DNA Methylation in Amyotrophic Lateral Sclerosis

Charlie Samantha Appleby-Mallinder BSc

Department of Neuroscience (SITraN)

Supervisors: Dr Robin Highley and Dr Paul Heath

SITraN

Sheffield Institute of Translational
Neuroscience

Acknowledgements

I would first like to thank my supervisors Robin and Paul for their unwavering support at both an academic and personal level. Paul – without you taking a chance on me when I was a blue haired undergraduate, and setting me loose in the lab, I would not be submitting this PhD. Your guidance, wisdom and advice have kept both the project and myself moving forward, while still understanding the importance of having a life outside the lab. Robin – our meetings that turned into talking about rubbish helped me keep what is left of my sanity! Thank you for teaching me everything I know about neuropathology, and mentoring me through what has been a tough three years. A nod also to Dr Julie Simpson and Dr Janine Kirby who have offered fantastic moral support throughout.

Thanks also go to our wonderful tech team. In particular, Matt Wyles, Catherine Gelsthorpe, Lynne Baxter and Dan Fillingham. It has been a delight to be in the lab with you all, be that to figure out why something on my project isn't working (this happened on many occasions), or just generally putting the world to rights. Without you all, my project would not have come to fruition. Thanks also go to my fellow PhD students and Dr Rachel Waller for taking me on regular coffee trips during the write up phase. It was therapeutic and I am so grateful to have that support!

Most importantly, I would like to thank ALS tissue donors and their families for their selfless donations, as without you, I would not be conducting my research. Thanks also go to the funders of my project, the Pathological Society and the British Neuropathological Society.

Many thanks also to my wonderful family and friends. In particular to my mum and dad, for encouraging and supporting me in my studies, and always pushing me to do my best.

During the writing up phase, I would like to thank my dogs and cats for providing cuddles and emotional support, and for helping keep me calm and focused. There is nothing better than a cuddle from a furry goofball to make you feel better!

My final and biggest thanks go to my wonderful, supportive and goofy husband Alex. There are not enough good words in the world to describe you. Thank you for putting up with me, helping me and for picking me up off the floor when I came home mentally and physically exhausted from PhD. I know I could not have got through this project without your unwavering love and support.

Abstract

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease primarily characterised by motor neurone (MN) degeneration and death. ALS can be sporadic (sALS) or familial, with a number of associated gene mutations, including *C9orf72* (C9ALS). Previous studies have found impaired gene expression, implicating pathways involved with RNA metabolism and inflammation/immune response. DNA methylation is an epigenetic mechanism whereby a methyl group is attached to a cytosine (5mC), usually resulting in gene expression repression. 5mC can further be oxidised to 5-hydroxymethylcytosine (5hmC). DNA methylation has been studied in other neurodegenerative diseases, but little work has been conducted in ALS.

The aim of this thesis is to elucidate DNA methylation's role (if any) in the decline of MNs without the interactions from other cell types, which may mask MN-specific DNA methylation changes.

Immunohistochemical analysis found higher levels of 5mC and 5hmC in ALS in the residual lower motor neurones (LMNs) of the spinal cord, with C9ALS displaying the highest global methylation. Interestingly, in LMNs with TDP43 pathology, a loss of 5mC and 5hmC from the nucleus was observed.

LMNs were then extracted from a subset of the same cases using laser capture microdissection (LCM). Following this, DNA was extracted from the LMNs and underwent analysis using the MethylationEPIC array. Results indicated a global hypermethylation in C9ALS, with both hypermethylation and hypomethylation detected at the single gene level in ALS. GO and pathway analysis implicated RNA metabolism changes. The MethylationEPIC dataset was then compared to pre-existing mRNA expression data, with overlapping hits undergoing pathway analysis. Changes were found in cell signalling, inflammation and immune response and cell death/apoptosis.

In conclusion, DNA methylation is a contributory factor in ALS. Data presented in this thesis suggests that hypermethylation is a prominent factor. Further studies are warranted to further understand the role of DNA methylation in ALS.

Table of Contents

List of Figures

List of Tables

Abbreviations

Abbreviation	Full name
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
ABC	Avidin/biotinylated enzyme complex
AD	Alzheimer's disease
AFCx	Association frontal cortex
AH	Anterior horn
ALS	Amyotrophic Lateral Sclerosis
ANG	Angiogenin
ANK1	Ankyrin1
BBB	Blood brain barrier
BS-NGS	Bisulphite converted-next generation sequencing
C9ALS	*C9orf72*-associated ALS
C9orf72	Chromosome 9 open reading frame 72
CAM	Charlie Appleby-Mallinder
Cas9	CRISP-associated protein-9 nuclease
CNS	Central nervous system
CRISPR	Clustered regularly interspaced short-palindromic repeats
CRISPR/Cas9	Clustered regularly interspaced short-palindromic repeats/CRISP-associated protein-9 nuclease
DAB	3,3-diaminobenzidine
DAVID	Database for annotation, visualisatoin and integrate discovery
DC	Dorsal column
DDR	DNA damage response
DLB	Dementia with Lewy bodies
DMR	Differentially methylated region
DNMT	DNA methyltransferases
DPR	Dipeptide repeat protein
DSB	Double stranded break
ELISA	Enzyme-linked immunosorbent essay
EWAS	Epigenome-wide association study
FACS	Fluorescence activated cell sorting
FACx	Frontal assocation cortex
fALS	Familial amyotrophic lateral sclerosis
FFPE	Formalin fixed paraffin embedded
FTD	Frontotemporal dementia
FTD/ALS	Frontotemporal dementia/Amyotrophic lateral sclerosis
FTLD	Frontotemporal lobar degeneration
FTLD-DLB	Frontotemporal lobar degeneration-dementia with Lewy bodies
FTLD-TDP	Frontotemporal lobar degeneration-TAR DNA-binding protein
FUS	Fused in sarcoma
GO	Gene ontology

GREAT	Genomic Regions Enrichment of Annotations
H2O2	Hydrogen peroxide
HD	Huntington's disease
hESCs	Human embryonic stem cells
HIV1	Human immunodeficiency virus type 1
hnRNP	Heterogeneous nuclear ribonucleoprotein
HRE	Hexanucleotide repeat expansion
IHC	Immunohistochemistry
iPSC	Induced pluripotent stem cell
JRH	J. Robin Highley
LCM	Laser capture microdissection
LC-MS	Liquid chromatography-mass spectrometry
LCT	Lateral corticospinal tract
LMN	Lower motor neurone
MCx	Motor cortex
miRNA	Micro RNA
MN	Motor neurone
m*SOD1*	Mutant *SOD1*
Ox-BS	Oxidative bisulphite sequencing
PANTHER	Protein analysis through evolutionary relationships
PD	Parkinson's disease
PET	Positron emission tomography
PM	Post-mortem
qRT-PCR	Quantitative-reverse transcriptase-polymerase chain reaction
RAN	Repeat-associated non-ATG translation
RBP	RNA binding protein
RNA-seq	RNA sequencing
RNP	Ribonucleoprotein particle
RPM	Random primer mix
RRBS	Reduced representation bisulphite conversion
RRM1	RNA recognition motif 1
RRM2	RNA recognition motif 2
RT	Room temperature
RT-PCR	Reverse transcriptase-polymerase chain reaction
sALS	Sporadic amyotrophic lateral sclerosis
SAMP8	Senescence-accelerated mouse 8
SBTB	Sheffield brain tissue bank
SMA	Spinal muscular atrophy
SMRT	Single-molecule real-time
SOD1	Superoxide dismutase 1
SSB	Single stranded break
TARDBP	TAR DNA binding protein
TB	Toluidine blue
TBS	Tris-buffered saline
TDG	Thymine DNA glycosylase

TDP43	TAR DNA-binding protein 43 kDa
TdT	Terminal deoxynucleotidyl transferase
TET	Ten-eleven translocation enzymes
TET-BS	Tet-assisted bisulphite sequencing
TSC	Tri-sodium citrate
UMN	Upper motor neurone
UTR	Untranslated region
WGBS	Whole-genome bisulphite sequencing

Chapter 1: Introduction

This project aims to determine whether abnormal DNA methylation in motor neurones and glia is involved in the pathogenesis of amyotrophic lateral sclerosis (ALS).

1.1 Amyotrophic Lateral Sclerosis

ALS is the most common motor neurone disease. It is characterised by progressive motor function loss brought on by degeneration and death of motor neurones (MNs), first described in 1869 (Charcot and Joffroy, 1869). Degeneration is seen in both upper motor neurones (UMN) and lower motor neurones (LMN), leading to muscle weakness, spasticity and atrophy, with death typically occurring three years after disease onset. According to the criteria usually used for ALS diagnosis, El Escorial, the incidence of ALS is 1.5-2 per 100,000 people, per year, with a prevalence of 5-7 per 100,000 people worldwide (Ludolph *et al.*, 2015).

The cause of ALS is unknown. Risk factors include having an immediate family member diagnosed with the disease, a lifestyle with vigorous prolonged exercise (Harwood *et al.*, 2009) and advanced age. The mean age of onset varies from 50 to 65 years of age, with only 5% of cases displaying onset below 30 years of age (Zarei *et al.* 2015).

1.2 Current treatments in ALS

There are currently two drug treatments available, Riluzole and Edaravone (Jaiswal 2019). Riluzole was introduced into the clinic for ALS in 1994, after randomised controlled trials showed a modest increase in patient survival by 2-3 months (Miller *et al.* 2012). Edaravone, approved by Japan in 2015 and the USA in 2017, followed this. Edaravone modestly slows the decline in motor function in ALS patients (Edaravone Writing Group 2017). The mechanisms of action for both drugs are poorly understood, and no other treatments are available for ALS (Petrov *et al.* 2017). Due to the modest effects of these drugs, it is of high importance that research into the complex nature of ALS is carried out in order to provide more effective treatments.

1.3 Causes and genetics of ALS

ALS is classified as either familial ALS (fALS) or sporadic ALS (sALS): fALS is defined as having a family history of ALS, potentially with a known associated gene mutation. sALS refers to cases with no known family history of the disease, potentially with no identified mutation, with the potential for reduced penetrance. There appears to be overlap between these, with some disease-causing mutations being observed in some sALS cases. This may

be due to a number of reasons, including: *de novo* mutations, family members having died before they developed the disease or contact has been lost with family members who may carry the mutant gene. Sporadic cases account for the majority of cases, (90-95%), with males being more affected than females, at a ratio of 1.5:1 (Love, *et al.*, 2015). The approximate age of disease onset for sALS is 60 years, with approximately 25% of first diagnoses being made under 50 years old. The age of onset in fALS is approximately ten years earlier than sALS. fALS makes up 5-10% of diagnoses, defined as specific associated genetic dominant inheritance of a mutated gene or having an intermediate family member with the disease, although some non-autosomal dominant cases have been reported (Zarei *et al.* 2015). Some cases are familial, but the causative gene is unknown.

The first mutations identified in fALS were in the gene encoding an antioxidant ubiquitous enzyme, superoxide dismutase 1 (*SOD1*). This gene has been implicated in 20% of fALS cases and 5% of sALS cases. The mutations result in structural instability, with misfolding of the enzyme. This creates aggregates in the MNs within the central nervous system (CNS). Other gene mutations known to harbour causative mutations in ALS include TAR DNA binding protein (*TARDBP*) and fused in sarcoma (*FUS*), which account for 5-10% of all fALS cases. More recently, a mutation in the chromosome 9 open reading frame (*C9orf72*) gene was discovered to be implicated in ALS. This is found in 37% of fALS cases and 7% of sALS cases, making it the most commonly mutated gene (Rademakers, 2012). Mutated *C9orf72* is also associated with frontotemporal dementia (FTD), which defines a group of disorders where the frontal and/or temporal lobes degenerate. This leads to changes in behaviour, language and movement, with FTD sometimes being seen in patients also suffering from ALS, known as FTD/ALS.

1.3.1 TDP43

TAR DNA-binding protein 43 kDa (TDP43) is coded for by the *TARDBP* gene (Ou *et al.*, 1995) and acts as a transcriptional repressor. *TARDBP*'s chromosomal location is 1p36.22, from base pairs 11,012,622 to 11,030,528 (see figure 1.1). TDP43 includes six cysteine residues, with four of these (Cys 173, Cys 175, Cys 198 and Cys 244) being located in two RNA recognition motifs (RRM1 and RRM2). The remaining two cysteine residues (Cys 39 and Cys 50) are located in the N-terminal domain (Valle and Carrì, 2017). TDP43 contains repetitive functional domains. These are referred to as low complexity or prion-like domains, which are defined as low complexity sequences found in RNA proteins that have been shown to drive protein aggregation in neurodegeneration, and to some extent, normal function as well (Hill *et al.*, 2016). Mutations in the *TARDBP* gene are rare and account for

approximately 1% of sALS cases and 4% of fALS cases (Scotter, Chen and Shaw, 2015). However, TDP43 proteinopathy is frequently seen in ALS (see section 1.3.2).

Figure 1.1: Visual representation of the chromosomal location of the *C9orf72* and *TARDBP* genes
(A) The *C9orf72* gene has a chromosomal location of 9p21.1, with a molecular location of base pair 27,546,546 to base pair 27,573,866. (**B)** The *TARDBP* gene has a chromosomal location of 1p36.22, with a molecular location of base pair 11,012,622 to base pair 11,030,528 (Genome Decoration Page, NCBI).

TDP43 has multiple nucleic acid processing functions and belongs to the heterogeneous nuclear ribonucleoprotein (hnRNP) family and has multiple mRNA targets: Sephton *et al.*, 2011 characterised the transcriptome-wide binding sites of TDP43 in homogenates of mouse brain by RNA immunoprecipitation followed by sequencing. This revealed thousands of RNA species, many of which were RNA processing genes. TDP43 plays a role in mRNA stability regulation by recruiting CNOT7/CAF1 deadenylase on mRNA 3'UTR. This leads to poly (A) tail deadenylation, and therefore shortening (Fukushima *et al.*, 2019). The length of the Poly (A) tail is important during mRNA translation, with a longer poly (A) tail increasing the duration of the mRNA translation process. The rate-limiting step of the message decaying is the shortening of the poly (A) tail. Hence, this could have a damaging effect on mRNA translation (Rubin and Halim 1993).

While TDP43 has both RNA and DNA processing roles, the former has been well studied, with the latter relatively neglected. Thus, a literature search using Google Scholar on 14/08/2019 was conducted, which yielded the following results:

- Key words: "TDP43" AND "RNA processing" AND "ALS" yielded 4,280 results
- Keywords: "TDP43" AND "DNA processing" AND "ALS" yielded 52 results. However, on closer inspection of these hits, only RNA processing roles of TDP43 were discussed

TDP43 is mainly localised to the nucleus, but also shuttles to and from the cytoplasm. Within the nucleus, TDP43 regulates splicing of both non-coding and protein-coding genes involved in neuronal survival in neurodegeneration (Tollervey *et al.*, 2011) and mitochondrial homeostasis maintenance by regulating mitochondrial transcript processing (Izumikawa *et al.*, 2017). In the cytoplasm, TDP43 is also involved in transcription regulation, and stress granule formation in response to oxidative insult (Higashi *et al.*, 2013), as well as having a role in normal skeletal muscle formation and regeneration. Stress granules are mRNA-arresting ribonucleoprotein particles (RNPs) which form after transient cell stress to stop translation and aid in the regulation of cell metabolism (Vogler *et al.*, 2019). TDP43 also regulates its own mRNA transcription levels through a negative feedback loop, with TDP43 binding via the 3' untranslated region (UTR) to the *TARDBP* mRNA transcript (Ayala *et al.*, 2011).

While TDP43's RNA processing roles are well documented, the role TDP43 plays in DNA processing is less well understood. It is thought to act through homologous DNA pairing through its DNA binding domain. In 1995, TDP43 was identified as a cofactor that binds to a regulatory element in the long terminal repeat of the human immunodeficiency virus type 1 (HIV-1), resulting in in vitro transcription being repressed from this element (Ou *et al.*, 1995).

However, studies carried out after this point focused solely on the RNA-related roles of TDP43. Therefore, DNA processing represents an under-researched area that could be of high significance in understanding the role TDP43 plays in ALS.

1.3.2 TDP43 pathology in ALS

TDP43 has been posited to have many pathological roles in ALS, including changes in protein stability and degradation, impaired cytoskeletal function and altered homeostasis of DNA and RNA binding proteins (Peters, 2015).

A hyper-phosphorylated, ubiquitinated and cleaved form of TDP43 is associated with ALS. It is observed in almost all cases of sALS, and most cases of mutation-associated ALS (with the exception of cases with mutations of FUS and SOD1) (Manuela Neumann *et al.*, 2006), henceforth referred to as pathologic TDP43. This pathologic TDP43 is found in the form of pre-inclusions, defined as cells displaying diffuse cytoplasmic TDP43 with concomitant loss of nuclear TDP43, and cytoplasmic inclusions, defined as dense aggregates in the cytoplasm, again with loss of nuclear TDP43. These inclusions are seen in a subset of residual neurones and glia. Unlike cytoplasmic stress granule formation, when pathologic TDP43 inclusions form, it is believed that the cell will not recover.

Oxidative stress, excitotoxicity and neuroinflammation have been shown to cause TDP43 to delocalise from the nucleus, where it is normally found, into the cytosol, where it forms large aggregates and oligomers, which are associated with synaptic loss and neuronal death (Bozzo *et al.*, 2016; Ederle and Dormann, 2017).

The four stages of TDP43 proteinopathy in ALS are summarised in table 1.1 (Brettschneider *et al.*, 2013). These four stages pertain to the spread of inclusions through the brain. Other classification systems for TDP43 pathology do exist (Tan *et al.*, 2015; Nelson *et al.*, 2019). However the system described in by Brettschneider *et al.*, 2013 is the most relevant to ALS.

The links between ALS pathogenesis and TDP43 protein function are yet to be elucidated. However, one possible mechanism includes a toxic loss of function occurring in nuclear TDP43 due to disruption of pathways in which TDP43 is involved. Another possible mechanism is that the TDP43 cytoplasmic inclusions produced in ALS MNs represent a toxic gain of function, leading to MN death (Hill *et al.*, 2016).

Table 1.1: The four stages of TDP43 proteinopathy in ALS.	
Stage	**Areas TDP43 proteinopathy is observed**
I	•Motor cortex: neurones from layers II, III, V and VI •Neocortex: projection neurones •Spinal cord ventral horns: large motor neurones •Cranial nerve nuclei: V, VII, X, XI and XII
II	•Prefrontal cortex •Some precerebellar nuclei •Reticular formation •Substantia nigra pars compacta: dopaminergic neurones
III	•Prefrontal granular cortex layers; all apart from layer IV •Parietal and temporal cortices: sensory areas (pyramidal cells) •Inferior colliculus •Striatum nuclei •Claustrum projection neurones
IV	•Anteriomedial temporal lobe neuronal groups •Hippocampal neuronal groups: dentate fascia granular cells and pyramidal neurones of Ammon's horn •Cerebellum: some involvement in dentate nuclei

1.3.3 *FUS*

FUS is a ubiquitously expressed gene, encoded by 15 exons, that belongs to the hnRNP family and has many features in common with TDP-43. Structurally, *FUS* contains an RNA recognition motif, C-terminus, a zinc finger motif, and multiple Arg-Gly-Gly repeats. It is mainly localised to the nucleus, but shuttles between nucleus and cytoplasm (Mackenzie, Rademakers and Neumann, 2010). Nuclear *FUS* is involved in the regulation of transcription, as well as splicing, of thousands of target genes. Within the cytoplasm, FUS is part of the formation of RNA transport granules and stress granules (Ayala *et al.*, 2011).

FUS accounts for approximately 5% of fALS cases and 1% of sALS cases, with 58 pathogenic *FUS* mutations identified, with most of these mutations found in the conserved C-terminus (Lagier-Tourenne *et al.*, 2013), which is required for DNA/RNA binding and the modulation of alternative splicing (Zinszner *et al.*, 1997). Mutant *FUS* is associated with misfolded RNA-binding proteins and the formation of aggregates in the cytoplasm, similar to in TDP43 pathology (Ederle and Dormann, 2017). These characteristics make it highly similar, both structurally and functionally, to TDP43. These similarities also therefore implicate nucleic acid processing in ALS disease pathogenesis.

1.3.4 *C9orf72*

C9orf72 is a gene involved in endosomal trafficking regulation and forms part of the *C9orf72*-SMCR8 complex, which regulates autophagy (Yang *et al.*, 2016). This gene can be mutated in ALS, with *C9orf72* mutations taking the form of a hexanucleotide repeat expansion comprising a GGGGCC repeat, found in the first intron of the gene, between two non-coding exons. It is located on 9p21.2 from base pair 27,546,546 to base pair 27,573,866 (*Homo sapiens* annotation release 108,GRCh38.p7, Genome Decoration Page, NCBI) (see figure 1.1). Three transcripts of *C9orf72* exist. Variants 1 and 3 contain different non-coding first exons, exon 1b for variant 1 and exon 1a for variant 3. These are fused to coding exons 2-11, with both variants encoding the same 481 amino acid protein. Variant 2 contains exon 1a fused to coding exons 2-5, encoding a 222 amino acid protein (DeJesus-Hernandez *et al.*, 2011) (see figure 1.2, adapted from: Haeusler *et al.*, 2016). A normal number of repeats is three, with an intermediate number being 4-27 repeats, which has less certain pathological significance. ALS caused by *C9orf72* mutations (C9ALS) is normally defined by greater than 30 repeats (Love *et al.*, 2015).

Figure 1.2: *C9orf72* gene variants
Three transcripts of *C9orf72* exist. Variants 1 and 3 contain different non-coding first exons, exon 1b for variant 1 and exon 1a for variant 3. These are fused to coding exons 2-11, with both variants encoding the same 481 amino acid protein. Variant 2 contains exon 1a fused to coding exons 2-5, encoding a 222 amino acid protein. Adapted from: (Haeusler *et al.* 2016).

1.3.4.1 *C9orf72* pathophysiology in ALS

Three principal aetiological mechanisms for *C9orf72* related ALS are posited: loss of function through haploinsufficiency; RNA toxicity from transcript accumulation and proteotoxicity (see figure 1.3).

Figure 1.3: *C9orf72*-ALS overview
Many genetic causes of ALS have been identified. Diagnosis falls either under fALS or sALS categories. *C9orf72* has been implicated in some fALS cases and is thought to have some involvement in sALS. Three hypothesis are currently being considered as to how *C9orf72* expansion repeat relates to ALS. These are: haploinsufficiency, RNA toxicity and proteotoxicity.

1.3.4.2 Haploinsufficiency in *C9orf72*-ALS

There are lower levels of C9orf72 protein found in C9ALS cases, and thus reduced expression levels. With respect to haploinsufficiency, cellular roles for *C9orf72* have been suggested to include the regulation of autophagy and endosomal trafficking in primary neurones and neuronal cells (Farg *et al.*, 2014). On knockdown of *C9orf72*, an increase in the autophagosome marker light chain 3 (LC3) was found. This suggests that *C9orf72* may be involved in the regulation of endocytosis and autophagy (Webster *et al.*, 2016). *C9orf72* variant 2 showed decreased transcription levels in cells from mutation carriers (DeJesus-Hernandez *et al.*, 2011), suggesting loss through haploinsufficiency as a possible aetiological mechanism.

C9orf72 knockout mouse models have been developed to test the haploinsufficiency hypothesis of *C9orf72* pathogenesis (Burberry *et al.*, 2016; Sudria-Lopez *et al.*, 2016). Evidence of this hypothesis was seen in lower C9orf72 protein levels in C9ALS/FTD patients (Waite *et al.*, 2014). Sudria-Lopez *et al.*, 2016 showed that full ablation of *C9orf72* in all

tissues results in reduced survival rates, as well as disordered immunity. However, no effect on motor function was detected, as well as no ALS-type pathology: No MN degeneration or gliosis was present, as well as no abnormal ubiquitination or TDP43 pathology. This is similar to findings by Burberry *et al.*, 2016, who also noted lower survival rates, with no changes to number of spinal cord (SC) MNs, and no gross changes in SC or brain. In this mouse model, an increase in inflammatory cytokines was observed, as well as autoimmunity, with increased levels of GFAP-immunopositive glia detected. The inability to develop a mouse model that accurately replicates the pathology of ALS indicates that the role of pathological studies of human tissue is of vital importance.

1.3.4.3 *C9orf72*-ALS and RNA toxicity

RNA toxicity may result from the accumulation of bidirectionally-transcribed sense (GGGGCC) and anti-sense (CCCCGG) transcripts which appear to sequester RNA binding proteins, leading to defective processing of pre-messenger RNA (Cooper-Knock *et al.*, 2015).

RNA containing the *C9orf72* expansion repeat forms nuclear RNA foci, which are toxic. These interact with and sequester RNA binding proteins, leading to transcriptome defects. Both sense and anti-sense RNA transcripts are produced. A similar mechanism is involved in other neurodegenerative diseases caused by repeat expansions, including Huntington's disease (HD) (Wang *et al.*, 2013) and myotonic muscular dystrophy (Machuca-tzili *et al.* 2005).

1.3.4.4 *C9orf72*-ALS and proteotoxicity

Proteotoxicity is impairment of cell function as a result of protein misfolding. In the context of *C9orf72*-ALS, this is caused by the production of dipeptide repeat proteins (DPRs) from an unconventional form of translation from expanded nucleotide repeats. This process is referred to as repeat-associated non-ATG (RAN) translation, it occurs in all reading frames, resulting in the production of five DPRs in *C9forf72*-related FTD/ALS. These DPRs are aggregation-prone and can accumulate in the CNS (Ash *et al.*, 2013). The DPRs are also associated with toxicity, in the form of TDP43 proteinopathy, microgliosis and extramotor involvement (Moens *et al.*, 2019).

1.4 The neuropathology of ALS

Some gross brain abnormalities are observed in ALS, including macroscopic atrophy of the SC anterior nerve roots. There may also be atrophy of the pyramids in the medulla and, on occasion, the precentral gyrus. Some reduction in the white matter of the corticospinal tract can occur. At the microscopic level, neuronal and axonal loss is seen. This is characterised

by degeneration and loss of the large MNs in the anterior horn (AH) of the SC and the motor nuclei of the brainstem. Myelinated axon loss is present in both the lateral and anterior corticospinal tracts of SC and there may be decrease in AH size (Saberi *et al.*, 2016). TDP43 proteinopathy is also a major neuropathologic marker of ALS, and is previously discussed in section 1.3.2.

1.4.1 Extramotor structures affected in ALS

All cases of ALS show TDP43 pathology in motor regions, with some cases displaying TDP43 pathology in nonmotor regions. Clinically, this continuum of pathology from motor system to motor and extra motor pathology to more dominant nonmotor pathology can be observed in patients who display motor symptoms, and those that present with cognitive symptoms, which is normally classified as FTD. Frontotemporal lobar degeneration-TDP (FTLD-TDP) cases display prominent nonmotor pathology in the frontal and temporal lobes (Arai *et al.*, 2006). Some cases of FTLD-DTP also have additional motor pathology. Some patients present with both motor and cognitive symptoms, and can be diagnosed with FTD/ALS(Mackenzie, Rademakers and Neumann, 2010). There is significant correlation between the location of TP43 pathology and the clinical phenotype of ALS and FTD (Tan *et al.*, 2015).

Neuronal TDP43 pathology has been observed in the hippocampus and neocortex (Geser *et al.*, 2008), as well as in the hippocampus and neocortex (Neumann *et al.*, 2006). Extramotor involvement in ALS pathology is noted in stage II of pathology, as determined by (Brettschneider *et al.*, 2013) (discussed in section 1.3.2), with the prefrontal cortex showing signs of pathology. This spreads to involve further extramotor regions in stages III and IV. Microglial activation can also be observed in extramotor regions, and will be further discussed in section 1.4.2.

1.4.2 Neuroinflammation and glial pathology in ALS

Neuroinflammation in ALS is characterised by astroglial activation, T-lymphocyte infiltration, overproduction of inflammatory cytokines and microglial activation.

Microglia are monocytes, forming the active immune defence system of the CNS, and can have both neuroprotective and neurotoxic functions. Microglia become activated and undergo morphology changes, broadly classified as either M1 (classical activation/pro-inflammatory) or M2 (anti-inflammatory), which can be further sub-divided into alternative activation and acquired deactivation. However, recent studies have suggested that this classification system is too simplistic, with some microglia not falling into either of these categories (Friedman *et al.*, 2018; Böttcher *et al.*, 2019). This was determined through single cell analysis of mouse and human cortex, where phenotypes that were transcriptionally

distinct from the M1/M2 classifications were found. Microglia have multiple functions, including: CNS development, innate immune function (pathogen recognition), phagocytosis, cytotoxicity and inflammation/immune response modulation (Boche *et al*, 2013). In ALS, microglia are activated. Widespread microglial activation has been observed in living ALS brains using positron emission tomography (PET), with a correlation between intensity of microglial activation in motor cortex (MCx) and severity of clinical MN deficits observed (Liu and Wang, 2017). This microglial pathology also correlates with UMN degeneration severity in ALS, with activated microglia releasing proinflammatory cytokines and reactive oxygen species in response to neuronal stress, leading to increased inflammation in ALS brains (Saberi *et al.*, 2016).

Microglial pathology is also present in nonmotor and extramotor regions in ALS (Brettschneider *et al.*, 2012), with PET analysis showing microglial activation in the cerebral cortex (particularly the MCx), supplementary motor regions and temporal cortex (Corcia *et al.*, 2012). Astrocytes have a number of roles, including supporting endothelial cells of the blood brain barrier (BBB), extracellular ion balance maintenance, providing nutrients to nervous tissue and repair of the brain/CNS after injury, with a trophic effect on neurones (Saberi *et al.*, 2016). Astrocytes have been implicated in ALS pathology (Allen, Shaw and Ferraiuolo, 2017). Pathologic astrocytes display increased immunoreactivity for GFAP and the calcium binding protein S100β, as well as expressing inflammatory markers (Saberi *et al.*, 2016).

1.5 DNA damage in ALS

DNA damage is a mechanism proposed to be involved with the pathogenesis of a number of neurodegenerative diseases, including ALS. DNA damage can be caused by the cell itself, referred to as endogenous DNA damage. This can be caused in a number of ways, including: oxidative damage, apoptosis, excision repair and depurination (disruption of the bond between deoxyribose in the DNA backbone and either base A or G, causing the base to be removed, resulting in a depurinated sugar). DNA damage can also be caused by external influences that the cell is exposed to. This is referred to as exogenous DNA damage, and includes; physical damage caused to the cell, such as radiation, or through chemical agents, such as cytotoxic drugs (Brown and Jackson, 2015). This damage can have a number of forms, including damage to the sugars or bases of DNA, single stranded breaks (SSBs) and double stranded breaks (DSBs). DSBs may be lethal to the cell and may not be capable of being repaired (Kuo and Yang, 2008).

Histones are proteins that form a major component of chromatin. The function of a histone is to package and order DNA into nucleosomes (structural subunits), with histones found to

play a role in gene regulation. There are currently five known families of histones. One of these families is H2A. This core histone molecule is made up of individual histone proteins: H2A, which has the greatest number of variants (H2A.1, H2A.2, H2A.X and H2A.Z), as well as H2B, H3 and H4. This histone core, along with DNA, make up the nucleosome complex (see figure 1.4). H2AX is a variant of the H2A protein, with H2AX making up around 10% of the total H2A protein in humans. The H2AX protein has been found to be incorporated into histones throughout the DNA and is unique in its structure due to its carboxyl tail. The carboxyl tail sequence is highly conserved, with one serine at position 139, and one glutamine residue at position 40. Together, the serine and glutamine residue make up the SQ motif. In the presence of DNA damage, H2AX is phosphorylated on serine at position 139 to produce γH2AX. The γH2AX protein is one major marker of DSBs, it is recruited to sites of DNA damage, where it then recruits other components of the DNA repair machinery.

DNA damage response (DDR) is activated by a cell with the aim of repairing damage. However, if the DNA cannot be repaired, this induces apoptosis. In one study, IHC identified an upregulation of DDR markers, including γH2AX, in lumbar MNs from C9ALS cases (Farg *et al.*, 2017). This is in contrast to Vazquez-Villaseñor *et al.*, 2019, who conducted IHC in the MCx and frontal association cortex (FACx). γH2AX+ neuronal and glial nuclei quantification in sALS indicated no difference in γH2AX expression. Potential conflict between these findings could be caused by a number of factors. This includes that different antibodies for γH2AX were used in each study. Another is that different regions were focused on, and finally, one study focuses on C9ALS cases, with the other focusing on sALS cases. It could be that differences in DNA damage exist between sALS and C9ALS cases.

Many other genes known to have mutations in some cases of ALS including *SETX*, *EWSR1* and *TAF15,* have roles in DNA damage and repair, suggesting a role for DNA integrity in ALS pathophysiology.
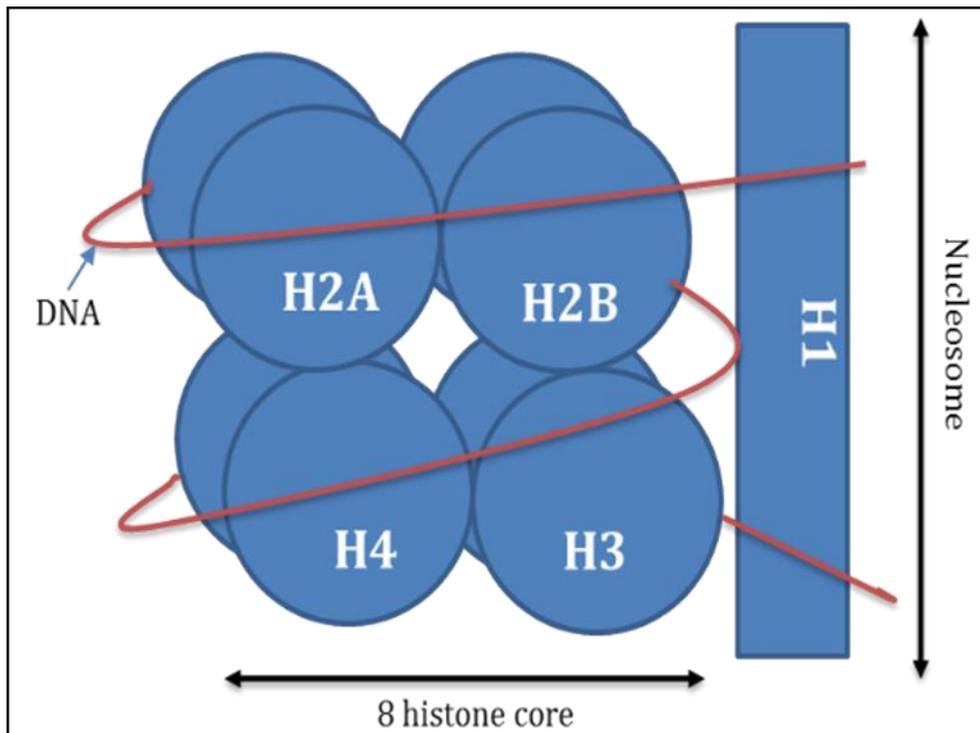
Figure 1.4: The core histone complex
The core histone molecule is made up of individual histone proteins;
H2A, which has the greatest number of variants, H2B, H3 and H4.
This histone core, along with DNA, make up the nucleosome complex.
H2AX is a variant of the H2A protein, with H2AX making up around
10% of the total H2A protein in humans. The H2AX protein has been
found to be randomly incorporated into histones throughout the DNA.
It is unique in its structure due to its carboxyl tail. The carboxyl tail
sequence is highly conserved, with one serine at position 139, and
one glutamine residue at position 40. Together, the serine and
glutamine residue make up the SQ motif. In the presence of DNA
damage, H2AX is phosphorylated on serine at position 139 to produce
γH2AX.

1.6 Epigenetics: DNA methylation

Epigenetic changes are defined as stable alterations that are made to either DNA or histone proteins that alter gene expression, but do not affect the base sequence. Epigenetic mechanisms are involved in gene expression regulation and are required for specific cell-type gene expression (Tammen, Friso and Choi, 2013). Little research has been conducted into the possibility of epigenetic pathology in ALS. DNA methylation is the most studied epigenetic mechanism in cancer and other forms of neurodegeneration including Alzheimer's disease (AD) (Qazi *et al.*, 2018), and is the process of attaching a methyl group to a cytosine. This usually results in gene silencing caused by the methylated cytosine directly preventing the binding of transcription factors to the gene promoter by changing the chromatin structure (see figure 1.5). DNA methylation has also been shown to be involved in promoting genomic stability (Robertson and Wolffe, 2000).

DNA methyltransferases (DNMTs) are responsible for the attachment of methyl groups to cytosine nucleotides in DNA and have two classes: *De novo* DNMTs are able to newly methylate cytosines, setting up DNA methylation patterns (DNMT3a and DNMT3b), while maintenance DNMTs maintain these DNA methylation patterns (DNMT1).

DNMT3a has previously been shown to have higher DNA methylation activity than DNMT3b (Takeshima *et al.*, 2006), and its role has not yet been elucidated in ALS. Hence, the priority of the DNMT IHC in this study focused on DNMT1 and DNMT3a, with the aim to move on to DNMT3b. The role of DNMTs in neurodegeneration is discussed further in section 1.12.

Cytosine methylation results in the formation of 5-methylcytosine (5mC). This can be oxidised by ten-eleven translocation (TET) enzymes to 5-hydroxymethylcytosine (5hmC). Further oxidation results in further demethylation into 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Ito *et al.*, 2011). Terminal deoxynucleotidyl transferase (TdT) or thymine DNA glycosylase (TDG) convert 5fC and 5caC back to unmethylated cytosine by base excision repair (Zhang *et al.*, 2012) (see figure 1.6).

5mC is implicated in gene expression repression, synaptic plasticity, gene imprinting, chromatin structure maintenance and X-chromosome inactivation. It is enriched in genes encoding proteins with function in neurones and other nervous system cells (Jin *et al.*, 2011). 5hmC is present in high levels in fully differentiated neurones (Coppieters *et al.*, 2014) and low levels in stem cells (Orr *et al.*, 2012), suggesting 5hmC is present in areas of low cell proliferation. Global 5hmC loss has been detected in cancers (Pfeifer, Kadam and Jin, 2013), suggesting that 5hmC cannot be maintained in proliferating cells (Jin *et al.*, 2011). 5hmC is mostly absent from non-gene-encoding regions of DNA (Jin *et al.*, 2011), and is largely found at CpG sites and promoter regions, as is 5mC. CpG sites are DNA regions

where a cysteine is followed by a guanine, separated by a phosphate (Illingworth and Bird, 2009).

5hmC levels have been shown to be generally present at 10% of the levels of 5mC in the genome (Branco, Ficz and Reik, 2012), but with higher levels present in the CNS (Globisch *et al.*, 2010). 5fC and 5caC levels are even less abundant in the genome than both 5mC and 5hmC (Ito *et al.*, 2011).

Figure 1.5: DNA methylation as a repressive modulator
DNA methylation is the most studied epigenetic mechanism in cancer and neurodegeneration, and is the process of attaching a methyl group to a cytosine. This usually results in gene silencing, and is referred to as a repressive modulator. This silencing is caused by the methylated cytosine directly preventing the binding of transcription factors to the gene promoter by changing the chromatin structure.

Figure 1.6: DNA methylation mechanism

5mC describes a cytosine nucleotide present within the DNA that has been modified by the addition of a CH_3 methyl group. This can be oxidised by TET enzymes, to actively demethylate 5mC into 5hmC. Further oxidation by TET enzymes results in further demethylation into 5fC and 5caC (Ito *et al.* 2011). 5fC and 5caC can be converted back into the unmethylated form of cytosine by TdT or TDG (Zhang *et al.* 2012).

1.7 DNA methylation and ageing

There are well-established alterations in gene expression in the brain with age (Dillman *et al.*, 2017), which may be due to alterations in methylation. Thus, Kraus and colleagues (Kraus *et al.*, 2015) quantified the number of neurones positive for nuclear 5hmC in neurologically healthy cases at post mortem at a variety of age ranges: foetus (range 17-30 weeks), adolescent (range 15-23 years), adult (range 39-42 years), elderly (range 57-68) and aged (range 78-81 years). It was found that the cortex had 50% greater numbers of positively-stained neuronal nuclei, while the white matter had 200% greater numbers of 5hmC-positive (5hmC$^+$) glia in the aged group compared to the other groups. The internal granular cell layer (largely neurones) and the molecular cell layer of cerebellum, composed of only glial cells, showed a significant increase of 5hmC from foetus until adulthood, where levels then started to decrease. Purkinje cells exhibited 5hmC positive nuclear staining in all age groups. As this study suggests there are varying levels of 5hmC over a lifespan, this raises the possibility of changes in 5hmC levels over time may play a role in ageing and neurodegeneration. Variations in 5hmC levels over time, plus the relatively higher levels of 5hmC in the brain compared to the rest of the body suggests it could be a target for dysregulation in abnormal ageing, and therefore a potential contributory factor in neurodegenerative diseases (López, Fernández and Fraga, 2017). A study of 5hmC in mouse cerebellum found increased levels of 5hmC genes associated with neurodegeneration in aged mice (Song *et al.*, 2011). Accordingly, as an increase in DNA hydroxymethylation is associated with ageing, it raises the possibility of vulnerability to age-related neurodegenerative diseases.

1.8 Epigenetic clocks

Epigenetic drift is defined as alterations of DNA methylation within cells associated with age, with a general net loss of DNA methylation during ageing (Horvath and Raj, 2018). Epigenetic age estimators, referred to as epigenetic clocks, indicate how biologically aged a specific tissue or blood is in comparison to the chronological age of an individual based on the pattern of gene methylation seen. Horvath's clock is the most well-known of the 'epigenetic clocks' (Horvath, 2013). Epigenetic age estimators use sets of CpGs, known as clock CpGs, plus an algorithm to estimate the age of a tissue. Epigenetic clock algorithms have been used to study age-related neurodegenerative conditions, including Alzheimer's disease (AD), dementia, Huntington's disease (HD) and Parkinson's disease (Horvath and Raj, 2019), with findings suggesting an accelerated epigenetic age in those with neurodegenerative conditions compared to those who were neurologically healthy. This suggests that increased epigenetic age could be a factor in other neurodegenerative diseases, including ALS.

1.9 DNA methylation in neurodegeneration

Abnormal global (within tissue as a whole) and local (within a specific cell population) DNA methylation patterns have previously been associated with several neurodegenerative diseases, including AD and HD (Pook, 2012; Lu *et al.*, 2013).

The majority of epigenetic studies have been carried out in AD. IHC studies of 5mC and 5hmC in AD have shown both hypo and hypermethylation in different brain regions, making it difficult to determine the global trend of methylation in AD/neurodegeneration for the brain overall (Roubroeks *et al.*, 2017). A study by Ellison *et al.* 2017 conducted gas chromatography/mass spectrometry (GC/MS) in a number of brain regions, analysing brains from both early and late stage AD. Findings show that there were global changes in both 5mC and 5hmC levels in the early stage brains, but that late stage brains showed similar 5mC and 5hmC levels to controls. A decrease in 5mC was found in cases displaying dementia with Lewy bodies (DLB) and frontotemporal lobar degeneration (FTLD) compared to controls. An increase in 5hmC was found in pre-clinical AD, in cases displaying mild cognitive impairment, and in FTLD/DLB cases compared to controls.

Lunnon *et al.*, 2014 found hypermethylation in the chromosomal region associated with the Ankyrin 1 (*ANK1*) gene in entorhinal cortex, superior temporal gyrus and prefrontal cortex of AD brains using the Illumina 450k array platform. This has 450,000 probes to detect DNA methylation at various sites. *ANK1* was also found to be hypermethylated in other studies of different brain regions, such as in the entorhinal cortex (Smith *et al.*, 2019) and the temporal cortex (De Jager *et al.*, 2014). Further, a study of the prefrontal cortex and superior frontal gyrus found HOXA to be hypermethylation. This is of interest, as HOX genes help maintain akyrin locus expression (Smith *et al.*, 2018). Another study, focusing on late-onset AD superior temporal gyrus, showed a number of differentially methylated regions (DMRs), mostly displaying hypermethylation (Watson *et al.*, 2016). Together, these findings indicate that not only do DNA methylation levels vary in different regions of the AD brain, but changes in DNA methylation also vary based on early or late stage AD.

Aside from AD, studies of HD have shown loss of 5hmC and increased 5mC (Villar-Menéndez *et al.*, 2013), in the putamen of HD patients, as well as genome-wide 5hmC loss in the striatum and cortex of HD mouse brain (Wang *et al.*, 2013). Further studies are needed to identify if this is a causal factor in disease onset, or has arisen as a consequence of disease progression.

1.10 DNA Methylation in ALS

Thirty six percent of *C9orf72* expansion carriers display expansion-specific hypermethylation, observed in the 5' CpG island in the promoter region of *C9orf72*, around 200 base pairs upstream of the 5' end of the GGGGCC repeat (Xi *et al.*, 2012).

High variability in global methylation status across the genome has been detected between different familial groups with *C9orf72* mutations, as determined by predicting the two CpG islands immediately flanking the *C9orf72* repeat. However, similar methylation levels in *C9orf72* were detected within familial groups (Xi *et al.*, 2013). A correlation was observed between a higher degree of methylation and shorter disease duration. This disease modifying mechanism is similar to that seen in Huntington's disease, which is also caused by a repeat expansion and is similarly associated with differential methylation of the *HDD* gene (Villar-Menéndez *et al.*, 2013). No hypermethylation was detected in persons with a normal number of repeats, defined as up to 43 in this study.

CpG islands and sites near to promoter regions of nearby genes in the region of the *C9orf72* expansion repeat were also hypermethylated, suggesting more CpG islands are present due to the *C9orf72* repeat expansion, resulting in increased methylation as a result of the repeat, leading to lower mRNA expression. These findings have to be regarded with a little caution, as this study used homogenated tissue, which can mask each cell type's individual contribution to disease. Looking at DNA methylation of specific cell types would be of use to determine each cell type's role in the onset of disease: ALS is characterised by neuronal cell death, astrogliosis and microgliosis. This may result in altered cell proportions, which could affect methylation profiles observed when the various cell types are pooled without compensation for intergroup differences in cellular composition.

Ebbert *et al.*, 2017 conducted a study of human post-mortem (PM) cerebellum and frontal cortex, and performed reduced representation bisulphite sequencing (RRBS). This method combines the use of restriction enzymes with bisulphite conversion, resulting in the enrichment of areas of the genome with high CpG content. Differentially methylated DNA methylation profiles were found, including in *TARDBP* and *RANGAP1*, a gene that interacts with *C9orf72* expanded RNA. This implicates ALS-related pathologies in extra motor and non-motor regions, and could represent an 'early-phase' of ALS, as the motor regions are generally more affected, with neurodegeneration more evident. Collectively, these studies suggest a role of DNA methylation in C9ALS.

1.11 Gene methylation and sporadic ALS

Gene methylation may be significant to ALS aetiology outside the context of *C9orf72* mutations: The first study to investigate the role DNA methylation played in ALS was an epigenome wide association study (EWAS) (Morahan *et al.*, 2009). Frontal cortex from ten sALS patients and 10 controls underwent DNA methylation analysis using Affymetrix GeneChip Human Tiling 2.0R arrays. A number of genes were found to be differentially methylated, with pathway analysis showing altered DNA methylation in genes involved in calcium homeostasis, oxidative stress and neurotransmission. Aside from demonstrating altered methylation, albeit with the limitation of using tissue homogenates, this study demonstrates that regions outside of the motor system are affected in ALS.

A second study to use EWAS in ALS, has shown global increases in 5mC and 5hmC levels in post mortem sALS spinal cord, but not in blood (Figueroa-Romero *et al.*, 2012). This suggests that the DNA methylation profiles in the CNS differ from those of blood. This highlights that blood biomarkers may not be a good indicator of crucial DNA methylation changes in ALS and illustrates the need for understanding cell specificity of DNA methylation patterns. Genome-wide expression profiling of total RNA was also conducted in this study. Gene enrichment analysis of these data identified several biological functional categories that had also been identified in 5mC profiling, allowing comparisons between the 5mC methylation patterns and the mRNA expression analysis data. 112 epigenomes were identified as showing concordant direction of DNA methylation and mRNA expression changes. These had functional enrichment for immune response and neurone adhesion.

1.12 DNMTs and neuronal degeneration

Feng *et al.*, 2010 knocked down *DNMT1* and *DNMT3a* in the forebrain excitatory neurones of adult mice and found that while there was no neuronal loss, the size of the neurones were significantly smaller than that of the control type mice. Behavioural changes related to learning and memory were found, with DNA methylation levels significantly decreased, and altered expression in genes related to synaptic plasticity. This study highlights the relevance of studying DNMTs in neurodegeneration.

Chestnut *et al.*, 2011 tested the hypothesis that DNMTs could mediate neuronal cell death. DNMT3a expression was experimentally promoted in NSC34 cell culture, which led to degeneration. Camptothecin was then used to induce apoptosis. In response to this, levels of DNMT1 increased fivefold, with levels of DNMT3a increasing twofold, as detected by DNMT enzyme-linked immunosorbent assay (ELISA). An accumulation of 5mC was also seen in the nucleus of these cells, detected by immunofluorescence microscopy. This finding suggests that neuronal degeneration and death is regulated by DNMTs.

Oh *et al.*, 2016 isolated human bone marrow mesenchymal stromal cells from ALS patients and found that DNMT1 was excessively expressed. When these cells were treated with the DNMT inhibitor RG108, PCR analysis confirmed that there was increased expression of the anti-senescence genes *TERT, VEGF* and *ANG*, and a decrease of the expression of senescence genes *ATM* and *p21*. This suggests that excessive DNMT1 (and by association, potentially DNA methylation) could be associated with increased senescence. Štalekar *et al.*, 2015 knocked down TDP43 in a neuroblastoma cell line, and found that expression of DNMT3a was significantly downregulated. Together, this suggests that DNMT1 and DNMT3a are potential players in the pathogenesis of ALS, and that they may have a functional involvement with TDP43, a protein known to be mutated in ALS. Of note is the availability of pre-existing drugs targeting DNA methylation, in the form if DNMT inhibitors, providing possible candidate treatments (Foulks *et al.*, 2012; J. Li *et al.*, 2017).

1.13 Concluding remarks: What next for DNA methylation in ALS?

Only a small amount of work has been conducted thus far to elucidate the effects of DNA methylation in ALS. Work performed on other neurodegenerative diseases, in particular AD, gives promise of a possible association between epigenetics and ALS that warrants further investigation. The most pressing matter is the ability to conduct experiments on individual cell types. The contribution of each cell type present within the brain and CNS would be expected to have different DNA methylation patterns, with all cells contributing to disease. Therefore, determining each cell type's specific role in relation to disease is important. DNA methylation patterns in individual cell types within the normal ageing brain also need to be studied in order to understand which DNA methylation patterns are present within healthy individuals. This would aid further studies on DNA methylation in neurodegeneration. Aside from DNA methylation, a small study has been conducted as part of this thesis to assess DNA damage in ALS. Data from this supports prior findings of a lack of DNA damage in neurones in ALS (Vazquez-Villaseñor *et al.*, 2019), and therefore the main focus of this thesis will be ALS and DNA methylation.

1.14 Hypotheses and Aims

This study has the following aims:

<u>Overall aim 1:</u> Assess global DNA methylation, DNA hydroxymethylation and DNA damage levels using immunohistochemistry.

1. To conduct a pathological investigation of the levels of DNA methylation and hydroxymethylation in control, sALS and C9ALS CNS spinal cord (SC), motor cortex (MCx) and anterior frontal cortex (AFCx) using IHC for 5mC and 5hmC. This will be conducted for anterior horn (AH) LMNs, as well as glia of the AH, lateral corticospinal tract (LCT) and dorsal column (DC) in SC, as well as glia in the white matter and UMNs in the grey matter of MCx and neurones in AFCx.

2. To investigate the methylation status of LMN cells with and without TDP43 proteinopathy in the anterior horn using the same methods as in aim 1.

3. To use the spread of TDP43 proteinopathy in the brain to study DNA methylation in controls and sALS with TDP proteinopathy versus those that have not developed TDP43 proteinopathy in the anterior frontal cortex . The sALS cases without TDP43 proteinopathy may also provide data on any early changes that occur, as an intermediate between control and sALS with TDP43 pathology cases. *C9orf72*-ALS cases will also be assessed in this region. These cases are characterised by significant TDP43 pathology in this locus.

4. To investigate the localisation and levels of γH2AX in SC for control, sALS and C9ALS cases using IHC.

5. To assess the relationship between γH2AX status and TDP43 proteinopathy in MNs, also using IHC.

These studies will test the following hypotheses in relation to aim 1:

1. There is greater expression of 5mC and 5hmC in ALS cases than in healthy controls in SC AH MNs and glia.

2. The same effect seen in SC will also be seen in MCx/upper motor neurones and AFCx.

3. Higher expression of 5mC and 5hmC will be observed in MNs displaying TDP43 proteinopathy

4. There will be no difference in γH2AX expression in glial and MNs cells in ALS versus controls

5. γH2AX expression will be higher in MNs with TDP43 proteinopathy compared to MNs not displaying TDP43 proteinopathy.

<u>Overall aim 2:</u> Quantify DNA methylation at single nucleotide resolution

1. To assess DNA methylation in MNs from control, sALS and C9ALS SC at the gene level using the Illumina Methylation 850k EPIC BeadChip arrays. MNs will be collected from PM formalin fixed paraffin embedded (FFPE) SC tissue using laser capture microdissection (LCM). DNA will then be extracted before undergoing bisulphite conversion and microarray analysis using the MethylationEPIC BeadChip. Data will be analysed using the R package RnBeads.

2. To attempt validation of the findings of the MethylationEPIC BeadChip using the creation of a bisulphite converted- next generation sequencing methyl library (BS-NGS) from the same cases.

3. To compare DNA methylation patterns from the MethylationEPIC BeadChip with mRNA expression data. The data collected from this will be used to assess the effect of DNA methylation on mRNA expression in diseased MNs.

These studies will test the following hypotheses in relation to aim 2:

1. Hypermethylation will be observed in ALS cases versus healthy controls.

2. Any methylation changes in MNs between ALS cases and healthy controls will be more marked in *C9orf72*-ALS cases versus sALS cases, with this methylation affecting the mRNA transcriptome on these neurones, therefore impacting upon disease process.

Chapter 2: Materials and Methods

2.1 Suppliers

Abcam PLC, Discovery Drive Cambridge Biomedical Campus, Cambridge CB2 0AX, UK.

Abgent Inc., 9765 Clairemont Mesa Blvd, Suite C, San Diego, CA 92124, USA.

Applied Biosystems Inc., 850 Lincoln Centre Drive, Foster City, CA 94404, USA.

Beckman Coulter, Oakley Ct Kingsmead Business Park London Road High Wycombe Buckinghamshire HP11 1JU.

Eppendorf Ltd., Eppendorf House, Gateway 1000 Whittle Way, Arlington Business Park, Stevenage SG1 2FP, UK.

Fisher Scientific Inc., Bishop Meadow Road, Loughborough, Leicestershire, LE11 5RG, UK.

GenWay Biotech. Inc., 6777 Nancy Ridge Drive San Diego, CA 92121, USA.

Hamamatsu Photonics UK Ltd., 2 Howard Court Tewin Road, Welwyn Garden City, Hertfordshire, AL7 1BW, UK.

Illumina®., Watson Building 11 , Granta park, Great Abington, Cambridge, CB21 6GP, UK.

Leica Microsystems Ltd., Davy Avenue, Knowhill, Milton Keynes, MK5 8LB, UK.

Life Technologies Ltd., 3 Fountain Drive, Inchinnan Business Park, Paisley, PA4 9RF, UK.

Proteintech Group., 4th Floor, 196 Deansgate, Manchester, M3 3WF, UK.

Qiagen., Skelton House, Lloyd St N, Manchester M15 6SH.

R&D Systems, 614 McKinley Place NE, Minneapolis, MN 55413, USA.

Santa Cruz Biotechnology, Bergheimer Str. 89-2, 69115 Heidelberg, Germany.

Thermo Scientific., 168 Third Avenue, Waltham, MA 02451, USA.

Vector laboratories Ltd., 3 Accent Park, Bakewell Road, Orton Southgate, Peterborough, PE2 6XS, UK.

Zymo Research., Mülhauser Str. 9, 79110 Freiburg im Breisgau, Germany.

2.2 Immunohistochemistry

A table of standard laboratory solution preparations used is summarised in appendix II.

2.2.1 Human tissue for immunohistochemistry

Formalin-fixed paraffin-embedded (FFPE) human PM SC, MCx and AFCx was obtained from the Sheffield Brain Tissue Bank (SBTB). For this study, sections were used from three groups: controls, sALS and ALS caused by *C9orf72* mutations (C9ALS). Groups were age- and sex-matched, as far as was possible. Case details are summarised in table 2.1 for γH2AX MNs and γH2AX glia, table 2.2 for SC DNA methylation and hydroxymethylation, table 2.3 for FCx DNA methylation and hydroxymethylation and table 2.4 for MCx DNA

methylation and hydroxymethylatino. For the γH2AX study, full demographics for the cohort were unavailable. This resulted in being unable to statistically control for variables within this study. While this is not ideal, it is representative of the limitations of working with human tissue. Previous findings within the laboratory group found no difference in γH2AX was observed between ALS and controls (Vazquez-Villaseñor *et al.*, 2019) (discussed in section 1.5). However, it was considered important to determine if there is a difference within the ALS group for MNs displaying TDP43 pathology, and those that were not. This latter stage, focusing only on ALS cases, have full demographic details available.

Ethical approval for this study has been granted by the Management Committee of SBTB, which in turn has ethical approval to provide tissue for research under the provision to act as a Research Tissue Bank, as approved by the Scotland A Research Ethics Committee (ref. 08/MRE00/103) (see appendix I).

Table 2.1: Case details for gamma H2AX MN and glial cell counts
NK = not known
Key:
Cases highlighted in black were used for MN studies only.
Cases highlighted in red were used for MN and glial studies.

Case ID	Sex	Age	Disease
144/1991	M	NK	Control
2108/1999	M	NK	Control
098/2007	M	67	Control
808/1990	M	NK	Control
309/1990	M	82	Control
019/1991	M	54	Control
041/2008	F	60	sALS
014/2011	M	51	sALS
072/2005	M	66	sALS
094/2006	M	71	sALS
005/2010	M	40	sALS
099/2009	M	79	sALS
023/2010	F	42	sALS
081/2009	M	59	C9ALS
039/2011	M	72	C9ALS
066/2008	F	59	C9ALS
273/1999	M	68	C9ALS
118/2001	M	66	C9ALS
098/2002	F	64	C9ALS
041/2004	M	64	C9ALS

Averages (plus standard deviations) Control = 67.7(SD=14.0) sALS = 58.5 (SD=14.8) C9ALS = 64.6 (SD=4.9)	Glial cases only averages (plus standard deviations) Control = 67 (SD=0) sALS = 62 (SD=8.6) C9ALS = 64.5 (SD=6.6)
Age T-Tests Control v sALS = 0.39 Control v C9ALS = 0.60 sALS v C9ALS = 0.31	Glial cases only age T-Tests Control v sALS = 0.29 Control v C9ALS = 0.48 sALS v C9ALS = 0.66
Sex splits: Control = 6M, 0F sALS = 5 M, 2F C9ALS = 5 M, 2F Overall = 16M, 4F	Sex splits: Control = 4M, 0F sALS = 3M, 1F C9ALS = 3M, 1F Overall = 10M, 2F

Table 2.2: Cases details for samples used in spinal cord DNA methylation and hydroxymethylation studies

NK = not known
NA = not applicable

Case ID	Sex	Age (years)	Disease duration (months)	Disease Status
012/2007	M	63	NA	Control
080/1992	F	62	NA	Control
085/2007	F	59	NA	Control
098/2007	M	67	NA	Control
023/1992	F	75	NA	Control
035/1996	F	87	NA	Control
071/1992	F	84	NA	Control
293/1991	M	65	NA	Control
039/1997	M	53	NA	Control
005/2007	M	63	NA	Control
141/2003	M	75	7	sALS
055/2012	F	72	NK	sALS
024/2008	F	63	46	sALS
027/2008	F	65	44	sALS
039/2005	F	61	48	sALS
087/1992	F	75	37	sALS
200/1997	M	73	47	sALS
034/2005	M	63	38	sALS
072/2005	M	66	10	sALS
094/2006	M	72	9	sALS
063/2008	F	69	26	C9ALS
041/2004	M	64	20	C9ALS
083/2010	M	79	137	C9ALS
046/2010	F	67	43	C9ALS
053/1996	F	65	40	C9ALS
073/2007	F	65	24	C9ALS
098/2002	F	64	42	C9ALS
273/1999	M	68	14	C9ALS
039/2011	M	72	26	C9ALS
118/2001	M	66	31	C9ALS

Averages (including standard deviations)
Control = 67.8 (SD=10.9)
sALS = 68.5 (SD=5.4)
C9ALS = 67.9 (SD=4.6)

Age T-tests
Control v sALS p= 0.86
Control v C9ALS p= 0.98
sALS v C9ALS p= 0.79

Sex splits:
Control = 5M, 5F
sALS = 5M, 5F
C9ALS = 5M, 5F
Overall = 15M, 15F

Disease duration averages (including standard deviations)
sALS = 40.3 (SD = 35.3)
C9ALS = 31.8 (SD = 17.8)
Combined ALS groups = 36.3 (SD = 28.8)

Table 2.3: Case details for samples used in frontal cortex DNA methylation and hydroxymethylation studies

sALSnoPath = sALS with no TDP43 pathology, sALSPath = sALS with TDP43 pathology.

