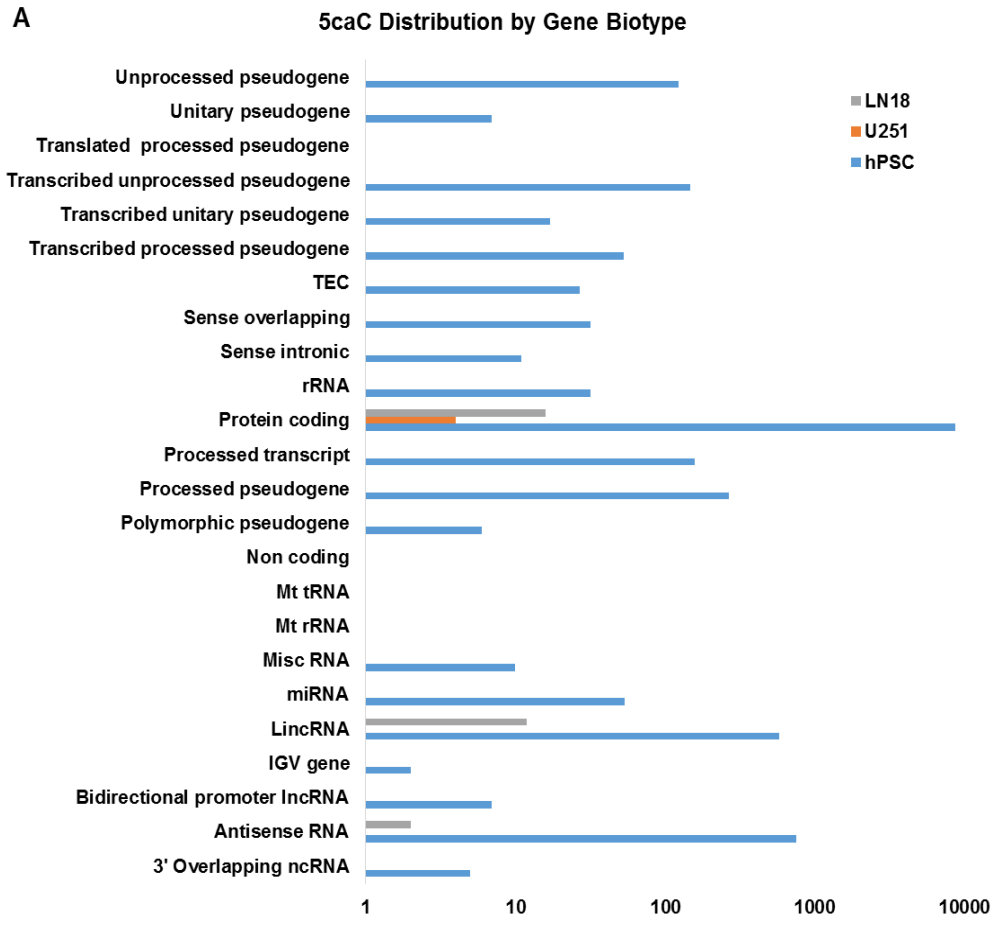
Methylated DIP for 5caC was coupled with Next Generation Sequencing methods (5caC DIP-Seq) to investigate the distribution and semi-quantitatively, the magnitude of 5caC genomic presence and, coupled with RNA Sequencing methods, assess the influence of 5caC presence on transcription. 5caC DIP-Seq was performed on U251 and LN18 GBM cell lines and RBLPAT hiPSC lines as 5caC controls.

5.31 5caC Enrichment in GBM Cell Lines Occurs at Transcriptional Regulatory Regions, within Gene Bodies and on Repetitive Sequences

Analysis of 5caC-DIP-Seq reads demonstrated that 5caC enrichment in GBM cell lines U251 and LN18 was heavily underrepresented throughout genomic features compared to RBLPAT hiPSCs (figs 5.5A-B & fig 5.6A-C). Within gene biotypes, greatest 5caC enrichment was observed in protein –coding genes for LN18, U251 and RBLPAT however 5caC peaks were observed in Linc RNA genes in LN18 but not U251 (fig 5.5A). RBLPAT cells displayed high 5caC enrichment at multiple gene biotypes including but not exclusive to protein coding, unprocessed pseudogenes, Linc RNAs and antisense RNA. Protein coding genes were the only biotype where 5caC peak were detected in U251 whereas peaks were

detected in protein coding and antisense RNA in LN18. No 5caC peaks were detected at mitochondrial transfer RNA (tRNA) or ribosomal RNA (rRNA) or non-coding gene biotypes. Within gene features, 5caC peaks were observed throughout features such as transcripts, 5' and 3' untranslated regions (5' and 3'UTRs), introns, exons, intragenic sequences, start and stop codons, coding sequences and 3kb upstream and downstream regions (fig 5.5B). U251 and LN18 5caC peaks were predominately detected within intragenic sequences of gene features however peaks were also reported in transcripts, introns, exons and at 3kb upstream and downstream. 5caC peaks were also located on regulatory sequences (fig 5.6A). U251 and LN18 5caC enrichment occurred at CCCTC (CTCF) binding sites, open chromatin and transcription factor (TF) binding sites at similar frequencies to each other but fewer than RBLPAT peaks. Contrary to RBLPAT, no U251 or LN18 5caC peaks were discovered at promoters, promoter flanking regions or enhancers. Interestingly, the highest frequency of 5caC peaks of any genomic feature were detected at centromeric satellite repeat sequences in U251 whereas highest 5caC levels were present on intragenic sequences within gene bodies for LN18 (fig 5.6B). Amongst repetitive sequences, U251 5caC peaks were also observed at SINE and *Alu* elements, telomeric repeats, simple repeats and general satellite sequences. LN18 5caC peaks were also observed at endogenous retroviral (ERV) long terminal repeats (LTRs) such as LTR ERVL, LTR ERV1 and mammalian apparent endogenous retroviral long terminal repeat (LTR ERVL-MaLTR). Visualisation of consensus peaks using the

Integrative Genomics Viewer (IGV) illustrated the depletion of 5caC peaks in U251 and LN18 relative to the abundance observed in RBLPAT cells (fig 5.6C). Interestingly, while unique peaks existed for U251 and LN18, the majority of peaks aligned to the same genomic location.



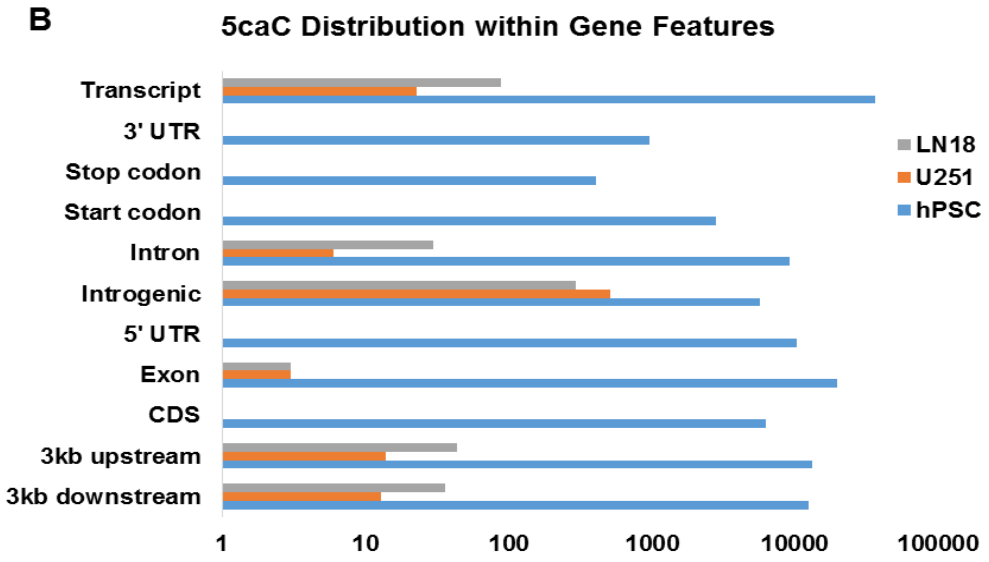
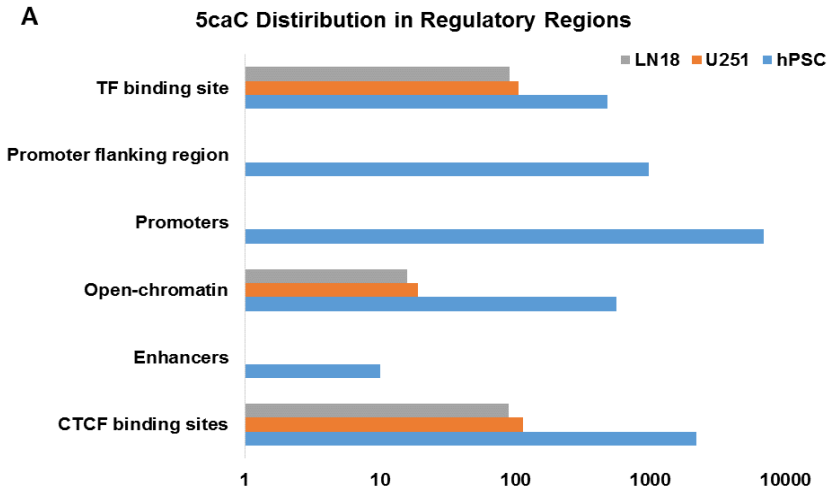


Figure 5.5 Profiles of 5caC genomic distribution between cell pluripotent stem cells and GBM cell lines. (A) 5caC distribution across gene biotypes (B) 5caC distribution within gene features.



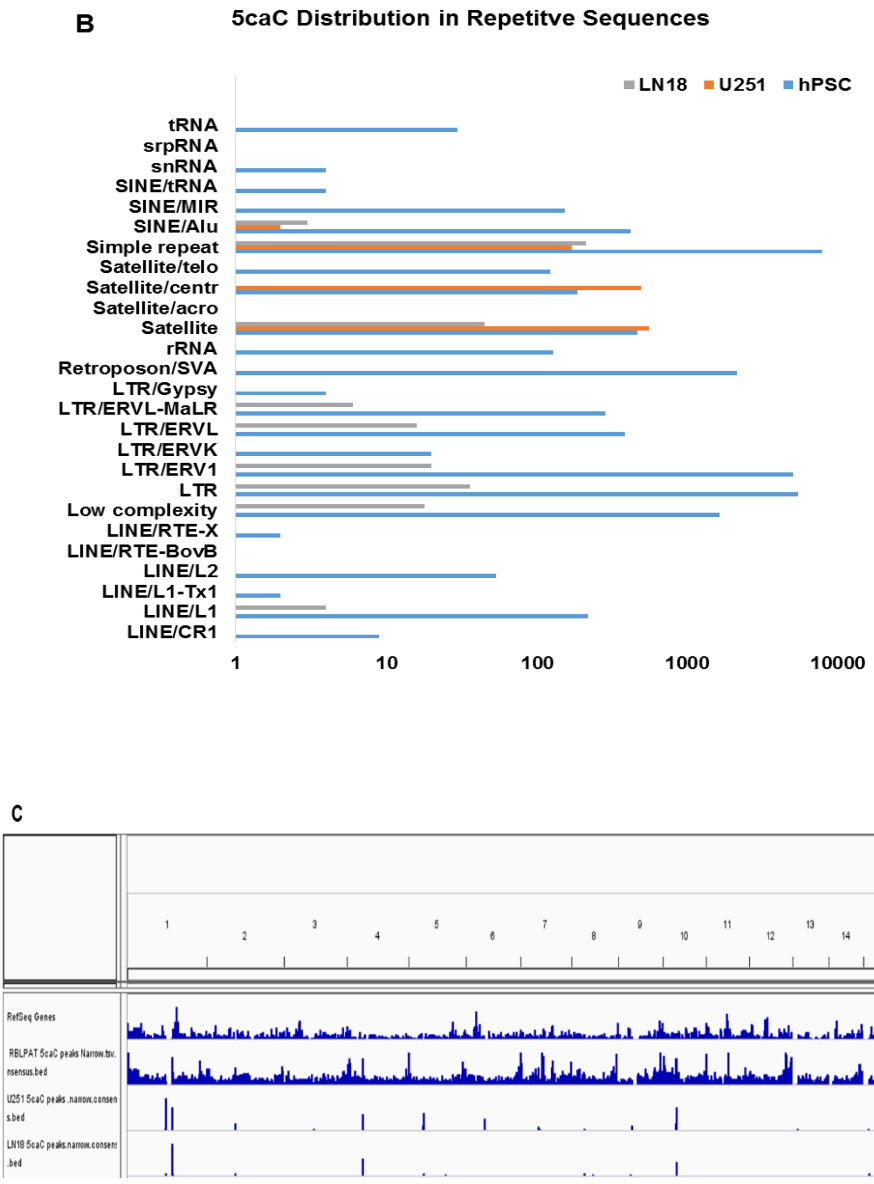


Figure 5. 6 Profiling 5caC genomic distribution in pluripotent and GBM cell lines in gene regulatory and repetitive sequences. (A) 5caC distribution in regulatory features (B) distribution across multiple repeat sequence biotypes (C) IGV genome browser window depicting 5caC consensus peak depletion in U251 and LN18 relative to RBLPAT cells.

5.32 5caC Genomic Peaks in U251 and LN18 Predominately Align to Centromeric Repeats

Closer inspection of U251, LN18 and RBLPAT consensus narrow peaks (fig 5.7A-B) highlights the congregation of consensus peak clusters around the chromosome 1 centromere (red dashed box). At a resolution of 7,800kb (fig 5.7B), U251, LN18 and RBLPAT possess an aligning peak cluster (red dashed box) however multiple unique RBLPAT and U251 peak clusters are present within the gene-poor region. This trend of peak distribution was observed in all chromosomes.

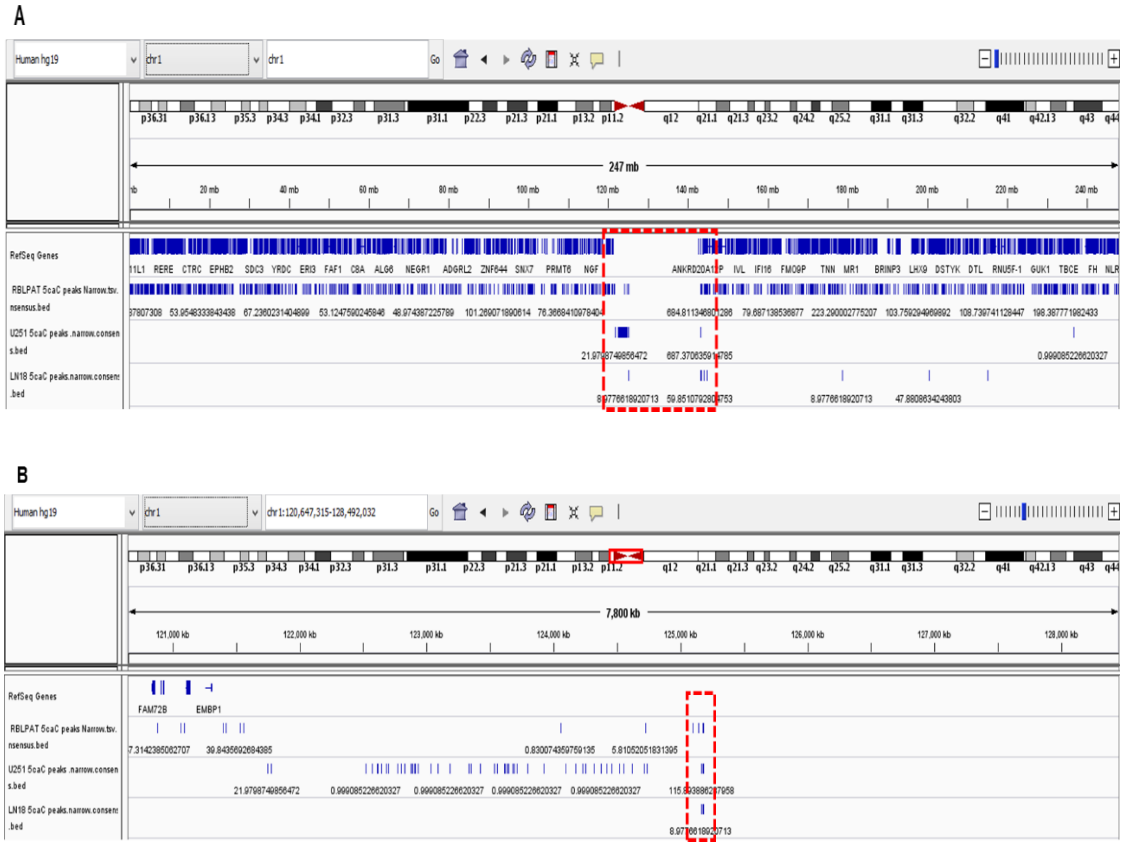


Figure 5.7 Distribution of 5caC consensus peaks across cell lines. (A-B) 5caC consensus narrow peaks exhibit enrichment at centromeres in U251 and LN18 but anti-correlate with RBLPAT which shows depletion (red dashed boxes).

To gauge visually how 5caC peaks within regulatory and gene body sequences were represented, constitutive hiPSC genes *SOX2* (fig 5.8A-B) and *NANOG* (fig 5.8A-B) were interrogated relative to U251 and LN18 cell lines. RBLPAT 5caC peaks were observed within approximately 500bp of the transcriptional start site (TSS) of *SOX2* and spanned approximately 300-400bp. Peaks were highly enriched for 5caC over input DNA at the same coordinates. Fewer reads and hence smaller peaks were observed in both U251 and LN18 relative to RBLPAT and both GBM cell lines displayed inconsistent 5caC peak distribution between biological duplicates compared to RBLPAT. The second biological repeats in each case (U251 5caC_2 and LN18 5caC_2) possessed higher and more numerous peaks though both biological repeats for both GBM cell lines were higher than their respective inputs. Similar trends in RBLPAT 5caC peak height and distribution were observed within the range of the *NANOG* TSS and exon 1 (not shown). Peaks highly enriched over the input and spanning greater than 500bp were detected in RBLPAT whereas inconsequential peaks at the same coordinates were reported in U251 and LN18.

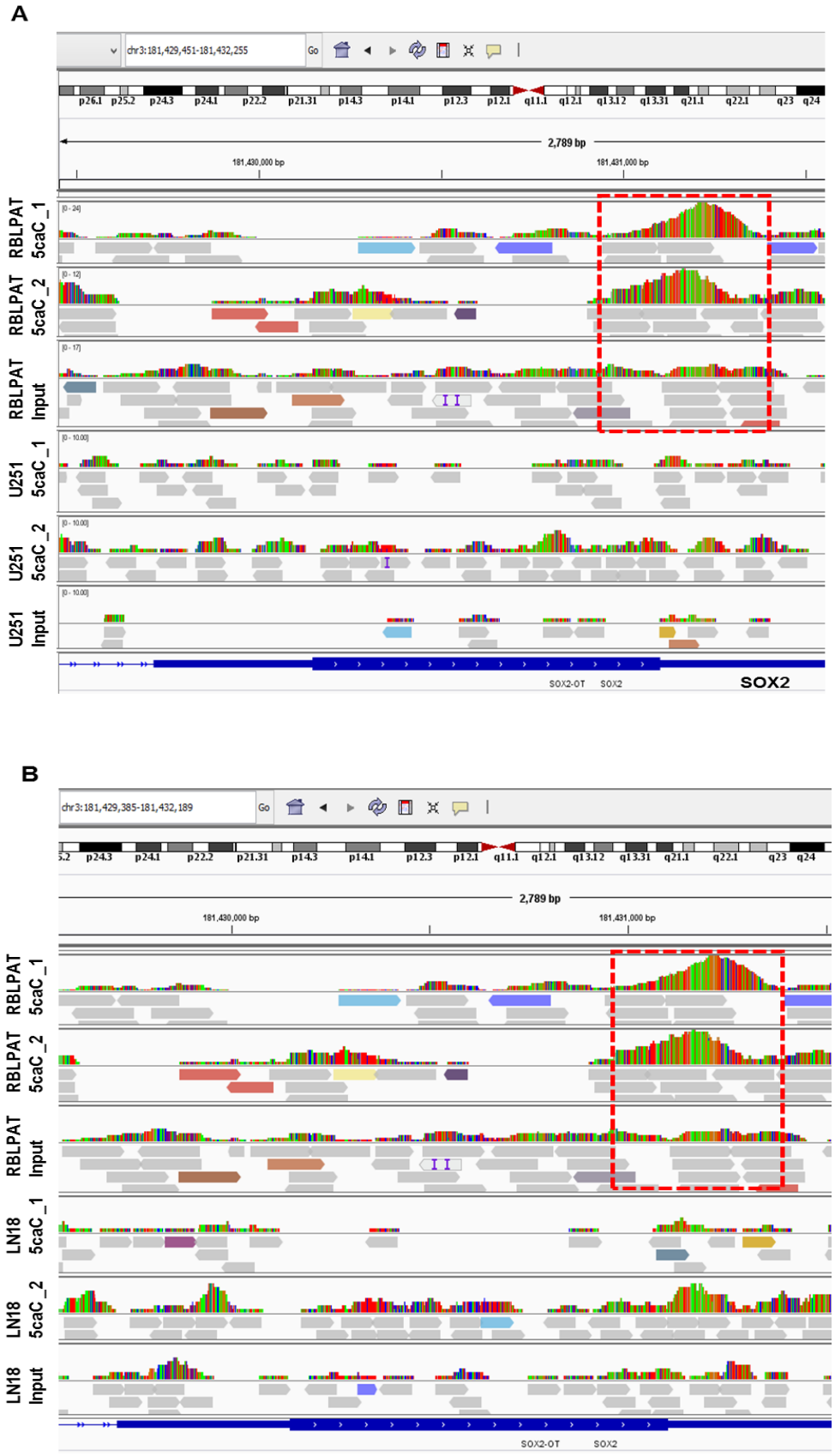
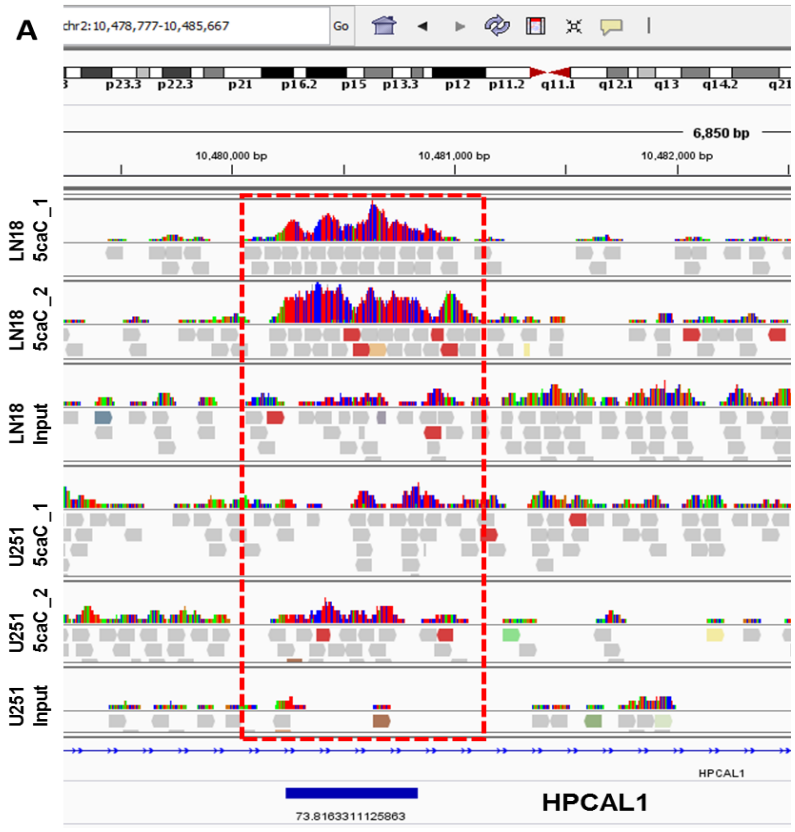


Figure 5.8 Consensus 5caC positive control peaks in RBLPAT cells. (A-B) 5caC narrow peaks displayed enrichment at markedly greater levels than input in RBLPAT (red dashed boxes) compared to U251 or LN18, n = 2.

5.33 LN18 and U251 Possess Distinct 5caC Enrichment Sites

LN18 and U251 exhibited differential distribution of 5caC peaks throughout the genome (fig 5.9A-B). LN18 5caC peaks were strongly enriched over input levels and spanned an approximate distance of 900-950bp in Hippocalcin-Like 1 (*HPCAL1*) introns. U251 5caC peaks at the same positions were smaller relative to LN18 but higher than U251 input levels. This pattern was mirrored at Synuclein Alpha Interacting Protein (*SNCAIP*) introns with non-negligible 5caC peaks observable (approximately 600bp) but fewer and heavily reduced in magnitude compared to LN18 5caC peaks at the same positions.



.

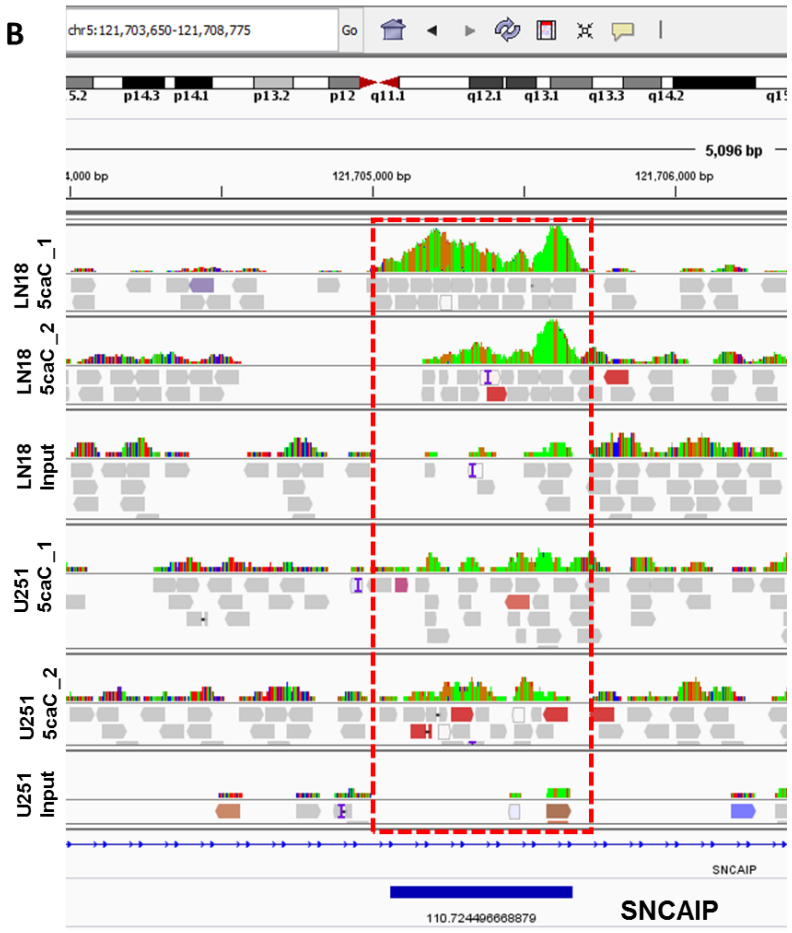


Figure 5.9 Differential 5caC genomic occupancy within genes in LN18 cells. (A) 5caC consensus peaks within *HPCAL1* (red dashed box) were unique to LN18. (B) Differential 5caC enrichment within SNCAIP gene sequence (red dashed box) was unique to LN18.

5.34 5caC and WT1 Peaks Align in U251

Large 5caC peaks spanning close to 1.5kb were observed at the overlapping sequences of Galectin 8 (*LGALS8*) exon 8 and HEAT-Repeat containing 1 (*HEATR1*) exon 45 in U251 (fig 5.10A-B). 5caC peaks were observed in LN18 however they appeared indistinct from LN18 input levels and were hence disregarded. Unexpectedly, WT1 ChIP

peaks mapped exactly to 5caC DIP peaks at the overlapping *LGALS8* – *HEATR1* exonic sequence and the peaks were similar in magnitude and coverage (fig 5.10B). Closer inspection of WT1-5caC peaks at 220bp resolution revealed the substantial enrichment of peak heights over the U251 input, n = 2.

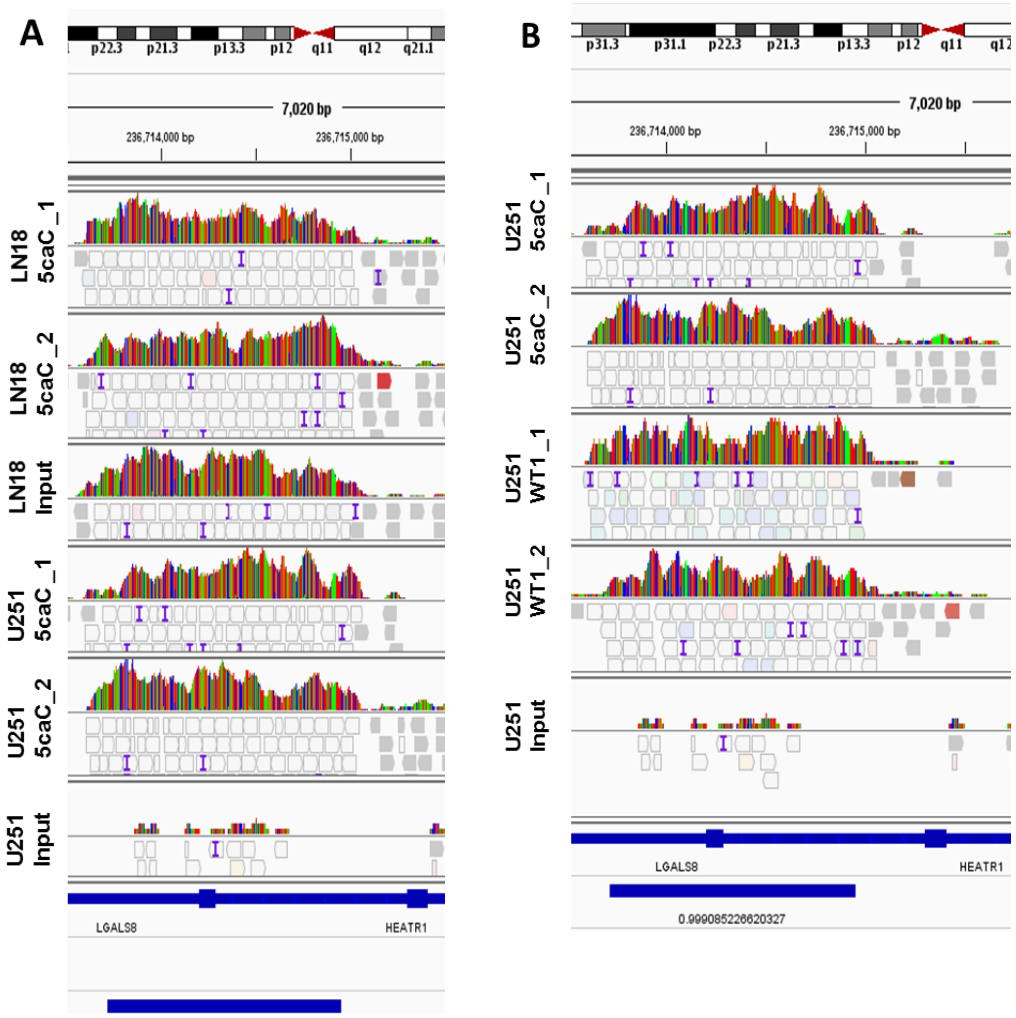


Figure 5.10 Differential 5caC genomic occupancy between U251 and LN18 aligns with WT1 enrichment in U251. (A) U251 and LN18 display differential 5caC enrichment within *LGALS8* and *HEATR1* sequences as LN18 peaks are

disregarded as negligible relative to its input. (B) 5caC consensus peaks align with WT1 ChIP peaks at LGALS8 and HEATR1 sequences, n = 2.

Contra V3 powered by the TRANSFAC database detected multiple WT1 binding sites at the approximately 200bp resolution *LGALS8-HEATR1* 5caC and WT1 peak in U251 (fig 5.11A-D). Predicted WT1 TFBS were depicted based on positional weight matrices (PWMs) which illustrate statistically inferred possible WT1 binding sequences (fig 5.10C). The WT1 TFBS PWN designated WT1_Q6.M01118 (fig 5.10C top) was predominately represented (green) in the 5caC peak sequences in *LGALS8-HEATR1* overlap. Other PWM such as M6535_1.02 (orange) and M02036 (blue) provided by the Homo sapiens and TRANSFAC databases respectively did not detect WT1 binding sites represented by their respective motifs in the WT1-5caC peak-called region. PWM analysis recognised WT1 binding sites within the area spanned by WT1-5caC peaks as 5'-GTGAGGCTC-3' and 5'-CTCTCCCAG-3' on reverse and forward strands respectively. Cross-referencing against the UCSC genome browser demarcated the position of predicted WT1 binding sites (green blocks) within the 5caC/WT1 peaks at *LGALS8* exon 8 and *HEATR1* exon 45, thus verifying the WT1 ChIP peak calling result (fig 5.10D). Interestingly, aligned 5caC peaks were observed within Lamin B1 (*LMNB1*) introns in U251 and LN18 however no WT1 peaks at these coordinates were observed despite the presence of a WT1 binding site within the 5caC peak locations as predicted by TRANSFAC M01118 PWM (data not shown).

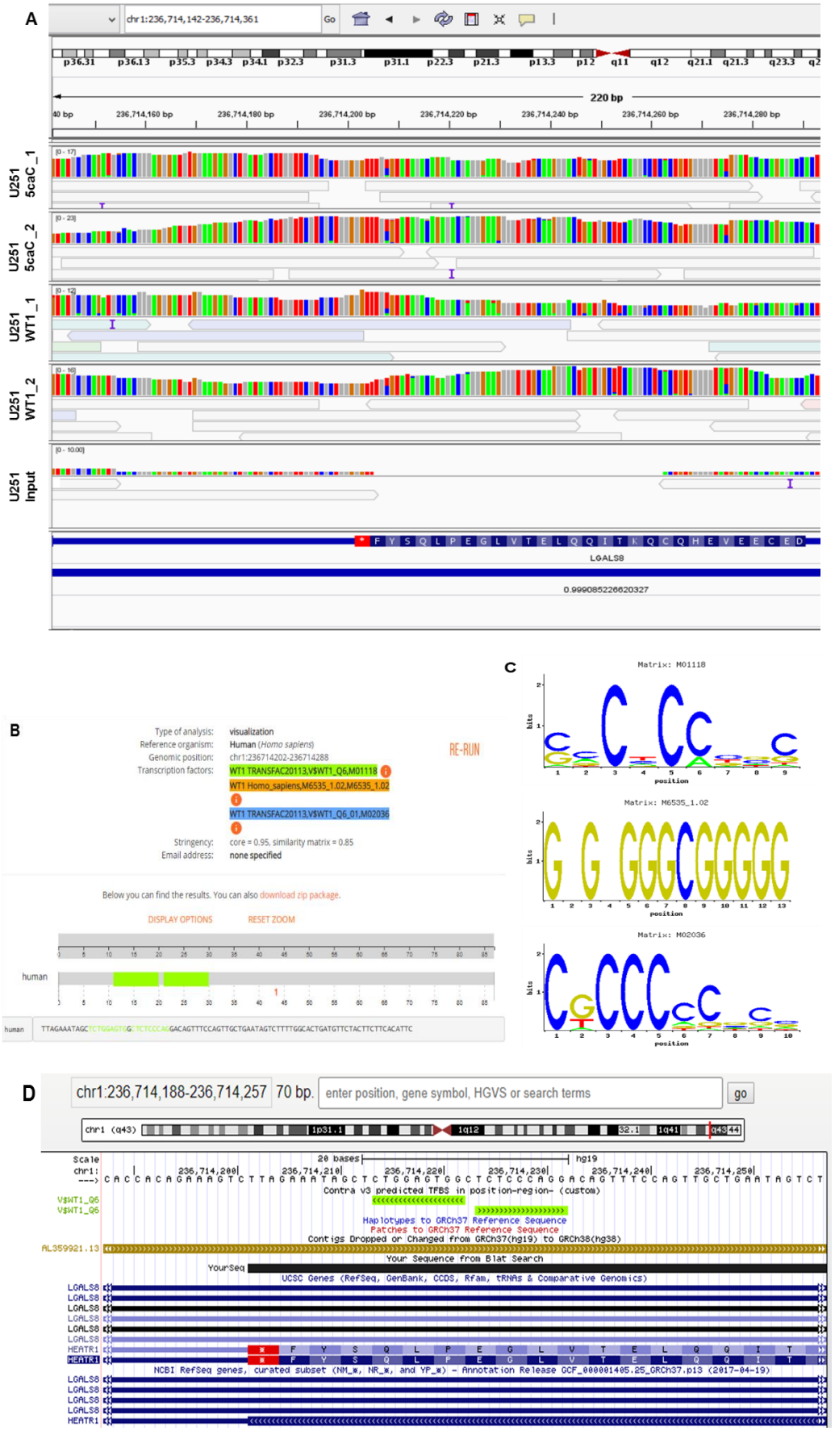

A
chr1:236,714,142-236,714,361
Go
220 bp
U251 5caC_1
U251 5caC_2
U251 WT1_1
U251 WT1_2
U251 Input
LGALS8
0.999085226620327
B
Type of analysis: visualization
Reference organism: Human (Homo sapiens)
Genomic position: chr1:236714202-236714288
Transcription factors: WT1 TRANSFAC20113,V$WT1_Q6,M01118
WT1 Homo_sapiens,M6535_1.02,M6535_1.02
WT1 TRANSFAC20113,V$WT1_Q6_01,M02036
Stringency: core = 0.95, similarity matrix = 0.85
Email address: none specified
RE-RUN
Below you can find the results. You can also download zip package.
DISPLAY OPTIONS
RESET ZOOM
human
C
Matrix: M01118
Matrix: M6535_1.02
Matrix: M02036
bits
position
D
chr1:236,714,188-236,714,257
70 bp.
enter position, gene symbol, HGVS or search terms
go
Contra v3 predicted TFBS in position-region- (custom)
V$WT1_Q6
Haplotypes to GRCh37 Reference Sequence
Patches to GRCh37 Reference Sequence
Contigs Dropped or Changed from GRCh37(hg19) to GRCh38(hg38)
AL359921.13
Your Sequence from Blat Search
YourSeq
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
LGALS8
HEATR1
NCBI RefSeq genes, curated subset (NM_*, NR_*, and YP_*) - Annotation Release GCF_000001405.25_GRCh37.p13 (2017-04-19)

Figure 5.11 *In silico* Verification of WT1 ChIP peaks. (A) 220bp resolution of U251 WT1 / 5caC aligned peaks at *LGALS8* sequence. (B) Contra V3 transcription factor binding site prediction of WT1 binding within LGALS8. Green blocks represent WT1 binding sites with predicted sequence in green. (C) Positional weight matrices of predicted WT1 binding sequences, M01118 was most prominently predicted and was denoted by the green block predicted by Contra V3. (D) UCSC genome browser representation of WT1 M01118 binding motif coordinates within LGALS8 (green arrowed tracks), n = 2.

5.35 LN18 and U251 Share Common 5caC Genomic Occupancy

Genes possessing consensus 5caC narrow peaks exclusively in either U251, LN18 or present in both were analysed to determine the distribution of 5caC presence between the two GBM cell lines (fig 5.12). Across the whole genome, 17 5caC peaks were observed exclusively within LN18 which included genes such as *HPCAL1*, *SNCAIP*, PRELI Domain-containing 2 (*PRELID2*), Alkylglycerol mono-oxygenase (*AGMO*) and CUB and Sushi Multiple Domains (*CSMD1*). *LGALS8*, *HEATR1*, Double Homeobox 4 (*DUX4*), Rho Associated Coiled-Coil Containing Protein Kinase 1 Pseudogene 1(*ROCK1P1*), Family With Sequence Similarity 230 Member C (*FAM230C*) and Actinin Alpha 2 (*ACTN2*) 5caC peaks were exclusively reported in U251 whilst peaks within genes Arf-GAP With Dual PH Domain-Containing Protein 1 (*ADAP1*), *LMNB1*, Dmx Like 1 (*DMXL1*), Phosphatidylinositol-3,4,5-Trisphosphate Dependent Rac Exchange Factor 2 (*PREX2*), Tumour

Necrosis Factor receptor (TNFR)-Associated Factor 6 (*TRAF6*) and others were common between U251 and LN18.