Case ID	Disease Status	Sex	Age (years)	Case ID	Disease Status	Sex	Age (years)
085/2007	Control	F	59	006/2002	sALSnoPath	F	64
135/2018	Control	F	85	345/1990	sALSPath	M	68
224/2016	Control	F	82	110/1992	sALSPath	M	69
009/2018	Control	M	69	336/1990	sALSPath	M	63
098/2007	Control	M	68	026/1994	sALSPath	M	59
012/2007	Control	M	63	096/2006	sALSPath	M	51
033/2012	Control	M	78	086/2008	sALSPath	M	66
070/2007	Control	M	26	014/2011	sALSPath	M	51
007/2009	Control	M	39	015/1998	sALSPath	F	86
005/2007	Control	M	63	024/2008	sALSPath	F	63
010/1996	Control	F	65	050/2008	sALSPath	F	70
096/2008	sALSnoPath	F	69	098/2008	sALSPath	F	75
212/1999	sALSnoPath	M	64	059/2009	sALSPath	F	80
102/2005	sALSnoPath	M	64	005/2010	sALSPath	M	40
049/2006	sALSnoPath	M	91	073/2007	C9ALSPath	F	64
064/2009	sALSnoPath	M	67	046/2010	C9ALSPath	F	67
074/2009	sALSnoPath	M	69	069/2009	C9ALSPath	M	48
094/2009	sALSnoPath	M	63	081/2009	C9ALSPath	M	59
099/2009	sALSnoPath	M	79	039/2011	C9ALSPath	M	72
064/2010	sALSnoPath	M	78	083/2010	C9ALSPath	M	79
039/2005	sALSnoPath	F	57	118/2001	C9ALSPath	M	66
037/2004	sALSnoPath	F	66	098/2002	C9ALSPath	F	64
049/2005	sALSnoPath	F	64	063/2008	C9ALSPath	F	69
041/2008	sALSnoPath	F	60	066/2008	C9ALSPath	F	59
075/2008	sALSnoPath	F	61	006/2006	C9ALSPath	F	75
010/2009	sALSnoPath	F	60	040/2011	C9ALSPath	F	67
041/2010	sALSnoPath	F	69				

<div align="center">

Averages (plus standard deviations)
Control = 71.7 (SD=4.3)
sALS no TDP43 pathology = 75.8 (SD=6.0)
sALS with TDP43 pathology = 76.0 (SD=5.4)
C9ALS = 75.0 (SD=6.6)

Age T-Tests
Control v sALS no TDP43 pathology p= 0.43
Control v sALS with TDP43 pathology p= 0.83
Control v C9ALS p= 0.68
sALS no TDP43 pathology v sALS with TDP43 pathology p= 0.49
sALS no TDP43 pathology v C9ALS p= 0.61
sALS with TDP43 pathology v C9ALS p= 0.81

Sex splits:
Control = 7M, 4F
sALS = 16M, 14F
C9ALS = 5M, 7F
Overall = 28M, 25F

</div>

Table 2.4: Case details for samples used in motor cortex DNA methylation and hydroxymethylation studies

Case ID	Sex	Age (years)	Disease Status	Case ID	Sex	Age (years)	Disease Status
010/1996	F	65	Control	005/2010	M	40	sALS
135/2018	F	85	Control	041/2008	F	60	sALS
224/2016	F	82	Control	059/2009	F	80	sALS
012/2007	M	63	Control	074/2009	M	69	sALS
033/2012	M	78	Control	091/2008	M	53	sALS
085/2007	F	59	Control	094/2009	M	63	sALS
098/2007	M	67	Control	099/2009	M	79	sALS
005/2007	M	63	Control	023/2010	F	42	sALS
094/2006	M	72	sALS	118/2001	M	66	C9ALS
200/1997	M	73	sALS	083/2010	M	79	C9ALS
141/2003	M	75	sALS	273/1999	M	68	C9ALS
072/2005	M	66	sALS	046/2010	F	67	C9ALS
034/2005	M	63	sALS	066/2008	F	59	C9ALS
039/2005	F	57	sALS	040/2011	F	67	C9ALS
002/2006	F	64	sALS	053/1996	F	61	C9ALS
004/2006	F	64	sALS	073/2007	F	64	C9ALS
024/2008	F	63	sALS	045/2006	M	46	C9ALS
027/2008	F	70	sALS	006/2006	F	75	C9ALS
102/2017	M	53	sALS	063/2008	F	63	C9ALS
026/1994	M	58	sALS	041/2004	M	62	C9ALS
025/2013	F	72	sALS				

Averages (plus standard deviations)
Control = 70.3 (SD=9.9)
sALS = 63.6 (SD=10.7)
C9ALS = 64.8 (SD=9.2)

Age T-Tests
Control v sALS = 0.14
Control v C9ALS = 0.19
sALS v C9ALS = 0.75

Sex splits:
Control = 4M, 4F
sALS = 12M, 9F
C9ALS = 5M, 7F
Overall = 21M, 20F

2.2.2 Immunohistochemistry – an overview

IHC was conducted to visualise the cellular localisation antigens of interest in cells and tissues. The standard avidin/biotinylated enzyme complex (ABC) staining method was utilised, in conjunction with 3,3'-diaminobenzidine (DAB) (both Vector Laboratories, UK) for visualisation. All work was carried out at room temperature (RT), unless otherwise stated.

5 µm sections of SC were cut onto positively charged slides. Sections were cut sequentially onto two or three slides, as described in figure 2.1 and dried at 37 ˚C overnight. Sectioning in

this manner allows the same MN nuclei to be visualised in multiple sections. For frontal and MCx, one 5 μm section per slide was used. One of two conditions was used for antigen retrieval. These conditions are summarised in figure 2.2.

Figure 2.1: Diagrammatic representation of the comparison between 5mC/5hmC and TDP43 (A) and γH2AX and TDP43 (B) stained spinal cord in sequential sections.
Tissue was sectioned onto sequential slides, allowing the same MN nucleus to be visualised in sequential sections so that cell counting semi-quantitative analysis can occur.

TSC pH6.5/microwave		pH6/pressure cooker
1. Dewax in two changes of xylene for 5 mins each 2. Rehydration in a graded series of ethanols (100%, 100%, 95%, 70% (Fisher Scientific, UK)) for 5 mins each 3. Wash in distilled water 4. Submerge in tri-sodium citrate (TSC) pH6.5 5. Heat at 800W in microwave for 10 mins 6. Allow to cool to RT 7. Rinse in tap water	OR	1. Dewax in xylene for 5 mins 2. Transfer to pressure cooker and submerge in A. Menarini Antigen Access pH6 (A. Menarini diagnostics, Berkshire, UK) 3. Run at 125 °C and 20 psi for 30 s 4. Allow to cool to RT 5. Rinse in tap water

THEN: Blocking endogenous peroxidase activity

Immersion in 3% hydrogen peroxide in methanol (Fisher Scientific, UK) for 20 minutes to saturate endogenous peroxidases.

Figure 2.2: Details of the antigen retrieval methods used in immunohistochemistry
An overview of the two methods for antigen retrieval used in conjunction with primary antibodies.

2.2.3 Immunohistochemistry

IHC was carried out for γH2AX, TDP43, 5mC and 5hmC. Conditions for all primary antibodies used are summarised in table 2.5. Mrs Lynne Baxter conducted the immunostaining for γH2AX. Antibody optimisation was also carried out for numerous DNMT1 and DNMT3a primary antibodies, also summarised in table 2.5.

Sections were subjected to IHC using the standard ABC technique. The mouse Vectastain Elite kit was used for 5mC, with the rabbit Vectastain Elite kit used for γH2AX, TDP43 and 5hmC (Vector Laboratories, UK). The substrate DAB (Vector Laboratories, UK) was used in conjunction with all Vectastain Elite kits (Vector Laboratories, UK).

Positive controls were used to assess the efficiency of the staining. Either negative controls or IgG controls were conducted, with the primary antibody step omitted from one section in each IHC run.

1) **Section blocking**: either 1.5% normal horse serum (150 μL horse serum concentrate in 10 mL Tris-buffered saline (TBS)) for 5mC, or 1.5% goat serum (150 μL goat serum concentrate in 10 mL TBS) for γH2AX, TDP43 and 5hmC for 30 min. erum was removed from sections using gentle tapping.

2) **Primary antibody incubation**: 150 μL mouse monoclonal 5mC antibody was added to the section at 1:100 dilution overnight RT (Genway Biotech, GWB-BD5190). 150 μL TDP43 rabbit monoclonal antibody was applied to section at 1:4,000 dilution for 1 hour RT (Proteintech, cat no. 10782-2-AP), with 150 μL 5hmC rabbit monoclonal antibody applied at 1:32,000 for 1 hour at RT (Abcam, cat no. ab214728).

3) Sections were washed in TBS for 5 min.

4) **Secondary antibody incubation**: 0.5% biotinylated secondary antibody was applied for 30 min (50 μL in 1.5% horse serum for 5mC, 50 μL in 1.5% goat serum for γH2AX, TDP43 and 5hmC).

5) Sections were washed in TBS for 5 min.

6) **ABC incubation**: ABC reagent was applied (100 μL reagent 'A' and 100 μL reagent 'B' in 10 mL TBS, prepare at least 30 minutes before use) and incubated for 30 min.

7) Sections were washed in TBS for 5 min.

8) **DAB incubation**: DAB substrate (Vector Laboratories, UK) (100 μLμL buffer, 200 μL DAB and 100 μL hydrogen peroxide with 5 mL dH_2O) was prepared and incubated on sections for 2-10 min.

9) Sections were washed in dH_2O to quench any further enzymatic action.

10) Sections were counterstained with Harris' haematoxylin (Leica, UK) for 2 min, followed by rinsing in tap water.

11) Sections were submerged in Scott's tap water to blue before rinsing in tap water.

12) **Rapid dehydration**: Sections were dehydrated in graded alcohols (70%, 95%, 100% and 100%) for 30 seconds each.

13) Sections were cleared in xylene for 5 min and mounted using DPX (Leica, UK) and glass cover slips (Fisher Scientific, UK).

14) Sections were incubated overnight at 37°C to set.

Table 2.5: Primary antibody details.
Summary of the manufacturer, species, clonality and optimised immunohistochemistry conditions for all primary antibodies used. This includes what dilution antibodies are used at, length and temperature of incubation and preferred antigen retrieval method.

Antibody	Species	Clonality	Dilution	Primary antibody incubation conditions	Antigen retrieval method	Supplier
5-methylcytosine (5mC)	Mouse	Monoclonal	1:100	Overnight at RT	TSC pH 6.5/microwave	Genway Biotech, GWB-BD5190
TDP43	Rabbit	Polyclonal	1:4000	1 hour at RT	TSC pH6.0/pressure cooker	Proteintech, cat no. 10782-2-AP
5-hydroxymethylcytosine (5hmC)	Rabbit	Monoclonal	1:32,000	1 hour at RT	TSC pH 6.0/pressure cooker	Abcam, cat no. ab214728
γH2AX	Rabbit	Polyclonal	1:1,000	1 hour at RT	EDTA pH8/pressure cooker	R&D systems, cat no. AF2288

DNMT primary antibody details.
Summary of the manufacturer, species and clonality for all DNMT primary antibodies used.

Antibody	Species	Clonality	Supplier
DNA Methyltransferase 1 (DNMT1)	Rabbit	Polyclonal	Abcam, cat no. ab19905
DNA Methyltransferase 1 (DNMT1)	Mouse	Monoclonal	Novus Biologicals, cat no. 60B1220.1
DNA Methyltransferase 1 (DNMT1)	Mouse	Monoclonal	Santa Cruz Biotechnology, cat no. sc-271729
DNA Methyltransferase 1 (DNMT1)	Rabbit	Polyclonal	Proteintech, cat no. 24206-1-AP
DNA Methyltransferase 1 (DNMT1)	Rabbit	Polyclonal	Abgent, cat no. AP1032b
DNA Methyltransferase 1 (DNMT1)	Sheep	Polyclonal	R&D Systems, cat no. AF6110
DNA Methyltransferase 3a (DNMT3a)	Goat	Polyclonal	Santa Cruz Biotechnology, cat no. sc-10232
DNA Methyltransferase 3a (DNMT3a)	Rabbit	Polyclonal	Abcam, cat no. ab4897
DNA Methyltransferase 3a (DNMT3a)	Rabbit	Polyclonal	Abgent, cat no. AP1034a

2.2.4 Antibody optimisation

Optimisation was carried out for 5mC, 5hmC, DNMT1 and DNMT3a antibodies to assess the most appropriate antigen retrieval method and antibody dilution. A number of conditions were trialled, summarised in table 2.6.

Table 2.6: Antibody optimisation trials summary

Antibody	5mC Genway	5hmC Abcam	DNMT1 Abcam	DNMT1 Novus	DNMT1 Santa Cruz	DNMT1 Proteintech	DNMT1 Abgent	DNMT1 R&D systems	DNMT3a Santa Cruz	DNMT3a Abcam	DNMT3a Abgent
Dilutions trialled	• 1:50 • 1:100 • 1:200 • 1:500 • 1:1,000	• 1:1,000 • 1:2,000 • 1:5,000 • 1:10,000 • 1:32,000 • 1:64,000					• 1:50 • 1:100 • 1:200 • 1:500 • 1:1,000				
Antigen retrieval methods trialled	• TSC pH6.5/microwave • A. Menarini Antigen Access pH6 solution/pressure cooker • A.Menarini Antigen Access pH9 solution/pressure cooker										
Primary antibody incubation periods trialled	• 1hour/RT • Overnight/RT	• 1hour/RT • Overnight/RT					• 1hour/RT • Overnight/RT • Overnight/4°C				

2.2.5 Antibody specificity

Antibody specificity checks were carried out for both 5mC and 5hmC to ensure that any immunopositive staining observed was specific. Pre-absorption of the antibodies was carried out, as well as DNase treatments. Positive and negative controls were also used throughout.

2.2.6 Slide scanning

All slides that underwent IHC for γH2AX, TDP43, 5mC, and 5hmC were scanned using the Hamamatsu NanoZoomer slide scanner (Hamamatsu Photonics, Japan) to create a digital whole slide images.

2.2.7 Inter-rater reliability testing

Cohen's kappa coefficient (Cohen 1960) was used to assess the robustness of the method and ensure reproducibility. This method allows for a measure of inter-rater agreement to be calculated. This method takes the possibility of chance agreement into account, and is therefore a more robust indication of inter-rater reliability than simple percentage agreement. A series of both MNs and glia were classified by two independent observers (JRH and CAM). Inter-rater reliability testing was conducted for both 5mC and TDP43 staining, as well as in all three groups. This is summarised in figure 2.3.

An inter-rater reliability test (Cohen's Kappa) was conducted to assess the reliability, validity and reproducibility of cell counting. This calculation is a chance-corrected version of the observed agreement. High agreement between observers indicates diagnosis consensus as well as consensus on the interchangeability and reliability of the ratings. All cell counting was conducted over seven cases, including three control cases, two sALS cases and two C9ALS cases, with a total of 140 cells graded independently by the two raters (JRH and CAM) per assessment. Inter-rater reliability levels were ≥0.82, indicating high reproducibility between raters (see figure 2.3). An example of γH2AX immunopositivity can be seen in figure 3.1.

An inter-rater reliability test in the form of Cohen's Kappa calculation was conducted to assess the reliability, validity and reproducibility of cell counting. This calculation is a chance-corrected version of the observed agreement. High agreement between observers indicates diagnosis consensus as well as consensus on the interchangeability and reliability of the ratings. All cell counting was conducted over seven cases, including three control cases, two sALS cases and two C9ALS cases, with a total of 140 cells graded independently by the two raters (JRH and CAM) per assessment. Inter-rater reliability levels were ≥0.81, indicating high reproducibility between raters (table 2.3).

$$k = \frac{p_o - p_e}{1 - p_e}$$

A

B

0.1-0.2	Slight agreement
0.21-0.4	Fair agreement
0.41-0.6	Moderate agreement
0.61-0.8	Substantial agreement
0.81-1.0	Near perfect agreement

C

Spinal cord: 5mC/5hmC/TDP43		Spinal cord: γH2AX/TDP43		Frontal cortex: 5mC/5hmC		Motor cortex: 5mC/5hmC	
Case ID	Disease Status	Case ID	Disease Status	Case ID	Disease Status	Case ID	Disease Status
005/2007	Control	098/2007	Control	035/1996	Control	033/2012	Control
012/2007	Control	808/1990	Control	098/2007	Control	098/2007	Control
071/1992	Control	144/1991	Control	293/1991	Control	085/2007	Control
200/1997	sALS	094/2006	sALS	034/2005	sALS	034/2005	sALS
039/2005	sALS	072/2005	sALS	039/2005	sALS	072/2005	sALS
063/2008	C9ALS	066/2008	C9ALS	041/2004	C9ALS	046/2010	C9ALS
273/1999	C9ALS	273/1999	C9ALS	273/1999	C9ALS	118/2001	C9ALS

D

Area	Test	Cohen's Kappa
Spinal cord	γH2AX glia	0.82
	γH2AX motor neurones	0.86
Spinal cord	TDP43 motor neurones	0.96
	5hmC glia	0.98
	5hmC motor neurones	0.91
Motor cortex	5mC glia	0.82
	5mC motor neurones	0.81
	5hmC glia	0.90
	5hmC motor neurones	0.84
Anterior frontal cortex	5mC glia	0.88
	5mC motor neurones	0.90
	5hmC glia	0.89
	5hmC motor neurones	0.90

Figure 2.3: Cohen's kappa coefficient equation and grading criteria (Cohen 1960) for inter-rater reliability assessment and case details for comparisons.
(A) The kappa coefficient calculation is used to assess the reliability of different raters, taking into account the risk of raters guessing correctly by chance. Po = relative observed agreement among raters, and Pe = hypothetical probability of chance agreement. **(B)** A number between 0 and 1 will be obtained, which indicates the reliability of the grading used, with 0 indicating no agreement, up to 1, indicating perfect agreement. **(C)** Inter-rater analysis was carried out for the cases and areas, summarised in table format. **(D)** Outcomes of the Cohen's kappa tests suggest all comparisons are rated as 'near perfect agreement'.

2.2.8 Cell counting

Quantitative analysis was conducted for the assessment of MNs and glia using the Hamamatsu NanoZoomer NDP.view 2 digital pathology software (Hamamatsu Photonics, Japan). MNs were counted in the AH, with glia being counted in the AH, LCT and DC of SC. Neurones and glia were also assessed in the fourth and fifth layer of motor cortex and AFCx. All immunohistochemical staining and cell counting was conducted blind to prevent unconscious bias.

All total cell counts were converted to percentages in order to account for variances in total cell numbers within each case to allow for direct comparisons.

Sequential sections stained for γH2AX, TDP43, 5mC and 5hmC and were matched, and the same MN was identified in adjacent tissue sections, as shown in figure 2.1. MN nuclei were graded to assess nuclear immunopositive and immunonegative staining, as described in figure 2.4.

Global immunopositive cell counts were taken for MNs and glia of γH2AX, TDP43, 5mC and 5hmC in SC, with cell counts for 5mC and 5hmC also taken for motor cortex and AFCx as indicated in figure 2.5.

Figure 2.4: Representation of image analysis grading for sequential staining.
(**A**) Nucleus displays immunopositive staining for 5mC / 5hmC / γH2AX but immunonegative staining for TDP43. (**B**) Nucleus displays immunonegative staining for 5mC / 5hmC / γH2AX but immunopositive staining for TDP43. (**C**) Nucleus displays immunopositive staining for both 5mC / 5hmC / γH2AX and TDP43. (**D**) Nucleus displays immunonegative staining for both 5mC / 5hmC / γH2AX and TDP43.

Figure 2.5: Diagram of human spinal cord cross-section, anterior frontal cortex and motor cortex.
(A) MN cell counting was conducted within the anterior horn area. Glia cell counting was conducted within the anterior horn, dorsal column and lateral corticospinal tract areas. Cell counting was also conducted in the pre-frontal cortex (indicated in yellow, (**B**)) and motor cortex (indicated in red, (**C**)).

2.2.9 Statistical analysis of cell counting data

Statistical analysis was carried out using either SPSS version 24 or 25. One-way ANOVA was used to test differences in global levels of methylation between the three groups (controls, sALS and C9ALS), as well as the TDP43 pathology levels between the three groups, as data was found to be normally distributed. One-way ANOVA was also used to assess any potential relationships between γH2AX/5mC/5hmC status in the MNs and TDP43 pathology. All one-way ANOVA analysis was followed by Tukey post-hoc tests to identify any significant differences in expression of γH2AX/5mC/5hmC and TDP43 between control, sALS and C9ALS cases. Sex differences were investigated using multiple *t*-tests in order to identify if any significant differences in expression were present. For all experiments, data was normally distributed, resulting in either ANOVA or multiple *t*-tests being appropriate statistical tests.

2.3 Preparation for microarray and next generation sequencing

2.3.1 Cases used in MethylationEPIC array

FFPE human PM SC was obtained from the SBTB. For this study, sections were used from three groups: controls, sALS and C9ALS, with n=6 for each cohort (total n=18). Cases were age and sex matched, as far as was possible. Details of cases used are shown in table 2.7.

Table 2.7: Summary of case details used for MethylationEPIC array			
Case ID	**Sex**	**Age**	**Disease Status**
085/2007	F	59	Control
035/1996	F	87	Control
039/1997	M	53	Control
005/2007	M	63	Control
012/2007	M	63	Control
087/1992	F	75	sALS
200/1997	M	73	sALS
034/2005	M	63	sALS
094/2006	M	72	sALS
099/2009	M	80	sALS
041/2004	M	64	C9ALS
053/1996	F	65	C9ALS
073/2007	F	65	C9ALS
098/2002	F	64	C9ALS
045/2006	M	46	C9ALS
039/2011	M	72	C9ALS
<u>Averages (including standard deviations)</u> Control = 65.0 (SD=13.0) sALS = 72.6 (SD=6.2) C9ALS = 62.7 (SD=8.7)			
<u>Age T-Tests</u> Control v sALS p= 0.27 Control v C9ALS p= 0.73 sALS v C9ALS p= 0.06			
<u>Sex splits:</u> Control = 3M, 2F sALS = 4M, 1F C9ALS = 3M, 3F Overall = 10M, 6F			

2.3.2 Toluidine Blue stain

1) 10 µm sections of SC were mounted onto uncharged slides.

2) Sections were dewaxed in two changes of xylene for 5 min each.

3) Sections were rehydrated through graded alcohols: (100%, 100%, 95% and 70%) for 5 min each.

4) Sections were rinsed in distilled water.

5) This was followed by transferring to 0.01% toluidine blue (TB) for 1 min.

6) Sections were rinsed in distilled water.

7) This was followed by dehydration through graded alcohols (70%, 95%, 100%, 100%) for 30 s each.

8) Sections were cleared in xylene for 10 min.

9) Finally, sections were dried for 1 hour in a fume hood.

2.3.3 Laser capture microdissection

LCM was used to extract LMNs from the AH. The principles of LCM are illustrated in figure 2.6. The Arcturus VERITAS™ laser capture microdissection system (Applied Biosystems, UK), in conjunction with Arcturus Capsure Macro LCM Caps (Life Technologies, UK), was used with the following settings: 30µm spot size, 70 mW laser power, 2500 µsec pulse, 1 hit, 200 mV laser intensity and 0 msec delay. MNs were first visualised using the TB stain, then picked and imaged, all at x20 magnification. Approximately 10,000 cells in total were collected, across six caps, per case.
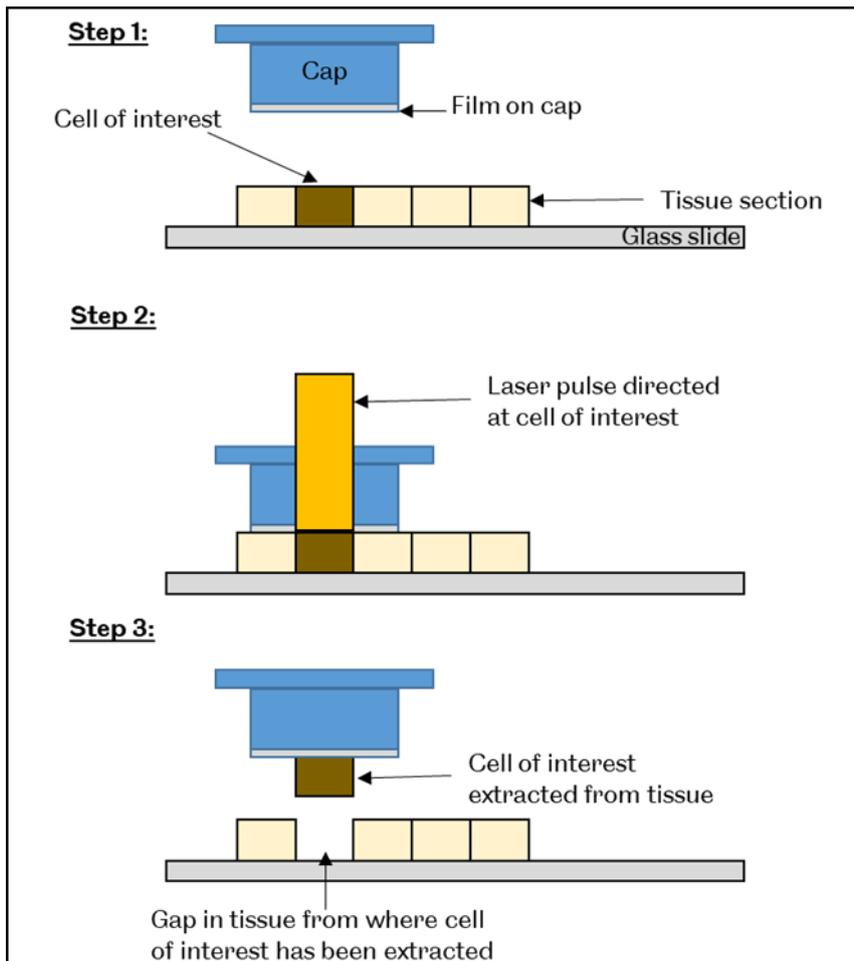
Figure 2.6: Laser Capture Microdissection workflow
Prior to LCM collection, tissue sections were stained with toluidine blue, enabling MNs to be visualised within the tissue. (**Step 1**) A cap with a thin film on the underside of the cap is placed over the stained tissue. (**Step 2**) MNs are visualised via a microscope through the cap. AI laser is then fired through the cap at the motor neurones. This melts the film onto the cells for collection. (**Step 3**) When all MNs have been identified, the cap is removed from the tissue. The MNs targeted are adhered to the film, and are therefore removed from the tissue when the cap is lifted off the tissue. The film can then be peeled off the cap and subjected to DNA extraction.

2.3.4 DNA extraction

Total DNA was extracted from the MNs collected on the LCM cap using the Zymo quick-DNA™ FFPE kit (Zymo research, Germany). As samples had already undergone deparaffinisation and rehydration during the TB staining process, these steps were not carried out. Instead, the protocol was begun from the tissue digestion stages:

1. LCM caps containing extracted MNs were placed in sterile 1.5 mL centrifuge tubes (Eppendorf, UK) along with 45µL nuclease free water, 45 µL 2× digestion buffer and 10 µL Proteinase K. The digestion buffer breaks down cell and nuclear walls, releasing the DNA. Proteinase K digests contaminating proteins and nucleases that contribute to DNA degradation.

2. Incubated at 55°C for 4 hours followed by 94°C for 20 minutes, to halt digestion.

3. Added 5 µL RNase A to each sample and incubated for 5 minutes at RT.

4. Added 350 µL genomic lysis to each buffer (a nucleic acid protector) to each sample.

5. Transferred mixture to a Zymo-Spin™ IIC column in a collection tube.

6. Centrifuged at 10,000 ×g for 1 min, discarding the flowthrough.

7. Added 400 µL wash buffer 1 (containing a protein denaturant), centrifuged at 10,000 xg for 1 min, discarding the flowthrough.

8. Added 700 µL wash buffer 2 (containing Tris hydrochloride to increase cell membrane permeability), centrifuged at 12,000 ×g for 1 min, discarding the flowthrough.

9. Added a further 200 µL wash buffer 2 and centrifuged at 12,000 ×g for 1 min, discarding the flowthrough.

10. Transfered spin column to new, sterile 1.5 mL tube and added 25 µL elution buffer to the spin column membrane. Incubated for 5 min at RT.

11. Eluted DNA by centrifuging at 16,000 ×g for 30 s.

12. Repeated steps 10-11 to give a total elution volume of 50 µL.

2.3.5 Bisulphite conversion

Bisulphite conversion allows unmethylated and methylated cytosines to be distinguished: Methylated cytosines remain unchanged, whereas unmethylated cytosines are converted to uracil (summarised in figure 2.7). Bisulphite conversion was carried out using the Zymo EZ DNA Methylation-Direct Kit (Zymo Research, Germany).

1) Approximately 250 ng of DNA was used per sample, and made up to a total volume of 20 µL with nuclease free water.

2) Added 130 µL of CT conversion agent to each sample and mixed via gentle pipetting. This agent contains sodium bisulphite which allows for the conversion of unmethylated cytosines to uracil.

3) Incubate at 95°C for 30 seconds to denature the DNA, followed by an incubation at 50°C for 1 hour. This cycle of 95°C/50°C was repeated for 16 cycles to allow for the sulphonation and hydrolytic deamination steps required for bisulphite conversion.

4) Added 600 µL of M-binding buffer (to aid DNA solubility) each spin column, placed in a collection tube.

5) Loaded samples into their corresponding spin columns and inverted to mix, followed by centrifugation at 10,000 ×g for 30 s. Discarded flow through.

6) Added 100 µL wash buffer and centrifuged at 10,000 ×g for 30 s, discarding flow through.

7) Added 200 µL desulphonation buffer (this allows the desulphonation process to occur, resulting in the conversion of uracil sulphonate to uracil) and incubated for 20 min at RT. Centrifuged at 10,000 ×g for 30 s, discarding flow through.

8) Added 200 µL wash buffer and centrifuged at 10,000 ×g for 30 s. Repeated this step.

9) Transferred spin column to new, sterile 1.5 mL tube and added 10 µL elution buffer to column membrane and incubated for 1 min at RT.

10) Centrifuged at 16,000 ×g for 30 s to elute. Repeated steps 9-10 to give a total elution volume of 20 µL.

Figure 2.7: Bisulphite Conversion Workflow.
Bisulphite conversion is carried out to enable methylated cytosine to be distinguished from unmethylated cytosine. During this process, methylated cytosine remains unchanged. Unmethylated cytosine is converted to uracil. (**Step 1**) Unmethylated cytosine is converted to cytosine sulphonate through a sulphonation reaction. (**Step 2**) Cytosine sulphonate is converted to uracil sulphonate through hydrolytic deamination. (**Step 3**) Cytosine sulphonate is converted to the final product uracil through a desulphonation reaction.

2.3.6 DNA quality and quantity checks

2.3.6.1 NanoDrop

A NanoDrop 1000 spectrophotometer (Thermoscientific, UK) was used as per manufacturer's instructions to assess DNA quantity and quality. DNA quantity was measured in ng/μL, and quality assessed by the 260/280 absorbance ratio. This ratio measures the absorbance of UV light at wavelengths 260nm and 280nm. The 260nm measurement refers to the amount of DNA in the sample, with the 280nm measurement referring to the amount of protein present in a sample. An absorbance ratio of 1.8 is viewed as an indicator of 'pure' DNA (ThermoScientific NanoDrop Spectrophotometer TO42 technical bulletin). NanoDrop measurements were taken after both the DNA extraction and bisulphite conversion.

2.3.6.2 Agilent HS DNA chip

To determine both the quality and quantity of DNA within the methyl library, an Agilent high sensitivity DNA chip was used. This allows for the sizing of fragments within the library to be

calculated, along with quantification for dsDNA samples ranging from 50 to 7,000 base pairs. Chip analysis was carried out on the Agilent 2100 Expert software as per manufacturer's instructions. All samples indicated a high enough quantity and suitable quality of DNA to be taken forward for DNA methylation analysis using the MethylationEPIC BeadChip.

1. Left DNA dye concentrate and gel matrix to equilibrate to RT for 30 min in the dark before vortexing the dye concentrate and transferring 15 µL of the concentrate to the gel matrix.
2. Vortexed the gel matrix for 10 s, then transferred to a spin filter and centrifuged at RT for 10 min at 2240 xg.
3. Added 9 µL gel-dye mix to the 'G' well surrounded by a circle on the chip.
4. Placed the chip was in the priming station with the plunger set at 1 mL. Pressed the plunger down and held for 60 s before releasing.
5. After waiting a further 5 s, pulled the plunger up to the 1 mL mark.
6. Added 9 µL of gel-dye mix to remaining wells marked 'G'.
7. Added 5 µL DNA marker to the ladder well and each of the 11 sample wells.
8. Added 1 µL DNA ladder to the ladder well, and 1 µL sample into the relevant sample wells.
9. Placed the chip into the vortexer for 60 s at 2400 rpm.
10. Transferred the chip to the 2100 bioanalyser, where analysis of the chip was then run.

2.4 Illumina Infinium MethylationEPIC BeadChip

The Illumina® Infinium® MethylationEPIC BeadChip is a method allowing for methylation profiling using bisulphite converted DNA and whole-genome amplification in conjunction with BeadChip technology to measure signal intensity in order to determine methylation status at specific CpG loci. Single base extension is carried out on DNA hybridised to the BeadChips, which incorporates labelled nucleotides. C and and G nulceotides are labelled with biotin, whilst A and T nucleotides are labelled with dinitrophenyl.

Methylation status at CpG sites can be determined using beta values, which are calculated using the equation detailed in figure 2.8. A summary of reagent preparation prior to conducting the protocols detailed from section 2.4.1 and beyond are provided in appendix IV. Details on reagents used in the MethylationEPIC protocol can be found in appendix VI, with a summary of control used in appendix VII.

$$\beta \text{ value} = \frac{\text{Intensity of methylated signal}}{(\text{Intensity of unmethylated signal} + \text{intensity of unmethylated signal} + 100)}$$

Figure 2.8: Equation for calculating degree of methylation (β value)
This calculates the degree of DNA methylation observed in a gene, promoter or CpG island.

2.4.1 DNA Amplification

1. Transferred 20 µL of bisulphite converted DNA to the corresponding well in a 96 well plate.
2. Added 20µL of MA1 and 4 µL 0.1 N NaOH into each well. MA1 minimises sample evaporation, while NaOH denatures and neutralises the samples prior to amplification.
3. Sealed the plate and vortexed at 1,600rpm for 1 min.
4. Centrifuged at 280 ×g for 1min.
5. Incubated for 10 min at RT to allow denaturation and neutralisation.
6. Added 68µL of random primer mix (RPM) to each well with 75µL MSM, a multi-sample amplification master mix.
7. Resealed plate and vortexed at 1,600 rpm for 1 min.
8. Centrifuged at 280 ×g for 1 min.
9. Transferred plate to Illumina hybridisation oven at 37°C for 24 hours. During this process, DNA is incubated and isothermally amplified in order to increase the amount of DNA available by several thousand folds.

2.4.2 DNA fragmentation

1. Removed samples from hybridisation oven and centrifuge at 280 ×g for 1 min.
2. Added 50µL FMS to each well. FMS contains potassium chloride (for fragmentation) and sodium chloride (to stabilise the DNA).
3. Resealed plate and vortexed at 1,600rpm for 1 min.
4. Centrifuged at 280 xg for 1 min.
5. Incubated plate at 37°C for 1 hour for fragmentation.

2.4.3 DNA precipitation and resuspension

1. Centrifuged the plate at 280 ×g for 1min. Removed the seal and added 100 µL PM1. PM1 contains ammonium acetate, which is a salt often used in samples with high dNTPs and oligosaccharide content, and aids precipitation of the fragmented DNA.
2. Resealed the plate, and vortexed at 1,600 rpm for 1min.
3. Incubated the plate at 37°C for 5 min, then centrifuged at 280 xg for 1 min.
4. Added 300 µL 100% 2-propanol and resealed the plate.
5. Inverted the plate ten times, and incubated at 4°C for 30 min, allowing isopropanol precipitation to occur.
6. Centrifuged at 4°C at 3,000 xg for 20 min to allow the precipitated DNA fragments to collect and pellet at the bottom of the wells.
7. Unsealed the plate and inverted onto paper towels to decant the supernatant. Firmly tappred the plate onto the towels for 1 min to remove any excess liquid.
8. Left the plate inverted on a tube rack for 1 hour at RT to dry the DNA pellet. A blue pellet containing the DNA could be observed at the bottom of each well.
9. Added 46 µL RA1 to resuspend. Resealed the plate by placing the plate onto the heat sealer and pressing until all wells create distinguishable indentations in the foil.
10. Transferred plate to the hybridisation oven at 48°C for 1 hour. RA1 contains several components, namely formamide, which is used as a protectant in tissue preservation, and sodium chloride, used for DNA stability. This incubation period allowed the pelleted DNA to be resuspended in solution prior to hybridisation.
11. Vortexed at 1,600 rpm for 1 min, then centrifuged at 260 ×g for 1 min.

2.4.4 DNA hybridisation to BeadChip

1. Placed plate on a heat block at 95°C for 20 min to denature the DNA prior to hybridisation.
2. Whilst the plate was incubating, the hybridisation chambers were assembled. Gaskets were placed into the chambers, and 400 µL PB2 was into the humidifying reservoirs. PB2 contains polypropylene glycol, which is a humectant, and keeps samples from drying out by attracting and retaining moisture from the air. Placed lids on the chambers immediately to avoid evaporation.
3. Left the plate to cool to RT for 30 min, then pulse centrifuged at 280 ×g for 1 min. Removed the seal.
4. Place BeadChips in the chamber inserts, orientated so that the barcode end of the BeadChip matches the barcode symbol on the chamber insert. Added 25 µL of sample into the corresponding well of the BeadChip.

5. Loaded the chamber inserts containing the BeadChips into the chamber, with the lid placed on top. Locked the Chambers and transferred them to the hybridisation oven and incubated overnight for 16 hours. During this incubation, the fragmented DNA annealed to locus-specific 50mers, which are attached to the BeadChip.

2.4.5 Washing, Extension and Staining of the BeadChip

2.4.5.1 Washing

1. BeadChips were removed from the hybridisation oven and cooled to RT for 30 min.
2. During this period, two wash dishes were filled with 200 mL PB1 each. Placed 150 mL PB1 into the multi-sample BeadChip alignment fixture in preparation for the wash steps.
3. Removed the BeadChips from the chambers. Peeled off and discard the seal on the BeadChip. Transferred the BeadChips to the first wash dish containing PB1 and moved up and down for 1 min. Repeated this process in the second wash dish containing PB1. During this process, any unhybridised or non-specifically hybridised DNA was washed away.
4. Placed BeadChips in the alignment fixtures required for the staining process. Placed spacers on top of BeadChips, followed by a glass back plate. Use of the spacer allows a small inlet to allow the staining procedure to occur. Clamped each BeadChip to secure.

2.4.5.2 Single base extension

Single base extension enables the incorporation of fluorescently labelled ddNTPs at the 3' CpG site. This allows the unmethylated to methylated cytosine conversions to be measured.

1. Placed BeadChip into the chamber rack holder, which was set to 44°C. Added 150 µL RA1, a wash solution, into the reservoirs of the BeadChips and incubated for 30 s. Repeated this step five times.
2. Added 450 µL of XC1, a buffer solution containing the salt sodium phosphate dibasic, to each reservoir and incubated for 10 min.
3. Added 450 µL XC2 to each reservoir, and incubated for 10 min. This solution contains glycerol and sucrose, which are both viscous and prepare the BeadChip for the addition of staining solution, and ensures that solutions applied coat all of the BeadChip.
4. Added 250 µL TEM to each reservoir and incubated for 15 min. TEM contains magnesium chloride, a cofactor needed for DNA polymerase to work correctly. This

incubation starts the extension process, with single base extension of the oligos on the BeadChip. The DNA hybridised to the BeadChip acts as the template, and detectable labels are incorporated into the extended DNA. This allows the methylation levels to be determined when the BeadChips are scanned.

5. Added 450 µL Formamide/1 mM EDTA to the reservoirs and incubated for 1 min to wash residual TEM off the BeadChips. Repeated this step to ensure all TEM was removed from the BeadChips.

6. Incubated the BeadChips for 5 min, then added 450 µL XC3 and incubate for 1 min. Repeated this step.

7. Set the chamber rack holder to 32°C before commencing BeadChip staining.

2.4.5.3 Staining

1. Added 250 µL STM, a two-colour master mix used to stain the BeadChip, to each reservoir and incubated for 10 min.

2. Washed with 450 µL XC3 for 1 min each, followed by a 5 min incubation.

3. Added 250 µL ATM, an anti-stain colour master mix, and incubated for 10 min, followed by two washes with 450 µL XC3 for 1 min each, then a 5 minute incubation.

4. Added 250 µL STM incubated for 10 min, followed by washing twice with 450 µL XC3 for 1 min, then incubating for 5 min.

5. Added 250 µL ATM and incubated for 10 min, again followed by washing twice with 450 µL XC3 for 1 min, followed by a 5 min incubation.

6. Added 250 µL STM for a last time, and incubated for 10 min. Conducted the final wash two steps in 450 µL XC3 for 1 min each, followed by a 5 min incubation.

7. Removed BeadChips from the chamber racks and placed on the lab bench.

2.4.5.4 Washing and coating

1. Removed BeadChips from the apparatus, and placed in 310 mL PB1 to wash. Moved BeadChips up and down in the PB1 ten times to remove all residual staining and washing residues. Left the BeadChips to soak in PB1 for 5 min.

2. Transferred BeadChips to 310 mL XC4, and moved up and down 10 times in the solution, followed by soaking for 5 min. XC4 is very viscous, and was used to coat the BeadChips prior to scanning.

3. Removed BeadChips from XC4 and placed on a tube rack to dry in a fume hood for 50 min.

2.4.5.5 Imaging the BeadChip

1. Once dry, BeadChips were placed in the Illumina® HiSeq® 2500 scanner.
2. Logged into the Illumina account, select 'access by BeadChip', and added BeadChip identification codes. Once the HiScan found the BeadChips, scanning occurred. During the scanning process, a laser is used to excite the fluorophores of the single-based extension products on the beads. The light emitted from the fluorophores is recorded by the scanner, with the proportion of DNA methylation being calculated by comparing the ratios of methylated cytosine signal to unmethylated signals. IDAT files were created, logging all this information.

2.4.6 Horvath Epigenetic Clock Analysis

The DNA methylation age of the three experimental groups (control, sALS and C9ALS) were analysed using the Horvath DNA Methylation Age Calculator (https://horvath.genetics.ucla.edu/html/dnamage). Beta values generated from the RnBeads data analysis was used as input for the calculator. Whilst the calculator is set up for the 27K and 450K arrays, it is still suitable for use in conjunction with the MethylationEPIC arrays, with a few changes made to the R code beforehand (see appendices; Horvath, personal communication).

2.5 Zymo Pico Methyl-Seq™ Library Formation for Next Generation Sequencing

2.5.1 Human spinal cord tissue

Table 2.8 details cases used in this experiment. FFPE human PM SC was obtained from the SBTB. For this study, sections from were used from three groups: controls, sALS and C9ALS, with n=4 for each cohort (total n=12). Cases were age and sex matched, as far as was possible.

Table 2.8: Case details for BS-NGS			
Case ID	**Sex**	**Age**	**Disease Status**
039/1997	M	53	Control
005/2007	M	63	Control
035/1996	F	87	Control
085/2007	F	59	Control
200/1997	M	73	sALS
087/1992	F	75	sALS
094/2006	M	72	sALS
034/2005	M	63	sALS
041/2004	M	64	C9ALS
039/2011	M	72	C9ALS
053/1996	F	65	C9ALS
098/2002	F	64	C9ALS
Averages (including standard deviations) Control = 65.5 (SD=14.9) sALS = 70.8 (SD=5.3) C9ALS = 66.3 (SD=3.9)			
Age T-Tests Control v sALS p= 0.53 Control v C9ALS p= 0.93 sALS v C9ALS p= 0.22			
Sex splits: Control = 2M, 2F sALS = 3M, 1F C9ALS = 2M ,1F Overall = 7M, 4F			

2.5.2 Bisulphite converted DNA repair

Bisulphite conversion of DNA results in degradation and loss of DNA. To counter this degradation, a repair reaction was conducted.

1. For each sample, a reaction with the following was set up in a 0.2 mL tube on ice: 5 µL bisulphite converted DNA, 2 µL nuclease free water, 10 µL DNA repair buffer and 2 µL DNA repair enzyme mix.
2. Samples were vortexed and microcentrifuged before placing them in the thermocycler with the programme run as detailed: 30 °C for 30 min to allow for double stranded synthesis and fragment end polishing, followed by 68 °C for 15 min to denature the DNA repair enzymes and end the reaction.

2.5.3 Random Primer Amplification

1. Samples were first primed and amplified. For each reaction: 2 µL PrepAmp Buffer (5x), 1 µL PrepAmp primer, and 7 µL bisulphite converted DNA. Mixed via pipetting.

2. Incubated on the following cycle: 98 °C for 2 min, 8 °C holding temperature, followed by temperature of steps at 16, 22, 28, 36 and 36.5 °C for 1 min each, with the cycle ending at 37 °C for 8 min.

3. During the 8 °C holding temperature step, added the following PrepAmp mixtures: 1 µL PrepAmp buffer (5x), 3.75 µL PrepAmp Pre mix and 0.3 µL PrepAmp Polymerase. Hold the reaction mix at 8 °C for 4 min, before continuing with the cycle described above.

4. Repeated this cycle, with only 0.3 µL PrepAmp Polymerase being added in the 8 °C hold phase, instead of the PrepAmp mix.

2.5.4 Purification

1. Mixed samples at a 4:1 ratio of DNA binding buffer to sample via pipette, then transferred to a spin column and centrifuged at 10,000 ×g for 30 s to bind the sample to the column.

2. Added 200 µL DNA wash buffer, then centrifuged at 10,000 ×g for 30 s. Repeated this step to ensure thorough washing of the sample.

3. Conducted a dry spin to remove all residual liquid.

4. Transferred the column to a 1.5 mL tube, and added 6.3 µL DNA elution buffer to the column membrane. Incubated for 1 min at RT to allow the elution buffer to bind to the DNA.

5. Eluted samples by centrifugation at 10,000 xg for 30 s. Repeated steps 4 and 5 to create a total of 12.6 µL eluate.

2.5.5 Adapter ligation and amplification

1. Added adapters to the newly amplified DNA samples: 12.5 µL LibraryAmp Master Mix (2x), 1 µL LibraryAmp primers and 11.5 µL sample after purification.

2. Incubated samples at 94 °C for 30 s, followed by ten cycles of: 94 °C for 30 s, 45 °C for 30 s, 55 °C for 30 s, 68 °C for 1 min. Then, incubated at 68 °C for 5 min.

2.5.6 Index Primer ligation and Amplification

1. Assigned samples a specific index (see table in appendix XII) to allow for identification when sequencing.

2. Added 12 µL sample to 12.5 µL LibraryAmp master mix (2×) and 0.5 µL assigned Index Primer, to give a total reaction volume of 25 µL.

3. Incubated at 94 °C, followed by conducting the following cycle ten times: 94 °C for 30 s, 58 °C for 30 s, and 68 °C for 1 min. Then, incubated at 68 °C for 5 min.
4. Conducted the Agilent high sensitivity DNA chip, as detailed in section 2.3.6.2.

2.5.7 Methyl Library Sequencing – Adapter Removal

After undergoing Agilent high sensitivity DNA chip analysis, adapter peaks that could potentially interfere with sequencing and data interpretation were removed with AmPure XT magnetic beads (Beckman Coulter, A63881).

1. Allowed beads to warm to RT for 5 min.
2. Mixed DNA sample with the beads at a 1:1 ratio, with thorough mixing via pipetting.
3. Transferred the sample/bead mixes to a U-bottomed 96 well plate and incubated at RT for 5 min before transferring to a magnetic stand for 2 min.
4. Discarded the supernatant and washed bead pellets in 200 µL 70% ethanol for 30 s.
5. Removed and discarded the ethanol.
6. Removed the plate from the magnetic stand and resuspended pellets in 40 µL nuclease free water. Pellets were disrupted by pipetting and thoroughly mixing with the water.
7. Incubated sample/bead mixes at RT for 2 min, before transferring to the magnetic stand.
8. Incubated for a further 1 min to allow beads to pellet.
9. Removed eluate and transferred to a 0.2 mL tube.

2.5.8 Illumina sequencing

Samples were then sequenced on the Illumina® HiSeq® 2500, using rapid mode. This mode was chosen due to availability of the scanner, and that rapid mode is most suited to short read applications, which is the case for this cohort.

2.5.9 BS-NGS Quality Control and Pre-Analysis

Quality control was performed on the Bisulphite NGS datasets using the 'FastQC' function in Galaxy (Andrews 2010, http://www.bioinformatics.babraham.ac.uk/projects/fastqc). This analysis performs a series of tests, including the total number of reads, GC content and sequence length see table 2.9 for more details. Whole genome bisulphite sequencing was conducted, with paired end reads, with a read length of 66bp. Samples were multiplexed for sequencing, followed by demultiplexing as part of the data analysis processing.

The tool 'Bismark' was used to map the sequencing reads from the BS-NGS data to the human genome hg38 (Krueger and Andrews, 2011) and determine the levels of CpG methylation.

Table 2.9: Summary of the quality control functions carried out on BS-NGS dataset using the FastQC function in Galaxy

FastQC function	Description of function
Per sequence quality scores	This is a series of statistical tests presented as box and whisker plots to determine the quality of the sequencing reads. The statistical analyses are carried out over the whole length of the read, with plots falling within the green areas determined as good quality, yellow areas indicating adequate quality, and the red areas indicating poor quality.
Plot sequence quality scores	This calculation looks at the total number of reads versus the average quality score over the whole read, with a good quality graph showing most of the distribution in the upper ranges of the plot.
Plot sequence quality scores	Over-represented sequences: to be classed as over-represented, a sequence has to account for ≥ 0.1% of total reads. Most of the sequences identified with this analysis are adapter reads and can therefore be discounted.

2.6 RnBeads analysis of MethylationEPIC and BS-NGS datasets

Bioconductor and the programming language R (version 3.5.0), using the RnBeads package in conjunction with a GUI vignette, known as the RnBeads Data Juggler, were used to analyse the BeadChip and BS-NGS datasets (Assenov *et al.*, 2014; Müller *et al.*, 2019). The RnBeads package allows for single base pair resolution analysis of DNA methylation datasets. In the case of the MethylationEPIC BeadChips, IDAT files containing the data collected from the scanned BeadChips was loaded into the Data Juggler. For the BS-NGS dataset, the BED files containing the sequencing data was loaded into the data juggler. A .csv file containing case details was also loaded into the data juggler (see appendix VII). Parameters were set in order to conduct analysis as detailed in table 2.10.

Table 2.10: RnBeads parameters used for analysis of MethylationEPIC and BS-NGS datasets.	
Parameter	**Description of parameter**
Analysis	The latest version of the human genome annotation (HGM) was chosen (hg38) (215291 [UID] 1341698 [GenBank] 1362718 [RefSeq]).
Quality control	This was enabled in order to assess the efficiency of the bisulphite conversion, staining, extension and BeadChip hybridisation. Negative control probe quality control analysis was also conducted.
Pre-processing	This was enabled, along with the Greedycut algorithm, which filters out any probes or samples with high fractions of unreliable measurements. The normalisation method 'beta mixture quantile dilation' (BMIQ) (Teschendorff *et al.*, 2013) was used.
Differential methylation	This was enabled to allow for the methylation status of genes, promoters and CpG islands to be ascertained and compared between the three groups (control, sALS and C9ALS) in order to identify differential methylation.

As part of the RnBeads analysis, beta-mixture quantile (BMIQ) normalisation was used. This is an intra-samples normalisation method that corrects the bias of type-II probe values found on the MethylationEPIC array (Liu and Siegmund, 2016). Type I probes take up double the amount of space on the BeadChip in comparison to type II probes, but are required to enable sufficient measurement of methylation at dense CpG regions. Type II probes have a lower dynamic range, but take up less space on the BeadChip. A combination of type I and type II probes on the BeadChip allows for a balance of high sensitivity with a large number of probes (Pidsley *et al.*, 2016). Probe filtering to remove cross-hybridising probes and the removal of SNP-enriched probes was also carried out as part of the RnBeads analysis pipeline. Quantile-Quantile plots (QQ plots) were analysed, indicating normal distribution of data generated using the MethylationEPIC array (appendix X). Average beta values for controls were 0.29, and 0.36 for ALS.

2.6.1 Significance sorting for MethylationEPIC and BS-NGS datasets

All analysis was sorted based on differentially methylated promoters, with a false-discovery rate adjusted significance cut-off at α= 0.05. This was then further sorted by identifying which promoters were classed as either hypomethylated or hypermethylated in ALS versus control. It was decided that the focus should be on analysis of promoters. Primarily, this was due to promoters being where gene transcription is initiated, with promoters controlling mRNA polymerase binding to DNA in order to transcribe DNA into mRNA, and eventually, a functional protein. As DNA methylation has been established to alter gene expression, it was decided that priority should be given to gene promoters. If more time was available, further analysis would have been conducted in genes and CpG islands.

2.6.2 Gene ontology analysis

The classification system 'Protein analysis through evolutionary relationships' (PANTHER) (Mi *et al.*, 2019) was used to conduct analysis on molecular function and biological processes on the promoters identified as differentially methylated in the ALS groups compared to controls in both the MethylationEPIC and the pre-existing mRNA expression dataset (Highley *et al.*, 2014). Prior gene ontology analysis was also conducted within the Highley *et al.*, 2014 paper using the database for annotation, visualisation and integrate discovery (DAVID) (https://david.ncifcrf.gov/home.jsp) (Huang *et al.*, 2009). Functions in RnBeads also allowed for gene ontology analysis, with this function being enabled.

2.6.3 mRNA expression data

Pre-existing mRNA expression data (Highley *et al.*, 2014) for control, sALS and C9ALS cohorts was used as a comparison for the MethylationEPIC dataset. Case details for the mRNA expression data are summarised in table 2.11.

Table 2.11: Summary of case details from Highley *et al.* 2014.			
Case ID	**Gender**	**Age**	**Disease Status**
056/1990	M	51	Control
080/1992	F	62	Control
114/1993	M	78	Control
129/1994	M	63	Control
039/1997	M	53	Control
005/2007	M	63	Control
088/1989	F	59	sALS
193/1990	M	70	sALS
088/1996	M	68	sALS
155/1996	F	58	C9ALS
045/2006	M	46	C9ALS
081/2009	M	60	C9ALS

The mean age of the control group was 61.7 (SD=9.6), 65.7 (SD=5.9) for sALS and 54.7 (SD=7.6) for C9ALS. The combined mean age of sALS and C9ALS cases was 60.2 (SD=8.5). *t*-tests were carried out determine if any significant difference in age could be observed between the groups. No significant difference was detected (control v sALS p=0.54, control v C9ALS p=0.31, control v ALS combined p=0.78, sALS v C9ALS p=0.12).

In this study, six control cases, three sALS cases and three C9ALS cases were used. LMNs were extracted from SC sections using LCM, followed by RNA extraction. These samples then underwent mRNA expression analysis using the Affymetrix GeneChip Human Exon 1.0 ST Arrays, with analysis carried out using the Partek Genomics Suite Analysis software. The dataset for this study is freely available on the GEO database (series record: GSE33855). 3195 genes were found to be differentially expressed in ALS cases versus controls. Of these, 1537 (48.1%) of cases were downregulated, with 1658 (51.9%) upregulated in ALS (p<0.05).

2.6.4 Comparison of MethylationEPIC to mRNA expression data

Venny 2.1.0 was also used to identify any overlap in promoters that were significantly differentially methylated in the MethylationEPIC dataset and differentially expressed genes (regardless of p-value) of the mRNA expression dataset. This was decided as the expression data is purely a validation of the MethylationEPIC in order to establish if direction of change matches, rather than the statistical significance of this. Of the cohort identified as significantly different for methylation, with corresponding changes in expression, a comparison was carried out to identify if the direction of change was the same. Promoters identified as hypermethylated were expected to see a decrease in expression, whilst promoters identified as hypomethylated were expected to show an increase in expression.

Chapter 3: Histological characterisation of DNA damage in ALS spinal cord

3.1 Introduction

DNA damage is posited as part of the pathogenesis of neurodegeneration, with one of these damages being in the form of double stranded breaks (DSBs) (Merlo *et al.*, 2016). These DSBs have the potential to be lethal to a cell. Once a DSB is detected, repair machinery, including H2AX, is recruited to the site of damage. H2AX then becomes phosphorylated to form γH2AX and recruits further repair machinery to the site. Therefore, γH2AX is seen as a suitable marker for DSBs.

Previous findings of studies into DNA damage in ALS have been contradictory, with one study finding an upregulation of γH2AX expression in lumbar motor neurones within the spinal cord of C9ALS cases compared to controls (Farg *et al.*, 2017). This is in contrast to more recent findings by Vazquez-Villaseñor *et al.*, 2019. In this study, IHC analysis was performed on neuronal and glial populations in motor and frontal cortices from sALS cases, with no difference in γH2AX expression found. Together, this suggests further experimental studies are required to further understand the role, if any, DNA damage plays in ALS.

The first aim of this thesis was to conduct a pathological investigation into the localisation and levels of DNA damage in motor neurones (MNs) and glia of the spinal cord (SC). Additionally, the DNA damage status of SC MNs displaying TDP43 pathology were investigated.

3.2 Hypotheses tested

1) No difference in DNA damage will be observed in ALS in MNs and glia of the SC when overall levels of damage are assessed, independently of TDP43 status. This hypothesis was tested by conducting IHC in conjunction with the DNA damage marker γH2AX, scoring MNs and glia for immunopositivity.

2) There will be higher levels of DNA damage in MNs displaying TDP43 pathology compared to MNs without TDP43 pathology. To test this hypothesis, whole slide images of adjacent sections stained for γH2AX and TDP43 were aligned such that the same MN could be identified in adjacent sections. MN nuclei were then graded as immunopositive or immunonegative for each marker.

3.3 Methods Overview

Detailed methods for this chapter can be found in section 2.2. In summary, sequential sections of FFPE human cervical spinal cord were immunohistochemically stained with γH2AX and TDP43 using standard ABC-HRP staining techniques. Sections were then scanned using the Hammamtsu NanoZoomer slide scanner (Hamamatsu Photonics, Japan). Case details can be seen in table 3.1. Sequential sections stained for γH2AX and TDP43 respectively, were matched up, and the same MN identified in each section. MN nuclei were then graded as immunopositive or immunonegative for both γH2AX and TDP43. Total cell counts for MN were conducted for both γH2AX and TDP43 in the AH in order to assess global levels of expression, as well as total cell counts for γH2AX glia in the AH, LCT and DC. Statistical analysis was carried out in the form of student t-tests and ANOVA, where appropriate. Values were considered to be significant at $p < 0.05$, unless otherwise stated.

Table 3.1: Case details for gamma H2AX MN and glial cell counts

NK = not known

Key:

Cases highlighted in black were used for MN studies only.

Cases highlighted in red were used for MN and glial studies.

Case ID	Sex	Age	Disease
144/1991	M	NK	Control
2108/1999	M	NK	Control
098/2007	M	67	Control
808/1990	M	NK	Control
309/1990	M	82	Control
019/1991	M	54	Control
041/2008	F	60	sALS
014/2011	M	51	sALS
072/2005	M	66	sALS
094/2006	M	71	sALS
005/2010	M	40	sALS
099/2009	M	79	sALS
023/2010	F	42	sALS
081/2009	M	59	C9ALS
039/2011	M	72	C9ALS
066/2008	F	59	C9ALS
273/1999	M	68	C9ALS
118/2001	M	66	C9ALS
098/2002	F	64	C9ALS
041/2004	M	64	C9ALS

3.4 Results

γH2AX is a marker of double stranded DNA (dsDNA) breaks. The immunostaining pattern observed in the anterior horn of the spinal cord reveals localisation of γH2AX immunostaining to the nucleus of motor neurones, where nuclear DNA is present. Some cytoplasmic γH2AX can be observed in motor neurones, potentially reflective of dsDNA breaks in mitochondrial DNA (figure 3.1). γH2AX immunostaining can also be found localised to the nucleus of glial cells.

Figure 3.1: Representation of the grading system used for slide analysis in TDP43 and γH2AX sequential immunohistochemistry.
(A) MN nucleus displaying γH2AX immunopositive staining, (B) MN nucleus displaying γH2AX immunonegative staining.

3.4.1 No difference in γH2AX levels in ALS spinal cord LMNs

Total MN cell counts (irrespective of TDP43 status) in the AH were conducted, assessing nuclei for γH2AX immunopositivity, a marker of DNA damage. No differences were observed between any groups (see figure 3.2): Control cases displayed immunopositive nuclear levels of 50.5%±26.3, sALS was 52.1%±19.9, with C9ALS at 73.2%±19.7. No significant difference was detected between control v sALS (p=0.93), control v C9ALS (p=0.21) or sALS v C9ALS (p=0.17), as determined by Student t-tests (SPSS v24). Post-hoc power analysis was conducted on the cohort (G*power v3.1.9.4, Faul et $al.$, 2007), with a power of 0.61 recorded. A figure of 1.0 represents an ideal cohort, suggesting this study was underpowered. Predicted power analysis determined a cohort of 55 samples would yield a power of 0.87.

Figure 3.2: No differences in the expression of γH2AX were observed in motor neurones of the anterior horn between groups (mean ±SD).
(p≥ 0.17 for all comparisons).
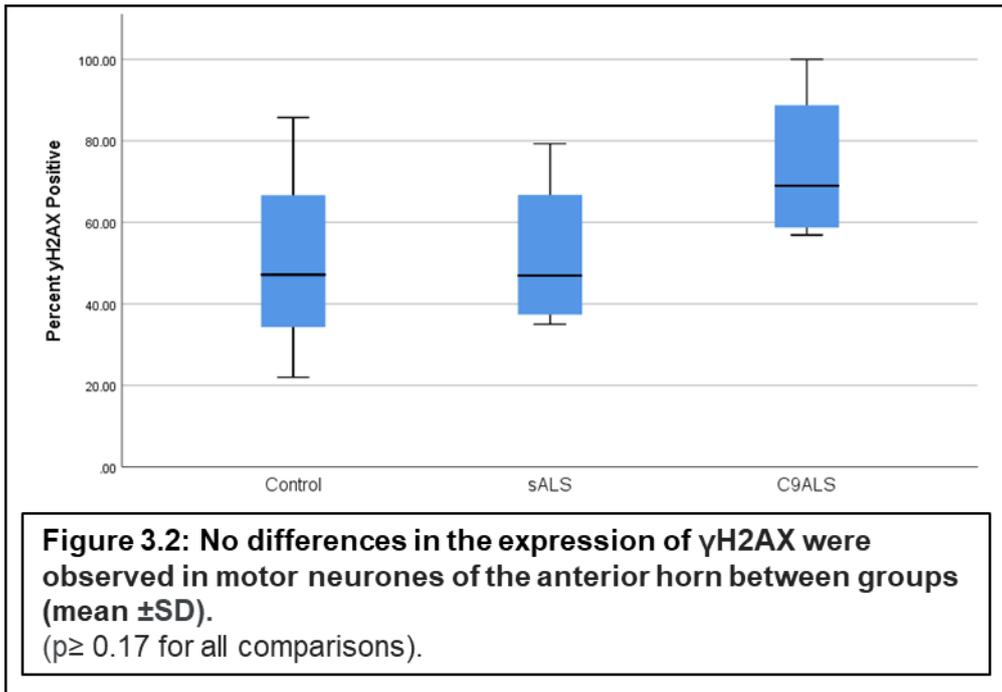
3.4.2 γH2AX levels in glia in spinal cord do not differ between controls and ALS

No difference was observed in the percentage of γH2AX-positive glia in any of the regions of SC assessed (p≥ 0.13 for all comparisons) (see table 3.2 and figure 3.3).

Figure 3.3: No intergroup differences in the expression of γH2AX between groups were detected in glia of the anterior horn, lateral corticospinal tract or dorsal column of the cervical spinal cord (mean ±SD). (p≥ 0.13 for all comparisons).
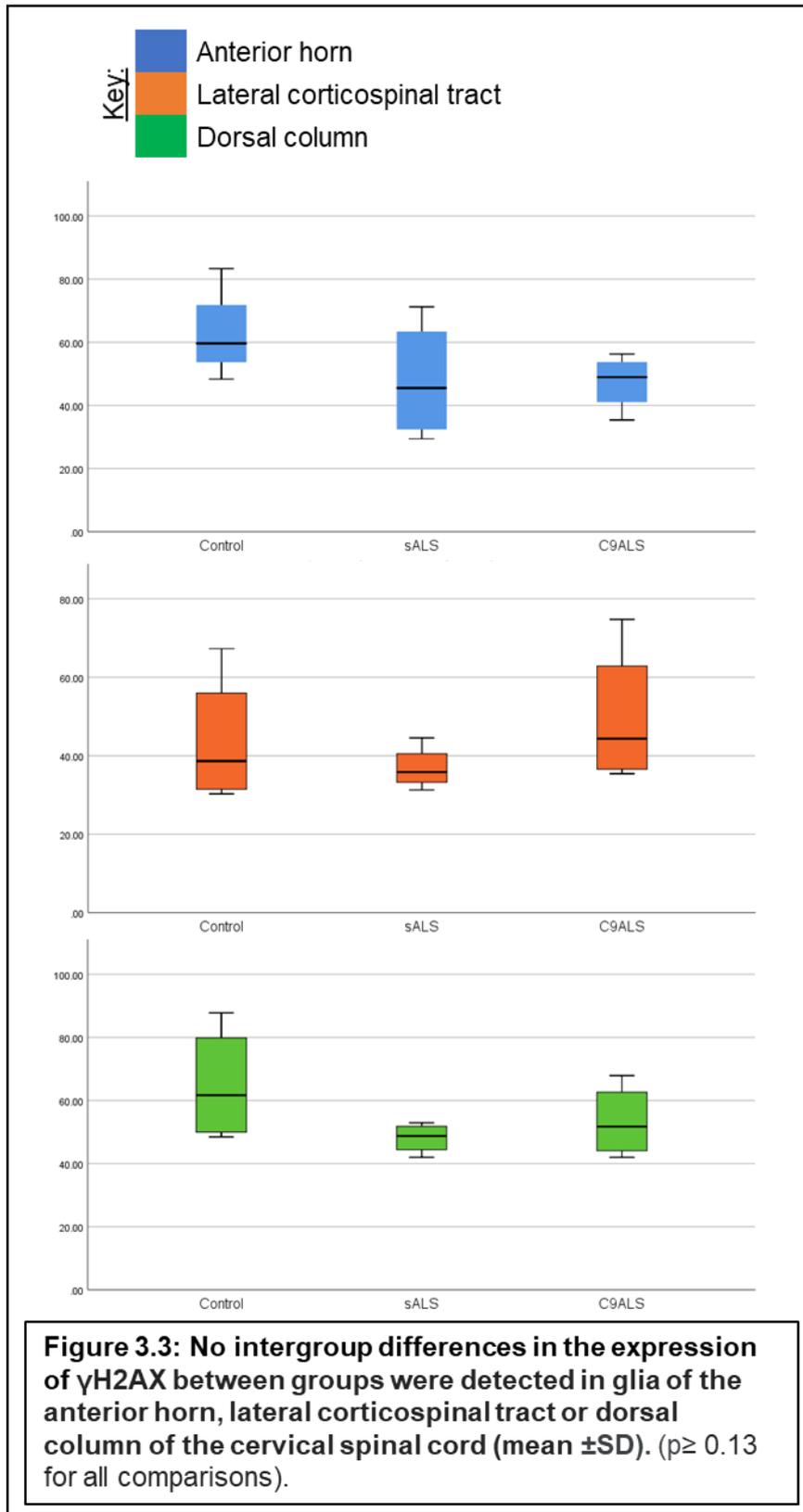
Table 3.2: Percentage immunopositive values (mean plus standard deviation) of glial cell counts for γH2AX in the anterior horn, lateral corticospinal tract and dorsal column in spinal cord, plus statistical analysis.

Target	Area	Average immunopositivity in percentage (plus SD)			P-values		
		Control	sALS	C9ALS	Control v sALS	Control v C9ALS	sALS v C9ALS
γH2AX	Anterior horn	62.7 (14.8)	47.9 (19.2)	47.4 (8.9)	0.27	0.13	0.96
	Lateral Corticospinal Tract	64.9 (18.5)	48.1 (4.8)	53.3 (11.7)	0.13	0.33	0.44
	Dorsal column	43.7 (16.9)	36.9 (5.8)	49.7 (18.0)	0.47	0.64	0.22

3.4.3 TDP43 pathology does not relate to DNA damage in ALS LMNs of the spinal cord

MNs identified as immunopositive for γH2AX were then assessed for TDP43 pathology to identify if there was any relationship between DNA damage and TDP43 pathology. No difference in γH2AX was observed between MNs positive or negative for TDP43 pathology (p=0.64) (figure 3.4). For MNs identified as having no TDP43 pathology (i.e. nuclei were positive for TDP43), γH2AX expression levels were 25.7%±18.4. For MNs displaying TDP43 pathology (nuclei were negative for TDP43), γH2AX levels were 28.7%±13.8.

Figure 3.4: No difference in percentage of γH2AX immunopositive motor neurone nuclei detected between immunopositive or immunonegative TDP43 motor neurone nuclei in ALS cases (mean ±SD).
No difference in γH2AX was observed between MNs positive or negative for TDP43 pathology (p=0.64).
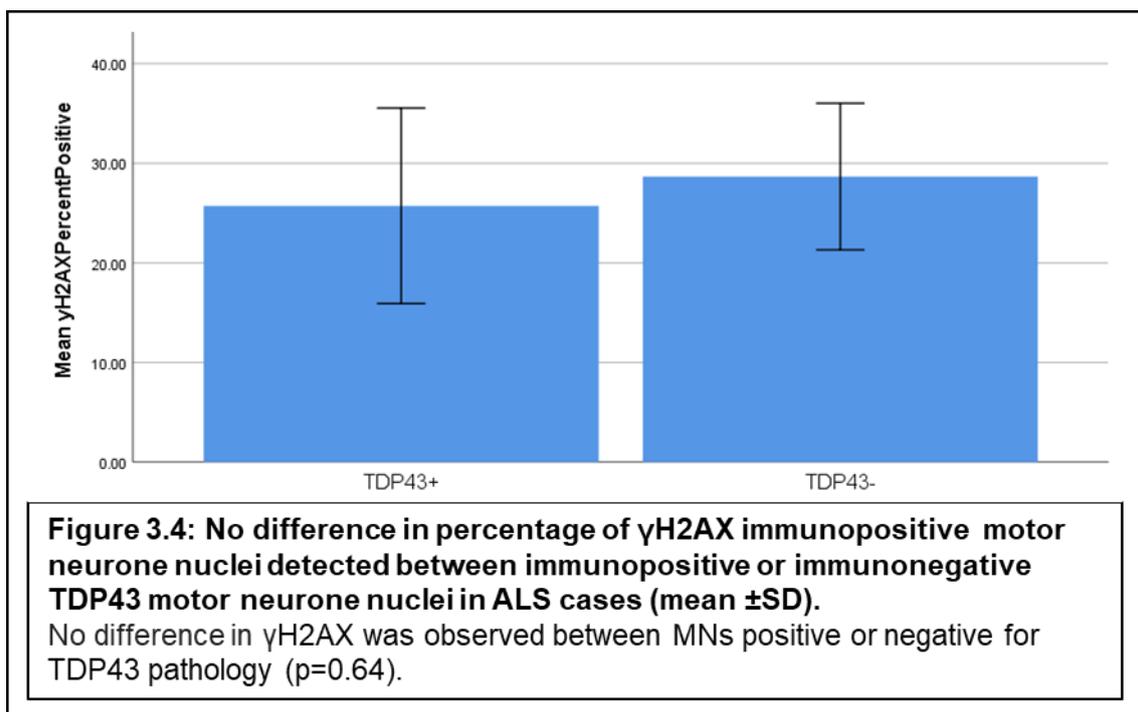
3.5 Discussion

In this chapter, a pathological characterisation of DNA damage in ALS LMNs and glia of the SC has been conducted. Using IHC, assessments of global levels of DNA damage were conducted using the marker γH2AX. The results suggest that this was not an issue for ALS LMNs and glia of the SC.

Walker *et al.*, 2017 analysed MNs of the AH in cervical SC in C9ALS cases. Compared to controls, higher levels of γH2AX immunopositivity were found in the C9ALS cohort. However, only 50 MNs were assessed per case (compared to 200 in the current study), with case numbers being low in this cohort (n=6 for each group). Post-hoc power analysis detailed in section 3.4.1 indicated this study was underpowered, which suggests that the samples size was not large enough in this case. Differing distributions of both age and sex of samples could also be a contributory factor in the non-concordance of the data presented in this chapter when compared to Walker *et al.* 2017, with the study presented in this thesis having a predominantly male cohort. No outliers were identified when analysing the data presented in figures 3.2, 3.3 and 3.4. When comparing this to the findings of Walker et al. 2017, the findings of this thesis did not replicate these findings. In the Walker paper, large variation is seen in the data. For example, one control has less than 5% γH2AX positive MNs, while another control sample displayed nearly 90% γH2AX MN positivity.

High variability between human samples is a frequent occurrence due to many factors. This includes post-mortem delay, but mainly variance seen naturally due to differing environmental factors. Batch variability can also be problematic. However, this wasn't the case for this study as all experiments were carried out in the same batch. One drawback of this study is the potential unknown impact of all the control samples being males, and the ages being unknown. This makes it difficult to assess the impact of this on the findings of the study. Ideally, replication of this experiment with a sex and age matched cohort is preferable. Due to tissue availability, this was not an option in this study. Age and sex statistical analysis on this cohort was carried out wherever possible to mitigate the issues identified with this study (see table 2.1 for statistics related to this). The advantages and disadvantages of using IHC to study DNA pathology are discussed at length in section 4.5.9.

A study of sALS cases which conducted IHC for both γH2AX and TDP43, determined that with increased nuclear TDP43 loss, an increase in γH2AX was also observed (Mitra *et al.*, 2019). However, this study did not assess TDP43 pathology and γH2AX status within the same MN. Rather, they conducted a field cell count and then compared the overall levels of both TDP43 pathology and γH2AX status. While this still useful, it cannot account for the specificity of determining what exactly is happening at a single cell level.

The results presented in this chapter are in agreement with findings by Vazquez-Villaseñor *et al.*, 2019. This study assessed the levels of γH2AX in neurones and glia of the MCx and frontal association cortex and found no changes in the levels of γH2AX when comparing ALS to controls. Further weight to this validation of the findings by Vazquez-Villaseñor *et al.*, could also be attributed to the use of the same antibody (R&D systems).

Results presented in this study suggest that no difference in DNA damage levels were observed in LMNs or glia of the SC. An increase in DNA damage is associated with increased age (Konopka and Atkin, 2018), with all groups indicating some level of DNA damage for both MNs and glia. However, these were not significantly higher in the ALS groups versus controls.

When comparing the results discussed in this chapter to others indicating increased levels of DNA damage in ALS, another factor to consider is the impact of using end-stage disease tissue. PM tissue is usually from individuals who are at the later stages of disease, and this was the case for the cohort used in this study. It could be that only the MNs and glia most resistant to DNA damage remained at the time of death, and the other MNs who have succumbed to ALS and died off were more susceptible to DNA damage.

Another limitation of using PM tissue is the availability of samples. For this study, low cohort numbers were used (n=6 for each cohort). Low cohort numbers are also a common feature in the other studies discussed above (Walker *et al.*, 2017; Mitra *et al.*, 2019). Ideally, larger cohorts would be used, as this would be more representative of the general ALS MN and glial populations. High variability is also a common issue when using PM human tissue, with humans naturally having variable levels due to genetic, epigenetic and environmental factors. This is reflected in the current study, with high variability being observed in all cohorts. This variability may mask genuine changes occurring within certain cell types, with associations only emerging when looking at larger cohorts. This was partially addressed in this study through the use of matching single MNs for both TDP43 pathology and DNA damage, as it focused in the individual cell, rather than focusing on a global phenomenon.

The most extensive and thorough study, and therefore arguably the most representative of the ALS neuronal population, was that conducted by Vazquez-Villaseñor *et al.*, 2019. This gives weight to the findings of this study, as although the study by Vazquez-Villaseñor *et al.* was not conducted in SC, as was the focus of characterisation conducted in this chapter, it did focus on tissues affected in ALS. MCx is highly affected in ALS, with frontal association cortex showing signs of neurodegeneration, which could contribute to cognitive decline in ALS. Overall, this chapter provides additional weight to existing studies that DNA damage levels in MNs and glia of the SC is not increased in ALS when compared to controls.

DNA damage has also been posited as a potential contributory factor in other neurodegenerative diseases, such as Alzheimer's disease. An immunohistochemical analysis in human post-mortem hippocampus and frontal cortex displayed an increase in γH2AX in both neurones and astrocytes in cases of mild cognitive impairment and Alzheimer's disease when compared to controls, which was replicated in a second cohort. However, as was the case in the study presented in this thesis and in the Walker *et al.* 2017 paper, the primary cohort used was small, with two male controls, three mild cognitive impairment cases (one male, two female) and eight Alzheimer's disease cases (seven male, one female). The secondary cohort in this study expanded the cohort size to eight controls (seven female, one male), seven mild cognitively impaired cases (three male, one female) and eight Alzheimer's cases (six male, two female) (Shanbhag *et al.*, 2019), with large variation seen in presented data. This study again highlights the issues of variability and small cohort sizes are also present in the wider field of neurodegenerative studies.

What this chapter does establish is the varied nature of using human post-mortem tissue and the limitations that come with this, such as batch variation, sample variability and the importance of age and sex matching samples. These factors could go some way in explaining the lack of consensus both is ALS and in the wider neurodegeneration field as to the role of DNA damage. Further studies are required to address these issues in order to establish the role, if any, DNA damage plays in neurodegeneration.

Chapter 4: Histological characterisation of DNA methylation and hydroxymethylation in ALS

4.1 Introduction

This thesis aimed to elucidate DNA pathology in ALS. After finding in the previous chapter that DNA damage is not significantly increased, other DNA pathologies were explored in the form of DNA methylation.

DNA methylation (5mC) is the addition of a methyl group to cytosine, usually resulting in gene expression repression. 5mC can be further oxidised to DNA hydroxymethylation (5hmC). Increased levels of both 5mC and 5hmC have been implicated in neurodegenerative disease, most notably in Alzheimer's disease. A few studies have been conducted looking into changes in DNA methylation and hydroxymethylation in ALS, with the general consensus being hypermethylation and hyperhydroxymethylation (Roubroeks *et al.*, 2017). Whilst many of these studies have been conducted in human tissues, these studies focused on homogenates, with the involvement of each cell type not being elucidated. This raises issues of cell-specific changes potentially being masked. Therefore, it is of value to elucidate if any cell-specific changes in DNA methylation and hydroxymethylation are present in ALS glia and neurones.

The first aim of this chapter was to conduct a pathological investigation into the localisation and levels of DNA methylation in neurones and glia of the SC, MCx and AFCx. Additionally, the methylation status of SC MNs within ALS cases with and without TDP43 pathology were investigated.

4.2 Hypotheses

1) Increased DNA methylation and hydroxymethylation will be observed in ALS in all areas (SC, MCx and AFCx). This will be more marked in C9ALS cases. This hypothesis was tested by scoring MNs and glia for immunopositivity for 5mC and 5hmC.

2) There will be higher levels of DNA methylation and hydroxymethylation in MNs displaying TDP43 pathology. To test this hypothesis, whole slide images of adjacent sections stained for 5mC and TDP43 or 5hmC and TDP43 were aligned such that the same MN could be identified in adjacent sections. MN nuclei were then graded as immunopositive or immunonegative for each marker.
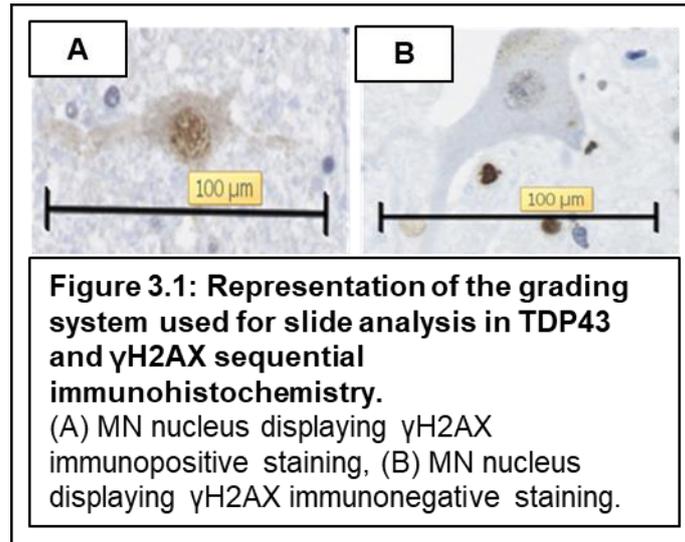
4.3 Methods Overview

Detailed methods for this chapter can be found in section 2.2. In summary, sequential sections of FFPE human cervical spinal cord were immunohistochemically stained with 5mC, 5hmC and TDP43 using standard ABC-HRP staining techniques. Sections were then scanned using the Hammamtsu NanoZoomer slide scanner (Hamamatsu Photonics, Japan). Sequential sections stained for 5mC/5hmC and TDP43 respectively, were matched up, and the same MN identified in each section. MN nuclei were then graded as immunopositive or immunonegative for both 5mC/5hmC and TDP43. Total cell counts for MN were conducted for both 5mC/5hmC and TDP43 in the AH in order to assess global levels of expression, as well as total cell counts for 5mC/5hmC glia in the AH, LCT and DC. To test if there were any sex differences in the expression of 5mC/5hmC in MNs and glia, male and female cases within each experimental group were statistically compared using one-way ANOVA in conjunction with Tukey post-hoc analysis. Values were considered to be significant at $p < 0.05$, unless otherwise stated. An overview of cases used in the studies relevant to this chapter can be seen in table 4.1.

Table 4.1: Case details for samples used in the DNA methylation and DNA hydroxymethylation immunohistochemistry studies

Key:
Black for one analyses per sample
Red for two analyses per sample
Green for three analyses per sample

Case ID	Sex	Age (years)	Disease Status	Spinal Cord	Motor Cortex	Frontal Cortex	Comments
012/2007	M	63	Control	Y	Y	Y	
080/1992	F	62	Control	Y			
085/2007	F	59	Control	Y	Y	Y	
098/2007	M	67	Control	Y	Y	Y	
023/1992	F	75	Control	Y			
035/1996	F	87	Control	Y			
071/1992	F	84	Control	Y			
293/1991	M	65	Control	Y			
039/1997	M	53	Control	Y			
005/2007	M	63	Control	Y	Y	Y	
224/2016	F	82	Control		Y	Y	
009/2018	M	69	Control			Y	
010/1996	F	65	Control		Y	Y	
007/2009	M	39	Control			Y	
070/2007	M	26	Control			Y	
033/2012	M	78	Control		Y	Y	
135/2018	F	85	Control		Y	Y	

Table 4.1: Case details for samples used in the DNA methylation and DNA hydroxymethylation immunohistochemistry studies continued…
Key:
Black for one analyses per sample
Red for two analyses per sample
Green for three analyses per sample

Case ID	Sex	Age (years)	Disease Status	Spinal Cord	Motor Cortex	Frontal Cortex	Comments
059/2009	F	80	sALS		Y	Y	Path
005/2010	M	40	sALS		Y	Y	Path
004/2006	F	64	sALS		Y		
102/2017	M	53	sALS		Y		
025/2013	F	72	sALS		Y		
091/2008	M	53	sALS		Y		
023/2010	F	42	sALS		Y		
141/2003	M	75	sALS	Y	Y		
055/2012	F	72	sALS	Y			
024/2008	F	63	sALS	Y	Y	Y	Path
027/2008	F	65	sALS	Y			
039/2005	F	61	sALS	Y	Y	Y	NoPath
087/1992	F	75	sALS	Y			
200/1997	M	73	sALS	Y	Y		
034/2005	M	63	sALS	Y	Y		
072/2005	M	66	sALS	Y	Y		
094/2006	M	72	sALS	Y	Y		
096/2008	F	69	sALS			Y	NoPath
212/1999	M	64	sALS			Y	NoPath
102/2005	M	64	sALS			Y	NoPath
049/2006	M	91	sALS			Y	NoPath
064/2009	M	67	sALS			Y	NoPath
074/2009	M	69	sALS		Y	Y	NoPath
094/2009	M	63	sALS		Y	Y	NoPath

Table 4.1: Case details for samples used in the DNA methylation and DNA hydroxymethylation immunohistochemistry studies continued…
Key:
Black for one analyses per sample
Red for two analyses per sample
Green for three analyses per sample

Case ID	Sex	Age (years)	Disease Status	Spinal Cord	Motor Cortex	Frontal Cortex	Comments
099/2009	M	79	sALS		Y	Y	NoPath
064/2010	M	78	sALS			Y	NoPath
037/2004	F	66	sALS			Y	NoPath
049/2005	F	64	sALS			Y	NoPath
041/2008	F	60	sALS		Y	Y	NoPath
075/2008	F	61	sALS			Y	NoPath
010/2009	F	60	sALS			Y	NoPath
041/2010	F	69	sALS			Y	NoPath
006/2002	F	64	sALS		Y	Y	NoPath
345/1990	M	68	sALS			Y	Path
110/1992	M	69	sALS			Y	Path
336/1990	M	63	sALS			Y	Path
026/1994	M	59	sALS		Y	Y	Path
096/2006	M	51	sALS			Y	Path
086/2008	M	66	sALS			Y	Path
014/2011	M	51	sALS			Y	Path
015/1998	F	86	sALS			Y	Path
050/2008	F	70	sALS			Y	Path
098/2008	F	75	sALS			Y	Path

Table 4.1: Case details for samples used in the DNA methylation and DNA hydroxymethylation immunohistochemistry studies continued…

Key:
Black for one analyses per sample
Red for two analyses per sample
Green for three analyses per sample

Case ID	Sex	Age (years)	Disease Status	Spinal Cord	Motor Cortex	Frontal Cortex	Comments
063/2008	F	69	C9ALS	Y	Y	Y	
041/2004	M	64	C9ALS	Y	Y		
083/2010	M	79	C9ALS	Y	Y	Y	
046/2010	F	67	C9ALS	Y	Y	Y	
053/1996	F	65	C9ALS	Y	Y		
073/2007	F	65	C9ALS	Y	Y	Y	
098/2002	F	64	C9ALS	Y		Y	
273/1999	M	68	C9ALS	Y	Y		
039/2011	M	72	C9ALS	Y		Y	
118/2001	M	66	C9ALS	Y		Y	
069/2009	M	48	C9ALS			Y	
081/2009	M	59	C9ALS			Y	
066/2008	F	59	C9ALS		Y	Y	
006/2006	F	75	C9ALS		Y	Y	
040/2011	F	67	C9ALS		Y	Y	
045/2006	M	46	C9ALS		Y		

4.4 Results

Within the anterior horn of the spinal cord, motor neurones displayed immunopositivity for 5mC and 5hmC, localised to the nucleus (figure 4.7). Immunopositivity within the nucleus of glial cells could also be observed in the anterior horn, lateral corticospinal tracts and dorsal column regions of the spinal cord. In the motor cortex and frontal cortex, nuclear immunostaining for 5mC and 5hmC is observed in neurones in grey matter, and glia of the grey and white matter (figures 4.8 and 4.9).

Within the anterior horn of the spinal cord, TDP43 immunopositivity can be seen localised to the nucleus of motor neurones in control cases (figure 4.12, panel a). Within sALS and C9LAS cases within the anterior horn of spinal cord, TDP43 is mainly localised to the nucleus. However, in motor neurones displaying TDP43 pathology, a loss of nuclear TDP43 and TDP43 cytoplasmic aggregates is observed (figure 4.12, panel b and c). Within the motor cortex, TDP43 was localised to the nucleus of neurones in controls (figure 13, panel a). for sALS and C9ALS cases, the majority of neurones displayed TDP43 immunopositivity in the nucleus. However, a subset of neurones displayed TDP43 pathology, with TDP43-positive aggregation in the cytoplasm and loss of nuclear TDP43 (figure 4.13, panel b, c and d).

4.4.1 Antibody optimisation

In order to establish the optimal conditions for IHC and subsequent analysis, a series of conditions were tested, as detailed below:

- *Tissue source*: FFPE
- *Antigen retrieval method*: Solutions at pH 6, pH 6.5 and pH 9 were tested, along with either pressure cooker or microwave methods
- *Primary antibody incubation conditions*: 1 hour/RT, overnight/RT and overnight/4 ºC
- *Primary antibody dilutions*: Ranges trialled from 1:50 to 1:64,000

Antibody optimisation was conducted for 5mC, 5hmC, DNMT1 and DNMT3a (figure 4.1 for 5mC, 4.2 for 5hmC, figure 4.3 for DNMT1 and 4.4 for DNMT3a).
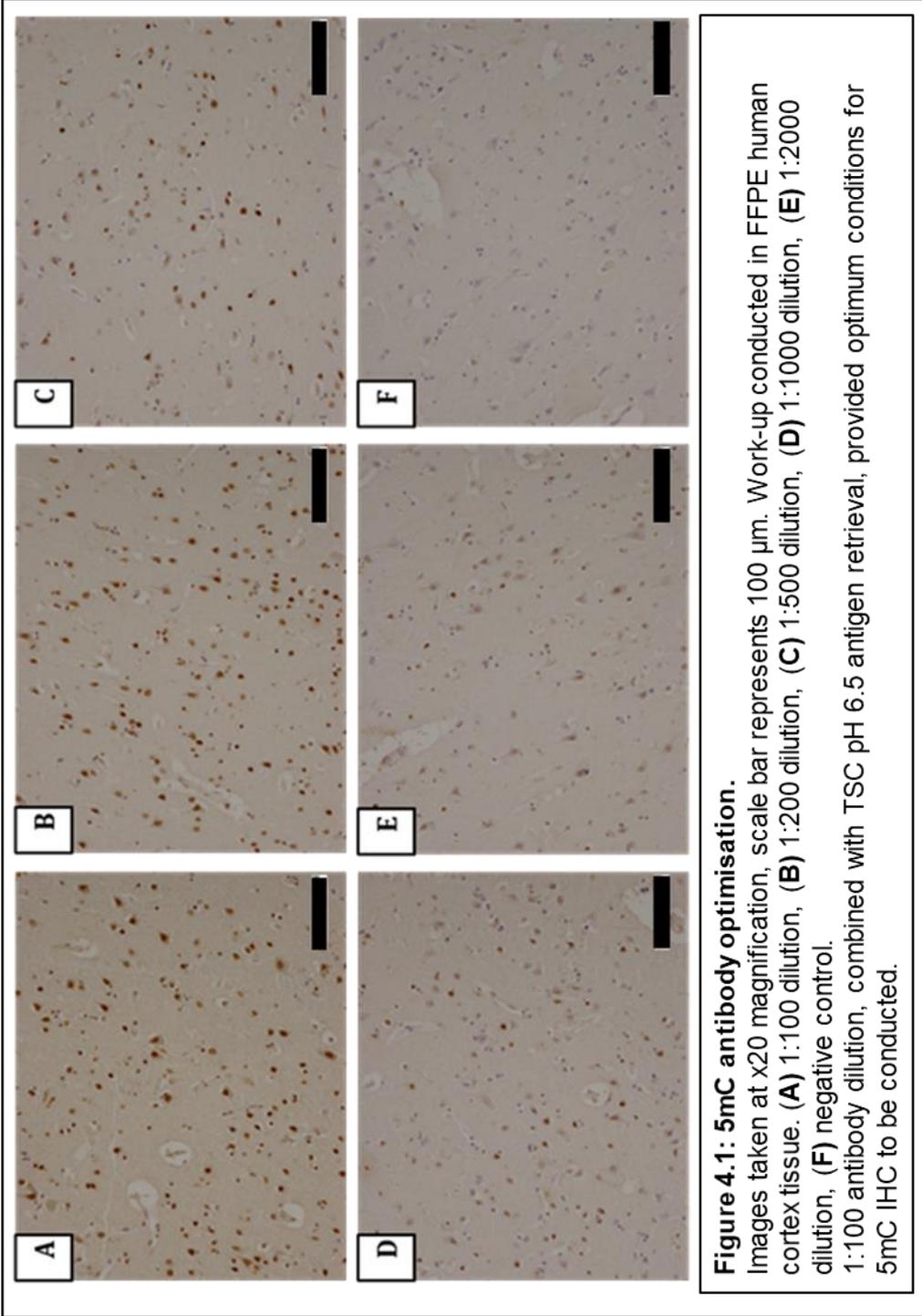
Figure 4.1: 5mC antibody optimisation.
Images taken at x20 magnification, scale bar represents 100 µm. Work-up conducted in FFPE human cortex tissue. (**A**) 1:100 dilution, (**B**) 1:200 dilution, (**C**) 1:500 dilution, (**D**) 1:1000 dilution, (**E**) 1:2000 dilution, (**F**) negative control.
1:100 antibody dilution, combined with TSC pH 6.5 antigen retrieval, provided optimum conditions for 5mC IHC to be conducted.
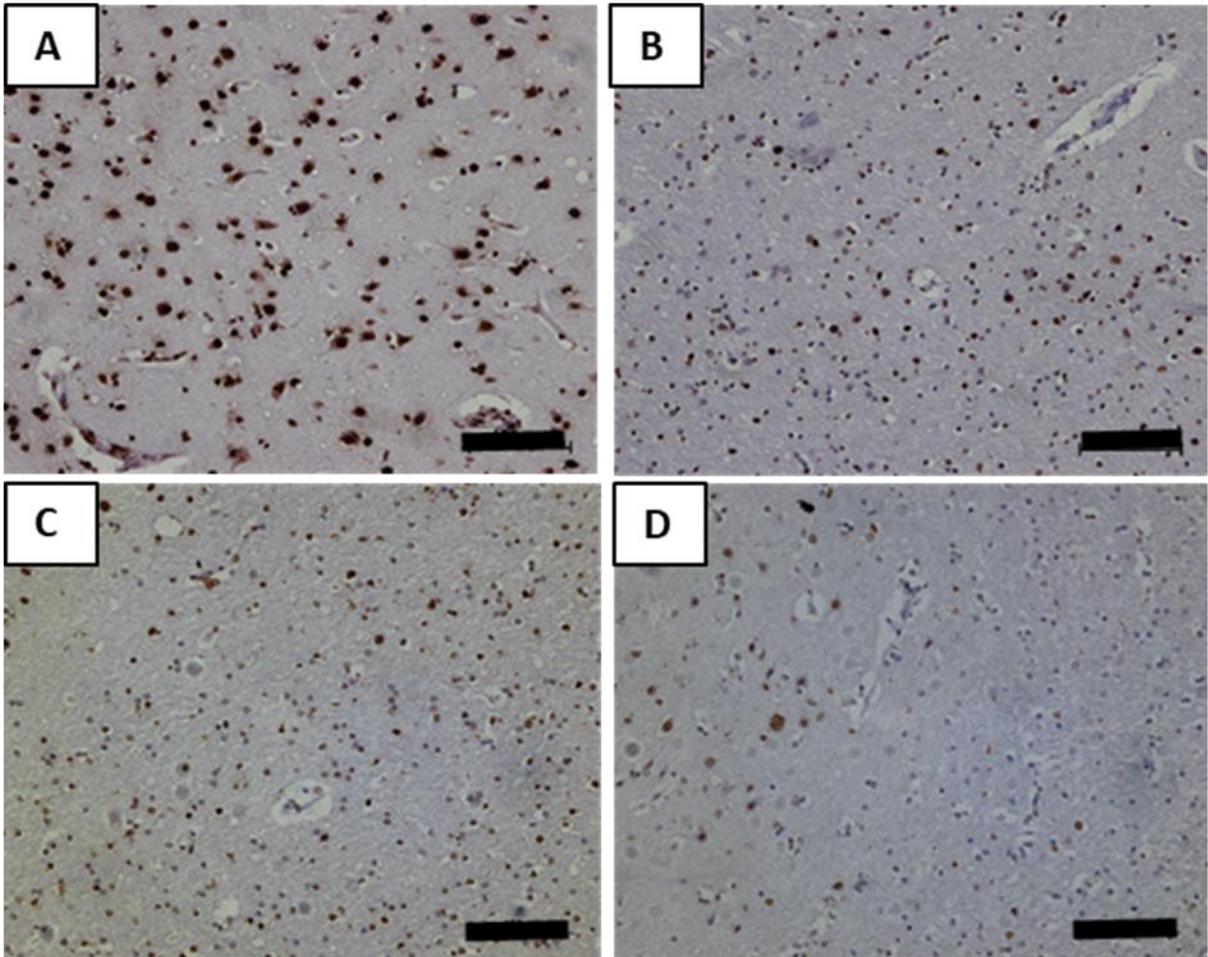
Figure 4.2: 5hmC antibody optimisation examples
Images taken at x20 magnification, scale bar represents 100 μm.

pH6/pressure cooker antigen retrieval at 1:500 dilution (**A**): A complete saturation of immunopositive staining can be observed in cells, with dense brown staining observed, with no variation in immunostaining intensity. Non-specific background staining can also be observed.

pH9/pressure cooker antigen retrieval at 1:1000 dilution (**B**): A reduction in non-specific background can be observed, with the haematoxylin counterstain being visible. However, saturation of immunopositive cells can still be observed, with very dark, dense staining. pH6/pressure cooker antigen retrieval was selected as the most appropriate, and further dilution optimisation was carried out.

A dilution of 1:16000 (**C**) was conducted, which again resulted in little non-specific background staining. However, the immunospecific staining observed was still at a very saturated level.

1:32,000 (**D**) dilution in conjunction with pH6/pressure cooker antigen retrieval was selected as the most appropriate, and was used to conduct immunohistochemistry on the cohorts. As can be observed in the example, haematoxylin blue counterstain can be observed, indicating no non-specific background staining. Further, variation in immunostaining intensity (visualised in brown using DAB) can be observed in cells. Immunonegative cells can also be easily visualised.

Antigen retrieval method

TSC pH6.5/microwave   pH6/pressure cooker   pH9/pressure cooker
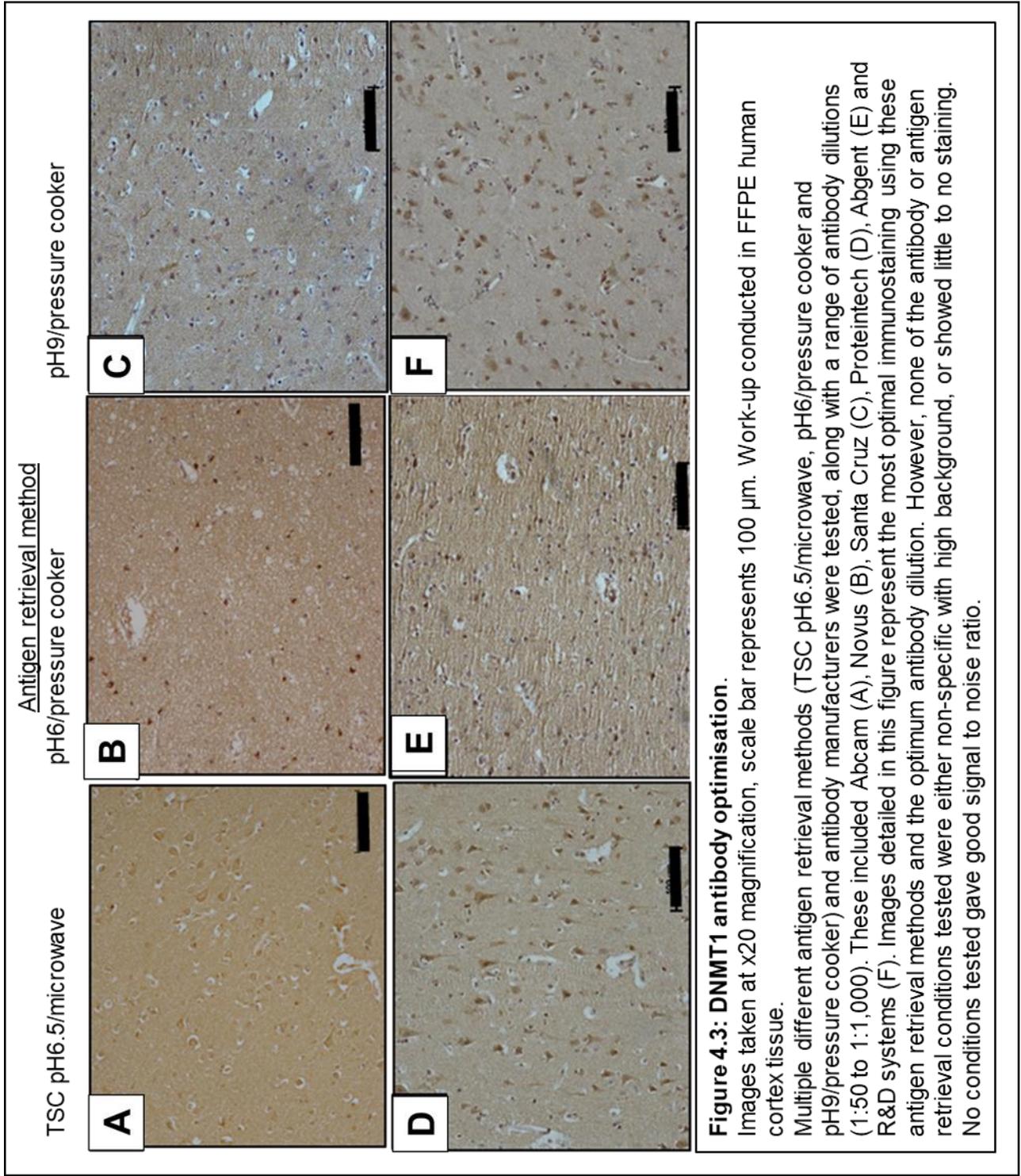
Figure 4.3: DNMT1 antibody optimisation.

Images taken at x20 magnification, scale bar represents 100 µm. Work-up conducted in FFPE human cortex tissue.

Multiple different antigen retrieval methods (TSC pH6.5/microwave, pH6/pressure cooker and pH9/pressure cooker) and antibody manufacturers were tested, along with a range of antibody dilutions (1:50 to 1:1,000). These included Abcam (A), Novus (B), Santa Cruz (C), Proteintech (D), Abgent (E) and R&D systems (F). Images detailed in this figure represent the most optimal immunostaining using these antigen retrieval methods and the optimum antibody dilution. However, none of the antibody or antigen retrieval conditions tested were either non-specific with high background, or showed little to no staining. No conditions tested gave good signal to noise ratio.
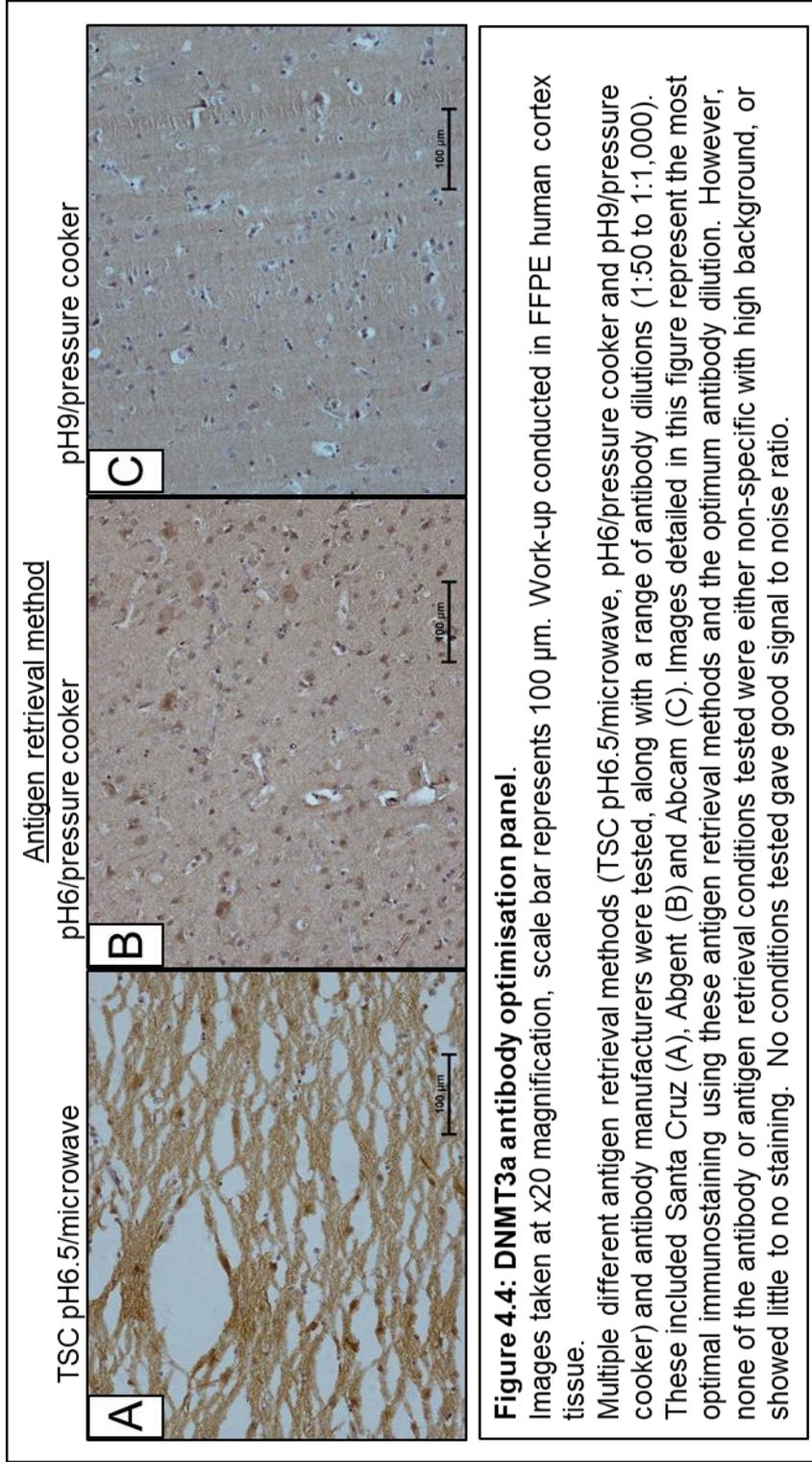
Figure 4.4: DNMT3a antibody optimisation panel.
Images taken at x20 magnification, scale bar represents 100 µm. Work-up conducted in FFPE human cortex tissue.

Multiple different antigen retrieval methods (TSC pH6.5/microwave, pH6/pressure cooker and pH9/pressure cooker) and antibody manufacturers were tested, along with a range of antibody dilutions (1:50 to 1:1,000). These included Santa Cruz (A), Abgent (B) and Abcam (C). Images detailed in this figure represent the most optimal immunostaining using these antigen retrieval methods and the optimum antibody dilution. However, none of the antibody or antigen retrieval conditions tested were either non-specific with high background, or showed little to no staining. No conditions tested gave good signal to noise ratio.

A number of factors were considered when establishing the optimum conditions for each antibody. This included assessment of levels of immunostaining, localisation of the immunostaining and levels of non-specific background staining. Optimal conditions were established for 5mC and 5hmC, with other conditions being discounted as they either have no immunolabelling of nuclei or non-specific immunolabelling.

Detailed below are the optimal conditions for 5mC and 5hmC antibodies:

- *5mC*: pH 6.5 tri-sodium citrate (TSC) / microwave antigen retrieval, 1:100 primary antibody dilution overnight at RT on FFPE tissue.
- *5hmC*: pH 6 pressure cooker antigen retrieval, 1:32,000 primary antibody dilution overnight at RT on FFPE tissue.

Antibody optimisation was unsuccessful for DNMT1 and DNMT3a, despite extensive trials. For all DNMT antibodies used, none showed nuclear-specific staining, with high background detected. Despite conducting a thorough range of antigen retrieval techniques, antibody dilutions and incubation times, antibodies for both DNMT1 and DNMT3a failed to provide clear immunoreactivity. Therefore, this part of the project was not carried forward for analysis in ALS v control cohorts.

4.4.2 Specificity testing of 5mC and 5hmC

5mC and 5hmC immunostaining was localised to the nucleus, with immunostaining of both neurones and glia displayed. Specificity testing was conducted to ensure that immunopositive staining was specific.

DNase treatment was carried out to denature all DNA. This should result in no DNA being present for the 5mC and 5hmC antibodies to bind to, resulting in no positive immunostaining. Separately, antibody pre-absorption was also carried out to completely saturate the antibody by binding it to DNA. This should result in no antibody receptors being left to bind to DNA when applied to the tissue. If any binding of the antibody to the tissue was found to result in immunopositive staining, this would suggest that the antibody also binds to other targets. No immunopositive staining was observed in the DNase or pre-absorbed experiments, suggesting that antisera bind specifically with minimal artefactual cross labelling. This is summarised in figure 4.5 for 5mC and figure 4.6 for 5hmC.

Representative examples of immunostaining can be observed for 5mC and 5hmC in the anterior horn of spinal cord for control, sALS and C9ALS cases in figure 4.7. Figure 4.8 provides representative images of 5mC in frontal and motor cortices in all experimental groups, with figure 4.9 providing examples of 5hmC immunostaining in frontal and motor cortices for all experimental groups.
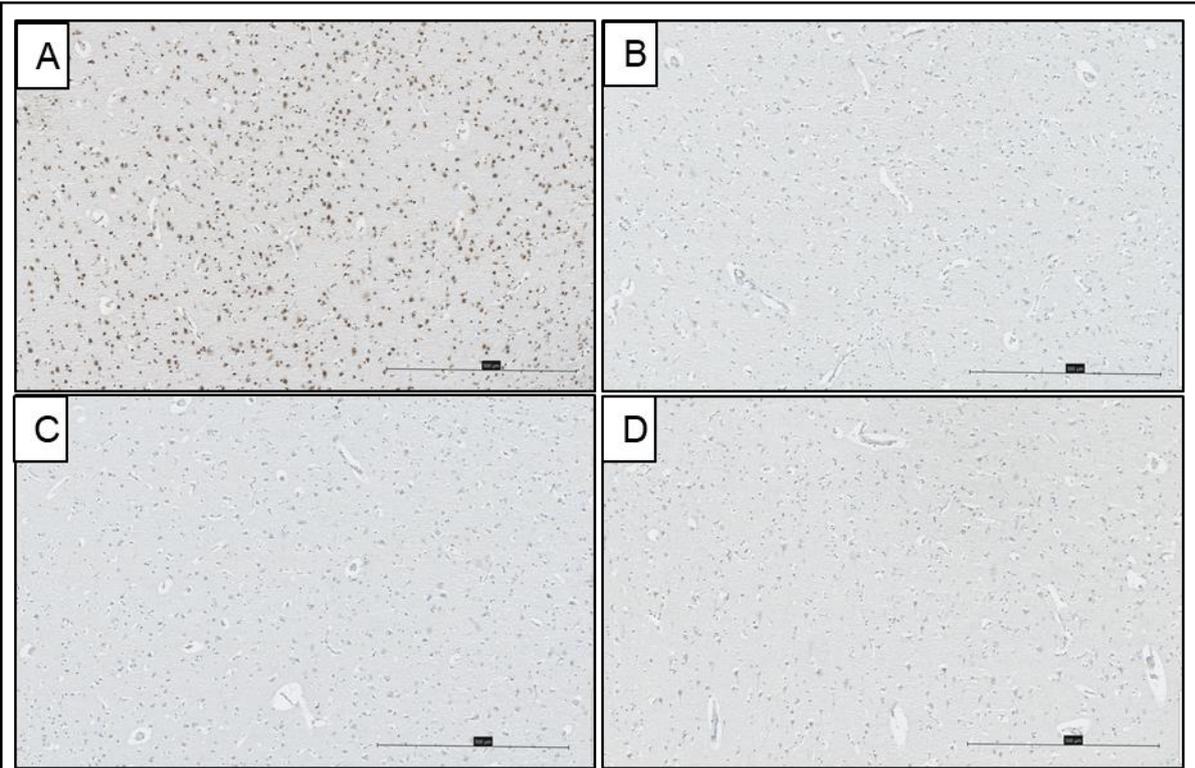
Figure 4.5: 5mC antibody specificity testing show no non-specific binding
Images taken at x20 magnification, scale bar represents 100 μm.
Immunopositive staining was observed for the positive control (A), with no
immunopositive staining detected for the negative control (B), the 5mC pre-absorbed
antibody treatment (C) or the DNase treated section (D).
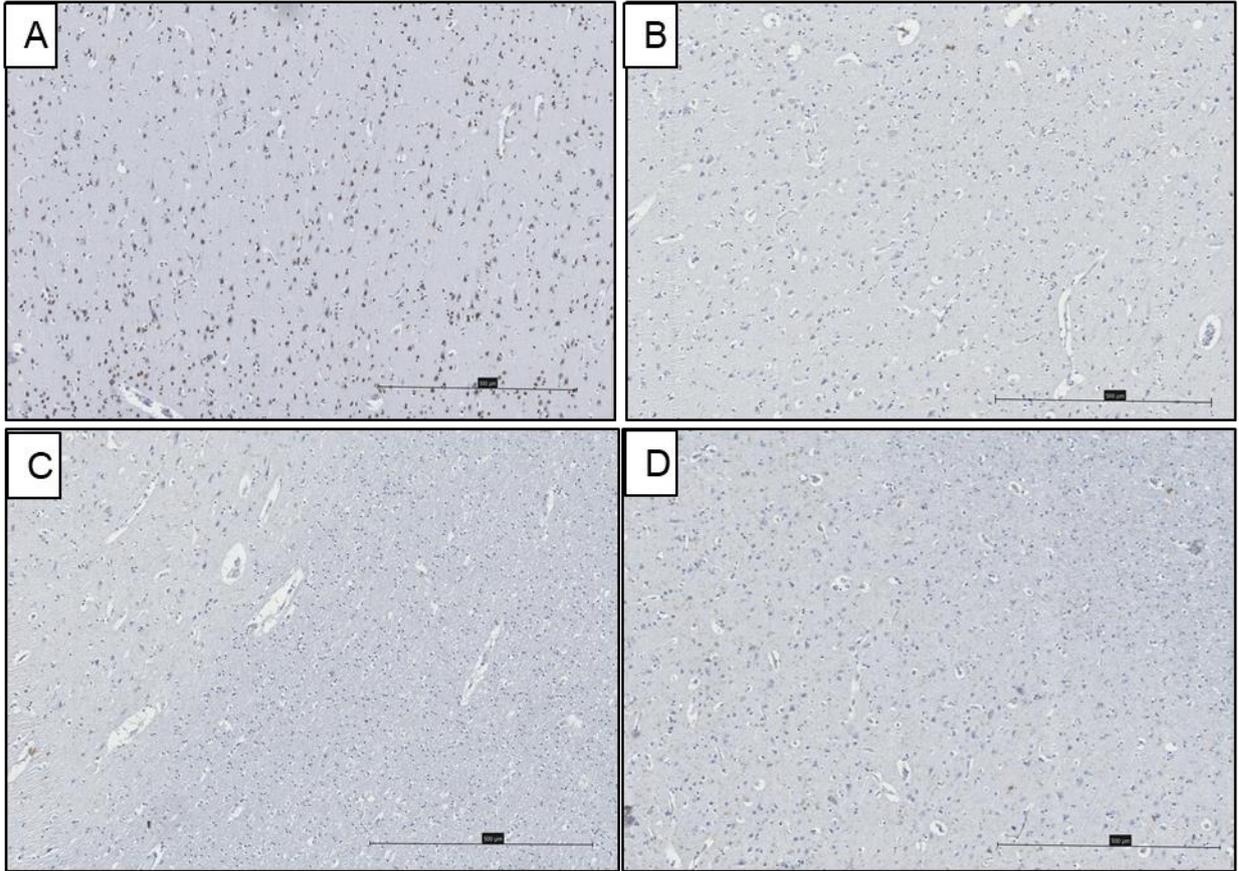
Figure 4.6: 5hmC antibody specificity testing show no non-specific binding
Images taken at x20 magnification, scale bar represents 100 μm.
Immunopositive staining was observed for the positive control (A), with no immunopositive staining detected for the negative control (B), the 5mhC pre-absorbed antibody treatment (C) or the DNase treated section (D).
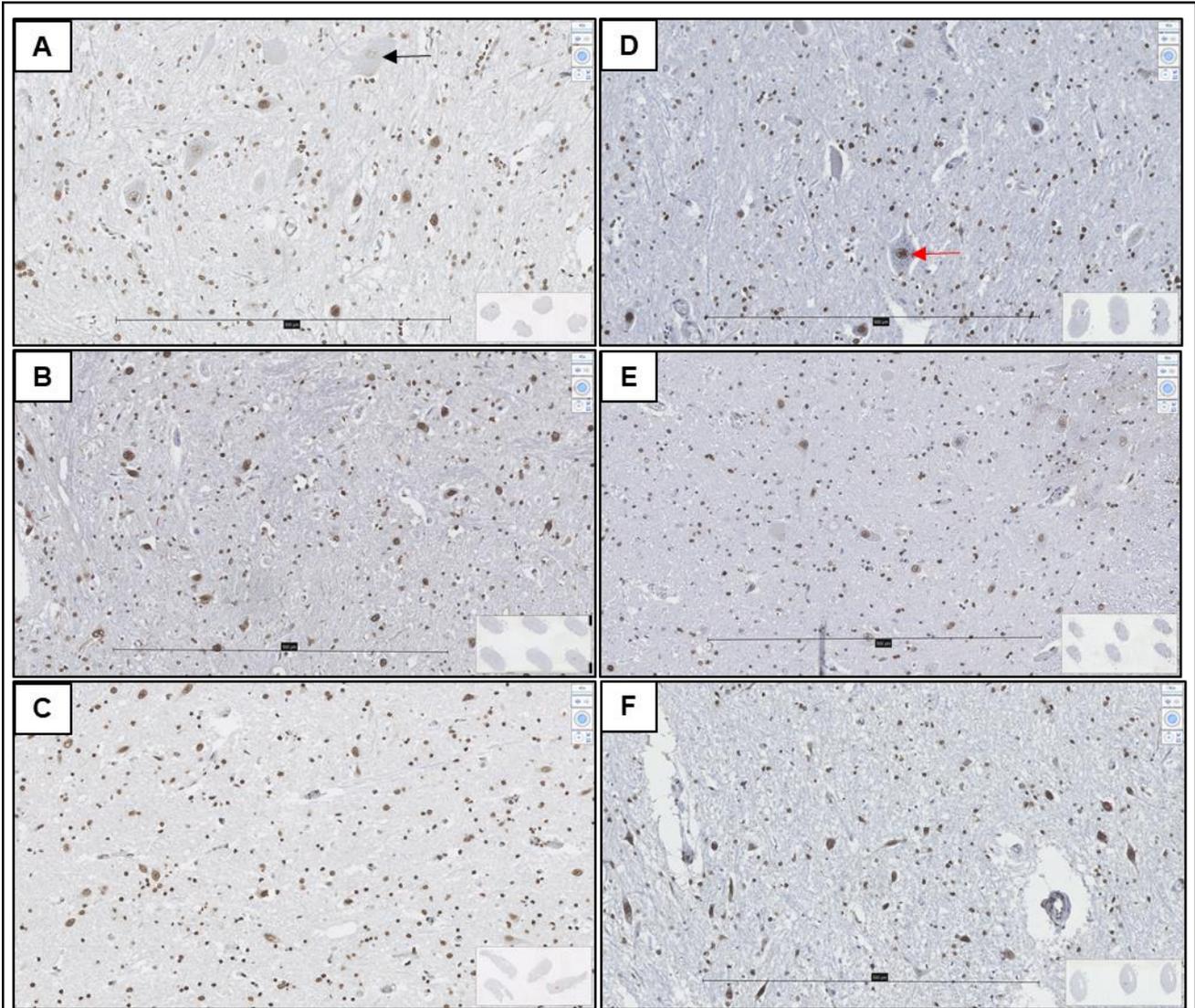
Figure 4.7: Representative images of 5mC and 5hmC immunohistochemical staining in the anterior horn of cervical spinal cord.
Images taken at x40 magnification, scale bar represents 500 µm.
5mC: (A) Control case (085/2007), with a motor neurone displaying nuclear immunonegativity (indicated by black arrow). (B) sALS case (035/2005). (C) C9ALS case (083/2010).
5hmC: (D) Control case (293/1991), with a motor neurone displaying nuclear immunopositivity (indicated by red arrow). (E) sALS case (141/2003). (F) C9ALS case (053/1996).

Figure 4.8: Representative images of 5mC immunohistochemical staining in motor cortex and frontal cortex. Images taken at x20 magnification, scale bar represents 500 μm. *Motor Cortex:* (A) Control case (085/2007), (B) sALS case (141/2003), (C) C9ALS case (046/2010). *Frontal Cortex:* (D) Control case (007/2009), (E) sALS case (005/2010), (F) C9ALS case (006/2006).
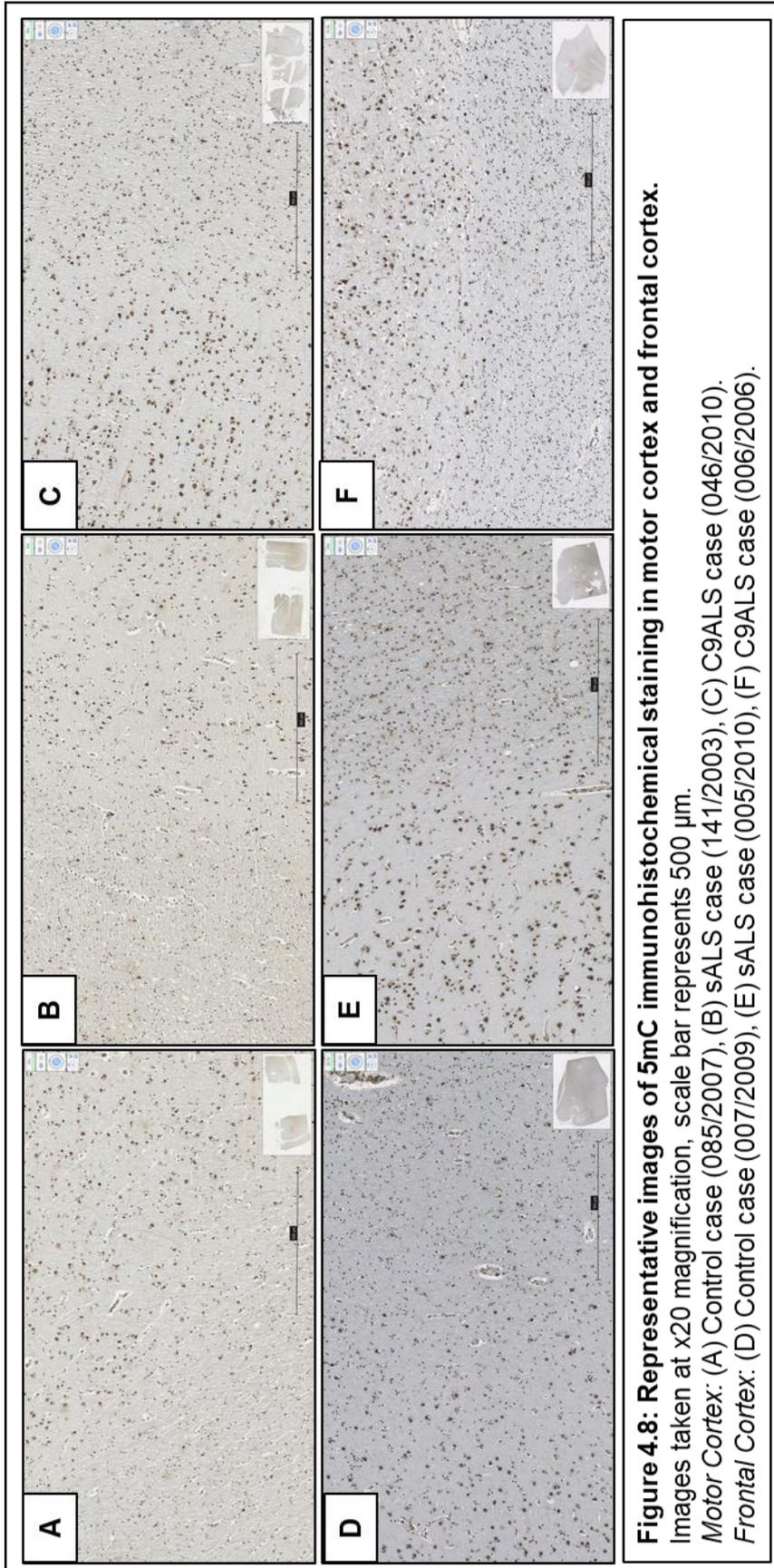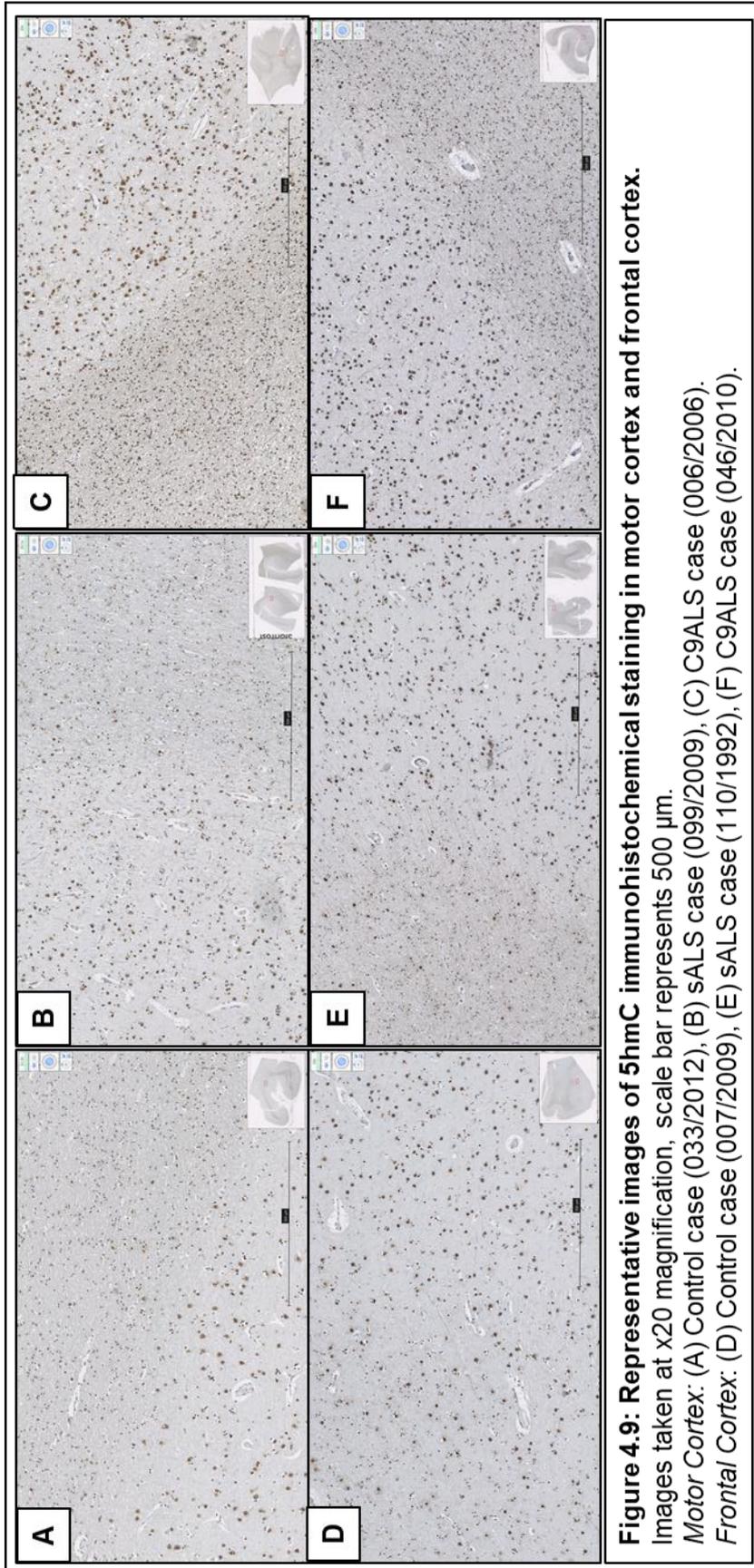
Figure 4.9: Representative images of 5hmC immunohistochemical staining in motor cortex and frontal cortex. Images taken at x20 magnification, scale bar represents 500 μm. *Motor Cortex:* (A) Control case (033/2012), (B) sALS case (099/2009), (C) C9ALS case (006/2006). *Frontal Cortex:* (D) Control case (007/2009), (E) sALS case (110/1992), (F) C9ALS case (046/2010).