A

LN18 5caC	U251 5caC	LN18+ U251 5caC
HPCAL1	LGALS8	ADAP1
SNCAIP	HEATR1	LMNB1
PRELID2	DUX4	DMXL1
AGMO	ROCK1P1	PREX2
CSMD1	ACTN2	TRAF6
K1AA1549	FAM230C	MIR100HG
CPA6		ZNF609
MDM2		MMP28
NTN1		KCNK2
NBPF20		HUNK
RALGPS2		ADAP1
PIK3R5		DACH2
EFCAB6-AS1		FOXK1
KCNU1		C8orf34
CCDC26		ANKDD1A
LINC02208		NAALAD2
SVOPL		SHOX
		PACSIN2
		MIR3681HG
		AC060834.2
		AC109466.1
		AC087386.1
		AC138701.2
		LINC02301
		ASIC2
		AP005262.2

B

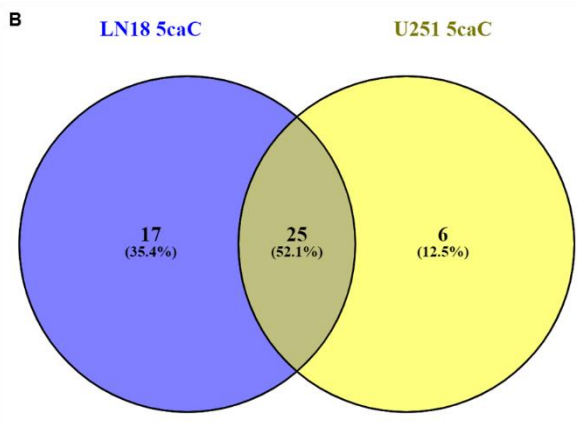


Figure 5.12 5caC peak distribution within genes in U251 and LN18 cell lines. (A) Identities of protein coding genes, miRNA and LincRNA sequences common (grey) and unique to LN18 (blue) and U251 (yellow). (B) Proportion of 5caC peaks common or unique to LN18 and U251, n =3.

5.36 A Subset of Genes Possess WT1 and 5caC Aligned at Identical Genomic Coordinates

Interrogation of 5caC and WT1 narrow consensus peaks revealed that multiple 5caC exclusive peaks were present on genes such as *LGALS8*, *HEATR1* and Double Homeobox 4 Like 17 (*DUX4L17*) whereas WT1 exclusive peaks were reported on genes Annexin A6 (*ANXA6*), Potassium Voltage-Gated Channel Subfamily Q Member 5 (*KCNQ5*),

BCL2-Like 13 (*BCL2L13*) and *LINE1* repetitive elements *L1MA2* and *L1MB4* (fig 5.13A-B). Unexpectedly, 9 genes displayed both 5caC and WT1 genomic occupancy which represented 25% of total genes interrogated for 5caC and WT1 enrichment. Transcriptome analysis of genes possessing 5caC or WT1 enrichment or both identified genes significantly altered in their expression in the C5A *wt1* mutant relative to control U251 under neurobasal conditions (fig 5.13C). A subset of genes (marked with an asterisk) which included *LGALS8* ($p = 0.0094$) and *KCNQ5* ($p = 0.011$) displayed significant differential transcriptional expression. Genes microRNA *MIR3681HG* and *LMNB1* exhibited 5caC peaks only whereas *KCNQ5* possessed WT1 only enrichment peaks in U251. Curiously, *ANXA6* and *BCL2L13* which both possessed exclusive WT1 enrichment were down-regulated in C5A but not significantly (*ANXA6*, $p = 0.083$; *BCL2L13*, $p = 0.26$). Interestingly, *LGALS8* which displayed both 5caC and WT1 occupancy produced the greatest fold change in expression (-1.53) of any gene in the analysis.

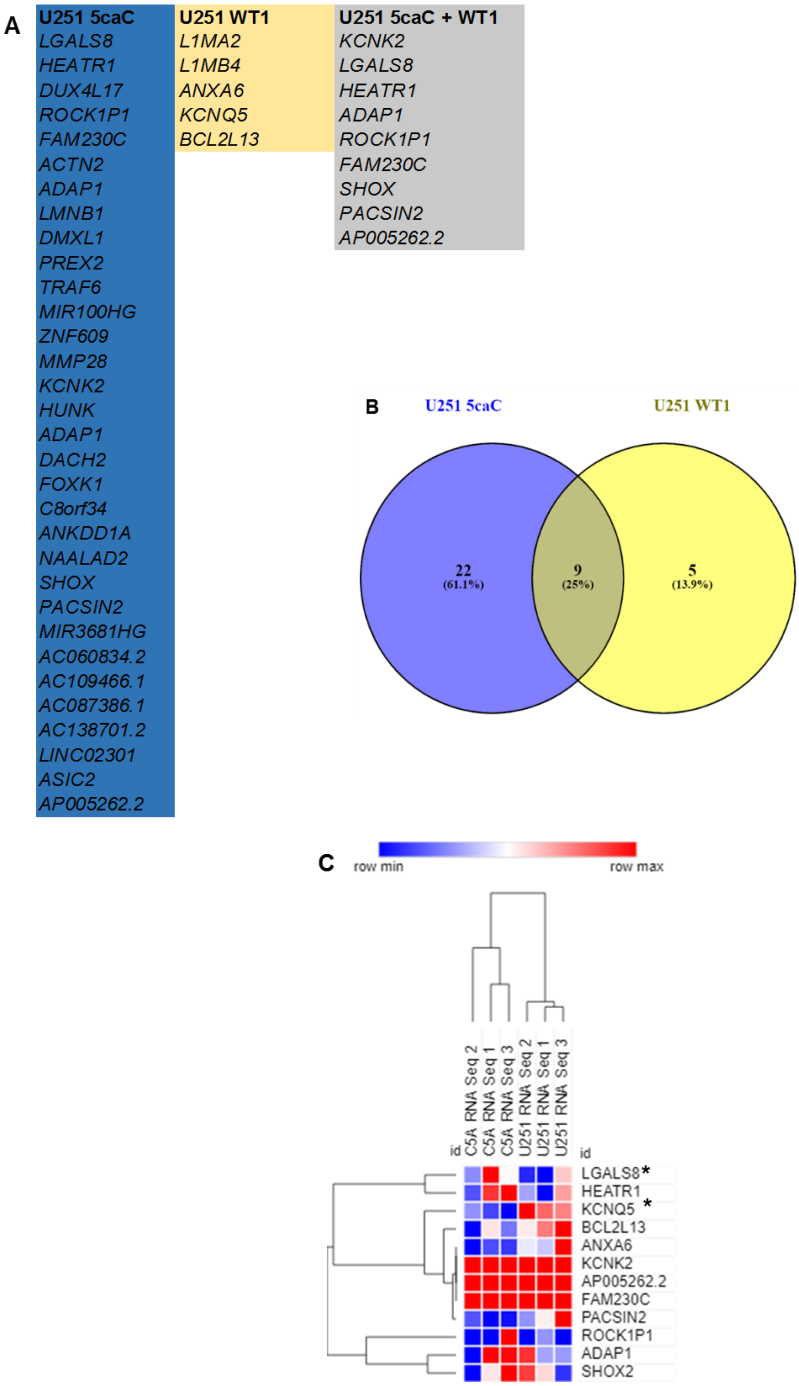


Figure 5.13 Distribution of 5caC and WT1 peaks within U251. (A) Identities of 5caC and WT1 peaks within U251. (B) Proportion of common and unique 5caC and WT1 peaks. (C) Transcriptional profile of genes possessing 5caC and WT1 peaks in U251 vs C5A cell lines. * denotes significantly different transcript levels, n =3.

5.37 5caC Presence Does Not Faithfully Correlate with Transcriptional Up-Regulation of Genes

To understand the relationship between 5caC and transcription, 5caC DIP-Seq narrow consensus peaks were examined to determine if their enrichment on stemness, differentiation and GBM pathology genes correlated with their relative expression (fig 5.14A-C).

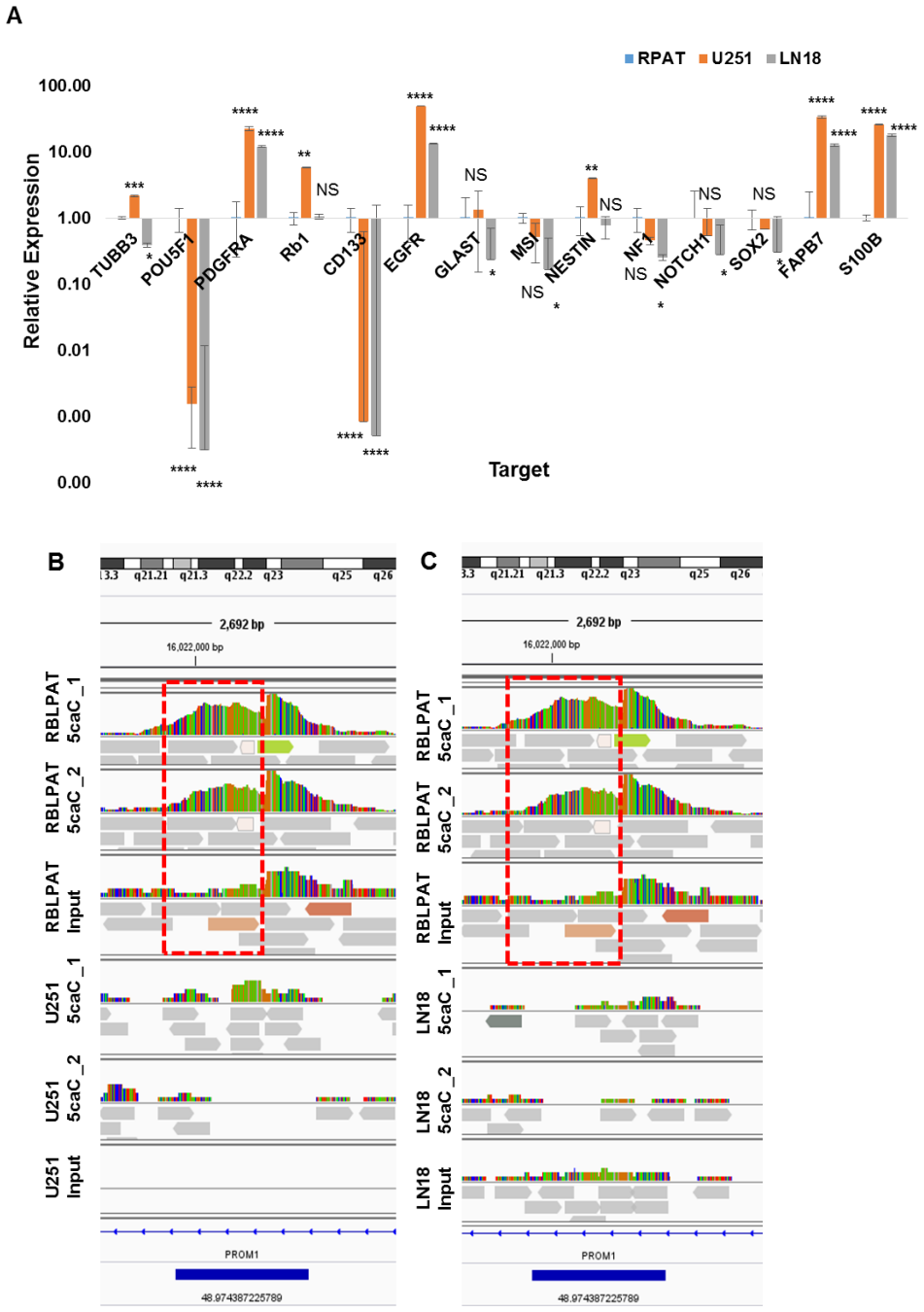


Figure 5.14 Correlation of transcriptional activity with 5caC genomic occupancy. (A) QPCR analysis of stemness and neuronal / glial lineage and GBM pathway genes in U251 and LN18 relative to RBLPAT, mean ± SD, unpaired, two tailed Kruskal-Wallis test, NS $p \geq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$, n = 3. (B-C) 5caC consensus peaks enrichment in RBLPAT (red dashed box) relative to U251 and LN18 within the *PROM1* (CD133) sequence, n =2.

Analysis of transcriptional activity was performed by qPCR of multiple stemness, differentiation and GBM genes in U251 and LN18 relative to RBLPAT hiPSCs. Although qPCR analysis depicted highly significant differences in transcriptional activity between RBLPAT and GBM cell lines e.g. *POU5F1* (Oct4) was 3 (0.0016, $p < 0.0001$) and 4 (0.00031, $p < 0.0001$) orders of magnitude decreased in U251 and LN18 respectively relative to RBLPAT, 5caC narrow consensus peaks were only observed at intronic regions of *Prominin1* (*PROM1*) which encodes CD133 surface protein. No consensus peaks were consistently observed at any of the genes up-regulated in GBM cell lines e.g. *PDGFRA* or *EGFR* despite their 22.6 and 12.1-fold and 49.4 and 13.3-fold increase in expression in U251 an LN18 respectively. This analysis is limited by lack of 5caC DIP-Seq data for LN18 and U251 in which 5caC depletion has occurred e.g. via *TDG* over-expression. Thus RBLPAT hiPSCs were used as a positive control for 5caC presence on stem-related genes.

5.4 Transcriptomic Analysis of C5A *wt1* Mutants

Transcriptome analysis of the C5A *wt1* mutant was performed to examine the mutational effects of the CRISPR cas9 mediated mutagenesis of the *WT1* DNA binding domain on the Transcriptome, with particular emphasis of DNA demethylation machinery, neural stem cell multipotency and GBM pathologically relevant genes.

5.41 C5A *wt1* Mutants Highly Express a Novel *wt1* Isoform

Examination of *wt1* expression in C5A mutants versus U251 control cells (fig 5.15A) demonstrated the consistency of expression of 12 *WT1* transcripts (201- 212) which exhibited no fold change between C5A and U251. Curiously, further investigation into *WT1* transcript structure uncovered the existence and high expression of a novel unidentified transcript (fig 5.15B) at 29222 showing high expression ranging from 0.87 to 1.33 in magnitude. The novel isoform appeared to possess 1 exon spanning from the position of exon 3 to exceeding exon 5. This isoform was reported in 2 out of 3 C5A RNA-Seq biological repeats and was not observed in any control U251 samples.

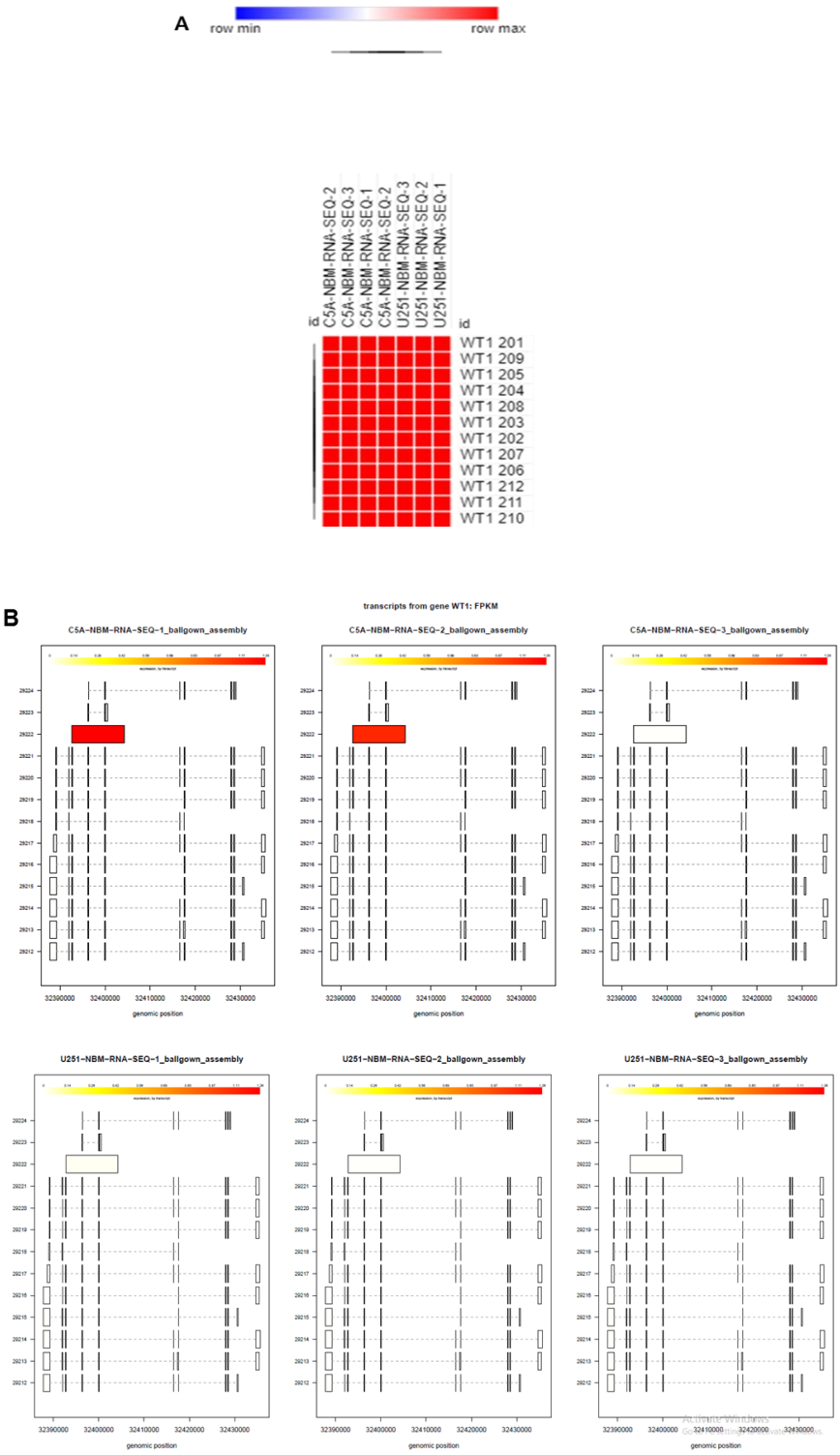


Figure 5.15 Analysis of WT1 isoform expression between U251 and C5A cell lines. (A) RNA-Seq analysis of 12 WT1 isoforms in U251 vs C5A. (B) Ball gown assembly plots depicting *WT1* transcript variants and their expression levels in U251 controls vs C5A *wt1* mutants. Red boxes denote a highly

expressed truncated isoform present in 2/3 C5A biological repeats but absent in all U251 control repeats, n = 3.

5.42 Multiple Genes Demonstrate Strong Differential Expression between C5A *wt1* Mutant and U251 Control Cells

Genes of interest relevant to DNA demethylation, GBM pathology, neural stem cell self-renewal and differentiation in addition to cell cycle related and signalling pathway mechanisms were examined and ordered by supervised hierarchical clustering (fig 5.16). Genes such as *TET3*, *EGR1*, Axin-2 (*AXIN2*) and *WT1* in the heatmap represented targets significantly transcriptionally up-regulated in C5A compared to U251. Significantly down-regulated targets included Integrin beta-1 (*ITGB1*), Cadherin 12 (*CDH12*), *NOTCH1* and *TET1*. Targets E-Cadherin (*CDH1*), GLI Family Zinc Finger 1 (*GLI1*), *DUX4*, *FAM230C* and *CDKN2A* displayed no fold change in expression whilst the remaining target genes displayed non-significant fluctuations in transcriptional expression. Considering the thousands of genes within the transcriptome analysis, computation of expression of all transcripts simultaneously in one heatmap proved impractical. Thus the first 150 genes were selected and plotted by unsupervised hierarchical analysis to determine differentially expressed targets between C5A and U251 (fig 5.16B). Genes exhibiting differential expression were over-represented relative to those with non-consistent and non-significant changes. To understand the relationships between differentially regulated genes between C5A and U251,

Biological Process (BP) Gene Ontology (GO) was performed on the differentially expressed genes (fig 5.17A-B). Many genes within the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) proteomic network displayed overlapping gene ontologies however the highest represented terms included 'regulation of response to external stimuli', 'anatomical structure development', 'response to stimulus', 'extracellular structure organisation', developmental processes, 'multicellular organismal processes' and 'multicellular organism development'.

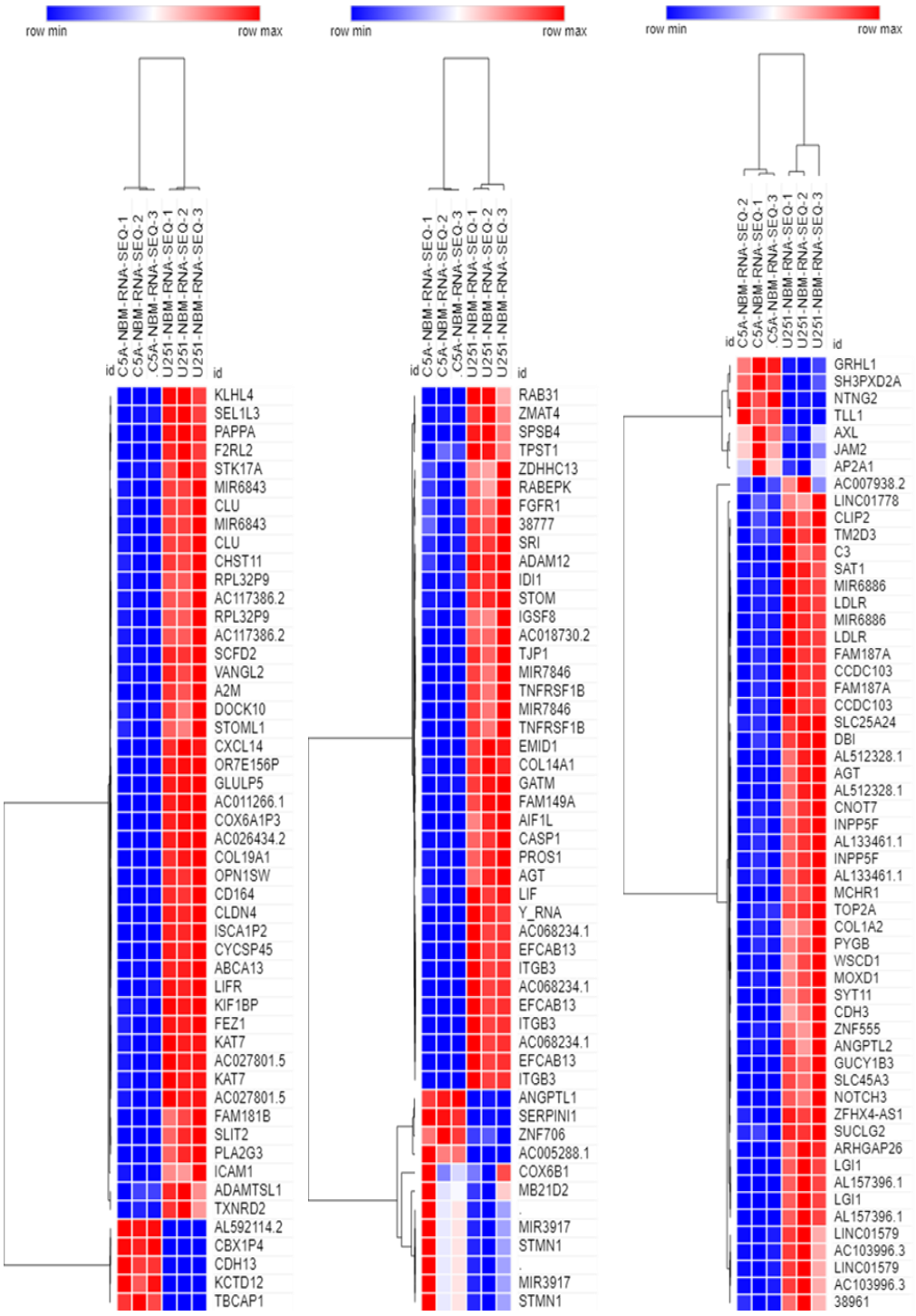


Figure 5.16 Transcriptome analysis of wt1 mutagenesis in U251 and C5A cell lines under neurobasal conditions. All targets exhibit significantly different transcription levels, $p < 0.05$ 0.0001, $n = 3$.

Proteomic interaction networks were produced to illustrate the relationship between differentially up regulated genes (fig 5.18A) and differentially down regulated genes contextualised around WT1 presence (fig 5.18B). No relationships were evident in up-regulated genes as no visible edges connecting nodes were observed. In contrast, while many protein nodes were isolated, direct connections relating to WT1 (red dashed circle) were detected for Leukaemia Inhibitory Factor (LIF) and Tight Junction Protein 1 (TJP1). LIF connected directly to Leukaemia Inhibitory Factor Receptor (LIFR) and TJP1 connected directly to Intercellular Adhesion Molecule 1 (ICAM1) which represented a central node in multiple branching sub-networks. Interestingly, multiple proteins in the ICAM1 network e.g. ICAM1, TJP1, Integrin beta-3 (ITGB3), Collagen Type I Alpha 2 Chain (COL1A2), Collagen Type XIV Alpha 1 Chain (COL14A1), Collagen Type XIX Alpha 1 Chain (COL19A1) and Cadherin 3 (CDH3) represented cell adhesion and extracellular matrix components.

A

Biological Process (GO)			
GO-term	*description*	*count in gene set*	*false discovery rate*
GO:0006887	exocytosis	14 of 774	0.0073
GO:0032940	secretion by cell	15 of 959	0.0088
GO:0032101	regulation of response to external stimulus	13 of 732	0.0088
GO:0072359	circulatory system development	13 of 807	0.0146
GO:0045055	regulated exocytosis	12 of 691	0.0146
GO:0048856	anatomical structure development	37 of 5085	0.0212
GO:0050896	response to stimulus	49 of 7824	0.0217
GO:2000257	regulation of protein activation cascade	4 of 54	0.0222
GO:0050727	regulation of inflammatory response	8 of 338	0.0222
GO:0048731	system development	32 of 4144	0.0222
GO:0043062	extracellular structure organization	8 of 339	0.0222
GO:0032502	developmental process	38 of 5401	0.0222
GO:0032501	multicellular organismal process	43 of 6507	0.0222
GO:0030449	regulation of complement activation	4 of 52	0.0222
GO:0007275	multicellular organism development	34 of 4726	0.0301
GO:0048861	leukemia inhibitory factor signaling pathway	2 of 4	0.0341
GO:0030198	extracellular matrix organization	7 of 296	0.0341
GO:0001775	cell activation	13 of 1024	0.0341
GO:1905907	negative regulation of amyloid fibril formation	2 of 5	0.0420

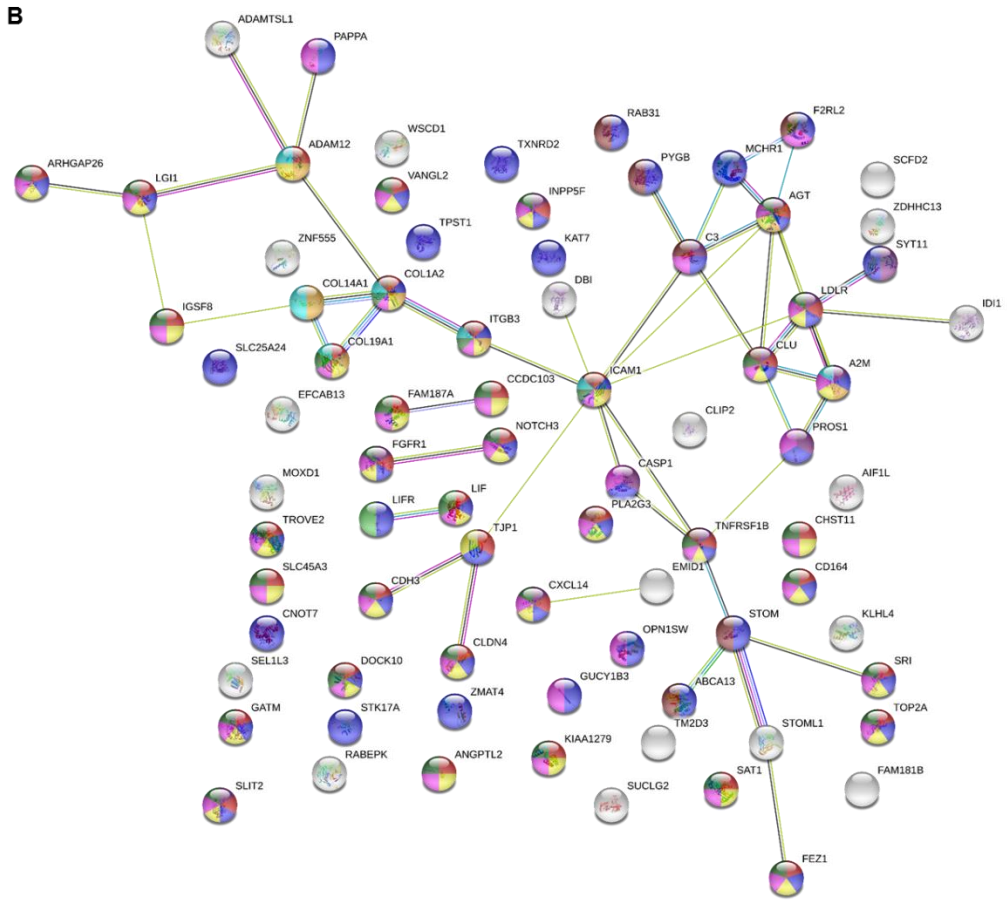


Figure 5.17 GO analysis of differentially regulated genes in U251 vs C5A under neurobasal conditions. (A) Biological process gene ontology of differentially regulated genes. Purple highlighted processes illustrate the process hits most frequently associated with the interrogated genes. Each biological process is represented by a coloured sphere. (B) STRING protein interactome map of differentially regulated genes. Coloured nodes correspond to biological processes from the GO list.

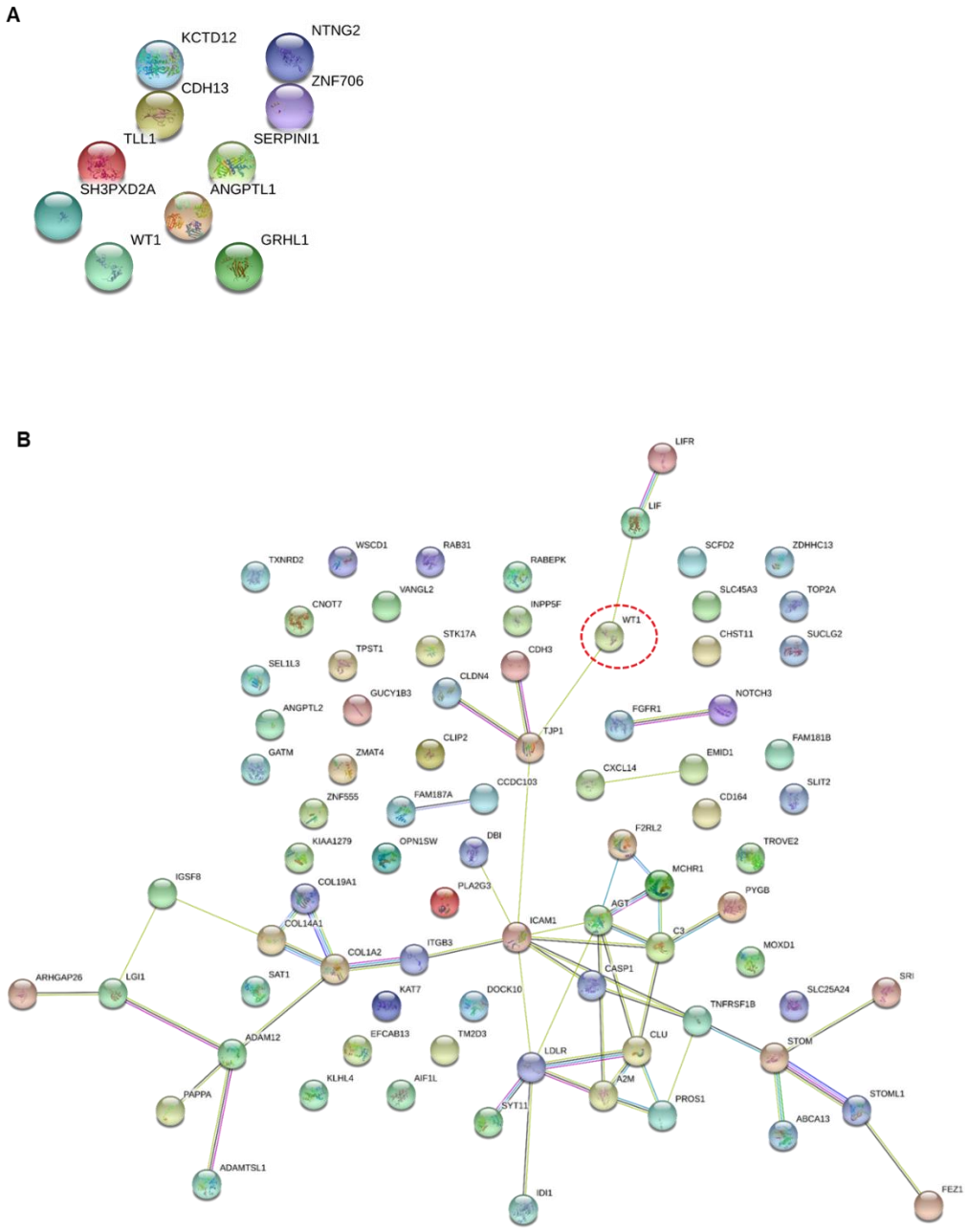


Figure 5.18 Examination of *in silico* protein interaction network between differentially regulated genes between U251 and C5A. (A) STRING interactome network of up-regulated genes in C5A relative to U251. Lack of edges between nodes demonstrates no interaction observed between proteins in this network. (B) STRING interactome network between transcriptionally down-regulated genes in C5A relative to U251. WT1 (red dashed circle interacts directly with LIF and TJP1.

5.5 Functional Characterisation of GBM *wt1* Mutants

To understand the potential functional implications of *wt1* mutagenesis on U251 and LN18 DNA demethylation and overall cell biology, functional assays including stem cell enrichment, cell proliferation, differentiation and *in vivo* orthotopic xenograft tumour formation studies were performed.

5.51 Mutagenesis of *WT1* May Affect DNA Demethylation Dynamics during Neural Stem Cell Enrichment

Immunofluorescent confocal microscopic investigation of U251, C5A, LN18 and C7 cell lines conditioned in neurobasal conditions for stem cell enrichment highlighted an effect of *WT1* mutagenesis on active DNA demethylation mechanisms (fig 5.19A-C). Significant 5hmC signal intensity reduction was observed between control and *wt1* mutant cell lines with a 27.3% reduction ($p = 0.0004$) reported in C5A compared U251 and a 28.1% decrease ($p = 0.0014$) detected in C7 relative to LN18. Transcriptome analysis of DNA demethylation and methylation genes in neurobasal versus serum conditions highlighted clusters of genes including *TET3*, *TDG*, *DNMT1* and *DNMT3B* which showed higher expression in serum than neurobasal media while *TET1* and *DNMT3A* were strongly up-regulated in neurobasal conditions. *TET2* exhibited inconsistent change in expression between conditions and *WT1* was

consistently highly expressed. Unexpectedly, genes displaying the greatest fold change in expression in C5A mutants under neurobasal conditions relative to U251 were *TET1* (1.60- fold change, p = 0.015) and *TET3* (-1.66-fold change, p = 0.016).

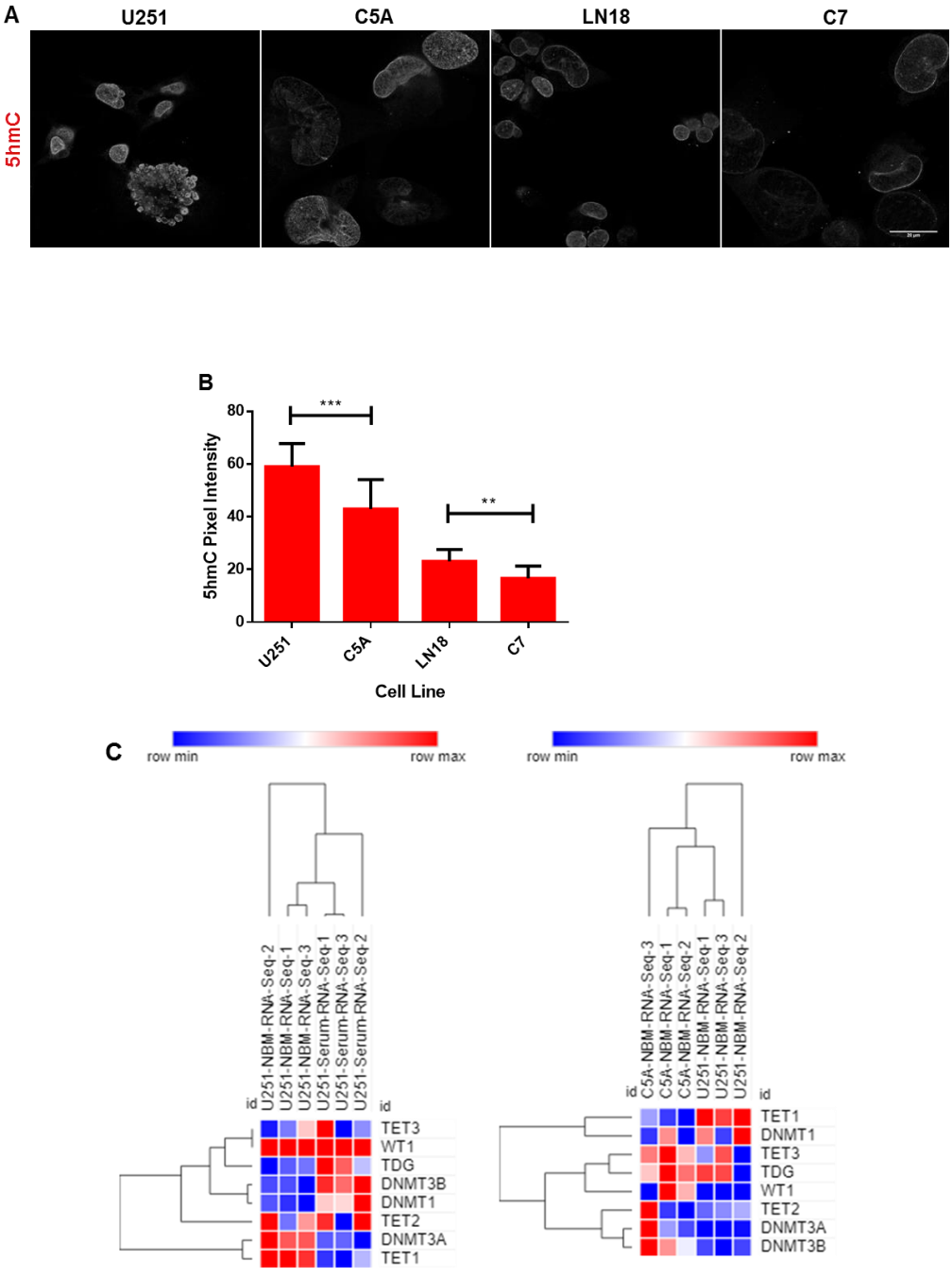
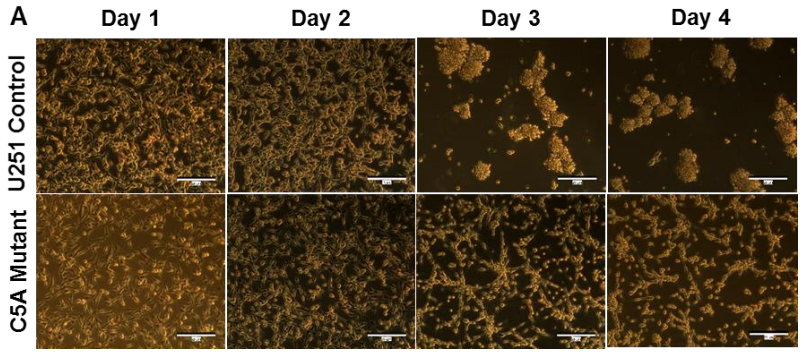


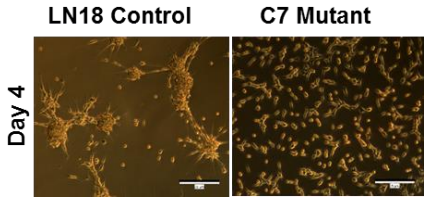
Figure 5. 19 Characterisation of the effects of wt1 mutagenesis on DNA demethylation under neural stem cell enrichment conditions. (A) Immunostaining of 5hmC magnitude and distribution in *wt1* mutant GBM cell lines versus their wildtype controls. Staining and images acquired at identical settings, scale bar represents 200µm. (B) Quantification of 5hmC immunofluorescent signal between *wt1* mutant and control cell lines. Mean ± SD, unpaired two tailed Student's t-test, ** $p < 0.01$; *** $p < 0.001$, $n = 3$. (C) Transcriptome analysis of DNA methylation and demethylation related genes in C5A compared to wild type controls, $n = 3$.