4.4.3 Negative controls

Negative controls were used to determine the levels of background staining for IHC and assess if any cross-reactivity has taken place. Negative controls involved the omission of the primary antibody from the initial incubation step, along with the use of an isotype control at the same concentration of the same host species as the antibody. Isotype controls were used to determine the level of non-specific background that could be attributed to the use of primary antibodies. No background staining was observed in the negative controls for any of the antibodies used. Only nuclei counterstained with haematoxylin was present (see appendix III).

4.4.4 Methylation does not correlate with age

Analysis was conducted on the control cohort to determine if levels of DNA methylation in neurones correlates with age at time of death. This was conducted in SC, MCx and AFCx. Linear regression analysis was used, with no correlation detected for both 5mC and 5hmC DNA methylation markers (figure 4.10; for 5mC, R^2= 0.0226 for SC, 0.003 for MCx and 0.0221 for AFCx for 5mC; for 5hmC, R^2=0.0362 for SC, 0.0206 for MCx and 0.0021 for AFCx). While this was not an original aim or hypothesis, this calculation was conducted to add reassurance that aberrant age effects were not responsible for any changes in DNA methylation observed.

Figure 4.10: No correlation between degree of methylation and age in control cohort.
No correlation between methylation levels and age at death for controls were detected, as measured by regression analysis. Spinal cord 5mC (A), spinal cord 5hmC (B), motor cortex 5mC (C), motor cortex 5hmC (D), frontal cortex 5mC (E) and frontal cortex 5hmC (F) showed no associations between percentage positive of 5mC or 5hmC positivity and survival.

4.4.5 5mC and 5hmC levels in spinal cord lower motor neurones are higher in ALS compared to controls

First, a count of MNs that were positive and negative for both 5mC and 5hmC (irrespective of TDP43 status) was performed in the AH. These cell counts were converted into percentages to allow for direct comparison between cases that had differing numbers of MNs present. These 5mC and 5hmC data (visualised in figure 4.11) suggest that DNA methylation and hydroxymethylation levels are significantly higher in residual MNs of the AH in ALS when compared to controls ($p \leq 0.01$). No difference was observed between the two ALS groups ($p \geq 0.53$) (table 4.2).

Considerations for technical reasons as to why a significant difference in 5mC/5hmC expression was observed between control and ALS samples have also been considered.

These include tissue quality, number of motor neurones present in samples and specific motor neurone populations within each sample. Tissue quality can influence how well antibodies can bind to tissue, however in this case, statistical analysis of tissue differences indicated no differences of note between tissues. Another way these potential differences were mitigated was using tissue from the same brain bank. This results in all samples being processed via the same methods/protocols, limiting the effects of different handling and tissue processing systems on the findings of this chapter.

As ALS is characterised by motor neurone degeneration and death, it is feasible to assume that some control cases may have more motor neurones present that ALS cases, resulting in more opportunity to gather large numbers of motor neurones from control samples. This was controlled for in this instance by assessing multiple cross sections of tissue from each case, and limiting the number of motor neurones counted to approximately 200 motor neurones per case. Another consideration is the motor neurone population present within samples. It is possible that specific motor neurones present in control samples may not be present in ALS samples. It is feasible that within the ALS cases, motor neurones which are particularly susceptible have perished, and only motor neurones that are more resistant have remained, which could result in differences between control and ALS cases. This is discussed in section 4.5.3. Even with the above considerations, as the effect is seen in both 5mC and 5hmC, this acts as a semi-validation, and gives more weight to the findings of this chapter being genuine.

Figure 4.11: Differences in global levels of 5mC and 5hmC in motor neurones of the anterior horn are present between control and disease (mean ±SD).
n= 10 for each group, total n=30.
A)Significantly higher 5mC is observed in sALS and C9ALS cases than controls (p=0.0001 and 0.0002 respectively). No difference was observed between the two diseased groups (p=0.48).
B)Significantly higher 5hmC is observed in sALS and C9ALS cases than controls (p=0.0011 and 0.00072 respectively). No difference was observed between the two diseased groups (p=0.88).

Table 4.2: Statistical analysis of MN counts for 5mC and 5hmC in the anterior horn of spinal cord.

Target	Average immunopositivity in percentage (plus SD)			Outcome of *t*-tests, p=		
	Control	sALS	C9ALS	Control v sALS	Control v C9ALS	sALS v C9ALS
5mC	79.5 (6.1)	89.5 (3.3)	90.9 (3.1)	0.01	0.003	0.53
5hmC	81.6 (5.7)	90.4 (3.8)	90.7 (4.7)	0.001	0.0007	0.88

4.4.6 Relationship between TDP43 pathology and 5mC and 5hmC status

Having established higher levels of methylation in the LMNs of both sALS and C9ALS cases compared to controls, it was next decided to investigate whether there was any effect of TDP43 pathology on DNA methylation at the individual cell level. This was to establish if any relationship between TDP43 pathology and DNA methylation status is present in ALS MNs within the SC. Examples of immunopositive and immunonegative MNs are seen in figure 4.12, figure 4.13 and figure 4.14. Adjacent sections immunostained for 5mC and TDP43 and adjacent sections immunostained for 5hmC and TDP43 were digitised and then aligned such that the same MN nuclei was visualised in multiple sections, and assessed for both DNA methylation markers and TDP43 pathology, as seen in figure 4.15 for 5mC and figure 4.16 for 5hmC. MNs identified as immunopositive for the relevant methylation marker (5mC or 5hmC) were then assessed for TDP43 pathology. High levels (mean±SD) of 5mC and 5hmC were present in MN nuclei expressing TDP43 (72.7%±15.5 for 5mC; 86.9%±6.0 for 5hmC). However, a significantly lower percentage of MN nuclei displayed immunopositivity for 5mC and 5hmC in neurones with TDP43 pathology where nuclear TDP43 was absent (27.7±37.9, p=0.00002 for 5mC, and 50.7%±=16.6, p<0.00001 for 5hmC) (figure 4.17).
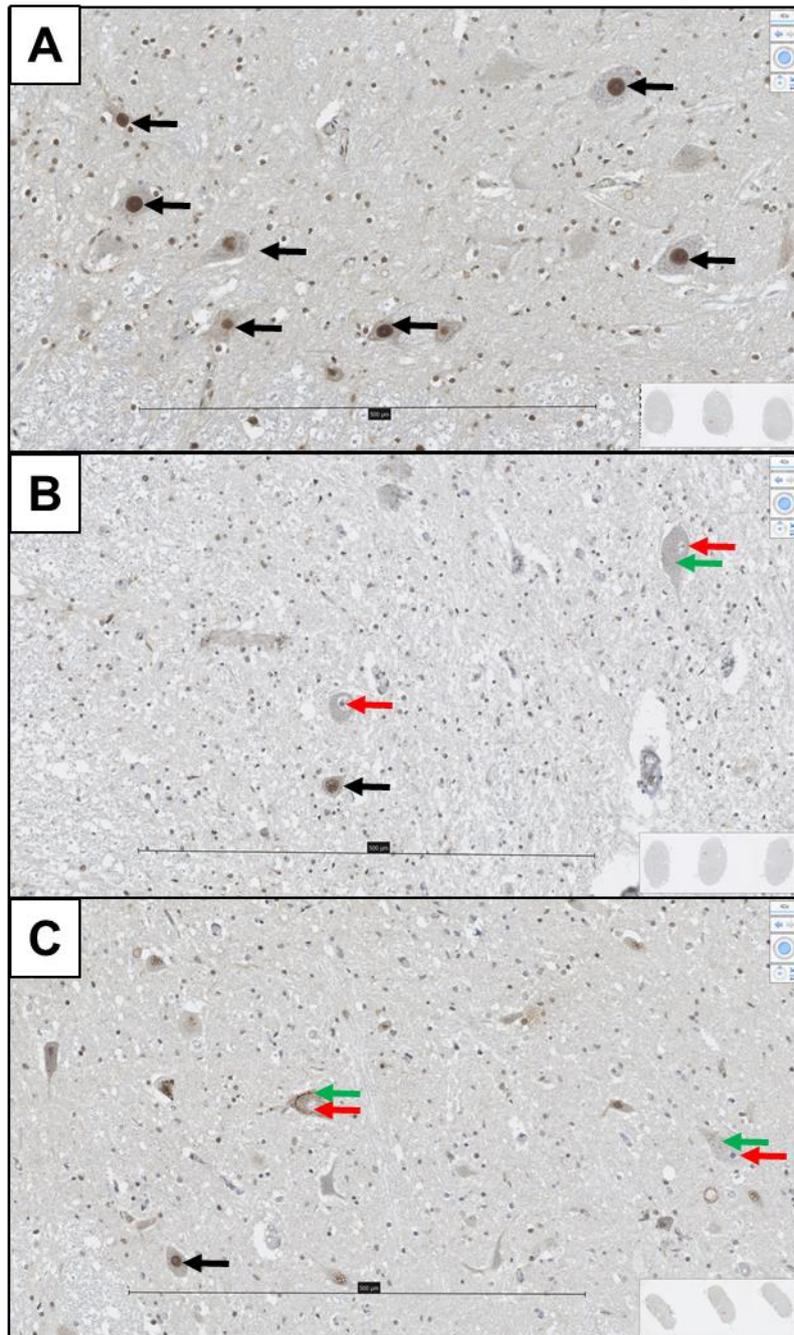
Figure 4.12: Representative immunohistochemical staining of TDP43 pathology in motor neurones of the anterior horn.

Images taken at x40 magnification, scale bar represents 500 μm.
(A) Control case (039/1997). Immunohistochemical staining displays TDP43 located in the nucleus, as indicated by the black arrows. No Cytoplasmic immunostaining is seen, indicating no TDP43 pathology is observed in motor neurones. (B) sALS case (200/1997). TDP43 is located in the nucleus for some motor neurones, an example as indicated by the black arrow. Some motor neurones are displaying TDP43 pathology, with loss of TDP43 from the nucleus (indicated by red arrows), with the presence of small TDP43 cytoplasmic aggregates (shown by green arrow). (C) C9ALS case (046/2010). TDP43 can be seen located in the nucleus (indicated by black arrow). TDP43 pathology is shown to a great extent, with large cytoplasmic TDP43 aggregates (indicated by green arrows) and complete loss of TDP43 from the nucleus (shown by red arrows).
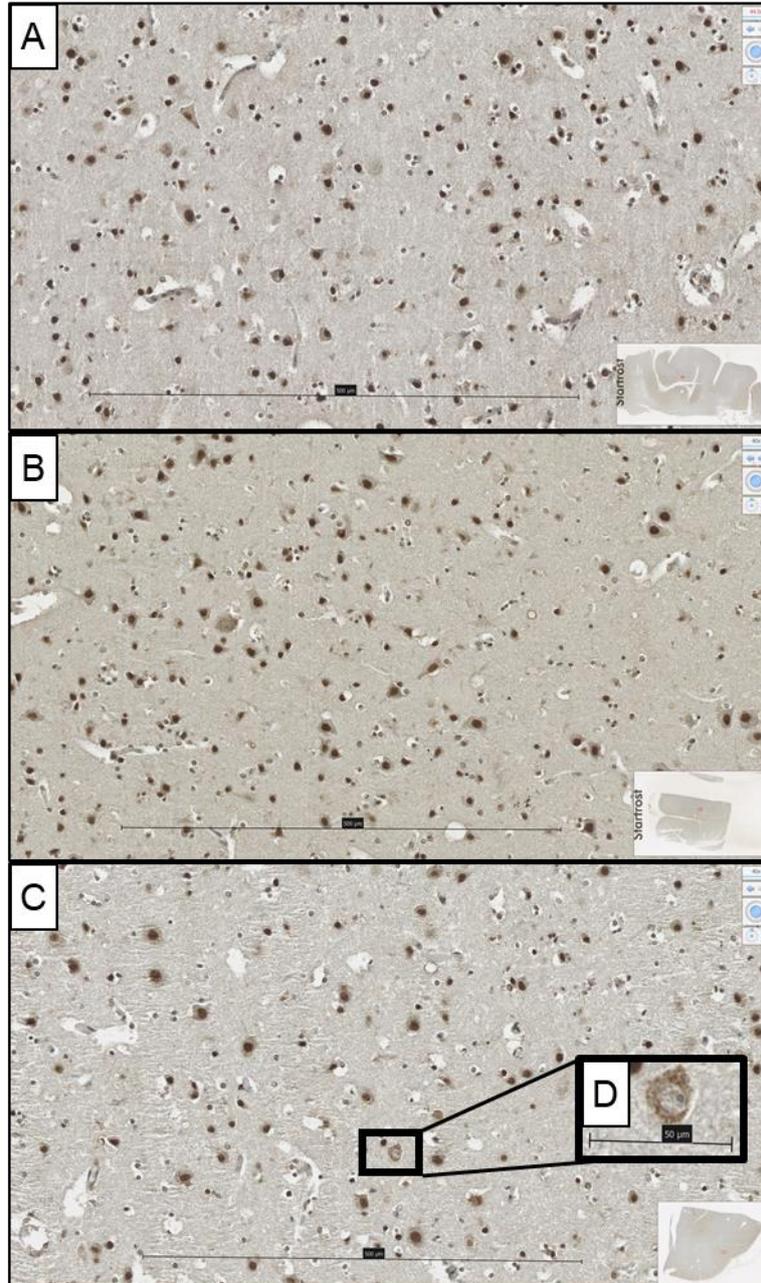
Figure 4.13: Examples of TDP43 immunohistochemistry in the motor cortex of control, sALS and C9ALS cases.
Images taken at x40 magnification, scale bar represents 500 µm.
(A) Control case (005/2007). Nuclear TDP43 immunopositive staining can be observed, with no cytoplasmic aggregation within neurones. (B) sALS case (141/2003). Nuclear TDP43 can be observed in the motor neurones. (C) C9ALS (118/2001). Motor neurones show mainly nuclear TDP43 immunostaining, with some TDP43 pathology, as shown in (D), a zoomed in image at x80 magnification with the scale bar representing 50 µm, which demonstrates loss of nuclear TDP43 and the formation of cytoplasmic TDP43 aggregates.
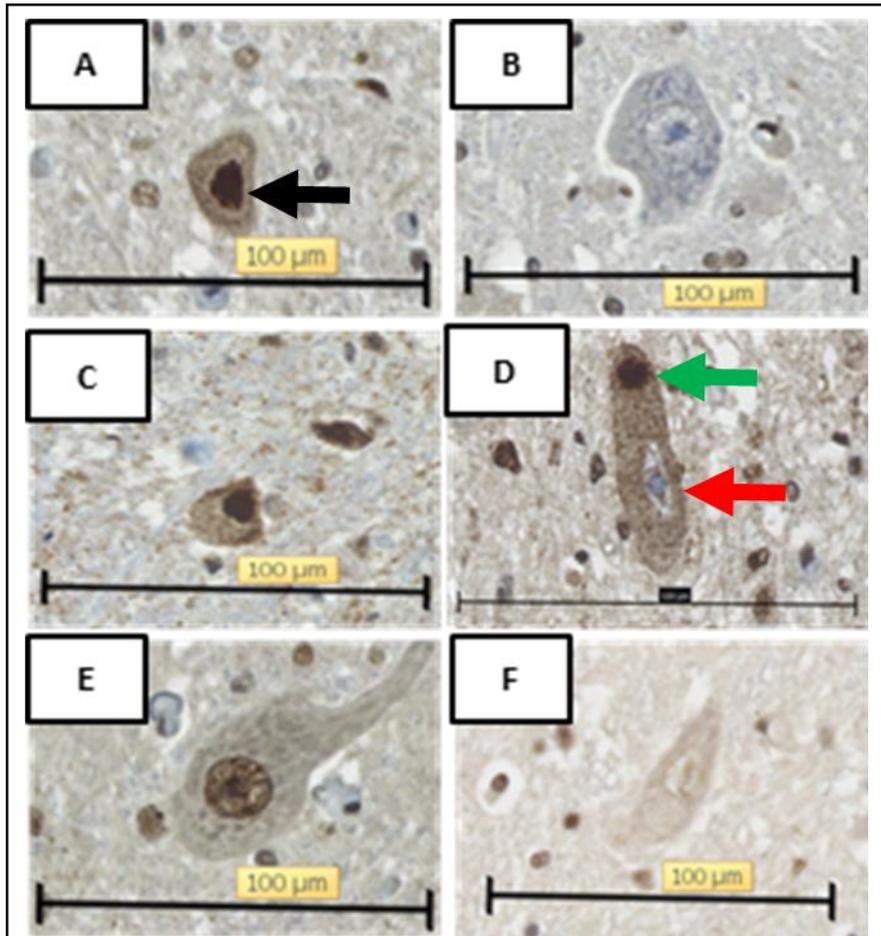
Figure 4.14: Representation of grading system used for slide analysis in TDP43 and 5mC/5hmC sequential staining.
Images taken at x80 magnification, scale bar represents 100 μm.
(**A**) MN nucleus displaying 5hmC immunopositive staining (indicated by the back arrow). (**B**) MN nucleus displaying 5hmC immunonegative staining. (**C**) MN nucleus displaying TDP43 immunopositive staining. (**D**) MN nucleus displaying TDP43 immunonegative staining, with loss of nuclear TDP43 indicated by the red arrow, and TDP43 cytoplasmic aggregation indicated by the gree arrow. (**E**) MN nucleus displaying 5mC immunopositive staining. (**F**) MN nucleus displaying 5mC immunonegative staining.
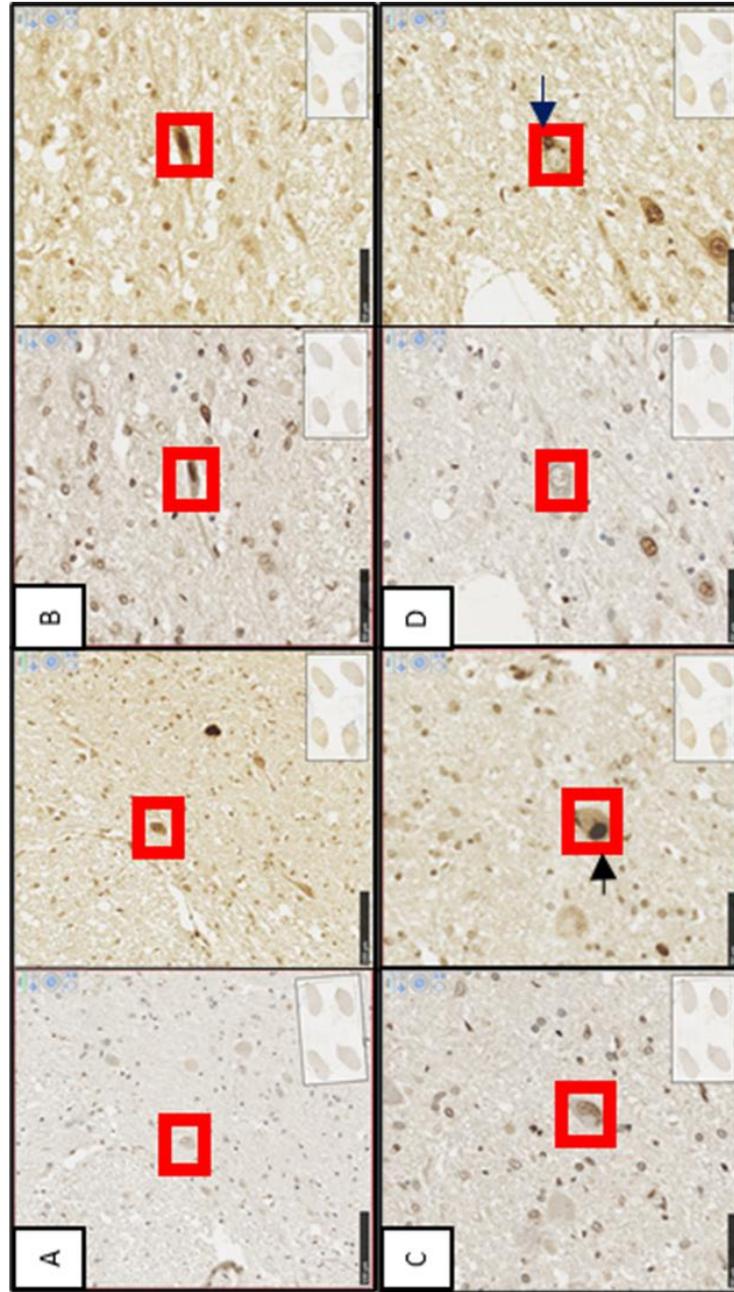
Figure 4.15: Motor neurone nuclei of the anterior horn in cervical spinal cord can be visualised and graded for 5mC and TDP43 immunopositivity in sequential sections.

Motor neurones (MNs) were identified using the Hamamatsu slide viewing software at x40 magnification, scale bar represents 100 µm, with the same MN being identified in sequential sections stained for 5mC and TDP43, respectively. Once the same MN had been identified, the nuclei was graded as either immunopositive or immunonegative for both 5mC and TDP43. Examples of MN nuclei in each of the four categories investigated can be observed. Red boxes identify the matched MN identified in the sequential sections. Scale bar represents 50 µM. Case details: All images from case 045/2006 (C9ALS). An image of the whole slide can be observed in the bottom right of each image.

A) Matched MN nuclei displaying as immunonegative for 5mC for TDP43, but immunopositive for TDP43
B) Matched MN nuclei displaying as immunopositive for both 5mC and TDP43
C) Matched MN nuclei displaying as immunopositive for 5mC, but immunonegative for TDP43. A TDP43 immunopositive cytoplasmic aggregate can be observed (black arrow)
D) Matched MN nuclei displaying as immunonegative for both 5mC and TDP43. A TDP43 immunopositive cytoplasmic aggregate can be observed (black arrow)
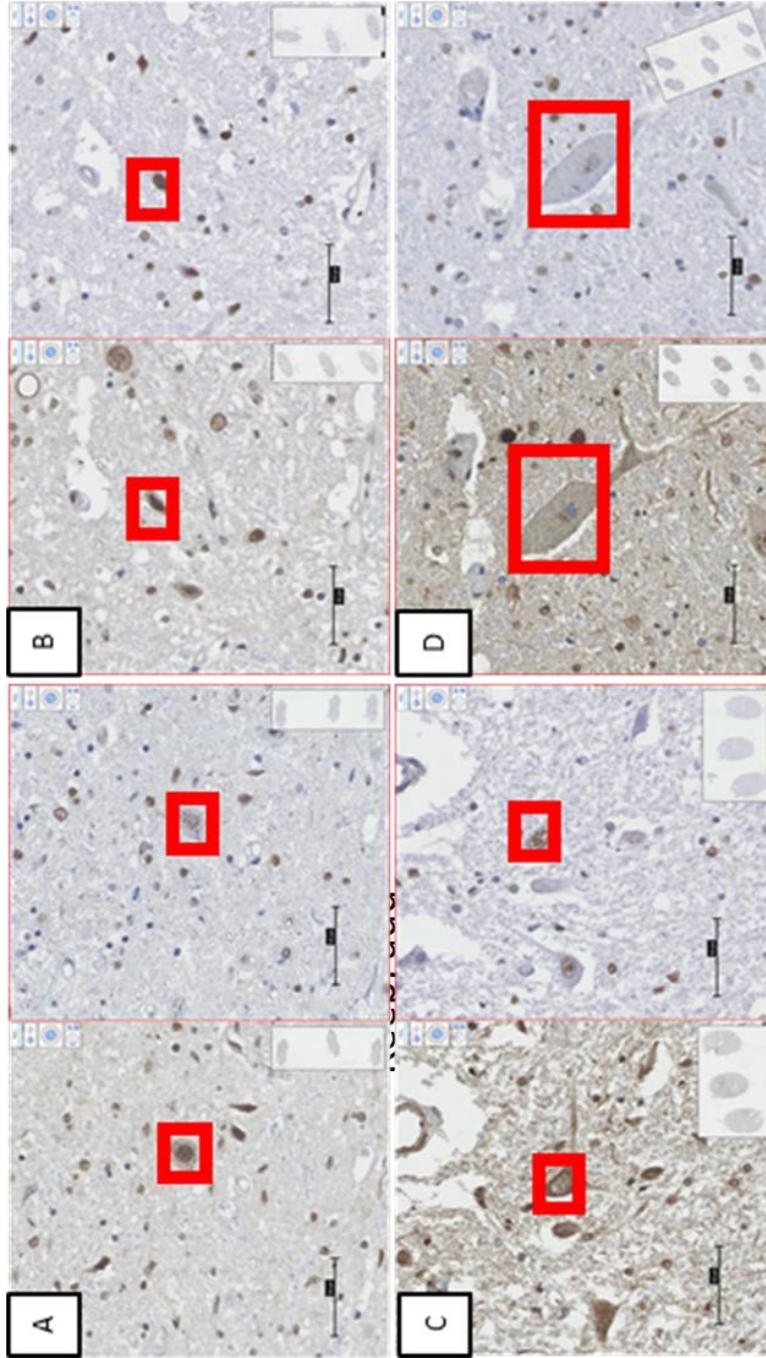
Figure 4.16: Motor neurone nuclei of the anterior horn in cervical spinal cord can be visualised and graded for 5hmC and TDP43 immunopositivity in sequential sections.

Motor neurones (MNs) were identified using the Hamamatsu slide viewing software at x40 magnification, scale bar represents 100 µm, with the same MN being identified in sequential sections stained for 5hmC and TDP43, respectively. Once the same MN had been identified, the nuclei was graded as either immunopositive or immunonegative for both 5hmC and TDP43. Examples of MN nuclei in each of the four categories investigated can be observed. Red boxes identify the matched MN identified in the sequential sections. Scale bar represents 50 µM. Case details: Images from cases: 073/2007 (A), 094/2006 (B), 273/1999 (C) and 141/2003 (D). An image of the whole slide can be observed in the bottom right of each image.

A)Matched MN nuclei displaying as immunopositive for TDP43, but immunonegative for 5hmC

B)Matched MN nuclei displaying as immunopositive for both TDP43 and 5hmC

C)Matched MN nuclei displaying as immunopositive for 5hmC, but immunonegative for TDP43. Matched MN nuclei displaying as immunonegative for both TDP43 and 5hmC.

D)Matched MN nuclei displaying as immunonegative for both TDP43 and 5hmC.

Figure 4.17: 5mC and 5hmC loss from the nucleus of motor neurones is related to TDP43 motor neurone nuclear loss (mean ±SD).

(A) 5mC. High levels of methylation were present in motor neurone nuclei positive for TDP43. However, a significantly lower percentage of motor neurone nuclei displayed immunopositivity for 5mC when nuclear TDP43 was lost, p=0.00002.

(B) 5hmC. High levels of methylation were present in motor neurone nuclei expressing TDP43. However, a significantly lower percentage of motor neurone nuclei displayed immunopositivity for 5hmC when nuclear TDP43 was lost, p=<0.00001.

4.4.7 5mC and 5hmC levels in glia in spinal cord do not differ between controls and ALS

Counts of 5mC and 5hmC positive and negative glia were similarly carried out in the AH, LCT and DC of the SC. Two-tailed *t*-tests were conducted, with no significant differences observed for any intergroup comparisons in any of the regions investigated (see figure 4.18, plus table 4.3). No sex differences were observed in any regions: AH (all $p \geq 0.22$), LCT (all $p \geq 0.41$), DC (all $p \geq 0.80$). Comparisons were also conducted, comparing the levels of DNA methylation between regions. No differences were observed between any regions ($p \geq 0.05$).
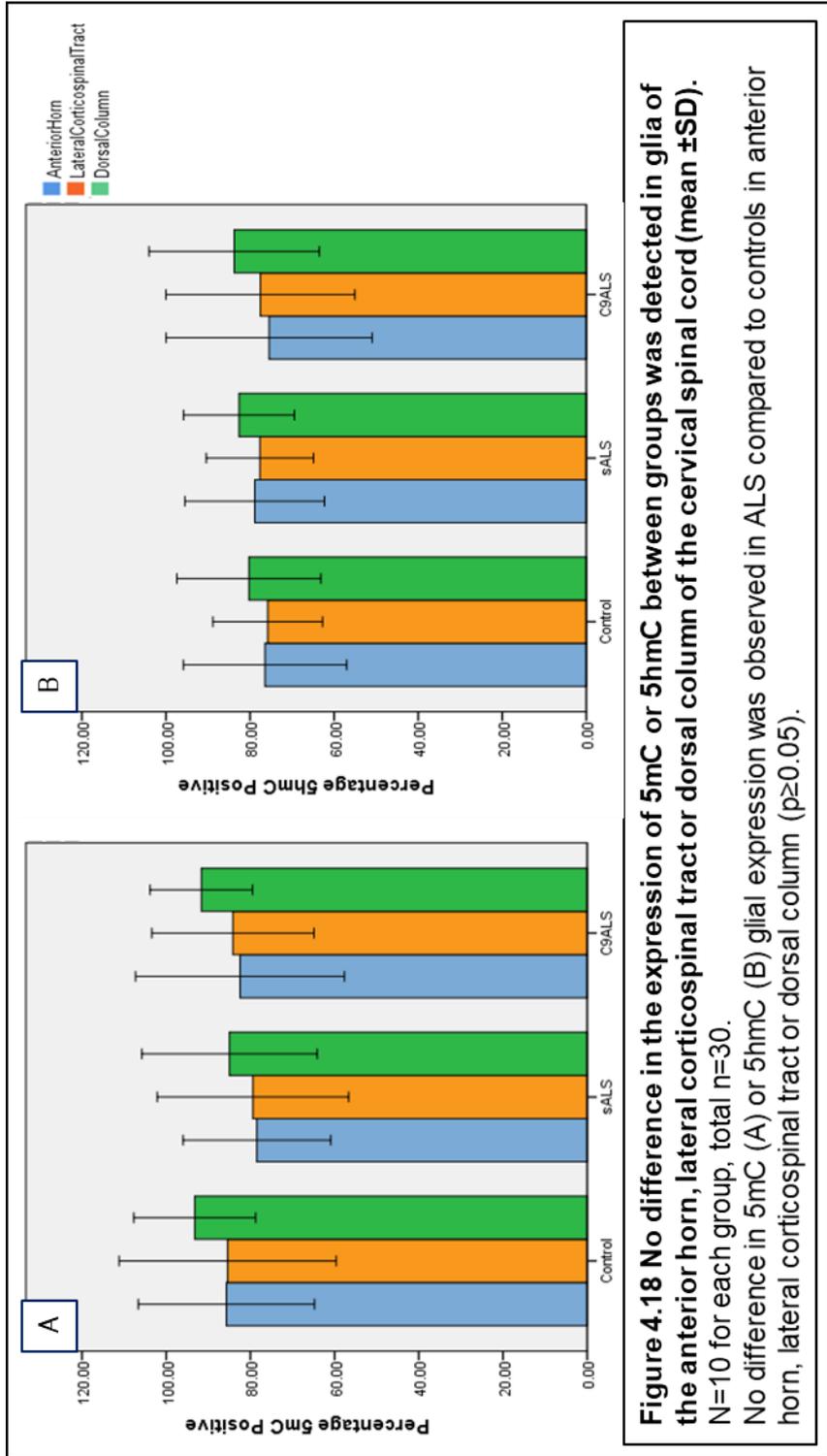
Figure 4.18 No difference in the expression of 5mC or 5hmC between groups was detected in glia of the anterior horn, lateral corticospinal tract or dorsal column of the cervical spinal cord (mean ±SD). N=10 for each group, total n=30.

No difference in 5mC (A) or 5hmC (B) glial expression was observed in ALS compared to controls in anterior horn, lateral corticospinal tract or dorsal column (p≥0.05).

Table 4.3: Statistical analysis of glial cell counts for 5mC and 5hmC in the anterior horn, lateral corticospinal tract and dorsal column of spinal cord.

Target	Area	Average immunopositivity in percentage (plus SD)			Outcome of *t*-tests, p=		
		Control	sALS	C9ALS	Control v sALS	Control v C9ALS	sALS v C9ALS
5mC	Anterior horn	85.7 (10.5)	82.5 (12.4)	78.5 (8.8)	0.11	0.54	0.42
	Lateral Corticospinal Tract	95.4 (12.9)	84.2 (9.6)	79.4 (11.4)	0.28	0.81	0.33
	Dorsal column	93.2 (7.2)	91.7 (6.1)	85.0 (10.4)	0.05	0.61	0.10
5hmC	Anterior horn	76.4 (9.7)	78.9 (8.3)	75.5 (12.3)	0.52	0.85	0.48
	Lateral Corticospinal Tract	75.8 (6.5)	77.7 (6.4)	77.5 (11.2)	0.52	0.68	0.98
	Dorsal column	80.3 (8.6)	82.6 (6.6)	83.8 (10.1)	0.98	0.42	0.77

4.4.8 No difference in 5mC or 5hmC neurone expression in motor cortex or anterior frontal cortex

p62 is a ubiquitin binding protein that functions as an autophagosome cargo protein that targets proteins that bind to it, resulting in selective autophagy. All AFCx cases used for the current IHC study had been assessed for p62 pathology. This pathology is indicative of TDP43 pathology in sALS (Cykowski *et al.*, 2018; Rossor *et al.*, 2019). Therefore, the p62 pathology status of sALS cases were used as an indicator of TDP43 pathology. All C9ALS cases displayed p62 pathology in the MCx and AFCx and previous (unpublished) observations by our group have failed to find C9ALS cases that are negative for TDP43 pathology, C9ALS cases were not stratified. In addition, for both sALS and C9ALS cases, all motor cortices displayed inclusion pathology. Therefore, the MCx data were not stratified in this manner.

Having assessed the methylation and hydroxymethylation status of neurones and glia in the SC, we next wished to assess this in the MCx and the AFCx (outside the pyramidal tract). For the MCx cohort, no intergroup differences were present for either 5mC (p≥0.24) or 5hmC (p≥0.22) (table 4.4 and figure 4.19).

Initially, for the AFCx cohort, analysis was conducted combining the sALS cases regardless of inclusion pathology (n=11 for controls, n=38 for sALS, n=12 for C9ALS). Intergroup comparisons for 5mC and 5hmC were non-significant (p≥0.20), with the exception of the comparison between control v sALS 5mC immunopositive levels (p=0.04) and sALS v C9ALS 5hmC immunopositive levels (p=0.02) (also seen in table 4.4 and figure 4.19).

Thus, the anterior AFCx data was split into those displaying inclusion pathology (n=13) and those that were not (n=17). Eight cases had no inclusion data, so were excluded from further analysis. There were no significant difference observed for all comparisons (p≥ 0.05) (table 4.5 and figure 4.20). No sex differences were observed in MCx neurones (p=0.09 for 5mC, p=0.33 for 5hmC), with no sex differences present in FCx neurones (p=0.06 for 5mC, 0.09 for 5hmC).
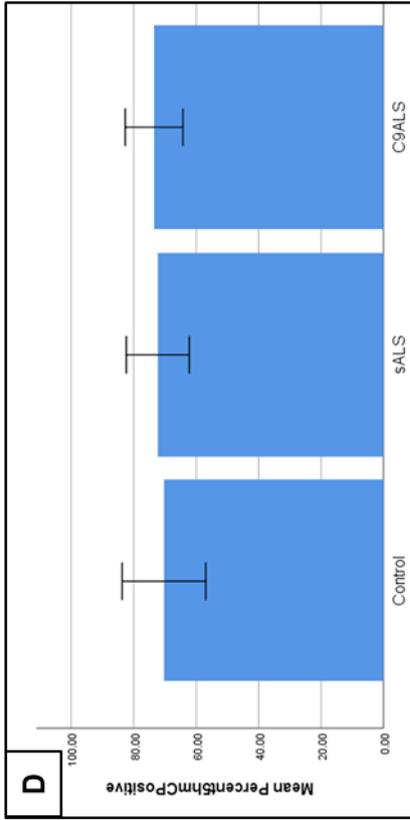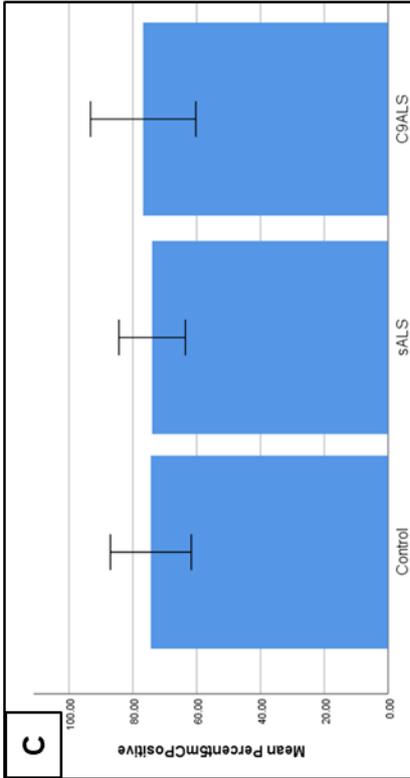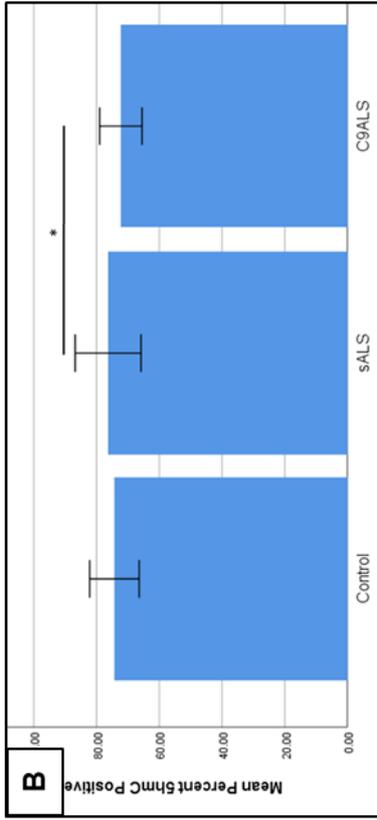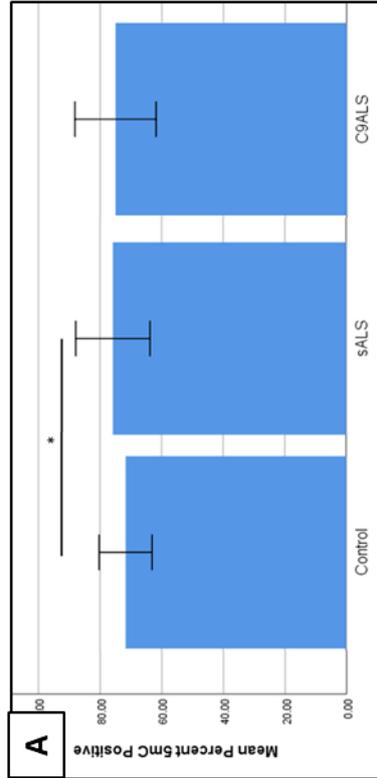
Figure 4.19: 5mC/5hmC MN immunopositivity show no difference in ALS versus controls for motor cortex and frontal cortex

(A) Frontal cortex 5mC: A significant difference was observed between control and sALS (p=0.04). (B) Frontal cortex 5hmC: A significant difference was observed between sALS and C9ALS (p=0.02). No differences in 5mC (C) or 5hmC (D).

Table 4.4: Statistical analysis of motor neurone cell counts for 5mC and 5hmC in motor and anterior frontal cortex.

Target	Area	Average immunopositivity in percentage (plus SD)			Outcome of *t*-tests, p=		
		Control	sALS	C9ALS	Control v sALS	Control v C9ALS	sALS v C9ALS
5mC	Motor cortex	74.4 (6.3)	74.0 (5.2)	76.8 (8.3)	0.86	0.50	0.24
	Frontal cortex	71.7 (4.3)	75.8 (6.0)	75.0 (6.1)	0.04	0.59	0.83
5hmC	Motor cortex	70.3 (6.7)	72.3 (5.0)	73.5 (4.6)	0.40	0.22	0.51
	Frontal cortex	74.3 (4.0)	76.3 (5.2)	72.3 (3.4)	0.24	0.20	0.02

Figure 4.20: 5mC/5hmC MN barcharts for anterior frontal cortex sorted for inclusion pathology (mean ±SD).
Frontal cortex cases were split into those displaying TDP43 pathology, and those not displaying TDP43 pathology. For 5mC (**A**) and 5hmC (**B**), no inter-group differences were observed (p≥ 0.05).
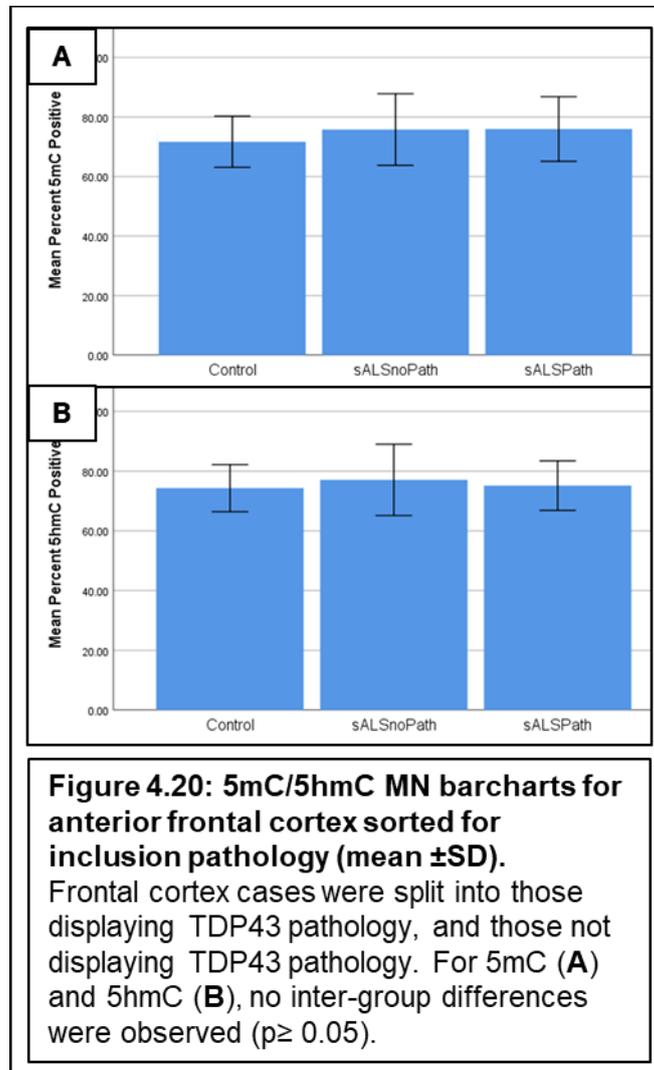
Table 4.5: Statistical analysis of neuronal cell counts for 5mC and 5hmC anterior frontal cortex sorted for TDP43 pathology. sALSnoPath = sALS with no TDP43 pathology, sALSPath = sALS with TDP43 pathology

Target	Area	Average immunopositivity in percentage (plus SD)			Outcome of t-tests, p=		
		Control	sALS noPath	sALS Path	Control v sALS noPath	Control v sALS Path	sALS noPath v sALS Path
5mC	Anterior frontal cortex (TDP43 pathology)	71.7 (4.3)	75.8 (6.0)	76.0 (5.4)	0.06	0.05	0.93
5hmC	Anterior frontal cortex (TDP43 pathology)	64.3 (4.0)	77.1 (6.0)	75.2 (4.2)	0.19	0.61	0.33

4.4.9 No difference in 5mC or 5hmC glial expression in motor cortex or anterior frontal cortex

For 5mC and 5hmC levels, there were no significant differences were observed between between controls and ALS cohorts in either the MCx or AFCx (figure 4.21; For AFCx, p≥0.41 for 5mC and p≥0.24 for 5hmC; for MCx, p≥0.19 for 5mC, and p≥0.17 for 5hmC; table 4.6). As was the case for the neurones, glia showed no differences between sexes (p=0.36 for MCx 5mC, p=0.09 for MCx 5hmC, p=0.40 for AFCx5mC, AFCx p=0.91 for 5hmC).
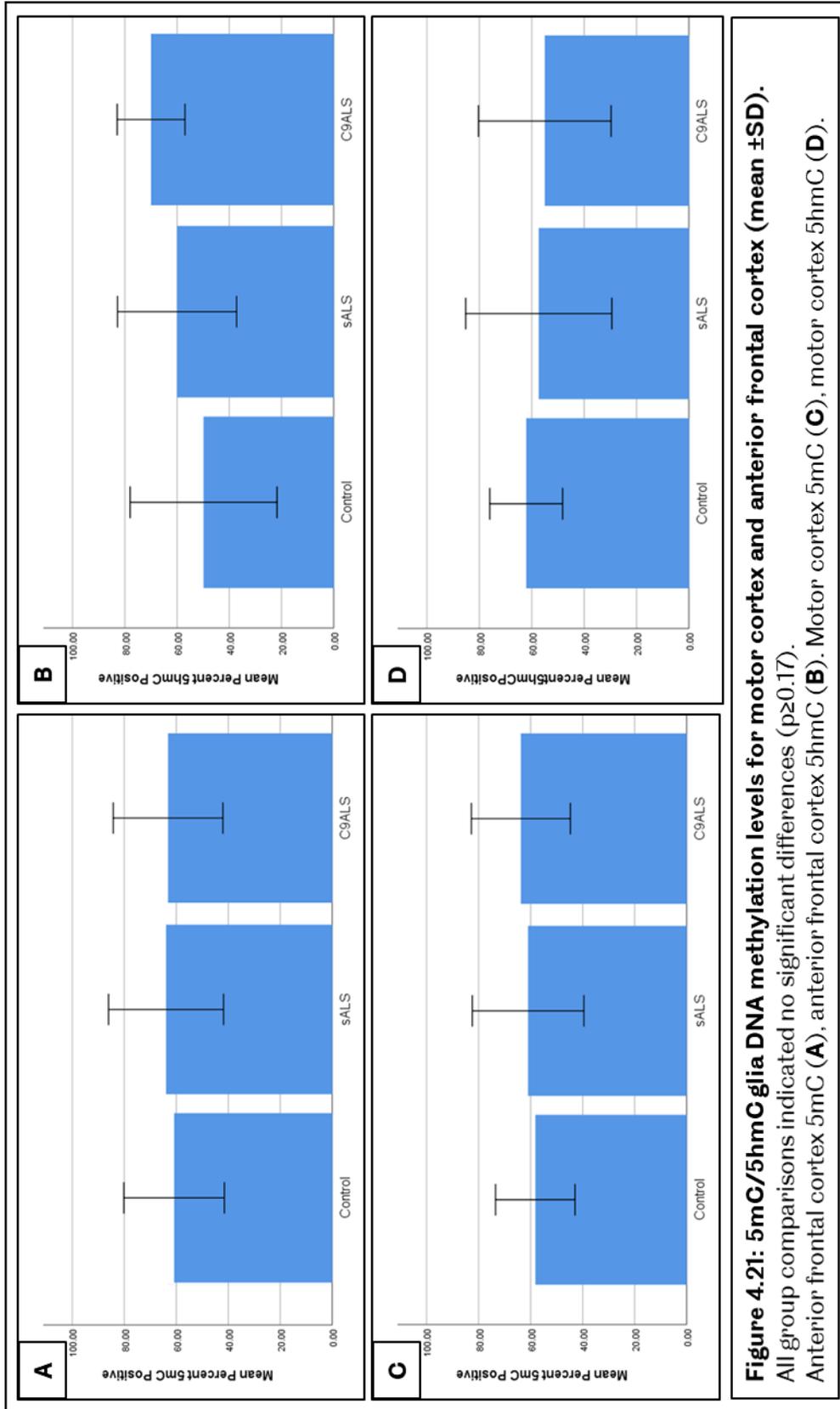
Figure 4.21: 5mC/5hmC glia DNA methylation levels for motor cortex and anterior frontal cortex (mean ±SD).
All group comparisons indicated no significant differences (p≥0.17).
Anterior frontal cortex 5mC (**A**), anterior frontal cortex 5hmC (**B**). Motor cortex 5mC (**C**), motor cortex 5hmC (**D**).

Table 4.6: Statistical analysis of glial cell counts for 5mC and 5hmC in motor and anterior frontal cortex.

Target	Area	Average immunopositivity in percentage (plus SD)			Outcome of *t*-tests, p=		
		Control	sALS	C9ALS	Control v sALS	Control v C9ALS	sALS v C9ALS
5mC	Motor cortex	58.2 (7.6)	61.0 (10.7)	63.8 (9.5)	0.51	0.19	0.46
	Anterior frontal cortex	60.8 (9.7)	63.9 (11.0)	63.2 (10.8)	0.41	0.59	0.83
5hmC	Motor cortex	62.1 (7.0)	57.3 (14.0)	55.0 (12.6)	0.37	0.17	0.64
	Anterior frontal cortex	49.8 (14.1)	60.0 (11.4)	62.4 (11.2)	0.24	0.44	0.93

4.4.10 DNA methylation levels are higher in the neurones of the spinal cord compared to motor cortex and frontal cortex

Cell type-specific DNA methylation level differences between regions were also analysed. DNA methylation levels were compared between SC, MCx and AFCx. For this analysis, all cases from each cohort were pooled. For both neuronal cells, the highest levels of DNA methylation were seen in the SC, with comparisons between SC v AFCx and SC v MCx both being significantly higher (p≤0.005) (table 4.7). When focusing solely on the control group, this pattern was still observed (p≤0.0003).

Table 4.7: Higher levels of DNA methylation are present in spinal cord neurones when compared to motor cortex and anterior frontal cortex neurones

Target	Area	Mean (plus SD)	Outcome of *t*-tests, p=	
5mC	Spinal cord	86.1 (6.2)	Spinal cord v motor cortex	2E-10
	Motor cortex	74.9 (6.4)	Spinal cord v anterior frontal cortex	1E-12
	Anterior frontal cortex	74.9 (6.0)	Motor cortex v anterior frontal cortex	0.97
5hmC	Spinal cord	87.6 (6.3)	Spinal cord v motor cortex	5E-17
	Motor cortex	72.2 (5.3)	Spinal cord v anterior frontal cortex	1E-17
	Anterior frontal cortex	75.2 (4.9)	Motor cortex v anterior frontal cortex	0.005

4.4.11 Disease survival duration does not correlate with degree of methylation

Within the cohort of ALS cases, there was a variation in survival rates. In this case, survival rate referred to the length of time between initial diagnosis of ALS and death. In order to establish if there is any correlation between DNA methylation status and survival rates in neurones, linear regression analysis was conducted in SC. No correlation between DNA methylation levels and survival rates was detected (figure 4.22). For 5mC, R^2= 0.0003, and R^2=0.2289 for 5hmC in for SC.

Figure 4.22: No correlation between degree of methylation and survival rates in spinal cord motor neurones.
No correlation between methylation levels and survival rates were detected, as measured by regression analysis. Spinal cord 5mC (A), spinal cord 5hmC (B).
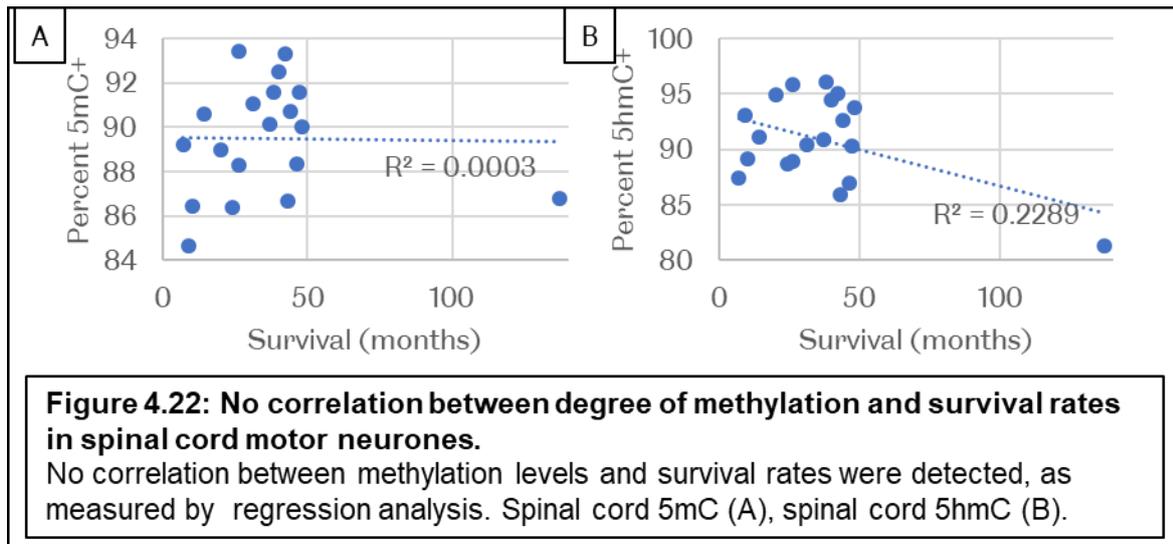
4.5 Discussion

In this chapter, a pathological characterisation of DNA methylation in ALS has been undertaken. Using IHC, assessments of levels of DNA methylation were conducted, using the DNA methylation markers 5mC and 5hmC. The results provide evidence supporting the role of LMN-specific DNA methylation and hydroxymethylation in ALS. Additionally, further investigation into the LMNs of the AH in SC related TDP43 pathology to loss of DNA methylation from the nucleus. Mutual validation was observed through the use of two separate markers (5mC and 5hmC).

4.5.1 Residual motor neurones of the spinal cord display significantly higher levels of DNA methylation in ALS

The first hypothesis was that an increase in DNA methylation would be observed in ALS, with the highest levels in C9ALS. In the initial analysis of MNs in the SC, this was the case for both markers of DNA methylation used (5mC and 5hmC), with significantly higher levels of both DNA methylation and hydroxymethylation observed in sALS and C9ALS cases

compared to controls. However, no differences in expression was observed when comparing sALS and C9ALS.

Other studies have previously established that DNA methylation changes occur in ALS, but all have limitations. Figueroa-Romero *et al.*, 2012 conducted a study of DNA methylation in sALS SC, determined by ELISA plates for 5mC and 5hmC. They found that, as in the study discussed in this chapter, an increase in DNA methylation was present for both markers of methylation. However, this was a tissue homogenate, and this thesis shows the cell-specific nature of DNA methylation changes. Another consideration is the cell proportions in this study, as ALS is characterised by MN loss, so this will be reflected in the use of homogenates. Blood analysis was also conducted, showing no differences. This also indicates the importance of cell-specificity.

4.5.2 TDP43 pathology associates with loss of DNA methylation from the LMN nucleus

The second hypothesis of this chapter was that DNA methylation would be associated with TDP43 pathology in LMNs of the SC within the ALS groups. In cells displaying TDP43 pathology (i.e. loss of nuclear TDP43 staining), a loss of DNA methylation from the nucleus was found. This was the case for both 5mC and 5hmC. ALS is primarily a disease of the MNs, so it is to be expected that for the ALS samples, some MNs will have already undergone cell neurodegeneration and death. Even though a relationship between DNA hypomethylation and TDP43 pathology was found, there is no confirmation of the direction of causation. One possibility is that only resilient MNs are observed in the tissue, which display DNA hypermethylation, with other less resilient MNs dying off before the patient died. This highlights another limitation of using PM tissue, as it is representative of end stage disease, with no indication of the DNA methylation changes, if any, taking place during the course of the disease.

It is prudent to point out that the MNs displaying TDP43 pathology were only a small subset in comparison to the whole MN population. This is relevant, as when neurones as a whole are examined, the majority will not display TDP43 pathology. This means that when the whole population of MNs is analysed without respect to TDP43 status (as in 4.5.1 above) the data will be biased towards DNA methylation levels seen in neurones without TDP43 pathology.

A number of potential reasons for the association with nuclear TDP43 loss and hypomethylation can be posited. Firstly, TDP43 pathology may affect cells with hypomethylation more than cells without TDP43 pathology, and hypermethylation plays a neuroprotective role. Another possibility is that the TDP43 pathology causes the drop in DNA

methylation. Finally, another currently unknown factor could be responsible for both the TDP43 pathology and DNA methylation loss. Unfortunately, the direction of causation is unable to be determined with the use of PM tissue. It is possible to conduct mechanistic experiments in model systems to address this, as both the levels of TDP43 pathology and DNA methylation can be manipulated in order to establish causation. This is discussed further in section 6.2.

As seen in figure 4.11, figure 4.17 and section 4.4.5, some variation in 5mC/5hmC and TDP43 expression was observed when compared with expression levels observed in the total cell counts of 5mC/5hmC and TDP43. Only a subset of motor neurone nuclei was able to be visualised and matched up in the TDP43 association studies as nuclei were only able to be visualised in subsequent sections for a selection of the total numbers of motor neurones present. This reduced the number of motor neurones assessed in the TDP43 pathology/DNA methylation and TDP43/DNA hydroxymethylation association studies. Further, the number of motor neurones displaying TDP43 pathology is relatively small, resulting in a small number of cells being assessed in the TDP43 pathology association studies.

4.5.3 DNA methylation is higher in spinal cord MNs compared to neurones of the motor cortex and anterior frontal cortex

To assess whether levels of DNA methylation are higher in the SC, levels of DNA methylation were compared between SC, AFCx and MCx. Significantly higher levels were found in the SC (table 4.7). Together, these findings suggest a LMN/SC specific effect, again raising the possibility of selective vulnerability.

Selective vulnerability refers to the tendency for a given pathological process (ALS in this instance) to affect some CNS regions or cell populations more than others. Vulnerable neurones are more likely to succumb to cytotoxic events, such as cell death related signalling, neuroinflammation and synaptic toxicity (Fu *et al.*, 2018). TDP43 pathology mapping studies, such as those discussed in section 1.3.2, suggest that protein aggregates accumulate in regions of primary vulnerability, before spreading to areas of secondary vulnerability.

A number of potential causes have been posited in relation to selective vulnerability. A risk factor associated with many neurodegenerative diseases, ALS included, is ageing. This is the consequence of aged neurones coming to the end of their lifespan. This lifespan may vary between neurones, dependent on their function, past exposure to stress and their genetic predisposition. Protein homeostasis is also implicated, with some neurones

theorised to be more affected by how TDP43 protein aggregates disrupt normal protein homeostasis networks. In the IHC studies presented in this chapter, different levels of methylation and hydroxymethylation in spinal cord are observed, compared to the brain. Therefore, the spinal cord LMNs have methylation pathology in spinal cord, whereas no methylation pathology is seen in neurones of the MCx or AFCx. It may be that spinal cord requires greater methylation (for reasons unknown) and that may make it selectively vulnerable to TDP43 pathology. Exploring the pathological mechanisms that methylation protects is an interesting possibility to research (discussed further in section 6.4).

4.5.4 Methylation and hydroxymethylation in the motor cortex and anterior frontal lobe

No intergroup differences were found in MCx or AFCx (figure 4.19). The original hypothesis was that there would be intergroup differences in MCx and AFCx. Based on this assumption, the intention was to use the sporadic ALS inclusion positive and negative cases as a way to look at whether DNA methylation differences were seen before or after the arrival of inclusion pathology in the anterior frontal cortex. However, given that the MCx data show that dysmethylation is not seen in the forebrain, the premise for this strategy was not sound. Therefore only archival p62 data was used, rather than conducting further TDP43 IHC in the AFCx. Furthermore, it was found that the presence of absence of inclusion pathology in the frontal lobes did not affect methylation. One hypothesis for this is that TDP43 pathology affects a lesser percentage of neurones in the forebrain than in LMNs, another is the potential role of selective vulnerability (discussed in section 4.5.3).

Overall, a change in DNA methylation and hydroxymethylation was observed in ALS in the LMNs of SC, which were not observed in either the MCx or AFCx. This could imply a different pathologic process and would also suggest that further studies in DNA methylation and hydroxymethylation in ALS should focus on the SC.

4.5.5 Levels of DNA methylation in glia are unchanged in ALS

Glial cells were counted in the AH, LCT and DC of the SC, as this encompassed glial cells in both grey and white matter. Counting was also conducted on glial cells in the white matter of AFCx and MCx. In all regions analysed, no difference was detected in the levels of DNA methylation between control and ALS. This further supports the notion that DNA methylation is a LMN specific phenomenon.

While ALS is primarily a disease of the MNs, it is also well established that glial cells play a fundamental role in the disease, in both neuroprotective and neurotoxic roles (Pehar *et al.*, 2018). Glia make up 90% of all CNS cells. Astrogliosis and microglial activation are present in most neurodegenerative diseases, and can contribute to the pro-inflammatory phenotype

found in ALS (Staszewski and Prinz, 2014). It has also been shown to lead to alterations in neuronal-glial interactions, which lead to synaptic function decline and ultimately, neuronal death (Li *et al.*, 2011). Most studies have focused on DNA methylation changes in glia in neurodevelopment (Neal and Richardson, 2018). No studies to date have been conducted to assess DNA methylation levels of glial cells in ALS, but studies have been performed in AD, with contradictory results: One study found global hypermethylation in AD (Rao *et al.*, 2012), another found global hypomethylation in AD (Mastroeni *et al.*, 2011), and a third study found no global differences in DNA methylation in AD (Schwob *et al.,* 1990).

With this large variability in findings for AD, and lack of publications for ALS glia and DNA methylation in mind, an analysis of glial cells in ALS is warranted. It is important to note that the studies presented in this chapter are only a measure of global DNA methylation, and it is probable that at the gene level, changes in DNA methylation will be present in glial cells. Following from this, a limitation of the current study is that multiple glial subtypes have been grouped together. Given that DNA methylation levels are different in neurones in regions focused on within this study, it would be of value to perform separate cell counts of these different cellular subtypes to fully understand DNA methylation patterns of glial cells in ALS. There is also evidence to suggest that methylation is potentially a bigger issue for non-mitotic cells versus mitotic cells (discussed in section 1.6). As neurones are post-mitotic, it is not surprising that methylation changes are observed in ALS. A potential way to conduct this would be through double staining for both a glia marker and DNA methylation marker.

4.5.6 DNA methylation and the normal ageing brain

To ensure that any significant findings were due to ALS, and not due to DNA methylation changes associated with age, linear regression analysis was conducted in this study, finding that no changes were present (section 4.4.4).

Both increases and decreases in DNA methylation are found to occur with increasing age, dependent on the particular tissue or gene studied (Richardson, 2003), with DNA methylation in the brain and CNS being significantly higher than other regions of the body (Wilson *et al.*, 1987).

In general, a trend for reduced DNA methylation with increasing age in the brain is posited (Catania and Fairweather 1991). Ageing at the cell/organ level can be characterised by reductions in response to intercellular signals, which could in turn result in changes to normal pathways for gene expression. A potential mechanism for this could be DNA methylation. A hypothesis favoured by Catania and Fairweather in 1991 suggested that

methylated regions of DNA could become selectively lost during ageing, with DNMTs also becoming less efficient over time.

A study conducted by Hernandez *et al.*, in 2011 investigated methylation changes occurring in different regions of the brain, analysed using the Illumina DNA methylation 27k arrays, one of the predecessors of the current MethylationEPIC BeadChip array. In this study, the frontal cortex, cerebellum, pons and temporal lobe underwent analysis, with the age range of donors being vast (1 years to 102 years), in order to establish any changes in DNA methylation associated with chronological ageing. Interestingly, this study identified ten loci displaying a positive correlation with DNA methylation and ageing, which is in disagreement the general consensus.

Together, what these findings do establish is the complexity of DNA methylation changes that occur in the brain and CNS in normal ageing. Therefore, it was important to establish that any changes observed in DNA methylation levels between the control and ALS cohorts was actually a result of healthy v disease changes, and not changes associated with age. Linear regression analysis was also conducted on the association of survival rates and DNA methylation levels, again to establish if any changes to DNA methylation in the ALS cohort could be explained by another factor. This analysis also showed no correlation between survival rates and DNA methylation levels (figure 4.22).

The effects of age on methylation may be a potential mechanism for the increasing prevalence of neurodegeneration in general, and ALS in particular, with age. The studies discussed above find changes in methylation with age, but this was not found in the cohort used in this chapter. It is probable that no effect was seen in the cohort used for this study as a very narrow range of ages were used, whereas the studies discussed previously in this section detail changes over a prolonged period of time. Another factor is that IHC is a very course measure of methylation across the genome, whereas the studies that found age-related epigenetic drift involved larger numbers of cases, and used techniques allowing for gene-specific DNA methylation changes to be analysed.

Together, this gave confidence that any changes seen were because of DNA methylation changes that have occurred due to ALS, and not as a consequence of chronological age or disease duration.

4.5.7 Study limitations

As with any study, this study has a number of limitations. These include the small number of controls for the MCx cohort (n=8). Post-hoc power analysis was performed for the spinal cord study (power of 0.87), frontal cortex study (power of 0.83) and motor cortex (power of

0.77) using G*power v3.1.9.4 (Faul *et al.*, 2007). A 'perfect' study would have a power of 1.0, which is not the case for the cohorts used in this study. However, the recorded powers do indicate that the studies are of a reasonable size, with powers of 0.87, 0.83 and 0.77 indicating sufficient power to be appropriate for the context of the study (human samples, time available for analysis). In order to increase the power of these studies and to make the findings more representative of the general population, increasing the cohort size and carrying out replication in samples from another brain bank could be of value.

Variability between raters can also be a potential source of lack of reproducibility, as individual cells are identified/interpreted differently (Alafuzoff *et al.*, 2006). In this case, inter-rater reliability tests were conducted to ensure raters were assessing and determining cell immunopositivity status in reproducible ways. These tests revealed that inter-rater reliability was high, suggesting this limitation was reduced, and had an added level of robustness.

It is important to note that two samples used in the control cohort were aged 26 and 39, which is lower than in the ALS cohort. No lower cut off for age was enforced in this study, as there was low availability of control tissue available, which reduced the number of limitations that could be placed on the study. To ensure that age was not a contributory factor in any significant results, statistical analysis was carried out, with no significant difference found when comparing the cohorts (table 2.3). After consideration, it was decided that having a larger cohort of control samples was of higher importance than removal of younger samples from the control cohort.

IHC can be viewed as a course measure of 5mC and 5hmC expression. This is further compounded by the limitations of DAB as a visual representation of immunopositivity. All cells analysed will have varying levels of 5mC and 5hmC present within them. However, IHC visualised with DAB cannot be classed as quantitative due to this only being a measure of the absence or presence of DAB immunostaining. Therefore, this is only a course measure of 5mC and 5hmC present within a cell. This highlights the issue of sensitivity, with the potential for low levels of 5mC and 5hmC within the cell not being detected due to the lack of sensitivity associated with DAB. This is addressed by the use of the MethlationEpic array discussed in chapter 5, which assesses DNA methylation at the single gene level.

The major limitation in this study, and with the use of PM tissue in general, is that is it representative of 'end stage' disease. Therefore, these findings may not reflect the DNA methylation status of 'early stage' ALS. A number of efforts have been used to account for this. The use of frontal cortex can be viewed as a representation of 'early stage' ALS, as it is one of the last regions to be affected.

Even with the limitations discussed above, FFPE tissue and histological studies are of high value. FFPE cases are generally well characterised and are neuropathologically assessed for signs of ALS, with data including age, sex and disease survival usually available for cases. It is important to characterise cells from PM sources, as the cells are taken from their native environment, and deductions made from these are more representative of the heterogeneous tissues of the body. This is not possible using cell models.

Another issue when using PM tissue is that it is only ever possible to find a correlation, and correlation does not imply the causation of the effect seen. This can be addressed through the manipulation of disease models, such as cell culture, zebrafish models and mouse models. This is discussed at length in section 6.2.

The final issue is that when conducting the analysis of the difference in DNA methylation and hydroxymethylation in LMNs with and without TDP43 pathology, the effect observed was the opposite of what was hypothesised. Considerations for this include that the hypothesis tested was prior to the awareness of the study by Štalekar *et al.*, 2015, which saw TDP43 proteinopathy with decreases in DNMT3a expression. In retrospect, had we been aware of this study prior to hypothesis making, it may have resulted in a change of hypothesis. This finding highlights that replication is necessary.

4.5.8 Suggestions for future works

Suggestions for future work would include replicating the findings of this chapter in another cohort from a different brain bank, as well as replicating these findings in the cohort used in this study by another independent rater.

Once DNMT antibodies are more developed and have been applied to both IHC and neurodegeneration, it would be of use to determine levels of DNMTs in neurones. DNMTs are catalysts for the methylation of cytosine, so it would be plausible that if DNA methylation were lost from the nucleus, that DNMT levels would also be less in the nucleus. A study to determine if any association between DNMT expression and TDP43 pathology was present would be of use, and if this mimics the findings of the study in this chapter, where TDP43 pathology was associated with DNA methylation loss from the nucleus. A study of similar design, where DNMT status and DNA methylation status in the same nucleus was identified would also help to understand the association of DNA methylation and its catalysts. Another alternative is conducting reverse transcriptase-polymerase chain reaction (RT-PCR) analysis on LMNs extracted from frozen tissue using laser capture microdissection.

A study of the hippocampal AD brain showed an increase in 5mC and 5hmC levels, while also exhibiting a decrease in 5fC and 5caC (Bradley-Whitman and Lovell, 2013), which are

the intermediates in the demethylation pathway (figure 1.6). Potential reasons for this are the absence of a mechanism to restore normal DNA methylation levels, which results in 5mC and 5hmC accumulation. As an increase of 5mC and 5hmC levels is detected in LMNs in the study described in this chapter, as was shown in the study by Bradley-Whitman and Lovell, it would also be interesting to observe if a decrease in 5fC and 5caC is present in ALS LMNs.

In order to understand the causative factors in the correlation found between TDP43 pathology and loss of DNA methylation from the nucleus, experimental manipulation needs to be carried out. One way to do this would be to manipulate TDP43 levels, either by knockdown or introducing a transgene, and then assessing if any changes in DNA methylation levels occur. Conversely, DNA methylation levels could be manipulated by drugs targeting DNMTs, knockdown, or upregulation of DNMTs and assess what happens to TDP43, potentially under conditions of cell stress. This is discussed further in section 6.2.3.

4.5.9 Antibody optimisation, antibody specificity and the limitations of immunohistochemistry

In this study, all DNMT antibodies trialled failed to produce specific immunoreactivity with minimal non-specific staining. Many factors can affect the variability in successful IHC. These include the IHC methods carried out, the type of tissue used and the variety of antibodies available. IHC variability can be caused by the use of either FFPE or frozen tissue. Both types of tissue are fixed, but their properties do differ from each other. Frozen tissue is fixed using acetone; with FFPE tissue fixed using formalin. With these fixation processes comes differences in epitope availability, and therefore antibody recognition. Benefits of using frozen tissue mean that this is overcome. However, frozen tissue does have its limitations, including specialised storage (and potential for tissue degradation if stored incorrectly) and poor morphology once IHC has taken place. This makes for increased difficulty in assessing IHC. These limitations are addressed in the use of FFPE.

Throughout this study, both monoclonal and polyclonal antibodies have been utilised, with benefits and limitations to both. Monoclonal antibodies tend to be more specific, as they should only recognise one epitope on a protein. The caveat of this is that FFPE tissue is vulnerable to the loss of this single epitope due to both the fixation process, antigen retrieval and post-mortem degradation, and cross-reactivity, and non-specific reactivity can still occur. This could result in IHC failing, with no immunostaining taking place. In contrast to this, polyclonal antibodies recognise many epitopes. However, with this less specific binding comes the risk of false positives. This is due to cross-reactivity which results in high background signal.

There are many potential reasons as to why the DNMT antibodies could not be optimised. One is incompatibility with tissue from the brain bank used in this study. A solution to this would be to trial the antibodies in tissue from another brain bank, or use tissue from the same brain bank from a region unrelated to the brain/CNS, to determine if this is a brain/CNS specific phenomenon. However, the possibility of the two points raised above is low, as a large number of antibodies from a variety of manufacturers were tested. With this being the case, it would be reasonable to assume that some of them would be compatible. Another consideration is that DNMTs are expressed in such low levels in neurones and glia that they are undetectable, with the most likely reason being that the DNMT antibodies are simply not adequate yet. This conclusion was reflected in the immunostaining detailed in this study. Non-specific staining was found in the majority of antibodies and conditions trialled, highlighting that the DNMT antibodies currently available are not yet of a useable quality to establish genuine patterns of immunostaining, and potentially that the antibodies are of poor quality and are incapable of binding to their target even in optimal conditions.

A web search was conducted using the supplier website 'references' function to identify if any of the DNMT antibodies used had now been published in the neurodegeneration field to compare immunostaining to (summarised in table 4.8). All DNMT antibodies were mainly published in cancer fields, with a focus on the use of western blotting. Few publications were for the use of IHC. One study used the Abcam DNMT1 antibody for IHC (Mastroeni *et al.*, 2010), but no images showing the DNMT1 IHC were published, and no evidence of specificity testing was present in the publication.

Table 4.8: Searches conducted for DNMT antibodies tested		
Antibody	Supplier	Search findings
DNMT1	Abcam, cat no. ab19905	• Mostly in cancer • One study in AD, using IHC. However, no images of immunostaining are present (Mastroeni *et al.*, 2010)
DNMT1	Novus Biologicals, cat no. 60B1220.1	Mostly in cancer
DNMT1	Santa Cruz Biotechnology, cat no. sc-271729	Mostly in cancer
DNMT1	Proteintech, cat no. 24206-1-AP	Mostly in cancer, all western blot analysis
DNMT1	Abgent, cat no. AP1032b	Mostly in cancer, all western blot analysis
DNMT1	R&D Systems, cat no. AF6110	No publications
DNMT3a	Santa Cruz Biotechnology, cat no. sc-10232	Discontinued, no data
DNMT3a	Abcam, cat no. ab4897	Mostly in cancer
DNMT3a	Abgent, cat no. AP1034a	Mostly in cancer

If there were no limitations on tissue, further antigen retrieval processes could be trialled, such as enzymatic antigen retrieval and formic acid antigen retrieval. For enzymatic antigen retrieval, sections are immersed in a solution containing the enzyme Trypsin, with Trypsin digesting proteins into smaller peptides by selectively cleaving proteins at the C-terminal side of lysine and arginine amino acid residues. Formic acid antigen retrieval breaks the cross-links formed during the tissue fixation process, which may have helped unmask antigens relevant to the DNMT antibodies trialled.

Consideration was given to conducting further optimisation of DNMT antibodies. However, it was decided not to conduct further trials on frozen tissue. This was due to a number of factors, one being the availability of frozen tissue. Fewer samples were available as frozen tissue, which would have reduced the size of the cohort considerably. Due to the extensive nature of the DNMT antibody optimisations already carried out on FFPE tissue, it was deemed that the use of precious frozen tissue to further this optimisation trial was unjustified, as potentially large amounts of tissue would be used in the optimisation alone, even before applying this to a cohort.

Further, one of the aims of this study was to determine the TDP43 pathology status of LMNs. The TDP43, 5mC and 5hmC antibodies were all optimised in FFPE, which allowed the same MN to be visualised in multiple sections. Therefore, the comparison between TDP43 pathology/DNA methylation status and DNMT immunopositivity would not be able to be determined within the same MN when using FFPE for the TDP43, 5mC and 5hmC and frozen tissue for DNMTs.

In conclusion, without adequate DNMT antibodies, detection of changes in levels of DNMTs within the brain/CNS remain challenging. Obtaining biopsies of tissue throughout the course of disease would also be of benefit, and help to counteract the end-stage nature of post mortem tissue, whilst providing a picture of DNA methylation changes in ALS over time.

4.5.10 Final remarks

In conclusion, the findings of this study suggest greater levels of DNA methylation and hydroxymethylation in ALS LMNs, with a loss of DNA methylation from LMN nuclei correlating with TDP43 pathology.

The study described in this chapter details a comprehensive pathological characterisation of DNA methylation in ALS in two motor areas severely affected in ALS (SC and MCx), and one extra-motor region (AFCx) that shows signs of neurodegeneration during ALS, and can therefore be indicative of 'early' and 'disease progression' signs of disease.

Samples sizes were reasonable for the scale of the project (SC n=30, AFCx n=61, MCx n=41), and are equal to or larger than cohorts previously assessed for DNA methylation using IHC by others (Roubroeks *et al.*, 2017). For the TDP43 pathology/DNA methylation study, the cohort size was large. However, a possible improvement on this study would be to replicate this analysis in a cohort derived from a different brain bank to ensure reproducibility.

Chapter 5: Genome-wide analysis of DNA methylation in ALS lower motor neurones

5.1 Introduction

The pathological studies carried out in chapter 4 suggest higher levels of DNA methylation and DNA hydroxymethylation are present in ALS MNs, with this effect not seen in glial cells or neurones of the motor and frontal cortices. This suggested a MN-specific phenomenon. In order to further understand DNA methylation changes in ALS MNs, MNs were extracted from the AH of human FFPE PM cervical spinal cord by laser capture microdissection, resulting in a neuronally enriched DNA sample, which has then undergone analysis by Illumina® Infinium® MethylationEPIC BeadChip. A flow diagram of the steps conducted in this chapter are seen in figure 5.1. A number of studies on DNA methylation in neurodegeneration used the HumanMethylation450 BeadChip array, which contained over 450,000 probes to measure DNA methylation levels at the single-gene level, the precursor to the MethylationEPIC array (Taskesen *et al.*, 2017). The HumanMethylaiton450 array was the precursor to the MethylationEPIC array, which contains over 850,000 probes. Hence, the decision was made to use the MethylationEPIC array to allow for potential comparisons between the dataset generated in this study compared to datasets available from previous studies originating from other laboratory groups.

It was found by IHC in section 4.4.5 that there were greater levels of methylation in sALS and C9ALS in residual LMNs of the SC compared to controls, an effect not seen in glial cells or in neurones of the motor or anterior frontal cortices. In section 5.4.2, using the Illumina® Infinium® MethylationEPIC BeadChip platform, greater levels of residual LMN DNA methylation in C9ALS cases are demonstrated, compared to control cases with intermediate (but not significantly different) levels of methylation in sALS. Concerning the previous chapter and this chapter, the overall increase in DNA methylation seen using the MethylationEPIC platform is validated by the IHC data in section 4.4.5.

Further, attempts to validate some of the individual differentially methylated genes identified on the Illumina platform using the Zymo Methyl-Seq™ Library will be also be discussed. To ensure a direct comparison, twelve of the sixteen samples used in the MethylationEPIC analysis were used. This was the maximum number of samples available for this study due to resources and time constraints (laser capture microdissection is very time consuming).

Total hits	
Control v non-control	*33,698 hits identified*

Filter for significant hits (p<0.05)

These hits were then sorted based on known gene names. Of these, 613 hits were found to have a gene name. 522 hits were classed as 'NA', with no known gene name. These hits were then identified using GeneCards. Of these 522, 119 could be identified, with 301 being novel transcripts with no gene name.

Therefore, the 613 hits, plus the 119 hits were taken forward, totalling 732.

Assessment of direction of DNA methylation change

402 hits hypermethylated
330 hypomethylated

Overlap with mRNA expression data

All 33,698 DNA methylation hits compared to all 22,511 mRNA expression hits to identify overlap between cohorts

Identification of significantly differentially methylated and expressed promoters overlap

The 732 significantly different DNA methylation promoter hits were compared to the mRNA hits to identify overlap between cohorts, with 279 overlapping hits.

Direction of change analysis on overlapping hits

The 279 overlapping hits were then assessed to see if the hypermethylated hits matched to downregulation, and hypomethylated hits corresponded to upregulation.

136 hits matched the expected direction of change.

Figure 5.1: Overview of analysis conducted for MethylationEPIC and mRNA expression analysis.

The aims of this chapter were as follows:

1. To validate the finding of increased global DNA methylation in residual LMNs using IHC (section 4.4.5)
2. To identify if there were any specific genes or pathways dysregulated in residual LMNs
3. To compare any gene-level dysregulation found to pre-existing mRNA expression data to establish if the DNA methylation changes correlate to the expected changes to mRNA expression

5.2 Hypotheses

1. Higher levels of DNA methylation will be observed in ALS groups compared to controls. To test this, the total number of methylated sites will be assessed and compared between each of three experimental groups: Controls, sALS and ALS associated with *C9orf72* mutations (C9ALS).
2. This increase in DNA methylation will be more marked in C9ALS cases compared to sALS cases.
3. This increase in DNA methylation in ALS cases will affect the neuronal mRNA transcriptome, and thus impact on the disease process. To test this hypothesis, promoters identified as significantly differentially methylated in ALS will be compared to pre-existing mRNA expression data.
4. There will be greater levels of methylation in the ALS groups, being most marked in the C9ALS cases.
5. A significant number of differentially methylated genes identified using the Illumina platform will be found to be differentially methylated using the Zymo Methyl-Seq™ Library.

5.3 Methods Overview

A more detailed description of methodologies employed in this chapter are described in chapter 2. In summary, MNs were stained with toludine blue and extracted from the anterior horn of human FFPE cervical spinal cord using laser capture microdissection (LCM). DNA was then extracted, quantified and bisulphite converted. Samples then underwent DNA methylation analysis using the Illumina® Infinium® MethylationEPIC BeadChip. Briefly, this entailed a series of steps including DNA amplification and fragmentation, followed by precipitation and resuspension. DNA was then hybridised to the BeadChip, stained and scanned. A summary of this process can be seen in fig.

During the scanning process, an IDAT file was created for each BeadChip, detailing the methylation status of DNA. IDAT files were then analysed using the package RnBeads in the programming language R to identify any significantly different DNA methylation expression between the three experimental groups; control, C9ALS and sALS. Significance values were set at $p<0.05$, unless otherwise stated. An overview of cases used in studies presented in this chapter can be seen in tables 5.1.

Table 5.1: Summary of case details used for MethylationEPIC array and BS-NGS studies

Key: Cases highlighted in red were used in the BS-NGS study

Case ID	Sex	Age	Disease Status
085/2007	F	59	Control
035/1996	F	87	Control
039/1997	M	53	Control
005/2007	M	63	Control
012/2007	M	63	Control
087/1992	F	75	sALS
200/1997	M	73	sALS
034/2005	M	63	sALS
094/2006	M	72	sALS
099/2009	M	80	sALS
041/2004	M	64	C9ALS
053/1996	F	65	C9ALS
073/2007	F	65	C9ALS
098/2002	F	64	C9ALS
045/2006	M	46	C9ALS
039/2011	M	72	C9ALS

Averages (including standard deviations)
Control = 65.0 (SD=13.0)
sALS = 72.6 (SD=6.2)
C9ALS = 62.7 (SD=8.7)

Age T-Tests
Control v sALS p= 0.27
Control v C9ALS p= 0.73
sALS v C9ALS p= 0.06

BS-NGS averages (including standard deviations)
Control = 65.5 (SD=14.9)
sALS = 70.8 (SD=5.3)
C9ALS = 66.3 (SD=3.9)

BS-NGS age T-Tests
Control v sALS p= 0.53
Control v C9ALS p= 0.93
sALS v C9ALS p= 0.22

5.4 Results

5.4.1 RnBeads quality control

Quality control was conducted and summarised in table 5.2. The Greedycut algorithm was used to remove CpGs/samples containing the largest fractions of measurements deemed unreliable. This is calculated based on detection p-values and/or read coverage. The Greedycut algorithm removed 496,306 unreliable probes. No samples were removed. QQ plots were used to assess normal distribution of the samples, which was found to be the case (see appendix X). The Greedycut algorithm was chosen as part of the quality control process for a number of reasons. At the time of analysis, there were few studies published using the MethylationEPIC array, with no prior knowledge of conducting analysis on this array within the department. When deciding which method was most appropriate, considerations were given to a number of factors, such as the availability of published literature detailing the analysis procedure, the level of expertise needed to conduct the analysis and the level of support available from various sources when carrying out the analysis. Specific to the Greedycut algorithm, this was chosen as it was part of the standard RnBeads analysis protocol recommended for use with the MethylationEPIC array. The Greedycut algorithm also provided a balance between filtering the data to an adequate level and conducting the analysis in a reasonable timeframe. Firstly, it was part of the standard RnBeads analysis protocol for the MethylationEPIC array. Samples were deemed to be unreliable by the Greedycut algorithm based on the read coverage and detection p-values, with Greedycut filtering out probes with a detection p-value of >0.05. The removal of 496,306 probes, while high, is reflective of the DNA quality extracted from the FFPE tissue, with the bisulphite conversion process degrading the DNA further. While these probes could have been included in further analysis, it was deemed that a smaller amount of better quality and more reliable results was preferable over a larger amount of less representative results.

Table 5.2: Summary of the quality control measurements for the MethylationEPIC samples, analysed using RnBeads.

Quality control assessment	Details	Comments/interpretation
Bisulphite conversion I and Bisulphite II conversion	• Sample dependent control • Samples showing higher signal intensities are normally removed	• Some samples show higher than expected intensities for both red and green probes • Samples showing high intensities were not removed, as the intensities observed are expected when using low quality DNA, such as that from FFPE tissue
Hybridisation	• Sample independent control • Synthetic targets are present in the hybridisation buffer at three concentrations: low, medium and high, resulting in three intensity intervals being observed	• Green channel shows expected intensity intervals • Red channel shows similar pattern, but at lower intensities, suggesting less effective hybridisation in this channel
Extension	• Sample independent control • Monitors the extension efficiency of A, T, C, and G nucleotides	• No samples show significantly different intensities when compared to other samples in the cohort • Low signal intensities detected for the t green channel. However, the intensities are comparable between the samples
Staining	• Sample independent control • Monitors efficiency of staining step • Independent of hybridisation and extension steps • Expected signal intensities should be approximately 1,000-2,000	• No samples show significantly different intensities when compared to other samples in the cohort • All intensities match the expected patterns for both red and green channels • Low intensity staining, suggesting weak staining, detected for biotin (5K) red channel, DNP (high) green channel and biotin (high) red
Negative control	• Sample dependent control • Monitors signal at bisulphite-converted sequences not containing CpGs • Used to estimate the overall background of the signals in both red and green channels • The detection p value for each probe are estimated based on the intensities of the negative control probes • Negative control probe intensities should be comparable, and ideally distributed around a low mean of approximately 1,000 between the channels • Deviations from this intensity indicated samples of lower quality	• Higher than ideal intensities observed for some of the samples • However, these samples were not removed from analysis, as the intensities observed are reflective of the input DNA quality

5.4.2 Comparison of total number of methylated regions

The total number of methylated DNA probes was calculated and compared between the three experimental groups (n=6 in each group, total n=18). The C9ALS cases had the highest number of methylated probes (mean of 419,966 probes methylated, SD=98,783), followed by sALS cases (mean of 309,017 probes methylated, SD=57,157), with control cases displaying the lowest number of methylated probes (mean of 263,207 methylated probes, SD=38,856). A significant difference (p=0.011) was detected between control and C9ALS groups in the number of probes methylated revealing a greater number of probes in the latter (figure 5.2). Significant differences were not detected when comparing control and sALS groups, or sALS and C9ALS groups, which were at intermediate levels between the other two. Post-hoc power analysis was conducted using G*Power v3.1.9.4 (Faul *et al.*, 2007) to assess the achieved power of the test, which was 0.87. In a perfectly powered study, a figure of 1.0 is found. Analysis was also conducted to assess the ideal sample size, which was 24 samples, compared to the 16 samples that were used in the array. If 24 samples were used, the expected power of the experiment was 0.96. In conclusion, this study was slightly underpowered. This was unavoidable due to the time consuming nature of LCM, and the finite length of this project.

Figure 5.2: A significant difference in methylation levels is present between control and C9ALS in MethylationEPIC array data (mean ±SD).
A significant difference in number of methylated sites was detected between control and C9ALS groups (p=0.011).
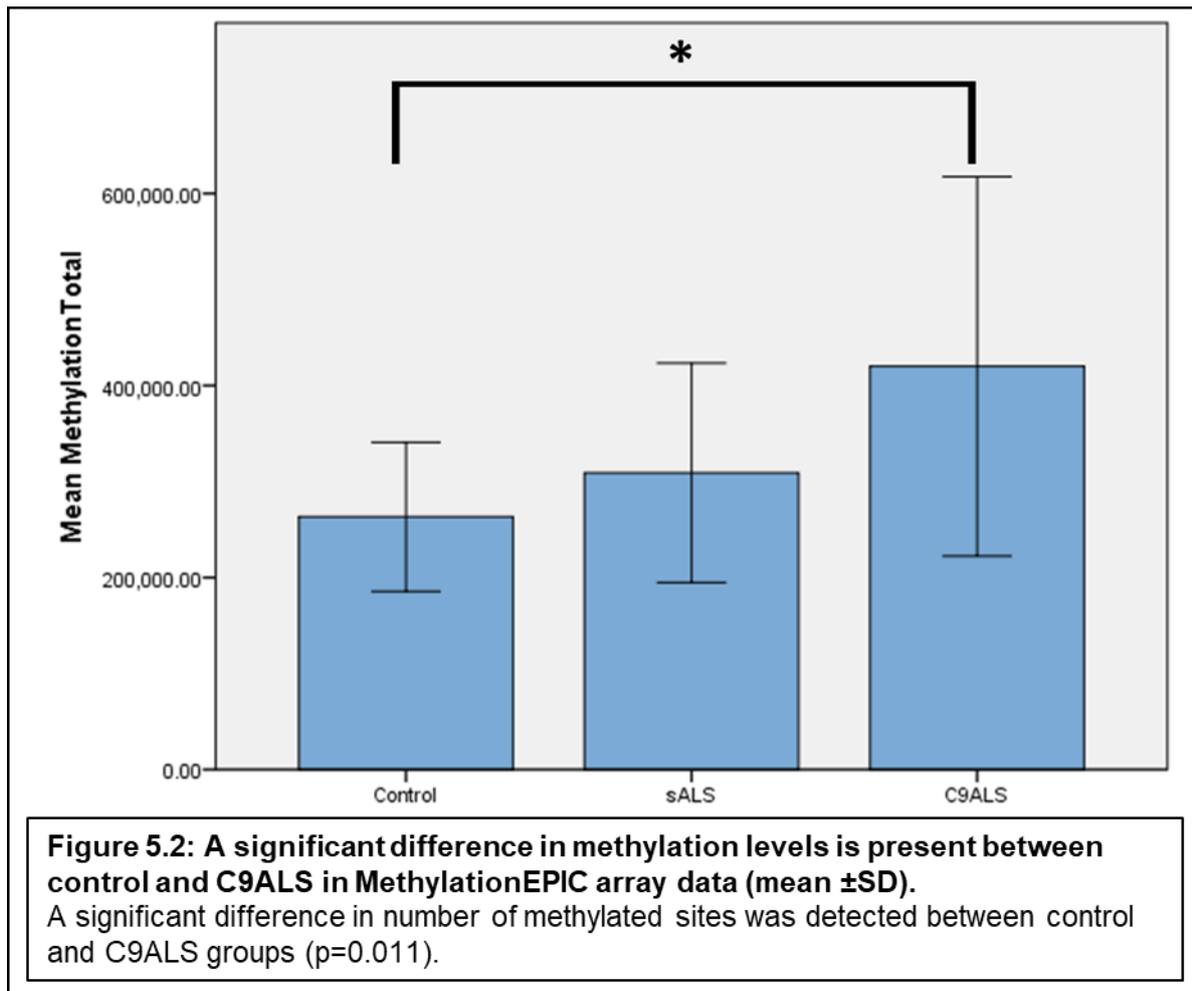
5.4.3 ALS has significantly differentially methylated promoter regions

Analysis presented in the previous section found increased numbers of methylated sites in C9ALS cases compared to controls, with intermediate, but not significant levels in sALS cases. The next step was to look at the significance of this in specific genes. This is a more focused analysis of gene promoters, compared to the analysis conducted in the previous section, which identifies all methylated sites, irrespective of location or significance in the genome. The analysis presented in this section focuses only on promoters of known genes.

RnBeads analysis documents were first converted from csv files to excel files. The sort function was used to order the hits according to 'comb.p.val'. This is the normalisation corrected p significance value. All hits where p<0.05 were taken forward for further analysis. The significant hits were then sorted according to the official gene symbol. The control versus disease (sALS plus C9ALS pooled) analysis indicated 613 hits were significantly differentially methylated in ALS versus controls. A further 522 hits were identified which had no gene name, and were therefore recorded as 'NA'. Searches using GeneCards

(https://www.genecards.org/) were conducted on these 522 hits to identify if any gene name could be found. A further 119 hits were identified, with 403 identified as novel transcripts with no known gene name or associated information. For this reason, only the 613 originally identified and the 119 subsequently identified as genes were taken forward for further analysis, totalling 732, with the unidentified 403 removed from further analysis.

Of the 732 total hits identified, 402 promoters were identified as significantly hypermethylated in ALS, with 330 hits identified as significantly hypomethylated in ALS (figure 5.3, section b). The degree of dysmethylation was more profound in the hypermethylated promoters, with a mean increase in methylation of 21.8% in ALS compared to controls, with hypomethylated promoters showing a mean decrease in methylation of 9.4% in ALS versus controls.

Figure 5.3: MethylationEPIC significantly differentially methylated promoters are mostly hypermethylated and protein coding.
(A) Of the 732 promoters, 378 were identified as protein coding, 202 as pseudogenes and 122 identified as RNA genes.
(B) Of 732 significantly differentially methylated promoters (p<0.05), 402 were hypermethylated, and 330 were hypomethylated.
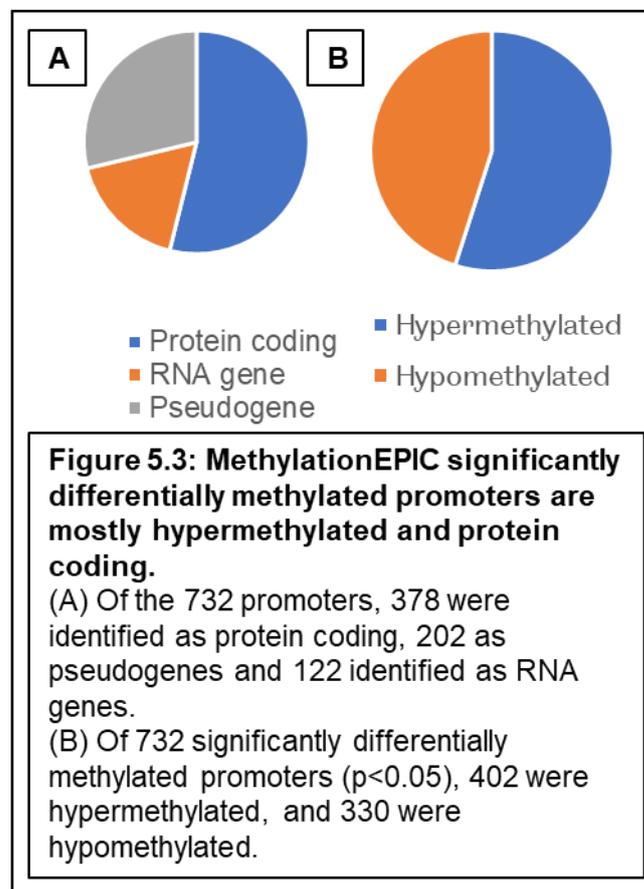
5.4.4 Promoter type analysis

The human gene database GeneCards (https://www.genecards.org/) was used to identify the promoter types, based on the 732 hits identified from the RnBeads analysis. Of the 732 promoters, 378 were identified as protein coding, 202 as pseudogenes and 122 identified as RNA genes (figure 5.3, section a).

5.4.5 Promoter GO enrichment and pathway analysis

Gene ontology (GO) analysis was conducted as part of the RnBeads analysis. The GO analysis conducted used a hypergeometric test, which addresses the hierarchical structure of the ontology. This corrects for the possibility of multiple pathways being annotated by the same genes being covered by parent terms (Falcon and Gentleman, 2007). The GO pathways identified were then split into promoters that were hypermethylated in ALS, and those that were hypomethylated (table 5.3). Benjamini-Hochberg corrected p-values were calculated, with only those that had a p-value of <0.01 considered. A number of pathways involved with neurodegeneration were found. For hypermethylation, most hits implicate RNA processing, nucleic acid processing and RNA splicing. This GO analysis was used as it was part of the RNBeads standard pipeline, and therefore makes for easier comparison between other datasets analysed using the RnBeads method, and the method controls for genes found in multiple pathways being covered by a parent term.

Table 5.3: Gene ontology enrichment analysis for significantly differentially hypermethylated and hypomethylated promoters identified in MethylationEPIC analysis.

GOMFID	p-value	Term
Hypermethylated		
GO:0000395	0	mRNA 5'-splice site recognition
GO:0000245	0	spliceosomal complex assembly
GO:0071826	1.00E-04	ribonucleoprotein complex subunit organization
GO:0022613	1.00E-04	ribonucleoprotein complex biogenesis
GO:0000377	2.00E-04	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile
GO:0098884	4.00E-04	postsynaptic neurotransmitter receptor internalization
GO:0008380	5.00E-04	RNA splicing
GO:0006397	9.00E-04	mRNA processing
GO:0110077	0.0016	vesicle-mediated intercellular transport
GO:1904906	0.0016	positive regulation of endothelial cell-matrix adhesion via fibronectin
GO:0070194	0.0032	synaptonemal complex disassembly
GO:0071840	0.0041	cellular component organization or biogenesis
GO:1905180	0.0048	positive regulation of cardiac muscle tissue regeneration
GO:0043312	0.0061	neutrophil degranulation
GO:0106027	0.0063	neuron projection organization
GO:0034622	0.0063	cellular protein-containing complex assembly
GO:0090673	0.0064	endothelial cell-matrix adhesion
GO:1903416	0.0064	response to glycoside
GO:0098657	0.0067	import into cell
GO:0042119	0.0067	neutrophil activation
GO:0030100	0.0074	regulation of endocytosis
GO:0000320	0.008	re-entry into mitotic cell cycle
GO:0016321	0.008	female meiosis chromosome segregation
GO:0045321	0.0081	leukocyte activation
GO:0031503	0.0086	protein-containing complex localization
GO:0002275	0.0089	myeloid cell activation involved in immune response
GO:0002444	0.0094	myeloid leukocyte mediated immunity
GO:0097500	0.0096	receptor localization to non-motile cilium
GO:0099505	0.0096	regulation of presynaptic membrane potential
GO:0022607	0.01	cellular component assembly
Hypomethylated		
GO:0097274	0.0011	Urea homeostasis
GO:0071947	0.0027	Protein deubiquitination involved in ubiquitin-dependent protein catabolic process
GO:0023041	0.0037	Neuronal signal transduction
GO:0015867	0.0048	ATP transport
GO:0036444	0.0053	Calcium import into the mitochondrion
GO:0048681	0.0064	Negative regulation of axon regeneration
GO:0015865	0.0075	Purine nucleotide transport
GO:0022038	0.0085	Corpus callosum development

5.4.6 Comparison of MethylationEPIC promoters with gene expression data

MethylationEPIC promoters identified as significantly differentially methylated were compared to pre-existing mRNA expression data (Highley *et al.*, 2014) to determine the relationship between the two datasets. Promoters identified as hypermethylated were compared to mRNA expression hits identified as downregulated, with promoters identified as hypomethylated being compared to mRNA expression data identified as upregulated.

In order to establish if the mRNA expression and MethylationEPIC platforms were compatible, with overlapping targets, the probes on the arrays (promoters/genes) were assessed to see the level of overlap, and ultimately, the suitability of comparing data from the two platforms. First, all of the gene promoters assessed for methylation (regardless of p-value) in the MethylationEPIC dataset were compared to all the hits in the mRNA expression analysis to determine if there was any overlap in the probes assessed in both systems. This determined that there was an overlap of 14,193 hits between the systems (figure 5.4, section a). The 732 promoters found to be significantly differently methylated in the MethylationEPIC array were compared to the entire mRNA platform (regardless of mRNA p-value). This was carried out to establish if the expected direction was reflected in the datasets (figure 5.4, section b). Of these, 279 were found to be present in the mRNA expression dataset. These 279 hits were then assessed to determine if they matched the expected direction of change (hypermethylation with downregulation, hypomethylation with upregulation). In total, 136 hits matched the expected direction of change.

Figure 5.4: Comparison of overlapping hits between significantly differentially methylated promoters from MethylationEPIC and significantly different mRNA expression.

(A) All hits identified in the MethylationEPIC array were compared to all the hits in the mRNA expression array to identify overlap between the cohorts. (B) Comparison of MethylationEPIC to mRNA expression data identified 279 overlapping hits.

5.4.7 GO analysis comparison between MethylationEPIC and mRNA expression

In the paper by Highley *et al.*, 2014, gene ontology analysis was conducted using DAVID, and is summarised in table 5.4. In both analyses on genes with increased expression and decreased expression in ALS, RNA splicing was implicated. This is complementary to the findings from the MethylationEPIC dataset, which found RNA metabolism dysregulation (table 5.3).

Table 5.4: Summary of case details from Highley *et al*. 2014.
Key: Cases also used in MethylationEPIC studies highlighted in red.

Case ID	Gender	Age	Disease Status
056/1990	M	51	Control
080/1992	F	62	Control
114/1993	M	78	Control
129/1994	M	63	Control
039/1997	M	53	Control
005/2007	M	63	Control
088/1989	F	59	sALS
193/1990	M	70	sALS
088/1996	M	68	sALS
155/1996	F	58	C9ALS
045/2006	M	46	C9ALS
081/2009	M	60	C9ALS

Gene ontology analysis of mRNA expression data
Used with permission from Highley *et al*. 2014.

	Motor neurones	
	Term	P
Differentially spliced genes	Ribonucleotide binding	7.14×10^{-11}
	Cell projection organization	7.89×10^{-07}
	Actin cytoskeleton organization	9.53×10^{-06}
	Protein localization	2.30×10^{-06}
	Macromolecule catabolic process	6.39×10^{-05}
Genes with increased expression	Positive regulation of nitrogen compound metabolic process	0.0003
	Cell adhesion	0.0097
	RNA splicing	0.0475
Genes with decreased expression	Cell morphogenesis	0.0045
	Striated muscle adaptation	0.0039
	Steroid biosynthetic process	0.0015
	RNA splicing	0.0046
	Cellular macromolecule catabolic process	0.0186

5.4.8 PANTHER pathway analysis comparison

PANTHER pathway analysis was carried out (http://pantherdb.org/) on both the MethylationEPIC datasets (separate analysis for hypermethylation and hypomethylation) and mRNA expression dataset (separate analysis for upregulation and downregulation) to identify pathways associated with the gene promoters/mRNA expression identified as significantly differentially methylated/significantly differentially expressed in ALS.

In total, PANTHER pathway analysis has 177 pathways in its repositories. For the hypermethylated MethylationEPIC dataset, 18 pathways were identified, with 17 of these pathways also picked up in the mRNA downregulated expression pathway analysis (table 5.5). For the hypomethylated MethylationEPIC dataset, 51 pathways were identified as being affected. Of these, 43 pathways were also present in the mRNA upregulated expression pathway analysis (table 5.6). This indicates a high level of overlap in the pathways found to be implicated in each dataset. Other pathway analyses repositories were considered but not used, as PANTHER was found to be the most up to date (last revised April 2018) compared to other repositories such as DAVID, which was last revised October 2016. PANTHER has been used in a number of publications using the MethylationEPIC's precursor, the HumanMethylation450 BeadChip array. In order to facilitate potential dataset comparisons in the future, PANTHER was deemed to be an appropriate pathway analysis tool to facilitate any future analyses.

For the hypermethylation MethylationEPIC dataset, multiple pathways involved in neurodegeneration were implicated (AD-presenilin pathway, HD), as well as inflammation (B cell activation, cadherin, interleukin and inflammation mediated by chemokine and cytokine signalling) and cell signalling (Notch signalling). For the hypomethylation MethylationEPIC dataset, pathways involved in neurodegeneration (AD-presenilin pathway, HD and Parkinson's disease (PD)), inflammation (B cell and T cell activation, TGF beta signalling) and cell death (apoptosis signalling, p53 pathway and TGF beta signalling) were found.

Together, the findings for both hypermethylation and hypomethylation implicate pathways previously associated with ALS, and suggest a key role for RNA metabolism dysmethylation and dysregulation.

Table 5.5: PANTHER pathway analysis of EPIC hypermethylated promoters.

This was compared to PANTHER pathway analysis on mRNA expression data, with those pathways appearing in both datasets highlighted in red.

Pathway
5HT3 type receptor mediated signalling pathway (P04375)
Alzheimer disease-presenilin pathway (P00004)
Angiogenesis (P00005)
B cell activation (P00010)
Cadherin signalling pathway (P00012)
DNA replication (P00017)
EGF receptor signalling pathway (P00018)
FGF signalling pathway (P00021)
Gonadotropin-releasing hormone receptor pathway (P06664)
Huntington disease (P00029)
Inflammation mediated by chemokine and cytokine signalling pathway (P00031)
Integrin signalling pathway (P00034)
Interleukin signalling pathway (P00036)
Nicotinic acetylcholine receptor signalling pathway (P00044)
Notch signalling pathway (P00045)
PDGF signalling pathway (P00047)
VEGF signalling pathway (P00056)
Wnt signalling pathway (P00057)

Table 5.6: PANTHER pathway analysis of EPIC hypomethylated promoters.
This was compared to PANTHER pathway analysis on mRNA expression data, with those pathways appearing in both datasets highlighted in red.

Pathway	
5-Hydroxytryptamine degradation (P04372)	Hypoxia response via HIF activation (P00030)
5HT1 type receptor mediated signalling pathway (P04373)	Inflammation mediated by chemokine and cytokine signalling pathway (P00031)
ATP synthesis (P02721)	Insulin/IGF pathway-mitogen activated protein kinase kinase/MAP kinase cascade (P00032)
Alzheimer disease-presenilin pathway (P00004)	Insulin/IGF pathway-protein kinase B signalling cascade (P00033)
Angiogenesis (P00005)	Integrin signalling pathway (P00034)
Angiotensin II-stimulated signalling through G proteins and beta-arrestin (P05911)	Interleukin signalling pathway (P00036)
Apoptosis signalling pathway (P00006)	Muscarinic acetylcholine receptor 1 and 3 signalling pathway (P00042)
Axon guidance mediated by netrin (P00009)	N-acetylglucosamine metabolism (P02756)
Axon guidance mediated by semaphorins (P00007)	PDGF signalling pathway (P00047)
B cell activation (P00010)	PI3 kinase pathway (P00048)
Blood coagulation (P00011)	Parkinson disease (P00049)
CCKR signalling map (P06959)	Pyridoxal-5-phosphate biosynthesis (P02759)
Cadherin signalling pathway (P00012)	Ras Pathway (P04393)
Circadian clock system (P00015)	Salvage pyrimidine ribonucleotides (P02775)
Cytoskeletal regulation by Rho GTPase (P00016)	Serine glycine biosynthesis (P02776)
EGF receptor signalling pathway (P00018)	T cell activation (P00053)
Endothelin signalling pathway (P00019)	TGF-beta signalling pathway (P00052)
FGF signalling pathway (P00021)	Toll receptor signalling pathway (P00054)
General transcription by RNA polymerase I (P00022)	Ubiquitin proteasome pathway (P00060)
Gonadotropin-releasing hormone receptor pathway (P06664)	VEGF signalling pathway (P00056)
Hedgehog signalling pathway (P00025)	Vasopressin synthesis (P04395)
Heme biosynthesis (P02746)	Vitamin B6 metabolism (P02787)
Heterotrimeric G-protein signalling pathway-Gi alpha and Gs alpha mediated pathway (P00026)	Wnt signalling pathway (P00057)
Heterotrimeric G-protein signalling pathway-Gq alpha and Go alpha mediated pathway (P00027)	p53 pathway feedback loops 2 (P04398)
Histamine H1 receptor mediated signalling pathway (P04385)	p53 pathway (P00059)
Huntington disease (P00029)	

5.4.9 Comparison of MethylationEPIC and mRNA expression at the single gene level

The 136 hits identified in section 5.4.6 were further assessed. Of these hits, 29 were hypermethylated/downregulated. Of these 29 hits, 25 were protein coding. Analysis into the protein coding hit functions was conducted using GeneCards to assess if they matched the pathways identified (table 5.7). This identified genes involved in inflammation and immune response (*ITLN2, SIGLEC1, SH3RF1*), cell signalling mediation (*CD151, TSPAN31*), cell death (*HSPB7, RPS19BP1*) and RNA metabolism/transcriptional regulation (*RNF17, ZNF273, UBP1, NCOA4*).

This process was repeated with the hits that were hypomethylated/upregulated, of which 104 were protein coding (table 5.8). Genes involved in inflammation and immune response were present (*FYN, CCL18, BCL10, CNR2, ISG15, CLEC4A, CXCL1, GIMAP5, KLRG1* and *MSC*), along with cell signalling response (*FYN, NOXA1, RHEBL1, PIK3C2A, PREX2* and *ITGA8*). Hits for cell death were also present (*NCAM1, PCSK9, TEAD2, BCL10, COMMD4, SIDT2, RHEBL1, PIK3C2A, CYCS, TRAF2, SET* and *KLF4*), along with RNA metabolism/transcriptional regulation (*HBP1, ARNTL, TEAD2, ARID3B, NFE2L2, ZNF383, ZNF417, HNRNPUL1, CASZ1, SF3A1, PRPF4B, SET, KLF4* and *MSC*). Also of interest is the hypomethylation of *SERPINA1* (20.3% hypomethylation in ALS versus controls), which has previously been implicated in ALS (Ebbert *et al.*, 2017).

Table 5.7: Hypermethylated/downregulated protein coding genes identified in MethylationEPIC and mRNA expression analysis

Gene Symbol	Description
ITLN2	Potential role in **defence system** against pathogens
HSPB7	Encodes a small heat shock family B member that can heterodimerize with similar heat shock proteins Encoded protein may be a tumour suppressor in the **p53 pathway**
AQP7	Localizes to plasma membrane and allows movement of water, glycerol and urea across cell membranes Role in body energy homeostasis under conditions that promote lipid catabolism, giving rise to glycerol and free fatty acids
RNF17	**Regulation of transcriptional activity** of MYC (gene that upregulates transcript elongation of actively transcribed genes)
ZNF273	Encodes a nuclear protein with 13 C2H2-type zinc fingers and a KRAB domain, involved in **transcriptional regulation**
TSPAN31	Cell-surface protein with four hydrophobic domains, **mediates signal transduction events** that play a role in the regulation of cell development, activation, growth and motility
KIF2B	Involved in spindle assembly and chromosome movement, associated with KIF2B include Charcot-Marie-Tooth Disease, Type 4C
SIGLEC1	Macrophage-restricted adhesion molecule that **mediates sialic-acid dependent binding to lymphocytes, including granulocytes, monocytes, natural killer cells, B-cells and CD8 T-cells**
CDH9	Type II classical cadherin from the cadherin superfamily, mediates calcium-dependent cell-cell adhesion
RPS19BP1	Direct regulator of SIRT1. Enhances SIRT1-mediated deacetylation of p53/TP53, thereby participating in **inhibition of p53/TP53-mediated transcriptional activity**
TUBGCP2	Gamma-tubulin complex is necessary for microtubule nucleation at the centrosome
KLK6	The encoded protease may participate in the **cleavage of amyloid precursor protein and alpha-synuclein**, thus implicating this protease in Alzheimer's and Parkinson's disease, respectively, may be involved in regulation of axon outgrowth following spinal cord injury
NT5DC2	No information

Table 5.7: Hypermethylated/downregulated protein coding genes identified in MethylationEPIC and mRNA expression analysis continued…

Gene Symbol	Description
PRCP	Activator of the cell matrix-associated prekallikrein
SRI	Encodes a calcium-binding protein with multiple E-F hand domains that relocates from the cytoplasm to the sarcoplasmic reticulum in response to elevated calcium levels
RERG	Inhibits cell proliferation and binds GDP/GTP and possesses intrinsic GTPase activity
CD151	**Mediates signal transduction events** that play a role in the regulation of cell development, activation, growth and motility
CCDC158	No information
UBXN8	ATPase complex cofactor involved in endoplasmic reticulum-associated degradation (ERAD) for misfolded lumenal proteins, possibly by tethering VCP to the endoplasmic reticulum membrane
UBP1	**Transcriptional activator** in a promoter context-dependent manner
MBD5	This protein contains a PWWP domain (Pro-Trp-Trp-Pro motif), which consists of 100-150 amino acids and is found in numerous proteins that are involved in cell division, growth and differentiation and binds to chromatin Mutations in this gene cause an autosomal dominant type of cognitive disability. Haploinsufficiency of this gene is associated with a syndrome involving microcephaly, intellectual disabilities, severe speech impairment, and seizures
LRRTM4	Possible role in development and maintenance of vertebrate nervous system Exhibits strong synaptogenic activity, restricted to excitatory presynaptic differentiation
SH3RF1	Involved in protein sorting at the trans-Golgi network May act as a scaffold for the c-Jun N-terminal kinase signaling pathway, **facilitating the formation of a functional signaling module** Regulates the **differentiation of CD4(+) and CD8(+) T-cells and promotes T-helper 1 (Th1) cell differentiation** Regulates the activation of MAPK8/JNK1 and MAPK9/JNK2 in CD4(+) T-cells and the activation of MAPK8/JNK1 in CD8(+) T-cells
GPR162	No information
NCOA4	Encoded protein interacts with the androgen receptor in a ligand-dependent manner to enhance its **transcriptional activity**
IP6K2	This protein is likely responsible for the conversion of inositol hexakisphosphate (InsP6) to diphosphoinositol pentakisphosphate

Table 5.8: Hypomethylated/upregulated protein coding genes identified in MethylationEPIC and mRNA expression analysis

Gene symbol	Description
USP54	No information
NDST2	Essential bifunctional enzyme that catalyses both the N-deacetylation and the N-sulphation of glucosamine (GlcNAc) of the glycosaminoglycan in heparan sulphate. Modifies the GlcNAc-GlcA disaccharide repeating sugar backbone to make N-sulphated heparosan, a prerequisite substrate for later modifications in heparin biosynthesis. Plays a role in determining the extent and pattern of sulfation of heparan sulfate. Required for the exosomal release of SDCBP, CD63 and syndecan
RSAD1	May be a heme chaperone, appears to bind heme. Homologous bacterial proteins do not have oxygen-independent coproporphyrinogen-III oxidase activity (Probable). Binds 1 [4Fe-4S] cluster. The cluster is **coordinated with 3 cysteines and an exchangeable S-adenosyl-L-methionine**
TMC5	Probable ion channel
HBP1	**Transcriptional repressor** that binds to the promoter region of target genes. Plays a role in the regulation of the cell cycle and of the Wnt pathway. Binds preferentially to the sequence 5'-TTCATTCATTCA-3'. Binding to the H1F0 promoter is enhanced by interaction with RB1. Disrupts the interaction between DNA and TCF4.
ITPR2	Receptor for inositol 1,4,5-trisphosphate, a second messenger that mediates the release of intracellular calcium. This release is regulated by cAMP both dependently and independently of PKA
SLC12A4	Mediates the coupled movement of potassium and chloride ions across the plasma membrane
FYN	Non-receptor tyrosine-protein kinase that plays a role in many biological processes including regulation of cell growth and survival, cell adhesion, integrin-mediated signalling, cytoskeletal remodelling, cell motility, immune response and axon guidance Participates in the downstream **signalling pathways** that lead to **T-cell differentiation and proliferation following T-cell receptor (TCR) stimulation**
CCL18	**Attracts naive T-lymphocytes toward dendritic cells** and activated macrophages in lymph nodes, has chemotactic activity for naive T-cells, CD4+ and CD8+ T-cells and thus may play a role in both humoral and **cell-mediated immunity responses**
C7orf26	No information
SLC38A7	Mediates sodium-dependent transport of amino acids, preferentially L-glutamine
STARD6	May be involved in the intracellular transport of sterols or other lipids and cholesterol homeostasis
NCAM1	Cell adhesion molecule, transmembrane protein that are involved in the binding of a cell to another cell or to the extracellular matrix. They have roles in cell proliferation, differentiation, motility, trafficking, **apoptosis** and tissue architecture
PCSK9	Crucial role in the regulation of plasma cholesterol homeostasis **Regulates neuronal apoptosis** via modulation of LRP8/APOER2 levels and related anti-apoptotic signalling pathways
SGSM3	May play a cooperative role in NF2-mediated growth suppression of cells
DSEL	No information
RTN4IP1	This gene encodes a mitochondrial protein that interacts with reticulon 4, which is a potent **inhibitor of regeneration following spinal cord injury**
QRSL1	Allows the formation of correctly charged Gln-tRNA(Gln) through the transamidation of misacylated Glu-tRNA(Gln) in the mitochondria
SLC16A5	Proton-linked monocarboxylate transporter
AFM	May be involved in the transport of vitamin E across the blood-brain barrier

Gene symbol	Description
Table 5.8: Hypomethylated/upregulated protein coding genes identified in MethylationEPIC and mRNA expression analysis continued…	
BAAT	Involved in bile acid metabolism
PCIF1	**Cap-specific adenosine methyltransferase** that catalyses formation of N(6),2'-O-dimethyladenosine cap (m6A(m)) by methylating the adenosine at the second transcribed position of capped mRNAs. N6-methylation of m6A(m) promotes the translation of capped mRNAs
SLC1A5	Sodium-dependent amino acids transporter
HTR1F	G-protein coupled receptor for 5-hydroxytryptamine (serotonin)
ARNTL	**Transcriptional activator** which forms a core component of the circadian clock
TRAPPC6A	Role in vesicular transport during the biogenesis of melanosomes
TEAD2	**Transcription factor** which plays a key role in the **Hippo signaling pathway**
CDK5RAP2	Potential regulator of CDK5 activity via its interaction with CDK5R1, plays a role in neurogenesis
C11orf24	No information
VGLL4	May act as a specific coactivator for the mammalian TEF
THAP2	No information
BCL10	Involved in **adaptive immune response** and **promotes apoptosis**, pro-caspase-9 maturation and activation of NF-kappa-B via NIK and IKK
CNR2	Heterotrimeric G protein-coupled receptor for endocannabinoid 2-arachidonoylglycerol mediating inhibition of adenylate cyclase. May function in **inflammatory response**
RNF31	Involved in protein-DNA and protein-protein interactions
QTRT1	This gene encodes the catalytic subunit of tRNA-guanine transglycosylase
ARID3B	Roles in embryonic patterning, cell lineage gene regulation, cell cycle control, **transcriptional regulation** and possibly in chromatin structure modification
ISG15	Key role in the **innate immune response** to viral infection either via its conjugation to a target protein
NOXA1	Functions in the production of reactive oxygen species (ROS) which participate in a variety of biological processes including host defence, hormone biosynthesis, oxygen sensing and **signal transduction**
NFE2L2	**Transcription activator** that binds to antioxidant response (ARE) elements in the promoter regions of target genes
MIS12	Part of the MIS12 complex which is required for normal chromosome alignment

Gene symbol	Description
Table 5.8: Hypomethylated/upregulated protein coding genes identified in MethylationEPIC and mRNA expression analysis continued…	
Gene symbol	**Description**
WAPAL	Regulator of sister chromatid cohesion in mitosis which negatively regulates cohesin association with chromatin
METTL7A	Probable **methyltransferase**
NCAPG	Regulatory subunit of the condensin complex, a complex required for conversion of interphase chromatin into mitotic-like condense chromosomes
BSDC1	No information
COMMD4	May modulate activity of cullin-RING E3 ubiquitin ligase, **downregulates activation of NF-kappa-B**
TATDN2	Putative deoxyribonuclease
ZNF383	**Inhibits transcription** of some MAPK signalling pathway genes
RPS25	Catalyses protein synthesis
SIDT2	Mediates the translocation of RNA and DNA across the lysosomal membrane during RNA and DNA **autophagy**
CLEC4A	The encoded type 2 transmembrane protein may play a role in **inflammatory and immune response**
ISLR2	Required for axon extension during neural development
UPP2	Catalyses the reversible phosphorylytic cleavage of uridine and deoxyuridine to uracil and ribose- or deoxyribose-1-phosphate
ZNF417	May be involved in **transcriptional regulation**
CXCL1	This protein plays a role in **inflammation** and as a chemoattractant for neutrophils
NT5C1A	Dephosphorylates the 5' and 2'(3')-phosphates of deoxyribonucleotides
ARF1	GTP-binding protein involved in protein trafficking among different compartments
HNRNPUL1	Acts as a basic **transcriptional regulator, plays a role in mRNA processing and transport**
OR7C1	Olfactory receptor
RBBP4	Core histone-binding subunit that may target chromatin assembly factors, **chromatin remodelling factors and histone deacetylases** to their histone substrates in a manner that is regulated by nucleosomal DNA
CASZ1	The protein encoded by this gene is a **zinc finger transcription factor**

Table 5.8: Hypomethylated/upregulated protein coding genes identified in MethylationEPIC and mRNA expression analysis continued…	
Gene symbol	**Description**
TSEN15	Non-catalytic subunit of the **tRNA-splicing** endonuclease complex
GIMAP5	Plays a role **in T lymphocyte development** and the optimal generation of CD4/CD8 double-positive thymocytes
RHEBL1	Binds GTP and exhibits intrinsic GTPase activity. May activate **NF-kappa-B-mediated gene transcription**. Promotes **signal transduction through MTOR**
KLRG1	Plays an **inhibitory role on natural killer (NK) cells and T-cell functions** upon binding to their non-MHC ligands
M6PR	Transport of phosphorylated lysosomal enzymes from the Golgi complex and the cell surface to lysosomes
TNR	The encoded protein is restricted to the central nervous system. The protein may play a role in neurite outgrowth, neural cell adhesion and modulation of sodium channel function
DIAPH3	Involved in actin remodelling and regulate cell movement and adhesion
IRS1	May mediate the control of various cellular processes by insulin
PIK3C2A	Roles in **signalling pathways** involved in cell proliferation, oncogenic transformation, **cell survival**, cell migration, and intracellular protein trafficking
ADAM11	Probable ligand for integrin in the brain
TMEM186	No information
PREX2	Functions as a RAC1 guanine nucleotide exchange factor, role in **insulin-signalling** pathways
SF3A1	This gene encodes a subunit of the splicing factor 3a protein complex, critical role in **spliceosome assembly and pre-mRNA splicing**
EHMT1	The protein encoded by this gene is a **histone methyltransferase** that methylates the lysine-9 position of histone H3
CYCS	Electron carrier protein, plays a role in **apoptosis**
MSC	**Transcription repressor** capable of inhibiting the transactivation capability of TCF3/E47. May play a role in **regulating antigen-dependent B-cell differentiation**
PRPF4B	Has a role in **pre-mRNA splicing**
RBP4	Delivers retinol from the liver stores to the peripheral tissues
ANKRD45	No information
CNTN4	Mediates cell surface interactions during nervous system development

Table 5.8: Hypomethylated/upregulated protein coding genes identified in MethylationEPIC and mRNA expression analysis continued…	
Gene symbol	**Description**
HMCN1	Promotes cleavage furrow maturation during cytokinesis in preimplantation embryos
C9orf96	No information
SFMBT1	Histone-binding protein, which is part of various corepressor complexes
ITGA8	Mediates numerous cellular processes including cell adhesion, cytoskeletal rearrangement, and **activation of cell signalling pathways**, neuronal receptor for TNC: it mediates cell-cell interactions and regulates neurite outgrowth of sensory and motor neurones
EFCAB5	No information
SMAGP	May play a role in epithelial cell-cell contacts
PSAT1	Catalyses the reversible conversion of 3-phosphohydroxypyruvate to phosphoserine and of 3-hydroxy-2-oxo-4-phosphonooxybutanoate to phosphohydroxythreonine
TOM1	May be involved in intracellular trafficking. Probable association with membranes
TRAF2	Regulates activation of NF-kappa-B and JNK and plays a central role in the regulation of **cell survival and apoptosis**
LOXL3	Mediates the oxidation of peptidyl lysine residues to allysine in target proteins
AHCTF1	Required for the assembly of a functional nuclear pore complex (NPC) on the surface of chromosomes as nuclei form at the end of mitosis
LRRC55	Auxiliary protein of the large-conductance, voltage and calcium-activated potassium channel
SET	The protein encoded by this gene inhibits acetylation of nucleosomes, especially histone H4, by histone acetylases, involved in **apoptosis, transcription**, nucleosome assembly and histone chaperoning
VAV1	Couples tyrosine kinase signals with the activation of the Rho/Rac GTPases, thus leading to cell differentiation and/or proliferation
KLF4	**Transcription factor**; can act both as activator and as repressor, contributes to the **down-regulation of p53/TP53 transcription**
PES1	Component of the PeBoW complex, which is required for maturation of 28S and 5.8S ribosomal RNAs and formation of the 60S ribosome
OR8A1	Odorant receptor
RXFP1	Receptor for relaxins
TCN2	This gene encodes a member of the vitamin B12-binding protein family and is a transport protein
SEMA3D	Induces the collapse and paralysis of neuronal growth cones. Could potentially act as repulsive cues toward specific neuronal populations
SERPINA1	Inhibitor of serine proteases, **previously implicated in ALS**
ACSM5	No information
CTNNA3	May be involved in formation of stretch-resistant cell-cell adhesion complexes
OR52E2	Odorant receptor

5.4.10 Horvath epigenetic clock analysis

The Horvath epigenetic clock algorithm was originally created for use in conjunction with the 450k array, but can be modified for use with MethylationEPIC arrays (Horvath and Raj, 2018). It relies on a series of probes related to biological ageing, originally identified in the oncology field. Due to the nature of using FFPE tissue, not all the probes necessary for calculating the Horvath were present in the dataset. Of the 28, 587 probes used to calculate DNA methylation age, 15,711 (55%) of probes were missing. This results in a less accurate determination of epigenetic age. Therefore, results of this calculation were discounted.