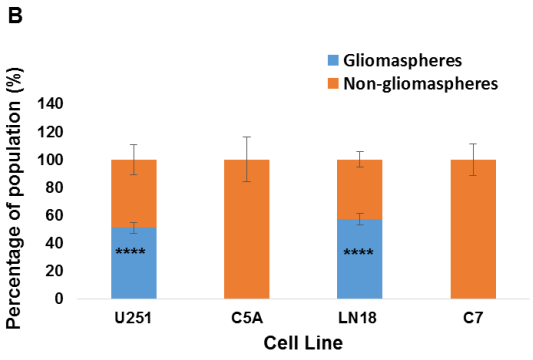
5.52 C5A and C7 *wt1* Display Diminished Gliomasphere Formation Capacity

To understand the implications of reduced TET1 expression and 5hmC presence in *wt1* mutants, C5A and U251 lines were assayed for neural stem cell enrichment potential by way of gliomasphere formation capacity. Neurobasal stem cell enrichment of C5A and C7 *wt1* mutant lines relative to their respective U251 and LN18 controls demonstrated the effect of *wt1* mutagenesis on gliomasphere formation over time (fig 5.20A-D). Neurobasal media conditioning produced gliomaspheres by day 4 of the assay in U251 and LN18 however no spheres were observed in either C5A or C7 mutant lines. C5A and C7 mutants remained attached to the flask surface throughout the assay whilst cell detachment from the flask surface was initially observed on day 3 for both U251 and LN18 lines. Branching connections between cellular projections were observed in C5A populations and to a lesser degree in C7 cultures. Gliomaspheres accounted for 51.16 and 57.14% of the cell population in U251 and LN18

respectively. No glioma spheres were observed in C5A or C7 (z score – 37.82, $p < 0.0001$ and z score – 32.61, $p < 0.0001$ respectively). Additionally, neural stem cell marker *NESTIN* displayed significant 7.7-fold reduction ($p = 0.0023$) in expression and E-cadherin encoded by *CDH1* gene was 25-fold down-regulated ($p < 0.0091$) in C5A mutants relative to U251. Neural stem cell related genes *SOX2* and *OLIG2* were up-regulated while *NOTCH1*, *POU5F1* and *NANOG* showed reduced expression but none by statistically significant margins. Similarly, C7 mutants displayed a 5-fold ($p = 0.015$) reduction in *NESTIN* expression, a 6.25-fold reduction in *NOTCH1* ($p = 0.0038$) and 3.57-fold decrease ($p = 0.0006$) in *NANOG* magnitudes but a dramatic 1666.67-fold decrease ($p < 0.0001$) in *CDH1* levels. No significant expression change was observed for *SOX2*, *OLIG2* or *POU5F1*.







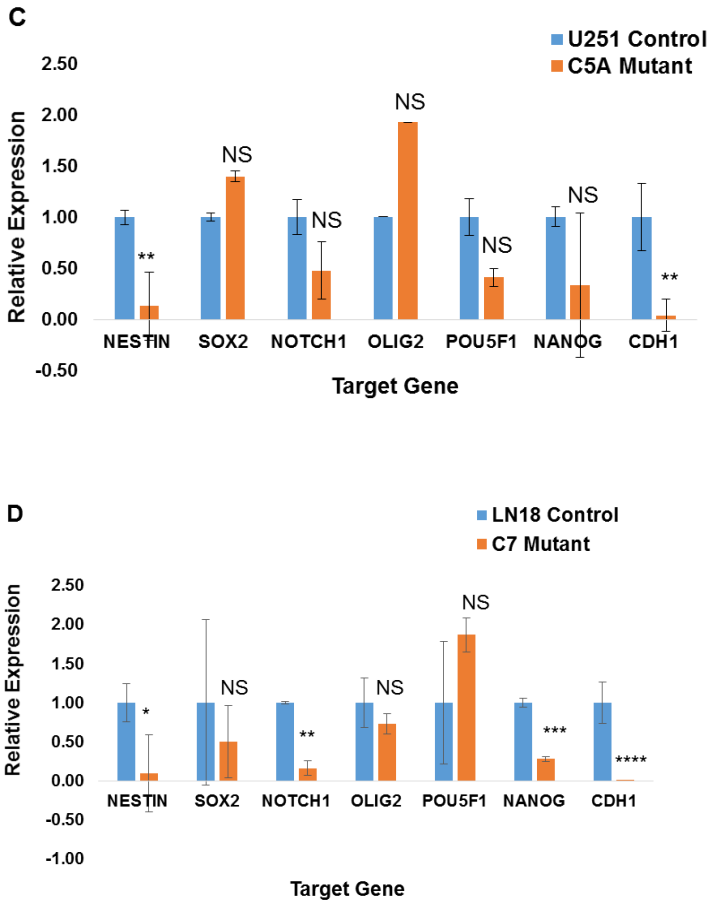
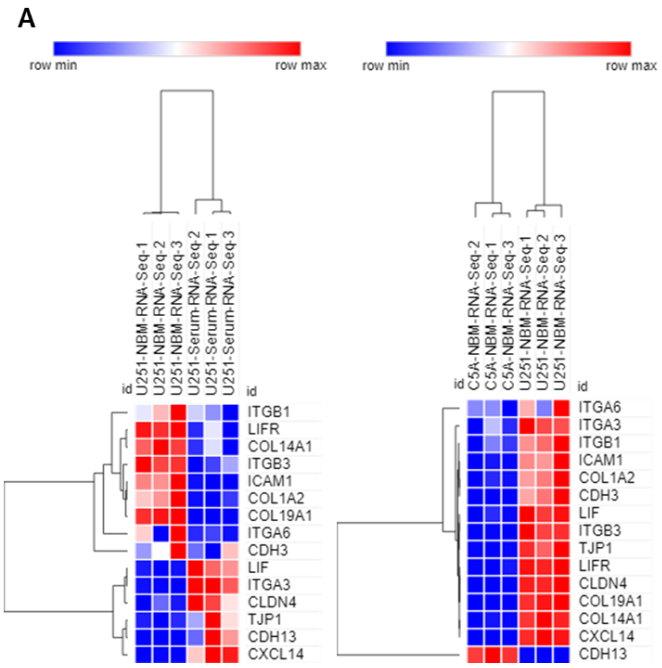


Figure 5.20 Effects of wt1 mutagenesis on gliomasphere enrichment. (A) Brightfield images of gliomasphere formation over time in wt1 mutants vs controls, images acquired at identical settings, scale bar represents 25µm. (B) Proportion of gliomaspheres observed per cell line. Mean ± SD, unpaired, two tailed z-test for proportions, **** $p < 0.0001$. (C-D) QPCR transcriptional analysis of neural stemness gene expression in *wt1* mutants vs controls. Mean ± SD, unpaired two tailed Mann-Whitney tests, NS $p \geq 0.05$; * $p < 0.05$; ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, $n = 3$.

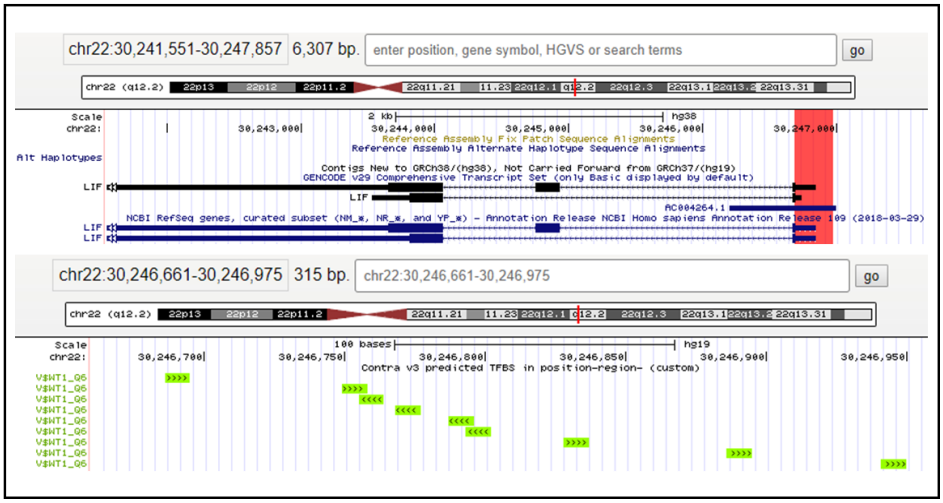
5.53 Cell Adhesion, Extracellular Matrix and Stemness Genes are Strongly Down-Regulated in C5A *wt1* Mutants

In light of the discovery of diminished neurosphere formation capacity in C5A mutants relative to U251, closer examination of C5A vs U251 neurobasal transcriptome profiles for the differentially expressed genes pertaining to cell adhesion, stemness and extracellular matrix (ECM) bioprocesses highlighted in section 8.42 was conducted (fig 5.21A-C). Supervised hierarchical clustering demonstrated a clear disparity in gene expression with *ITGB1*, *LIFR*, *COL14A1*, *ITGB3*, *ICAM1*, *COL1A2* and *COL19A1* displaying strong significant ($p < 0.05$) up-regulation in neurobasal conditions compared to serum. Genes *LIF*, Integrin Alpha 3 (*ITGA3*), Claudin 4 (*CLDN4*), *TJP1*, *CDH13* and Platelet factor 4 (*CXCL4*) showed diametrically opposed patterns of expression to the up-regulated neurobasal cluster and *ITGA6* and *CDH3* were not significantly up-regulated ($p = 0.15$ and $p = 0.93$ respectively) in neurobasal conditions. Remarkably, although differential expression was observed between neurobasal and serum conditions, all genes with the exception of *CDH13* showed significantly reduced expression in C5A mutants relative to U251, both in neurobasal conditions. Both *LIF* and *LIFR* possessed WT1 binding sites (fig 5.21 B-C) present within 500bp of their respective exon 1 (red highlighted regions) and PWM analysis predicted M01118 (green) motifs to predominate within *LIF* upstream regulatory regions whereas M02036 WT1 motifs (blue) were predicted in *LIFR* upstream regions.

A


row min
row max
row min
row max
U251-NBM-RNA-Seq-1
U251-NBM-RNA-Seq-2
U251-NBM-RNA-Seq-3
U251-Serum-RNA-Seq-2
U251-Serum-RNA-Seq-1
U251-Serum-RNA-Seq-3
id
ITGB1
LIFR
COL14A1
ITGB3
ICAM1
COL1A2
COL19A1
ITGA6
CDH3
LIF
ITGA3
CLDN4
TJP1
CDH13
CXCL14
C5A-NBM-RNA-Seq-2
C5A-NBM-RNA-Seq-1
C5A-NBM-RNA-Seq-3
U251-NBM-RNA-Seq-1
U251-NBM-RNA-Seq-2
U251-NBM-RNA-Seq-3
ITGA6
ITGA3
ITGB1
ICAM1
COL1A2
CDH3
LIF
ITGB3
TJP1
LIFR
CLDN4
COL19A1
COL14A1
CXCL14
CDH13

B


chr22:30,241,551-30,247,857
6,307 bp.
enter position, gene symbol, HGVS or search terms
go
chr22:30,246,661-30,246,975
315 bp.
Contra v3 predicted TFBS in position-region- (custom)
V$WT1_Q6

C

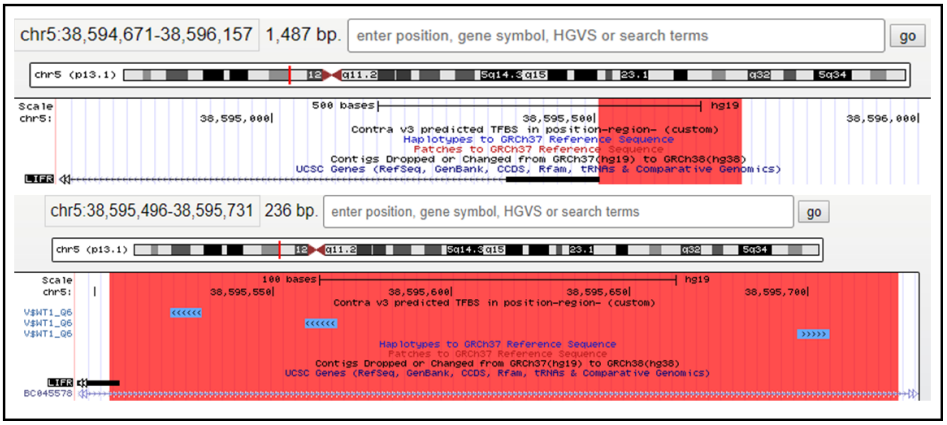
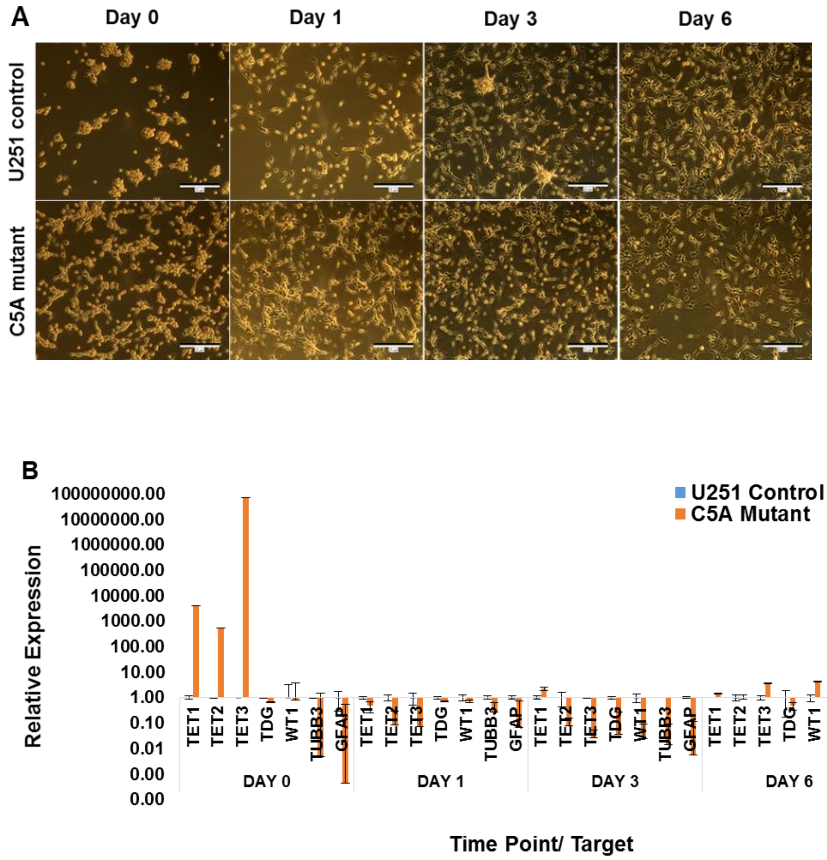

chr5:38,594,671-38,596,157
1,487 bp.
enter position, gene symbol, HGVS or search terms
go
chr5:38,595,496-38,595,731
236 bp.
Contra v3 predicted TFBS in position-region- (custom)
Haplotypes to GRCh37 Reference Sequence
Patches to GRCh37 Reference Sequence
Contigs Dropped or Changed from GRCh37(hg19) to GRCh38(hg38)
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
LIFR

Figure 5.21 Profiling of the effect of wt1 mutagenesis on neurobasal condition-regulated target genes in U251. (A) Transcriptome analysis of cell adhesion, stemness and extracellular matrix genes influenced by neurobasal media and wt1 mutagenesis, n =3. (B-C) UCSC Genome Browser tracks demonstrating WT1 transcription factor binding sites within LIF (B) and LIFR (C) promoter sequences near transcriptional start sites (red shaded regions). Green and blue arrowed tracks represent WT1 positional weight matrices M0118 and M02036 respectively.

5.54 C5A and C7 *wt1* Mutants Exhibit Diminished *TUBB3* and *GFAP* Expression Patterns during Differentiation

To further functionally characterise *wt1* mutants, C5A and C7 cell lines were differentiated to neuronal and glial lineages over a 6 day programme (fig 5.22 A-D). Morphologically, fewer branching cellular processes were observed in C5A mutants relative to U251, most discernable by day 6 of the neuronal programme. Overall, no significant morphological changes were observed between U251 control and C5A mutant cells. *TET1*, *TET2* and *TET3* were markedly up-regulated on the last day of neurobasal enrichment (day 0), with *TET3* in particular showing a $6.9x10^7$ fold elevation. Neuronal and glial specific markers *TUBB3* and *GFAP* were down-regulated by 210.1 and 2325.5-fold respectively. Overall, DNA demethylation genes *TET1*, *TET2*, *TET3* and *TDG* were decreased throughout the neuronal differentiation though lowest and highest expression of *TET3*, *WT1*, *TDG*, *TUBB3* and *GFAP* were reported at days 3 and 6 respectively. Negligible changes in expression were observed during glial differentiation in C5A cells relative

to U251. Fewer branching connections were observed during glial differentiation in C5A cells compared to U251 and fewer surviving cells were reported in C5A relative to U251 by day 6 of differentiation. Marginal changes in gene expression of DNA demethylation genes and GFAP were detected throughout the course of the glial programme. Greater morphological disparity existed between C7 and LN18 during both neuronal and glial differentiation for C7 mutants compared to C5A and U251 (fig 5.22A-D).



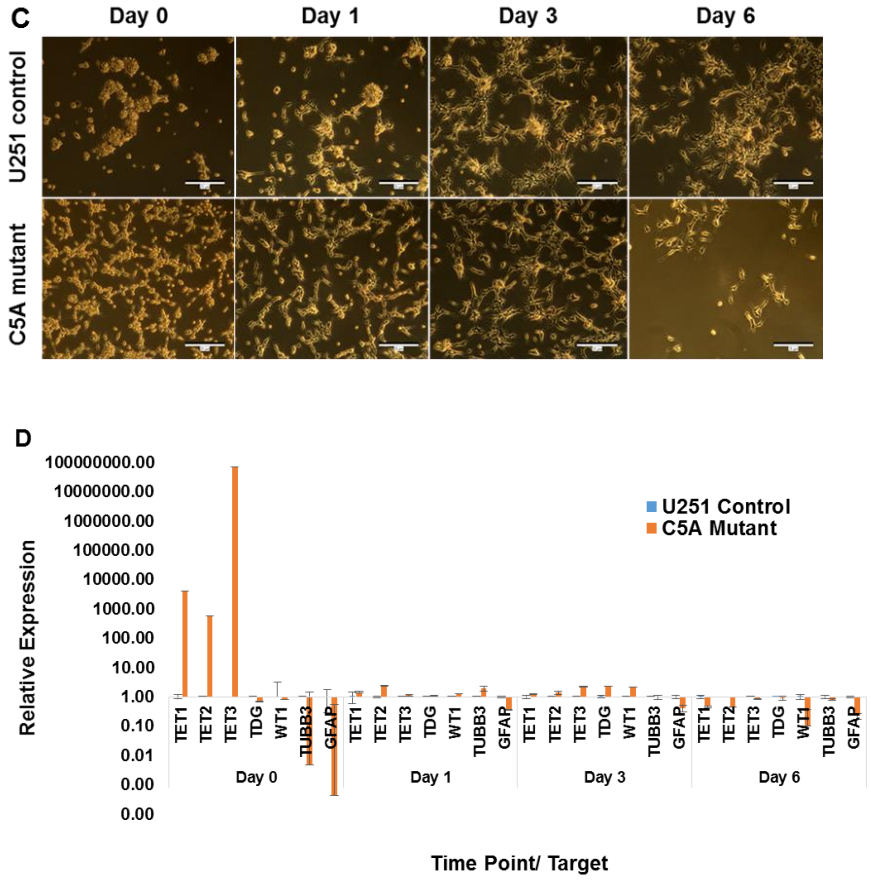


Figure 5.22 Effects of wt1 mutagenesis on GBM cell line differentiation to neuronal and glial lineages. (A&C) Morphology of U251 vs C5A neurobasal conditioned cells (day 0) during differentiation programmes (days 1-6). Images acquired at identical settings, scale bar represents 25µm. (B &D) QPCR analysis of DNA demethylation, *WT1*, *TUBB3* and *GFAP* genes during differentiation, n = 1.

Fewer C7 mutant cells survived the neuronal differentiation programme than LN18 however more distinct neurite outgrowth was observed in C7 mutants on days 1, 3 and 6 compared to LN18. During the neurobasal conditioning phase (day 0) *TET1*, *TET2*, *TET3* and *TDG* were up-regulated but no substantial changes in expression for any gene occurred during the neuronal programme. Again fewer C7 cells survived

the glial differentiation programme than LN18 but longer cellular processes were observed in C7 relative to LN18, most evident on days 1, 3 and 6. In contrast to neuronal differentiation *TET3* showed marked up-regulation on days 1 and 3 (2.8 and 3.8-fold increases respectively) and, notably, *WT1* and *EGR1* displayed 6.1 and 2.8-fold and 5.2 and 3.8-fold elevated expression on days 1 and 3 respectively. Curiously, *GFAP* expression remained down-regulated throughout the glial programme.

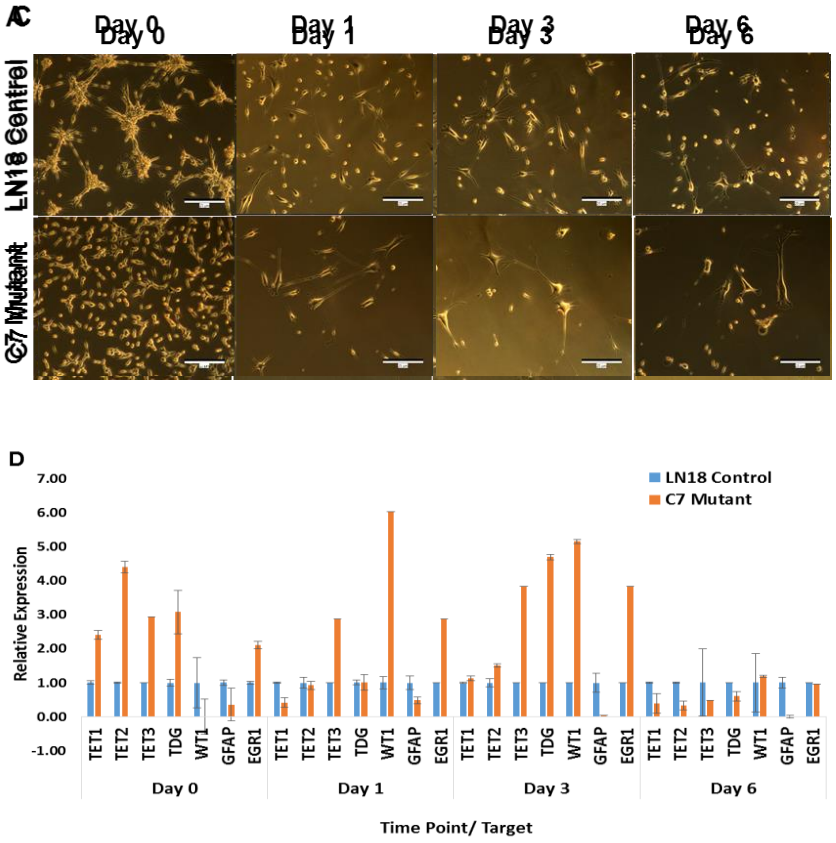


Figure 5.23 Effects of *wt1* mutagenesis on GBM cell line differentiation to neuronal and glial lineages. (A&C) Morphology of LN18 vs C7 neurobasal conditioned cells (day 0) during differentiation programmes (days 1-6). Images acquired at identical settings, scale bar represents 25µm. (B & D) QPCR analysis of DNA demethylation, *WT1*, *TUBB3* and *GFAP* genes during differentiation, n = 1.

5.55 C5A and C7 *wt1* Do Not Display Notable Changes in Cellular Proliferation

10 day cell proliferation assays were performed on C5A and C7 against their respective controls U251 and LN18 cell lines (fig 5.24A-B). No significant difference in cell number was observed at any of the sampling days for either C5A or C7 against their controls ($p = 0.825$ and $p = 0.665$ respectively).

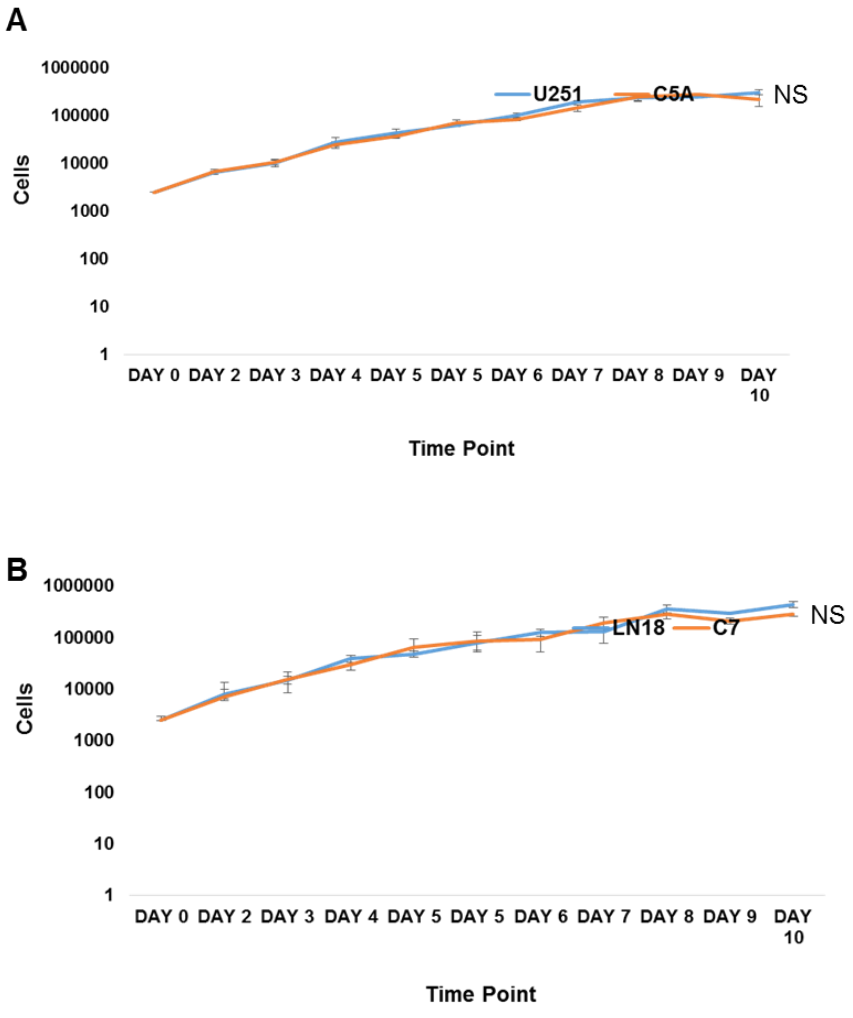


Figure 5.24 Effect of wt1 mutagenesis on cell proliferation. (A-B) Growth curves illustrating cell proliferation rates in C5A and C7 *wt1* mutants relative to

their *WT1* wild type U251 and LN18 controls respectively. Mean ± SD, unpaired two tailed Student's t-tests, NS $p \geq 0.05$, $n = 3$.

5.56 Reduced Size and Frequency of Tumour Formation in C5A *wt1* Orthotopic Xenograft Mice

Injection of C5A mutant cell lines into nude mice cerebellums was performed to determine the functional effects of *wt1* mutagenesis on GBM pathology in an *in vivo* physiologically representative environment (fig 5.25 A-B). Tumours were detected at the base of the cerebellum (red dashed circles) in 100% (4 out of 4) control mice injected with U251 however, remarkably, tumours were observed in 75% (3 out of 4) of C5A administered mice (appendix). Tumours in C5A injected mice were dramatically smaller and displayed significantly reduced proliferation, indicated by ki67 staining detected at 2.84-fold reduction ($p < 0.0001$) in C5A injected mice relative to U251 injected ones.

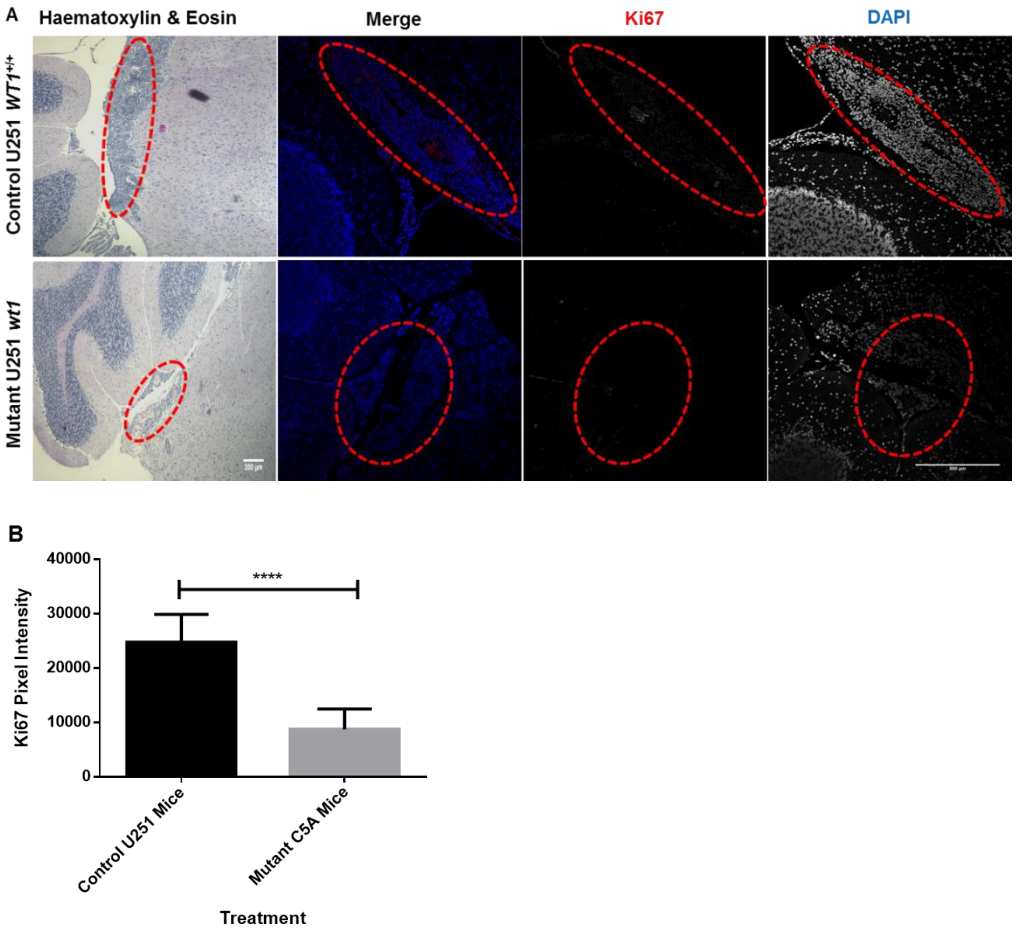


Figure 5.25 Effect of wt1 mutagenesis on *in vivo* tumour formation. (A) Haematoxylin and eosin staining and immunofluorescence interrogation of U251 vs C5A tumour engraftment (red dashed circles) and Ki67 staining. H&E images acquired at identical settings, immunostaining images captured at identical settings, scale bars represents 200µm. (B) Quantification of Ki67 in engrafted mouse tumours derived from U251 versus C5A injection. Mean ± SD, unpaired two tailed Mann-Whitney test, **** $p < 0.0001$, $n = 3$.

5.6 Technical Discussion

This chapter focused on characterising both descriptively and functionally the effects of CRISPR cas9 mediated mutagenesis of the *WT1* DNA binding domain in C5A and C7 mutant cells derived from U251 and LN18 control GBM cell lines respectively. The combination of Immunofluorescence staining coupled with confocal microscopy, MeDIP-Seq, ChIP-Seq, transcriptome profiling and *in vivo* orthotopic xenograft studies have provided powerful tools for analysing and understanding possible roles of WT1 activity in DNA demethylation and the potential phenotypic variations consequential in wt1 DNA binding null mutants.

Immunostaining and confocal analysis demonstrated C5A and C7 *wt1* mutants to display an augmented distribution of WT1 cellular presence compared to control U251 and LN18 lines. While U251 and LN18 exhibited uniform nuclear and cellular distribution of WT1, the observed clusters of red aggregates pooling at nuclear periphery and exhibiting nuclear exclusion may not necessarily result due to *wt1* zinc finger mutation. The WT1 primary antibody (GTX) utilised for immunofluorescent interrogation recognises a 139 amino acid epitope corresponding to residues 359 – 497 of the C terminal zinc finger domain encoded by exons 7 to 10.C5A and C7 *wt1* mutants (based on Sanger sequencing results and *in silico* predictions) possess truncated zinc finger domains in which partial deletion of zinc finger 2 and full loss of zinc fingers 3 and 4 have occurred, accounting for approximately one third (37.4%) of the GTX antibody epitope structure. Therefore partial

epitope deletion may significantly reduce or altogether abrogate antibody-epitope binding dynamics and result in reduce specificity of WT1 recognition in these mutants. However, partial epitope deletion is insufficient to explain the observation of aggregates localising at nuclei periphery as DNA binding defunct *wt1* proteins. Determination of altered specificity of the GTX primary antibody would require interrogation of WT1 cellular distribution utilising a WT1 N terminal primary antibody specific to the non-CRISPR-targeted region in C5A and C7 mutants relative to their respective U251 and LN18 controls. Observing no significant difference in both magnitude and distribution of the WT1 signal intensity between control and *wt1* mutant cells using such an antibody would confirm a non-specific and altered affinity of the C terminal GTX antibody however, a reproducible result consistent with both WT1 N and C terminal primary antibodies would lend credence to the notion of WT1 DNA binding domain truncation in C5A and C7 *wt1* mutants. Ultimately, verification of said truncation would require WT1 proteomic extraction, purification and species identification, possibly via liquid chromatography coupled mass-spectrometry.

Concordant with staining results, transcriptional analysis via qPCR of *WT1* expression in which overall down-regulation of transcription was detected in C5A and C7 mutants relative to their controls. Curiously, the opposite profiles for exon 1 and exons 8-10 demonstrated by C5A and C7 is problematic to delineate owing to large variability between biological repeats. However, the consistency between repeats of exon 1

expression in C7 and exons 8-10 in C5A may be reflective of successful *wt1* null mutation and therefore perturbed capacity to self-regulate its transcription through a positive feedback loop. Although CRISPR-cas9 mediated targeting of exons 8 and 9 was successful through stable incorporation of the EF-1a promoter-eGFP-IRES-blasticidin selection cassette, mutation of the non-targeted allele produced a stop codon predicting to result in premature translational termination and hence does not affect transcription. Consequently, qPCR primers still detect exons 8-10 within cDNA templates albeit at significantly lower levels than controls. Additionally, random hexamers, which were utilised for random priming of RNA transcripts during reverse transcription, cannot differentiate between poly-adenylated and non-poly-adenylated transcripts, suggesting that templates destined for degradation may have been amplified and erroneously contributed to inflated transcriptional levels.
The significantly altered 5caC distribution in *wt1* mutants exposes an important caveat for consideration in the interpretation of these and other downstream functional studies performed on these engineered cell lines.

The selection of CRISPR cas9 edited *wt1* mutants was contingent on the purification of clonal lines originating from one parent ancestor cell and cultured to achieve clonal density. Considering the heterogeneous nature of U251 and LN18 cell lines from which C5A and C7 were respectively derived, inferences regarding 5caC distribution and gene expression analysis in mutants may be unrepresentative of general trends occurring in parent cell lines. Circumvention of this limitation may

be achieved through clonal selection of U251 and LN18 parent cell lines into multiple distinct subcultures via limiting dilution assay or flow cytometry. Clonal *wt1* mutants could then by assayed against distinct clonal U251 and LN18 lines.

Next Generation Sequencing of 5caC-DIP data demonstrated the dramatic depletion of 5caC peaks present in U251 and LN18 GBM cell lines relative to RBLPAT hiPSCs. Cell lines RBLPAT, U251 and LN18 were subjected to the same protocols of DNA extraction and sonication with the equal concentrations of DNA per sample exposed to the same standard 5caC antibody at titres optimised for each cell line during immunoprecipitation.

The vast discrepancy in 5caC peaks between GBM cell lines and RBLPAT was unexpected possibly due to the exaggerated perspective of 5caC magnitude and nuclear distribution projected by immunofluorescence confocal microscopy. As this technique employs HRP-conjugated secondary antibodies and subsequent enzymatic-based signal amplification, an exponential increase in 5caC signal magnitude is produced which is necessary to generate 5caC signals detectable with confocal microscopes. Though serviceable as a tool for the semi-quantitative assessment of oximC levels in a comparative manner, immunofluorescent microscopy may be discordant with representing 5caC genomic distribution. This is further exemplified by the preponderance of 5caC consensus peaks present at alpha-satellite centromeric repeats concurrent with the depletion of 5caC peaks

throughout the genome, a phenomenon which should be observable microscopically.

Conscious of the 5caC disparity observed in different cell lines, the 5caC DIP-Seq protocol was optimised using the same standard 5caC primary antibody for all cell lines at concentrations sufficient to generate equal 5caC DIP-Seq DNA library concentrations for sequencing across all cell lines. Additionally, the sequencing of the “no antibody” and “IgG isotype” controls did not produce significant sequencing reads which verified the presence of DNA fragments as a result of 5caC precipitation. Therefore, the distinct 5caC peaks detected in LN18, U251 and RBLPAT enriched over their respective inputs coupled with the anti-correlation of 5caC GBM repetitive sequence peaks observable between U251 and RBLPAT suggests the presence of legitimately annotated peaks between the cell lines. Within cell lines LN18 and RBLPAT, many 5caC annotated consensus peaks did possess identical enrichment in input libraries as in the experimental sample libraries and were therefore disregarded from our analysis however these confounding factors were not observed in U251 samples at the same genomic locations. Similarly to 5caC peak distributions, WT1 ChIP peaks in RBLPAT did not align to 5caC consensus peaks in RBLPAT despite their dramatic coverage within the genome and proximity to each other. A subset of 5caC and WT1 peaks exhibited alignment in U251 and inspection of map quality of sequence reads fell within acceptable quality control ranges indicating accurate peak calling and mapping.

WT1 transcripts expression were normalised by fragments per kilobase million (FPKM) in which forward and reverse reads of 150 bases per read were utilised to sequence RNA transcripts. The transcriptome analysis misleading demonstrated no significant fold-change occurring between *WT1* transcript expressions however Ballgown assembly plots revealed the existence and high expression of a novel WT1 isoform, an observation not possible to detect by heatmap clustering alone. The expression of the novel transcript in only 2 of 3 biological replicates is perplexing as all replicates were treated identically, harvested at the same time and sequencing libraries prepared in parallel. Unfortunately, although Ballgown assembly analysis does indicate the high expression a truncated transcript present in 2 out of 3 C5A mutant repeats, elucidation of which exons are present in the novel isoform is indiscernible. Additionally, heatmap analysis of U251 *WT1* gene expression in neurobasal versus serum conditions and subsequently control U251 versus C5A mutants in neurobasal conditions suggest that although *WT1* gene expression was unaffected under serum to neurobasal conditions, *WT1* mutagenesis resulted in its transcriptional up-regulation. This observation is unreflective of the novel truncated anomalous transcript presence and hence transcriptional analysis insufficiently describes *wt1* mutagenesis in C5A cell lines.