5.4.11 Quality control for BS-NGS

Initial quality control analysis was conducted using the FastQC programme on Galaxy (https://usegalaxy.org/). The programme Bismark (https://www.bioinformatics.babraham.ac.uk/projects/bismark/) was then used to align the methylation sequencing data to the human genome sequence (hg38), generating BED files which were then analysed using RnBeads. FastQC analysis indicated that all samples sequenced were classed as good, as represented in figure 5.5. However, Bismark sequence alignment to hg38 was poor for all samples. This is demonstrated in table 5.9. The column labelled 'sequence without alignment' (highlighted in red) gives an indication of how many sequences could not be aligned using the reference human genome sequence (hg38). As seen in table 5.9, large percentages of sequences were not aligned. As poor alignment to hg38 was observed, a consideration to alignment with hg37 was given, as there was a possibility that the data could be better aligned to this version of the human genome sequence. However, as all the samples showed poor alignment, and the alignment percentages were small, this suggests earlier issues at stages such as library formation and sequencing, possibly due to the input DNA not being of sufficient quality to perform sequencing, meaning that any analysis carried out on this dataset would be unreliable. Due to this reason, alignment to hg37 was not attempted.

Figure 5.5: FastQC analysis of BS-NGS data and Bismark sequence alignment.
(A) Example of quality of sequencing reads. All samples had quality scores in the green area, indicating that they were of sufficient quality. (B) Sequences were aligned to the hg38, with the 'unique alignments' indicating the mapping percentage of the BS-NGS reads to hg38. Low to medium levels of mapping were observed.

Table 5.9: BS-NGS data was poorly aligned to the reference genome								
Case ID	Disease Status	Total sequences analysed	Alignments with unique mapping		Sequences without alignments		Sequences not uniquely mapped	
LP005/2007	Control	9621492	2184926	22.7%	6837937	71.1%	598629	6.2%
LP035/1996	Control	11999802	2034504	17.0%	9517151	79.3%	448147	3.7%
LP039/1997	Control	28067459	1801532	6.4%	25303917	90.2%	962010	3.4%
LP085/2007	Control	21211712	537453	2.5%	20496917	96.6%	177342	0.8%
LP034/2005	sALS	20169720	4796194	23.8%	13515381	67.0%	1858145	9.2%
LP087/1992	sALS	19861625	3706984	18.7%	14760762	74.3%	1393879	7.0%
LP094/2006	sALS	20126102	1602393	8.0%	16935324	84.1%	1588385	7.9%
LP200/1997	sALS	18392471	1010958	5.5%	16920888	92.0%	460625	2.5%
LP098/2002	C9ALS	20986118	786710	3.7%	19473015	92.8%	726393	3.5%
LP053/1996	C9ALS	17319189	2911648	16.8%	12765708	73.7%	1641833	9.5%
LP041/2004	C9ALS	39055677	13836430	35.4%	22251797	57.0%	2967450	7.6%
LP039/2011	C9ALS	20821661	4025889	19.3%	13872604	66.6%	2923168	14.0%

5.4.12 Total methylation sites for BS-NGS

Cytosine methylation was calculated using Bismark, with controls at 65.25%±6.2, sALS 71.3%±10.3 and C9ALS 70.7%±8.0. No significant difference in methylation levels were detected between any groups. Control v sALS (p=0.36), control v C9ALS (p=0.37) and sALS v C9ALS (p=0.94) (figure 5.6). Post-hoc power calculations were performed using G*power v3.1.9.4 (Faul *et al.*, 2007), indicating that the sample size of 12 used in this study was 0.71, indicating the study was underpowered. Increasing the study cohort to 20 samples would have increased the power of the study to 0.92. However, as LCM is a very time consuming procedure and BS-NGS is expensive, it was not feasible to increase the cohort size for this experiment.

Figure 5.6: No difference in methylation levels detected between groups for BS-NGS (mean ±SD)
Control v sALS (p=0.36), control v C9ALS (p=0.37) and sALS v C9ALS (p=0.94).
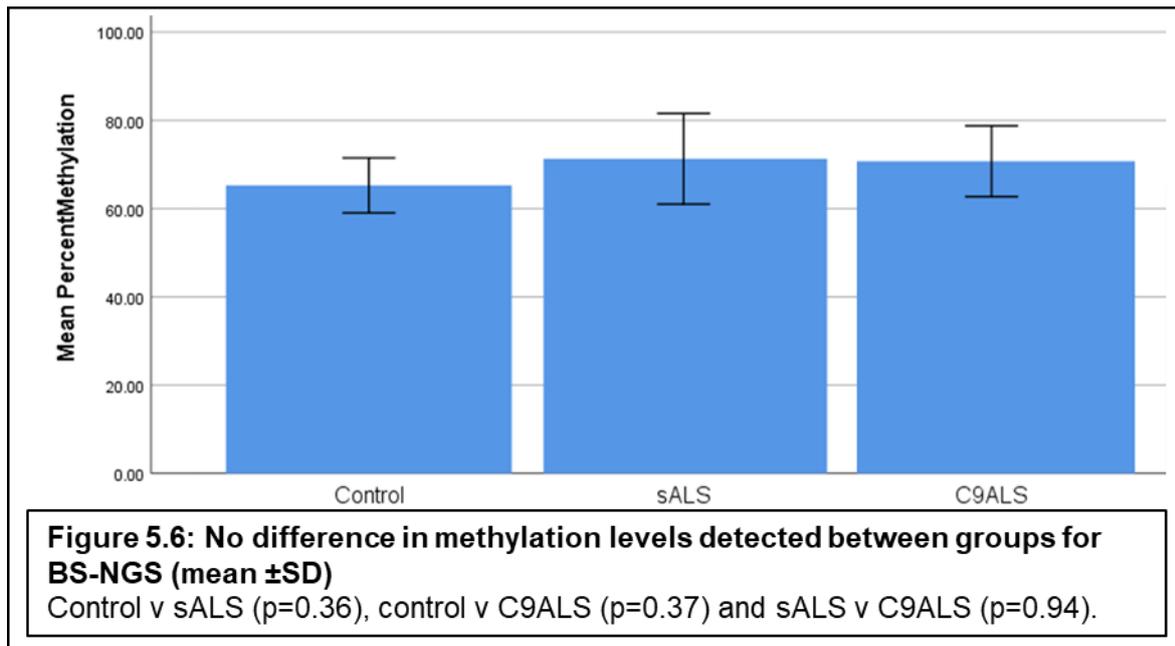
5.5 Discussion

In chapter 4, it was established that residual LMNs in sALS and C9ALS cases showed hypermethylation compared to controls. For C9ALS cases, this was validated in the findings presented in this chapter, with global methylation levels showing C9ALS being significantly more methylated than controls. There was an intermediate level of methylation for sALS cases that lay between that of these groups that was not statistically significant. Methylation was further analysed at the gene level, with 732 promoters being identified as significantly differentially methylated, 402 of these promoters were identified as hypermethylated (figure 5.3), with a mean increase in methylation of 21.8% in ALS. Combined with findings of changes at the single gene level, GO analysis and PANTHER pathway analysis also found dysregulation in RNA metabolism. These findings were validated using PANTHER pathway analysis, which showed strong overlap between the datasets at the pathway/GO level. This is in keeping with the existing literature which highlights the importance of focusing on the general mechanisms implicated in disease before continuing analysis at the single gene level (Pita-Juárez *et al.*, 2018). After implicated pathways had been established, comparisons of MethylationEPIC and mRNA expression data were made at the single gene level, which supported the findings of the pathway analysis, with hypermethylation of genes involved in inflammation/immune response, apoptosis/cell death and cell signalling (table 5.7).

5.5.1 Increased DNA methylation in C9ALS

The total number of methylated sites were calculated to determine a measurement of global DNA methylation levels. The findings support those seen in the previous chapter, with a

significant global increase in DNA methylation in C9ALS cases. sALS cases showed an intermediate, level of DNA methylation, but this was not significantly different to other groups. Overall, this finding indicates a global hypermethylation in C9ALS, with intermediate increases in sALS (although not significant). This is supported in the literature, with global increases in 5mC and 5hmC in sALS SC (Figueroa-Romero *et al.*, 2012) and in the cerebellum of C9FTD/ALS patients (Belzil, Nouri-mahdavi and Raoof, 2008). Increased DNA methylation has also been observed in *SOD1*-ALS (Coppedè *et al.*, 2017). Further, global levels of 5mC were increased in sALS, as well as C9ALS cases (which showed the highest increase in 5mC levels), as well as in spinocerebellar ataxia type 1 and type 2 in blood (Hamzeiy *et al.*, 2018). Together, these findings implicate global hypermethylation in neurodegenerative disease of MNs, and a strong case for hypermethylation specifically in ALS.

5.5.2 Changes in DNA methylation are present in ALS at the single gene level

Studies on the DNA methylation status of particular genes in ALS is sparse. Therefore, genes found to be downregulated at the RNA levels in datasets within published literature could be seen as potentially indicative of DNA methylation changes. Below will be considerations for genes that show similar changes in both mRNA expression and DNA methylation. For example, in both the mRNA expression dataset from Highley *et al.*, 2014 (from LMNs extracted using LCM from ALS cases) and the MethylationEPIC dataset described in this chapter, the gene *RNF17* was dysregulated. This gene is involved in transcriptional activity regulation, and has previously been found to be downregulated in the hippocampus of APP$_{swe}$/PS1dE9 double transgenic AD mice (Wan Nasri *et al.*, 2019). While this is not in ALS, it does suggest a potential dysregulated role in neurodegeneration, and is consistent with the findings discussed in this chapter.

When considering changes in methylation at the single gene level, little difference is observed between sALS and C9ALS, with few significantly differentially methylated promoters identified. This suggests that similar promoters are being affected in both sALS and C9ALS versus controls, but that the dysmethylation of these similar promoters is more severe for the C9ALS cases. GO analysis of separated lists for hypermethylation and hypomethylation in ALS revealed DNA methylation changes in genes associated with aberrant RNA metabolism, a process which is heavily implicated in ALS (Mackenzie, Rademakers and Neumann, 2010). This is discussed further in section 5.5.2.1.

5.5.2.1 GO analysis: RNA metabolism and ALS

GO analysis of the MethylationEPIC hypermethylated promoters in combined sALS and C9ALS by RnBeads heavily implicated RNA metabolism (section 5.4.5), with analysis implicating alternate RNA splicing in the mRNA expression dataset. Four pathways were found to include RNA splicing. RNA splicing is the removal of introns (non-coding sequences) from pre-mRNA, and the subsequent joining of exons (protein coding sequences), to allow for mRNA translation by ribosomes in the nucleus to take place. These are: mRNA 5' splice site recognition (GO:0000395), spliceosomal complex assembly (GO: 0000245), RNA splicing, via transesterification on reactions with bulged adenosine as nucleophile (GO:0000377) and RNA splicing (GP: 0006397). Two pathways were found to implicate ribonucleoprotein complexes. Ribonucleoprotein complexes are macromolecular complexes made up of RNA and RNA binding proteins. These are: ribonucleoprotein complex subunit organisation (GO: 0071826) and ribonucleoprotein complex biogenesis (GO:0022613). mRNA processing was also implicated (GO:0006397). Broadly, mRNA processing involves the processing of primary mRNA transcripts to create functional mRNAs. This process can include 3' cleavage and polyadenylation, 5' capping, mRNA splicing and RNA editing. Together, this suggests broad-reaching dysmethylation in RNA metabolism is present in ALS LMNs.

mRNA processing is made up of a number of steps, including: splicing, polyadenylation, editing, transport and translation, which are all highly dynamic processes. These processing steps are closely involved with RNA binding proteins (RBPs). RBPs are responsible for transcriptome maintenance by controlling the regulation of RNA processing and transport. They also modulate co-transcriptional and post-transcriptional transcript processing. RBPs bind RNA molecules at specific sequences/secondary structures to facilitate RNA processing in both the nucleus and cytoplasm (Nussbacher *et al.*, 2015). Finally, RBPs provide multi-functional roles by associating with many different protein complexes, and can therefore influence many processing steps of its RNA targets. With this wide reaching impact, and the dynamic nature of mRNA processing, even small changes in RBP expression can cause an amplification effect on subsequent RNA processing steps, such as expression, splicing and translation of RNA transcripts. This could lead to global dysfunction.

RBPs have previously been implicated in ALS, such as in aggregation, including TDP43 pathology, and sequestration by transcripts/abnormal proteins with pathological repeat expansions, such as that of *C9orf72* (Nussbacher *et al.*, 2019).

5.5.2.2 Overlap of MethylationEPIC with mRNA expression

The main aim of this chapter was to ascertain the changes in DNA methylation occurring in LMNs, and which pathways/processes were implicated. Secondly, considerations on elucidating the relationship between mRNA expression and DNA methylation changes were also sought, as well as corroboration with the findings of the previous chapter. A further aim was to compare MethylationEPIC with mRNA expression data at both the individual gene level, and at the level of cellular processes. This was to observe if any key promoters or pathways were found to be affected in both datasets.

While there was little relationship between DNA methylation and mRNA expression at the level of individual genes, a strong overlap between mRNA expression and MethylationEPIC PANTHER pathway analysis was found (section 5.4.8, table 5.5 and table 5.6). This implicated pathways and processes associated with inflammation (B cell and T cell activation, interleukin signalling and notch signalling), and cell death (apoptosis, p53 pathway and TGF beta signalling). These processes will be discussed in relation to ALS below.

5.5.2.3 Inflammation and immune response

PANTHER analysis of hypermethylated hits identified B cell activation, cadherin signalling, interleukin signalling and inflammation mediated by chemokine and cytokine signalling pathways. Genes found to be hypermethylated were also found to be involved in inflammation and immune response (*ITLN2, SIGLEC1, SH3RF1*). Further, PANTHER analysis of hypomethylated hits identified B cell activation, T cell activation and TGF beta signalling (table 5.8). At the single gene level, hypomethylation was seen in genes involved in inflammation and immune response (*FYN, CCL18, BCL10, CNR2, ISG15, CLEC4A, CXCL1, GIMAP5, KLRG1* and *MSC*). This heavily implicates dysmethylation of pathways and genes involved in inflammation in ALS.

Inflammation has been implicated in ALS, with microglia (the monocytes of the CNS), B cells (which mature into antibody-producing cells) and T cells (involved in cellular immunity) all being implicated in pathogenesis (Allen, Shaw and Ferraiuolo, 2017). Global inflammatory responses have been found in the mutant *SOD1* (m*SOD1*) mouse model, and higher levels of *Cox-2* (which plays a key role in inflammation) has also been found in PM human sALS SC (Almer *et al.*, 2001). The role of B cells in ALS pathogenesis has not been thoroughly explored; the role of T cells has been investigated. One study has found T-helper cells in close proximity to degenerating corticospinal tracts, with T-helper and T-suppressor cells found in the ventral horns of post-mortem sALS SC (Engelhardt, Tajti and Appel, 1993), with others finding an increase in CD4[+] T-helper cells in ALS induced pluripotent stem cells

(iPSCs) (Song *et al.*, 2016) and a decrease in CD8[+] T-suppressor cells (Chen *et al.*, 2014). Takeuchi *et al.*, 2010 found increased circulating B and T cells in the m*SOD1* mouse model, with Cady *et al.*, 2014 finding increased expression in the microglial activating gene *TREM2* in sALS SC and *SOD1* G93A mice, as detected by qualitative reverse transcriptase-polymerase chain reaction (qRT-PCR). Finally, a study of m*SOD1* mice which artificially depleted mature B and T cells showed reduced microglial activation at end stage disease, along with enhanced disease progression (Alexianu, Kozovska and Appel, 2001). What is of interest, is that the majority of studies into gene expression related to inflammation and immune response is the variety of gene expression changes, with both downregulation and upregulated found (McCauley and Baloh, 2019). This is comparable with the data found using the MethylationEPIC in this chapter, where both hypermethylation and hypomethylation of pathways and genes associated with inflammation and immune response were found. Together, these results show that immune response is differentially regulated in the LMNs of ALS, both in terms of DNA methylation and corresponding mRNA expression analysis.

While inflammation and immune response is generally associated with glial cells, there is precedent for neurones to exhibit inflammatory/immune response-associated pathway responses. One study extracted MNs from the spinal cord of post-mortem frozen human ALS tissue and conducted mRNA expression analysis using the Affymetrix Human Genome U133 Plus 2.0 GeneChip. Upon analysis, it was found that one module was enriched for the GO category 'immune system process', finding inflammation responses in MNs (Cooper-Knock *et al.*, 2017).

Although it is atypical to observe inflammation-related pathways in a motor neurone-enriched sample, consideration was given to the genes highlighted in both single-gene and pathway analysis indicating other unknown functions of these genes. As can be observed in table 5.7, many of these genes also perform other functions separate from immune response that could be differentially methylated and expressed in MNs as part of ALS pathogenesis. However, there is potential for other cell types traditionally associated with inflammation and immune response, such as astrocytes and microglia, acting as a contaminating factor in the LCM-neuronally enriched samples, as tissue immediately surrounding the MNs could have been collected. The limitations of this are discussed at length in section 5.5.5.

5.5.2.4 Apoptosis and cell death

PANTHER pathway analysis of hypomethylated hits implicated a number of pathways involved in apoptosis and cell death (table 5.6). These pathways included apoptosis signalling, the p53 pathway and TGF beta signalling, all of which will be considered below.

Also, hypomethylation of genes associated with apoptosis and cell death were also found (*NCAM1, PCSK9, TEAD2, BCL10, COMMD4, SIDT2, RHEBL1, PIK3C2A, CYCS, TRAF2, SET* and *KLF4*). *HSPB7* and *RPS19BP1* were found to be hypermethylated, with *RPS19BP1* being a direct regulator of *SIRT1*, which is involved in p53/TP53 apoptosis.

p53 is a transcription factor capable of inducing apoptosis. This is through the upregulation of pro-apoptotic genes and the downregulation of anti-apoptotic genes. qRT-PCR analysis of ventral horns from ALS human PM SC isolated by LCM found increases in p53 expression (Eve, Dennis and Citron, 2007). This study also conducted microarray gene expression analysis on ventral horns isolated by LCM from the SC of the Wobbler mouse model. While not a model of ALS, its characteristics include rapid MN death and motor deficits. Increase p53 expression was observed, which was also validated in qRT-PCR and western blot. Immunocytochemistry in PM ALS MNs of the SC and MCx detected the presence of p53, which was not seen in controls (Martin, 2000). IHC on ALS human PM spinal cord also showed an increase in p53 in MNs, but this was not observed in the UMNs of the MCx. This is supportive to findings of the previous chapter, where changes in DNA methylation and hydroxymethylation were seen in LMNs of the SC, but not in neurones of the MCx (section 4.4.5).

Within the MethylationEPIC dataset, hypomethylation of *NCAM1* and *BCL10* were found, which are involved in TGF beta signalling. TGF beta is a regulator of survival and apoptosis in cells. It acts via specific receptors, which activate multiple intracellular pathways. This results in the phosphorylation of Smad2/3 proteins. These proteins then associate with the mediator Smad4. This complex then translocates to the nucleus, where it binds to DNA and regulates the transcription of multiple genes. Overproduction of TGF beta in the m*SOD1* mouse model has been found to accelerate ALS disease progression (Endo *et al.*, 2015). Increased TGF beta levels were also found in the serum of ALS patients (Iłżecka, Stelmasiak and Dobosz, 2002).

In summary, both hypermethylation and hypomethylation in pathways and at the individual gene level implicate changes in cell death and apoptosis in ALS, which correlates to existing literature on the role of cell death and apoptosis in ALS.

5.5.2.5 Cell signalling

At the single gene level, *CD151* and *TSPAN31* were hypermethylated, with Notch signalling also being implicated in the hypermethylated hits. Downregulation of *CD151* mRNA was also observed by Jiang *et al.*, 2005, in motor neurones from human post mortem sALS spinal cord. *CD151* was also downregulated in the lumbar SC of SOD1-ovberexpressing transgenic mice (D'Arrigo *et al.*, 2010). A Drosophila model of spinal muscular atrophy (SMA), a MN

disease, found *TSPAN31* was downregulated in the SMA mutants compared to wild type (Lotti *et al.*, 2012).

Notch is a transmembrane receptor and a master regulator in the CNS, where it coordinates signalling cascades, modulates cell fate decisions and mediates cell-to-cell communication. Our analysis suggests that genes associated with Notch signalling are hypermethylated. No studies have been conducted in ALS to determine the DNA methylation status of Notch signalling in ALS, but studies have been conducted on gene expression. Increases in gene expression associated with Notch signalling were found in the NSC34 cell line, which had been stably transfected with human mutant *SOD1* G93A. IHC on the m*SOD1* G93A mouse model, focused on genes associated with Notch signalling (*Notch1, Jagged1, Hey1, Hes1* and *Maml1*), were found to be increased in ALS compared to wild-type controls (Wang *et al.*, 2015). With hypermethylation being an indicator of downregulation of gene expression, this finding is in opposition to that of our study. However, other studies have been found to detect a downregulation in Notch signalling in ALS. This was determined via qRT-PCR, focusing on gene expression of Notch-associated genes (*HES1, HEY1, NFKappaB and* cyclophilins), which found decreases in gene expression in the middle frontal gyrus of PM human C9ALS tissue and in induced pluripotent stem cell (iPSC)-derived cortical neurones of C9ALS patients (Yang *et al.*, 2015). The discrepancy in findings could be down to the use of different techniques and tissues used in analysis. However, the findings by Yang *et al.*, 2015 are in PM tissue, as is the case with the studies described in this chapter, along with iPSCs. Both of these materials are derived from human origin, whereas the study by Wang *et al.*, 2015 used a model of ALS, which is not as representative of humans.

Together, this supports the findings of the study presented in this chapter, and gives potential targets at the single gene level.

5.5.2.6 Calcium homeostasis

Within the MethylationEPIC data, a slight hypomethylation/upregulation was observed for *ITPR2* (hypomethylation in ALS of 1.6% (p=0.04), with mRNA expression showing a fold change of 1.24 (although non-significant). *ITPR2* is involved in glutamate-mediated neurotransmission in neurones, and is highly expressed in MNs (Bosch *et al.*, 1999). It is the main regulator of intracellular calcium and is a receptor for inositol 1,4,5-triphosphate, which is a second messenger that mediates intracellular calcium release,
ITPR2 has previously been implicated in ALS, with analysis of sALS blood showing upregulation (van Es *et al.*, 2007). Increases in the regulation of *ITPR2* can have an effect on neuronal vulnerability due to higher concentrations of intracellular calcium. These

increases can lead to selective degeneration and death of MNs, with cells that overexpress *ITPR2* being shown to have increased cell death (Gutstein and Marks, 1997).

Within the MethylationEPIC dataset*, SRI* is found to be hypermethylated (12.2% higher DNA methylation in ALS (p=0.012)). *SRI* encodes a calcium-binding protein which relocates from the cytoplasm to the endoplasmic reticulum when intracellular calcium levels are raised. If this gene is hypermethylated/downregulated, it could be that the protein is no longer being produced to the same levels, thus allowing intracellular calcium levels to increase. Combined with the increased *ITPR2*, this could indicate than dysmethylation has a role in calcium homeostasis dysregulation in LMNs affected by ALS.

Calcium ions are held in reserve within the endoplasmic reticulum. Calcium ions are released into the cytosol upon extracellular signalling which detect lowered levels of cytosolic calcium ions. This signalling is activated by neurotransmitters. Calcium homeostasis is important in neurones, with calcium ions found to be involved in a number of functions. These include controlling dendritic responses to neurotransmitters, gene expression regulation through signalling to the nucleus and neurotransmitter release initiation from presynaptic terminals. Calcium ions are also involved in the regulation of neuronal excitability, cell proliferation and cell death (Gleichmann and Mattson, 2011). Calcium homeostasis dysregulation has been highly implicated in ALS pathogenesis, and is involved in MN selective vulnerability and in defective proteins associated with ALS. *SOD1* aggregation has previously been linked to calcium overload in the m*SOD1* G93A mouse model (Tateno *et al.*, 2004) and in a murine cell culture expressing human *SOD1* (Tradewell *et al.*, 2011).

TDP43 aggregation has also been implicated in calcium homeostasis dysregulation, with increased calcium levels found to activate the calcium ion-dependent calpain protrease, which cleaves TDP43 at the C terminal. This forms the N-terminal segments, which are aggregation prone and found in ALS (Aggad *et al.*, 2014; Yamashita and Kwak, 2014). One study generated iPSCs from ALS patients, with calcium imaging revealing mutant TDP43 showing increased basal intracellular calcium levels. Further, this study also showed elevated expression of voltage gated calcium channels in iPSCs derived from patients with mutant *C9orf72*, as measured by qRT-PCR (Bursch *et al.*, 2019).

5.5.2.7 *SERPINA1*

SERPINA1 is a serine protease inhibitor, which is part of a larger family of serine protease inhibitors, which includes *SERPINA3. SERPINA3* has previously been implicated in both AD (Kalsheker, 1996; Kamboh *et al.*, 2006) and PD (Wang *et al.*, 2001). *SERPINA1* has been shown to be both differentially expressed and methylated in ALS by Ebbert *et al.*, 2017. In this study, it was found that the methylome profiles of sALS and C9ALS cases overlapped,

which was also the case in the data presented in this chapter. This study conducted both RNA sequencing (RNA-seq), qRT-PCR and RRBS to determine gene expression and DNA methylation changes, respectively. *SERPINA1* had significant increases in gene expression and hypomethylation in sALS frontal cortex and sALS and C9ALS cerebellum. This is the same effect observed in our analyses, which showed a 1.0 fold increase in gene expression (although not significant), and a 20.3% decrease in DNA methylation (p=0.006). Combined, these results indicate that serine protease inhibitors may obstruct neuronal function in neurodegenerative disease.

Considerations were given to comparing the datasets produced by Ebbert *et al.*, 2017 as the datasets were carried out in human post-mortem sALS and C9ALS brain. However, a number of factors prevented this comparison. The Ebbert study focused on frontal cortex and cerebellum homogenates, areas not typically associated with ALS pathology. This raises the issue of masking of cell-specific responses and differences in ALS pathology generally observed when comparing a highly affected area such as spinal cord with less affected areas such as frontal cortex and cerebellum samples, with comparisons therefore not reflective of region-specific responses.

5.5.2.8 Limitations of comparing MethylationEPIC and mRNA expression data

Many of the promoters were not validated using this comparison. This can be attributed to a number of reasons. First, it is not just one step between methylation and mRNA expression, with multiple other regulatory systems involved in the interim stages. These include mRNA transcription, alternative splicing, RNA transport, mRNA stabilisation and nonsense mediated decay (Butti and Patten, 2019). Another factor to consider is the influence of other epigenetic mechanisms. This includes histone modifications and microRNA (miRNA), which are both involved in transcription regulation. Histone modifications are also highly implicated in mRNA expression, along with chemical modifications. Histone modifications can result in both loosening of the DNA, allowing for transcription to occur, and tighten DNA, repressing transcription. Modifications on the N-terminal of histone tails include acetylation, phosphorylation and ubiquitination. These modifications can influence how tightly bound the DNA is packed, therefore also affecting transcription. The cohorts used in the analysis are different from each other, so natural variation will be observed as a consequence of this.

Another consideration is that the mRNA expression dataset is now five years old, with newer, more robust expression analysis tools now available (discussed further in section 6.4.2). RNA levels in general are prone to high levels of fluctuation, and these fluctuations can be rapid. This is in contrast to age-related DNA methylation changes, which are

generally regarded as more stable alterations. The mean differences in methylation between <35yo and >55yo is only 3.2% (Horvath and Raj, 2019), based on the average change observed based on the subset of CpGs used to calculate epigenetic age, and it is therefore unlikely that these small changes in will be reflected at the RNA/protein levels. Further, epigenetic changes in general occur in small number of cells, with different types of cells exhibiting different changes, which might not be seen in analysis of a collection of cells. One solution to this is to do single cells analysis (discussed in section 5.5.9). It may be that the cell has no desire to transcribe a given gene. Thus the promoter can be as unmethylated as possible, however if the relevant transcription factor has not been produced as the cell has no need for that protein, it will not be expressed. Equally, while a gene may be methylated to some degree, it may still be transcribed if sufficient transcription factor is produced. Finally, while the individual genes did not show complete correlation, the GO/PANTHER analyses demonstrated that there was dysmethylation which affected similar processes to altered mRNA expression, such that the cell was using the same pathways.

5.5.3 Study limitations

Limitations of using PM tissue have been discussed previously in sections 3.7 and 4.5.7, with the limitation of using LCM discussed in section 5.5.4. Here, further limitations specific to the MethylationEPIC findings of this chapter will be discussed.

First, the study was underpowered (see section 5.4.2), as assessed by G*Power. In ideal circumstances, had ample time and resources allowed, greater number of samples would have been used. The use of FFPE tissue is also uncommon in DNA methylation analysis, with subsequent extracted DNA being of reasonable, but not perfect quality (as evidenced by the lack of ability to accurately calculate the DNA methylation age of samples, see section 5.4.10).

As per the analysis conducted in chapter 4 (section 4.4.6), separating out LMNs that displayed TDP43 pathology from those who did not, and conducting separate DNA methylation analysis on these, would have been of benefit. However, this is unfeasible due to the relative small population of LMNs displaying TDP43 pathology compared to LMNs not displaying TDP43 pathology. In order to gain enough material to conduct this, large amounts of spinal cord tissue would be required, which is not possible for such a precious resource.

It was established in section 4.5.3 that higher DNA methylation and hydroxymethylation were seen in neurones compared to glia. It would have been useful to extend the MethylationEPIC study to include analysis of glial cells, as it is likely that DNA methylation changes at the single gene level would be present, especially considering the potential role of

dysmethylation in inflammation and immune response (section 5.5.2.3), in which glial cells are involved with.

Methylation levels observed in the MethylationEPIC data relate to methylation types combined (5mC, 5hmC), whereas the IHC studies were focused solely on 5mC and 5hmC. This could be why different results are observed in these experiments. Further IHC studies focused on 5hmC may help to elucidate this. A potential solution to this is to use modified bisulphite sequencing to determine 5hmC levels. This can be done through the use of Tet-assisted bisulphite sequencing (TET-BS), or oxidative bisulphite sequencing (Ox-BS). TET-BS uses the Tet enzyme to distinguish between 5mC and 5hmC. In this method, 5hmC is protected by glycosylation, whereas 5mC is converted to 5caC using Tet. When sequencing, any Cs in the sequence are interpreted as 5hmC. The benefit of using this method is that it provides a direct measurement of 5hmC, but it does rely on the Tet enzyme, which does not have 100% efficiency, and it is also very expensive to produce (Yu *et al.*, 2012). Ox-BS provides measurements of both 5mC and 5hmC. Normal bisulphite conversion is conducted to determine 5mC levels, as well as an additional bisulphite run containing an oxidative step to allow for measurement of 5hmC. In this step, 5fC is sensitive to deamination by bisulphite treatment. So, bisulphite conversion of the oxidised DNA converts 5fC to uracil, with any remaining Cs interpreted as 5mC. After this, both the regular BS DNA and the ox-BS DNA are sequenced and compared, with 5hmC levels inferred from the differences in 5mC and 5fC levels. However, this method required two runs of sequencing, which can be costly, and is not a direct measurement of 5hmC (Booth *et al.*, 2013).

5.5.4 Laser captured MNs from FFPE tissue are compatible with the MethylationEPIC BeadChip

Most analyses using the MethylationEPIC method are on samples where large quantities of tissue or cells are available. For instance, cell cultures or PM tissue homogenates. The combination of using FFPE tissue with LCM-MNs has not previously been attempted. This chapter indicates that this method is viable but time consuming, allowing for cell-specific methylation analysis to be conducted. Flow cytometry could address this, as it has previously been used to sort specific neuronal populations (Martin *et al.*, 2017). In this paper, a method is described whereby FFPE human frontal cortex tissue is tagged with NeuN, a neuronal marker, followed by flow cytometry to isolate this population. While this technique is useful, the NeuN marker doesn't distinguish between MNs and other types of neurones, and this method was only found to be successful in FFPE samples that were fixed a maximum of eight weeks prior to the experiment and involved large quantities of tissue in comparison to

the study described in the chapter, which is unrepresentative of the majority of FFPE brain tissue available and is incompatible with human post-mortem spinal cord.

5.5.5 Benefits and limitations of laser capture microdissection

LCM is a widely used and tested method of extracting a specific cell type from PM tissue in order to produce an enriched sample of a particular cell type. This has also been the case in this study thus far, with MNs successfully being extracted from FFPE SC tissue (see appendix III for image of laser captured MNs). Bisulphite conversion on these samples was also successful (see section 5.4.1), which allows further analysis to take place.

However, some considerations have to be taken into account when conducting this further analysis. As FFPE tissue was used throughout this project, the limitations of this type of tissue need to be acknowledged. FFPE produces poorer quality DNA when compared to frozen tissue due to the DNA fragmentation that occurs during the tissue fixation, processing and embedding. However, there are also many benefits to using FFPE tissue in a project such as this one, including that FFPE tissue is easier to store than frozen tissue. Recent advances of techniques using FFPE tissue have also been developed. In the future, the Illumina® Infinium® MethylationEPIC BeadChip should be used to create DNA methylation profiles from a large cohort, to increase the power of the experiment, with MNs collected from the FFPE tissue. This method is now adapted for use with this tissue type (Illumina® Infinium® MethylationEPIC BeadChip data sheet), requiring an input of 250 ng. The NanoDrop 1000 spectrophotometer analysis of the DNA extraction and bisulphite conversion data suggests ample DNA is present within the samples for this to be successful (section 5.4.1).

5.5.6 Horvath clock analysis

Horvath clock methylation analysis could not be accurately predicted for this dataset. This could be because of a number of factors. On factor to consider is that the Horvath equation is optimised for the 450k array, and some of the probes used to calculate methylation age are not present in the MethylationEPIC array. However, a study by McEwen *et al.*, 2018 suggests that the Horvath clock is appropriate for use with the EPIC array. In this paper, DNA methylation analysis was carried out on human primary monocytes using both the 450k array and the MethylationEPIC array, followed by carrying out Horvath clock analysis. Methylation age prediction between he 450k and MethylationEPIC was highly correlated, suggesting the Horvath lock is applicable to the MethylationEPIC array.

More likely responsible for the Horvath algorithm not being appropriate was the low quality DNA as the sample input. This led to nearly half of the probes on the MethylationEPIC array

being discarded in the Greedycut filtering stage. This again lowered the number of probes used to calculate methylation age. Therefore, the results of this calculation were not reliable, as they have not been calculated using all probes that the algorithm relies on for accuracy.

5.5.7 Compatibility of archival FFPE LCM tissue and BS-NGS

The outcomes of the BS-NGS section of this chapter suggest that the Zymo Methyl-Seq™ Library method is not compatible with DNA from LCM/FFPE. This was characterised by the poor alignment with the reference genome, and the low number of sequence reads generated from the NGS data. This could be due to a number of factors. Firstly, the compatibility with low quality, low input archival DNA samples could affect the ability of the library indexes to bind to DNA present in the sample. This would significantly affect the amount of DNA that proceeds through all stages to the final sequencing. The method employed in this study requires many washes and steps where DNA could potentially be lost, with shorter DNA fragments more likely to be lost compared to longer sections of DNA. This leaves FFPE more vulnerable to this kind of loss, as DNA from FFPE tissue is usually shorter than frozen tissue due to the tissue fixation process, which can cause the fragmentation of DNA. Poor alignment with the reference genome would also indicate this, as fewer reads and many small fragments of DNA are difficult to map. This method also uses PCR amplification, which can introduce bias, which could have resulted in the mis-amplification of DNA, thus meaning alignment could not occur.

5.5.8 Limitations and benefits of frozen versus FFPE tissue

Frozen tissue has been most commonly used in sequencing/array-based post-mortem experiments. There are a number of reasons for this, including that frozen tissue has not undergone any tissue fixation processes, meaning no DNA cross-linking and reducing the chance of introducing contaminants into experiments. DNA and RNA extracted from frozen tissue is generally of higher quality than FFPE tissue. During the FFPE tissue fixation process, DNA becomes fragmented and is subject to degradation during this process. The formalin fixation causes these DNA breaks, resulting in a decline in DNA quality. Over time, this degradation continues, with shorter DNA fragments. For example, a 20 year old FFPE sample would have poorer quality, more fragmented DNA that a 2 year old FFPE sample, where the DNA fragments would be longer, and therefore the DNA would be of a higher quality. Another consideration is the comparison of datasets, as it is preferable to comp-are datasets where samples have undergone similar processing techniques, resulting in it being preferable to compare datasets generated from frozen material to be compared to other datasets also originating from frozen material.

However, the use of FFPE in array and sequencing studies is gaining popularity as analysis tools improve. The benefits of using FFPE tissue in studies of this type are vast. For example, there is generally greater availability of FFPE tissue in brain banks when compared to the more 'precious' resource of frozen tissue. This can result in obtaining larger experimental cohorts, and can give scope to more easily sex and age matching experimental groups, with archival FFPE tissue more likely to be available. Other practicalities of using FFPE tissue include not requiring specialist storage facilities, as is the case with frozen tissue requiring freezer storage.

5.5.9 Options for future studies

There is precedent in the cancer field of using FFPE tissue in conjunction with sequencing technologies, such as FFPE tissue from human tumour biopsies undergoing reduced representation bisulphite sequencing (RRBS) (Ludgate *et al.*, 2017). While this is based on the use of tissue homogenates with a large DNA input, it does suggest that technologies are being developed to better utilise FFPE tissue in sequencing-based experiments. A relatively new technology that could alleviate the need for error-prone PCR amplification steps is nanopore technology, such as that designed by Oxford Nanopore Technologies, which has a raw base-called error rate of <5% (Amarasinghe *et al.*, 2020). This is a technique that allows for the detection of base modifications through the use of electrolytic currents. Long, single molecules of DNA can be sequenced with no bisulphite conversion or PCR amplification required. Simpson *et al.*, 2017 details a method for the successful identification and analysis of the DNA methylation marker 5mC in *Escherichia coli* DNA using nanopore technology. The competitor for nanopore technology is the PacBio single-molecule real-time (SMRT) long-read sequencing method. This technique detects fluorescent events corresponding to the addition of a specific nucleotide by a polymerase attached to a well, with a raw base-called error rate of <1% (Amarasinghe *et al.*, 2020). However, these techniques may currently be incompatible with FFPE tissue, as FFPE tissue usually has short fragments of DNA, with the nanopore and SMRT techniques optimised to read long pieces of DNA. Another current caveat is the relatively large amount of input DNA required for both techniques, which would be feasible for homogenates. This thesis has shown that DNA methylation changes are neurone-specific. A potential solution for this is to conduct the long-read sequencing on tissue homogenates and then apply bioinformatics tools to filter for neuronal specific markers. While these techniques are still in their infancy, as the techniques become more refined and the sample input need and cost of the procedures are reduced, these techniques could become a viable option for the study of PM tissue, albeit frozen rather than FFPE.

A further possible option is single cell -omics, which would allow for DNA methylation and mRNA expression analysis to be conducted in the same cell. However, this still relies on bisulphite conversion and PCR amplification. Single cell -omics was conducted by Linker *et al.*, 2019, where single differentiating human induced pluripotent stem cells underwent both single-cell RNA sequencing and single cell bisulphite conversion and sequencing. While the starting material use in this study was of significantly higher quality, with DNA still being intact, it does indicate that this method is becoming more widely used, and is suitable for low DNA input. Bisulphite conversion is potentially not well suited towards single cell omics, due the DNA degradation associated with the bisulphite conversion process, as well as the error-prone PCR amplification process after this (Karemaker and Vermeulen, 2018).

Within the cancer field, a study has been published in which single ovarian cancer cells were extracted from FFPE tissue using LCM. Following this, whole genome amplification, PCR purification and bisulphite conversion were conducted prior to sequencing. Of the 12 cells separately isolated, three were successful in this process, with differing DNA methylation profiles for each cell, indicating that DNA methylation is heterogeneous even within the same cell populations (Q. Li *et al.*, 2017). Another study using neurones isolated from frozen PM human frontal cortex identified 21 neuronal subpopulations based on the DNA methylation profiles of the cells (Luo *et al.*, 2017). Together, these studies indicate that current methods are becoming increasingly useful in the study of PM tissues, and the potential for FFPE/LCM techniques like the one utilised in this thesis, for conducting single cell DNA methylation analysis.

5.5.10 Final remarks

Results presented in this chapter corroborate the findings of the 5mC and 5hmC IHC presented in chapter 4, with an increase in global DNA methylation associated with C9ALS in lower motor neurones. Further to this, DNA methylation changes in gene promoters were also found in ALS, implicating pathways involved in RNA metabolism. This hypermethylation implicated cell signalling, cell death/apoptosis and inflammation/immune response, which have all previously been implicated in ALS pathogenesis.

Concerning the suitability of archival FFPE, LCM samples and their viability as a DNA source for BS-NGS, this chapter highlights the current limitations in conducting cell-specific pathological studies in FFPE PM tissue using BS-NGS, and details potential new technologies to address this.

Chapter 6: Conclusions and future work

ALS is a fatal neurodegenerative disease characterised by motor neurone (MN) degeneration and subsequent death. The current project aimed to define the DNA methylation-specific changes in MNs and glia in both sporadic ALS and ALS due to mutations of *C9orf72*. This included conducting an immunohistochemistry (IHC)-based pathological study of DNA methylation and hydroxymethylation in spinal cord, motor cortex and anterior frontal cortex (SC, MCx and AFCx). The association between TDP43 pathology and DNA methylation and hydroxymethylation was also investigated in lower MNs of the SC. Further, in order to establish gene-level DNA methylation changes occurring in ALS MNs, LMNs were extracted from SC tissue and subjected to analysis in the form of the MethylationEPIC BeadChip. The study also aimed to understand the relationship between findings of the MethylationEPIC BeadChip and mRNA expression data.

6.1 Main findings of the project

- IHC for DNA markers 5mC and 5hmC showed greater DNA methylation and hydroxymethylation levels in residual LMNs of the SC in ALS. This was not found in glial cells, or in neurones of the MCx or AFCx. This suggests cell- and region-specific DNA methylation and hydroxymethylation changes occur in ALS.

- This was further seen in the MethylationEPIC BeadChip array for ALS cases due to mutations of *C9orf72* (C9ALS), where an increase in the global methylation levels of LMNs isolated from the SC via LCM. Significantly greater DNA methylation were observed in the C9ALS group when compared to controls, with sALS at an intermediate level (although this was non-significant).

- Analyses of individual genes using the MethylationEPIC platform also revealed DNA methylation changes in ALS. The majority of changes noted in ALS showed an increase in methylation in promoters of protein coding genes, suggesting potential functional impacts. However, significant hypomethylation was also detected at the gene level.

- Comparing the findings of the MethylationEPIC array to pre-existing mRNA expression data at the GO/pathway level found overlaps in the involved pathways, implicating cell signalling, apoptosis/cell death and inflammation/immune response.

- TDP43 pathology corresponds with loss of DNA methylation from LMN nuclei, observed using IHC. TDP43 has unknown roles in DNA binding, suggesting that an interaction with DNA methylation is plausible. This potentially also relates to the dysregulation in RNA metabolism found.

6.2 Models of ALS and DNA methylation

A potential issue with conducting pathology-based studies is understanding the direction of causation. In this project, a relationship between ALS and hypermethylation is present. However, it is impossible to know if ALS causes hypermethylation, hypermethylation causes ALS, or that a completely different mechanism is causing both ALS and hypermethylation.

The underlying pathogenesis of degeneration in MNs is mostly unknown. It is therefore useful to investigate the pathogenic features that lead to MN degeneration, and ultimately, death. This is a current issue with PM tissue, as it only gives an indication of the DNA methylation status of cells that have survived, and not those that have perished during the course of disease. Using models can overcome this, with investigations into both the onset and progression of disease feasible. Unlike with the use of human post-mortem tissue, animals can be sacrificed at various time points within the disease, which allows any changes in DNA methylation to be detected and mapped over time. This gives an indication of any changes that occur with age (when studying the wild type group), and any changes associated with disease progression detected in the modified group. Another benefit is that any environmental factors, such as diet, weight and exercise can be managed and are more homogeneous than with humans.

Genetic models of disease are the most common. A potential issue in using genetic models of disease is that of overexpression-specific effects, which are usually needed in order for the model to exhibit signs of neurodegeneration prior to death. This therefore is not a complete mimic of the disease state in humans. Consideration in how to overcome this include the use of Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISP-associated protein-9 nuclease (cas9) (CRISPR-Cas9) and induced pluripotent stem cell (iPSC)-derived neurones. Below are sections discussing for the most used models in ALS currently. The mutations that these models are based on only account for 5-10% of diagnosed cases, with most diagnoses of ALS defined as sALS. This highlights the need of models to replicate sALS.

CRISPR/Cas9 is a genomic editing tool. Originally, CRISPR/Cas makes up part of the bacterial natural defence system, which is now being used as an RNA-guided editing tool for DNA. Cas9 is an RNA-guided DNA endonuclease enzyme. It creates double stranded breaks at particular genomic locations, which trigger DNA repair pathways. This contributes to the desired genomic modification. The method has been used to create a zebrafish model of ALS by introducing point mutations in *TARDBP* and *FUS* genes (Armstrong *et al.*, 2016), and a *C9orf72* deficient mouse model of ALS (Sullivan *et al.*, 2016). In regard to DNA

methylation, this system could be used to study the DNA methylation changes in wild type versus mutation, to identify what pathways are dysregulated.

6.2.1 Zebrafish models

The use of zebrafish is well-documented in ALS. They are useful as they have similar gene homology to mammals (approximately 70%) and have had their DNA methylation profiles characterised (Howe *et al.*, 2013). This makes them a good option for vertebrate-specific aspects of DNA methylation. Compared to mouse models, zebrafish models are a simple system, with rapid development and large egg numbers. Zebrafish are also easy to genetically modify, with genes previously implicated in neurodegeneration being highly conserved between humans and zebrafish (Morrice, Gregory-Evans and Shaw, 2018). However, there are some limitations with the use of zebrafish models, mainly being that they are not mammals and do not have UMNs. The zebrafish genome codes for eight DNMT orthologs, including DNMT1 and DNMT3, with similar protein structures to human orthologs. Having both DNMT1 and DNMT3 orthologs also indicates some level of both *de novo* and maintenance of DNA methylation in zebrafish (Goll and Halpern, 2011).

Unfortunately, no studies to date have been published considering DNA methylation changes in ALS, or the neurodegeneration field more generally, using the zebrafish model. However, in our lab, it has successfully been shown that 5mC and 5hmC antibodies detailed in this study are compatible with zebrafish (unpublished works). This potentially highlights the use of zebrafish in studying DNA methylation in ALS. The use of zebrafish to study DNA methylation in conjunction with TDP43 proteinopathy may be problematic, as unlike humans, zebrafish have two TARDBP genes (TARDBP and TARRBPL) (Hewamaddumal *et al.*, 2013). Considerations as to suitable models to study this are discussed in section 6.2.3.

6.2.2 Mouse models

Mouse models of DNA methylation have started to be developed. These are mostly in the cancer field, with genetic manipulation targeting genes that are DNA methylation regulators in the form of DNMT inhibitors (Conerly and Grady, 2010). However, steps have been taken to progress this into the neuroscience field (Landgrave-Gómez, Mercado-Gómez and Guevara-Guzmán, 2015). A rat model of epilepsy studied the global DNA methylation changes within the hippocampus, which showed hypermethylation. This was comparable to studies of human epilepsy hippocampal tissues (Kobow *et al.*, 2013). Fuso *et al.*, 2011 manipulated the TgCRND8, a model of AD which overexpresses mutant human APP, mouse line to cause B vitamin deficiency, with this resulting in hypomethylation, as determined by

BS-sequencing. Interestingly, B vitamin metabolism was found to be hypomethylated in PANTHER pathway analysis (table 5.6).

The CRISPR/Cas9 system has recently been used to target DNA methylation machinery. Tet1 fusion proteins targeted to methylated promoter sequences caused increased transcription activation, while Cas9 targeting of DNMT3a fusion proteins to unmethylated promoter sequences caused gene silencing. Further, this study isolated mouse post-mitotic cortical neurones from embryos, using Cas9 to conduct targeted demethylation of BDNF promoters, which resulted in increased BDNF expression (Liu *et al.*, 2016).

A mouse model of neurodegeneration (senescence-accelerated mouse P8 (SAMP8), characterised by learning and memory deterioration with age) has also been studied in the context of DNA methylation. SAMP8 mice that had environmental enrichment showed increase 5mC levels, with an increase in expression of antioxidant genes, and a decrease in expression of inflammatory diseases (Griñan-Ferré *et al.*, 2016). Carrying out a similar study in an ALS-specific mouse model would be of use, with the findings of the SAMP8 acting as a validation to any dysmethylation observed.

6.2.3 Cell culture models

iPSC neurones derived from ALS patients is another consideration in studying DNA methylation changes in ALS. However, it is still being established if neurones derived from iPSCs retain their original DNA methylation profiles.

A study by de Boni *et al.*, 2018 investigated how suited iPSCs are to epigenetic modelling. In this study, the DNA methylation patterns and gene expression of mature neurones derived from hESCs were compared to DNA methylation patterns and gene expression of mature neurones derived from iPSCs. The 450k DNA methylation array (the precursor to the MethylationEPIC) was used, in conjunction with the HT12v4 gene expression array. Results found that DNA methylation patterns and gene expression were highly preserved between the hESCs and iPSCs, both at the global level and the single gene level. While this may be useful for developmental studies, it does suggest that this may not be representative of aged neurones. Choi *et al.*, 2015 also found the same effect, with RNA sequencing only identifying 49 differentially expressed mRNA, with reduced representation bisulphite conversion showing little changes in DNA methylation patterns. Further, Teichroeb, Betts and Vaziri, 2011 found that the transcriptome of neuronal iPSCs and hESCs had a match rate of 99.6%, indicating high conservation. However, when translating this to comparisons with native human tissue, large transcriptional differences were found between iPSCs and primary midbrain dopamine neurones (Xia *et al.*, 2016). This indicates that there are discrepancies that require futher studies to establish the differences in DNA methylation profiles of mature native neurones versus those derived from fibroblasts and redifferentiated.

Once the DNA methylation profile retention of neurones derived from iPSCs is established, it could provide a useful model for understanding the association of TDP43 proteinopathy and DNA methylation. While not focused on DNA methylation, Tank *et al.*, 2018 conducted a genome-wide study of RNA stability using RNA sequencing in iPSCs derived from sALS and C9ALS patients. The role of TDP43 accumulation was also hypothesised to contribute to RNA instability in patient ALS iPSC-derived neurones. This was tested by overexpressing TDP43 in control neurone iPSCs and conducting RNA sequencing. 1,330 hits were identified, with 75% of these described as destabilised, compared to neuronal iPSCs that did not have TDP43 overexpression. Destabilisation of both ribosomal and mitochondrial transcripts were found, with liquid chromatography-mass spectrometry (LC-MS) also showing reductions in mitochondrial components. Compensatory increases in protein synthesis were also present. Validation of the RNA sequencing were then conducted on sALS and C9ALS SC, conducting qRT-PCR, which again found reduced abundance of RNAs that were related to the ribosome and mitochondrial oxidative phosphorylation. This approach could be used to study TDP43 proteinopathy in ALS, and its possible association with DNA methylation. Conducting DNA methylation analysis on neuronal iPSCs that have overexpression of TDP43 versus those that are wild type, followed by comparing the DNA methylation profiles of both, would give some indication of dysmethylation associated with TDP43 expression.

Considerations for this current thesis include the finding by others that age-specific gene expression signatures are not maintained during the reprogramming of fibroblasts into iPSCs (Mertens *et al.*, 2015). The implicated pathways (ribosome biogenesis and oxidative phosphorylation) are essential for cells that are metabolically active, such as MNs (Jung, Yoon and Holt, 2012; Smith, Shaw and De Vos, 2017). Mitotic cells, such as glia or iPSCs can potentially address abnormalities in these pathways through cell division, but this has larger implications for post-mitotic MNs, which cannot perform this function. Due to the origins of iPSCs, namely the cell reprogramming, there is potential for epigenetic remodelling to occur, which would raise the issue of viability of iPSCs to model DNA methylation in ALS.

In conclusion, the use of iPSCs to study DNA methylation changes is still being investigated, and careful considerations need to be taken when using this model in DNA methylation studies.

6.3 Cell specificity

In order to understand a particular cell type's role in disease pathogenesis, the ability to extract cells of interest from human tissue is of vital importance. LCM is a technology which allows for isolation of particular cell types, and has successfully been shown to isolate MNs

from human tissue in neurodegenerative disease (Batra *et al.,* 2016). In ALS, LCM has previously been used to isolate MNs in order to investigate mRNA expression changes (Highley *et al.*, 2014). However, this technology has not been employed when studying DNA methylation changes. In this study, a method for isolating an enriched MN population and conducting and analysing gene-level DNA methylation changes was developed. An advantage of working with enriched cell populations over the use of whole-tissue homogenates is that the individual pathological mechanisms of disease specific to particular cell types can be elucidated. These differences may be masked when using whole tissue homogenates.

It has been successfully demonstrated in this study that cell-specific differences in DNA methylation occur (hypermethylation and hyper-hydroxymethylation in ALS LMNs). Previous studies have also established the difference in DNA methylation in neuronal versus non-neuronal cell populations. PM human prefrontal cortices from neurologically healthy donors underwent fluorescence activated cell sorting (FACS) using the neuronal marker NeuN. This was followed by pyrosequencing, with differential expression of transcription factors found in neurones when compared to the non-neuronal population (Iwamoto *et al.*, 2011). Kessler *et al.*, 2016 analysed cell-specific signatures using the 450k DNA methylation array on PM neurologically healthy human cortex, finding that neurones exhibited higher levels of DNA methylation than the non-neuronal population. This was then verified in mouse cortex using pyrosequencing.

Price *et al.*, 2018 conducted whole genome bisulphite sequencing (WGBS) on human PM prefrontal cortex using NeuN/FACS and also found differences in DNA methylation between neurones and non-neuronal cells. PM human prefrontal cortex was also sorted using FACS to separate out different types of neurones (medial ganglionic eminence-derived inhibitory GABAergic interneurons and excitatory glutamatergic neurones) and oligodendrocytes. These cell populations then underwent WGBS, with highly distinct profiles being found for each cell type (Kozlenkov *et al.*, 2018). Four brain regions from neurologically healthy PM brains (anterior cingulate gyrus, hippocampus, prefrontal cortex and nucleus accumbens) underwent FACS sorting for neuronal and non-neuronal populations. Findings showed significant differences in DNA methylation patterns in both neuronal and non-neuronal populations and between the different brain regions (Rizzardi *et al.*, 2019). The finding that different neurones from different regions exhibited distinct DNA methylation profiles could go some way in explaining why differences in DNA methylation were observed between the LMNs of the SC compared to the UMNs of the MCx and neurones of the AFCx (section 4.5.3). This was also seen in a study of post mortem human AD occipital lobe, whereby flow cytometry was used to sort neurones from other cell types (mainly glia), again using NeuN.

The 450K DNA methylation array was then used, which detected genes with cell-type specific DNA methylation signatures for both neuronal and glial populations (Gasparoni *et al.*, 2018).

Combined with the findings of the studies detailed in this thesis, this literature highlights that differences in DNA methylation signatures at both the global and single gene level are present between neurones and non-neuronal cells. This is highlighted in both neurologically healthy brain/CNS tissues, as well as in neurodegenerative disease. Further to this, there is emerging evidence that even within neurones, there are differences in DNA methylation (Kozlenkov *et al.*, 2018). This is supported by the IHC findings discussed in section 4.5.3, which found differential expression of both methylation and hydroxymethylation in LMNs of the SC compared to UMNs of the MCx and neurones of the AFCx. A limitation of the above studies, as well as a limitation of IHC studies of glia (section 4.4.7), is that the different glial cell populations were not separated out into astrocytes, microglia and oligodendrocytes. It is probable that differences in these cell populations will also be present, due to their differing roles within the brain and CNS. What is also of note is that many different techniques, tissues, laboratories and cohorts (neurologically healthy, AD and ALS) were used. However, what the above literature does highlight is that complimentary finding are present. This suggests a genuine phenomenon, rather than findings being as a result of experimental procedure.

6.4 Beyond the scope of this project: future works

A multi-pronged approach, much like the one used in this project, is the most viable option for further studies into the role DNA methylation plays in ALS. These include further, expanded studies on PM tissues used in conjunction with experimental models, which allow for manipulations of both TDP43 and DNA methylation levels. Each are discussed below.

6.4.1 Immunohistochemical studies

Repeating the IHC studies in another cohort would be of use. This would enable determination of if this effect is specific to the cohort used within this study, or if this is representative of the DNA methylation and hydroxymethylation status of LMNs in ALS. Further studies on the cohort used in this project is also warranted.

Antibodies for 5fC and 5caC, DNA demethylation markers (summarised in figure 1.6) are beginning to come to market. This gives the potential for methylation status determination of a cell all the way through the methylation pathway. Once DNMT antibodies have become available, DNMT characterisation in LMNs would be a useful addition, as it is directly

involved in the DNA methylation pathway, and may be aberrant, and therefore the cause of altered methylation changes.

6.4.2 MethylationEPIC and mRNA expression studies

The main limitation with the study presented is the small numbers of cases that underwent MethylationEPIC analysis. This was due to a number of factors: availability of tissues, the time consuming nature of LCM, and limited financial resources. Expanding the cohort size would increase the robustness of findings presented. Power analysis on the cohort used in the MethylationEPIC experiment indicated that the experiment was underpowered (section 5.4.6). Increasing the cohort size would increase the power of the experiment, and thus and deductions made from this analysis.

Although mRNA expression data was available for comparison in this study, the data available was from a different cohort. In addition, better methods for analysing mRNA expression are now available, as the dataset used in this study was five years old. RNA-seq is the whole transcriptome sequencing of RNA to establish gene expression changes. It has many benefits over the array method used for comparison in this study. A number of studies have been conducted to compare arrays and RNA-seq to determine gene expression changes, with RNA-seq outperforming arrays, with higher sensitivity, better detection of low abundance transcripts and higher agreement with qPCR validation (Wang *et al.*, 2014). RNA-seq is also better for determining alternative-splicing and discovering novel splicing events at the genome-wide level, whereas arrays have poor validation for splicing events. However, they are well suited to detecting splicing events on well-described transcription regions. However, cost is a factor, with RNA-seq being more expensive (Romero *et al.*, 2018).

It would be useful to LCM extract MNs, and then divided the same sample into two, one for MethylationEPIC analysis and one for mRNA expression analysis, so that any deductions are directly comparable. Other useful analyses could also be carried out on the MethylationEPIC dataset generated during this project. This includes the use of the 'genomic regions enrichment of annotations tool' (GREAT). This analysis technique can be used to analyse the annotations of nearby genes to allow for analysis of non-coding genomic regions (McLean *et al.*, 2010).

Another viable option would be to perform analysis of differentially methylated regions (DMRs), which comprise multiple consecutive methylated CpG sites. DMRs are defined as genomic regions that have varying DNA methylation statuses among samples, and have the possibility of functional regions that may be involved in gene transcription regulation. Analysis of DMRs have the potential to identify and therefore allow characterisation of

networks associated with ALS disease development and progression. Many packages are available for such analysis, for example, DMRcate (Peters *et al.*, 2015). If time permitted, this would have been a suitable next step for further analysis.

A future advancement in the field could be the study of mitochondrial DNA methylation and its effect in ALS. Mitochondrial DNA methylation is not covered in the MethylationEPIC array, so this is a reassurance of a lack of mitochondrial DNA contamination in the MethylationEPIC array dataset, however it does highlight that other techniques need to be utilised to study mitochondrial DNA methylation.

Mitochondrial impairment and increased oxidative stress are features associated with neurodegeneration. Impaired meth levels of the mitochondria regulatory region (D-loop region) in both animal models, PM brain regions or circulating blood levels of patients with AD, PD and ALS (Devall *et al.*, 2017). In general, very little research has been conducted in the area of mitochondrial DNA methylation in neurodegeneration. Mitochondrial DNA doesn't have histone-like packaging proteins. Therefore, it is more susceptible to oxidative damage and is therefore prone to a higher rate of mutation when compared to nuclear DNA.

However, it is harder to study mitochondrial DNA than nuclear DNA due to methodological limitations. Mitochondrial DNA has no CpG islands and is organized into tightly packed nucleoprotein complexes known as nucleoids (Coppedè and Stoccoro, 2019). One study has detected mitochondrial DNA methylation levels across varying human brain regions using methylated DNA immunoprecipitation sequencing which is based in the affinity capture of methylated DNA eith an antibody specific to 5mC (Devall *et al.*, 2017).

Mitochondrial DNA methylation and hydroxymethylation can currently be measured through ELIsA, antibodies, BS-NGS and pyrosequencing. BS-NGS may be problematic as mitochondrial DNA has a circular structure, so it would have to be linearized before it could be sequenced. Other factors to consider are the high DNA input currently required to carry out these studies, as well as issues with cell-type isolation issues and the high cost and study complexity associated with mitochondrial DNA methylation studies (Coppedè and Stoccoro, 2019).

6.4.3 ALS/DNA methylation experimental models

The correlation between loss of both DNA methylation and hydroxymethylation from the nuclei of LMN and TDP43 pathology within the LMNs of SC in ALS has been established (section 4.5.2).

Within the MethylationEPIC data, GO analysis highlighted differential methylation within RNA metabolism (section 5.5.2.1). TDP43 pathology has been implicated in multiple RNA

metabolism processes, including splicing, transport, RNA stability (Prasad *et al.*, 2019). A third of transcribed RNAs contain TDP43 binding sites (Polymenidou *et al.*, 2011; Tollervey *et al.*, 2011), which recapitulates the key functions of TDP43 in RNA splicing and transport regulation. Combining this knowledge with the RNA dysregulation found within the MethylationEPIC dataset suggests that TDP43 could be involved with DNA methylation.

Currently, no cell model studies have been conducted into the relationship between TDP43 and DNA methylation. However, a study conducted by Štalekar *et al.*, 2015 did knockdown TDP43 in a neuroblastoma cell line, with DNMT3a being downregulated. These cells will have lost TDP from the nucleus, with loss of DNMT3a potentially indicating a loss of DNA methylation from the nucleus, as was seen in the IHC studies presented in section 4.5.2.

Lower expression of DNMT3a upon knockdown of TDP43 could represent loss of DNA methylation from the nucleus, as DNMTs are a key enzyme in the methylation of cytosine. This was observed in the LMNs of the SC in ALS, where loss of methylation and hydroxymethylation from the nucleus correlated with TDP43 pathology (section 4.5.2). TDP43's DNA binding roles are not well understood, but it is possible that it plays a role in DNA methylation, which is why the correlation between TDP43 pathology and loss of methylation and hydrozymethylation is observed. A consideration of the study conducted by Štalekar *et al.*, 2015 is that the validity of using neuroblastoma cell lines for DNA methylation studies into ALS have not been studied, and evidence presented in section 6.2.3 suggests that, following further studies to understand DNA methylation profiles, the most viable cell model to use for DNA methylation studies in ALS is neurones derived from iPSCs.

A cell culture of neurones derived from iPSCs, where TDP43 is knocked down and DNA methylation changes are detected would be useful to determine causality, and validate the findings of IHC studies presented in this thesis (section 4.5.2). Also conducting this in reverse, with DNA methylation levels manipulated through the use of DNMT inhibitors such as RG108 or approved DNA methylation-targeted drugs (Da Cotsa *et al.*, 2017), with analysis then conducted on levels of nuclear TDP43 would be useful. This would help to determine if there is a direct causation of DNA methylation and TDP43, or if there is another external factor affecting both TDP43 and DNA methylation. Further, it would be interesting to see if TDP43 binds to any of the dysmethylated genes identified in this study (chapter 5), as it has been found that TDP43 binds methyl CpG binding protein 2 (*MECP2*) RNA (Sephton *et al.*, 2011), which gives weight to the findings of chapter 4, that TDP43 proteinopathy associates with loss of methylation and hydroxymethylation from LMN nuclei.

6.5 Overall conclusions

Global hypermethylation is observed in LMNs of the ALS SC in both IHC and microarray studies. At the single gene-level, both hypermethylation and hypomethylation changes are present in ALS LMNs, with GO analysis implicating RNA metabolism dysregulation. Overlaps in hypermethylation/downregulation of gene expression and hypomethylation/upregulation of gene expression in the MethylationEPIC and mRNA expression datasets found changes in pathways involved in apoptosis, cell signalling and inflammation/immune response, all of which have been found to be dysregulated in PM tissues and in animal and cell models of ALS. Further, TDP43 pathology correlated with loss of both methylation and hydroxymethylation in LMNs. Combined with the RNA dysregulation found in the MethylationEPIC dataset, a role of TDP43 in DNA methylation is possible.

In general terms, DNA methylation has been shown to be altered in ALS in this study as well as others. This is also the case for other neurodegenerative diseases, suggesting possible overlaps in pathways that warrant further analysis. What has also been shown is that different cell types exhibit different levels of DNA methylation, so homogenates are potentially not good for determining each cell type's role, but are a good indicator of global changes occurring in tissue.

In conclusion, developing a clearer understanding of mechanisms of DNA methylation in ALS may be of importance to potential therapeutic treatments. Drugs specifically targeting epigenetic mechanisms are now on the market, and can be potentially of use in ALS.

References

Aggad, D. *et al.* (2014) 'TDP-43 toxicity proceeds via calcium dysregulation and necrosis in aging Caenorhabditis elegans motor neurons', *Journal of Neuroscience*, 34(36), pp. 12093–12103. doi: 10.1523/JNEUROSCI.2495-13.2014.

Alafuzoff, I. *et al.* (2006) 'Interlaboratory comparison of assessments of Alzheimer disease-related lesions: a study of the BrainNet Europe Consortium.', *Journal of Neuropathology & Experimental Neurology*, 65(8), pp. 740–757. doi: 10.1097/01.jnen.0000229986.17548.27.

Alexianu, M. E., Kozovska, M. and Appel, S. H. (2001) 'Immune reactivity in a mouse model of familial ALS correlates with disease progression', *Neurology*, 57(7), pp. 1282–1289. doi: 10.1212/wnl.57.7.1282.

Allen, C. F., Shaw, P. J. and Ferraiuolo, L. (2017) 'Can Astrocytes Be a Target for Precision Medicine?', in El-Khamisy, S. (ed.) *Personalised Medicine: Lessons from Neurodegeneration to Cancer*. Cham: Springer International Publishing, pp. 111–128. doi: 10.1007/978-3-319-60733-7_7.

Almer, G. *et al.* (2001) 'Increased expression of the pro-inflammatory enzyme cyclooxygenase-2 in amyotrophic lateral sclerosis', *Annals of Neurology*, 49(2), pp. 176–185. doi: 10.1002/1531-8249(20010201)49:2<176::AID-ANA37>3.0.CO;2-X.

Amarasinghe, S. L. *et al.* (2020) 'Opportunities and challenges in long-read sequencing data analysis', *Genome Biology*. Genome Biology, 21(1), pp. 1–16. doi: 10.1186/s13059-020-1935-5.

Arai, T. *et al.* (2006) 'TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis', *Biochemical and Biophysical Research Communications*, 351(3), pp. 602–611. doi: 10.1016/j.bbrc.2006.10.093.

Armstrong, G. A. B. *et al.* (2016) 'Homology directed knockin of point mutations in the zebrafish tardbp and fus genes in ALS using the CRISPR/Cas9 system', *PLoS ONE*, 11(3), pp. 1–10. doi: 10.1371/journal.pone.0150188.

Ash, P. E. A. *et al.* (2013) 'Unconventional Translation of C9ORF72 GGGGCC Expansion Generates Insoluble Polypeptides Specific to c9FTD/ALS', *Neuron*. Elsevier Inc., 77(4), pp. 639–646. doi: 10.1016/j.neuron.2013.02.004.

Assenov, Y. *et al.* (2014) 'Comprehensive analysis of DNA methylation data with RnBeads', 11(11). doi: 10.1038/nmeth.3115.

Ayala, Y. M. *et al.* (2011) 'TDP-43 regulates its mRNA levels through a negative feedback loop', *EMBO Journal*, 30(2), pp. 277–288. doi: 10.1038/emboj.2010.310.

Batra, R. *et al.* (2016) 'Gene Expression Signatures of Sporadic ALS Motor Neuron Populations', *BioRxiv*, p.038448.

Belzil, V. V., Nouri-mahdavi, K. and Raoof, D. A. (2008) 'Characterisation of DNA hypermethylation in the cerebellum of c9FTD/ALS patients', *Computer*, 144(5), pp. 724–732. doi: 10.1038/jid.2014.371.

Boche, D., Perry, V. H. and Nicoll, J. A. R. (2013) 'Review: Activation patterns of microglia and their identification in the human brain', *Neuropathology and Applied Neurobiology*, 39(1), pp. 3–18. doi: 10.1111/nan.12011.

de Boni, L. *et al.* (2018) 'DNA methylation alterations in iPSC- and hESC-derived neurons: Potential implications for neurological disease modeling', *Clinical Epigenetics*. Clinical Epigenetics, 10(1), pp. 1–13. doi: 10.1186/s13148-018-0440-0.

Booth, M. J. *et al.* (2013) 'Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine', *Nature Protocols*, 8(10), pp. 1841–1851. doi: 10.1038/nprot.2013.115.

Bosch, L. *et al.* (1999) 'Calcium Handling Proteins in Isolated Spinal Motoneurons', *Life Sciences*, 65(15). Available at: http://ac.els-cdn.com/S0024320599004051/1-s2.0-S0024320599004051-main.pdf?_tid=0d951068-70ea-11e7-9bec-00000aacb35d&acdnat=1500953802_85d3c08a1d09349b7025473244c5181a.

Böttcher, C. *et al.* (2019) 'Human microglia regional heterogeneity and phenotypes determined by multiplexed single-cell mass cytometry', *Nature Neuroscience*. Springer US, 22(1), pp. 78–90. doi: 10.1038/s41593-018-0290-2.

Bozzo, F. *et al.* (2016) 'Structural insights into the multi-determinant aggregation of TDP-43 in motor neuron-like cells', *Neurobiology of Disease*. Elsevier Inc., 94, pp. 63–72. doi: 10.1016/j.nbd.2016.06.006.

Bradley-Whitman, M. A. and Lovell, M. A. (2013) 'Epigenetic changes in the progression of Alzheimer's disease', *Mechanisms of Ageing and Development*. Elsevier Ireland Ltd, 134(10), pp. 486–495. doi: 10.1016/j.mad.2013.08.005.

Branco, M. R., Ficz, G. and Reik, W. (2012) 'Uncovering the role of 5-hydroxymethylcytosine in the epigenome.', *Nature reviews. Genetics*. Nature Publishing Group, 13(1), pp. 7–13. doi: 10.1038/nrg3080.

Brettschneider, J. *et al.* (2012) 'Microglial activation correlates with disease progression and upper motor neuron clinical symptoms in amyotrophic lateral sclerosis', *PLoS ONE*, 7(6), pp. 13–15. doi: 10.1371/journal.pone.0039216.

Brettschneider, J. *et al.* (2013) 'Stages of pTDP-43 pathology in amyotrophic lateral sclerosis', *Annals of Neurology*, 74(1), pp. 20–38. doi: 10.1002/ana.23937.

Brown, J. S. and Jackson, S. P. (2015) 'Ubiquitylation, neddylation and the DNA damage response', *Open Biology*, 5(4), pp. 150018–150018. doi: 10.1098/rsob.150018.

Burberry, A. *et al.* (2016) 'Loss-of-function mutations in the C9ORF72 mouse ortholog cause fatal autoimmune disease', *Science Translational Medicine*, 8(347). doi: 10.1126/scitranslmed.aaf6038.

Bursch, F. *et al.* (2019) 'Altered calcium dynamics and glutamate receptor properties in iPSC derived motor neurons from ALS patients with C9orf72, FUS, SOD1 or TDP43 mutations', *Human Molecular Genetics*, 28(17), pp. 2835–2850. doi: 10.1093/hmg/ddz107.

Butti, Z. and Patten, S. A. (2019) 'RNA Dysregulation in Amyotrophic Lateral Sclerosis', *Frontiers in Genetics*, 9(January), p. 712. doi: 10.3389/fgene.2018.00712.

Cady, J. *et al.* (2014) 'TREM2 variant p.R47H as a risk factor for sporadic amyotrophic lateral sclerosis', *JAMA Neurology*, 71(4), pp. 449–453. doi: 10.1001/jamaneurol.2013.6237.

Catania, J. and Fairweather, D. S. (1991) 'DNA methylation and cellular ageing', *Mutation Research DNAging*, 256(2–6), pp. 283–293. doi: 10.1016/0921-8734(91)90019-8.

Charcot, J. M. and Joffroy, A. (1869) 'Deux cas d'atrophie musculaire progressive: avec lésions de la substance grise et des faisceaux antérolatéraux de la moelle épinière.', *Masson.*

Chen, X. *et al.* (2014) 'Evidence for peripheral immune activation in amyotrophic lateral sclerosis', *Journal of the Neurological Sciences*. Elsevier B.V., 347(1–2), pp. 90–95. doi: 10.1016/j.jns.2014.09.025.

Chestnut, B. a. *et al.* (2011) 'Epigenetic Regulation of Motor Neuron Cell Death Through

DNA Methylation', *Journal of Neuroscience*, 31(46), pp. 16619–16636. doi: 10.1523/JNEUROSCI.1639-11.2011.Epigenetic.

Choi, J. *et al.* (2015) 'A comparison of genetically matched cell lines reveals the equivalence of human iPSCs and ESCs', *Nature Biotechnology*. Nature Publishing Group, 33(11), pp. 1173–1181. doi: 10.1038/nbt.3388.

Cohen, J. (1960) '&dquo;two-legged', XX(1), pp. 37–46.

Conerly, M. and Grady, W. M. (2010) 'Insights into the role of DNA methylation in disease through the use of mouse models', *DMM Disease Models and Mechanisms*, 3(5–6), pp. 290–297. doi: 10.1242/dmm.004812.

Cooper-Knock, J. *et al.* (2015) 'The Spectrum of C9orf72-mediated Neurodegeneration and Amyotrophic Lateral Sclerosis', *Neurotherapeutics*, 12(2), pp. 326–339. doi: 10.1007/s13311-015-0342-1.

Cooper-Knock, J. *et al.* (2017) 'A data-driven approach links microglia to pathology and prognosis in amyotrophic lateral sclerosis', *Acta neuropathologica communications*. Acta Neuropathologica Communications, 5(1), p. 23. doi: 10.1186/s40478-017-0424-x.

Coppedè, F. *et al.* (2017) 'Increase in DNA methylation in patients with amyotrophic lateral sclerosis carriers of not fully penetrant SOD1 mutations', *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*. Informa UK Limited, trading as Taylor & Francis Group, 0(0), pp. 1–9. doi: 10.1080/21678421.2017.1367401.

Coppedè, F. and Stoccoro, A. (2019) 'Mitoepigenetics and Neurodegenerative Diseases', *Frontiers in Endocrinology*, 10(February), pp. 1–7. doi: 10.3389/fendo.2019.00086.

Coppieters, N. *et al.* (2014) 'Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain', *Neurobiology of Aging*. Elsevier Ltd, 35(6), pp. 1334–1344. doi: 10.1016/j.neurobiolaging.2013.11.031.

Corcia, P. *et al.* (2012) 'Molecular Imaging of Microglial Activation in Amyotrophic Lateral Sclerosis', *PLoS ONE*, 7(12), pp. 6–12. doi: 10.1371/journal.pone.0052941.

Da Cotsa, E. *et al.* (2017) 'DNA Methylation-Targeted Drugs', *The cancer journal*, 23(5), pp. 270–276.

Cykowski, M. D. *et al.* (2018) 'Phosphorylated TDP-43 (pTDP-43) aggregates in the axial skeletal muscle of patients with sporadic and familial amyotrophic lateral sclerosis'. Acta Neuropathologica Communications, 43, pp. 1–15. doi: 10.1186/s40478-018-0528-y.

D'Arrigo, A. *et al.* (2010) 'Transcriptional profiling in the lumbar spinal cord of a mouse model of amyotrophic lateral sclerosis: A role for wild-type superoxide dismutase 1 in sporadic disease?', *Journal of Molecular Neuroscience*, 41(3), pp. 404–415. doi: 10.1007/s12031-010-9332-2.

DeJesus-Hernandez, M. *et al.* (2011) 'Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS', *Neuron*. Elsevier Inc., 72(2), pp. 245–256. doi: 10.1016/j.neuron.2011.09.011.

Devall, M. *et al.* (2017) 'Regional differences in mitochondrial DNA methylation in human post-mortem brain tissue', *Clinical Epigenetics*. Clinical Epigenetics, 9(1), p. 47. doi: 10.1186/s13148-017-0337-3.

Dillman, A. A. *et al.* (2017) 'Transcriptomic profiling of the human brain reveals that altered synaptic gene expression is associated with chronological aging', (November), pp. 1–12. doi: 10.1038/s41598-017-17322-0.

Ebbert, M. T. W. *et al.* (2017) 'Conserved DNA methylation combined with differential frontal

cortex and cerebellar expression distinguishes C9orf72-associated and sporadic ALS, and implicates SERPINA1 in disease', *Acta Neuropathologica*, pp. 1–14. doi: 10.1007/s00401-017-1760-4.

Ederle, H. and Dormann, D. (2017) 'TDP-43 and FUS en route from the nucleus to the cytoplasm', 591, pp. 1489–1507. doi: 10.1002/1873-3468.12646.

Ellison, EM, Abner, EL, Lovell, M. (2017) 'Multiregional analysis of global 5-methylcytosine and 5- hydroxymethylcytosine throughout the progression of Alzheimer's disease', 118(24), pp. 6072–6078. doi: 10.1002/cncr.27633.Percutaneous.

Endo, F. *et al.* (2015) 'Astrocyte-Derived TGF-β1 Accelerates Disease Progression in ALS Mice by Interfering with the Neuroprotective Functions of Microglia and T Cells', *Cell Reports*. The Authors, 11(4), pp. 592–604. doi: 10.1016/j.celrep.2015.03.053.

Engelhardt, J. I., Tajti, J. and Appel, S. H. (1993) 'Lymphocytic Infiltrates in the Spinal Cord in Amyotrophic Lateral Sclerosis', *Archives of Neurology*, 50(1), pp. 30–36. doi: 10.1001/archneur.1993.00540010026013.

van Es, M. A. *et al.* (2007) 'ITPR2 as a susceptibility gene in sporadic amyotrophic lateral sclerosis: a genome-wide association study', *Lancet Neurology*, 6(10), pp. 869–877. doi: 10.1016/S1474-4422(07)70222-3.

Eve, D. J., Dennis, J. S. and Citron, B. A. (2007) 'Transcription factor p53 in degenerating spinal cords', *Brain Research*, 1150(1), pp. 174–181. doi: 10.1016/j.brainres.2007.02.088.

Falcon, S. and Gentleman, R. (2007) 'Using GOstats to test gene lists for GO term association', *Bioinformatics*, 23(2), pp. 257–258. doi: 10.1093/bioinformatics/btl567.

Farg, M. A. *et al.* (2014) 'C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking', *Human Molecular Genetics*, 23(13), pp. 3579–3595. doi: 10.1093/hmg/ddu068.

Farg, M. A. *et al.* (2017) 'The DNA damage response (DDR) is induced by the C9orf72 repeat expansion in amyotrophic lateral sclerosis', *Human molecular genetics*, 26(15), pp. 2882–2896. doi: 10.1093/hmg/ddx170.

Faul, F. *et al.* (2007) 'G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences', *Behavior research methods*, 39(2), p. 175. doi: 10.3758/BF03193146.

Feng, J. *et al.* (2010) 'Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons.', *Nature neuroscience*, 13(4), pp. 423–30. doi: 10.1038/nn.2514.

Figueroa-Romero, C. *et al.* (2012) 'Identification of Epigenetically Altered Genes in Sporadic Amyotrophic Lateral Sclerosis', *PLoS ONE*, 7(12). doi: 10.1371/journal.pone.0052672.

Foulks, J. M. *et al.* (2012) 'Epigenetic Drug Discovery:Targeting DNA Methyltransferases', *Journal of Biomolecular Screening*, 17(1), pp. 2–17. doi: 10.1177/1087057111421212.

Friedman, B. A. *et al.* (2018) 'Diverse Brain Myeloid Expression Profiles Reveal Distinct Microglial Activation States and Aspects of Alzheimer's Disease Not Evident in Mouse Models', *Cell Reports*. ElsevierCompany., 22(3), pp. 832–847. doi: 10.1016/j.celrep.2017.12.066.

Fu, H., Hardy, J. and Duff, K. E. (2018) 'Selective vulnerability in neurodegenerative diseases', *Nature Neuroscience*, 21(10), pp. 1350–1358. doi: 10.1038/s41593-018-0221-2.

Fukushima, M. *et al.* (2019) 'TDP-43 accelerates deadenylation of target mRNAs by recruiting Caf1 deadenylase', 593, pp. 277–287. doi: 10.1002/1873-3468.13310.

Fuso, A. *et al.* (2011) 'Changes in Presenilin 1 gene methylation pattern in diet-induced B vitamin deficiency', *Neurobiology of Aging*. Elsevier Inc., 32(2), pp. 187–199. doi: 10.1016/j.neurobiolaging.2009.02.013.

Gasparoni, G. *et al.* (2018) 'DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex', *Epigenetics and Chromatin*. BioMed Central, 11(1), pp. 1–19. doi: 10.1186/s13072-018-0211-3.

Geser, F. *et al.* (2008) 'Evidence of multisystem disorder in whole-brain map of pathological TDP-43 in amyotrophic lateral sclerosis', *Archives of Neurology*, 65(5), pp. 636–641. doi: 10.1001/archneur.65.5.636.

Gleichmann, M. and Mattson, M. P. (2011) 'Neuronal calcium homeostasis and dysregulation', *Antioxidants and Redox Signaling*, 14(7), pp. 1261–1273. doi: 10.1089/ars.2010.3386.

Globisch, D. *et al.* (2010) 'Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates', *PLoS ONE*, 5(12), pp. 1–9. doi: 10.1371/journal.pone.0015367.

Goll, M. and Halpern, M. (2011) 'DNA Methylation in Zebrafish', 101(3), pp. 193–218. doi: 110.1016/j.bbi.2017.04.008.

Griñan-Ferré, C. *et al.* (2016) 'Environmental enrichment modified epigenetic mechanisms in SAMP8 mouse hippocampus by reducing oxidative stress and inflammaging and achieving neuroprotection', *Frontiers in Aging Neuroscience*, 8(OCT), pp. 1–12. doi: 10.3389/fnagi.2016.00241.

Group, T. W. and Group, S. (2017) 'Articles Safety and efficacy of edaravone in well defined patients with amyotrophic lateral sclerosis : a randomised ', 16(July). doi: 10.1016/S1474-4422(17)30115-1.

Gutstein, D. and Marks, A. (1997) 'Role of inositol 1,4,5-trisphosphate receptors in regulating apoptotic signaling and heart failure', *Heart and vessels*, pp. pp53-57.

Haeusler, A. R., Donnelly, C. J. and Rothstein, J. D. (2016) 'The expanding biology of the C9orf72 nucleotide repeat expansion in neurodegenerative disease', *Nat Rev Neurosci*. Nature Publishing Group, 17(6), pp. 383–395. doi: 10.1038/nrn.2016.38.

Hamzeiy, H. *et al.* (2018) 'Elevated Global DNA Methylation Is Not Exclusive to Amyotrophic Lateral Sclerosis and Is Also Observed in Spinocerebellar Ataxia Types 1 and 2', *Neurodegenerative Diseases*, 18(1), pp. 38–48. doi: 10.1159/000486201.

Harwood, C. A., McDermott, C. J. and Shaw, P. J. (2009) 'Physical activity as an exogenous risk factor in motor neuron disease (MND): A review of the evidence', *Amyotroph Lateral Scler*, 2968(September 2008), pp. 1–14. doi: 10.1080/17482960802549739.

Hernandez, D. G. *et al.* (2011) 'Distinct DNA methylation changes highly correlated with chronological age in the human brain', *Human Molecular Genetics*, 20(6), pp. 1164–1172. doi: 10.1093/hmg/ddq561.

Hewamaddumal, C. A. A. *et al.* (2013) 'Tardbpl splicing rescues motor neuron and axonal development in a mutant tardbp zebrafish', *Human Molecular Genetics*, 22(12), pp. 2376–2386. doi: 10.1093/hmg/ddt082.

Higashi, S, Kabuta, T, Nagai, Y, Tsuchiya, Y, Akiyama, H, Wada, K. (2013) 'Results TDP-43 associates with stalled ribosomes under', *Neurochemistry, Journal O F*, pp. 288–300. doi: 10.1111/jnc.12194.

Highley, J. R. *et al.* (2014) 'Loss of nuclear TDP-43 in amyotrophic lateral sclerosis (ALS)

causes altered expression of splicing machinery and widespread dysregulation of RNA splicing in motor neurones', pp. 670–685. doi: 10.1111/nan.12148.

Hill, S. J. *et al.* (2016) 'Two familial ALS proteins function in prevention/repair of transcription-associated DNA damage.', *Proceedings of the National Academy of Sciences of the United States of America*, p. 201611673. doi: 10.1073/pnas.1611673113.

Horvath, S. (2013) 'DNA methylation age of human tissues and cell types DNA methylation age of human tissues and cell types'.

Horvath, S. and Raj, K. (2018) 'DNA methylation-based biomarkers and the epigenetic clock theory of ageing', *Nature Reviews Genetics*. Springer US, 19, pp. 371–384. doi: 10.1038/s41576-018-0004-3.

Horvath, S. and Raj, K. (2019) 'DNA methylation-based biomarkers and the epigenetic clock theory of ageing', *Nature Reviews Genetics*. Springer US. doi: 10.1038/s41576-018-0004-3.

Howe, K. *et al.* (2013) 'The zebrafish reference genome sequence and its relationship to the human genome The generation of maps used in the initial assemblies and the production of clone tiling paths were carried out', *Nature*, 496(7446), pp. 498–503. doi: 10.1038/nature12111.

Huang, D. W. *et al.* (2009) 'Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources', *Cold Spring Harbor Perspectives in Biology*, 5(1), pp. a015073–a015073. doi: 10.1101/cshperspect.a015073.

Illingworth, R. S. and Bird, A. P. (2009) 'CpG islands - "A rough guide"', *FEBS Letters*. Federation of European Biochemical Societies, 583(11), pp. 1713–1720. doi: 10.1016/j.febslet.2009.04.012.

Iłżecka, J., Stelmasiak, Z. and Dobosz, B. (2002) 'Transforming growth factor-beta 1 (TGF-beta 1) in patients with amyotrophic lateral sclerosis', *Cytokine*, 20(5), pp. 239–243. doi: 10.1006/cyto.2002.2005.

Ito, S. *et al.* (2011) 'Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine.', *Science*, 333(6047), pp. 1300–3. doi: 10.1126/science.1210597.

Iwamoto, K. *et al.* (2011) 'Neurons show distinctive DNA methylation profile and higher interindividual variations compared with non-neurons', *Genome Research*, 21(5), pp. 688–696. doi: 10.1101/gr.112755.110.

Izumikawa, K. *et al.* (2017) 'TDP-43 stabilises the processing intermediates of mitochondrial transcripts', *Scientific Reports*. Springer US, (June), pp. 1–14. doi: 10.1038/s41598-017-06953-y.

De Jager, P. L. *et al.* (2014) 'Alzheimer's disease: Early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci', *Nature Neuroscience*. Nature Publishing Group, 17(9), pp. 1156–1163. doi: 10.1038/nn.3786.

Jaiswal, M. K. (2019) 'Riluzole and edaravone : A tale of two amyotrophic lateral sclerosis drugs', (July 2018), pp. 733–748. doi: 10.1002/med.21528.

Jiang, Y. M. *et al.* (2005) 'Gene expression profile of spinal motor neurons in sporadic amyotrophic lateral sclerosis', *Annals of Neurology*, 57(2), pp. 236–251. doi: 10.1002/ana.20379.

Jin, S. G. *et al.* (2011) 'Genomic mapping of 5-hydroxymethylcytosine in the human brain', *Nucleic Acids Research*, 39(12), pp. 5015–5024. doi: 10.1093/nar/gkr120.

Jung, H., Yoon, B. C. and Holt, C. E. (2012) 'Axonal mRNA localization and local protein synthesis in nervous system assembly, maintenance and repair', *Nature Reviews*

Neuroscience, 13(5), pp. 308–324. doi: 10.1038/nrn3210.

Kalsheker, N. A. (1996) 'α1-Antichymotrypsin', *International Journal of Biochemistry and Cell Biology*, 28(9), pp. 961–964. doi: 10.1016/1357-2725(96)00032-5.

Kamboh, M. *et al.* (2006) 'Alpha-1-antichymotrypsin (ACT or SERPINA3) polymorphism may affect age-at-onset and disease duration of Alzheimer's disease', 27(10), pp. 1435–1439. doi: 10.1109/TMI.2012.2196707.Separate.

Karemaker, I. D. and Vermeulen, M. (2018) 'Single-Cell DNA Methylation Profiling: Technologies and Biological Applications', *Trends in Biotechnology*. Elsevier Ltd, 36(9), pp. 952–965. doi: 10.1016/j.tibtech.2018.04.002.

Kessler, N. J. *et al.* (2016) 'CpG methylation differences between neurons and glia are highly conserved from mouse to human', *Human Molecular Genetics*, 25(2), pp. 223–232. doi: 10.1093/hmg/ddv459.

Kobow, K. *et al.* (2013) 'Deep sequencing reveals increased DNA methylation in chronic rat epilepsy', *Acta Neuropathologica*, 126(5), pp. 741–756. doi: 10.1007/s00401-013-1168-8.

Konopka, A. and Atkin, J. (2018) 'The Emerging Role of DNA Damage in the Pathogenesis of the C9orf72 Repeat Expansion in Amyotrophic Lateral Sclerosis', *International Journal of Molecular Sciences*, 19(10), p. 3137. doi: 10.3390/ijms19103137.

Kozlenkov, A. *et al.* (2018) 'A unique role for DNA (hydroxy)methylation in epigenetic regulation of human inhibitory neurons', *Science Advances*, 4(9). doi: 10.1126/sciadv.aau6190.

Kraus, T. F. J., Guibourt, V. and Kretzschmar, H. A. (2015) '5-Hydroxymethylcytosine, the ???Sixth Base???, during brain development and ageing', *Journal of Neural Transmission*. Springer Vienna, 122(7), pp. 1035–1043. doi: 10.1007/s00702-014-1346-4.

Krueger, F. and Andrews, S. R. (2011) 'Bismark : a flexible aligner and methylation caller for Bisulfite-Seq applications', 27(11), pp. 1571–1572. doi: 10.1093/bioinformatics/btr167.

Kuo, L. J. and Yang, L.-X. (2008) 'Gamma-H2AX - a novel biomarker for DNA double-strand breaks.', *In vivo (Athens, Greece)*, 22(3), pp. 305–9. doi: 0258-851X/2008.

Lagier-Tourenne, C. *et al.* (2013) 'Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration.', *Proceedings of the National Academy of Sciences of the United States of America*, 110(47), pp. E4530-9. doi: 10.1073/pnas.1318835110.

Landgrave-Gómez, J., Mercado-Gómez, O. and Guevara-Guzmán, R. (2015) 'Epigenetic mechanisms in neurological and neurodegenerative diseases', *Frontiers in Cellular Neuroscience*, 9(FEB), pp. 1–11. doi: 10.3389/fncel.2015.00058.

Li, C. *et al.* (2011) 'Astrocytes: Implications for Neuroinflammatory Pathogenesis of Alzheimers Disease', *Current Alzheimer Research*, 8(1), pp. 67–80. doi: 10.2174/156720511794604543.

Li, J. *et al.* (2017) 'Epigenetic targeting drugs potentiate chemotherapeutic effects in solid tumor therapy', *Scientific Reports*. Springer US, 7(1), pp. 1–13. doi: 10.1038/s41598-017-04406-0.

Li, Q. *et al.* (2017) 'Heterogeneous DNA methylation status in same-cell subpopulations of ovarian cancer tissues', *Tumor Biology*, 39(6), pp. 0–7. doi: 10.1177/1010428317701650.

Linker, S. M. *et al.* (2019) 'Combined single-cell profiling of expression and DNA methylation reveals splicing regulation and heterogeneity', *Genome Biology*. Genome Biology, 20(1), pp. 1–14. doi: 10.1186/s13059-019-1644-0.

Liu, J. and Siegmund, K. D. (2016) 'An evaluation of processing methods for HumanMethylation450 BeadChip data', *BMC Genomics*. BMC Genomics, 17(1), pp. 1–11. doi: 10.1186/s12864-016-2819-7.

Liu, J. and Wang, F. (2017) 'Role of neuroinflammation in amyotrophic lateral sclerosis: Cellular mechanisms and therapeutic implications', *Frontiers in Immunology*, 8(AUG), pp. 1–12. doi: 10.3389/fimmu.2017.01005.

Liu, X. S. *et al.* (2016) 'Editing DNA Methylation in the Mammalian Genome', *Cell*. Elsevier Inc., 167(1), pp. 233-247.e17. doi: 10.1016/j.cell.2016.08.056.

López, V., Fernández, A. F. and Fraga, M. F. (2017) 'The role of 5-hydroxymethylcytosine in development, aging and age-related diseases', *Ageing Research Reviews*. Elsevier B.V., 37, pp. 28–38. doi: 10.1016/j.arr.2017.05.002.

Lotti, F. *et al.* (2012) 'An SMN-dependent U12 splicing event essential for motor circuit function', *Cell*. Elsevier, 151(2), pp. 440–454. doi: 10.1016/j.cell.2012.09.012.

Love S, Budka, H, Ironside, JW, Perry, A. (2015) 'No Title', in *Greenfield's Neuropathology*. 9th edn, pp. 817–845.

Lu, H. *et al.* (2013) 'DNA methylation, a hand behind neurodegenerative diseases', *Frontiers in Aging Neuroscience*, 5(DEC), pp. 1–16. doi: 10.3389/fnagi.2013.00085.

Ludgate, J. L. *et al.* (2017) 'A streamlined method for analysing genome-wide DNA methylation patterns from low amounts of FFPE DNA', *BMC Medical Genomics*. BMC Medical Genomics, 10(1), p. 54. doi: 10.1186/s12920-017-0290-1.

Ludolph, A. *et al.* (2015) 'A revision of the El Escorial criteria - 2015', *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, (April), pp. 1–2. doi: 10.3109/21678421.2015.1049183.

Lunnon, K. *et al.* (2014) 'Cross-tissue methylomic profiling strongly implicates a role for cortex-specific deregulation of ANK1 in Alzheimer's disease neuropathology', *Nature neuroscience*, 17(9), pp. 1164–1170. doi: 10.1038/nn.3782.Cross-tissue.

Luo, C. *et al.* (2017) 'Elements in Mammalian Cortex', 604(August), pp. 600–604.

Machuca-tzili, L., Brook, D. and Hilton-jones, D. (2005) 'CLINICAL AND MOLECULAR ASPECTS OF THE MYOTONIC DYSTROPHIES : A REVIEW', 1(July), pp. 1–18. doi: 10.1002/mus.20301.

Mackenzie, I. R. A., Rademakers, R. and Neumann, M. (2010) 'TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia', *The Lancet Neurology*. Elsevier Ltd, 9(10), pp. 995–1007. doi: 10.1016/S1474-4422(10)70195-2.

Martin, D. *et al.* (2017) 'Neurocytometry: Flow Cytometric Sorting of Specific Neuronal Populations from Human and Rodent Brain', *ACS Chemical Neuroscience*, 8(2), pp. 356–367. doi: 10.1021/acschemneuro.6b00374.

Martin, L. J. (2000) 'p53 is abnormally elevated and active in the CNS of patients with amyotrophic lateral sclerosis', *Neurobiology of Disease*, 7(6), pp. 613–622. doi: 10.1006/nbdi.2000.0314.

Mastroeni, D. *et al.* (2010) 'Epigenetic changes in Alzheimer's disease: decrements in DNA methylation', *PLoS ONE*, 32(7), pp. 736–740. doi: 10.1371/journal.pone.0178059.

Mastroeni, D. *et al.* (2011) 'Epigenetic mechanisms in Alzheimer's disease', *Neurobiology of Aging*, 32(7), pp. 1161–1180. doi: 10.1016/j.neurobiolaging.2010.08.017.

McCauley, M. E. and Baloh, R. H. (2019) 'Inflammation in ALS/FTD pathogenesis', *Acta

Neuropathologica. Springer Berlin Heidelberg, 137(5), pp. 715–730. doi: 10.1007/s00401-018-1933-9.

McEwen, L. M. *et al.* (2018) 'Systematic evaluation of DNA methylation age estimation with common preprocessing methods and the Infinium MethylationEPIC BeadChip array', *Clinical Epigenetics*. Clinical Epigenetics, 10(1), pp. 1–9. doi: 10.1186/s13148-018-0556-2.

McLean, C. Y. *et al.* (2010) 'GREAT improves functional interpretation of cis-regulatory regions', *Nature Biotechnology*. Nature Publishing Group, 28(5), pp. 495–501. doi: 10.1038/nbt.1630.

Merlo, D. *et al.* (2016) 'DNA Damage, Neurodegeneration, and Synaptic Plasticity', *Neural Plasticity*, 2016, pp. 1–3. doi: 10.1155/2016/1206840.

Mertens, J. *et al.* (2015) 'Directly Reprogrammed Human Neurons Retain AgingAssociated Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects', *Cell Stem Cel*, 17(6), pp. 705–718. doi: 10.1097/CCM.0b013e31823da96d.Hydrogen.

Mi, H. *et al.* (2019) 'PANTHER version 14 : more genomes , a new PANTHER GO-slim and improvements in enrichment analysis tools'. Oxford University Press, 47(November 2018), pp. 419–426. doi: 10.1093/nar/gky1038.

Mitra, J. *et al.* (2019) 'Motor neuron disease-associated loss of nuclear TDP-43 is linked to DNA double-strand break repair defects', *Proceedings of the National Academy of Sciences*, p. 201818415. doi: 10.1073/pnas.1818415116.

Moens, T. G. *et al.* (2019) 'C9orf72 arginine-rich dipeptide proteins interact with ribosomal proteins in vivo to induce a toxic translational arrest that is rescued by eIF1A', *Acta Neuropathologica*, 137(3), pp. 487–500. doi: 10.1007/s00401-018-1946-4.

Morahan, J. M. *et al.* (2009) 'A genome-wide analysis of brain DNA methylation identifies new candidate genes for sporadic amyotrophic lateral sclerosis.', *Amyotrophic lateral sclerosis : official publication of the World Federation of Neurology Research Group on Motor Neuron Diseases*, 10(5–6), pp. 418–429. doi: 10.3109/17482960802635397.

Morrice, J. R., Gregory-Evans, C. Y. and Shaw, C. A. (2018) 'Animal models of amyotrophic lateral sclerosis: A comparison of model validity', *Neural Regeneration Research*, 13(12), pp. 2050–2054. doi: 10.4103/1673-5374.241445.

Müller, F. *et al.* (2019) 'RnBeads 2 . 0 : comprehensive analysis of DNA methylation data'. Genome Biology, pp. 1–12.

Neal, M. and Richardson, J. R. (2018) 'Epigenetic regulation of astrocyte function in neuroinflammation and neurodegeneration', *Biochimica et Biophysica Acta - Molecular Basis of Disease*. Elsevier, 1864(2), pp. 432–443. doi: 10.1016/j.bbadis.2017.11.004.

Nelson, P. T. *et al.* (2019) 'Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report', *Brain : a journal of neurology*, 142(6), pp. 1503–1527. doi: 10.1093/brain/awz099.

Neumann, Manuela *et al.* (2006) 'Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis.', *Science*, 314(5796), pp. 130–133. doi: 10.1126/science.1134108.

Neumann, M. *et al.* (2006) 'Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis', *Science*, 314(5796), pp. 130–133. doi: 10.1126/science.1134108.

Nussbacher, J. K. *et al.* (2015) 'RNA binding proteins in neurodegeneration: Seq and you shall receive', *Trends in Neurosciences*, 38(4), pp. 226–236. doi: 10.1007/1-4020-0613-

6_15620.

Nussbacher, J. K. *et al.* (2019) 'Disruption of RNA Metabolism in Neurological Diseases and Emerging Therapeutic Interventions', *Neuron.* Elsevier Inc., 102(2), pp. 294–320. doi: 10.1016/j.neuron.2019.03.014.

Oh, Y. S., Kim, S. H. and Cho, G.-W. (2016) 'Functional Restoration of Amyotrophic Lateral Sclerosis Patient-Derived Mesenchymal Stromal Cells Through Inhibition of DNA Methyltransferase', *Cellular and Molecular Neurobiology.* Springer US, 36(4), pp. 613–620. doi: 10.1007/s10571-015-0242-2.

Orr, B. A. *et al.* (2012) 'Decreased 5-Hydroxymethylcytosine is associated with neural progenitor phenotype in normal brain and shorter survival in malignant glioma', *PLoS ONE*, 7(7). doi: 10.1371/journal.pone.0041036.

Ou, S. H. *et al.* (1995) 'Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs', *J Virol*, 69(6), pp. 3584–3596.

Pehar, M. *et al.* (2018) 'Role and Therapeutic Potential of Astrocytes in Amyotrophic Lateral Sclerosis', *Current Pharmaceutical Design*, 23(33), pp. 5010–5021. doi: 10.2174/1381612823666170622095802.

Peters, O. M., Ghasemi, M. and Brown Jr., R. H. (2015) 'Emerging mechanisms of molecular pathology in ALS Find the latest version : Emerging mechanisms of molecular pathology in ALS', *The Journal of Clinical Investigation*, 125(5), pp. 1767–1779. doi: https://doi.org/10.1172/JCI71601.

Peters, T. *et al.* (2015) 'De novo identification of differentially methylated regions in the human genome', *Epigenetics & Chromatin*, 8(6), pp. 150–159. doi: 10.1111/j.1467-842X.1988.tb00471.x.

Petrov, D. *et al.* (2017) 'ALS Clinical Trials Review : 20 Years of Failure . Are We Any Closer to Registering a New Treatment ?', 9(March), pp. 1–11. doi: 10.3389/fnagi.2017.00068.

Pfeifer, G. P., Kadam, S. and Jin, S.-G. (2013) '5-Hydroxymethylcytosine and Its Potential Roles in Development and Cancer.', *Epigenetics & chromatin*, 6(1), p. 10. doi: 10.1186/1756-8935-6-10.

Pidsley, R. *et al.* (2016) 'Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling', *Genome Biology.* Genome Biology, 17(1), pp. 1–17. doi: 10.1186/s13059-016-1066-1.

Pita-Juárez, Y. *et al.* (2018) 'The Pathway Coexpression Network: Revealing pathway relationships', *PLoS Computational Biology*, 14(3). doi: 10.1371/journal.pcbi.1006042.

Polymenidou, M. *et al.* (2011) 'Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43', *Nature Neuroscience*, 14(4), pp. 459–468. doi: 10.1097/CCM.0b013e31823da96d.Hydrogen.

Pook, M. a (2012) 'DNA methylation and trinucleotide repeat expansion diseases', *DNA Methylation - From Genomics to Technology*, 1, pp. 193–208. doi: 10.5772/34169.

Prasad, A. *et al.* (2019) 'Molecular Mechanisms of TDP-43 Misfolding and Pathology in Amyotrophic Lateral Sclerosis', *Frontiers in Molecular Neuroscience*, 12(February), pp. 1–36. doi: 10.3389/fnmol.2019.00025.

Price, A. *et al.* (2018) 'Divergent neuronal DNA methylation patterns across human cortical development: Critical periods and a unique role of CpH methylation', *Biological Sciences - Neuroscience*.

Qazi, T. J. *et al.* (2018) 'Epigenetics in Alzheimer's Disease: Perspective of DNA Methylation', *Molecular Neurobiology*. Molecular Neurobiology, 55(2), pp. 1026–1044. doi: 10.1007/s12035-016-0357-6.

Rademakers, R. (2012) 'C9orf72 repeat expansions in patients with ALS and FTD', *The Lancet Neurology*. Elsevier Ltd, 11(4), pp. 297–298. doi: 10.1016/S1474-4422(12)70046-7.

Rao, J. S. *et al.* (2012) 'Epigenetic modifications in frontal cortex from Alzheimer's disease and bipolar disorder patients', *Translational Psychiatry*. Nature Publishing Group, 2(7), pp. e132-7. doi: 10.1038/tp.2012.55.

Rg, M. *et al.* (2012) 'Riluzole for amyotrophic lateral sclerosis (ALS)/ motor neuron disease (MND) (Review)', (3). doi: 10.1002/14651858.CD001447.pub3.www.cochranelibrary.com.

Richardson, B. (2003) 'Impact of aging on DNA methylation', *Ageing Research Reviews*, 2(3), pp. 245–261. doi: 10.1016/S1568-1637(03)00010-2.

Rizzardi, L. F. *et al.* (2019) 'Neuronal brain-region-specific DNA methylation and chromatin accessibility are associated with neuropsychiatric trait heritability', *Nature Neuroscience*. Springer US, 22(2), pp. 307–316. doi: 10.1038/s41593-018-0297-8.

Robertson, K. D. and Wolffe, A. P. (2000) 'DNA methylation in health and disease', *Nature Reviews Genetics*, 1(1), pp. 11–19. doi: 10.1038/35049533.

Romero, J. P. *et al.* (2018) 'Comparison of RNA-seq and microarray platforms for splice event detection using a cross-platform algorithm', *BMC Genomics*. BMC Genomics, 19(1), pp. 1–14. doi: 10.1186/s12864-018-5082-2.

Rossor, A. M. *et al.* (2019) 'TDP43 pathology in the brain, spinal cord, and dorsal root ganglia of a patient with FOSMN', *Neurology*, 92(9), pp. E951–E956. doi: 10.1212/WNL.0000000000007008.

Roubroeks, J. A. Y. *et al.* (2017) 'Epigenetics and DNA methylomic profiling in Alzheimer's disease and other neurodegenerative diseases', *Journal of Neurochemistry*, 143(2), pp. 158–170. doi: 10.1111/jnc.14148.

Rubin, Harvey, N. and Halim, Mostafa, N. (1993) 'WHY , WHEN SHORTEN AND HOW DOES THE POLY (A) DURING mRNA TRANSLATION ? TAIL', *Interniational Journal of Biochemistry*, 25(3), pp. 287–295.

Saberi, S. *et al.* (2016) 'HHS Public Access', 33(4), pp. 855–876. doi: 10.1016/j.ncl.2015.07.012.

Schwob NG, Nalbantoglu J, Hastings KE, Mikkelsen T, C. N. (1990) 'DNA cytosine methylation in brain of patients with Alzheimer's disease.', *Ann Neurol*, 28:91–94.

Scotter, E. L., Chen, H. J. and Shaw, C. E. (2015) 'TDP-43 Proteinopathy and ALS: Insights into Disease Mechanisms and Therapeutic Targets', *Neurotherapeutics*, 12(2), pp. 352–363. doi: 10.1007/s13311-015-0338-x.

Sephton, C. F. *et al.* (2011) 'Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes', *Journal of Biological Chemistry*, 286(2), pp. 1204–1215. doi: 10.1074/jbc.M110.190884.

Shanbhag, N. M. *et al.* (2019) 'Early neuronal accumulation of DNA double strand breaks in Alzheimer's disease', *Acta Neuropathologica Communications*. Acta Neuropathologica Communications, 7(1), pp. 1–18. doi: 10.1186/s40478-019-0723-5.

Simpson, J. T. *et al.* (2017) 'Detecting DNA cytosine methylation using nanopore sequencing', *Nature Methods*. Nature Publishing Group, 14(4), pp. 407–410. doi: 10.1038/nmeth.4184.

Smith, A. R. *et al.* (2019) 'A cross-brain regions study of ANK1 DNA methylation in different neurodegenerative diseases', *Neurobiology of Aging*. Elsevier Inc, 74, pp. 70–76. doi: 10.1016/j.neurobiolaging.2018.09.024.

Smith, E. F., Shaw, P. J. and De Vos, K. J. (2017) 'The role of mitochondria in amyotrophic lateral sclerosis', *Neuroscience Letters*. Elsevier Ireland Ltd, (xxxx), p. 132933. doi: 10.1016/j.neulet.2017.06.052.

Smith, R. G. *et al.* (2018) 'Elevated DNA methylation across a 48-kb region spanning the HOXA gene cluster is associated with Alzheimer's disease neuropathology', *Alzheimer's and Dementia*. Elsevier Inc., 14(12), pp. 1580–1588. doi: 10.1016/j.jalz.2018.01.017.

Song, C. X. *et al.* (2011) 'Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine', *Nature Biotechnology*, 29(1), pp. 68–75. doi: 10.1038/nbt.1732.

Song, S. W. *et al.* (2016) 'Major histocompatibility complex class I molecules protect motor neurons from astrocyte-induced toxicity in amyotrophic lateral sclerosis', *Nature Medicine*. Nature Publishing Group, 22(4), pp. 397–403. doi: 10.1038/nm.4052.

Štalekar, M. *et al.* (2015) 'Proteomic analyses reveal that loss of TDP-43 affects RNA processing and intracellular transport', *Neuroscience*. IBRO, 293, pp. 157–170. doi: 10.1016/j.neuroscience.2015.02.046.

Staszewski, O. and Prinz, M. (2014) 'Glial epigenetics in neuroinflammation and neurodegeneration', *Cell and Tissue Research*, 356(3), pp. 609–616. doi: 10.1007/s00441-014-1815-y.

Sudria-Lopez, E. *et al.* (2016) 'Full ablation of C9orf72 in mice causes immune system-related pathology and neoplastic events but no motor neuron defects', *Acta Neuropathologica*. Springer Berlin Heidelberg, 132(1), pp. 145–147. doi: 10.1007/s00401-016-1581-x.

Sullivan, P. M. *et al.* (2016) 'The ALS/FTLD associated protein C9orf72 associates with SMCR8 and WDR41 to regulate the autophagy-lysosome pathway', *Acta neuropathologica communications*. Acta Neuropathologica Communications, 4(1), p. 51. doi: 10.1186/s40478-016-0324-5.

Takeshima, H. *et al.* (2006) 'Distinct DNA methylation activity of Dnmt3a and Dnmt3b towards naked and nucleosomal DNA', *Journal of Biochemistry*, 139(3), pp. 503–515. doi: 10.1093/jb/mvj044.

Takeuchi, S. *et al.* (2010) 'Induction of protective immunity by vaccination with wild-type apo superoxide dismutase 1 in mutant SOD1 transgenic mice', *Journal of Neuropathology and Experimental Neurology*, 69(10), pp. 1044–1056. doi: 10.1097/NEN.0b013e3181f4a90a.

Tammen, S. A., Friso, S. and Choi, S. W. (2013) 'Epigenetics: The link between nature and nurture', *Molecular Aspects of Medicine*, 34(4), pp. 753–764. doi: 10.1016/j.mam.2012.07.018.

Tan, R. H. *et al.* (2015) 'TDP-43 proteinopathies: Pathological identification of brain regions differentiating clinical phenotypes', *Brain*, 138(10), pp. 3110–3122. doi: 10.1093/brain/awv220.

Tank, E. M. *et al.* (2018) 'Abnormal RNA stability in amyotrophic lateral sclerosis', *Nature Communications*, 9(1). doi: 10.1038/s41467-018-05049-z.

Taskesen, E. *et al.* (2017) 'Susceptible genes and disease mechanisms identified in frontotemporal dementia and frontotemporal dementia with Amyotrophic Lateral Sclerosis by DNA-methylation and GWAS', *Scientific Reports*, 7(1), p. 8899. doi: 10.1038/s41598-017-09320-z.

Tateno, M. *et al.* (2004) 'Calcium-permeable AMPA receptors promote misfolding of mutant SOD1 protein and development of amyotrophic lateral sclerosis in a transgenic mouse model', *Human Molecular Genetics*, 13(19), pp. 2183–2196. doi: 10.1093/hmg/ddh246.

Teichroeb, J. H., Betts, D. H. and Vaziri, H. (2011) 'Suppression of the imprinted gene NNAT and X-Chromosome gene activation in isogenic human iPS cells', *PLoS ONE*, 6(10). doi: 10.1371/journal.pone.0023436.

Tollervey, J. R. *et al.* (2011) 'Characterising the RNA targets and position-dependent splicing regulation by TDP-43 ; implications for neurodegenerative diseases', 14(4), pp. 452–458. doi: 10.1038/nn.2778.Characterising.

Tradewell, M. L. *et al.* (2011) 'Calcium dysregulation, mitochondrial pathology and protein aggregation in a culture model of amyotrophic lateral sclerosis: Mechanistic relationship and differential sensitivity to intervention', *Neurobiology of Disease*. Elsevier Inc., 42(3), pp. 265–275. doi: 10.1016/j.nbd.2011.01.016.

Valle, C. and Carrì, M. T. (2017) 'Cysteine Modifications in the Pathogenesis of ALS', *Frontiers in Molecular Neuroscience*, 10(January), p. 5. doi: 10.3389/fnmol.2017.00005.

Vazquez-Villaseñor, I. *et al.* (2019) ' Expression of p16 and p21 in the frontal association cortex of ALS / MND brains suggest neuronal cell cycle dysregulation and astrocyte senescence in early stages of the disease ', *Neuropathology and Applied Neurobiology*, pp. 1–15. doi: 10.1111/nan.12559.

Villar-Menéndez, I. *et al.* (2013) 'Increased 5-methylcytosine and decreased 5-hydroxymethylcytosine levels are associated with reduced striatal A2AR levels in Huntington's disease', *NeuroMolecular Medicine*, 15(2), pp. 295–309. doi: 10.1007/s12017-013-8219-0.

Vogler, T. O. *et al.* (2019) 'muscle', 563(7732), pp. 508–513. doi: 10.1038/s41586-018-0665-2.TDP-43.

Waite, A. J. *et al.* (2014) 'Reduced C9orf72 protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal degeneration brain with the C9ORF72 hexanucleotide repeat expansion', *Neurobiology of Aging*. Elsevier Ltd, 35(7), pp. 1779.e5-1779.e13. doi: 10.1016/j.neurobiolaging.2014.01.016.

Walker, C. *et al.* (2017) 'C9orf72 expansion disrupts ATM-mediated chromosomal break repair', *Nature Neuroscience*. Nature Publishing Group, (June), pp. 1–14. doi: 10.1038/nn.4604.

Wan Nasri, W. N. *et al.* (2019) 'Tocotrienol Rich Fraction Supplementation Modulate Brain Hippocampal Gene Expression in APPswe/PS1dE9 Alzheimer's Disease Mouse Model', *Journal of Alzheimer's Disease*, 70, pp. 1–16. doi: 10.3233/jad-180496.

Wang, C. *et al.* (2014) 'The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance', *Nature Biotechnology*, 32(9), pp. 926–932. doi: 10.1038/nbt.3001.

Wang, F. *et al.* (2013) 'Genome-wide loss of 5-hmC is a novel epigenetic feature of huntington's disease', *Human Molecular Genetics*, 22(18), pp. 3641–3653. doi: 10.1093/hmg/ddt214.

Wang, S. Y. *et al.* (2015) 'Notch pathway is activated in cell culture and mouse models of mutant SOD1-related familial amyotrophic lateral sclerosis, with suppression of its activation as an additional mechanism of neuroprotection for lithium and valproate', *Neuroscience*, 301, pp. 276–288. doi: 10.1016/j.neuroscience.2015.06.002.

Wang, Y. C. *et al.* (2001) 'Genetic association analysis of alpha-1-antichymotrypsin

polymorphism in Parkinson's disease', *European Neurology*, 45(4), pp. 254–256. doi: 10.1159/000052138.

Watson, C. T. *et al.* (2016) 'Genome-wide12 DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease', *Genome Medicine*. Genome Medicine, 8(1), pp. 1–14. doi: 10.1186/s13073-015-0258-8.

Webster, C. P. *et al.* (2016) ' The C9orf72 protein interacts with Rab1a and the ULK 1 complex to regulate initiation of autophagy ', *The EMBO Journal*, 35(15), pp. 1656–1676. doi: 10.15252/embj.201694401.

Wilson$, V. L. *et al.* (1987) 'THE JOURNAL OF BIOLOGICAL CHEMISTRY Genomic 5-Methyldeoxycytidine Decreases with Age*', 262(21), p. 99469951.

Xi, Z. *et al.* (2012) 'Investigation of c9orf72 in 4 neurodegenerative disorders.', *Archives of neurology*, 69(12), pp. 1583–90. doi: 10.1001/archneurol.2012.2016.

Xi, Z. *et al.* (2013) 'Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion', *American Journal of Human Genetics*. The American Society of Human Genetics, 92(6), pp. 981–989. doi: 10.1016/j.ajhg.2013.04.017.

Xia, N. *et al.* (2016) 'Transcriptional comparison of human induced and primary midbrain dopaminergic neurons', *Scientific Reports*. Nature Publishing Group, 6(February), pp. 1–9. doi: 10.1038/srep20270.

Yamashita, T. and Kwak, S. (2014) 'The molecular link between inefficient GluA2 Q/R site-RNA editing and TDP-43 pathology in motor neurons of sporadic amyotrophic lateral sclerosis patients', *Brain Research*. Elsevier, 1584, pp. 28–38. doi: 10.1016/j.brainres.2013.12.011.

Yang, D. *et al.* (2015) 'FTD/ALS-associated poly(GR) protein impairs the Notch pathway and is recruited by poly(GA) into cytoplasmic inclusions', *Acta Neuropathologica*. Springer Berlin Heidelberg, 130(4), pp. 525–535. doi: 10.1007/s00401-015-1448-6.

Yang, M. *et al.* (2016) 'A C9ORF72/SMCR8-containing complex regulates ULK1 and plays a dual role in autophagy', *Science Advances*, 2(9), pp. 1–18. doi: 10.1126/sciadv.1601167.

Yu, M. *et al.* (2012) 'Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine', *Nature Protocols*, 12(7), pp. 2159–2170. doi: 10.1038/jid.2014.371.

Zarei, S., Carr, K., Reiley, L., Diaz, K., Guerra, O., Altamirano, P. F., Pagani, W., Lodin, D., Orozco, G., Chinea, A. (2015) 'A comprehensive review of amyotrophic lateral sclerosis', *Surgical neurology international*, 6.

Zhang, L. *et al.* (2012) 'Thymine DNA glycosylase specifically recognises 5-carboxylcytosine-modified DNA', 8(4), pp. 328–330. doi: 10.1038/nchembio.914.Thymine.

Zinszner, H. *et al.* (1997) 'TLS (FUS) binds RNA in vivo and engages in nucleo-cytoplasmic shuttling.', *Journal of cell science*, 1750(15), pp. 1741–50. doi: 10.1038/SJ.ONC.1200854.

Appendices

Appendix I: Ethical approval (16/006, 16/007, 18/007).

**AUTHORISATION TO USE TISSUE RESOURCE FROM THE
SHEFFIELD BRAIN TISSUE BANK (SBTB)**

FOLLOWING CONSIDERATION BY THE SBTB MANAGEMENT BOARD:

Proposed Study Title Control blocks to be used for histochemistry optimisation studies at SITraN

Head of proposed study

Title   Dr   Initials   JR

Surname   Highley

Position.   Senior Clinical Lecturer in Neuropathology

Organisation.   SITraN

Address.   385a Glossop Road, Sheffield, S10 2HQ

Telephone No   +44 (0) 114 222 2244   Fax No.   +44(0)114 222 2290

Email   robin.highley@sheffield.ac.uk

SBTB PROJECT REQUEST NUMBER: 16/006

This project was reviewed by the SBTB Management Board and approval to release tissue
under REC   **08/MRE00/103** was granted.

Professor P G Ince
Director SBTB

Date: 3/10/16

SBTB tissue authority / ver1: 01/02/2008

AUTHORISATION TO USE TISSUE RESOURCE FROM THE SHEFFIELD BRAIN TISSUE BANK (SBTB)

FOLLOWING CONSIDERATION BY THE SBTB MANAGEMENT BOARD:

Proposed Study Title   DNA pathology in Amyotrophic Lateral Sclerosis

Head of proposed study

Title   Dr   Initials   JR

Surname   Highley

Position.   Senior Clinical Lecturer in Neuropathology

Organisation.   SITraN

Address.   385a Glossop Road, Sheffield, S10 2HQ

Telephone No   +44 (0) 114 222 2244   Fax No.   +44(0)114 222 2290

Email   robin.highley@sheffield.ac.uk

SBTB PROJECT REQUEST NUMBER: 16/007

This project was reviewed by the SBTB Management Board and approval to release tissue under REC   **08/MRE00/103** was granted.

Professor P G Ince
Director SBTB

Date: 3/10/16

212

AUTHORISATION TO USE TISSUE RESOURCE FROM THE SHEFFIELD BRAIN TISSUE BANK (SBTB)

FOLLOWING CONSIDERATION BY THE SBTB MANAGEMENT BOARD:

Proposed Study Title: Generation of Tissue Microarrays for neurodegeneration research using brain bank material.