Examination of DNA demethylation dynamics in U251 vs C5A cell lines under neurobasal conditions was limited to 5hmC semi-quantitative

interrogation of magnitude and distribution as technical issues with the Tyramide Signal Amplification system buffer were experienced, rendering 5caC signal fluorescence as non-specific and therefore unusable.

Functional assessment of the effects *wt1* zinc finger mutagenesis in C5A at the transcriptional level by qPCR were limited by the variation between biological replicates which was also corroborated by RNA-Seq data, confirming the ambiguity of overall transcriptional fold-change for a subset of genes. Unfortunately, owing to time constraints, functional analysis of gene expression of C5A and C7 *wt1* mutants relative to their controls during neuronal and glial differentiation was performed in biological triplicate, however only one repeat was analysed. Therefore accurate statistical analysis could not be calculated. As previously mentioned with regards to clonal selection of CRISPR-cas9 mediated *wt1* mutants, all results generated from experimental assays conducted with C5A and C7 *wt1* mutants relative to U251 and LN18 controls bear the caveat of being unrepresentative owing to the cellular heterogeneity of control GBM cell lines. Consequently, physical differences of tumours formed in C5A injected mice relative to U251 treated ones cannot be claimed to be resultant solely on *wt1* mutagenesis. Additionally, observation of tumour size through microscopic examination of harvested brain sections does produce a multidimensional perspective of tumour size. Tumour could have been harvested and weighed from control and mutant specimens to better understand tumour growth rates.

General Discussion

6.1 Profiling OximC Dynamics in Brain Tumour Cell Lines

From our interrogation of 5hmC and 5caC levels in a range of cell lines and tissues by both immunofluorescence coupled confocal microscopy and mass spectrometry, a spectrum of oximC magnitudes and distributions is apparent with healthy adult somatic cells at one end, embryonic cells at the other and adult and paediatric brain neoplasms and foetal brain cells populating the middle. This pattern of distribution is somewhat expected as mouse E3.5 and human E6.0 embryonic stem cells possess the highest global 5caC levels (Smith *et al*, 2012) at which point subsequently undergo re-methylation following epiblast formation and uterine wall implantation at which a somatic ground state of methylation is established (Smith *et al*, 2012; Hackett & Surani, 2013; Seisenberger *et al*, 2013). 5caC was detectable in both paediatric (UW228-3, DAOY, BXD-1425EPN and DKFZ-EP1NS) and adult brain tumour cell lines (LN18, LN229, U87MG and U251). This result was

unexpected for the following reasons (1) despite pluripotent hESCs exhibiting DNA methylome profiles of global hypomethylation, and possessing the highest global 5caC magnitudes in the context of human development, 5caC magnitudes are depleted relative to genomic cytosine levels (Ito *et al*, 2011, Smith *et al*, 2012), (2) the preponderance of 5hmC and absence of 5caC in adult mammalian brain tissue previously reported by other groups (Kriaucionis & Heintz, 2009; Globisch *et al*, 2010; Jin *et al*, 2011; Nagae *et al*, 2011 Khare *et al*, 2012; Kraus *et al*, 2012) and (3) observations of significant reduction in 5hmC magnitudes in malignant tissues relative to their healthy matches (Jin *et al*, 2011; Yang *et al*. 2013; Gackowski *et al*, 2016).

Interestingly, using highly sensitive and direct oximC quantification via two-dimensional ultra-performance chromatography coupled tandem mass spectrometry (2D-UPLC-MS/MS), Gackowski and colleagues detected 5caC at non-negligible levels in primary colorectal carcinoma tumours which exhibited a reduced magnitude compared to their normal tumour margin controls (Gackowski *et al*, 2016). The higher sensitivity of 2D-UPLC-MS/MS facilitates direct quantification of 5caC molecules without the requirement of chemical derivitisation methods common to traditional HPLC-MS/MS or gas chromatography coupled tandem mass spectrometry (GC-MS/MS) which therefore enables detection of oximCs in samples of limited quantity (Gackowski *et al*, 2016). The superiority in sensitivity of 2D-UPLC-MS/MS suggests that 5caC and 5fC modifications may actually be present in tissues and cell types previously shown to be

negative for these epigenetic marks by conventional detection methods. Curiously, our analyses employed the same 2D-UPLC-MS/MS interrogation methods which revealed comparable 5hmC and 5fC magnitudes between adult glioblastoma LN18, paediatric medulloblastoma UW228-3 and colorectal cancer HCT_116 cell lines however 5caC was undetected in the latter. Considering the differences observed between Gackowski *et al* and our data regarding 5caC detectability in colorectal cancer tissue and cell lines, this may be reflective of a non-representative epigenetic profile between primary malignant colorectal material and cultured colorectal cancer cell lines (Gackowski *et al*, 2016), a phenomenon not inclusive to brain tumour derived cells and tissue (Eleftheriou *et al*, 2015; Abakir, Wheldon & Ruzov, 2016).

Transcriptionally, elevated levels of *TET2* and *TET3* expression concurrent with down-regulated *DNMT1*, *DNMT3a* and *DNMT3b* in adult GBM cell lines relative to healthy adult fibroblasts may contribute to 5caC persistence in these cells .This seems counterintuitive considering DNMT mediated generation of the 5mC substrate is required for oximC production (Tahiliani *et al*, 2009; Shen *et al*, 2013) however, a precedent is set for this as a transcriptional shift involving *DNMT* down-regulation and *TET* gene expression elevation occurs as a result of fibroblast reprogramming to iPSCs (Smith *et al*, 2012; Gao *et al*, 2013; He *et al*, 2019). The adult somatic GF06 fibroblasts were a primary cell line derived from patient breast tissue and due to their early passage level

(less than 10 passages) may have exhibited abnormal transcriptional and methylation profiles owing to their transition to *in vitro* culture conditions, rendering them an in appropriate 5caC negative control (Nestor *et al*, 2012). GBM cell line profiles were distinct from RBLPAT iPSC lines transcriptionally, by oximC immunostaining and mass spectrometric analysis. GBM cell lines LN18, LN229, U87MG and U251 exhibited 5hmC and 5caC signal intensities comparable to those of healthy foetal neural stem cell line FB83 but transcriptional levels of DNA demethylation machinery genes did not correlate with confocal microscopy analysis results. Paediatric brain tumour cell lines UW228-3 and DKFZ-EP1NS exhibited a TET-dependent mechanism of 5caC generation, with 5hmC and 5caC reduction evident on *TET1* knockdown.

Our comparisons of paediatric and adult brain tumour cell lines relative to HeLa cervical cancer line showed greater similarity in transcriptional profiles of DNA demethylation genes than when compared to RBLPAT iPSC or GF06 adult fibroblasts however no consistency between transcription profiles, oximC immunostaining intensity or mass spectrometry was observed. This discordance between transcriptional activity, immunostaining and mass spectrometric quantification in the characterisation of oximC levels has been experienced by other groups (Lorsbach *et al*, 2003; Nestor *et al*, 2012).

Interestingly, oximC staining intensity disparities were observed within paediatric brain tumour, adult GBM and foetal neural stem cell lines to

the extent that 5caC signal intensities appeared heterogeneous within the populations. Further examination of oximC profiles revealed that this "pseudo-heterogeneity" may be attributable to non-Gaussian distributions of oximC signal intensities within cells. Indeed, discovery of this phenomenon provided the impetus of questioning whether intrinsic factors e.g. stemness capacity and proliferative status or extrinsic factors such as culture conditions may influence oximC profiles within individual cells and the cellular population *en mass*. Through implementation of neurobasal media for neural stemness enrichment and neuronal and glial lineage differentiation programmes, oximC magnitudes and distributions both within populations and at the intracellular level are susceptible to these extrinsic cues. This work highlighted a dynamic and modular pattern of oximC presence which was consistent with trends reported previously in different systems (Wheldon *et al*, 2014; Lewis *et al*, 2017).

Our Next Generation Sequencing profiling of 5caC genomic presence reflected the scale of global 5caC occupancy observable in RBLPAT iPSCs compared with LN18 and U251 GBM cell lines, commensurate with oximC quantifications detected by mass spectrometry and immunostaining. We discovered a preponderance of 5caC peaks clustered at repetitive sequences of acrocentric, telocentric and centromeric alpha-satellites, SINE and simple repeats in GBM cell lines. These results display concordance with findings reported by Shen and colleagues who discovered 5fC and 5caC over-representation at major satellite repeat sequences in mouse ESCs (Shen *et al*, 2013). In contrast,

they found 5hmC predominately occurring at SINE and LTR sequences and *TDG* shRNA mediated knock down resulted in increased genomic occupancy of 5fC and 5caC at non-repetitive sequences in ESCs (Shen *et al*, 2013). Similar results were achieved by other groups (Raiber *et al*, 2012; Lu *et al* 2013, Song *et al*, 2013; Lu *et al*, 2015) with Lu and colleagues in particular detecting 5caC modifications at promoter regions at single base resolutions in mouse ESCs (Lu *et al*, 2015).

Interestingly, these independent groups all employed a DNA immunoprecipitation-based methodology of 5fC and 5caC detection and isolation, validated by dot blot assays and immunofluorescent analysis upon TDG knockdown, thus lending credence to the sensitivity and specificity of this physical oximC detection method (Raiber *et al*, 2012; Lu *et al* 2013, Song *et al*, 2013; Lu *et al*, 2015). The over-representation of 5caC within gene bodies of protein coding genes and at gene regulatory features including open chromatin, CTCF and transcription factor binding sites and intron/exon boundaries not only suggests a transcriptional elongation role but may also indicate an involvement in splicing (Khare *et al*, 2012; Shen *et al*, 2013) possibly through 5caC-protein interaction mechanism.

6.21 Caveats and Limitations

The difficulty of contextualising oximC presence in our brain tumour cell lines against the backdrop of previously published literature lies with the selection of appropriate control samples. The vast majority of oximC genomic profile characterisations have been performed using mouse and

human embryonic stem cells owing to observed 5hmC and 5caC enrichment during early development (Tahiliani *et al*, 2009; Ito *et al*, 2010; Cimmino *et al*, 2011; He *et al*, 2011; Inoue *et al*, 2011; Ito *et al*, 2011; Le *et a*l, 2011; Pastor *et al*, 2011; Wu *et al*, 2011; Xu *et al*, 2011 Raiber *et al*, 2012; Song *et al*, 2013; Sun *et al*, 2013; Booth *et al*, 2014; Bachman *et al*, 2015). While investigations exploring oximC magnitudes in cancers have been conducted employing similar experimental techniques to the ones used in our work, (Ko *et al*, 2010; Haffner *et al*, 2011; Jin *et al*, 2011; Yang *et al*, 2013; Shim *et al*, 2014; Park *et al*, 2015; Strand *et al*, 2015; Uribe-Lewis *et al*, 2015; Huang *et al*, 2016; Bhattacharyya *et al*, 2017; Storebjerg *et al*, 2018; Wilkins *et al*, 2018; Bergamaschi *et al*, 2019; Carrella *et al*, 2019; Kristensen *et al*, 2019; Sudhir *et al*, 2019), they utilise normal healthy tissue match samples as controls for characterising the oximC profiles observable in malignant neoplasms.

Considering the objective of this work is to elucidate the molecular mechanisms underpinning DNA demethylation, specifically with respect to potential protein "oximC reader" associations and not necessarily from a "disease versus healthy state" standpoint, healthy tissue match samples did not, in our opinion, constitute the appropriate controls nor were they easy to access. 5caC enrichment was initially discovered in glioma tissue samples (Eleftheriou *et al*, 2015) and we subsequently detected it in commercial GBM and paediatric brain tumour cell lines. Through empirical interrogation, we observed 22 post-coital week foetal

brain-derived neural stem cells to represent an appropriate control owing to the presence of 5caC in both the parental foetal brain tissue and its subcultured cell line, FB83. Consequently, oximC patterns of magnitude and distribution do not faithfully represent general trends observable in the current literature.

Nevertheless, rigorous internal controls and quality assurance have been implemented wherever possible. These include but are not limited to *in silico* screening and empirical determination of appropriate qPCR primer anneal temperatures, ensuring consistency of RNA quantities across samples for cDNA synthesis, implementation of cDNA synthesis controls, consistency of confocal microscopy settings to ensure samples are interrogated under equal conditions, standardisation of image analysis fields of view and representative cell numbers, performing all experiments in biological triplicate where possible and feasible and performing quality control checks of DNA and RNA integrity, sample quality, quantity and peak distribution appropriate for each step of NGS DIP, ChIP and RNA-Seq library preparation.

Although multiple studies detail the absence of 5caC enrichment in mammalian brain tissue despite the abundance of 5hmC observed, particularly in cerebellum (Kriaucionis *et al*, 2009; Guo *et al*, 2011; Jin *et al*, 2011a; Szulwach *et al*, 2011; Khare *et al*, 2012; Kraus *et al*, 2012; Hahn *et al*, 2013; Bachman *et al*, 2015; Abakir, Wheldon & Ruzov, 2016), the detection of 5caC by immunohistochemical methods (Condliffe *et al*,

2014; Jarmasz *et al*, 2019) and LC-MS/MS quantification (Zhang *et al*, 2016; Gao *et al*, 2017) has also been reported. Both Zhang *et al* and Gao *et al* groups utilised LC-MS/MS technologies for the detection of 5caC and implemented synthetically generated 5caC standards to identify and quantify the oximC modification in different tissues. Both groups utilised ammonia-based mobile phases optimised to a concentration of 0.1% to achieve the sensitivity required to separate and subsequently detect greater 5caC magnitudes (Zhang *et al*, 2016; Gao *et al*, 2017) than had previously been reported by Globisch, Pfaffeneder, and He (Globisch *et al*, 2010; He *et al*, 2011; Pfaffeneder *et al*, 2011) who used HPLC-MS and TLC-MS technologies.

Although studies conducted by Zhang and Gao groups contrast with the plethora of previously reported literature, their work does not conflict with our findings as their mass spectrometric analysis was performed on rat and mice brain tissue, neither of which were interrogated by mass spectrometry in this body of work. Additionally, although Condliffe *et al* and Jarmasz *et al* detected 5caC via immunostaining methods, the former analysed human adult healthy control and Alzheimer's disease brain tissue sections while the latter interrogated human, porcine and murine brain tissues post-mortem (Condliffe *et al*, 2014; Jarmasz *et al*, 2019). Considering these studies were performed on tissue sections, they are not directly comparable to malignant cell lines utilised in this thesis owing to transcriptional and epigenetic disparities observable in tissues upon *in vitro* cell culture conditions (Nestor *et al*, 2012).

Although switching FB83 and GBM cell line culture conditions from undefined serum to neurobasal media demonstrated the modularity of 5hmC and 5caC patterns present within individual cells and their wider population, whether the oximC profile augmentations observable were attributable to a neural stem cell identity was unconfirmed.

Albeit controversial (owing to the ambiguity of definitive neural stem cell markers and Hoechst 33342 dye effluxing potential), conventional methods of neural / glioma stem cell procurement entail side population isolation and neurobasal media facilitated enrichment of neurosphere generating cells (Kim & Morshead, 2003; Shimoda, Ota & Okada, 2017). However, despite satisfactory compliance with these parameters non-stem-like cells may be present within the side population derived neurospheres (Broadley *et al*, 2011; Golebiewska *et al*, 2011). Hence assays for clonogenity, clonal neurosphere derivation and multi-lineage (astrocyte, neuron and oligodendrocyte) differentiation subsequent to side population isolation and complementary to NESTIN, CD133, CD15 and SOX2 positive staining (Broadley *et al*, 2011; Golebiewska *et al*, 2011; Golebiewska *et al*, 2013; Zhang *et al,* 2013) would need to be performed prior to examination 5hmC and 5caC profiles. Therefore exposure of GBM and FB83 cell lines was primarily performed to modulate cellular transcriptional profiles as a means of investigating the static or dynamic nature of oximC genomic presence.

Similarly, although 5hmC and 5caC magnitudes and distributions were detectable in serum and neurobasal media (the disparity between which was semi-quantitatively measurable), discernment of whether 5caC enrichment was produced aetiologically or consequentially of extrinsic transcriptional modulation is unconfirmed. Concomitantly, 5caC genomic occupancy cannot be correlated with gene expression profiles as only 5caC DIP-Seq in serum conditions was performed.

Delineation of 5caC involvement in transcriptional regulation would require a distinction between two different datasets, one in which 5caC enrichment was observable at gene specific loci whereas the other might exhibit reduced or depleted 5caC levels such that differential 5caC peaks could be identified. Endeavours towards this end could be achieved by 5caC-DIP-Seq on U251 in both neurobasal and serum conditions and RNA-Seq transcriptome analysis to reveal 5caC genomic occupancy between the two states and correlation with gene expression. Within the same culture conditions, 5caC contribution towards transcriptional regulation could be investigated through locus specific TDG-mediated targeting.TALENS-TET1 and dCas9-TET1 protein fusion systems have been employed to demethylate site-specific hypermethylated loci (Gallego-Bartoleme *et al*, 2018; Marx *et al*, 2018; Morita *et al*, 2018; Ou *et al*, 2018; Liu *et al*, 2019; Noack *et al*, 2019). While these methodologies can successfully demethylate locus specific regions resulting in transcriptional activation of genes e.g. *CDKN1C* encoding p57 (Ou *et al*, 2018), TET1 oxidises 5mC modifications to higher oxidised

forms, most stably 5hmC (Kriaucionis *et al*, 2009; Tahiliani *et al*, 2009; Ito *et al*, 2011) but may not efficiently and stably catalyse to 5fC and 5caC forms in control healthy somatic cells such as astrocytes which would be a fitting comparison against GBM cell lines. Therefore, hypothetically, a dCas9-TDG catalytic domain fusion protein could be generated to perform 5caC specific targeted removal at 5caC enriched sites elucidated from our 5caC-DIP Seq peaks and RNA-Seq could be performed to interrogate any 5caC-depletion-associated transcriptional changes.

The lack of base resolution sequencing following 5caC-DIP limits the precise characterisation of 5caC genomic occupancy such that CpG or non-CpG demethylation cannot be discerned. Base resolution Methylation-Assisted Bisulphite Sequencing (MABSeq) could be performed on U251 and LN18 cell lines to overcome this.

.

6.3 Functional Impact of WT1 on DNA Demethylation Dynamics

To investigate whether the 5caC enrichment observable in GBM cell lines LN18, LN229, U251 and U87MG, was related to the over expression of *WT1*, we implemented *WT1* siRNA knock down, WT1 zinc finger DNA binding domain over-expression, design and constructed WT1 zinc finger null mutant cell lines derived from U251 and LN18 and performed WT1 ChIP-Seq and 5caC DIP-Seq to determine WT1 and 5caC genomic

occupancy. While siRNA mediated *WT1* knockdown was only performed in U87MG cells, the resulting trend exhibited up-regulation of *TET1* and *TET2* and down-regulation of TDG. This trend was consistently observed with LN18 derived C7 and U251 derived C5A *wt1* zinc finger mutants and the opposite was *WT1* zinc finger domain over-expression. While *TET1* was down regulated in C5A mutants when analysed by RNA-Seq profiling, this examination was of cells cultured in neurobasal conditions which initially elevated *TET1* expression when switching from serum conditions however *TET1* levels significantly dropped in C5A cell lines, indicating differential expression correlating to *wt1* mutagenesis. Interestingly, immunofluorescence examination of wt1 expression in control U251 and LN18 vs U251 and C7 mutants depicted pools of red wt1 signal localised at the nuclear periphery. Considering C5A and C7 mutants were generated through CRISPR cas9 mediated mutagenesis aimed at disrupting the WT1 DNA-binding zinc finger domain, C5A and C7 mutants may be exhibiting truncated wt1 proteins. This would agree with RNA-Seq Ball gown assembly plots illustrating the emergence of a novel *wt1* isoform present in C5A mutant but not control cell lines. This mutant isoform, unable to bind DNA may accumulate outside the nucleus, possibly designated for proteolysis within lysosomes (El Ghouzzi *et al*, 2000; Grigorov *et al*, 2007).

To confirm this, C5A and C7 mutants vs control cells immunofluorescence could be performed with anti-Lamp3 antibodies demarcating lysomes (Grigorov *et al*, 2007). Western blots for WT1 using an N terminal antibody could be performed on nuclear vs cytoplasmic

fractions to determine CRISPR-Cas9 mediated mutagenesis has produced a truncated protein resulting in altered sub-cellular localisation between mutant and control cells. Such a methodology and outcome was utilised and able to discern the irregular speckled sub-nuclear distribution of wt1 zinc finger truncation mutants (Englert *et al*, 1995).

Interestingly, -KTS and +KTS have been shown to localise to distinct nuclear compartments with the –KTS isoform displaying a diffuse nuclear distribution similar to that of specificity protein 1 (SP1) or transcription factor II D (TFIID) but the +KTS isoform localised in nuclear compartments bound to nuclear RNA splicing machinery (Larsson *et al*, 1995). These findings suggested WT1 localisation may be splice variant dependent however, Englert *et al* truncated zinc finger mutants produced a speckled pattern, localised within the nucleus (Englert *et al*, 1995). This may be indicative of an N terminal nuclear localisation signal present and unperturbed by the DNA binding domain truncations (Englert *et al*, 1995; Larsson *et al*, 1995).

Unexpectedly, the fusion of zinc finger domains to a β-galactosidase reporter which could not localise to the nucleus in COS-7 monkey cells, exhibited nuclear enrichment when WT1 zinc fingers 1, 2 and 3 but not 4 were utilised in the fusion reporter (Bruening *et al*, 1996). This indicated the presence of DNA binding domain nuclear designation signals within zinc fingers 1 and 2/3 however the nuclear localisation ability of N terminal domains fused to β-galactosidase reporter was not assessed (Bruening *et al*, 1996). Therefore, while N terminal and C terminal nuclear

localisation signals may be present in our C5A and C7 wt1 zinc finger mutants, we cannot accurately discern the WT1 distribution without N terminal WT1 primary antibody interrogation followed by sub nuclear and cytoplasmic fractionation and western blot and/ or mass spectrometry.

Strikingly, the 5hmC and 5caC cellular distribution patterns detected following immunofluorescence staining in C5A and C7 *wt1* mutants relative to their controls resembled the sporadic 5hmC and 5caC pattern of LN229 in which WT1 was found to not be expressed at the transcriptional or protein level. This suggests that 5caC may interact with and be stabilised by WT1 when expressed however, 5caC enrichment is independent of it. Analysis of WT1 ChIP-Seq and 5caC DIP-Seq peak distributions support this postulation as unique WT1 and 5caC peaks were observed as well as loci at which WT1 and 5caC peaks aligned. Moreover, positional weight matrices significantly predicted the presence of WT1 cognate DNA binding sites within the sequence encapsulated by aligned WT1 and 5caC narrow peaks, implicating WT1 and 5caC genomic co-localisation. Importantly, considering the nature by which C5A and C7 *wt1* mutant cell lines were screened and selected, i.e. eGFP FACS sorted at one cell per well and expanded to achieve clonal density, the seemingly sporadic 5hmC and 5caC staining distribution intimates a modular and transitory paradigm of 5caC genomic occupancy as both mutant cell lines are genotypically homogeneous.

The unexpected phenotype of perturbed neurosphere formation capacity of the C5A and C7 *wt1* mutants relative to their controls may be reflective of diminished extracellular matrix gene expression required for neurosphere formation, as illustrated by the transcriptomic data. *WT1* displayed transcriptional up-regulation (detected by qPCR) during GBM cell neuronal and glial differentiation most notably on days 1 and 3 which tallied chronologically with highest 5caC enrichment. Mutant C5A and C7 cell lines displayed transcriptional perturbation during differentiation most notably in LN18 however biological repeats would be required to confirm these findings.

Though no significant changes in growth kinetics between *wt1* mutants and control cells cultured in serum conditions were observed, phenotypic differences between U251 and LN18 controls and their C5A and C7 mutants appear to manifest in neurobasal conditions. Neurobasal conditions have been demonstrated to recapitulate primary tumour physiology more faithfully than serum conditions (Lee *et al*, 2006). Commensurate with this reasoning were the observations of small and absent tumours with significantly reduced proliferative levels reported in *in vivo* mouse xenografts of mutant cell injections relative to U251 controls. While the smaller tumour size observed may be attributable to the reduced proliferation rate observed in mutants compared to controls, the decreased extracellular matrix gene expression in C5A mutants relative to U251 discovered in neurobasal conditions may also have

influenced tumour growth and engraftment (Penuelas *et al*, 2009; Lathia *et al*, 2010; Nakada *et al*, 2013; Ying *et al*, 2014; Malric *et al*, 2017).

6.31 Caveats and Limitations

The mass spectrometry results of 5hmC and 5caC between C5A and C7 *wt1* mutants and their U251 and C7 controls does not correlate with the immunostaining results of 5hmC and 5caC signal intensity. This may be due to vast disparities in oximC signal profiles observable within the mutant lines. Considering that cellular material generated for mass spectrometry analysis is derived from homogenisation of cell pellets prior to genomic DNA extraction and hydrolysis to individual nucleosides, mass spectrometric data represents oximC species abundance across all cells i.e. an average and does not distinguish between individual under-represented cells.

Next Generation Sequencing ChIP-Seq and 5caC DIP-Seq data revealed an alignment of WT1 and 5caC genomic peaks which possessed similar magnitudes and peak width. The lack of 5caC DIP-Seq data on C5A mutant cells limits the scope to which analysis of a WT1-5caC interaction can examined as a reduction in 5caC peaks in C5A *wt1* DNA binding domain mutant cells might suggest a stabilisation mechanism of 5caC genomic occupancy by WT1. C5A WT1 ChIP-Seq versus U251 WT1 ChIP-Seq both using an N-terminal WT1 antibody in addition to 5caC DIP-Seq on both cell lines would be required to determine an association between WT1 and 5caC. Moreover, the current

U251 WT1 ChIP-Seq and U251 5caC DIP-Seq data was generated utilising sequencing libraries which produced approximately 40 million reads but 75 base pair pair-ended read lengths. Therefore mapped reads do not represent base-resolution accuracy of cognate WT1 and 5caC genomic binding positions which limits the precision to which WT1 and 5caC alignment and hence association predictions can be ascribed. Furthermore, considering the sporadic distribution of 5caC in C5A cells generated from a clonal lineage, 5caC genomic occupancy may be dynamic and transitory. This suggests that WT1 and 5caC, although exhibiting peak enrichment at aligned genomic coordinates, may not share temporal alignment. Synchronisation of cells in serum free media or using the G1 arrestor Lovastatin (Banfalvi, 2017) could be performed 24 hours prior to the commencement of WT1 ChIP and 5caC DIP experiments as a means of temporally coordinating cells for analysis.

6.4 Significance

It has been well documented that reductions in 5hmC presence in cells and tissues have been associated with malignancy (Ko *et al*, 2010; Haffner *et al*, 2011; Jin *et al*, 2011; Yang *et al*, 2013; Shim *et al*, 2014; Park *et al*, 2015; Strand *et al*, 2015; Uribe-Lewis *et al*, 2015; Huang *et al*, 2016; Bhattacharyya *et al*, 2017; Storebjerg *et al*, 2018; Wilkins *et al*, 2018; Bergamaschi *et al*, 2019; Carrella *et al*, 2019; Kristensen *et al*, 2019; Sudhir *et al*, 2019). In many studies, mutation of the *IDH1/2* genes, especially the arginine 132 to histidine substitution (R132H) occurring in

IDH1 have been responsible for the generation of oncometabolite D-2-hydroxyglutarate (D2HG) which has been demonstrated to catalytically attenuate TET protein activity correlating with reduced 5hmC magnitudes (Figueroa *et al*, 2010; Xu *et al*, 2011; Turcan *et al*, 2012; Chou *et al*, 2016). While *IDH* gene mutations do associate with low grade CIMP gliomas and secondary glioblastoma often exhibiting a pro-neural subtype (Parsons *et al*, 2008; Hartmann *et al*, 2009; Yan *et al*, 2009; Noushmehr *et al*, 2010; Turcan *et al*, 2012; Cohen *et al*, 2013) less than 10% of primary *de novo* glioblastoma tumours exhibit *IDH1/2* mutations yet significant 5hmC reduction is also common to *IDH1/2* wild type tumours (Balss *et al*, 2008; Yan *et al*, 2009; Capper *et al*, 2010; Jin *et al*, 2011a; Byeon *et al*, 2012; Xu *et al*, 2019).

Despite TET protein activity inhibition by elevated D-2HG concentrations (Xu *et al*, 2011), *TET* gene expression is increased in *IDH* wild type tumours relative *IDH* mutant ones, implying that transcriptional levels do not correlate with protein activity (Xu *et al*, 2019). Interestingly, in spite of significant 5hmC reduction in *IDH2* mutant gastric cancer cells, 5fC and 5caC levels (as detected by dot blot assay) were not significantly different from *IDH2* wild type controls (Chou *et al*, 2016). The persistence of 5fC and 5caC levels in contrast to 5hmC reductions may be attributable to TDG sensitivity to α-ketoglutarate as Spallotta *et al* observed in analysis of diabetic versus non-diabetic mice cardiac mesenchymal stem cells (Spallotta *et al*, 2018).

Cardiac mesenchymal stem cells isolated from non-diabetic and diabetic (generated via high fat diet or streptozotocin administration) mice displayed differential 5mC, 5hmC and 5fC levels with all modifications elevated in diabetic mice (Spallotta *et al*, 2018). Intriguingly, not only were α-ketoglutarate levels reduced in diabetic mice relative to non-diabetic controls which correlated with reduced glucose metabolism gene expression demonstrated by RNA-Seq transcriptomics, TET1 exhibited nuclear exclusion in diabetic mice derived cardiac mesenchymal stem cells but not in non-diabetic controls (Spallotta *et al*, 2018).

Contextualising our data with the current literature

Remarkably, molecular modelling identified an Arginine 275 (Arg275) residue present within a RXXXXXR α-ketoglutarate recognition sequence analogous to ALKBH5 (Fu *et al*, 2014) as the discriminatory sensor of α-ketoglutarate within TDG, verified by CRISPR Cas9 medated mutagenesis (Spallotta *et al*, 2018). *IDH1/2* gene mutations can simulate the effects of hypoxia on *TET* genes, a feature common to tumours (Thienpont *et al*, 2016; Liu *et al*, 2019). Considering our GBM cell lines are *IDH1/2* wildtype (Dang *et al*, 2009; Shi *et al*, 2015; Cui *et al*, 2016; Birner *et al*, 2014) but do reflect hypoxic effects typified by *HIF1a* expression (Minet *et al*, 1999; Lu *et al*, 2005; Gariboldi *et al*, 2010; Hong *et al*, 2013), the findings by Spallotta *et al* may reflect an alternate, perhaps glucose metabolism dependent pathway influencing HIF-1a activation and 5caC enrichment in GBM cell lines (Semenza , 2000; Semenza, 2002). However, the elevated *TET* genes expression and

reduced *TDG* gene expression relative to HeLA and fibroblasts may not directly correspond to protein activity.

From analysis of our 5caC DIP-Seq data we generated a pattern of 5caC genomic occupancy which showed greatest enrichment within the gene bodies of protein coding genes, transcription factor binding sites, CTCF binding sites, and satellite repetitive sequences surround centromeres, telomeres and SINE/*Alu* elements in U251 and LN18 GBM cell lines. DIP-Seq for 5mC, 5hmC, 5fC and 5caC performed on foetal brain derived neural stem cells versus patient derived xenograft (PDX) glioblastoma stem cells (GSC) revealed global reductions in 5mC and 5hmC magnitudes in GSCs relative to NSCs, specifically occurring at within gene bodies, CGIs and promoters whereas 5fC and 5caC magnitudes were greater in GSCs compared to NSCs however the accumulated modifications increased evenly across the genome, not localised to specific features (Zhou *et al*, 2018).

However, 5mC, 5fC and 5caC modifications which were enriched at CpG 'open seas' in NSCs depreciated in magnitude and were elevated in magnitude at CGIs and transcriptional start sites in GSCs. Likewise, 5hmC enriched at CpG shores in NSCs accumulated at CGIs and transcriptional start sites in GSCs (Zhou *et al*, 2018). These results were validated by -reduced representation bisulphite sequencing (RRBS) for 5mC and Methylase-assisted sequencing (MAB-Seq) for 5hmC and 5fC/5caC, the latter of which could not be individually discriminated,

hence the utilisation of 5mC, 5hmC, 5fC and 5caC antibodies which added greater specificity between modifications (Zhou *et al*, 2018).

Analysis of site-specific DNA demethylation dynamics between the foetal neural stem cells and PDX derived GSCs revealed substantial loss in 5mC, 5hmC and 5fC/5caC being replaced by unmodified cytosine, specifically 2190 instances of 5mC loss and 884 instances of 5hmC/5fC/5caC loss (Zhou *et al*, 2018). Additionally, 384 5hmC loci in NSCs were replaced by 5mC in GSCs and 1392 to 5fC/5caC. While relatively few 5mC NSC sites were replace with 5hmC (651), many 5mC loci were oxidised to 5fC/5caC (3962) in GSCs. Overall this describes a theme of global 5mC and 5hmC reduction in GSCs relative to NSCs concomitant with accumulation of higher oxidised forms in GSCs suggestive of dramatic methylome reprogramming occurring during tumorigenesis which may significantly influence malignant cell transcriptomes (Zhou *et al*, 2018).

These findings reported by Zhou *et al* produce a more defined and robust analysis of the effect of oximC presence on transcriptional levels owing to their co-ordination of MAB-Seq, RRBS, and 5mC/5hmC/5fC/5caC DIP-Seq with RNA-Seq experiments between both NSCs and PDX-GSCs (Zhou *et al*, 2018). Unfortunately owing to budget constraints and sample limitations, multivariate analysis between multiple cell lines was unfeasible. However, despite their multi-modal analysis, the implementation of MAB-Seq and TAB-Seq methods in conjunction with

DIP-Seq assays still cannot elucidate the locations of 5fC and 5caC modifications at base resolution. Therefore oximC presence can only correlate with RNA-Seq generate transcriptome data, not provide direct functional evidence (Zhou *et al*, 2018). Interestingly, positional weight matrices calculated transcription factor binding motifs to align with 5fC/5caC enriched CpG sites which predicted Aryl hydrocarbon receptor (AHR) and E2F transcription factor 2 (E2F2) as 5fC and 5caC readers (Zhou *et al*, 2018). While 5mC and 5hmC were enriched at poised enhancers marked by H3K4me1, 5fC and 5caC showed 50% greater incidence than 5mC or 5hmC at active enhancers designated by H3K27ac histone marks suggesting 5fC and 5caC are positively correlated with transcriptional activation (Zhou *et al*, 2018).

Although our data lacks verification using chemical conversion or enzymatic based sequencing methods such as CAB-Seq or caMAB-Seq, our DIP Seq produced similar patterns of 5caC enrichment specifically at transcription factor binding sites, within protein-coding gene bodies and gene regulatory regions (Zhou *et al*, 2018).

Our observation of 5caC preponderance at repetitive satellite sequences in GBM cell lines was concordant with reports by Lu *et al*, who described 5fC and 5caC DNA occupancy at major satellite repeats in TDG depleted mouse embryonic stem cells using a more sophisticated DIP-CAB-Seq method than our 5caC-DIP-Seq method alone (Lu *et al*, 2015). Ergo, Lu *et al* generated base-resolution mapping of 5fC and 5caC detailing that

although both these oximC marks display similar genomic distributions predominating at intergenic, intron and exon sequencers, they both occupy unique CpG sites and non-CpG sites, as reflected in their major satellite distribution (Lu *et al*, 2015). Furthermore, 5fC and 5caC enrichment clustered around transcriptional start sites, aligned with positional weight matrix predicted pluripotency network transcription factor binding sites and co-localised with H3K27ac ChIP peaks (Lu *et al*, 2015). This was a more in-depth analysis than their previous study in which base-resolution and block-profiling of chemically labelled 5fC delineated its localisation to H3K4me1 associated poised enhancers (Song *et al*, 2013).

In mouse embryonic fibroblasts subjected to shRNA mediated *TDG* knockdown, genomic profiling of 5mC, 5hmC, 5fC and 5caC modifications within repetitive sequences using DIP-Seq methods demonstrated that 5mC modifications distributions were predominately found on SINE elements and fewer but approximately equal representation at LINES, LTRs, simple repeats and non-repetitive sequences (Papin *et al*, 2017). OximCs i.e 5hmC, 5fC and 5caC in contrast were greatly represented within simple repeat sequences and LINEs but vastly under-represented at SINEs, LTRs and other repeat sequences (Papin *et al*, 2017). When comparing mouse embryonic fibroblasts (MEFS) to mouse embryonic stem cells (mESCs), 5hmC, 5fC and 5caC were enriched in GC-rich regulatory sequences of LTR including interstitial activated particles (IAPs) in mESCs but not detected

in MEFs, instead replaced by 5mC (Papin *et al*, 2017). In contrast to LTRs, oximCs, particularly 5hmC and 5caC displayed significantly higher enrichment in MEFs as oppose to mESCs. Additionally, all oximCs displayed 50 fold greater enrichment at CpA simple repeats within gene coding regions than at CpG sites at coding regions and significantly greater abundance than CpA methylation (Papin *et al*, 2017).

These findings may be indicative of a DNA demethylation mechanisms active at repetitive elements that are initially active during pluripotent transcriptional profiles that subsequently require silencing of potentially deleterious retro-transposon mutagenic elements as development ensues (Deniz, Frost & Branco, 2019).

In summary, our 5caC genomic profiling while requiring DIP-CAB-Seq validation as well as generation of additional data sets, tallies with current literature descriptions of 5caC DNA occupancy which is suggestive of a positive influencer of transcriptional activation. However, this may be contingent on the permutations of transcription factor expressed within the nuclear milieu.

Cell cycle phase may influence oximC presence dynamics

In this thesis, a dynamic nature of 5caC enrichment was demonstrated not only by susceptibility to extrinsic factors such as cell culture media, but also via clonal selection of screened cells, implying an intrinsic influence on oximC cellular distributions. One such intrinsic factor may be cell cycle synchronicity. TDG expression is down-regulated during cell

cycle progression in human pluripotent stem cell-derived neural progenitors (hPSC-NPCs) and anti-correlates with 5fC and 5caC nuclear presence (Germanguz *et al*, 2018). TDG exhibited highest expression at G0 and G1 phases but was dramatically depleted in S-M phases while 5fC and 5caC displayed opposite distributions, displaying highest enrichment at S-M phases and lowest at G1 (Germanguz *et al*, 2018).

5hmC displayed a biphasic pattern of enrichment with peaks at G1, G2 and M phases but significant reduction at S phase. 5fC and 5caC magnitudes significantly accumulated on siRNA mediated *TDG* knockdown, consistent with previous reports (Wheldon *et al*, 2014) however 5hmC levels did not significantly fluctuate (Germanguz *et al*, 2018). These findings are concordant with observations of TDG mediated excision of higher 5mC oxidised derivatives 5fC and 5caC (He *et al*, 2011) as TDG expression peaks correlate with 5fC and 5caC toughs (Germanguz *et al*, 2018).

Interestingly, Methylase –Assisted bisulphite sequencing (MAB-Seq) which enables detection of 5fC and 5caC modifications revealed 5fC and 5caC accumulation at the *EGR1* transcriptional start site upon *TDG* knockdown which correlated with *EGR1* transcriptional down-regulation and a significant increase in the proportion of Ki67 positive proliferating cells (Germanguz *et al*, 2018).

The localisation of 5fC and 5caC to *ERG1* TSS highlights an intriguing notion of transcription factor orchestration of oximC distribution and

possibly DNA occupancy stabilisation contingent on cell cycle phase status. Consistent with this notion, Singh *et al* utilised the fluorescent ubiquitin-based cell cycle indicator (FUCCI) system to track individual human ESCs at different cell cycle phases based on fluorescent reports linked to G1 (CDT1-kusabira orange 2 red signal) and S-M (GEMININ-azami green signal) which were non-overlapping and more sensitive than Vibrant Dye Cycle flow cytometry methods (Singh *et al*, 2013). FUCCI sorting coupled with RNA-Seq and gene ontology analysis demonstrated a preponderance of powerful developmental regulator transcription factor expression e.g. *SOX17*, *GATA6* and *FOXA2* at late G1 which was depleted during S-M phase under stem-maintenance conditions (Singh *et al*, 2013).

During early differentiation, these developmental regulator expression levels still displayed late G1 phase peaks consistently throughout 4 day differentiation programme which displayed a biphasic pattern of expression, peaking on day 0, depreciating on day 2 and increasing to day 0 levels on day 4 (Singh *et al*, 2013). Interestingly, 5hmC levels as detected by mass spectrometry and dot blot assays on G1, S, G2 and M phase separated samples exhibited highest enrichment during late G1 phase and were heavily depleted in S phase but restored at G2. Not only did 5hmC enrichment patterns show phasic alignment with developmental regulator transcription factors at their highest expression levels, oxi-bisulphite sequencing of differentially hydroxymethylated regions revealed 5hmC occupancy at enhancers marked by H3K4me1

and H3K9ac and bivalent histone marks. These points suggest G1 may represent an epigenetically facilitated developmental transcription factor 'primed window' in which cells may embark on differentiation programmes (Singh *et al*, 2013). Taken together these studies implicate a cell cycle-phasic separation of 5hmC and 5fC/5caC distributions and unique genomic occupancies (Singh *et al*, 2013; Germanguz *et al*, 2018) lending credence to the notion of differential functionality of oximC modifications and not simply just DNA demethylation intermediates (Song and He, 2013).

Transcription factor – 5caC interactions may be mediated by TET2

TET2 dysregulation has been associated with multiple cancers correlating with significant reductions in 5hmC and promotion of a cancer stem cell phenotype (Langemeijer *et al*, 2009; Li *et al*, 2011; Huang *et al*, 2016; Puig *et al*, 2018). TET2 has also been associated with promoting B cell reprogramming to iPSC cell fates in a transcription factor recruitment-dependent manner (Doege *et al*, 2012; Hu *et al*, 2014; Di Stefano *et al*, 2014; Di Stefano *et al*, 2016).

Transcription factors such as CCAAT enhancer binding protein alpha (C/EBPa) can recruit TET2 to highly methylated enhancers in B cells (Sardina *et al*, 2018). Depending on the degree of enhancer methylation (high, medium or low), the recruitment of TET2 to these sites may progress from closed and inactive to primed (highly methylated), primed

to active (medium methylated) or sustained active and open (low methylated) as demarcated by active H2K27ac and H3K4me2 histone marks (Sardina *et al*, 2018). As expected, the activation of these enhances was preceded by 5mC down-regulation and 5hmC increase at these sites which culminated in B cell priming (B') for subsequent OSKM mediated reprogramming to B cell derived iPSCs (Sardina *et al*, 2018).

TET2 can also interact with 'intermediary proteins' which stabilise the interaction between TET2 and other transcription factors. Mammalian two-hybrid studies utilising the V16 activator bound to full length TET2 / TET2 N terminal domain / TET2 catalytic domain and prey bound to the GAL4 DNA binding domain revealed that SMAD1 interacting protein 1 (SNIP1) is essential for TET2 interactions and recruitment to DNA sequences by transcription factors such as c-MYC, Cell Division 5 Like (CDC5L), BCL2 associated transcription factor 1 (BCLAF1), MAX dimerization factor (MGA), NFKB repressing factor (NKRF) and HIC ZBTB transcriptional repressor 1 (HIC1) (Chen *et al*, 2018). CRISPR cas9 mutagenesis of TET2 and SNIP coupled with mammalian two-hybrid, co-immunoprecipitation and protein mass spectrometry validated the TET2-SNIP-c-MYC interactions and revealed TET2-c-MYC dependence on SNIP for transcription factor binding (Chen *et al*, 2018).

Formation of the TET2-SNIP-c-MYC complex correlated with 5hmC elevation and 5mC reduction at c-MYC target gene transcriptional start sites and up-regulation of target gene expression (Chen *et al*, 2018).

These studies suggest that aberrant transcription factor activity may contribute to aberrant oximC distributions and thus altered transcriptome profiles in malignancies.

Intriguingly, WT1 can recruit TET2 to its target gene promoters via direct interaction of WT1 DNA binding domain and TET2 catalytic domain without the requirement of a ternary 'adaptor' protein such as SNIP1 (Wang *et al*, 2015). *WT1* knockdown produced no significant effects on *TET2* target genes *TMEM116* or *CDKL4* expression but did, as expected, abrogate *WT1* target genes *BTRC*, *DACT1*, *TBL1X* expression which was also achieved by *TET2* catalytic domain mutagenesis (Wang *et al*, 2015).

Additionally, TET2 ChIP-qPCR analysis revealed significant increase in TET2 bound to WT1 target gene promoters upon *WT1* over expression in HEK293T cells. The *BTRC*, *DACT1* and *TBL1X* targets possessing bound TET2 at their promoters exhibited increased elevated transcriptional activity compared to control cells in which *WT1* was not over-expressed (Wang *et al*, 2015). Moreover, TET2 bound to *WT1* target gene CpG sites within promoters and TSS correlated with increased 5hmC, decreased 5mC and the enrichment of active chromatin associated H3K4me3 marks compared to controls (Wang *et al*, 2015).

TET2 catalytic domain residues were identified which upon substitution mutation (D1242V, G1256D, R1302G, A1505T, G1861R and T1884A abrogated WT1-binding ability but did not ablate TET2 dioxygenase activity as 5mC could still be catalysed to 5hmC and 5caC (Wang *et al*,

2015). Importantly, these mutations elicited the same effects on cell physiology as produced by *WT1* knockdown exemplified by failure of *tet2* mutant HL-60 leukemic cells to suppress malignant cell proliferation compared to TET2 wildtype, WT1 expressing cells. This implies that while catalytically active but WT1 binding-deficient tet2 may still retain full independent 5mC dioxygenase capacity, disruption of WT1 binding ability perturbs tet2 direction to pathologically relevant CpG loci which consequently precludes tet2 tumour suppressor functionality (Wang *et al*, 2015).

Similarly, Rampal *et al* studied the effects of *IDH1/2*, *TET2* and *WT1* mutations on 5mC and 5hmC genomic magnitudes and distributions (Rampal *et al*, 2014). In contrast to Wang *et al* who focused on WT1 target gene promoters (Wang *et al*, 2015), Rampal *et al* identified 5mC and 5hmC differentially enriched sites genome wide by bisulphite sequencing, reduced representation bisulphite sequencing and β-glucosyltransferase mediated biotin chemical labelling and streptavidin capture (Rampal *et al*, 2014). This revealed greatest 5hmC differential regions occurring within gene bodies and at distal enhancer regions as oppose to 5mC differential enrichment regions which predominated at gene promoters and TSS in Acute Myeloid Leukemic cells (Rampal *et al*, 2014). Transcription factor motif analysis correlated WT1 binding to 5hmC enriched regions which was confirmed by WT1 ChIP-Seq and 5hmC chemical tagging and sequencing which also exposed WT1 binding to 5hmC enriched sites independent of TET2 in AML cells (Rampal *et al*, 2014). Altogether, although WT1 and TET2 can directly

interact to facilitate DNA demethylation at WT1 specific target genes, the impetus for their association is still currently unknown.

WT1 may influence a glioma stem cell phenotype via chromatin remodelling

Considering that in malignant cell types exhibiting *WT1* over-expression, *WT1* depletion attenuates *in vitro* parameters associated with pathology (Menssen *et al*, 2000; Oji *et al*, 2002; Nishida *et al*, 2006; Kim *et al*, 2008, Vincent *et al*, 2010; Brett *et al*, 2013; McGregor *et al*, 2018), we speculate whether ectopic WT1 expression may be able to modulate chromatin conformational changes resulting in reinstatement of a foetal brain transcriptional paradigm in adult neural stem and progenitor cells resulting inTET2 mediated 5caC enrichment at enhancers, promoter and gene bodies of tumorigenic genes.

Radial glial progenitors (RGPs) have been implicated as the precursor for multiple neurons, astrocytes and oligodendrocyte types found to populate the mammalian cortex and originate from neuroepithelial stem cells (Kowalczyk *et al*, 2009; Taverna *et al*, 2014; Vasistha *et al*, 2014).

The ATPase catalytic subunit of SWItch/ Sucrose Non Fermentable (SWI/SNF) chromatin remodelling complex, Brahma-related gene 1 (*BRG1/SMARCA4*) is involved in regulation of RGP-derived transit amplifying progenitor proliferation, maintenance and isolation from neurogenic differentiation cues (Lessard *et al*, 2007; Matsumoto *et al*,

2016). At embryonic days 14.5-15.5, *BRG1* depletion perturbs the delicate neurogenic-gliogenic differentiation balance, reducing the efficiency of transit amplifying cell derived neuron and astrocyte production, possibly by dysregulation of sonic-hedgehog signalling (Lessard *et al*, 2007; Matsumoto *et al*, 2016).

Interestingly, the pleiotropic peptide Thymosin Beta 4 (Tβ4) can induce BRG1 mediated activation of *WT1* via interacting with transcription factor CCAAT binding protein alpha (C/EBPα) which recruits BRG1 to *WT1* promoter elements. The consequent *WT1* expression activates a foetal epicardium transcriptional profile in murine adult quiescent epicardial cells (Vieira *et al*, 2017). Curiously, both *TB4* and *BRG1* are highly expressed in glioblastoma and their expressions have been attributed in the maintenance of glioma stem cell proliferation, migration and invasion (Bai *et al*, 2012; Wirsching *et al*, 2013).

Indeed, extracellular matrix components pertinent to glioblastoma stem cell maintenance such as integrins beta 1, alpha 3, alpha 6, cell adhesion marker 1 (Mi *et al*, 2005; Nakada *et al*, 2013; Tilghman *et al*, 2013; Ying *et al*, 2014; Erhart *et al*, 2019) (*ITGB1*, *ITGA3*, *ITGA6*, *ICAM-1* and *CDH3* respectively) were all down regulated in C5A and C7 *wt1* mutant cell lines commensurate with our observations of abrogated gliomasphere formation.

WT1 and DNA demethylation machinery may influence cell cycle regulation

We also observed 5caC and WT1 narrow DIP-Seq and ChIP-Seq peaks aligning at centromeric alpha satellite repeats in U251. Centromeres are epigenetically designated (Hara *et al*, 2019; Nachemia-Arbely *et al*, 2019). WT1 possesses ability to delay anaphase promoting complex/cyclosome (APC/c) mediated degradation of the SECURIN/Cyclin B complex via physically interacting and inhibiting activity of the closed MAD2 protein (cMAD2) in the cMAD2-CDC20-BUB1R-BUB3-MAD3 mitotic checkpoint complex (Shandilya *et al*, 2014; Shandilya & Roberts, 2015). This WT1 mediated regulation results in a protracted metaphase and delayed anaphase exit ensuring spindle fibres are attached to kinetochores at the centromere prior to onset of anaphase (Shandilya *et al*, 2014; Shandilya & Roberts, 2015).

The proto-oncogene c-MYC displays direct MAD2 trans-activating ability at centromeres exhibiting temporal equivalence with WT1 (Menssen *et al*, 2007). Considering C/EBPa can localise to pericentromeric DNA sequences (Schaufele *et al*, 2001), can interact with TET2 (Sardina *et al*, 2018) which can bind to c-MYC (Chen *et al*, 2018) which possesses a 5caC-sensitive binding partner MAX (Wang *et al*, 2017), we speculate whether oximCs may feature in regulation of cell cycle checkpoint mechanisms.

In this body of work we reported differential magnitude distributions between 5hmC and 5caC during neuronal and glial differentiation of GBM cell lines which exhibited *WT1* expression peaks consistent with 5caC

enrichment patterns. Moreover, *WT1* expression peaks were consistently accompanied by *TDG* down-regulation during the differentiation programmes in GBM cell lines and foetal brain-derived neural stem cells. Furthermore, *TDG* was up-regulated relative to controls in C5A and C7 *wt1* zinc finger mutants and *WT1* siRNA mediated knockdown U87MG cells, both resulting in 5caC magnitude reduction. WT1 is expressed at low levels at G0 and G1 phases in slow cycling but highly expressed in S, G2 and M phase in rapid proliferating cells in myeloid leukemic cells (Kerst *et al*, 2008). Additionally, flow cytometric analysis of Non-Small Cell Lung Cancer cells revealed high WT1 protein expression in G1 phase which correlated with enhanced S entry of these cell via WT1-mediated elevation of cyclin D and phosphorylated retinoblastoma protein (p-pRb) (Xu *et al*, 2013).

We therefore postulate a model of WT1-5caC activity in U251 and LN18 glioblastoma cell lines in which WT1 is ectopically highly expressed and cyclical, increasing in abundance in a synchronous fashion to TDG activity but functions to repress *TDG* gene expression. While 5caC modifications accumulate during S-M phase, the ectopic expression of WT1 displaying high expression while simultaneously repressing *TDG* in G1 may stabilised 5caC enrichment in G1 phase, thus explain the increased 5caC distribution within the cell population in U251 and LN18 and depletion in C5A and C7 *wt1* mutants. However, before any postulations regarding WT1 activity pertaining to 5caC interactions in

transcriptional regulation, our current datasets require validation and further experimentation.

6.5 Future Work

Owing to time and budget constraints, multiple experiments to complete and validate the current dataset could not be conducted. Therefore, future work would entail:

.

- The FUCCI system could be implemented to sort cells into different cell cycle phases by FACs. Sorted cells could then be stained for 5caC and WT1 by immunofluorescence and imaged using confocal microscopy to assess the synchronicity of 5caC and WT1 presence dynamics. These synchronisation and single cell sorting studies could be performed in serum and neurobasal conditions on GBM cell lines.
- Determining the stem cell status of glioblastoma cell lines by immunofluorescence staining for NESTIN, SOX and NOTCH1 proved to be an unrefined method of analysis. To assess the levels of oximCs in GBM stem cells, isolation of glioma stem cells could be performed using Hoechst 33342 dye staining to identify side population cells cultured in neurobasal media which could then be FACs sorted. Interrogation of the stem phenotype could be performed using immunofluorescent staining for CD133, CD15, NESTIN, SOX2 markers with cells positive for all staining separated from negative or partially negative ones. Functional

assays of neurosphere formation over multiple passages, limiting dilution assays to assess clonogenicity and self-renewal capacity and multi-lineage differentiation potential (neurons, astrocytes and oligodendrocytes) could be performed. Putative GBM stem cells could be interrogated for characteristic neural stem cell marker gene expression including *NESTIN*, *SOX2*, *NOTCH1*, *CD133*, *CD15*, *MSI*, *EGFR*, *FGF2* and *OLIG2* (Lathia *et al*, 2015; Wang *et al*, 2018; Wang *et al*, 2019) while neuron, astrocyte and oligodendrocyte differentiation could be confirmed using *TUJ1*, *NEUN*, *NEUROGENIN*, *MAP2* (neuron), *GFAP*, *S100B* and *GLAST*, (astrocyte) and *OLIG1* and *SOX10* (oligodendrocyte) markers (Trojanowski, Walkenstein & Lee, 1986; Mullen *et al*, 1992; Soltani *et al*, 2005; Cahoy *et al*, 2008; Pollard *et al*, 2009; Frati *et al*, 2018; Nierode *et al*, 2018; Noda *et al*, 2018) by qPCR.

- Bisulphite, oxi-bisulphite, MAB-Seq and caMAB-Seq would be required to validate all our current and future planned sequencing work.
- Considering multiple extracellular matrix (ECM) genes were down-regulated in C5A mutants relative to U251 controls, cell migration and invasion assays could be performed e.g. matrigel scratch assay or Boyden chamber migration assay.
- Considering the dramatic effects on 5hmC and 5caC presence observed by over-expression of –KTS and + KTS WT1 zinc finger domain isoforms in LN229, over-expression of these constructs could be performed on C5A and C7 wt1 mutants to determine if

one or both of these isoforms could 'rescue' wild type phenotypes e.g. neurosphere formation observed in U251 and LN18.

- Finally, a CRISPR-TDG fusion protein utilising the protein-RNA and DNA binding domains of hCas9 tethered to the catalytic glycosylase domain of TDG could be generated to enable the targeted removal of 5caC enriched sites in brain tumour cell lines. Targets for 5caC erasure would constitute 5caC enriched loci within transcriptional regulatory regions pertaining to disease relevant genes for which small guide RNAs could be designed. Indeed, multiple guides could be transfected simultaneous for the targeting of manifold loci. Alternatively, multiple sgRNAs sequences under control of the same U6 promoter may be incorporated as an array into the Adeno-associated virus intergration site 1 (AAVS1) safe harbour locus such that their expression is constitutive. The Cas9-TDG fusion construct controlled by an inducible promoter could also be incorporated into the safe harbour locus such that targeting of specific genes could be induced at specific times e.g. during differentiation. 5caC-DIP Seq could be performed on controls versus CRISPR-TDG targeted cells in each cell line such as UW228-8 or LN18 and sequencing would be validate by base-resolution sequencing techniques previously mentioned. Additionally, considering the redundancy of WT1 and EGR1 transcription factors for binding to their cognate consensus sequence, this Cas9-TDG fusion protein could be employed to investigate whether 5caC may have evolved

as epigenetic 'beacon' for the preferential selection of one transcription factor over another for occupancy at the same DNA sequence. Indeed, Hashimoto *et al* demonstrated the increase affinity of WT1 for 5caC on sequences harbouring the WT1/EGR1 cognate binding sequence. Interestingly, under conditions of high 5caC density within the consensus sequence, EGR1 affinity significantly increased, suggesting a threshold level of transcription factor selection in response epigenetic modification concentrations (Hashimoto *et al*, 2014). Therefore, implementation of a Cas9-TDG fusion protein in LN18 or U251 GBM cell lines versus its expression in a wt1 DNA binding null mutant cell line such as C5A or C7 may facilitate the examination of whether EGR1 may transcriptionally activate its targets in the presence/absence of WT1 under condition of Cas9-TDG targeted 5caC modulation. This may suggest a functional role of terminal oxidised 5mC derivatives in transcriptional regulation.

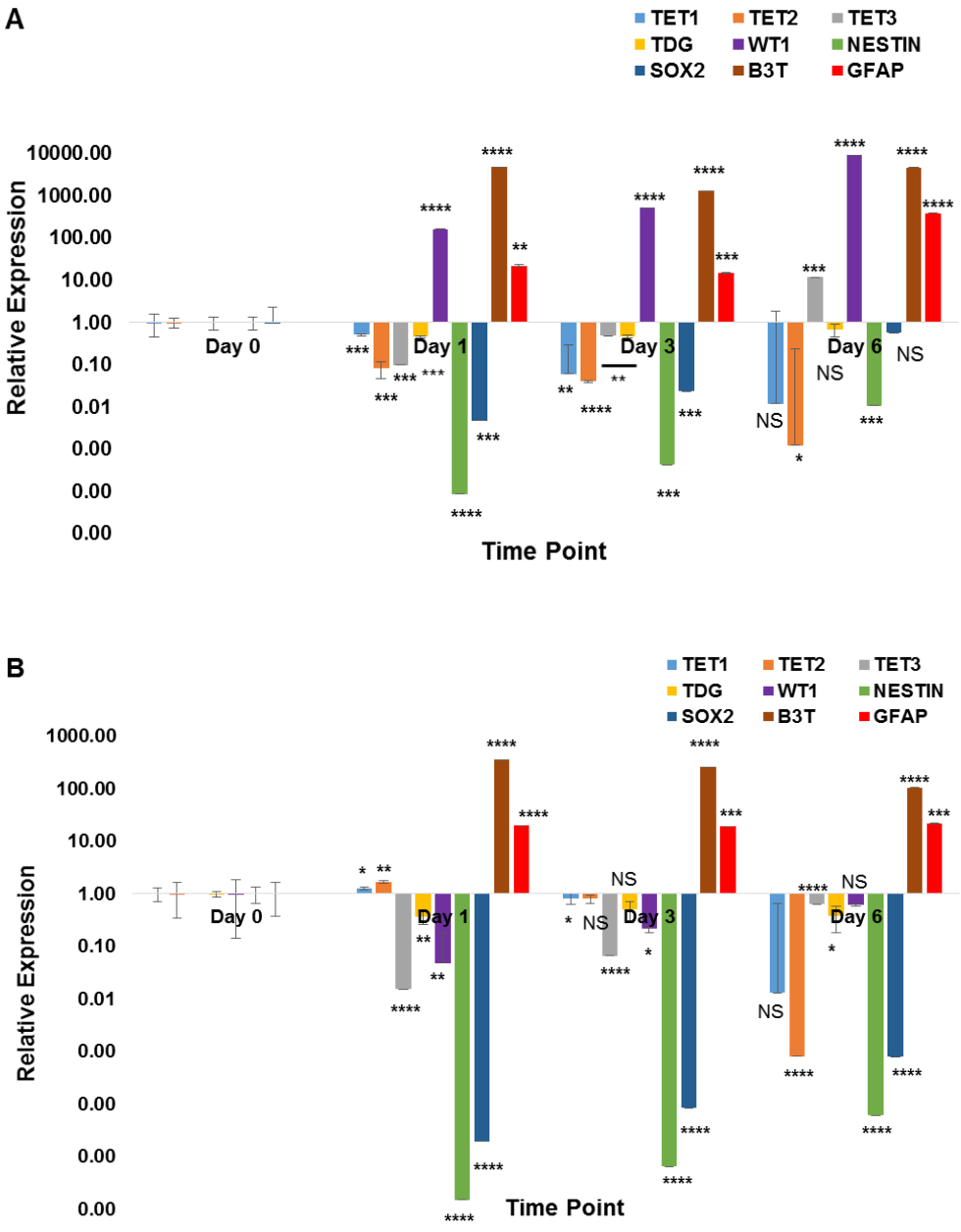
Concluding Remarks

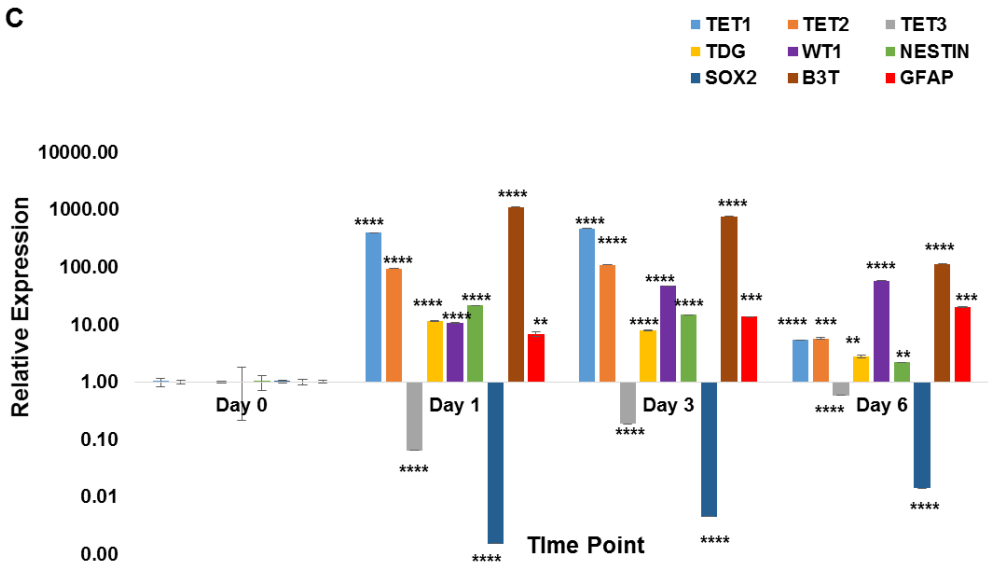
From the data generated in this thesis, the implementation of immunostaining, confocal microscopy and mass spectrometry have verified the presence and enrichment of 5caC in brain tumour cell lines at levels far below those of embryonic or reprogrammed pluripotent cells, but significantly exceeding those of adult healthy somatic cells. While the

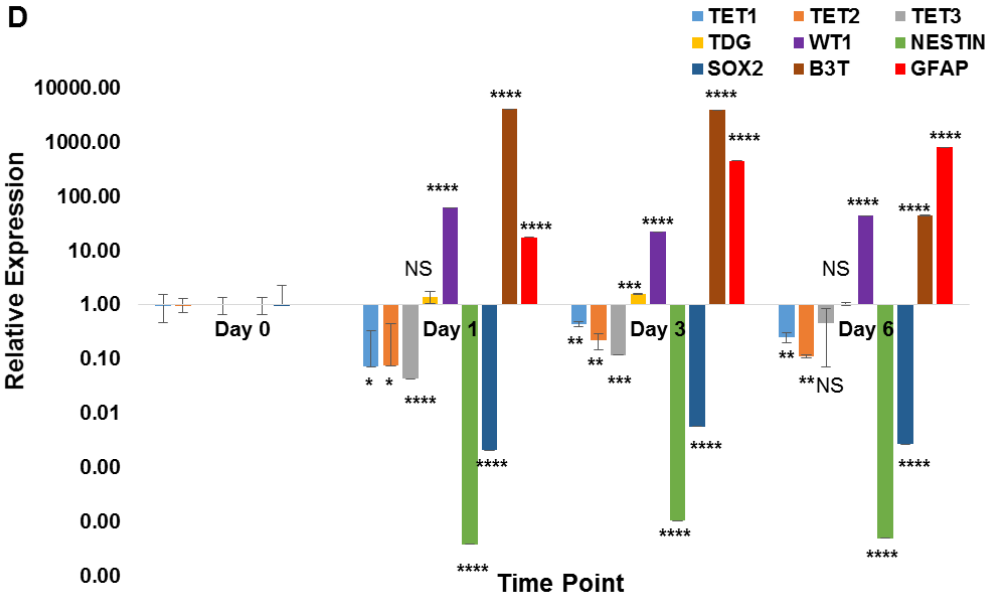
utilisation of appropriate tissue match controls and primary tumour tissue is required for a more accurate assessment of whether 5caC persistence in adult cells is a function of malignancy, the presence of 5caC above non-negligible levels in the cell lines interrogated here is not a trivial discovery. Manipulation cell culture conditions and introduction of paracrine signalling ligands during cell lineage specification has demonstrated that oximC magnitudes and distributions are modular and susceptible to extrinsic factors, however it is unknown their genomic occupancy induces or is a product of transcriptional profile fluctuations. Between GBM cell lines, 5caC genomic occupancy was detected at overlapping but also distinct loci, specific for each cell line and 5caC peaks aligned with those of WT1 in a minority of instances. The generated *wt1* zinc finger mutant GBM cell lines express a novel truncated mRNA isoform and these mutated cells display a phenotype of diminished neurosphere formation and extracellular matrix gene expression. This correlated with reduced number and size of tumours formed in orthotopically injected mice. In conclusion, while many subsequent experiments are required to complete and validate this dataset, this work represents a pilot study aimed at deciphering the role of WT1 as a putative 'reader' of 5caC in brain tumour cell lines but also, more widely, revealing the mechanisms governing transcription factor activity in the pathway of active DNA demethylation.

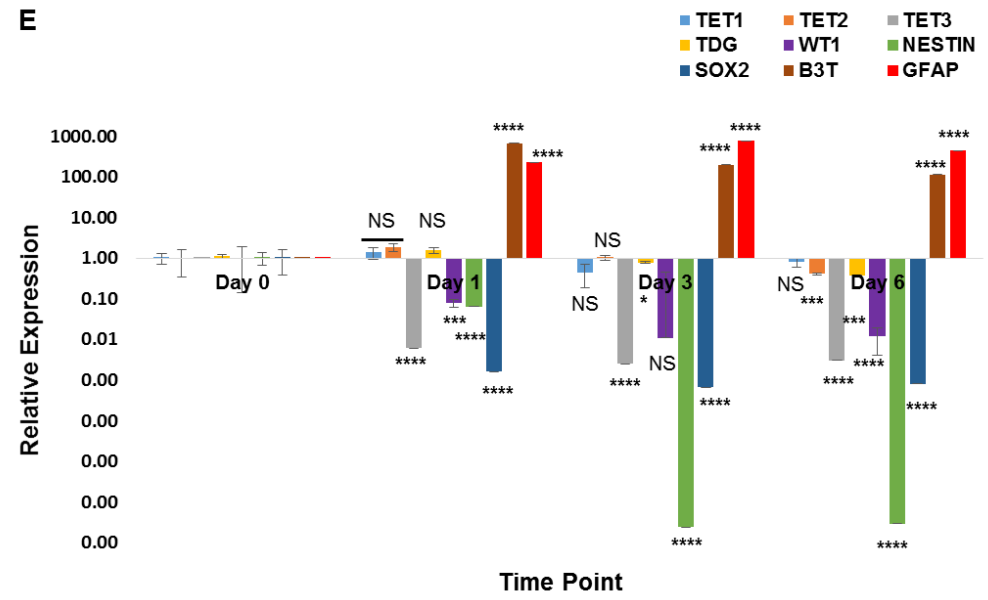
Appendix A

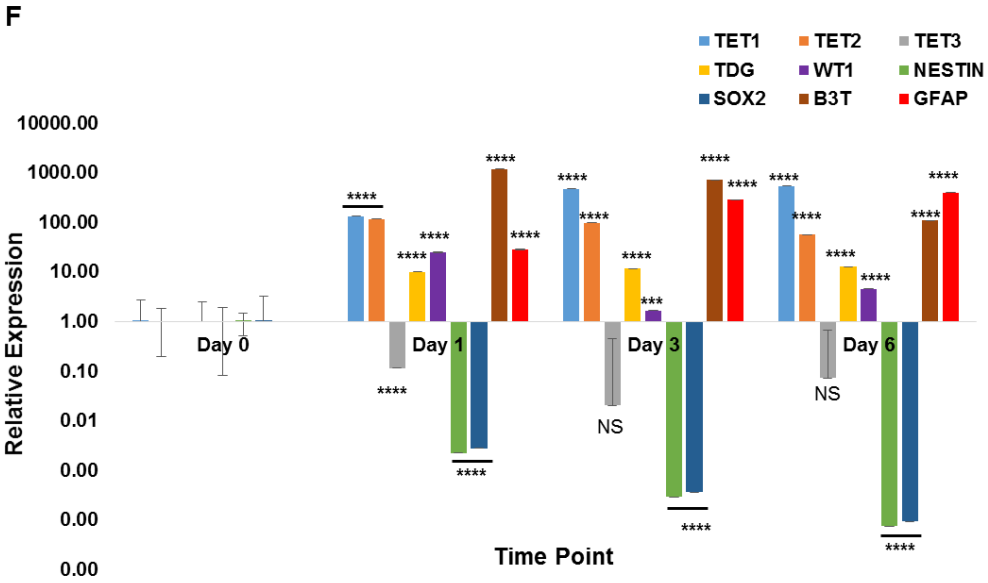
Glioblastoma and Foetal Brain Cell Line Differentiation

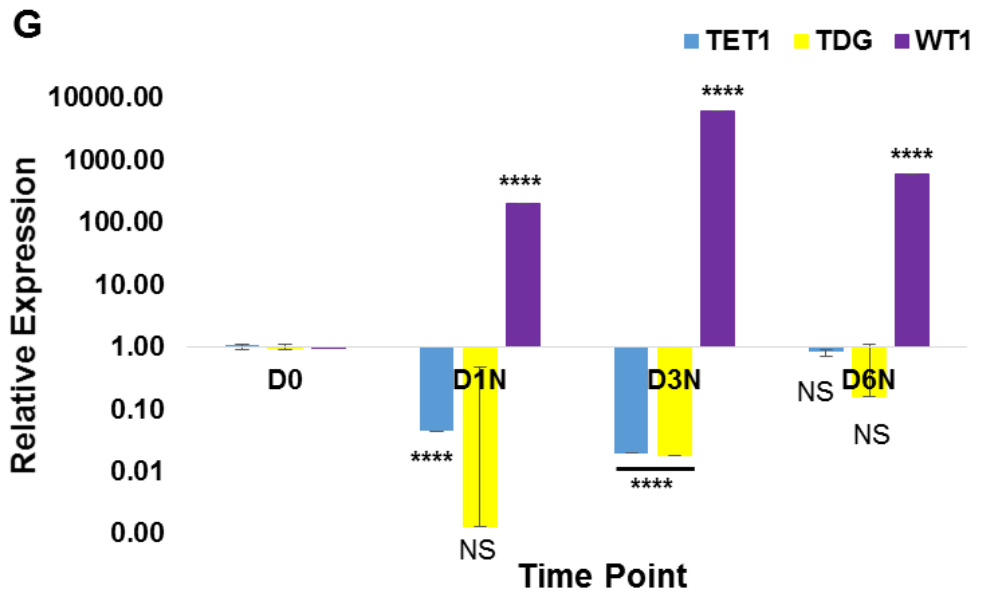



C
TET1
TET2
TET3
TDG
WT1
NESTIN
SOX2
B3T
GFAP
Relative Expression
10000.00
1000.00
100.00
10.00
1.00
0.10
0.01
0.00
Day 0
Day 1
Day 3
Day 6
Time Point


D
TET1
TET2
TET3
TDG
WT1
NESTIN
SOX2
B3T
GFAP
Relative Expression
10000.00
1000.00
100.00
10.00
1.00
0.10
0.01
0.00
0.00
NS
Day 0
Day 1
Day 3
Day 6
Time Point


E
TET1
TET2
TET3
TDG
WT1
NESTIN
SOX2
B3T
GFAP
Relative Expression
1000.00
100.00
10.00
1.00
0.10
0.01
0.00
0.00
0.00
0.00
0.00
Day 0
Day 1
Day 3
Day 6
NS
***
*
****
Time Point


F
TET1
TET2
TET3
TDG
WT1
NESTIN
SOX2
B3T
GFAP
Relative Expression
10000.00
1000.00
100.00
10.00
1.00
0.10
0.01
0.00
0.00
0.00
Day 0
Day 1
Day 3
Day 6
NS
****
***
Time Point


G
TET1
TDG
WT1
Relative Expression
10000.00
1000.00
100.00
10.00
1.00
0.10
0.01
0.00
D0
D1N
D3N
D6N
NS
****
Time Point

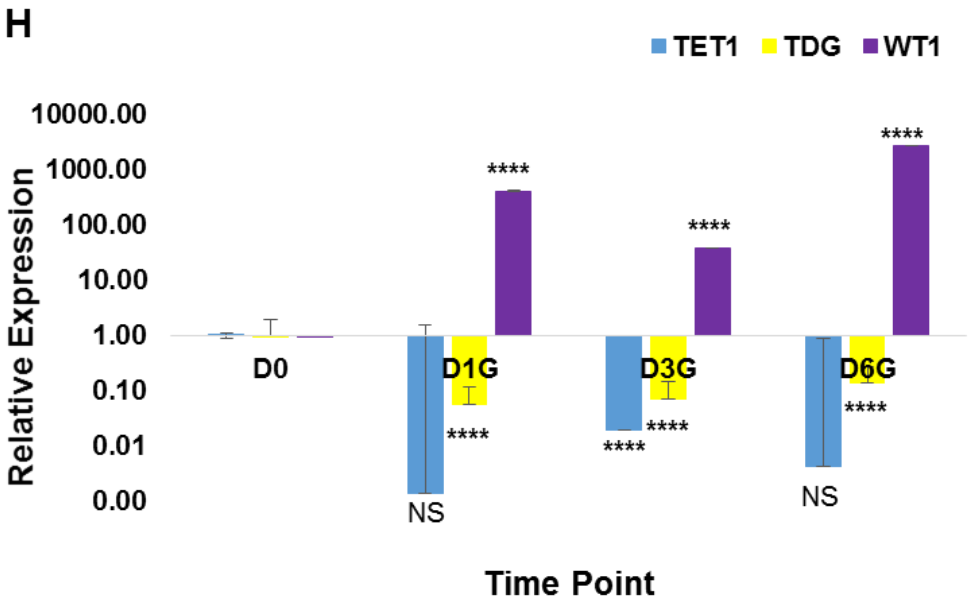


Figure A.1 Glioblastoma and FB83 cell line differentiation transcriptional profiles. (A-C) QPCR profiles of glioblastoma cell lines LN229, U251 and U87MG neural differentiation respectively. (D-F) QPCR profiles of LN229, U251 and U87MG glial differentiation respectively. (G-H) QPCR profiles of foetal brain cell line FB83 neural and glial differentiation respectively. D1N denotes Day 1 of neural differentiation and D1G day 1 of glial differentiation. Mean ± SD, Kruskal-Wallis test, NS $p \geq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$, **** $p < 0.0001$, $n = 3$.

Appendix B

Mouse Orthotopic Xenografts

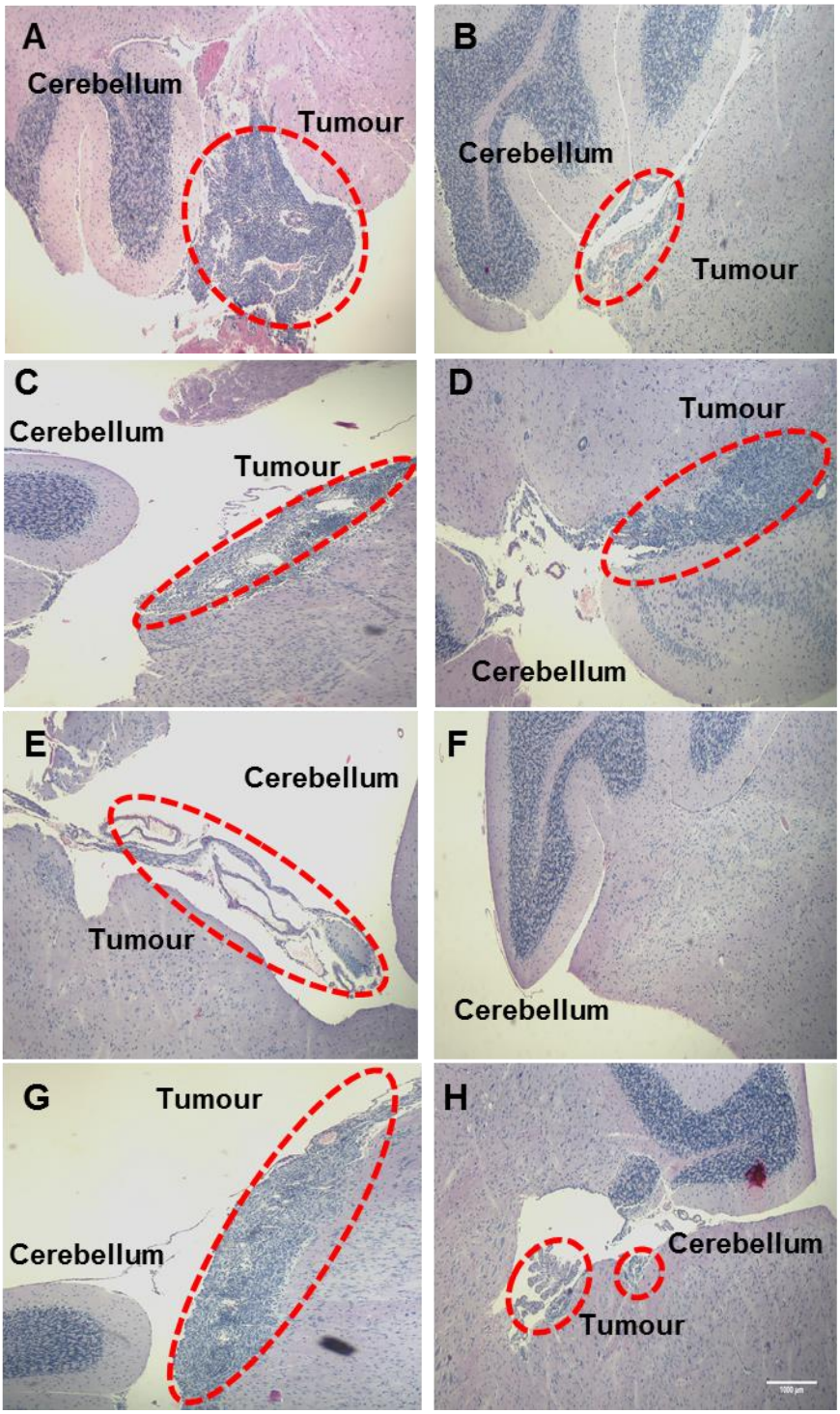


Figure B.1 Mouse orthotopic xenografts of U251 control vs C5A mutant derived tumours. (A, C, E & G) Control U251 injected mice. (B, D, F & H) C5A

mutant injected mice. Red dashed circles depict tumour masses, H&E staining, scale bar represents 1000µm.

Bibliography

ABAKIR, A., WHELDON, L., JOHNSON, A. D., LAURENT, P. & RUZOV, A. 2016. Detection of modified forms of cytosine using sensitive immunohistochemistry. *JoVE (Journal of Visualized Experiments)*, e54416.

ABAKIR, A., WHELDON, L. M. & RUZOV, A. 2016. Immunohistochemical Detection of Oxidized Forms of 5-Methylcytosine in Embryonic and Adult Brain Tissue. *Epigenetic Methods in Neuroscience Research*, 125-137.

ABLES, J. L., BREUNIG, J. J., EISCH, A. J. & RAKIC, P. 2011. Not (ch) just development: Notch signalling in the adult brain. *Nature Reviews Neuroscience,* 12, 269-283.

AHRINGER, J. 2000. NuRD and SIN3: histone deacetylase complexes in development. *Trends in Genetics,* 16, 351-356.

AKPA, M. M., IGLESIAS, D. M., CHU, L. L., CYBULSKY, M., BRAVI, C. & GOODYER, P. R. 2015. Wilms tumor suppressor, WT1, suppresses epigenetic silencing of the β-catenin gene. *Journal of Biological Chemistry,* 290, 2279-2288.

ALGAR, E., KHROMYKH, T., SMITH, S., BLACKBURN, D., BRYSON,

G. & SMITH, P. 1996. A WT1 antisense oligonucleotide inhibits proliferation and induces apoptosis in myeloid leukaemia cell lines. *Oncogene,* 12**,** 1005-1014.

ALLGAYER, J., KITSERA, N., BARTELT, S., EPE, B. & KHOBTA, A. 2016. Widespread transcriptional gene inactivation initiated by a repair intermediate of 8-oxoguanine. *Nucleic acids research,* 44**,** 7267-7280.

AMOUROUX, R., NASHUN, B., SHIRANE, K., NAKAGAWA, S., HILL, P. W., D'SOUZA, Z., NAKAYAMA, M., MATSUDA, M., TURP, A. & NDJETEHE, E. 2016. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nature cell biology,* 18**,** 225.

AMUNDSON, S., ZHAN, Q., PENN, L. & FORNACE JR, A. 1998. Myc suppresses induction of the growth arrest genes gadd34, gadd45, and gadd153 by DNA-damaging agents. *Oncogene,* 17**,** 2149-2154.

AN, J., RAO, A. & KO, M. 2017. TET family dioxygenases and DNA demethylation in stem cells and cancers. *Experimental & molecular medicine,* 49**,** e323.

ANNOVAZZI, L., CALDERA, V., MELLAI, M., RIGANTI, C., BATTAGLIA, L., CHIRIO, D., MELCARNE, A. & SCHIFFER, D. 2015. The DNA damage/repair cascade in glioblastoma cell lines after chemotherapeutic agent treatment. *International journal of oncology,* 46**,** 2299-2308.

ARITA, K., ARIYOSHI, M., TOCHIO, H., NAKAMURA, Y. & SHIRAKAWA, M. 2008. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature,* 455**,** 818-821.

ARMSTRONG, J. F., PRITCHARD-JONES, K., BICKMORE, W. A., HASTIE, N. D. & BARD, J. B. 1993. The expression of the Wilms' tumour gene, WT1, in the developing mammalian embryo. *Mechanisms of development,* 40**,** 85-97.

ARNOTT, S., BOND, P. & CHANDRASEKARAN, R. 1980. Visualization of an unwound DNA duplex. *Nature,* 287**,** 561.

BACHMAN, M., URIBE-LEWIS, S., YANG, X., BURGESS, H. E., IURLARO, M., REIK, W., MURRELL, A. & BALASUBRAMANIAN, S. 2015. 5-Formylcytosine can be a stable DNA modification in mammals. *Nature chemical biology,* 11**,** 555.

BAI, J., MEI, P.-J., LIU, H., LI, C., LI, W., WU, Y.-P., YU, Z.-Q. & ZHENG, J.-N. 2012. BRG1 expression is increased in human glioma and controls glioma cell proliferation, migration and invasion in vitro. *Journal of cancer research and clinical oncology,* 138**,** 991-998.

BALLESTAR, E. & WOLFFE, A. P. 2001. Methyl-CpG-binding proteins. *European Journal of Biochemistry,* 268**,** 1-6.

BALSS, J., MEYER, J., MUELLER, W., KORSHUNOV, A., HARTMANN, C. & VON DEIMLING, A. 2008. Analysis of the IDH1 codon 132 mutation in brain tumors. *Acta neuropathologica,* 116**,** 597-602.

BANFALVI, G. 2017. *Cell cycle synchronization*, Springer.

BAO, S., WU, Q., MCLENDON, R. E., HAO, Y., SHI, Q., HJELMELAND, A. B., DEWHIRST, M. W., BIGNER, D. D. & RICH, J. N. 2006. Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *nature,* 444**,** 756.

BARAU, J., TEISSANDIER, A., ZAMUDIO, N., ROY, S., NALESSO, V., HÉRAULT, Y., GUILLOU, F. & BOURC'HIS, D. 2016. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science,* 354**,** 909-912.

BARBAUX, S., NIAUDET, P., GUBLER, M.-C., GRÜNFELD, J.-P., JAUBERT, F., KUTTENN, F., FÉKÉTÉ, C. N., SOULEYREAU-THERVILLE, N., THIBAUD, E. & FELLOUS, M. 1997. Donor splice-site mutations in WT1 are responsible for Frasier syndrome. *Nature genetics,* 17**,** 467.

BARDELLA, C., AL-DALAHMAH, O., KRELL, D., BRAZAUSKAS, P., AL-QAHTANI, K., TOMKOVA, M., ADAM, J., SERRES, S., LOCKSTONE, H. & FREEMAN-MILLS, L. 2016. Expression of Idh1R132H in the murine subventricular zone stem cell niche recapitulates features of early gliomagenesis. *Cancer Cell,* 30**,** 578-594.

BARTON, S. C., ARNEY, K. L., SHI, W., NIVELEAU, A., FUNDELE, R., SURANI, M. A. & HAAF, T. 2001. Genome-wide methylation patterns in normal and uniparental early mouse embryos. *Human Molecular Genetics,* 10**,** 2983-2987.

BASSIL, C. F., HUANG, Z. & MURPHY, S. K. 2013. Bisulfite pyrosequencing. *Ovarian Cancer.* Springer.

BAYLIN, S. B. & OHM, J. E. 2006. Epigenetic gene silencing in cancer-a mechanism for early oncogenic pathway addiction? *Nature reviews. Cancer,* 6**,** 107.

BAYMAZ, H., FOURNIER, A., LAGET, S., JI, Z., JANSEN, P. W., SMITS, A. H., FERRY, L., MENSINGA, A., POSER, I. & SHARROCKS,

A. 2014. MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain. *Proteomics,* 14**,** 2179-2189.

BICKMORE, W., OGHENE, K., LITTLE, M., SEAWRIGHT, A., VAN HEYNINGEN, V. & HASTIE, N. 1992. Modulation of DNA binding specificity by alternative splicing of the Wilms tumor wt1 gene transcript. *Science,* 257**,** 235-237.

BIRD, A. 1999. DNA methylation de novo. *Science,* 286**,** 2287-2288.

BIRD, A. 2002. DNA methylation patterns and epigenetic memory. *Genes & development,* 16**,** 6-21.

BIRNER, P., PUSCH, S., CHRISTOV, C., MIHAYLOVA, S., TOUMANGELOVA-UZEIR, K., NATCHEV, S., SCHOPPMANN, S. F., TCHORBANOV, A., STREUBEL, B. & TUETTENBERG, J. 2014. Mutant IDH1 inhibits PI3K/Akt signaling in human glioma. *Cancer,* 120**,** 2440-2447.

BOLZER, A., KRETH, G., SOLOVEI, I., KOEHLER, D., SARACOGLU, K., FAUTH, C., MÜLLER, S., EILS, R., CREMER, C. & SPEICHER, M. R. 2005. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS biology,* 3**,** e157.

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**,** 2536-2539.

BOURNE, T. D., ELIAS, W. J., LOPES, M. B. S. & MANDELL, J. W. 2010. WT1 is not a reliable marker to distinguish reactive from neoplastic astrocyte populations in the central nervous system. *Brain Pathology,* 20**,** 1090-1095.

BRANCO, M. R., FICZ, G. & REIK, W. 2012. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nature Reviews Genetics,* 13**,** 7-13.

BRETT, A., PANDEY, S. & FRAIZER, G. 2013. The Wilms' tumor gene (WT1) regulates E-cadherin expression and migration of prostate cancer cells. *Molecular cancer,* 12**,** 3.

BROADLEY, K. W., HUNN, M. K., FARRAND, K. J., PRICE, K. M., GRASSO, C., MILLER, R. J., HERMANS, I. F. & MCCONNELL, M. J. 2011. Side population is not necessary or sufficient for a cancer stem cell phenotype in glioblastoma multiforme. *Stem cells,* 29**,** 452-461.

BROWN, S. E., FRAGA, M. F., WEAVER, I. C., BERDASCO, M. & SZYF, M. 2007. Variations in DNA methylation patterns during the cell cycle of HeLa cells. *Epigenetics,* 2**,** 54-65.

BRUENING, W., MOFFETT, P., CHIA, S., HEINRICH, G. & PELLETIER, J. 1996. Identification of nuclear localization signals within the zinc fingers of the WT1 tumor suppressor gene product. *FEBS letters,* 393**,** 41-47.

BYEON, S.-J., MYUNG, J. K., KIM, S. H., KIM, S.-K., PHI, J. H. & PARK, S.-H. 2012. Distinct genetic alterations in pediatric glioblastomas. *Child's Nervous System,* 28**,** 1025-1032.

CALL, K. M., GLASER, T., ITO, C. Y., BUCKLER, A. J., PELLETIER, J., HABER, D. A., ROSE, E. A., KRAL, A., YEGER, H. & LEWIS, W. H. 1990. Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. *Cell,* 60**,** 509-520.

CANVER, M. C., BAUER, D. E., DASS, A., YIEN, Y. Y., CHUNG, J.,

MASUDA, T., MAEDA, T., PAW, B. H. & ORKIN, S. H. 2014. Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *Journal of Biological Chemistry,* 289**,** 21312-21324.

CAPPER, D., JONES, D. T., SILL, M., HOVESTADT, V., SCHRIMPF, D., STURM, D., KOELSCHE, C., SAHM, F., CHAVEZ, L. & REUSS, D. E. 2018. DNA methylation-based classification of central nervous system tumours. *Nature,* 555**,** 469-474.

CAPPER, D., WEIßERT, S., BALSS, J., HABEL, A., MEYER, J., JÄGER, D., ACKERMANN, U., TESSMER, C., KORSHUNOV, A. & ZENTGRAF, H. 2010. Characterization of R132H mutation-specific IDH1 antibody binding in brain tumors. *Brain pathology,* 20**,** 245-254.

CARDOSO, L. C., DE SOUZA, K. R., REIS, D. O., HELENA, A., ANDRADE, R. C., BRITTO, A. C., DE LIMA, M. A., DOS SANTOS, A. C., DE FARIA, P. S. & FERMAN, S. 2013. WT1, WTX and CTNNB1 mutation analysis in 43 patients with sporadic Wilms' tumor. *Oncology reports,* 29**,** 315-320.

CARDOSO, M. C. & LEONHARDT, H. 1999. DNA methyltransferase is actively retained in the cytoplasm during early development. *The Journal of cell biology,* 147**,** 25-32.

CARELLA, A., TEJEDOR, J. R., GARCÍA, M. G., URDINGUIO, R. G., BAYÓN, G. F., SIERRA, M., LÓPEZ, V., GARCÍA-TORAÑO, E., SANTAMARINA-OJEDA, P. & PÉREZ, R. F. 2019. Epigenetic downregulation of TET3 reduces genome-wide 5hmC levels and

promotes glioblastoma tumorigenesis. *International journal of cancer*.

CARÉN, H., STRICKER, S. H., BULSTRODE, H., GAGRICA, S., JOHNSTONE, E., BARTLETT, T. E., FEBER, A., WILSON, G., TESCHENDORFF, A. E. & BERTONE, P. 2015. Glioblastoma stem cells respond to differentiation cues but fail to undergo commitment and terminal cell-cycle arrest. *Stem cell reports,* 5**,** 829-842.

CHALITCHAGORN, K., SHUANGSHOTI, S., HOURPAI, N., KONGRUTTANACHOK, N., TANGKIJVANICH, P., THONG-NGAM, D., VORAVUD, N., SRIURANPONG, V. & MUTIRANGURA, A. 2004. Distinctive pattern of LINE-1 methylation level in normal tissues and the association with carcinogenesis. *Oncogene,* 23**,** 8841-8846.

CHARLTON, J., IRTAN, S., BERGERON, C. & PRITCHARD-JONES, K. 2017. Bilateral Wilms tumour: a review of clinical and molecular features. *Expert Reviews in Molecular Medicine,* 19.

CHEN, Q., CHEN, Y., BIAN, C., FUJIKI, R. & YU, X. 2013. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature,* 493**,** 561-564.

CHENG, X. & ROBERTS, R. J. 2001. Base flipping. *eLS*.

CHIANG, P. K., GORDON, R. K., TAL, J., ZENG, G., DOCTOR, B., PARDHASARADHI, K. & MCCANN, P. P. 1996. S-Adenosylmethionine and methylation. *The FASEB Journal,* 10**,** 471-480.

CHIBA, Y., HASHIMOTO, N., TSUBOI, A., RABO, C., OKA, Y., KINOSHITA, M., KAGAWA, N., OJI, Y., SUGIYAMA, H. & YOSHIMINE, T. 2010. Prognostic value of WT1 protein expression level and MIB-1 staining index as predictor of response to WT1 immunotherapy in

glioblastoma patients. *Brain tumor pathology,* 27**,** 29-34.

CHU, V. T., WEBER, T., WEFERS, B., WURST, W., SANDER, S., RAJEWSKY, K. & KÜHN, R. 2015. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature biotechnology,* 33**,** 543.

CIERNIA, A. V. & LASALLE, J. 2016. The landscape of DNA methylation amid a perfect storm of autism aetiologies. *Nature Reviews Neuroscience,* 17**,** 411.

CIMMINO, L., ABDEL-WAHAB, O., LEVINE, R. L. & AIFANTIS, I. 2011. TET family proteins and their role in stem cell differentiation and transformation. *Cell stem cell,* 9**,** 193-204.

CLARK, A. J. 2006. The Expression and Function of Wilms' Tumor 1 in Malignant Glioma.

CLARK, A. J., DOS SANTOS, W. G., MCCREADY, J., CHEN, M. Y., VAN METER, T. E., WARE, J. L., WOLBER, S. B., FILLMORE, H. & BROADDUS, W. C. 2007. Wilms tumor 1 expression in malignant gliomas and correlation of+ KTS isoforms with p53 status.

CLARK, A. J., WARE, J. L., CHEN, M. Y., GRAF, M. R., VAN METER, T. E., DOS SANTOS, W. G., FILLMORE, H. L. & BROADDUS, W. C. 2010. Effect of WT1 gene silencing on the tumorigenicity of human glioblastoma multiforme cells. *Journal of neurosurgery,* 112**,** 18-25.

COHEN, A. L., HOLMEN, S. L. & COLMAN, H. 2013. IDH1 and IDH2 mutations in gliomas. *Current neurology and neuroscience reports,* 13**,** 345.

CONDLIFFE, D., WONG, A., TROAKES, C., PROITSI, P., PATEL, Y.,

CHOULIARAS, L., FERNANDES, C., COOPER, J., LOVESTONE, S. & SCHALKWYK, L. 2014. Cross-region reduction in 5-hydroxymethylcytosine in Alzheimer's disease brain. *Neurobiology of aging,* 35**,** 1850-1854.

COONS, S. W., JOHNSON, P. C., SCHEITHAUER, B. W., YATES, A. J. & PEARL, D. K. 1997. Improving diagnostic accuracy and interobserver concordance in the classification and grading of primary gliomas. *Cancer: Interdisciplinary International Journal of the American Cancer Society,* 79**,** 1381-1393.

COSKUN, V., TSOA, R. & SUN, Y. E. 2012. Epigenetic regulation of stem cells differentiating along the neural lineage. *Current opinion in neurobiology,* 22**,** 762-767.

CRAWFORD, D. J., LIU, M. Y., NABEL, C. S., CAO, X.-J., GARCIA, B. A. & KOHLI, R. M. 2016. Tet2 catalyzes stepwise 5-methylcytosine oxidation by an iterative and de novo mechanism. *Journal of the American Chemical Society,* 138**,** 730-733.

DALLOSSO, A. R., HANCOCK, A. L., BROWN, K. W., WILLIAMS, A. C., JACKSON, S. & MALIK, K. 2004. Genomic imprinting at the WT1 gene involves a novel coding transcript (AWT1) that shows deregulation in Wilms' tumours. *Human molecular genetics,* 13**,** 405-415.

DANG, L., WHITE, D. W., GROSS, S., BENNETT, B. D., BITTINGER, M. A., DRIGGERS, E. M., FANTIN, V. R., JANG, H. G., JIN, S. & KEENAN, M. C. 2009. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature,* 462**,** 739.

DAVIES, R. C., BRATT, E. & HASTIE, N. D. 2000. Did nucleotides or

amino acids drive evolutionary conservation of the WT1±KTS alternative splice? *Human molecular genetics,* 9**,** 1177-1183.

DAVIS, T. & VAISVILA, R. 2011. High sensitivity 5-hydroxymethylcytosine detection in Balb/C brain tissue. *JoVE (Journal of Visualized Experiments)***,** e2661-e2661.

DAWLATY, M. M., BREILING, A., LE, T., BARRASA, M. I., RADDATZ, G., GAO, Q., POWELL, B. E., CHENG, A. W., FAULL, K. F. & LYKO, F. 2014. Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Developmental cell,* 29**,** 102-111.

DAWLATY, M. M., BREILING, A., LE, T., RADDATZ, G., BARRASA, M. I., CHENG, A. W., GAO, Q., POWELL, B. E., LI, Z. & XU, M. 2013. Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Developmental cell,* 24**,** 310-323.

DAWLATY, M. M., GANZ, K., POWELL, B. E., HU, Y.-C., MARKOULAKI, S., CHENG, A. W., GAO, Q., KIM, J., CHOI, S.-W. & PAGE, D. C. 2011. Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell stem cell,* 9**,** 166-175.

DAWSON, M. A. & KOUZARIDES, T. 2012. Cancer epigenetics: from mechanism to therapy. *Cell,* 150**,** 12-27.

DEAN, W., SANTOS, F. & REIK, W. Epigenetic reprogramming in early mammalian development and following somatic nuclear transfer. Seminars in cell & developmental biology, 2003. Elsevier, 93-100.

DEATON, A. M. & BIRD, A. 2011. CpG islands and the regulation of

transcription. *Genes Dev,* 25**,** 1010-22.

DENIZ, Ö., FROST, J. M. & BRANCO, M. R. 2019. Regulation of transposable elements by DNA modifications. *Nature Reviews Genetics***,** 1.

DENSLOW, S. & WADE, P. 2007. The human Mi-2/NuRD complex and gene regulation. *Oncogene,* 26**,** 5433-5438.

DI STEFANO, B., COLLOMBET, S., JAKOBSEN, J. S., WIERER, M., SARDINA, J. L., LACKNER, A., STADHOUDERS, R., SEGURA-MORALES, C., FRANCESCONI, M. & LIMONE, F. 2016. C/EBPα creates elite cells for iPSC reprogramming by upregulating Klf4 and increasing the levels of Lsd1 and Brd4. *Nature cell biology,* 18**,** 371.

DI STEFANO, B., SARDINA, J. L., VAN OEVELEN, C., COLLOMBET, S., KALLIN, E. M., VICENT, G. P., LU, J., THIEFFRY, D., BEATO, M. & GRAF, T. 2014. C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. *Nature,* 506**,** 235.

DIOTEL, N., MÉROT, Y., COUMAILLEAU, P., GUEGUEN, M. M., SÉRANDOUR, A. A., SALBERT, G. & KAH, O. 2017. 5-hydroxymethylcytosine marks postmitotic neural cells in the adult and developing vertebrate central nervous system. *Journal of Comparative Neurology,* 525**,** 478-497.

DOEGE, C. A., INOUE, K., YAMASHITA, T., RHEE, D. B., TRAVIS, S., FUJITA, R., GUARNIERI, P., BHAGAT, G., VANTI, W. B. & SHIH, A. 2012. Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature,* 488**,** 652.

DRESSLER, G. R. 2009. Advances in early kidney specification,

development and patterning. *Development,* 136**,** 3863-3874.

DRESSLER, G. R., DEUTSCH, U., CHOWDHURY, K., NORNES, H. O. & GRUSS, P. 1990. Pax2, a new murine paired-box-containing gene and its expression in the developing excretory system. *Development,* 109**,** 787-795.

DUNCAN, B. K. & MILLER, J. H. 1980. Mutagenic deamination of cytosine residues in DNA.

DUNCAN, C. G., BARWICK, B. G., JIN, G., RAGO, C., KAPOOR-VAZIRANI, P., POWELL, D. R., CHI, J.-T., BIGNER, D. D., VERTINO, P. M. & YAN, H. 2012. A heterozygous IDH1R132H/WT mutation induces genome-wide alterations in DNA methylation. *Genome research,* 22**,** 2339-2355.

ECKHARDT, F., LEWIN, J., CORTESE, R., RAKYAN, V. K., ATTWOOD, J., BURGER, M., BURTON, J., COX, T. V., DAVIES, R. & DOWN, T. A. 2006. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nature genetics,* 38**,** 1378.

EHRLICH, M. 2002. DNA methylation in cancer: too much, but also too little. *Oncogene,* 21**,** 5400-5413.

EHRLICH, M., GAMA-SOSA, M. A., HUANG, L.-H., MIDGETT, R. M., KUO, K. C., MCCUNE, R. A. & GEHRKE, C. 1982. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic acids research,* 10**,** 2709-2721.

EHRLICH, M. & LACEY, M. 2013. DNA hypomethylation and hemimethylation in cancer. *Epigenetic Alterations in Oncogenesis.* Springer.

ELEFTHERIOU, M., PASCUAL, A. J., WHELDON, L. M., PERRY, C., ABAKIR, A., ARORA, A., JOHNSON, A. D., AUER, D. T., ELLIS, I. O. & MADHUSUDAN, S. 2015. 5-Carboxylcytosine levels are elevated in human breast cancers and gliomas. *Clinical epigenetics,* 7**,** 88.

ENGLERT, C., VIDAL, M., MAHESWARAN, S., GE, Y., EZZELL, R. M., ISSELBACHER, K. J. & HABER, D. A. 1995. Truncated WT1 mutants alter the subnuclear localization of the wild-type protein. *Proceedings of the National Academy of Sciences,* 92**,** 11960-11964.

ERHART, F., BLAUENSTEINER, B., ZIRKOVITS, G., PRINTZ, D., SOUKUP, K., KLINGENBRUNNER, S., FISCHHUBER, K., REITERMAIER, R., HALFMANN, A. & LÖTSCH, D. 2019. Gliomasphere marker combinatorics: multidimensional flow cytometry detects CD 44+/CD 133+/ITGA 6+/CD 36+ signature. *Journal of cellular and molecular medicine,* 23**,** 281-292.

ESTELLER, M., TORTOLA, S., TOYOTA, M., CAPELLA, G., PEINADO, M. A., BAYLIN, S. B. & HERMAN, J. G. 2000. Hypermethylation-associated inactivation of p14ARF is independent of p16INK4a methylation and p53 mutational status. *Cancer research,* 60**,** 129-133.

FAN, G., MARTINOWICH, K., CHIN, M. H., HE, F., FOUSE, S. D., HUTNICK, L., HATTORI, D., GE, W., SHEN, Y. & WU, H. 2005. DNA methylation controls the timing of astrogliogenesis through regulation of JAK-STAT signaling. *Development,* 132**,** 3345-3356.

FATEMI, M., HERMANN, A., GOWHER, H. & JELTSCH, A. 2002. Dnmt3a and Dnmt1 functionally cooperate during de novo methylation

of DNA. *European Journal of Biochemistry,* 269**,** 4981-4984.

FERNANDEZ, A. F., BAYÓN, G. F., SIERRA, M. I., URDINGUIO, R. G., TORAÑO, E. G., GARCÍA, M. G., CARELLA, A., LÓPEZ, V., SANTAMARINA, P. & PÉREZ, R. F. 2018. Loss of 5hmC identifies a new type of aberrant DNA hypermethylation in glioma. *Human molecular genetics,* 27**,** 3046-3059.

FICZ, G., BRANCO, M. R., SEISENBERGER, S., SANTOS, F., KRUEGER, F., HORE, T. A., MARQUES, C. J., ANDREWS, S. & REIK, W. 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature,* 473**,** 398-402.

FIGUEROA, M. E., ABDEL-WAHAB, O., LU, C., WARD, P. S., PATEL, J., SHIH, A., LI, Y., BHAGWAT, N., VASANTHAKUMAR, A. & FERNANDEZ, H. F. 2010. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer cell,* 18**,** 553-567.

FILION, G. J., ZHENILO, S., SALOZHIN, S., YAMADA, D., PROKHORTCHOUK, E. & DEFOSSEZ, P.-A. 2006. A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Molecular and cellular biology,* 26**,** 169-181.

FILLMORE, C. M. & KUPERWASSER, C. 2008. Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. *Breast cancer research,* 10**,** R25.

FISHER, C. L. & FISHER, A. G. 2011. Chromatin states in pluripotent, differentiated, and reprogrammed cells. *Current opinion in genetics &*

development, 21**,** 140-146.

FISHER, D. H. & GIESE, R. W. 1988. Determination of 5-methylcytosine in DNA by gas chromatography—electron-capture detection. *Journal of Chromatography A,* 452**,** 51-60.

FOURNIER, A., SASAI, N., NAKAO, M. & DEFOSSEZ, P.-A. 2012. The role of methyl-binding proteins in chromatin organization and epigenome maintenance. *Briefings in functional genomics,* 11**,** 251-264.

FRANCHINI, D.-M., SCHMITZ, K.-M. & PETERSEN-MAHRT, S. K. 2012. 5-Methylcytosine DNA demethylation: more than losing a methyl group. *Annual review of genetics,* 46**,** 419-441.

FROMMER, M., MCDONALD, L. E., MILLAR, D. S., COLLIS, C. M., WATT, F., GRIGG, G. W., MOLLOY, P. L. & 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,* 89**,** 1827-1831.

FU, Y., DOMINISSINI, D., RECHAVI, G. & HE, C. 2014. Gene expression regulation mediated through reversible m 6 A RNA methylation. *Nature Reviews Genetics,* 15**,** 293.

FUKS, F. 2005. DNA methylation and histone modifications: teaming up to silence genes. *Current opinion in genetics & development,* 15**,** 490-495.

GACKOWSKI, D., STARCZAK, M., ZARAKOWSKA, E., MODRZEJEWSKA, M., SZPILA, A., BANASZKIEWICZ, Z. & 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**,** 12128-12136.

GACKOWSKI, D., ZARAKOWSKA, E., STARCZAK, M., MODRZEJEWSKA, M. & OLINSKI, R. 2015. Tissue-specific differences in DNA modifications (5-hydroxymethylcytosine, 5-formylcytosine, 5-carboxylcytosine and 5-hydroxymethyluracil) and their interrelationships. *PLoS One,* 10**,** e0144859.

GAIANO, N., NYE, J. S. & FISHELL, G. 2000. Radial glial identity is promoted by Notch1 signaling in the murine forebrain. *Neuron,* 26**,** 395-404.

GAJ, T., GERSBACH, C. A. & BARBAS III, C. F. 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology,* 31**,** 397-405.

GALLEGO-BARTOLOMÉ, J., GARDINER, J., LIU, W., PAPIKIAN, A., GHOSHAL, B., KUO, H. Y., ZHAO, J. M.-C., SEGAL, D. J. & JACOBSEN, S. E. 2018. Targeted DNA demethylation of the Arabidopsis genome using the human TET1 catalytic domain. *Proceedings of the National Academy of Sciences,* 115**,** E2125-E2134.

GAMBICHLER, T., SAND, M. & SKRYGAN, M. 2013. Loss of 5-hydroxymethylcytosine and ten-eleven translocation 2 protein expression in malignant melanoma. *Melanoma research,* 23**,** 218-220.

GAO, Y., CHEN, J., LI, K., WU, T., HUANG, B., LIU, W., KOU, X., ZHANG, Y., HUANG, H. & JIANG, Y. 2013. Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA

methylation and hydroxymethylation in reprogramming. *Cell stem cell,* 12**,** 453-469.

GAO, Y., LI, J., CHEN, B., YANG, Q. & ZHANG, J. 2017. Detection of 5-Carboxyl Cytosine in Mammalian Cells DNA. *Gene Technol,* 6**,** 2.

GARIBOLDI, M. B., RAVIZZA, R. & MONTI, E. 2010. The IGFR1 inhibitor NVP-AEW541 disrupts a pro-survival and pro-angiogenic IGF-STAT3-HIF1 pathway in human glioblastoma cells. *Biochemical pharmacology,* 80**,** 455-462.

GERMANGUZ, I., PARK, J. C., CINKORNPUMIN, J., SOLOMON, A., OHASHI, M. & LOWRY, W. E. 2018. TDG regulates cell cycle progression in human neural progenitors. *F1000Research,* 7.

GKOUNTELA, S., ZHANG, K. X., SHAFIQ, T. A., LIAO, W.-W., HARGAN-CALVOPIÑA, J., CHEN, P.-Y. & CLARK, A. T. 2015. DNA demethylation dynamics in the human prenatal germline. *Cell,* 161**,** 1425-1436.

GLOBISCH, D., MÜNZEL, M., MÜLLER, M., MICHALAKIS, S., WAGNER, M., KOCH, S., BRÜCKL, T., BIEL, M. & CARELL, T. 2010. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PloS one,* 5**,** e15367.

GLOWACKA, W. K., JAIN, H., OKURA, M., MAIMAITIMING, A., MAMATJAN, Y., NEJAD, R., FAROOQ, H., TAYLOR, M. D., ALDAPE, K. & KONGKHAM, P. 2018. 5-Hydroxymethylcytosine preferentially targets genes upregulated in isocitrate dehydrogenase 1 mutant high-grade glioma. *Acta neuropathologica,* 135**,** 617-634.

GOLEBIEWSKA, A., BOUGNAUD, S., STIEBER, D., BRONS, N. H.,

VALLAR, L., HERTEL, F., KLINK, B., SCHRÖCK, E., BJERKVIG, R. & NICLOU, S. P. 2013. Side population in human glioblastoma is non-tumorigenic and characterizes brain endothelial cells. *Brain,* 136**,** 1462-1475.

GOLLA, J. P., ZHAO, J., MANN, I. K., SAYEED, S. K., MANDAL, A., ROSE, R. B. & VINSON, C. 2014. Carboxylation of cytosine (5caC) in the CG dinucleotide in the E-box motif (CGCAG| GTG) increases binding of the Tcf3| Ascl1 helix-loop-helix heterodimer 10-fold. *Biochemical and biophysical research communications,* 449**,** 248-255.

GRUENBAUM, Y., CEDAR, H. & RAZIN, A. 1981. Restriction enzyme digestion of hemimethylated DNA. *Nucleic acids research,* 9**,** 2509-2515.

GRUENBAUM, Y., CEDAR, H. & RAZIN, A. 1982. Substrate and sequence specificity of a eukaryotic DNA methylase.

GU, T.-P., GUO, F., YANG, H., WU, H.-P., XU, G.-F., LIU, W., XIE, Z.-G., SHI, L., HE, X. & JIN, S.-G. 2011. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature,* 477**,** 606-610.

GUO, F., LI, X., LIANG, D., LI, T., ZHU, P., GUO, H., WU, X., WEN, L., GU, T.-P. & HU, B. 2014. Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell stem cell,* 15**,** 447-458.

GUO, J. U., SU, Y., ZHONG, C., MING, G.-L. & SONG, H. 2011. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell,* 145**,** 423-434.

GUO, M., LI, X., ZHANG, L., LIU, D., DU, W., YIN, D., LYU, N., ZHAO, G., GUO, C. & 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 of flight mass spectrometry analysis. *Oncotarget,* 8**,** 91248.

GUY, J., HENDRICH, B., HOLMES, M., MARTIN, J. E. & BIRD, A. 2001. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nature genetics,* 27**,** 322-326.

HABER, D. A., BUCKLER, A. J., GLASER, T., CALL, K. M., PELLETIER, J., SOHN, R. L., DOUGLASS, E. C. & HOUSMAN, D. E. 1990. An internal deletion within an 11p13 zinc finger gene contributes to the development of Wilms' tumor. *Cell,* 61**,** 1257-1269.

HACKETT, J. A., SENGUPTA, R., ZYLICZ, J. J., MURAKAMI, K., LEE, C., DOWN, T. A. & SURANI, M. A. 2013. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science,* 339**,** 448-452.

HACKETT, J. A. & SURANI, M. A. 2013. DNA methylation dynamics during the mammalian life cycle. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 368**,** 20110328.

HAFFNER, M. C., CHAUX, A., MEEKER, A. K., ESOPI, D. M., GERBER, J., PELLAKURU, L. G., TOUBAJI, A., ARGANI, P., IACOBUZIO-DONAHUE, C. & NELSON, W. G. 2011. Global 5-hydroxymethylcytosine content is significantly reduced in tissue

stem/progenitor cell compartments and in human cancers. *Oncotarget,* 2**,** 627.

HAHN, M. A., QIU, R., WU, X., LI, A. X., ZHANG, H., WANG, J., JUI, J., JIN, S.-G., JIANG, Y. & PFEIFER, G. P. 2013. Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. *Cell reports,* 3**,** 291-300.

HARA, M. & FUKAGAWA, T. 2019. Centromere maintenance during DNA replication. *Nature cell biology,* 21**,** 669.

HARRINGTON, M. A., JONES, P. A., IMAGAWA, M. & KARIN, M. 1988. Cytosine methylation does not affect binding of transcription factor Sp1. *Proceedings of the National Academy of Sciences,* 85**,** 2066-2070.

HARTMANN, C., MEYER, J., BALSS, J., CAPPER, D., MUELLER, W., CHRISTIANS, A., FELSBERG, J., WOLTER, M., MAWRIN, C. & WICK, W. 2009. Type and frequency of IDH1 and IDH2 mutations are related to astrocytic and oligodendroglial differentiation and age: a study of 1,010 diffuse gliomas. *Acta neuropathologica,* 118**,** 469-474.

HASHIMOTO, H., HORTON, J. R., ZHANG, X. & CHENG, X. 2009. UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. *Epigenetics,* 4**,** 8-14.

HASHIMOTO, H., OLANREWAJU, Y. O., ZHENG, Y., WILSON, G. G., ZHANG, X. & CHENG, X. 2014. Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. *Genes & development,* 28**,** 2304-2313.

HASHIMOTO, H., ZHANG, X., VERTINO, P. M. & CHENG, X. 2015. The mechanisms of generation, recognition, and erasure of DNA 5-methylcytosine and thymine oxidations. *Journal of Biological Chemistry,* 290**,** 20723-20733.

HASHIMOTO, H., ZHANG, X., ZHENG, Y., WILSON, G. G. & CHENG, X. 2016. Denys-Drash syndrome associated WT1 glutamine 369 mutants have altered sequence-preferences and altered responses to epigenetic modifications. *Nucleic acids research,* 44**,** 10165-10176.

HASHIMOTO, N., TSUBOI, A., CHIBA, Y., IZUMOTO, S., OKA, Y., YOSHIMINE, T. & SUGIYAMA, H. 2009. Immunotherapy targeting the Wilms' tumor 1 gene product for patients with malignant brain tumors. *Brain and nerve= Shinkei kenkyu no shinpo,* 61**,** 805-814.

HASHIMSHONY, T., ZHANG, J., KESHET, I., BUSTIN, M. & CEDAR, H. 2003. The role of DNA methylation in setting up chromatin structure during development. *Nature genetics,* 34**,** 187-192.

HASTIE, N. D. 2017. Wilms' tumour 1 (WT1) in development, homeostasis and disease. *Development,* 144**,** 2862-2872.

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**,** 1983-1993.

HE, S., WANG, F., ZHANG, Y., CHEN, J., LIANG, L., LI, Y., ZHANG, M., YANG, X., PANG, H. & LI, Y. 2019. Hemi-methylated CpG sites connect Dnmt1-knockdown-induced and Tet1-induced DNA demethylation during somatic cell reprogramming. *Cell discovery,* 5**,** 11.

HE, Y.-F., LI, B.-Z., LI, Z., LIU, P., WANG, Y., TANG, Q., DING, J., JIA,

Y., CHEN, Z. & LI, L. 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science,* 333**,** 1303-1307.

HEGGE, B., SJØTTEM, E. & MIKKOLA, I. 2018. Generation of a PAX6 knockout glioblastoma cell line with changes in cell cycle distribution and sensitivity to oxidative stress. *BMC cancer,* 18**,** 496.

HEMBERGER, M., DEAN, W. & REIK, W. 2009. Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. *Nature reviews Molecular cell biology,* 10**,** 526.

HERMAN, J. G., GRAFF, J. R., MYÖHÄNEN, S., NELKIN, B. D. & BAYLIN, S. B. 1996. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proceedings of the National Academy of Sciences,* 93**,** 9821-9826.

HERMAN, J. G., MERLO, A., MAO, L., LAPIDUS, R. G., ISSA, J.-P. J., DAVIDSON, N. E., SIDRANSKY, D. & BAYLIN, S. B. 1995. Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. *Cancer research,* 55**,** 4525-4530.

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**,** 48350-48359.

HERZER, U., CROCOLL, A., BARTON, D., HOWELLS, N. & ENGLERT, C. 1999. The Wilms tumor suppressor gene wt1 is required for development of the spleen. *Current biology,* 9**,** 837-S1.

HINOUE, T., WEISENBERGER, D. J., LANGE, C. P., SHEN, H., BYUN, H.-M., VAN DEN BERG, D., MALIK, S., PAN, F., NOUSHMEHR, H. & VAN DIJK, C. M. 2012. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome research,* 22**,** 271-282.

HOHENSTEIN, P. & HASTIE, N. D. 2006. The many facets of the Wilms' tumour gene, WT1. *Human molecular genetics,* 15**,** R196-R201.

HONG, X.-Y., WANG, J. & LI, Z. 2013. AGR2 expression is regulated by HIF-1 and contributes to growth and angiogenesis of glioblastoma. *Cell biochemistry and biophysics,* 67**,** 1487-1495.

HOTCHKISS, R. D. 1948. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *Journal of Biological Chemistry,* 175**,** 315-332.

HSU, P. D., LANDER, E. S. & ZHANG, F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell,* 157**,** 1262-1278.

HU, X., ZHANG, L., MAO, S.-Q., LI, Z., CHEN, J., ZHANG, R.-R., WU, H.-P., GAO, J., GUO, F. & LIU, W. 2014. Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell stem cell,* 14**,** 512-522.

HUANG, H., JIANG, X., LI, Z., LI, Y., SONG, C.-X., HE, C., SUN, M., CHEN, P., GURBUXANI, S. & WANG, J. 2013. TET1 plays an essential oncogenic role in MLL-rearranged leukemia. *Proceedings of the National Academy of Sciences,* 110**,** 11994-11999.

HUANG, L., MOKKAPATI, S., HU, Q., RUTESHOUSER, E. C., HICKS,

M. J. & HUFF, V. 2016. Nephron Progenitor But Not Stromal Progenitor Cells Give Rise to Wilms Tumors in Mouse Models with β-Catenin Activation or Wt1 Ablation and Igf2 Upregulation. *Neoplasia,* 18**,** 71-81.

HUANG, Y., WANG, G., LIANG, Z., YANG, Y., CUI, L. & LIU, C.-Y. 2016. Loss of nuclear localization of TET2 in colorectal cancer. *Clinical epigenetics,* 8**,** 9.

HUGHES, L. A., MELOTTE, V., DE SCHRIJVER, J., DE MAAT, M., SMIT, V. T., BOVÉE, J. V., FRENCH, P. J., VAN DEN BRANDT, P. A., SCHOUTEN, L. J. & DE MEYER, T. 2013. The CpG island methylator phenotype: what's in a name? *Cancer research*.

HUTNICK, L. K., GOLSHANI, P., NAMIHIRA, M., XUE, Z., MATYNIA, A., YANG, X. W., SILVA, A. J., SCHWEIZER, F. E. & FAN, G. 2009. DNA hypomethylation restricted to the murine forebrain induces cortical degeneration and impairs postnatal neuronal maturation. *Human molecular genetics,* 18**,** 2875-2888.

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**,** 1670.

INOUE, K., TAMAKI, H., OGAWA, H., OKA, Y., SOMA, T., TATEKAWA, T., OJI, Y., TSUBOI, A., KIM, E. H. & KAWAKAMI, M. 1998. Wilms9 Tumor Gene (WT1) Competes With Differentiation-Inducing Signal in Hematopoietic Progenitor Cells. *Blood,* 91**,** 2969-2976.

INUI, M., MIYADO, M., IGARASHI, M., TAMANO, M., KUBO, A., YAMASHITA, S., ASAHARA, H., FUKAMI, M. & TAKADA, S. 2014.

Rapid generation of mouse models with defined point mutations by the CRISPR/Cas9 system. *Scientific reports,* 4**,** 5396.

IQBAL, K., JIN, S.-G., PFEIFER, G. P. & SZABÓ, 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**,** 3642-3647.

ITO, S., D'ALESSIO, A. C., TARANOVA, O. V., HONG, K., SOWERS, L. C. & ZHANG, Y. 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature,* 466**,** 1129.

ITO, S., SHEN, L., DAI, Q., WU, S. C., COLLINS, L. B., SWENBERG, J. A., HE, C. & ZHANG, Y. 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science,* 333**,** 1300-1303.

IZUMOTO, S., TSUBOI, A., OKA, Y., SUZUKI, T., HASHIBA, T., KAGAWA, N., HASHIMOTO, N., MARUNO, M., ELISSEEVA, O. A. & SHIRAKATA, T. 2008. Phase II clinical trial of Wilms tumor 1 peptide vaccination for patients with recurrent glioblastoma multiforme.

JACKSON, A. L., BARTZ, S. R., SCHELTER, J., KOBAYASHI, S. V., BURCHARD, J., MAO, M., LI, B., CAVET, G. & LINSLEY, P. S. 2003. Expression profiling reveals off-target gene regulation by RNAi. *Nature biotechnology,* 21**,** 635.

JACOBS, A. L. & SCHÄR, P. 2012. DNA glycosylases: in DNA repair and beyond. *Chromosoma,* 121**,** 1-20.

JARMASZ, J. S., STIRTON, H., BASALAH, D., DAVIE, J. R.,

CLARREN, S. K., ASTLEY, S. J. & DEL BIGIO, M. R. 2019. Global DNA Methylation and Histone Posttranslational Modifications in Human and Nonhuman Primate Brain in Association with Prenatal Alcohol Exposure. *Alcoholism: Clinical and Experimental Research,* 43**,** 1145-1162.

JIANG, F. & DOUDNA, J. A. 2017. CRISPR–Cas9 structures and mechanisms. *Annual review of biophysics,* 46**,** 505-529.

JIN, S.-G., JIANG, Y., QIU, R., RAUCH, T. A., WANG, Y., SCHACKERT, G., KREX, D., LU, Q. & PFEIFER, G. P. 2011. 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer research,* 71**,** 7360-7365.

JIN, S.-G., WU, X., LI, A. X. & PFEIFER, G. P. 2011. Genomic mapping of 5-hydroxymethylcytosine in the human brain. *Nucleic acids research***,** gkr120.

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**,** 816-821.

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**,** 13177.

JONES, P. L., VEENSTRA, G. C. J., WADE, P. A., VERMAAK, D., KASS, S. U., LANDSBERGER, N., STROUBOULIS, J. & WOLFFE, A.

P. 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature genetics,* 19**,** 187-191.

KANEKO, Y., OKITA, H., HARUTA, M., ARAI, Y., OUE, T., TANAKA, Y., HORIE, H., HINOTSU, S., KOSHINAGA, T. & YONEDA, A. 2015. A high incidence of WT1 abnormality in bilateral Wilms tumours in Japan, and the penetrance rates in children with WT1 germline mutation. *British journal of cancer,* 112**,** 1121.

KANG, Y.-J., BALTER, B., CSIZMADIA, E., HAAS, B., SHARMA, H., BRONSON, R. & YAN, C. T. 2017. Contribution of classical end-joining to PTEN inactivation in p53-mediated glioblastoma formation and drug-resistant survival. *Nature communications,* 8**,** 14013.

KASS, S. U., PRUSS, D. & WOLFFE, A. P. 1997. How does DNA methylation repress transcription? *Trends in Genetics,* 13**,** 444-449.

KELLINGER, M. W., SONG, C.-X., CHONG, J., LU, X.-Y., HE, C. & WANG, D. 2012. 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nature structural & molecular biology,* 19**,** 831-833.

KELLY, A. D., KROEGER, H., YAMAZAKI, J., TABY, R., NEUMANN, F., YU, S., LEE, J. T., PATEL, B., LI, Y. & HE, R. 2017. A CpG island methylator phenotype in acute myeloid leukemia independent of IDH mutations and associated with a favorable outcome. *Leukemia*.

KESHET, I., LIEMAN-HURWITZ, J. & CEDAR, H. 1986. DNA methylation affects the formation of active chromatin. *Cell,* 44**,** 535-543.

KHARE, T., PAI, S., KONCEVICIUS, K., PAL, M., KRIUKIENE, E., LIUTKEVICIUTE, Z., IRIMIA, M., JIA, P., PTAK, C. & XIA, M. 2012. 5-

hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nature structural & molecular biology,* 19**,** 1037-1043.

KIJIMA, N., HASHIMOTO, N., CHIBA, Y., FUJIMOTO, Y., SUGIYAMA, H. & YOSHIMINE, T. 2016. Functional roles of Wilms' tumor 1 (WT1) in malignant brain tumors. *Codon Publications***,** 261-273.

KIM, G. D., NI, J., KELESOGLU, N., ROBERTS, R. J. & PRADHAN, S. 2002. Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. *The EMBO journal,* 21**,** 4183-4195.

KIM, G.-H., CHOI, S. Y., OH, T.-I., KAN, S.-Y., KANG, H., LEE, S., OH, T., KO, H. M. & LIM, J.-H. 2019. IDH1R132H Causes Resistance to HDAC Inhibitors by Increasing NANOG in Glioblastoma Cells. *International journal of molecular sciences,* 20**,** 2679.

KIM, H. S., TAN, Y., MA, W., MERKURJEV, D., DESTICI, E., MA, Q., SUTER, T., OHGI, K., FRIEDMAN, M. & SKOWRONSKA-KRAWCZYK, D. 2018. Pluripotency factors functionally premark cell-type-restricted enhancers in ES cells. *Nature,* 556**,** 510.

KIM, M. K., MASON, J. M., LI, C.-M., BERKOFSKY-FESSLER, W., JIANG, L., CHOUBEY, D., GRUNDY, P. E., TYCKO, B. & LICHT, J. D. 2008. A pathologic link between Wilms tumor suppressor gene, WT1, and IFI16. *Neoplasia (New York, NY),* 10**,** 69.

KLEIHUES, P., BURGER, P. C. & SCHEITHAUER, B. W. 1993. The new WHO classification of brain tumours. *Brain pathology,* 3**,** 255-268.

KLOSE, R. J. & BIRD, A. P. 2006. Genomic DNA methylation: the mark

and its mediators. *Trends in biochemical sciences,* 31**,** 89-97.

KO, M., AN, J., BANDUKWALA, H. S., CHAVEZ, L., ÄIJÖ, T., PASTOR, W. A., SEGAL, M. F., LI, H., KOH, K. P. & LÄHDESMÄKI, H. 2013. Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature,* 497**,** 122-126.

KO, M., HUANG, Y., JANKOWSKA, A. M., PAPE, U. J., TAHILIANI, M., BANDUKWALA, H. S., AN, J., LAMPERTI, E. D., KOH, K. P. & GANETZKY, R. 2010. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature,* 468**,** 839-843.

KOH, K. P., YABUUCHI, A., RAO, S., HUANG, Y., CUNNIFF, K., NARDONE, J., LAIHO, A., TAHILIANI, M., SOMMER, C. A. & MOSTOSLAVSKY, G. 2011. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell stem cell,* 8**,** 200-213.

KOHLI, R. M. & ZHANG, Y. 2013. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature,* 502**,** 472.

KONSTANDIN, N., BULTMANN, S., SZWAGIERCZAK, A., DUFOUR, A., KSIENZYK, B., SCHNEIDER, F., HEROLD, T., MULAW, M., KAKADIA, P. & SCHNEIDER, S. 2011. Genomic 5-hydroxymethyicytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. *Leukemia,* 25**,** 1649-1653.

KRAUS, T. F., GLOBISCH, D., WAGNER, M., EIGENBROD, S., WIDMANN, D., MÜNZEL, M., MÜLLER, M., PFAFFENEDER, T., HACKNER, B. & FEIDEN, W. 2012. Low values of 5-

hydroxymethylcytosine (5hmC), the “sixth base,” are associated with anaplasia in human brain tumors. *International journal of cancer,* 131**,** 1577-1590.

KRAUS, T. F., KOLCK, G., GREINER, A., SCHIERL, K., GUIBOURT, V. & KRETZSCHMAR, H. A. 2015. Loss of 5-hydroxymethylcytosine and intratumoral heterogeneity as an epigenomic hallmark of glioblastoma. *Tumor Biology,* 36**,** 8439-8446.

KRIAUCIONIS, S. & HEINTZ, N. 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science,* 324**,** 929-930.

KRIEGSTEIN, A. & ALVAREZ-BUYLLA, A. 2009. The glial nature of embryonic and adult neural stem cells. *Annual review of neuroscience,* 32**,** 149-184.

KRISTENSEN, G., STRAND, S. H., RØDER, M. A., BERG, K. D., TOFT, B. G., HØYER, S., BORRE, M., SØRENSEN, K. D. & BRASSO, K. 2019. 5hmC Level Predicts Biochemical Failure Following Radical Prostatectomy in Prostate Cancer Patients with ERG Negative Tumors. *International journal of molecular sciences,* 20**,** 1025.

KUDO, Y., TATEISHI, K., YAMAMOTO, K., YAMAMOTO, S., ASAOKA, Y., IJICHI, H., NAGAE, G., YOSHIDA, H., ABURATANI, H. & KOIKE, K. 2012. Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. *Cancer science,* 103**,** 670-676.

LADOMERY, M. & DELLAIRE, G. 2002. Multifunctional zinc finger proteins in development and disease. *Annals of human genetics,* 66**,** 331-342.

LAFFAIRE, J., EVERHARD, S., IDBAIH, A., CRINIÈRE, E., MARIE, Y., DE REYNIÈS, A., SCHIAPPA, R., MOKHTARI, K., HOANG-XUAN, K. & SANSON, M. 2010. Methylation profiling identifies 2 groups of gliomas according to their tumorigenesis. *Neuro-oncology,* 13**,** 84-98.

LAITY, J. H., DYSON, H. J. & WRIGHT, P. E. 2000. Molecular basis for modulation of biological function by alternate splicing of the Wilms' tumor suppressor protein. *Proceedings of the National Academy of Sciences,* 97**,** 11932-11935.

LANE, N., DEAN, W., ERHARDT, S., HAJKOVA, P., SURANI, A., WALTER, J. & REIK, W. 2003. Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *genesis,* 35**,** 88-93.

LANGEMEIJER, S. M., KUIPER, R. P., BERENDS, M., KNOPS, R., ASLANYAN, M. G., MASSOP, M., STEVENS-LINDERS, E., VAN HOOGEN, P., VAN KESSEL, A. G. & RAYMAKERS, R. A. 2009. Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nature genetics,* 41**,** 838.

LARSON, M. H., GILBERT, L. A., WANG, X., LIM, W. A., WEISSMAN, J. S. & QI, L. S. 2013. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature protocols,* 8**,** 2180.

LARSSON, S. H., CHARLIEU, J.-P., MIYAGAWA, K., ENGELKAMP, D., RASSOULZADEGAN, M., ROSS, A., CUZIN, F., VAN HEYNINGEN, V. & HASTIE, N. D. 1995. Subnuclear localization of WT1 in splicing or transcription factor domains is regulated by alternative splicing. *Cell,* 81**,** 391-401.

LATHIA, J. D., GALLAGHER, J., HEDDLESTON, J. M., WANG, J., EYLER, C. E., MACSWORDS, J., WU, Q., VASANJI, A., MCLENDON, R. E. & HJELMELAND, A. B. 2010. Integrin alpha 6 regulates glioblastoma stem cells. *Cell stem cell,* 6**,** 421-432.

LATHIA, J. D., MACK, S. C., MULKEARNS-HUBERT, E. E., VALENTIM, C. L. & RICH, J. N. 2015. Cancer stem cells in glioblastoma. *Genes & development,* 29**,** 1203-1217.

LAWRENCE, M. S., STOJANOV, P., POLAK, P., KRYUKOV, G. V., CIBULSKIS, K., SIVACHENKO, A., CARTER, S. L., STEWART, C., MERMEL, C. H. & ROBERTS, S. A. 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature,* 499**,** 214.

LE, T., KIM, K.-P., FAN, G. & FAULL, K. F. 2011. A sensitive mass spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples. *Analytical biochemistry,* 412**,** 203-209.

LEE, J., KOTLIAROVA, S., KOTLIAROV, Y., LI, A., SU, Q., DONIN, N. M., PASTORINO, S., PUROW, B. W., CHRISTOPHER, N. & ZHANG, W. 2006. Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer cell,* 9**,** 391-403.

LEE, K., HAMM, J., WHITWORTH, K., SPATE, L., PARK, K.-W., MURPHY, C. N. & PRATHER, R. S. 2014. Dynamics of TET family expression in porcine preimplantation embryos is related to zygotic

genome activation and required for the maintenance of NANOG. *Developmental biology,* 386**,** 86-95.

LEES-MURDOCK, D. J., LAU, H.-T., CASTRILLON, D. H., DE FELICI, M. & WALSH, C. P. 2008. DNA methyltransferase loading, but not de novo methylation, is an oocyte-autonomous process stimulated by SCF signalling. *Developmental biology,* 321**,** 238-250.

LEIGHTON, G. & WILLIAMS JR, D. C. 2019. The Methyl-CpG Binding Domain 2 and 3 proteins and formation of the Nucleosome Remodeling and Deacetylase complex. *Journal of Molecular Biology*.

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**,** 865-873.

LERCHER, L., MCDONOUGH, M. A., EL-SAGHEER, A. H., THALHAMMER, A., KRIAUCIONIS, S., BROWN, T. & SCHOFIELD, C. J. 2014. Structural insights into how 5-hydroxymethylation influences transcription factor binding. *Chemical Communications,* 50**,** 1794-1796.

LESSARD, J., WU, J. I., RANISH, J. A., WAN, M., WINSLOW, M. M., STAAHL, B. T., WU, H., AEBERSOLD, R., GRAEF, I. A. & CRABTREE, G. R. 2007. An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron,* 55**,** 201-215.

LEU, S., VON FELTEN, S., FRANK, S., BOULAY, J.-L. & MARIANI, L. 2016. IDH mutation is associated with higher risk of malignant transformation in low-grade glioma. *Journal of neuro-oncology,* 127**,** 363-372.

LEVITT, R. J., GEORGESCU, M.-M. & POLLAK, M. 2005. PTEN-induction in U251 glioma cells decreases the expression of insulin-like growth factor binding protein-2. *Biochemical and biophysical research communications,* 336**,** 1056-1061.

LEVY, S. E. & MYERS, R. M. 2016. Advancements in next-generation sequencing. *Annual review of genomics and human genetics,* 17**,** 95-115.

LEWIS, L. C., LO, P. C. K., FOSTER, J. M., DAI, N., CORRÊA, I. R., DURCZAK, P. M., DUNCAN, G., RAMSAWHOOK, A., AITHAL, G. P. & DENNING, C. 2017. Dynamics of 5-carboxylcytosine during hepatic differentiation: Potential general role for active demethylation by DNA repair in lineage specification. *Epigenetics,* 12**,** 277-286.

LI, E., BESTOR, T. H. & JAENISCH, R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell,* 69**,** 915-926.

LI, X., OGHI, K. A., ZHANG, J. & KRONES, A. 2003. Eya protein phosphatase activity regulates Six1-Dach-Eya transcriptional effects in mammalian organogenesis. *Nature,* 426**,** 247.

LI, Y. & TOLLEFSBOL, T. O. 2011. DNA methylation detection: bisulfite genomic sequencing analysis. *Epigenetics Protocols.* Springer.

LI, Z., CAI, X., CAI, C.-L., WANG, J., ZHANG, W., PETERSEN, B. E., YANG, F.-C. & XU, M. 2011. Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood,* 118**,** 4509-4518.

LIANG, X., POTTER, J., KUMAR, S., ZOU, Y., QUINTANILLA, R.,

SRIDHARAN, M., CARTE, J., CHEN, W., ROARK, N. & RANGANATHAN, S. 2015. Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *Journal of biotechnology,* 208**,** 44-53.

LIEBELT, B. D., SHINGU, T., ZHOU, X., REN, J., SHIN, S. A. & HU, J. 2016. Glioma stem cells: signaling, microenvironment, and therapy. *Stem cells international,* 2016.

LIN, M.-J., LEE, T.-L., HSU, D.-W. & SHEN, C.-K. J. 2000. One-codon alternative splicing of the CpG MTase Dnmt1 transcript in mouse somatic cells. *FEBS letters,* 469**,** 101-104.

LITTLE, M., HOLMES, G., BICKMORE, W., VAN HEYNINGEN, V., HASTIE, N. & WAINWRIGHT, B. 1995. DNA binding capacity of the WT1 protein is abolished by Denys—Drash syndrome WT1 point mutations. *Human molecular genetics,* 4**,** 351-358.

LITTLE, M., PROSSER, J., CONDIE, A., SMITH, P. V., VAN HEYNINGEN, V. & HASTIE, N. 1992. Zinc finger point mutations within the WT1 gene in Wilms tumor patients. *Proceedings of the National Academy of Sciences,* 89**,** 4791-4795.

LIU, G., YUAN, X., ZENG, Z., TUNICI, P., NG, H., ABDULKADIR, I. R., LU, L., IRVIN, D., BLACK, K. L. & JOHN, S. Y. 2006. Analysis of gene expression and chemoresistance of CD133+ cancer stem cells in glioblastoma. *Molecular cancer,* 5**,** 67.

LIU, Y., XU, S., ZU, T., LI, F., SANG, S., LIU, C., AN, Y., MI, B., ORGILL, D. P. & MURPHY, G. F. 2019. Reversal of TET-mediated 5-hmC loss in hypoxic fibroblasts by ascorbic acid. *Laboratory*

Investigation, 1.

LOEB, D. M. 2006. WT1 influences apoptosis through transcriptional regulation of Bcl-2 family members. *Cell Cycle,* 5**,** 1249-1253.

LOEB, D. M., EVRON, E., PATEL, C. B., SHARMA, P. M., NIRANJAN, B., BULUWELA, L., WEITZMAN, S. A., KORZ, D. & SUKUMAR, S. 2001. Wilms' tumor suppressor gene (WT1) is expressed in primary breast tumors despite tumor-specific promoter methylation. *Cancer Research,* 61**,** 921-925.

LOEB, D. M. & SUKUMAR, S. 2002. The role of WT1 in oncogenesis: tumor suppressor or oncogene? *International journal of hematology,* 76**,** 117.

LORSBACH, R., MOORE, J., MATHEW, S., RAIMONDI, S., MUKATIRA, S. & DOWNING, J. 2003. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t (10; 11)(q22; q23). *Leukemia,* 17**,** 637.

LOTTAZ, C., BEIER, D., MEYER, K., KUMAR, P., HERMANN, A., SCHWARZ, J., JUNKER, M., OEFNER, P. J., BOGDAHN, U. & WISCHHUSEN, J. 2010. Transcriptional profiles of CD133+ and CD133− glioblastoma-derived cancer stem cell lines suggest different cells of origin. *Cancer research,* 70**,** 2030-2040.

LOUIS, D. N., OHGAKI, H., WIESTLER, O. D., CAVENEE, W. K., BURGER, P. C., JOUVET, A., SCHEITHAUER, B. W. & KLEIHUES, P. 2007. The 2007 WHO classification of tumours of the central nervous system. *Acta neuropathologica,* 114**,** 97-109.

LU, H., DALGARD, C. L., MOHYELDIN, A., MCFATE, T., TAIT, A. S. &

VERMA, A. 2005. Reversible inactivation of HIF-1 prolyl hydroxylases allows cell metabolism to control basal HIF-1. *Journal of Biological Chemistry,* 280**,** 41928-41939.

LU, X., HAN, D., ZHAO, B. S., SONG, C.-X., ZHANG, L.-S., DORÉ, L. C. & HE, C. 2015. Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics. *Cell research,* 25**,** 386.

LU, X., SONG, C.-X., SZULWACH, K., WANG, Z., WEIDENBACHER, P., JIN, P. & HE, C. 2013. Chemical modification-assisted bisulfite sequencing (CAB-Seq) for 5-carboxylcytosine detection in DNA. *Journal of the American Chemical Society,* 135**,** 9315-9317.

LV, M.-M., ZHU, X.-Y., CHEN, W.-X., ZHONG, S.-L., HU, Q., MA, T.-F., ZHANG, J., CHEN, L., TANG, J.-H. & ZHAO, J.-H. 2014. Exosomes mediate drug resistance transfer in MCF-7 breast cancer cells and a probable mechanism is delivery of P-glycoprotein. *Tumor Biology,* 35**,** 10773-10779.

MAHESWARAN, S., ENGLERT, C., BENNETT, P., HEINRICH, G. & HABER, D. A. 1995. The WT1 gene product stabilizes p53 and inhibits p53-mediated apoptosis. *Genes & Development,* 9**,** 2143-2156.

MAHFOUZ, M. M., PIATEK, A. & STEWART JR, C. N. 2014. Genome engineering via TALENs and CRISPR/Cas9 systems: challenges and perspectives. *Plant biotechnology journal,* 12**,** 1006-1014.

MALI, P., YANG, L., ESVELT, K. M., AACH, J., GUELL, M., DICARLO, J. E., NORVILLE, J. E. & CHURCH, G. M. 2013. RNA-guided human genome engineering via Cas9. *Science,* 339**,** 823-826.

MALRIC, L., MONFERRAN, S., GILHODES, J., BOYRIE, S., DAHAN, P., SKULI, N., SESEN, J., FILLERON, T., KOWALSKI-CHAUVEL, A. & MOYAL, E. C.-J. 2017. Interest of integrins targeting in glioblastoma according to tumor heterogeneity and cancer stem cell paradigm: an update. *Oncotarget,* 8**,** 86947.

MANSOURI, A., KARAMCHANDANI, J. & DAS, S. 2017. Molecular genetics of secondary glioblastoma. *Glioblastoma [Internet].* Codon Publications.

MARDIS, E. R. 2013. Next-generation sequencing platforms. *Annual review of analytical chemistry,* 6**,** 287-303.

MARTINOWICH, K., HATTORI, D., WU, H., FOUSE, S., HE, F., HU, Y., FAN, G. & SUN, Y. E. 2003. DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science,* 302**,** 890-893.

MARX, N., GRÜNWALD-GRUBER, C., BYDLINSKI, N., DHIMAN, H., NGOC NGUYEN, L., KLANERT, G. & BORTH, N. 2018. CRISPR-Based Targeted Epigenetic Editing Enables Gene Expression Modulation of the Silenced Beta-Galactoside Alpha-2, 6-Sialyltransferase 1 in CHO Cells. *Biotechnology journal,* 13**,** 1700217.

MATSUMOTO, S., BANINE, F., FEISTEL, K., FOSTER, S., XING, R., STRUVE, J. & SHERMAN, L. S. 2016. Brg1 directly regulates Olig2 transcription and is required for oligodendrocyte progenitor cell specification. *Developmental biology,* 413**,** 173-187.

MATSUMOTO, S., BANINE, F., FEISTEL, K., FOSTER, S., XING, R., STRUVE, J. & SHERMAN, L. S. 2016. Brg1 directly regulates Olig2

transcription and is required for oligodendrocyte progenitor cell specification. *Developmental biology,* 413**,** 173-187.

MAYO, M. W., WANG, C. Y., DROUIN, S. S., MADRID, L. V., MARSHALL, A. F., REED, J. C., WEISSMAN, B. E. & BALDWIN, A. S. 1999. WT1 modulates apoptosis by transcriptionally upregulating the bcl-2 proto-oncogene. *The EMBO journal,* 18**,** 3990-4003.

MCCLELLAND, M. & NELSON, M. 1985. The effect of site specific methylation on restriction endonuclease digestion. *Nucleic acids research,* 13**,** r201.

MCGREGOR, R. J., CHAU, Y.-Y., KENDALL, T. J., ARTIBANI, M., HASTIE, N. & HADOKE, P. W. 2018. WT1 expression in vessels varies with histopathological grade in tumour-bearing and control tissue from patients with breast cancer. *British journal of cancer,* 119**,** 1508.

MENKE, A. L., VAN DER EB, A. & JOCHEMSEN, A. 1998. The Wilms' tumor 1 gene: oncogene or tumor suppressor gene? *International review of cytology,* 181**,** 151-212.

MENSSEN, H., BERTELMANN, E., BARTELT, S., SCHMIDT, R., PECHER, G., SCHRAMM, K. & THIEL, E. 2000. Wilms' tumor gene (WT1) expression in lung cancer, colon cancer and glioblastoma cell lines compared to freshly isolated tumor specimens. *Journal of cancer research and clinical oncology,* 126**,** 226-232.

MERKLE, F. T., TRAMONTIN, A. D., GARCÍA-VERDUGO, J. M. & ALVAREZ-BUYLLA, A. 2004. Radial glia give rise to adult neural stem cells in the subventricular zone. *Proceedings of the National Academy of Sciences,* 101**,** 17528-17532.

MERLO, A., HERMAN, J. G., MAO, L., LEE, D. J., GABRIELSON, E., BURGER, P. C., BAYLIN, S. B. & SIDRANSKY, D. 1995. 5′ CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. *Nature medicine,* 1**,** 686.

MESSERSCHMIDT, D. M., KNOWLES, B. B. & SOLTER, D. 2014. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes & development,* 28**,** 812-828.

METSUYANIM, S., PODE-SHAKKED, N., SCHMIDT-OTT, K. M., KESHET, G., RECHAVI, G., BLUMENTAL, D. & DEKEL, B. 2008. Accumulation of malignant renal stem cells is associated with epigenetic changes in normal renal progenitor genes. *Stem Cells,* 26**,** 1808-1817.

MI, R., LUO, Y., CAI, J., LIMKE, T. L., RAO, M. S. & HÖKE, A. 2005. Immortalized neural stem cells differ from nonimmortalized cortical neurospheres and cerebellar granule cell progenitors. *Experimental neurology,* 194**,** 301-319.

MITRA, R. D., SHENDURE, J., OLEJNIK, J. & CHURCH, G. M. 2003. Fluorescent in situ sequencing on polymerase colonies. *Analytical biochemistry,* 320**,** 55-65.

MOHR, S. E., SMITH, J. A., SHAMU, C. E., NEUMÜLLER, R. A. & PERRIMON, N. 2014. RNAi screening comes of age: improved techniques and complementary approaches. *Nature reviews Molecular cell biology,* 15**,** 591.

MONK, M., BOUBELIK, M. & LEHNERT, S. 1987. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development,* 99**,** 371-382.

MORITA, S., HORII, T. & HATADA, I. 2018. Editing of DNA methylation using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Epigenome Editing.* Springer.

MULLEN, R. J., BUCK, C. R. & SMITH, A. M. 1992. NeuN, a neuronal specific nuclear protein in vertebrates. *Development,* 116**,** 201-211.

MÜNZEL, M., GLOBISCH, D., BRÜCKL, T., WAGNER, M., WELZMILLER, V., MICHALAKIS, S., MÜLLER, M., BIEL, M. & CARELL, T. 2010. Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angewandte Chemie International Edition,* 49**,** 5375-5377.

MUTSKOV, V. & FELSENFELD, G. 2004. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *The EMBO journal,* 23**,** 138-149.

NAGAE, G., ISAGAWA, T., SHIRAKI, N., FUJITA, T., YAMAMOTO, S., TSUTSUMI, S., NONAKA, A., YOSHIBA, S., MATSUSAKA, K. & MIDORIKAWA, Y. 2011. Tissue-specific demethylation in CpG-poor promoters during cellular differentiation. *Human molecular genetics,* 20**,** 2710-2721.

NAGAI, Y., MIYAZAWA, H., TANAKA, T., UDAGAWA, K., KATO, M., FUKUYAMA, S., YOKOTE, A., KOBAYASHI, K., KANAZAWA, M. & HAGIWARA, K. 2005. Genetic heterogeneity of the epidermal growth

factor receptor in non–small cell lung cancer cell lines revealed by a rapid and sensitive detection system, the peptide nucleic acid-locked nucleic acid PCR clamp. *Cancer research,* 65**,** 7276-7282.

NAKADA, M., NAMBU, E., FURUYAMA, N., YOSHIDA, Y., TAKINO, T., HAYASHI, Y., SATO, H., SAI, Y., TSUJI, T. & MIYAMOTO, K. 2013. Integrin α3 is overexpressed in glioma stem-like cells and promotes invasion. *British journal of cancer,* 108**,** 2516.

NAKAHARA, Y., OKAMOTO, H., MINETA, T. & TABUCHI, K. 2004. Expression of the Wilms' tumor gene product WT1 in glioblastomas and medulloblastomas. *Brain tumor pathology,* 21**,** 113-116.

NAKAMURA, T., ARAI, Y., UMEHARA, H., MASUHARA, M., KIMURA, T., TANIGUCHI, H., SEKIMOTO, T., IKAWA, M., YONEDA, Y. & OKABE, M. 2007. PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nature cell biology,* 9**,** 64-71.

NAKAMURA, T., LIU, Y.-J., NAKASHIMA, H., UMEHARA, H., INOUE, K., MATOBA, S., TACHIBANA, M., OGURA, A., SHINKAI, Y. & NAKANO, T. 2012. PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature,* 486**,** 415-419.

NAKATSUKA, S.-I., OJI, Y., HORIUCHI, T., KANDA, T., KITAGAWA, M., TAKEUCHI, T., KAWANO, K., KUWAE, Y., YAMAUCHI, A. & OKUMURA, M. 2006. Immunohistochemical detection of WT1 protein in a variety of cancer cells. *Modern pathology,* 19**,** 804.

NAN, X., NG, H.-H., JOHNSON, C. A., LAHERTY, C. D., TURNER, B. M., EISENMAN, R. N. & BIRD, A. 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase

complex. *Nature,* 393**,** 386-389.

NECHEMIA-ARBELY, Y., MIGA, K. H., SHOSHANI, O., ASLANIAN, A., MCMAHON, M. A., LEE, A. Y., FACHINETTI, D., YATES, J. R., REN, B. & CLEVELAND, D. W. 2019. DNA replication acts as an error correction mechanism to maintain centromere identity by restricting CENP-A to centromeres. *Nature cell biology,* 21**,** 743.

NESTOR, C. E., OTTAVIANO, R., REDDINGTON, J., SPROUL, D., REINHARDT, D., DUNICAN, D., KATZ, E., DIXON, J. M., HARRISON, D. J. & MEEHAN, R. R. 2012. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome research,* 22**,** 467-477.

NEUBERGER, M. S. & RADA, C. 2007. Somatic hypermutation: activation-induced deaminase for C/G followed by polymerase η for A/T. *The Journal of experimental medicine,* 204**,** 7-10.

NIERODE, G. J., GOPAL, S., KWON, P., CLARK, D. S., SCHAFFER, D. V. & DORDICK, J. S. 2019. High-throughput identification of factors promoting neuronal differentiation of human neural progenitor cells in microscale 3D cell culture. *Biotechnology and bioengineering,* 116**,** 168-180.

NISHIDA, S., HOSEN, N., SHIRAKATA, T., KANATO, K., YANAGIHARA, M., NAKATSUKA, S.-I., HOSHIDA, Y., NAKAZAWA, T., HARADA, Y. & TATSUMI, N. 2006. AML1-ETO rapidly induces acute myeloblastic leukemia in cooperation with the Wilms tumor gene, WT1. *Blood,* 107**,** 3303-3312.

NOACK, F., PATASKAR, A., SCHNEIDER, M., BUCHHOLZ, F.,

TIWARI, V. K. & CALEGARI, F. 2019. Assessment and site-specific manipulation of DNA (hydroxy-) methylation during mouse corticogenesis. *Life science alliance,* 2**,** e201900331.

NOUSHMEHR, H., WEISENBERGER, D. J., DIEFES, K., PHILLIPS, H. S., PUJARA, K., BERMAN, B. P., PAN, F., PELLOSKI, C. E., SULMAN, E. P. & BHAT, K. P. 2010. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer cell,* 17**,** 510-522.

NURMEMMEDOV, E., YENGO, R. K., UYSAL, H., KARLSSON, R. & THUNNISSEN, M. M. 2009. New insights into DNA-binding behavior of Wilms Tumor Protein (WT1)—A dual study. *Biophysical chemistry,* 145**,** 116-125.

O'BRIEN, L. L., GRIMALDI, M., KOSTUN, Z., WINGERT, R. A., SELLECK, R. & DAVIDSON, A. J. 2011. Wt1a, Foxc1a, and the Notch mediator Rbpj physically interact and regulate the formation of podocytes in zebrafish. *Developmental biology,* 358**,** 318-330.

OJI, Y., HASHIMOTO, N., TSUBOI, A., MURAKAMI, Y., IWAI, M., KAGAWA, N., CHIBA, Y., IZUMOTO, S., ELISSEEVA, O. & ICHINOHASAMA, R. 2016. Association of WT1 IgG antibody against WT1 peptide with prolonged survival in glioblastoma multiforme patients vaccinated with WT1 peptide. *International journal of cancer,* 139**,** 1391-1401.

OJI, Y., MIYOSHI, S., MAEDA, H., HAYASHI, S., TAMAKI, H., NAKATSUKA, S. I., YAO, M., TAKAHASHI, E., NAKANO, Y. & HIRABAYASHI, H. 2002. Overexpression of the Wilms' tumor gene

WT1 in de novo lung cancers. *International Journal of Cancer,* 100**,** 297-303.

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**,** 247-257.

ORR, B. A., HAFFNER, M. C., NELSON, W. G., YEGNASUBRAMANIAN, S. & EBERHART, C. G. 2012. Decreased 5-hydroxymethylcytosine is associated with neural progenitor phenotype in normal brain and shorter survival in malignant glioma. *PloS one,* 7**,** e41036.

OSTROM, Q. T., GITTLEMAN, H., FULOP, J., LIU, M., BLANDA, R., KROMER, C., WOLINSKY, Y., KRUCHKO, C. & BARNHOLTZ-SLOAN, J. S. 2015. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2008-2012. *Neuro-oncology,* 17**,** iv1.

OSWALD, J., ENGEMANN, S., LANE, N., MAYER, W., OLEK, A., FUNDELE, R., DEAN, W., REIK, W. & WALTER, J. 2000. Active demethylation of the paternal genome in the mouse zygote. *Current Biology,* 10**,** 475-478.

PAPIN, C., IBRAHIM, A., LE GRAS, S., VELT, A., STOLL, I., JOST, B., MENONI, H., BRONNER, C., DIMITROV, S. & HAMICHE, A. 2017. Combinatorial DNA methylation codes at repetitive elements. *Genome research,* 27**,** 934-946.

PARK, J.-L., KIM, H.-J., SEO, E.-H., KWON, O.-H., LIM, B., KIM, M., KIM, S.-Y., SONG, K.-S., KANG, G. H. & KIM, H. J. 2015. Decrease of

5hmC in gastric cancers is associated with TET1 silencing due to with DNA methylation and bivalent histone marks at TET1 CpG island 3′-shore. *Oncotarget,* 6**,** 37647.

PARSONS, D. W., JONES, S., ZHANG, X., LIN, J. C.-H., LEARY, R. J., ANGENENDT, P., MANKOO, P., CARTER, H., SIU, I.-M. & GALLIA, G. L. 2008. An integrated genomic analysis of human glioblastoma multiforme. *Science,* 321**,** 1807-1812.

PASTOR, W. A., ARAVIND, L. & RAO, A. 2013. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nature reviews Molecular cell biology,* 14**,** 341-356.

PASTOR, W. A., PAPE, U. J., HUANG, Y., HENDERSON, H. R., LISTER, R., KO, M., MCLOUGHLIN, E. M., BRUDNO, Y., MAHAPATRA, S. & KAPRANOV, P. 2011. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature,* 473**,** 394.

PELLETIER, J., SCHALLING, M., BUCKLER, A., ROGERS, A., HABER, D. & HOUSMAN, D. 1991. Expression of the Wilms' tumor gene WT1 in the murine urogenital system. *Genes & development,* 5**,** 1345-1356.

PEÑUELAS, S., ANIDO, J., PRIETO-SÁNCHEZ, R. M., FOLCH, G., BARBA, I., CUARTAS, I., GARCÍA-DORADO, D., POCA, M. A., SAHUQUILLO, J. & BASELGA, J. 2009. TGF-β increases glioma-initiating cell self-renewal through the induction of LIF in human glioblastoma. *Cancer cell,* 15**,** 315-327.

PFAFFENEDER, T., HACKNER, B., TRUß, M., MÜNZEL, M., MÜLLER, M., DEIML, C. A., HAGEMEIER, C. & CARELL, T. 2011. The

discovery of 5-formylcytosine in embryonic stem cell DNA. *Angewandte Chemie,* 123**,** 7146-7150.

PFEIFER, G. P. 2018. Defining driver DNA methylation changes in human cancer. *International journal of molecular sciences,* 19**,** 1166.

PFEIFER, G. P., KADAM, S. & JIN, S.-G. 2013. 5-hydroxymethylcytosine and its potential roles in development and cancer. *Epigenetics & chromatin,* 6**,** 1.

PIKAART, M. J., RECILLAS-TARGA, F. & FELSENFELD, G. 1998. Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes & development,* 12**,** 2852-2862.

POLLARD, S. M., YOSHIKAWA, K., CLARKE, I. D., DANOVI, D., STRICKER, S., RUSSELL, R., BAYANI, J., HEAD, R., LEE, M. & BERNSTEIN, M. 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**,** 568-580.

POLONI, A., GOTERI, G., ZIZZI, A., SERRANI, F., TRAPPOLINI, S., COSTANTINI, B., MARIANI, M., OLIVIERI, A., CATARINI, M. & CENTURIONI, R. 2013. Prognostic role of immunohistochemical analysis of 5 mc in myelodysplastic syndromes. *European journal of haematology,* 91**,** 219-227.

PRITCHARD-JONES, K., FLEMING, S., DAVIDSON, D., BICKMORE, W. & PORTEOUS, D. 1990. The candidate Wilms' tumour gene is involved in genitourinary development. *Nature,* 346**,** 194.

PRITCHARD-JONES, K., FLEMING, S., DAVIDSON, D., BICKMORE,

W., PORTEOUS, D., GOSDEN, C., BARD, J., BUCKLER, A., PELLETIER, J. & HOUSMAN, D. 1990. The candidate Wilms' tumour gene is involved in genitourinary development. *Nature,* 346**,** 194-197.

PUIG, I., TENBAUM, S. P., CHICOTE, I., ARQUÉS, O., MARTÍNEZ-QUINTANILLA, J., CUESTA-BORRÁS, E., RAMÍREZ, L., GONZALO, P., SOTO, A. & AGUILAR, S. 2018. TET2 controls chemoresistant slow-cycling cancer cell survival and tumor recurrence. *The Journal of clinical investigation,* 128.

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**,** 6139-6149.

RAIBER, E.-A., BERALDI, D., CUESTA, S. M., MCINROY, G. R., KINGSBURY, Z., BECQ, J., JAMES, T., LOPES, M., ALLINSON, K. & FIELD, S. 2017. Base resolution maps reveal the importance of 5-hydroxymethylcytosine in a human glioblastoma. *NPJ genomic medicine,* 2**,** 6.

RAIBER, E.-A., BERALDI, D., FICZ, G., BURGESS, H. E., BRANCO, M. R., MURAT, P., OXLEY, D., BOOTH, M. J., REIK, W. & 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**,** R69.

RAMANI, P. & COWELL, J. K. 1996. THE EXPRESSION PATTERN OF WILMS'TUMOUR GENE (WT1) PRODUCT IN NORMAL TISSUES AND PAEDIATRIC RENAL TUMOURS. *The Journal of pathology,* 179**,**

162-168.

RAMIRO, A. R. & BARRETO, V. M. 2015. Activation-induced cytidine deaminase and active cytidine demethylation. *Trends in biochemical sciences,* 40**,** 172-181.

RAMPAL, R., ALKALIN, A., MADZO, J., VASANTHAKUMAR, A., PRONIER, E., PATEL, J., LI, Y., AHN, J., ABDEL-WAHAB, O. & SHIH, A. 2014. DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia. *Cell reports,* 9**,** 1841-1855.

RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nature protocols,* 8**,** 2281.

RAPONI, E., AGENES, F., DELPHIN, C., ASSARD, N., BAUDIER, J., LEGRAVEREND, C. & DELOULME, J. C. 2007. S100B expression defines a state in which GFAP-expressing cells lose their neural stem cell potential and acquire a more mature developmental stage. *Glia,* 55**,** 165-177.

RAUSCHER, J., BESCHORNER, R., GIERKE, M., BISDAS, S., BRAUN, C., EBNER, F. H. & SCHITTENHELM, J. 2014. WT1 expression increases with malignancy and indicates unfavourable outcome in astrocytoma. *Journal of clinical pathology,* 67**,** 556-561.

RAY R, Z., POINTER, K. B., KUO, J. S. & DEMPSEY, R. J. 2014. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Neurosurgery,* 75**,** N9-N10.

RAZIN, A. & RIGGS, A. D. 1980. DNA methylation and gene function.

Science, 210**,** 604-610.

REDDY, J. C., MORRIS, J. C., WANG, J., ENGLISH, M. A., HABER, D. A., SHI, Y. & LICHT, J. D. 1995. WT1-mediated transcriptional activation is inhibited by dominant negative mutant proteins. *Journal of Biological Chemistry,* 270**,** 10878-10884.

REIK, W., KELSEY, G. & WALTER, J. 1999. Dissecting de novo methylation. *Nature genetics,* 23**,** 380-382.

REYA, T., MORRISON, S. J., CLARKE, M. F. & WEISSMAN, I. L. 2001. Stem cells, cancer, and cancer stem cells. *nature,* 414**,** 105.

RICHARD, D. J., SCHUMACHER, V., ROYER-POKORA, B. & ROBERTS, S. G. 2001. Par4 is a coactivator for a splice isoform–specific transcriptional activation domain in WT1. *Genes & development,* 15**,** 328-339.

RIU, E., CHEN, Z.-Y., XU, H., HE, C.-Y. & KAY, M. A. 2007. Histone modifications are associated with the persistence or silencing of vector-mediated transgene expression in vivo. *Molecular Therapy,* 15**,** 1348-1355.

ROBERTS, S. G. 2005. Transcriptional regulation by WT1 in development. *Current opinion in genetics & development,* 15**,** 542-547.

ROBERTSON, K. D. & JONES, P. A. 1998. The human ARF cell cycle regulatory gene promoter is a CpG island which can be silenced by DNA methylation and down-regulated by wild-type p53. *Molecular and cellular biology,* 18**,** 6457-6473.

ROSSELLA, F., POLLEDRI, E., BOLLATI, V., BACCARELLI, A. & FUSTINONI, S. 2009. Development and validation of a gas

chromatography/mass spectrometry method for the assessment of genomic DNA methylation. *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry,* 23**,** 2637-2646.

ROYER-POKORA, B., BUSCH, M., BEIER, M., DUHME, C., DE TORRES, C., MORA, J., BRANDT, A. & ROYER, H.-D. 2010. Wilms tumor cells with WT1 mutations have characteristic features of mesenchymal stem cells and express molecular markers of paraxial mesoderm. *Human molecular genetics,* 19**,** 1651-1668.

RUTESHOUSER, E. C. & HUFF, V. Familial Wilms tumor. American Journal of Medical Genetics Part C: Seminars in Medical Genetics, 2004. Wiley Online Library, 29-34.

RUTESHOUSER, E. C., ROBINSON, S. M. & HUFF, V. 2008. Wilms tumor genetics: mutations in WT1, WTX, and CTNNB1 account for only about one-third of tumors. *Genes, Chromosomes and Cancer,* 47**,** 461-470.

SADONI, N., LANGER, S., FAUTH, C., BERNARDI, G., CREMER, T., TURNER, B. M. & ZINK, D. 1999. Nuclear organization of mammalian genomes: polar chromosome territories build up functionally distinct higher order compartments. *The Journal of cell biology,* 146**,** 1211-1226.

SAKAUE, M., OHTA, H., KUMAKI, Y., ODA, M., SAKAIDE, Y., MATSUOKA, C., YAMAGIWA, A., NIWA, H., WAKAYAMA, T. & OKANO, M. 2010. DNA methylation is dispensable for the growth and

survival of the extraembryonic lineages. *Current Biology,* 20**,** 1452-1457.

SALLINEN, P., SALLINEN, S. L., HELEN, P., RANTALA, I., RAUTIAINEN, E., HELIN, H., KALIMO, H. & HAAPASALO, H. 2000. Grading of diffusely infiltrating astrocytomas by quantitative histopathology, cell proliferation and image cytometric DNA analysis: comparison of 133 tumours in the context of the WHO 1979 and WHO 1993 grading schemes. *Neuropathology and Applied Neurobiology,* 26**,** 319-331.

SANGER, F. 1988. Sequences, sequences, and sequences. *Annual review of biochemistry,* 57**,** 1-29.

SANTIVASI, W. L. & XIA, F. 2013. The role and clinical significance of DNA damage response and repair pathways in primary brain tumors. *Cell & bioscience,* 3**,** 10.

SANTOS, F., HENDRICH, B., REIK, W. & DEAN, W. 2002. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Developmental biology,* 241**,** 172-182.

SARDINA, J. L., COLLOMBET, S., TIAN, T. V., GÓMEZ, A., DI STEFANO, B., BERENGUER, C., BRUMBAUGH, J., STADHOUDERS, R., SEGURA-MORALES, C. & GUT, M. 2018. Transcription factors drive Tet2-mediated enhancer demethylation to reprogram cell fate. *Cell stem cell,* 23**,** 727-741. e9.

SARNAT, H. B., NOCHLIN, D. & BORN, D. E. 1998. Neuronal nuclear antigen (NeuN): a marker of neuronal maturation in the early human fetal nervous system. *Brain and Development,* 20**,** 88-94.

SAYEED, S. K., ZHAO, J., SATHYANARAYANA, B. K., GOLLA, J. P. & VINSON, C. 2015. C/EBPβ (CEBPB) protein binding to the C/EBP| CRE DNA 8-mer TTGC| GTCA is inhibited by 5hmC and enhanced by 5mC, 5fC, and 5caC in the CG dinucleotide. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms,* 1849**,** 583-589.

SCEBBA, F., DE BASTIANI, M., BERNACCHIA, G., ANDREUCCI, A., GALLI, A. & PITTO, L. 2007. PRMT11: a new Arabidopsis MBD7 protein partner with arginine methyltransferase activity. *The Plant Journal,* 52**,** 210-222.

SCHITTENHELM, J., MITTELBRONN, M., NGUYEN, T. D., MEYERMANN, R. & BESCHORNER, R. 2008. WT1 expression distinguishes astrocytic tumor cells from normal and reactive astrocytes. *Brain pathology,* 18**,** 344-353.

SCHOFIELD, C. J. & RATCLIFFE, P. J. 2004. Oxygen sensing by HIF hydroxylases. *Nature reviews Molecular cell biology,* 5**,** 343.

SCHOMACHER, L. 2013. Mammalian DNA demethylation: multiple faces and upstream regulation. *Epigenetics,* 8**,** 679-684.

SCHÜBELER, D. 2015. Function and information content of DNA methylation. *Nature,* 517**,** 321.

SEISENBERGER, S., ANDREWS, S., KRUEGER, F., ARAND, J., WALTER, J., SANTOS, F., POPP, C., THIENPONT, B., DEAN, W. & REIK, W. 2012. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Molecular cell,* 48**,** 849-862.

SEISENBERGER, S., PEAT, J. R., HORE, T. A., SANTOS, F., DEAN,

W. & REIK, W. 2013. Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 368**,** 20110330.

SEMENZA, G. L. 2000. Hypoxia, clonal selection, and the role of HIF-1 in tumor progression. *Critical reviews in biochemistry and molecular biology,* 35**,** 71-103.

SEMENZA, G. L. 2002. HIF-1 and tumor progression: pathophysiology and therapeutics. *Trends in molecular medicine,* 8**,** S62-S67.

SERI, B., GARCIA-VERDUGO, J. M., MCEWEN, B. S. & ALVAREZ-BUYLLA, A. 2001. Astrocytes give rise to new neurons in the adult mammalian hippocampus. *The Journal of Neuroscience,* 21**,** 7153-7160.

SETOGUCHI, H., NAMIHIRA, M., KOHYAMA, J., ASANO, H., SANOSAKA, T. & NAKASHIMA, K. 2006. Methyl-CpG binding proteins are involved in restricting differentiation plasticity in neurons. *Journal of neuroscience research,* 84**,** 969-979.

SHANDILYA, J. & ROBERTS, S. G. 2015. A role of WT1 in cell division and genomic stability. *Cell Cycle,* 14**,** 1358-1364.

SHARIF, J., MUTO, M., TAKEBAYASHI, S.-I., SUETAKE, I., IWAMATSU, A., ENDO, T. A., SHINGA, J., MIZUTANI-KOSEKI, Y., TOYODA, T. & OKAMURA, K. 2007. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature,* 450.

SHARMA, P. M., YANG, X., BOWMAN, M., ROBERTS, V. &

SUKUMAR, S. 1992. Molecular cloning of rat Wilms' tumor complementary DNA and a study of messenger RNA expression in the urogenital system and the brain. *Cancer research,* 52**,** 6407-6412.

SHEN, L., WU, H., DIEP, D., YAMAGUCHI, S., D'ALESSIO, A. C., FUNG, H.-L., ZHANG, K. & ZHANG, Y. 2013. Genome-wide analysis reveals TET-and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell,* 153**,** 692-706.

SHEN, Y., ZHANG, J., YU, T. & QI, C. 2018. Generation of PTEN knockout bone marrow mesenchymal stem cell lines by CRISPR/Cas9-mediated genome editing. *Cytotechnology,* 70**,** 783-791.

SHI, J., SUN, B., SHI, W., ZUO, H., CUI, D., NI, L. & CHEN, J. 2015. Decreasing GSH and increasing ROS in chemosensitivity gliomas with IDH1 mutation. *Tumor Biology,* 36**,** 655-662.

SHIM, E.-H., LIVI, C. B., RAKHEJA, D., TAN, J., BENSON, D., PAREKH, V., KHO, E.-Y., GHOSH, A. P., KIRKMAN, R. & VELU, S. 2014. L-2-Hydroxyglutarate: an epigenetic modifier and putative oncometabolite in renal cancer. *Cancer discovery,* 4**,** 1290-1298.

SHIMODA, M., OTA, M. & OKADA, Y. 2018. Isolation of cancer stem cells by side population method. *Cancer Stem Cells.* Springer.

SINGH, A. M., CHAPPELL, J., TROST, R., LIN, L., WANG, T., TANG, J., WU, H., ZHAO, S., JIN, P. & DALTON, S. 2013. Cell-cycle control of developmentally regulated transcription factors accounts for heterogeneity in human pluripotent cells. *Stem cell reports,* 1**,** 532-544.

SMITH, L. M., FUNG, S., HUNKAPILLER, M. W., HUNKAPILLER, T. J. & HOOD, L. E. 1985. The synthesis of oligonucleotides containing an

aliphatic amino group at the 5′ terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis. *Nucleic acids research,* 13**,** 2399-2412.

SMITH, Z. D., CHAN, M. M., MIKKELSEN, T. S., GU, H., GNIRKE, A., REGEV, A. & MEISSNER, A. 2012. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature,* 484**,** 339-344.

SMITH, Z. D. & MEISSNER, A. 2013. DNA methylation: roles in mammalian development. *Nature reviews. Genetics,* 14**,** 204.

SMOLL, N. R., SCHALLER, K. & GAUTSCHI, O. P. 2013. Long-term survival of patients with glioblastoma multiforme (GBM). *Journal of Clinical Neuroscience,* 20**,** 670-675.

SOLTANI, M. H., PICHARDO, R., SONG, Z., SANGHA, N., CAMACHO, F., SATYAMOORTHY, K., SANGUEZA, O. P. & SETALURI, V. 2005. Microtubule-associated protein 2, a marker of neuronal differentiation, induces mitotic defects, inhibits growth of melanoma cells, and predicts metastatic potential of cutaneous melanoma. *The American journal of pathology,* 166**,** 1841-1850.

SON, H., KANG, H., KIM, H. S. & KIM, S. 2017. Somatic mutation driven codon transition bias in human cancer. *Scientific reports,* 7**,** 14204.

SON, M. J., WOOLARD, K., NAM, D.-H., LEE, J. & FINE, H. A. 2009. SSEA-1 is an enrichment marker for tumor-initiating cells in human glioblastoma. *Cell stem cell,* 4**,** 440-452.

SONG, C.-X., SZULWACH, K. E., DAI, Q., FU, Y., MAO, S.-Q., LIN, L., STREET, C., LI, Y., POIDEVIN, M. & WU, H. 2013. Genome-wide

profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell,* 153**,** 678-691.

SONG, C.-X., SZULWACH, K. E., FU, Y., DAI, Q., YI, C., LI, X., LI, Y., CHEN, C.-H., ZHANG, W. & JIAN, X. 2011. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature biotechnology,* 29**,** 68-72.

SONG, C.-X., YI, C. & HE, C. 2012. Mapping recently identified nucleotide variants in the genome and transcriptome. *Nature biotechnology,* 30**,** 1107.

SONG, G., WANG, G., LUO, X., CHENG, Y., SONG, Q., WAN, J., MOORE, C., SONG, H., JIN, P. & QIAN, J. 2019. An All-to-All Approach to the Identification of Sequence-Specific Readers for Epigenetic DNA Modifications on Cytosine. *bioRxiv***,** 638700.

SPALLOTTA, F., CENCIONI, C., ATLANTE, S., GARELLA, D., COCCO, M., MORI, M., MASTROCOLA, R., KUENNE, C., GUENTHER, S. & NANNI, S. 2018. Stable oxidative cytosine modifications accumulate in cardiac mesenchymal cells from type2 diabetes patients: rescue by α-ketoglutarate and TET-TDG functional reactivation. *Circulation research,* 122**,** 31-46.

SPIERS, H., HANNON, E., SCHALKWYK, L. C., BRAY, N. J. & MILL, J. 2017. 5-hydroxymethylcytosine is highly dynamic across human fetal brain development. *BMC genomics,* 18**,** 738.

SPRUIJT, C. G., GNERLICH, F., SMITS, A. H., PFAFFENEDER, T., JANSEN, P. W., BAUER, C., MÜNZEL, M., WAGNER, M., MÜLLER, M. & KHAN, F. 2013. Dynamic readers for 5-(hydroxy) methylcytosine

and its oxidized derivatives. *Cell,* 152**,** 1146-1159.

STOLL, R., LEE, B. M., DEBLER, E. W., LAITY, J. H., WILSON, I. A., DYSON, H. J. & WRIGHT, P. E. 2007. Structure of the Wilms tumor suppressor protein zinc finger domain bound to DNA. *Journal of molecular biology,* 372**,** 1227-1245.

STOREBJERG, T. M., STRAND, S. H., HØYER, S., LYNNERUP, A.-S., BORRE, M., ØRNTOFT, T. F. & 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**,** 105.

STRAND, S. H., HOYER, S., LYNNERUP, A.-S., HALDRUP, C., STOREBJERG, T. M., BORRE, M., ORNTOFT, T. F. & SORENSEN, K. D. 2015. High levels of 5-hydroxymethylcytosine (5hmC) is an adverse predictor of biochemical recurrence after prostatectomy in ERG-negative prostate cancer. *Clinical epigenetics,* 7**,** 111.

STURM, D., WITT, H., HOVESTADT, V., KHUONG-QUANG, D.-A., JONES, D. T., KONERMANN, C., PFAFF, E., TÖNJES, M., SILL, M. & BENDER, S. 2012. Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer cell,* 22**,** 425-437.

SUDHIR, S., LOPEZ-BERTONI, H., LUGO-FAGUNDO, M., LAL, B. & LATERRA, J. 2019. SOX2-mediated 5hmC dysregulation in GBM stem cells. AACR.

SUN, Z., TERRAGNI, J., BORGARO, J. G., LIU, Y., YU, L., GUAN, S.,

WANG, H., SUN, D., CHENG, X. & ZHU, Z. 2013. High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. *Cell reports,* 3**,** 567-576.

SZEMES, M., DALLOSSO, A. R., MELEGH, Z., CURRY, T., LI, Y., RIVERS, C., UNEY, J., MÄGDEFRAU, A.-S., SCHWIDERSKI, K. & PARK, J. H. 2012. Control of epigenetic states by WT1 via regulation of de novo DNA methyltransferase 3A. *Human molecular genetics,* 22**,** 74-83.

SZULWACH, K. E., LI, X., LI, Y., SONG, C.-X., WU, H., DAI, Q., IRIER, H., UPADHYAY, A. K., GEARING, M. & LEVEY, A. I. 2011. 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nature neuroscience,* 14**,** 1607-1616.

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**,** e181-e181.

TAHILIANI, M., KOH, K. P., SHEN, Y., PASTOR, W. A., BANDUKWALA, H., BRUDNO, Y., AGARWAL, S., IYER, L. M., LIU, D. R. & ARAVIND, L. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science,* 324**,** 930-935.

TAKAI, H., MASUDA, K., SATO, T., SAKAGUCHI, Y., SUZUKI, T., SUZUKI, T., KOYAMA-NASU, R., NASU-NISHIMURA, Y., KATOU, Y. & OGAWA, H. 2014. 5-Hydroxymethylcytosine plays a critical role in glioblastomagenesis by recruiting the CHTOP-methylosome complex.

Cell reports, 9**,** 48-60.

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**,** 3445-3452.

TATIPARTI, K., SAU, S., KASHAW, S. & IYER, A. 2017. siRNA delivery strategies: a comprehensive review of recent developments. *Nanomaterials,* 7**,** 77.

TATSUMI, N., OJI, Y., TSUJI, N., TSUDA, A., HIGASHIO, M., AOYAGI, S., FUKUDA, I., ITO, K., NAKAMURA, J. & TAKASHIMA, S. 2008. Wilms' tumor gene WT1-shRNA as a potent apoptosis-inducing agent for solid tumors. *International journal of oncology,* 32**,** 701-711.

TAYLOR, M. D., POPPLETON, H., FULLER, C., SU, X., LIU, Y., JENSEN, P., MAGDALENO, S., DALTON, J., CALABRESE, C. & BOARD, J. 2005. Radial glia cells are candidate stem cells of ependymoma. *Cancer cell,* 8**,** 323-335.

TEFFERI, A., LIM, K., ABDEL-WAHAB, O., LASHO, T., PATEL, J., PATNAIK, M., HANSON, C. A., PARDANANI, A., GILLILAND, D. & LEVINE, R. 2009. Detection of mutant TET2 in myeloid malignancies other than myeloproliferative neoplasms: CMML, MDS, MDS/MPN and AML. *Leukemia,* 23**,** 1343-1345.

TEMPLE, S. 2001. The development of neural stem cells. *Nature,* 414**,** 112.

THAKKAR, J. P., DOLECEK, T. A., HORBINSKI, C., OSTROM, Q. T.,

LIGHTNER, D. D., BARNHOLTZ-SLOAN, J. S. & VILLANO, J. L. 2014. Epidemiologic and molecular prognostic review of glioblastoma. *Cancer Epidemiology and Prevention Biomarkers,* 23**,** 1985-1996.

THIENPONT, B., STEINBACHER, J., ZHAO, H., D'ANNA, F., KUCHNIO, A., PLOUMAKIS, A., GHESQUIÈRE, B., VAN DYCK, L., BOECKX, B. & SCHOONJANS, L. 2016. Tumour hypoxia causes DNA hypermethylation by reducing TET activity. *Nature,* 537**,** 63.

THURNHER, M. M. 2009. 2007 World Health Organization classification of tumours of the central nervous system. *Cancer imaging,* 9**,** S1.

TIAN, Y.-P., ZHU, Y.-M., SUN, X.-H. & LAI, M.-D. 2016. Multiple functions of ten-eleven translocation 1 during tumorigenesis. *Chinese medical journal,* 129**,** 1744.

TILGHMAN, J., YING, M. & LATERRA, J. 2013. SC-040. REGULATION OF GLIOMA STEM CELL GENE EXPRESSION AND PHENOTYPE BY KRÜPPEL-LIKE FACTOR 9 TRANSCRIPTION FACTOR. *Neuro-oncology,* 15.

TROJANOWSKI, J., WALKENSTEIN, N. & LEE, V. 1986. Expression of neurofilament subunits in neurons of the central and peripheral nervous system: an immunohistochemical study with monoclonal antibodies. *Journal of Neuroscience,* 6**,** 650-660.

TSAI, S. Q. & JOUNG, J. K. 2016. Defining and improving the genome-wide specificities of CRISPR–Cas9 nucleases. *Nature Reviews Genetics,* 17**,** 300.

TSUBOI, A., OKA, Y., OGAWA, H., ELISSEEVA, O. A., TAMAKI, H.,

OJI, Y., KIM, E. H., SOMA, T., TATEKAWA, T. & KAWAKAMI, M. 1999. Constitutive expression of the Wilms' tumor gene WT1 inhibits the differentiation of myeloid progenitor cells but promotes their proliferation in response to granulocyte-colony stimulating factor (G-CSF). *Leukemia research,* 23**,** 499-505.

TSUMURA, A., HAYAKAWA, T., KUMAKI, Y., TAKEBAYASHI, S. I., SAKAUE, M., MATSUOKA, C., SHIMOTOHNO, K., ISHIKAWA, F., LI, E. & UEDA, H. R. 2006. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes to Cells,* 11**,** 805-814.

TUNICI, P., BISSOLA, L., LUALDI, E., POLLO, B., CAJOLA, L., BROGGI, G., SOZZI, G. & FINOCCHIARO, G. 2004. Genetic alterations and in vivo tumorigenicity of neurospheres derived from an adult glioblastoma. *Molecular cancer,* 3**,** 25.

TURCAN, S., ROHLE, D., GOENKA, A., WALSH, L. A., FANG, F., YILMAZ, E., CAMPOS, C., FABIUS, A. W., LU, C. & WARD, P. S. 2012. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature,* 483**,** 479-483.

URIBE-LEWIS, S., STARK, R., CARROLL, T., DUNNING, M. J., BACHMAN, M., ITO, Y., STOJIC, L., HALIM, S., VOWLER, S. L. & LYNCH, A. G. 2015. 5-hydroxymethylcytosine marks promoters in colon that resist DNA hypermethylation in cancer. *Genome biology,* 16**,** 69.

VERHAAK, R. G., HOADLEY, K. A., PURDOM, E., WANG, V., QI, Y., WILKERSON, M. D., MILLER, C. R., DING, L., GOLUB, T. & MESIROV, J. P. 2010. Integrated genomic analysis identifies clinically

relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer cell,* 17**,** 98-110.

VICENT, S., CHEN, R., SAYLES, L. C., LIN, C., WALKER, R. G., GILLESPIE, A. K., SUBRAMANIAN, A., HINKLE, G., YANG, X. & SAIF, S. 2010. Wilms tumor 1 (WT1) regulates KRAS-driven oncogenesis and senescence in mouse and human models. *The Journal of clinical investigation,* 120**,** 3940.

VIEIRA, J. M., HOWARD, S., DEL CAMPO, C. V., BOLLINI, S., DUBÉ, K. N., MASTERS, M., BARNETTE, D. N., ROHLING, M., SUN, X. & HANKINS, L. E. 2017. BRG1-SWI/SNF-dependent regulation of the Wt1 transcriptional landscape mediates epicardial activity during heart development and disease. *Nature communications,* 8**,** 16034.

VIZE, P. D., SEUFERT, D. W., CARROLL, T. J. & WALLINGFORD, J. B. 1997. Model systems for the study of kidney development: use of the pronephros in the analysis of organ induction and patterning. *Developmental biology,* 188**,** 189-204.

WANG, D., HASHIMOTO, H., ZHANG, X., BARWICK, B. G., LONIAL, S., BOISE, L. H., VERTINO, P. M. & CHENG, X. 2016. MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. *Nucleic acids research,* 45**,** 2396-2407.

WANG, J., XU, S.-L., DUAN, J.-J., YI, L., GUO, Y.-F., SHI, Y., LI, L., YANG, Z.-Y., LIAO, X.-M. & CAI, J. 2019. Invasion of white matter tracts by glioma stem cells is regulated by a NOTCH1–SOX2 positive-feedback loop. *Nature neuroscience,* 22**,** 91.

WANG, L., GUNDELACH, J. H. & BRAM, R. J. 2017. Cycloheximide

promotes paraptosis induced by inhibition of cyclophilins in glioblastoma multiforme. *Cell death & disease,* 8**,** e2807.

WANG, L., ZHOU, Y., XU, L., XIAO, R., LU, X., CHEN, L., CHONG, J., LI, H., HE, C. & FU, X.-D. 2015. Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature*.

WANG, S., LONG, Y., WANG, J., GE, Y., GUO, P., LIU, Y., TIAN, T. & ZHOU, X. 2013. Systematic Investigations of Different Cytosine Modifications on CpG Dinucleotide Sequences: The Effects on the BZ Transition. *Journal of the American Chemical Society,* 136**,** 56-59.

WANG, T., PAN, Q., LIN, L., SZULWACH, K. E., SONG, C.-X., HE, C., WU, H., WARREN, S. T., JIN, P. & DUAN, R. 2012. Genome-wide DNA hydroxymethylation changes are associated with neurodevelopmental genes in the developing human cerebellum. *Human molecular genetics,* 21**,** 5500-5510.

WANG, T., WEI, J. J., SABATINI, D. M. & LANDER, E. S. 2014. Genetic screens in human cells using the CRISPR-Cas9 system. *Science,* 343**,** 80-84.

WANG, X., PRAGER, B. C., WU, Q., KIM, L. J., GIMPLE, R. C., SHI, Y., YANG, K., MORTON, A. R., ZHOU, W. & ZHU, Z. 2018. Reciprocal signaling between glioblastoma stem cells and differentiated tumor cells promotes malignant progression. *Cell Stem Cell,* 22**,** 514-528. e5.

WANG, Y., XIAO, M., CHEN, X., CHEN, L., XU, Y., LV, L., WANG, P., YANG, H., MA, S. & LIN, H. 2015. WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation. *Molecular cell,* 57**,** 662-673.

WATT, F. & MOLLOY, P. L. 1988. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes & development,* 2**,** 1136-1143.

WEBER, M., HELLMANN, I., STADLER, M. B., RAMOS, L., PÄÄBO, S., REBHAN, M. & SCHÜBELER, D. 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature genetics,* 39**,** 457.

WEISENBERGER, D. J. 2014. Characterizing DNA methylation alterations from the cancer genome atlas. *The Journal of clinical investigation,* 124**,** 17.

WEN, L., LI, X., YAN, L., TAN, Y., LI, R., ZHAO, Y., WANG, Y., XIE, J., ZHANG, Y. & SONG, C. 2014. Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. *Genome biology,* 15**,** 1.

WHELDON, L. M., ABAKIR, A., FERJENTSIK, Z., DUDNAKOVA, T., STROHBUECKER, S., CHRISTIE, D., DAI, N., GUAN, S., FOSTER, J. M. & CORRÊA JR, I. R. 2014. Transient accumulation of 5-carboxylcytosine indicates involvement of active demethylation in lineage specification of neural stem cells. *Cell reports,* 7**,** 1353-1361.

WILKINS, O. M., JOHNSON, K. C., HOUSEMAN, E. A., KING, J. E., MARSIT, C. J. & CHRISTENSEN, B. C. 2018. Normal breast tissue 5-hydroxymethylcytosine is enriched at DNA regulatory elements of breast cancer cells. AACR.

WILM, B. & MUÑOZ-CHAPULI, R. 2016. The role of WT1 in embryonic

development and normal organ homeostasis. *The Wilms' Tumor (WT1) Gene.* Springer.

WIRSCHING, H.-G., KRISHNAN, S., FLOREA, A.-M., FREI, K., KRAYENBÜHL, N., HASENBACH, K., REIFENBERGER, G., WELLER, M. & TABATABAI, G. 2013. Thymosin beta 4 gene silencing decreases stemness and invasiveness in glioblastoma. *Brain,* 137**,** 433-448.

WOSSIDLO, M., NAKAMURA, T., LEPIKHOV, K., MARQUES, C. J., ZAKHARTCHENKO, V., BOIANI, M., ARAND, J., NAKANO, T., REIK, W. & WALTER, J. 2011. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nature communications,* 2**,** 241.

WU, H., D'ALESSIO, A. C., ITO, S., XIA, K., WANG, Z., CUI, K., ZHAO, K., SUN, Y. E. & ZHANG, Y. 2011. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature,* 473**,** 389.

WU, S. C. & ZHANG, Y. 2010. Active DNA demethylation: many roads lead to Rome. *Nature reviews Molecular cell biology,* 11**,** 607-620.

WU, X. & LI, G. 2016. Prevalent Accumulation of Non-Optimal Codons through Somatic Mutations in Human Cancers. *PloS one,* 11**,** e0160463.

XI, Q., WANG, Z., ZAROMYTIDOU, A.-I., ZHANG, X. H.-F., CHOW-TSANG, L.-F., LIU, J. X., KIM, H., BARLAS, A., MANOVA-TODOROVA, K. & KAARTINEN, V. 2011. A poised chromatin platform for TGF-β access to master regulators. *Cell,* 147**,** 1511-1524.

XU, H., XIA, Y.-K., LI, C.-J., ZHANG, J.-Y., LIU, Y., YI, W., QIN, Z.-Y.,

CHEN, L., SHI, Z.-F. & QUAN, K. 2019. Rapid diagnosis of IDH1-mutated gliomas by 2-HG detection with gas chromatography mass spectrometry. *Laboratory Investigation,* 99**,** 588.

XU, W., YANG, H., LIU, Y., YANG, Y., WANG, P., KIM, S.-H., ITO, S., YANG, C., WANG, P. & XIAO, M.-T. 2011. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. *Cancer cell,* 19**,** 17-30.

XU, Y., WU, F., TAN, L., KONG, L., XIONG, L., DENG, J., BARBERA, A. J., ZHENG, L., ZHANG, H. & HUANG, S. 2011. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Molecular cell,* 42**,** 451-464.

YAMAGAMI, T., SUGIYAMA, H., INOUE, K., OGAWA, H., TATEKAWA, T., HIRATA, M., KUDOH, T., AKIYAMA, T., MURAKAMI, A. & MAEKAWA, T. 1996. Growth inhibition of human leukemic cells by WT1 (Wilms tumor gene) antisense oligodeoxynucleotides: implications for the involvement of WT1 in leukemogenesis. *Blood,* 87**,** 2878-2884.

YAMANOUCHI, K., OHTA, T., LIU, Z., OJI, Y., SUGIYAMA, H., SHRIDHAR, V., MATSUMURA, S., TAKAHASHI, T., TAKAHASHI, K. & KURACHI, H. 2014. The Wilms' Tumor Gene WT1− 17AA/− KTS Splice Variant Increases Tumorigenic Activity Through Up-Regulation of Vascular Endothelial Growth Factor in an In Vivo Ovarian Cancer Model. *Translational oncology,* 7**,** 580-589.

YAN, H., PARSONS, D. W., JIN, G., MCLENDON, R., RASHEED, B. A., YUAN, W., KOS, I., BATINIC-HABERLE, I., JONES, S. & RIGGINS, G. J. 2009. IDH1 and IDH2 mutations in gliomas. *New England Journal*

of Medicine, 360**,** 765-773.

YANG, B., SUN, H., LIN, W., HOU, W., LI, H., ZHANG, L., LI, F., GU, Y., SONG, Y. & LI, Q. Evaluation of global DNA hypomethylation in human prostate cancer and prostatic intraepithelial neoplasm tissues by immunohistochemistry. Urologic Oncology: Seminars and Original Investigations, 2013. Elsevier, 628-634.

YANG, H., LIU, Y., BAI, F., ZHANG, J., MA, S., LIU, J., XU, Z., ZHU, H., LING, Z. & YE, D. 2013. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene,* 32**,** 663-669.

YANG, Q., WU, K., JI, M., JIN, W., HE, N., SHI, B. & HOU, P. 2013. Decreased 5-hydroxymethylcytosine (5-hmC) is an independent poor prognostic factor in gastric cancer patients. *J Biomed Nanotechnol,* 9**,** 1607-1616.

YE ZHANG, S. S., BEN BARRES, MELANIE HAYDEN GEPHART 2016. GENT-08. HUMAN GLIOBLASTOMA ASTROCYTES RESEMBLE FETAL ASTROCYTE PRECURSOR CELLS AS REVEALED BY IMMUNOPANNING AND RNAseq. *NEURO-ONCOLOGY*.

YEUNG, J. T., HAMILTON, R. L., OKADA, H., JAKACKI, R. I. & POLLACK, I. F. 2013. Increased expression of tumor-associated antigens in pediatric and adult ependymomas: implication for vaccine therapy. *Journal of neuro-oncology,* 111**,** 103-111.

YIN, R., MAO, S.-Q., ZHAO, B., CHONG, Z., YANG, Y., ZHAO, C., ZHANG, D., HUANG, H., GAO, J. & LI, Z. 2013. Ascorbic acid

enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. *Journal of the American Chemical Society,* 135**,** 10396-10403.

YING, M., TILGHMAN, J., WEI, Y., GUERRERO-CAZARES, H., QUINONES-HINOJOSA, A., JI, H. & LATERRA, J. 2014. Kruppel-like factor-9 (KLF9) inhibits glioblastoma stemness through global transcription repression and integrin α6 inhibition. *Journal of Biological Chemistry,* 289**,** 32742-32756.

YOO, S. & BIEDA, M. C. 2014. Differences among brain tumor stem cell types and fetal neural stem cells in focal regions of histone modifications and DNA methylation, broad regions of modifications, and bivalent promoters. *BMC genomics,* 15**,** 724.

YOUSSOUFIAN, H., KAZAZIAN, H. H., PHILLIPS, D. G., ARONIS, S., TSIFTIS, G., BROWN, V. A. & ANTONARAKIS, S. E. 1986. Recurrent mutations in haemophilia A give evidence for CpG mutation hotspots. *Nature,* 324**,** 380-382.

YOUSSOUFIAN, H., KAZAZIAN JR, H., PHILLIPS, D. G., ARONIS, S., TSIFTIS, G., BROWN, V. A. & ANTONARAKIS, S. E. 1985. Recurrent mutations in haemophilia A give evidence for CpG mutation hotspots. *Nature,* 324**,** 380-382.

YU, X., LIANG, X., XIE, H., KUMAR, S., RAVINDER, N., POTTER, J., DU JEU, X. D. M. & CHESNUT, J. D. 2016. Improved delivery of Cas9 protein/gRNA complexes using lipofectamine CRISPRMAX. *Biotechnology letters,* 38**,** 919-929.

ZHANG, C., COMAI, L. & JOHNSON, D. L. 2005. PTEN represses

RNA Polymerase I transcription by disrupting the SL1 complex. *Molecular and cellular biology,* 25**,** 6899-6911.

ZHANG, L., LI, Z., CHEN, G., HUANG, Q., ZHANG, J., WEN, J., YE, X. & CAI, C. 2016. Validation and quantification of genomic 5-carboxylcytosine (5caC) in mouse brain tissue by liquid chromatography-tandem mass spectrometry. *Analytical Methods,* 8**,** 5812-5817.

ZHANG, P., LATHIA, J. D., FLAVAHAN, W. A., RICH, J. N. & MATTSON, M. P. 2009. Squelching glioblastoma stem cells by targeting REST for proteasomal degradation. *Trends in neurosciences,* 32**,** 559-565.

ZHANG, S., XIE, R., WAN, F., YE, F., GUO, D. & LEI, T. 2013. Identification of U251 glioma stem cells and their heterogeneous stem-like phenotypes. *Oncology letters,* 6**,** 1649-1655.

ZHANG, X., ZHANG, X., HU, S., ZHENG, M., ZHANG, J., ZHAO, J., ZHANG, X., YAN, B., JIA, L. & ZHAO, J. 2017. Identification of miRNA-7 by genome-wide analysis as a critical sensitizer for TRAIL-induced apoptosis in glioblastoma cells. *Nucleic acids research,* 45**,** 5930-5944.

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 & development,* 13**,** 1924-1935.

ZHONG, Y., WEI, J., FU, Y., SHAO, J., LIANG, Y., LIN, Y., LIU, J. & ZHU, Z. 2008. Toxicity of cationic liposome Lipofectamine 2000 in human pancreatic cancer Capan-2 cells. *Nan fang yi ke da xue xue*

bao= Journal of Southern Medical University, 28**,** 1981-1984.

ZHOU, D., ALVER, B. M., LI, S., HLADY, R. A., THOMPSON, J. J., SCHROEDER, M. A., LEE, J.-H., QIU, J., SCHWARTZ, P. H. & SARKARIA, J. N. 2018. Distinctive epigenomes characterize glioma stem cells and their response to differentiation cues. *Genome biology,* 19**,** 43.

ZHOU, X., REN, L., LI, Y., ZHANG, M., YU, Y. & YU, J. 2010. The next-generation sequencing technology: a technology review and future perspective. *Science China Life Sciences,* 53**,** 44-57.

ZLUVOVA, J., JANOUSEK, B. & VYSKOT, B. 2001. Immunohistochemical study of DNA methylation dynamics during plant development. *Journal of experimental botany,* 52**,** 2265-2273.

ZOU, X., MA, W., SOLOV'YOV, I. A., CHIPOT, C. & SCHULTEN, K. 2011. Recognition of methylated DNA through methyl-CpG binding domain proteins. *Nucleic acids research,* 40**,** 2747-2758.

ZULCH, K. J. 1979. Histological typing of tumours of the central nervous system. *International histological classification of tumours No 21.***,** 19-24.