Head of proposed study

Title: Dr   Initials: JR

Surname: Highley

Position: Senior clinical Lecturer in Neuropathology

Organisation: SITraN

Address: 385a Glossop Road, Sheffield, S10 2HQ

Telephone No: 0114 222 2244   Fax No: 0114 222 2290

Email: robin.highley@sheffield.ac.uk

SBTB PROJECT REQUEST NUMBER: 18/007

This project was reviewed by the SBTB Management Board and approval to release tissue under REC **08/MRE00/103** was granted.

Professor Stephen Wharton
Professor of Neuropathology and
Consultant Neuropatholpogist

Date: 25/6/18.

Appendix II: Laboratory solution recipes

Laboratory recipes : Histology solutions and reagents	
TBS (1x)	6.05g (50mM) Tris buffer (Fisher Scientific, Loughborough, UK) and 8.76g (150mM) Sodium chloride (Merck,) was made up to 1L in d.H2O, pH to 7.6.
Scott's tap water	8.75g sodium bicarbonate and 50.0g magnesium sulphate were dissolved in 2.5L water
Acid/Alcohol	20ml concentrate hydrochloric acid was diluted in 1780 mL 70% ethanol
EDTA 1mM (antigen retrieval)	0.15g Ethylenediaminetetraacetic acid (EDTA) was diluted in 500 mL distilled water, pH to 8.0
TSC (antigen retrieval)	1.5g tri-sodium citrate (TSC) was diluted in 500 mL distilled water, pH to 6.5
Toluidine blue (TB) 0.01%	1g toluidine blue in 1000 mL distilled water

Appendix III: Examples of positive and negative controls for 5mC, 5hmC and TDP43 immunohistochemistry
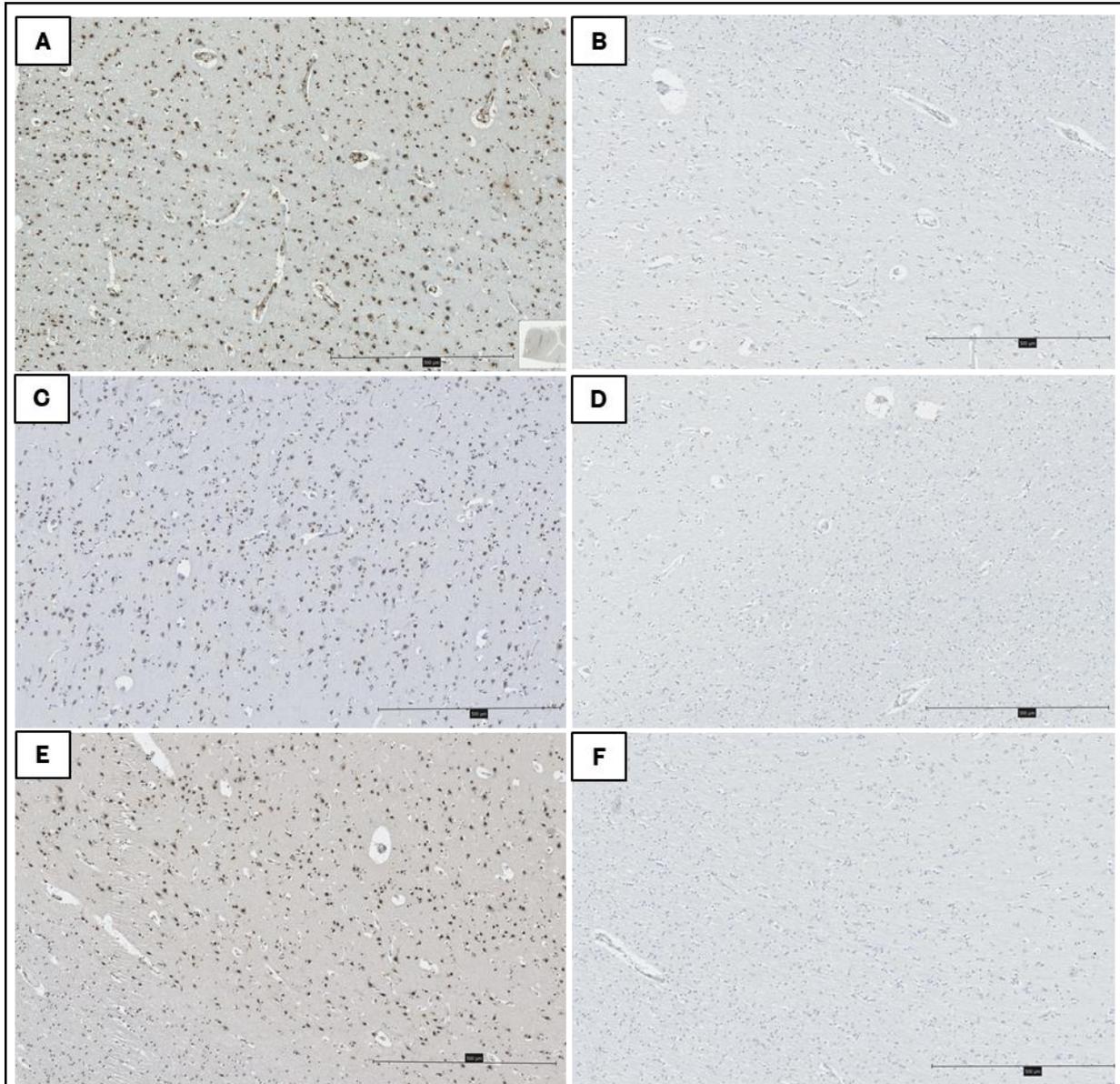
Examples of positive and negative controls for 5mC, 5hmC and TDP43 immunohistochemistry.
Images taken at x40 magnification, scale bar represents 500 μm.
All positive control examples highlight the nuclear-specific immunopositive staining for 5mC (A), 5hmC (C) and TDP43 (E). All negative controls show no immunopositive staining, with only blue haematoxylin counterstaining observed for 5mC (B), 5hmC (D) and TDP43 (F).

Appendix IV: Example of laser capture microdissection

Example of motor neurone extraction from anterior horn of spinal cord using laser capture microdissection.
Motor neurones (MNs) stained with toluidine blue are extracted from the anterior horn of cervical spinal cord using laser capture microdissection. MNs are indicated with red arrows, absence of MNs are indicated with blue arrows.
A) MNs are visualised within the tissue using toluidine blue staining. B) MNs targeted with infra red lasers are melted onto the LCM cap. A 'halo' (example circled in green) can be seen around these MNs, indicating the LCM cap film has adhered to the targeted MNs. C) Gaps in the tissue can be visualised where the MNs have been extracted. D) The targeted MNs can be visualised on the LCM cap.

Appendix V: Preparation of reagents for MethylationEPIC

Preparation of reagents for MethylationEPIC array		
Step	**Preparation of reagents**	**Preparation of equipment**
Amplification of bisulphite converted DNA	0.1 N NaOH: 5 mL 1N NaOH, plus 45 mL sterile water, invert to mix.	Preheat Illumina hybridisation oven to 37 °C.
Fragmentation of DNA	Thaw FMS to RT, and invert to mix.	Preheat heat block to 37 °C.
DNA precipitation and resuspension	Place paper towels on lab bench.	Set centrifuge to 4 °C.
Hybridisation of DNA to BeadChip	Thaw RA1 to RT, and invert to mix.	Preheat Illumina hybridisation oven to 48 °C, and set to 'rocker speed 5. Turn on heat sealer to warm up. Set heat block to 95 °C.
Washing, extension and staining of BeadChip	XC4: Add 330 mL 100% ethanol to XC4 to give a total volume of 350 mL and shake vigorously to mix. Thaw the following reagents to RT: XC1, XC2, TEM, STM, ATM. 95% Formamide/1 mM EDTA.	Preheat water cycler to 44 °C.
BeadChip imaging		Set up an Illumina account. Turn on HiScan scanner to calibrate. Download 'decode file client user guide (1337856)'. Download the BeadChip manifest file from Illumina. Locate the 'BCD manifest.csv' file. Open the 'decode.file client' software.

Appendix VI: MethylationEPIC solutions

Table depicting the chemical components and function of reagents/solutions used in the MethylationEPIC BeadChip array			
Name	Components	Stage	Reason/function
PB1	Sodium chloride	Prepare BeadChips for hybridisation	Hybridisation: forming dsDNA by joining two complementary ssDNA strands together NaCl aids this hybridisation as they help stabilise the DNA, with the salt ion charges reducing the electrostatic repulsion between the phosphate charges on DNA
PB2	Polypropylene glycol, water	Humidifying buffer for after hybridisation	Used as a humectant (opposite of a desiccant) to keep samples from drying out. Attracts and retains moisture from the air via absorption
XC1	Sodium phosphate dibasic	Xstain BeadChip solution	pH adjustment in buffer
XC2	Sucrose, glycerol	Xstain BeadChip solution	Glycerol and sucrose makes the sample more dense than the sample buffer, so the sample will remain in the bottom of a well rather than floating out. The difference between glycerol and sucrose is based upon their viscosity, Glycerol is superior to sucrose as a viscosity agent but Sucrose is better than glycerol for increasing density, because sucrose is less sticky.
XC3	Sodium hydroxide	Xstain BeadChip solution	Removal of stain
XC4	Water	Xstain BeadChip solution	Washing
MA1	Mineral oil	Multi-sample amplification mix 1	Minimise sample evaporation
FMS	Sucrose, glycerol, sodium chloride, potassium chloride	Fragmentation solution	Sucrose and glycerol (see XC2) Sodium chloride (see PB1) Potassium chloride-induces fragmentation in DNA
PM1	Ammonium acetate, water	Precipitation solution	Preferred salt in high dNTP's and oligosaccharides content solutions as these remain in solution. Avoid using if kinasing as ammonium ions inhibit polynucleotide kinase
TEM	Sucrose, glycerol, sodium hydroxide, potassium chloride, polyethylene glycol tert-octylphenyl ether (Triton X-100), magnesium chloride, potassium phosphate monobasic	Two-colour extension master mix	Sucrose and glycerol (see XC2) Sodium hydroxide (see XC3) Potassium chloride (see FMS) Magnesium chloride (see MSM) Trition X-100 is a detergent used to lyse cells, or to permeabilise cell membranes, solubilises membrane proteins, can form part of lysis buffer Potassium phosphate monobasic-buffering agent
ATM	Sucrose, sodium phosphate dibasic	Anti-stain two-colour master mix	Sucrose (see XC2) Sodium phosphate dibasic (see XC1)
STM	Sucrose, sodium phosphate dibasic	Superior two-colour master mix	Sucrose (see XC2) Sodium phosphate dibasic (see XC1)
RA1	Formamide, polyoxyethylenes orbitan monolaurate TWEEN?), potassium phosphate monobasic, sodium chloride	Resuspension, hybridisation, and wash solution	Formamide-cryoprotectant for tissue preservation, RNA stabiliser TWEEN? (see MSM) potassium phosphate monobasic-(see XC1) Sodium chloride (see PB1)
RPM	Random primer mix	Random primer mix	To allow amplification to take place

Appendix VII: Summary of controls used in MethylationEPIC

Summary of the controls used in the MethylationEPIC array

Control type	Purpose	Control dependency
Staining	To monitor staining step efficiency	Sample independent
Hybridisation	Using synthetic reference targets to monitor overall hybridisation performance. Targets are at low, medium and high concentrations	Sample independent
Extension	To monitor the efficiency of efficiency for A, T, C and G from a hairpin probe	Sample independent
Target removal	To monitor the stripping step efficiency after the extension phase	Sample independent
Bisulphite conversion I	To monitor the efficiency of bisulphite conversion for type I probes	Sample dependent
Bisulphite conversion II	To monitor the efficiency of bisulphite conversion for type II probes	Sample dependent
Specificity I	To monitor the allele specific extension of type I probes	Sample dependent
Specificity II	To monitor the allele specific extension of type II probes	Sample dependent
Non-polymorphic	This monitors the efficiency of all steps by querying a non-polymorphic base in the genome	Sample dependent
Negative	The negative control probes are used to estimate background signal for red and green channels and to monitor signal intensities at bisulphite converted sequences that have no CpGs present. The detection p-value for each probe is based on the measured intensities at the negative control probes	Sample dependent
Normalisation	This is for normalisation, whereby the controls are used to calculate scaling factors	Sample dependent

Appendix VIII: CSV file for MethylationEPIC analysis using RnBeads

1	[Header]						
2	Investigator Name	Scientist					
3	Project Name	Methylation					
4	Experiment Name	Methylation experiment					
5	Date						
6							
7	[Data]						
8	Sample_Name	Sample_Plate	Sample_Group	Pool_ID	Project	Sample_V	Sentrix_ID	Sentrix_Position
9	con1		Control				201465940042	R01C01
10	con2		Control				201465940042	R01C02
11	con3		Control				201465940042	R01C03
12	c91		C9ALS				201465940042	R01C04
13	c92		C9ALS				201465940042	R01C05
14	c93		C9ALS				201465940042	R01C06
15	s1		sALS				201465940042	R01C07
16	s2		sALS				201465940042	R01C08
17	s3		sALS				201465940028	R01C01
18	s4		sALS				201465940028	R01C02
19	s5		sALS				201465940028	R01C03
20	c94		C9ALS				201465940028	R01C04
21	c95		C9ALS				201465940028	R01C05
22	c96		C9ALS				201465940028	R01C06
23	con4		Control				201465940028	R01C07
24	con5		Control				201465940028	R01C08
25							

CSV file created in Excel detailing the case IDs, groupings and position of samples on the BeadChips, to allow for RnBeads analysis of the MethylationEPIC array.

Appendix IX: RnBeads script

Conducted on 08/02/2019

```
source(http://bioconductor.org/biocLite.R)
biocLite(c("RnBeads","RnBeads.hg38","RnBeads.hg19"))
library (RnBeads)
install.packages("Shiny")
install.packages("doParallel")
install.packages("devtools")
install.packages("RPMM")
devtools::install_github("daattali/shinyjs")
suppressPackageStartupMessages(library(RnBeads))
rnb.rub.dj()
```

2019-02-08 11:18:04	1.2	STATUS	STARTED RnBeads Pipeline
2019-02-08 11:18:04	1.2	INFO	Analysis Title: RnBeads Analysis
2019-02-08 11:18:05	1.2	INFO	Initialized report index and saved to index.html
2019-02-08 11:18:05	1.2	STATUS	STARTED Loading Data
2019-02-08 11:18:05	1.2	INFO	Number of cores: 8
2019-02-08 11:18:05	1.2	INFO	Loading data of type "idat.dir"
2019-02-08 11:18:05	1.2	STATUS	STARTED Loading Data from IDAT Files
2019-02-08 11:18:05	1.2	INFO	Added column barcode to the provided sample annotation table
2019-02-08 11:18:05	1.2	INFO	Detected platform: MethylationEPIC
2019-02-08 11:18:27	1.5	STATUS	COMPLETED Loading Data from IDAT Files
2019-02-08 11:20:13	1.0	STATUS	Loaded data from C:\Users\UOS\Documents\idat
2019-02-08 11:20:14	1.5	STATUS	Predicted sex for the loaded samples
2019-02-08 11:20:15	1.4	STATUS	Added data loading section to the report
2019-02-08 11:20:15	1.4	STATUS	Loaded 16 samples and 866895 sites
2019-02-08 11:20:15	1.4	INFO	Output object is of type RnBeadRawSet
2019-02-08 11:20:15	1.4	STATUS	COMPLETED Loading Data
2019-02-08 11:20:21	1.0	WARNING	Zip not found on this Windows system, this RnBSet object will not be saved. See the instructions for installing ZIP on Windows in the FAQ section of the RnBeads website.
2019-02-08 11:20:21	1.0	INFO	Initialized report index and saved to index.html
2019-02-08 11:20:27	1.0	STATUS	STARTED Quality Control
2019-02-08 11:20:27	1.0	INFO	Number of cores: 8
2019-02-08 11:20:27	1.0	STATUS	STARTED Quality Control Section
2019-02-08 11:20:52	1.2	STATUS	Added quality control box plots
2019-02-08 11:22:51	1.2	STATUS	Added quality control bar plots
2019-02-08 11:22:54	1.2	STATUS	Added negative control boxplots
2019-02-08 11:22:54	1.2	STATUS	COMPLETED Quality Control Section
2019-02-08 11:22:54	1.2	STATUS	STARTED Visualizing SNP Probe Data
2019-02-08 11:22:54	1.2	STATUS	STARTED Mixups Visualization Section
2019-02-08 11:23:00	1.4	STATUS	Added SNP Heatmap
2019-02-08 11:23:00	1.4	STATUS	Calculated Manhattan distances between samples based on SNP probes
2019-02-08 11:23:01	1.4	STATUS	Added SNP-based Distances
2019-02-08 11:23:01	1.4	STATUS	COMPLETED Mixups Visualization Section
2019-02-08 11:23:01	1.4	STATUS	COMPLETED Visualizing SNP Probe Data
2019-02-08 11:23:05	1.5	STATUS	COMPLETED Quality Control
2019-02-08 11:23:05	1.5	INFO	Initialized report index and saved to index.html
2019-02-08 11:23:11	1.0	STATUS	STARTED Preprocessing
2019-02-08 11:23:12	1.0	INFO	Number of cores: 8
2019-02-08 11:23:12	1.0	STATUS	STARTED Filtering Procedures I

2019-02-08 11:23:13   1.5 STATUS   STARTED Removal of SNP-enriched Sites
2019-02-08 11:23:13   1.5 STATUS   Removed 17371 sites using SNP criterion "3"
2019-02-08 11:23:13   1.5 STATUS   Saved removed sites to C:\Users\UOS\Documents\rnbeads_report\preprocessing_data/removed_sites_snp.csv
2019-02-08 11:23:13   1.5 STATUS   Added a corresponding section to the report
2019-02-08 11:23:13   1.5 STATUS   COMPLETED Removal of SNP-enriched Sites
2019-02-08 11:23:19   1.1 STATUS   Retained 16 samples and 849524 sites
2019-02-08 11:23:19   1.1 STATUS   COMPLETED Filtering Procedures I
2019-02-08 11:23:19   1.1 STATUS   STARTED Summary of Filtering Procedures I
2019-02-08 11:23:20   1.3 STATUS   Created summary table of removed sites, samples and unreliable measurements
2019-02-08 11:23:21   1.3 STATUS   Added summary table of removed and retained items
2019-02-08 11:23:21   1.3 INFO   Subsampling 117371 sites for plotting density distributions
2019-02-08 11:23:21   1.5 STATUS   Constructed sequences of removed and retained methylation values
2019-02-08 11:23:30   1.5 STATUS   Added comparison between removed and retained beta values
2019-02-08 11:23:30   1.5 STATUS   COMPLETED Summary of Filtering Procedures I
2019-02-08 11:23:30   1.5 STATUS   STARTED Manipulating the object
2019-02-08 11:25:35   1.5 STATUS   Removed 17371 sites (probes)
2019-02-08 11:25:35   1.5 INFO   Retained 849524 sites and 16 samples
2019-02-08 11:25:35   1.5 STATUS   COMPLETED Manipulating the object
2019-02-08 11:25:35   1.5 STATUS   STARTED Normalization Procedure
2019-02-08 11:29:40   1.6 STATUS   Performed normalization with method bmiq
2019-02-08 11:31:12   1.3 STATUS   Performed normalization with method "bmiq"
2019-02-08 11:31:24   1.5 STATUS   Added comparison between non-normalized and normalized beta values
2019-02-08 11:31:25   2.2 STATUS   Added histogram of observed beta shifts (magnitude of correction)
2019-02-08 11:31:26   2.2 STATUS   Added 2D histogram of observed beta values and shifts
2019-02-08 11:31:27   2.1 STATUS   Added normalization section
2019-02-08 11:31:27   2.1 STATUS   COMPLETED Normalization Procedure
2019-02-08 11:31:33   1.4 STATUS   STARTED Filtering Procedures II
2019-02-08 11:31:34   1.8 STATUS   STARTED Probe Context Removal
2019-02-08 11:31:34   1.8 STATUS   Removed 2985 probe(s) having not acceptable context
2019-02-08 11:31:34   1.8 STATUS   Saved removed sites to C:\Users\UOS\Documents\rnbeads_report\preprocessing_data/removed_sites_context.csv
2019-02-08 11:31:34   1.8 STATUS   Added a corresponding section to the report
2019-02-08 11:31:34   1.8 STATUS   COMPLETED Probe Context Removal
2019-02-08 11:31:34   1.8 STATUS   STARTED Removal of Sites on Sex Chromosomes
2019-02-08 11:31:34   1.9 STATUS   Removed 19438 site(s) on sex chromosomes
2019-02-08 11:31:34   1.9 STATUS   Saved removed sites to C:\Users\UOS\Documents\rnbeads_report\preprocessing_data/removed_sites_sex.csv
2019-02-08 11:31:34   1.9 STATUS   Added a corresponding section to the report
2019-02-08 11:31:34   1.9 STATUS   COMPLETED Removal of Sites on Sex Chromosomes
2019-02-08 11:31:35   1.9 STATUS   Retained 16 samples and 827101 sites
2019-02-08 11:31:35   1.9 STATUS   COMPLETED Filtering Procedures II

2019-02-08 11:31:35   1.9 STATUS   STARTED Summary of Filtering Procedures II
2019-02-08 11:31:41   1.6 STATUS   Created summary table of removed sites,
samples and unreliable measurements
2019-02-08 11:31:42   1.6 STATUS   Added summary table of removed and retained
items
2019-02-08 11:31:42   1.7 INFO   Subsampling 122423 sites for plotting density
distributions
2019-02-08 11:31:42   1.8 STATUS   Constructed sequences of removed and
retained methylation values
2019-02-08 11:31:52   1.5 STATUS   Added comparison between removed and
retained beta values
2019-02-08 11:31:52   1.5 STATUS   COMPLETED Summary of Filtering Procedures II
2019-02-08 11:31:52   1.5 STATUS   STARTED Manipulating the object
2019-02-08 11:33:55   1.5 STATUS   Removed 22423 sites (probes)
2019-02-08 11:33:55   1.5 INFO   Retained 827101 sites and 16 samples
2019-02-08 11:33:55   1.5 STATUS   COMPLETED Manipulating the object
2019-02-08 11:33:55   1.5 INFO   Imputation was skipped, data set may still contain
missing methylation values
2019-02-08 11:34:01   1.2 STATUS   COMPLETED Preprocessing
2019-02-08 11:34:07   1.1 WARNING   Zip not found on this Windows system, this
RnBSet object will not be saved. See the instructions for installing ZIP on Windows in the
FAQ section of the RnBeads website.
2019-02-08 11:34:07   1.1 INFO   Initialized report index and saved to index.html
2019-02-08 11:34:13   1.1 STATUS   STARTED Tracks and Tables
2019-02-08 11:34:13   1.1 INFO   Number of cores: 8
2019-02-08 11:34:14   1.1 STATUS   STARTED Generating Tracks and Tables
2019-02-08 11:34:14   1.1 STATUS   STARTED Exporting sites
2019-02-08 11:34:14   1.1 STATUS   STARTED Creating BED Files
2019-02-08 11:34:14   1.1 STATUS   Converting to GRangesList
2019-02-08 11:34:19   1.6 STATUS   Exporting sample con1
2019-02-08 11:34:23   1.5 STATUS   Exporting sample con2
2019-02-08 11:34:27   1.7 STATUS   Exporting sample con3
2019-02-08 11:34:31   1.6 STATUS   Exporting sample c91
2019-02-08 11:34:32   1.6 STATUS   Exporting sample c92
2019-02-08 11:34:36   1.5 STATUS   Exporting sample c93
2019-02-08 11:34:39   1.7 STATUS   Exporting sample s1
2019-02-08 11:34:43   1.6 STATUS   Exporting sample s2
2019-02-08 11:34:44   1.5 STATUS   Exporting sample s3
2019-02-08 11:34:48   1.7 STATUS   Exporting sample s4
2019-02-08 11:34:52   1.7 STATUS   Exporting sample s5
2019-02-08 11:34:55   1.7 STATUS   Exporting sample c94
2019-02-08 11:34:59   1.7 STATUS   Exporting sample c95
2019-02-08 11:35:03   1.6 STATUS   Exporting sample c96
2019-02-08 11:35:04   1.7 STATUS   Exporting sample con4
2019-02-08 11:35:08   1.7 STATUS   Exporting sample con5
2019-02-08 11:35:12   1.5 STATUS   COMPLETED Creating BED Files
2019-02-08 11:35:12   1.5 STATUS   STARTED Creating Track Hub -- bigBed
2019-02-08 11:35:12   1.5 WARNING   Skipped conversion bed -> bigBed for
region type sites; could not find the tool bedToBigBed
2019-02-08 11:35:12   1.5 STATUS   COMPLETED Creating Track Hub -- bigBed
2019-02-08 11:35:12   1.5 STATUS   STARTED Creating UCSC Track Hub --
bigWig
2019-02-08 11:35:12   1.5 WARNING   Skipped conversion bedGraph -> bigWig
for region type sites; could not find the tool bedGraphToBigWig

2019-02-08 11:35:12   1.5 STATUS   COMPLETED Creating UCSC Track Hub --
bigWig
2019-02-08 11:35:12   1.5 STATUS   COMPLETED Exporting sites
2019-02-08 11:35:12   1.5 STATUS   COMPLETED Generating Tracks and Tables
2019-02-08 11:35:12   1.5 STATUS   STARTED Writing export report
2019-02-08 11:35:29   1.3 STATUS   COMPLETED Writing export report
2019-02-08 11:35:30   1.3 STATUS   COMPLETED Tracks and Tables
2019-02-08 11:35:30   1.3 INFO   Initialized report index and saved to index.html
2019-02-08 11:35:30   1.3 STATUS   STARTED Covariate Inference
2019-02-08 11:35:30   1.3 INFO   Number of cores: 8
2019-02-08 11:35:31   1.4 STATUS   STARTED Age Prediction using predefined
predictor
2019-02-08 11:35:34   1.7 STATUS   COMPLETED Age Prediction using predefined
predictor
2019-02-08 11:35:34   1.7 STATUS   STARTED Adding Age Prediction Section to
Report
2019-02-08 11:35:52   1.5 STATUS   Added Comparison Plot
2019-02-08 11:35:55   1.5 STATUS   Added Error Plot
2019-02-08 11:35:55   1.5 STATUS   COMPLETED Adding Age Prediction Section to
Report
2019-02-08 11:35:56   1.6 STATUS   Calculated LUMP estimates
2019-02-08 11:35:56   1.6 STATUS   COMPLETED Covariate Inference
2019-02-08 11:35:56   1.6 WARNING   Zip not found on this Windows system, this
RnBSet object will not be saved. See the instructions for installing ZIP on Windows in the
FAQ section of the RnBeads website.
2019-02-08 11:35:56   1.6 INFO   Initialized report index and saved to index.html
2019-02-08 11:36:02   1.1 STATUS   STARTED Exploratory Analysis
2019-02-08 11:36:02   1.1 INFO   Number of cores: 8
2019-02-08 11:36:03   1.1 STATUS   Designed color mappings for probe type and CGI
status
2019-02-08 11:36:24   1.6 STATUS   STARTED Dimension Reduction Techniques
2019-02-08 11:36:35   1.6 INFO   Mapping 3 traits to point colors and types
2019-02-08 11:43:41   1.2 INFO   Principal components that explain at least 95 %
of the total variance: 13
2019-02-08 11:43:41   1.2 INFO   Saved percentage of total variance to
pca_variance_explained_1.csv
2019-02-08 11:43:42   1.2 INFO   Principal components that explain at least 95 %
of the total variance: 11
2019-02-08 11:43:42   1.2 INFO   Saved percentage of total variance to
pca_variance_explained_2.csv
2019-02-08 11:43:43   1.2 INFO   Principal components that explain at least 95 %
of the total variance: 10
2019-02-08 11:43:43   1.2 INFO   Saved percentage of total variance to
pca_variance_explained_3.csv
2019-02-08 11:43:44   1.2 INFO   Principal components that explain at least 95 %
of the total variance: 9
2019-02-08 11:43:44   1.2 INFO   Saved percentage of total variance to
pca_variance_explained_4.csv
2019-02-08 11:43:45   1.2 INFO   Principal components that explain at least 95 %
of the total variance: 12
2019-02-08 11:43:45   1.2 INFO   Saved percentage of total variance to
pca_variance_explained_5.csv
2019-02-08 11:43:46   1.2 STATUS   Created scatter plots and CDFs summarizing
the reduced dimensional representations
2019-02-08 11:43:46   1.2 STATUS   COMPLETED Dimension Reduction Techniques

2019-02-08 11:43:46   1.2   STATUS   STARTED Tests for Associations
2019-02-08 11:43:46   1.2   INFO   Testing the following traits for associations:
Sample_Group; Age; Sentrix_ID; Sentrix_Position; Predicted Male Probability; Genome-wide
methylation; predicted_ages; age_increase; Immune Cell Content (LUMP)
2019-02-08 11:43:46   1.2   STATUS   Created 10000 sample permutations
2019-02-08 11:46:30   1.2   STATUS   Computed correlations between principal
components and traits.
2019-02-08 11:46:41   1.2   STATUS   Computed pairwise correlations between traits.
2019-02-08 11:46:41   1.2   STATUS   COMPLETED Tests for Associations
2019-02-08 11:46:56   1.7   STATUS   STARTED Methylation Value Distributions -
Sample Groups
2019-02-08 11:47:13   1.7   STATUS   COMPLETED Methylation Value Distributions -
Sample Groups
2019-02-08 11:47:13   1.7   STATUS   STARTED Methylation Value Distributions -
Probe Categories
2019-02-08 11:48:08   1.8   STATUS   COMPLETED Methylation Value Distributions -
Probe Categories
2019-02-08 11:48:08   1.8   STATUS   STARTED Sample Clustering
2019-02-08 11:48:08   1.8   STATUS   STARTED Agglomerative Hierarchical
Clustering
2019-02-08 11:48:10   1.3   STATUS   Performed clustering on sites using
correlation as a distance metric
2019-02-08 11:48:11   1.3   STATUS   Performed clustering on sites using
manhattan as a distance metric
2019-02-08 11:48:12   1.3   STATUS   Performed clustering on sites using
euclidean as a distance metric
2019-02-08 11:48:12   1.3   STATUS   Performed clustering on promoters using
correlation as a distance metric
2019-02-08 11:48:12   1.3   STATUS   Performed clustering on promoters using
manhattan as a distance metric
2019-02-08 11:48:12   1.3   STATUS   Performed clustering on promoters using
euclidean as a distance metric
2019-02-08 11:48:13   1.5   STATUS   Performed clustering on tiling using
correlation as a distance metric
2019-02-08 11:48:13   1.5   STATUS   Performed clustering on tiling using
manhattan as a distance metric
2019-02-08 11:48:13   1.5   STATUS   Performed clustering on tiling using
euclidean as a distance metric
2019-02-08 11:48:13   1.6   STATUS   Performed clustering on genes using
correlation as a distance metric
2019-02-08 11:48:13   1.6   STATUS   Performed clustering on genes using
manhattan as a distance metric
2019-02-08 11:48:13   1.6   STATUS   Performed clustering on genes using
euclidean as a distance metric
2019-02-08 11:48:13   1.6   STATUS   Performed clustering on cpgislands using
correlation as a distance metric
2019-02-08 11:48:13   1.6   STATUS   Performed clustering on cpgislands using
manhattan as a distance metric
2019-02-08 11:48:13   1.6   STATUS   Performed clustering on cpgislands using
euclidean as a distance metric
2019-02-08 11:48:13   1.6   STATUS   COMPLETED Agglomerative Hierarchical
Clustering
2019-02-08 11:48:14   1.6   STATUS   STARTED Clustering Section
2019-02-08 11:48:14   1.6   STATUS   STARTED Generating Heatmaps
2019-02-08 11:48:14   1.7   STATUS   STARTED Region type: sites

2019-02-08 11:49:43   1.4 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage average . Discarding row dendrogram
2019-02-08 11:49:48   1.4 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage average . Discarding row dendrogram
2019-02-08 11:49:53   1.5 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage average . Discarding row dendrogram
2019-02-08 11:49:58   1.5 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage average . Discarding row dendrogram
2019-02-08 11:50:02   1.5 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage average . Discarding row dendrogram
2019-02-08 11:50:07   1.6 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage average . Discarding row dendrogram
2019-02-08 11:50:12   1.6 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage average . Discarding row dendrogram
2019-02-08 11:50:17   1.6 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage average . Discarding row dendrogram
2019-02-08 11:50:22   1.7 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage average . Discarding row dendrogram
2019-02-08 11:50:27   1.7 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage complete . Discarding row dendrogram
2019-02-08 11:50:31   1.8 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage complete . Discarding row dendrogram
2019-02-08 11:50:36   1.8 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage complete . Discarding row dendrogram
2019-02-08 11:50:41   1.4 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage complete . Discarding row dendrogram
2019-02-08 11:50:46   1.4 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage complete . Discarding row dendrogram
2019-02-08 11:50:50   1.5 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage complete . Discarding row dendrogram
2019-02-08 11:50:55   1.5 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage complete . Discarding row dendrogram
2019-02-08 11:51:00   1.5 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage complete . Discarding row dendrogram
2019-02-08 11:51:05   1.6 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage complete . Discarding row dendrogram
2019-02-08 11:51:10   1.6 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage median . Discarding row dendrogram
2019-02-08 11:51:15   1.6 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage median . Discarding row dendrogram
2019-02-08 11:51:19   1.7 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage median . Discarding row dendrogram
2019-02-08 11:51:24   1.7 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage median . Discarding row dendrogram
2019-02-08 11:51:29   1.8 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage median . Discarding row dendrogram
2019-02-08 11:51:34   1.8 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage median . Discarding row dendrogram
2019-02-08 11:51:38   1.4 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage median . Discarding row dendrogram
2019-02-08 11:51:43   1.4 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage median . Discarding row dendrogram
2019-02-08 11:51:48   1.5 WARNING   Could not cluster rows of the heatmap
with method manhattan and linkage median . Discarding row dendrogram

2019-02-08 11:51:53   1.5 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage average . Discarding row dendrogram
2019-02-08 11:51:58   1.5 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage average . Discarding row dendrogram
2019-02-08 11:52:03   1.6 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage average . Discarding row dendrogram
2019-02-08 11:52:07   1.6 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage average . Discarding row dendrogram
2019-02-08 11:52:12   1.6 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage average . Discarding row dendrogram
2019-02-08 11:52:17   1.7 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage average . Discarding row dendrogram
2019-02-08 11:52:22   1.7 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage average . Discarding row dendrogram
2019-02-08 11:52:27   1.8 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage average . Discarding row dendrogram
2019-02-08 11:52:31   1.8 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage average . Discarding row dendrogram
2019-02-08 11:52:36   1.4 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage complete . Discarding row dendrogram
2019-02-08 11:52:41   1.4 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage complete . Discarding row dendrogram
2019-02-08 11:52:46   1.5 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage complete . Discarding row dendrogram
2019-02-08 11:52:50   1.5 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage complete . Discarding row dendrogram
2019-02-08 11:52:55   1.5 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage complete . Discarding row dendrogram
2019-02-08 11:53:00   1.6 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage complete . Discarding row dendrogram
2019-02-08 11:53:05   1.6 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage complete . Discarding row dendrogram
2019-02-08 11:53:10   1.7 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage complete . Discarding row dendrogram
2019-02-08 11:53:15   1.7 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage complete . Discarding row dendrogram
2019-02-08 11:53:20   1.7 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage median . Discarding row dendrogram
2019-02-08 11:53:24   1.8 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage median . Discarding row dendrogram
2019-02-08 11:53:29   1.8 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage median . Discarding row dendrogram
2019-02-08 11:53:34   1.4 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage median . Discarding row dendrogram
2019-02-08 11:53:39   1.4 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage median . Discarding row dendrogram
2019-02-08 11:53:44   1.5 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage median . Discarding row dendrogram
2019-02-08 11:53:48   1.5 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage median . Discarding row dendrogram
2019-02-08 11:53:53   1.6 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage median . Discarding row dendrogram
2019-02-08 11:53:58   1.6 WARNING   Could not cluster rows of the heatmap with method euclidean and linkage median . Discarding row dendrogram
2019-02-08 11:54:03   1.6 STATUS   COMPLETED Region type: sites

2019-02-08 11:54:03	1.6	STATUS	STARTED Region type: promoters
2019-02-08 11:57:20	1.3	STATUS	COMPLETED Region type: promoters
2019-02-08 11:57:20	1.4	STATUS	STARTED Region type: tiling
2019-02-08 12:00:42	1.4	STATUS	COMPLETED Region type: tiling
2019-02-08 12:00:42	1.4	STATUS	STARTED Region type: genes
2019-02-08 12:04:00	1.4	STATUS	COMPLETED Region type: genes
2019-02-08 12:04:00	1.4	STATUS	STARTED Region type: cpgislands
2019-02-08 12:07:23	1.4	STATUS	COMPLETED Region type: cpgislands
2019-02-08 12:07:23	1.4	STATUS	Created 540 heatmaps based on the clustering results
2019-02-08 12:07:23	1.4	STATUS	COMPLETED Generating Heatmaps
2019-02-08 12:07:23	1.4	STATUS	STARTED Adding Color Legends
2019-02-08 12:08:22	1.4	STATUS	COMPLETED Adding Color Legends
2019-02-08 12:08:22	1.4	STATUS	STARTED Estimating Optimal Numbers of Clusters
2019-02-08 12:08:39	1.4	STATUS	Estimated number of clusters based on mean silhouette value
2019-02-08 12:08:39	1.4	STATUS	COMPLETED Estimating Optimal Numbers of Clusters
2019-02-08 12:08:39	1.4	STATUS	STARTED Overlapping Clusters with Sample Traits
2019-02-08 12:08:40	1.4	STATUS	Computed adjusted rand indices and saved to exploratory_analysis_data/adjusted_rand_indices_1.csv
2019-02-08 12:08:40	1.4	STATUS	Computed adjusted rand indices and saved to exploratory_analysis_data/adjusted_rand_indices_2.csv
2019-02-08 12:08:40	1.4	STATUS	Computed adjusted rand indices and saved to exploratory_analysis_data/adjusted_rand_indices_3.csv
2019-02-08 12:08:40	1.4	STATUS	Computed adjusted rand indices and saved to exploratory_analysis_data/adjusted_rand_indices_4.csv
2019-02-08 12:08:40	1.4	STATUS	Computed adjusted rand indices and saved to exploratory_analysis_data/adjusted_rand_indices_5.csv
2019-02-08 12:08:53	1.4	STATUS	COMPLETED Overlapping Clusters with Sample Traits
2019-02-08 12:08:54	1.4	STATUS	COMPLETED Clustering Section
2019-02-08 12:08:54	1.4	STATUS	COMPLETED Sample Clustering
2019-02-08 12:08:54	1.4	STATUS	COMPLETED Exploratory Analysis
2019-02-08 12:08:54	1.4	INFO	Initialized report index and saved to index.html
2019-02-08 12:09:00	1.1	STATUS	STARTED Differential Methylation
2019-02-08 12:09:00	1.1	INFO	Number of cores: 8
2019-02-08 12:09:00	1.1	STATUS	STARTED Analysis
2019-02-08 12:09:00	1.1	INFO	Using 0 permutation tests
2019-02-08 12:09:00	1.1	INFO	Using columns: Sample_Name,Sample_Group
2019-02-08 12:09:00	1.1	INFO	Using region types: promoters,tiling,genes,cpgislands
2019-02-08 12:09:07	1.1	STATUS	STARTED Retrieving comparison info
2019-02-08 12:09:07	1.1	STATUS	COMPLETED Retrieving comparison info
2019-02-08 12:09:07	1.1	STATUS	STARTED Computing differential methylation tables
2019-02-08 12:09:07	1.1	STATUS	STARTED Comparing: C9ALS vs. non.C9ALS (based on Sample_Group)
2019-02-08 12:09:07	1.1	STATUS	STARTED Computing Differential Methylation Table
2019-02-08 12:09:07	1.3	INFO	Conducting differential analysis using limma

2019-02-08 12:09:52	1.6	INFO	535 p-values are NA. They are treated as 1 in FDR adjustment
2019-02-08 12:09:54	1.8	STATUS	COMPLETED Computing Differential Methylation Table
2019-02-08 12:10:01	1.2	STATUS	STARTED Computing Differential Methylation Tables (Region Level)
2019-02-08 12:10:25	1.5	INFO	4 p-values are NA. They are treated as 1 in FDR adjustment
2019-02-08 12:10:25	1.6	STATUS	Computed table for promoters
2019-02-08 12:12:55	2.0	INFO	79 p-values are NA. They are treated as 1 in FDR adjustment
2019-02-08 12:12:56	1.8	STATUS	Computed table for tiling
2019-02-08 12:13:18	2.0	INFO	2 p-values are NA. They are treated as 1 in FDR adjustment
2019-02-08 12:13:18	2.0	STATUS	Computed table for genes
2019-02-08 12:13:34	2.2	STATUS	Computed table for cpgislands
2019-02-08 12:13:34	2.2	STATUS	COMPLETED Computing Differential Methylation Tables (Region Level)
2019-02-08 12:13:35	1.8	STATUS	COMPLETED Comparing: C9ALS vs. non.C9ALS (based on Sample_Group)
2019-02-08 12:13:35	1.8	STATUS	STARTED Comparing: Control vs. non.Control (based on Sample_Group)
2019-02-08 12:13:35	1.8	STATUS	STARTED Computing Differential Methylation Table
2019-02-08 12:13:35	2.1	INFO	Conducting differential analysis using limma
2019-02-08 12:14:22	2.0	INFO	536 p-values are NA. They are treated as 1 in FDR adjustment
2019-02-08 12:14:23	2.2	STATUS	COMPLETED Computing Differential Methylation Table
2019-02-08 12:14:32	1.4	STATUS	STARTED Computing Differential Methylation Tables (Region Level)
2019-02-08 12:14:59	1.9	INFO	4 p-values are NA. They are treated as 1 in FDR adjustment
2019-02-08 12:15:00	1.9	STATUS	Computed table for promoters
2019-02-08 12:17:34	2.1	INFO	79 p-values are NA. They are treated as 1 in FDR adjustment
2019-02-08 12:17:34	2.3	STATUS	Computed table for tiling
2019-02-08 12:17:56	2.2	INFO	2 p-values are NA. They are treated as 1 in FDR adjustment
2019-02-08 12:17:56	2.3	STATUS	Computed table for genes
2019-02-08 12:18:13	2.2	STATUS	Computed table for cpgislands
2019-02-08 12:18:13	2.2	STATUS	COMPLETED Computing Differential Methylation Tables (Region Level)
2019-02-08 12:18:13	2.3	STATUS	COMPLETED Comparing: Control vs. non.Control (based on Sample_Group)
2019-02-08 12:18:13	2.3	STATUS	STARTED Comparing: sALS vs. non.sALS (based on Sample_Group)
2019-02-08 12:18:13	2.3	STATUS	STARTED Computing Differential Methylation Table
2019-02-08 12:18:14	2.3	INFO	Conducting differential analysis using limma
2019-02-08 12:19:00	2.4	INFO	535 p-values are NA. They are treated as 1 in FDR adjustment

2019-02-08 12:19:01	2.4	STATUS	COMPLETED Computing Differential Methylation Table
2019-02-08 12:19:10	1.4	STATUS	STARTED Computing Differential Methylation Tables (Region Level)
2019-02-08 12:19:34	1.9	INFO	4 p-values are NA. They are treated as 1 in FDR adjustment
2019-02-08 12:19:35	1.9	STATUS	Computed table for promoters
2019-02-08 12:22:02	2.4	INFO	79 p-values are NA. They are treated as 1 in FDR adjustment
2019-02-08 12:22:03	2.0	STATUS	Computed table for tiling
2019-02-08 12:22:23	2.1	INFO	2 p-values are NA. They are treated as 1 in FDR adjustment
2019-02-08 12:22:24	2.1	STATUS	Computed table for genes
2019-02-08 12:22:40	2.2	STATUS	Computed table for cpgislands
2019-02-08 12:22:40	2.2	STATUS	COMPLETED Computing Differential Methylation Tables (Region Level)
2019-02-08 12:22:40	2.3	STATUS	COMPLETED Comparing: sALS vs. non.sALS (based on Sample_Group)
2019-02-08 12:22:40	2.3	STATUS	COMPLETED Computing differential methylation tables
2019-02-08 12:22:48	1.3	STATUS	STARTED Differential Methylation GO Enrichment Analysis
2019-02-08 12:22:48	1.3	STATUS	STARTED Comparison: C9ALS vs. non.C9ALS (based on Sample_Group)
2019-02-08 12:22:48	1.3	STATUS	STARTED Ontology: BP
2019-02-08 12:22:48	1.3	STATUS	STARTED Region Level
2019-02-08 12:22:48	1.3	STATUS	STARTED Region Type: promoters
2019-02-08 12:22:48	1.3	INFO	Rank cutoff: 100
2019-02-08 12:23:38	1.7	INFO	Rank cutoff: 500
2019-02-08 12:24:43	2.0	INFO	Rank cutoff: 1000
2019-02-08 12:25:56	1.8	INFO	Rank cutoff: 4 (auto-select)
2019-02-08 12:26:14	1.7	STATUS	COMPLETED Region Type: promoters
2019-02-08 12:26:14	1.7	STATUS	STARTED Region Type: tiling
2019-02-08 12:26:14	1.9	INFO	Not annotated with entrezID --> Skipped
2019-02-08 12:26:14	1.9	STATUS	COMPLETED Region Type: tiling
2019-02-08 12:26:14	1.9	STATUS	STARTED Region Type: genes
2019-02-08 12:26:15	1.9	INFO	Rank cutoff: 100
2019-02-08 12:27:00	1.8	INFO	Rank cutoff: 500
2019-02-08 12:28:03	2.0	INFO	Rank cutoff: 1000
2019-02-08 12:29:15	1.8	INFO	Rank cutoff: 3 (auto-select)
2019-02-08 12:29:45	2.0	STATUS	COMPLETED Region Type: genes
2019-02-08 12:29:45	2.0	STATUS	STARTED Region Type: cpgislands
2019-02-08 12:29:45	2.0	INFO	Not annotated with entrezID --> Skipped
2019-02-08 12:29:45	2.0	STATUS	COMPLETED Region Type: cpgislands
2019-02-08 12:29:45	2.0	STATUS	COMPLETED Region Level
2019-02-08 12:29:46	2.0	STATUS	COMPLETED Ontology: BP
2019-02-08 12:29:46	2.0	STATUS	STARTED Ontology: MF
2019-02-08 12:29:46	2.0	STATUS	STARTED Region Level
2019-02-08 12:29:46	2.0	STATUS	STARTED Region Type: promoters
2019-02-08 12:29:46	1.8	INFO	Rank cutoff: 100
2019-02-08 12:30:09	1.8	INFO	Rank cutoff: 500

2019-02-08 12:30:32	1.9	INFO	Rank cutoff: 1000
2019-02-08 12:30:57	1.9	INFO	Rank cutoff: 4 (auto-select)
2019-02-08 12:31:07	2.0	STATUS	COMPLETED Region Type: promoters
2019-02-08 12:31:07	2.0	STATUS	STARTED Region Type: tiling
2019-02-08 12:31:07	1.8	INFO	Not annotated with entrezID --> Skipped
2019-02-08 12:31:07	1.8	STATUS	COMPLETED Region Type: tiling
2019-02-08 12:31:07	1.8	STATUS	STARTED Region Type: genes
2019-02-08 12:31:07	1.9	INFO	Rank cutoff: 100
2019-02-08 12:31:35	2.0	INFO	Rank cutoff: 500
2019-02-08 12:31:59	1.8	INFO	Rank cutoff: 1000
2019-02-08 12:32:23	1.9	INFO	Rank cutoff: 3 (auto-select)
2019-02-08 12:32:43	1.9	STATUS	COMPLETED Region Type: genes
2019-02-08 12:32:43	1.9	STATUS	STARTED Region Type: cpgislands
2019-02-08 12:32:43	1.9	INFO	Not annotated with entrezID --> Skipped
2019-02-08 12:32:43	1.9	STATUS	COMPLETED Region Type: cpgislands
2019-02-08 12:32:44	1.9	STATUS	COMPLETED Region Level
2019-02-08 12:32:44	1.9	STATUS	COMPLETED Ontology: MF
2019-02-08 12:32:44	1.9	STATUS	COMPLETED Comparison: C9ALS vs. non.C9ALS (based on Sample_Group)
2019-02-08 12:32:44	1.9	STATUS	STARTED Comparison: Control vs. non.Control (based on Sample_Group)
2019-02-08 12:32:44	1.9	STATUS	STARTED Ontology: BP
2019-02-08 12:32:44	1.9	STATUS	STARTED Region Level
2019-02-08 12:32:44	1.9	STATUS	STARTED Region Type: promoters
2019-02-08 12:32:44	1.9	INFO	Rank cutoff: 100
2019-02-08 12:33:40	1.9	INFO	Rank cutoff: 500
2019-02-08 12:34:51	2.1	INFO	Rank cutoff: 1000
2019-02-08 12:36:09	2.2	INFO	Rank cutoff: 5 (auto-select)
2019-02-08 12:36:24	1.9	STATUS	COMPLETED Region Type: promoters
2019-02-08 12:36:24	1.9	STATUS	STARTED Region Type: tiling
2019-02-08 12:36:24	2.1	INFO	Not annotated with entrezID --> Skipped
2019-02-08 12:36:24	2.1	STATUS	COMPLETED Region Type: tiling
2019-02-08 12:36:24	2.1	STATUS	STARTED Region Type: genes
2019-02-08 12:36:24	2.1	INFO	Rank cutoff: 100
2019-02-08 12:37:17	2.1	INFO	Rank cutoff: 500
2019-02-08 12:38:22	2.2	INFO	Rank cutoff: 1000
2019-02-08 12:39:33	2.5	INFO	Rank cutoff: 3 (auto-select)
2019-02-08 12:39:54	2.5	STATUS	COMPLETED Region Type: genes
2019-02-08 12:39:54	2.5	STATUS	STARTED Region Type: cpgislands
2019-02-08 12:39:54	2.5	INFO	Not annotated with entrezID --> Skipped
2019-02-08 12:39:54	2.5	STATUS	COMPLETED Region Type: cpgislands
2019-02-08 12:39:54	2.5	STATUS	COMPLETED Region Level
2019-02-08 12:39:54	2.5	STATUS	COMPLETED Ontology: BP
2019-02-08 12:39:54	2.5	STATUS	STARTED Ontology: MF
2019-02-08 12:39:54	2.5	STATUS	STARTED Region Level
2019-02-08 12:39:54	2.5	STATUS	STARTED Region Type: promoters
2019-02-08 12:39:55	2.1	INFO	Rank cutoff: 100

231

```
2019-02-08 12:40:15    2.3   INFO                     Rank cutoff: 500
2019-02-08 12:40:38    2.5   INFO                     Rank cutoff: 1000
2019-02-08 12:41:00    2.3   INFO                     Rank cutoff: 5 (auto-select)
2019-02-08 12:41:09    2.5   INFO                     Could not conduct enrichment
analysis as associated genes are not in GO database.
2019-02-08 12:41:09    2.5   STATUS                   COMPLETED Region Type:
promoters
2019-02-08 12:41:09    2.5   STATUS                   STARTED Region Type: tiling
2019-02-08 12:41:09    2.4   INFO                     Not annotated with entrezID -->
Skipped
2019-02-08 12:41:09    2.4   STATUS                   COMPLETED Region Type: tiling
2019-02-08 12:41:09    2.4   STATUS                   STARTED Region Type: genes
2019-02-08 12:41:09    2.4   INFO                     Rank cutoff: 100
2019-02-08 12:41:31    2.4   INFO                     Rank cutoff: 500
2019-02-08 12:41:54    2.2   INFO                     Rank cutoff: 1000
2019-02-08 12:42:16    2.3   INFO                     Rank cutoff: 3 (auto-select)
2019-02-08 12:42:26    2.3   STATUS                   COMPLETED Region Type: genes
2019-02-08 12:42:26    2.3   STATUS                   STARTED Region Type: cpgislands
2019-02-08 12:42:26    2.3   INFO                     Not annotated with entrezID -->
Skipped
2019-02-08 12:42:26    2.3   STATUS                   COMPLETED Region Type:
cpgislands
2019-02-08 12:42:26    2.3   STATUS                  COMPLETED Region Level
2019-02-08 12:42:26    2.3   STATUS                 COMPLETED Ontology:  MF
2019-02-08 12:42:26    2.3   STATUS                COMPLETED Comparison:  Control vs.
non.Control (based on Sample_Group)
2019-02-08 12:42:26    2.3   STATUS                STARTED Comparison:  sALS vs. non.sALS
(based on Sample_Group)
2019-02-08 12:42:27    2.3   STATUS                   STARTED Ontology:  BP
2019-02-08 12:42:27    2.3   STATUS                    STARTED Region Level
2019-02-08 12:42:27    2.3   STATUS                     STARTED Region Type: promoters
2019-02-08 12:42:27    2.4   INFO                     Rank cutoff: 100
2019-02-08 12:43:14    2.5   INFO                     Rank cutoff: 500
2019-02-08 12:44:19    2.6   INFO                     Rank cutoff: 1000
2019-02-08 12:45:28    2.5   INFO                     Rank cutoff: 4 (auto-select)
2019-02-08 12:45:28    2.5   STATUS                   COMPLETED Region Type:
promoters
2019-02-08 12:45:28    2.5   STATUS                   STARTED Region Type: tiling
2019-02-08 12:45:28    2.6   INFO                     Not annotated with entrezID -->
Skipped
2019-02-08 12:45:28    2.6   STATUS                   COMPLETED Region Type: tiling
2019-02-08 12:45:28    2.6   STATUS                   STARTED Region Type: genes
2019-02-08 12:45:28    2.6   INFO                     Rank cutoff: 100
2019-02-08 12:46:22    2.4   INFO                     Rank cutoff: 500
2019-02-08 12:47:22    2.6   INFO                     Rank cutoff: 1000
2019-02-08 12:48:29    2.6   INFO                     Rank cutoff: 2 (auto-select)
2019-02-08 12:48:29    2.6   STATUS                   COMPLETED Region Type: genes
2019-02-08 12:48:29    2.6   STATUS                   STARTED Region Type: cpgislands
2019-02-08 12:48:30    2.6   INFO                     Not annotated with entrezID -->
Skipped
2019-02-08 12:48:30    2.6   STATUS                   COMPLETED Region Type:
cpgislands
2019-02-08 12:48:30    2.6   STATUS                  COMPLETED Region Level
2019-02-08 12:48:30    2.6   STATUS                 COMPLETED Ontology:  BP
2019-02-08 12:48:30    2.6   STATUS                 STARTED Ontology:  MF
```
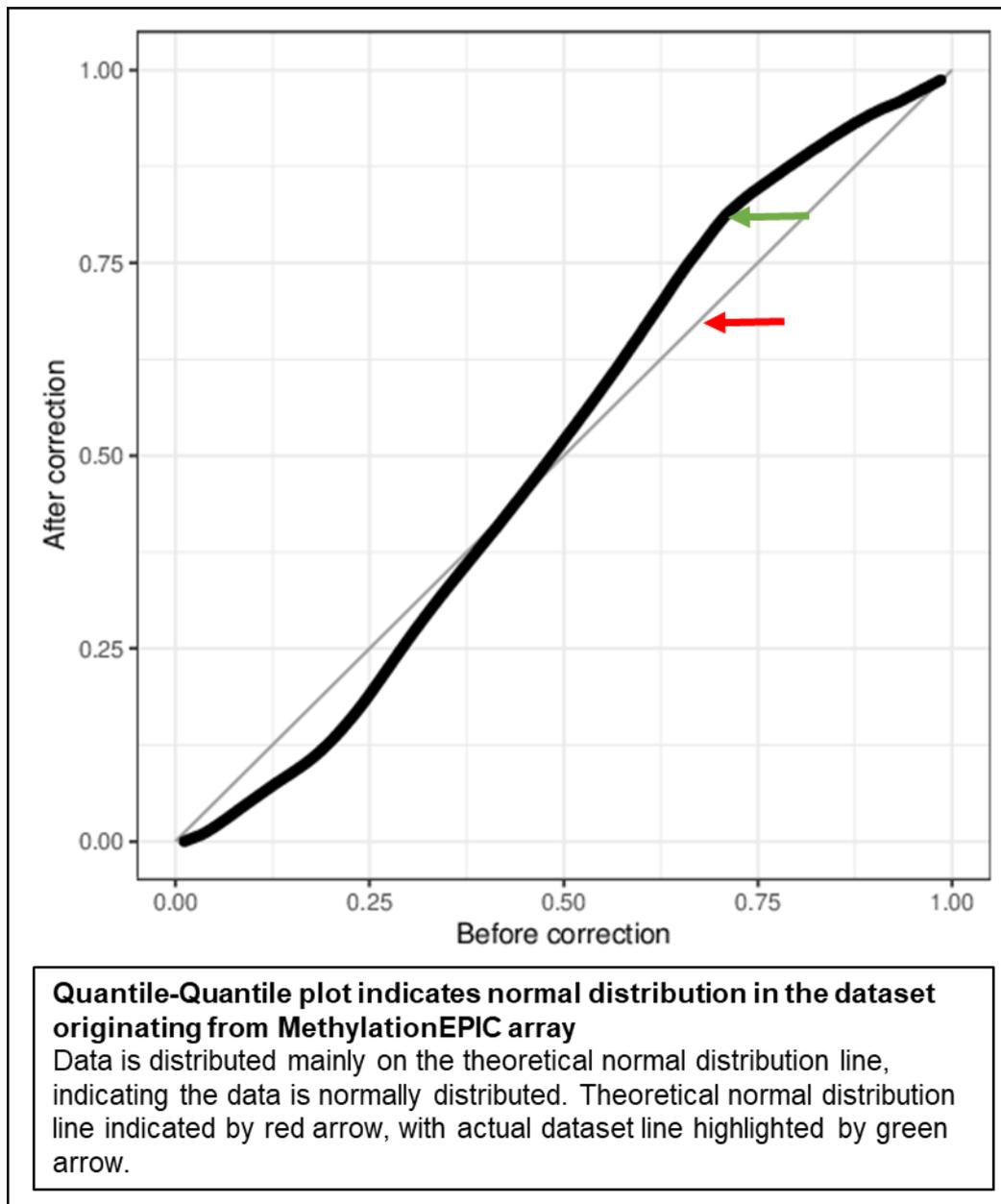
```
2019-02-08 12:48:30    2.6  STATUS            STARTED Region Level
2019-02-08 12:48:30    2.6  STATUS             STARTED Region Type: promoters
2019-02-08 12:48:30    2.7   INFO              Rank cutoff: 100
2019-02-08 12:48:51    2.8   INFO              Rank cutoff: 500
2019-02-08 12:49:14    2.6   INFO              Rank cutoff: 1000
2019-02-08 12:49:37    2.7   INFO              Rank cutoff: 4 (auto-select)
2019-02-08 12:49:37    2.7  STATUS             COMPLETED Region Type:
promoters
2019-02-08 12:49:37    2.7  STATUS             STARTED Region Type: tiling
2019-02-08 12:49:37    2.8   INFO              Not annotated with entrezID -->
Skipped
2019-02-08 12:49:37    2.8  STATUS             COMPLETED Region Type: tiling
2019-02-08 12:49:37    2.8  STATUS             STARTED Region Type: genes
2019-02-08 12:49:37    2.9   INFO              Rank cutoff: 100
2019-02-08 12:49:55    3.1   INFO              Rank cutoff: 500
2019-02-08 12:50:17    2.9   INFO              Rank cutoff: 1000
2019-02-08 12:50:38    2.7   INFO              Rank cutoff: 2 (auto-select)
2019-02-08 12:50:38    2.7  STATUS             COMPLETED Region Type: genes
2019-02-08 12:50:38    2.7  STATUS             STARTED Region Type: cpgislands
2019-02-08 12:50:38    2.7   INFO              Not annotated with entrezID -->
Skipped
2019-02-08 12:50:38    2.7  STATUS             COMPLETED Region Type:
cpgislands
2019-02-08 12:50:38    2.7  STATUS            COMPLETED Region Level
2019-02-08 12:50:38    2.7  STATUS           COMPLETED Ontology:  MF
2019-02-08 12:50:38    2.7  STATUS          COMPLETED Comparison:  sALS vs.
non.sALS (based on Sample_Group)
2019-02-08 12:50:38    2.7  STATUS         COMPLETED Differential Methylation GO
Enrichment Analysis
2019-02-08 12:50:50    2.2  STATUS      COMPLETED Analysis
2019-02-08 12:50:50    2.2  STATUS      STARTED Saving temp objects for debugging
2019-02-08 12:50:50    2.2 WARNING        Zip not found on this Windows system, this
RnBDiffMeth object will not be saved. See the instructions for installing ZIP on Windows in
the FAQ section of the RnBeads website.
2019-02-08 12:51:32    2.2  STATUS      COMPLETED Saving temp objects for debugging
2019-02-08 12:51:32    2.2  STATUS      STARTED Report Generation
2019-02-08 12:51:32    2.2  STATUS       Added introductory section
2019-02-08 12:51:32    2.2  STATUS       STARTED Adding Site Level Information
2019-02-08 12:51:32    2.2  STATUS        STARTED Selection of rank cutoffs
2019-02-08 12:51:34    2.8  STATUS        COMPLETED Selection of rank cutoffs
2019-02-08 12:51:34    2.8  STATUS        STARTED Adding scatterplots
2019-02-08 13:00:36    2.2  STATUS        COMPLETED Adding scatterplots
2019-02-08 13:00:45    2.2  STATUS        STARTED Adding volcano plots
2019-02-08 13:08:04    2.2  STATUS        COMPLETED Adding volcano plots
2019-02-08 13:08:04    2.2  STATUS        STARTED Adding tables
2019-02-08 13:10:45    2.7  STATUS        COMPLETED Adding tables
2019-02-08 13:10:45    2.7  STATUS       COMPLETED Adding Site Level Information
2019-02-08 13:10:45    2.7  STATUS       STARTED Adding Region Level Information
2019-02-08 13:10:45    2.7  STATUS        STARTED Selection of rank cutoffs
2019-02-08 13:10:46    2.6  STATUS        COMPLETED Selection of rank cutoffs
2019-02-08 13:10:46    2.6  STATUS        STARTED Adding scatterplots
2019-02-08 13:14:10    2.2  STATUS        COMPLETED Adding scatterplots
2019-02-08 13:14:16    2.2  STATUS        STARTED Adding volcano plots
2019-02-08 13:18:21    2.2  STATUS        COMPLETED Adding volcano plots
2019-02-08 13:18:21    2.2  STATUS        STARTED Adding tables
```

2019-02-08 13:19:07   2.8 STATUS   COMPLETED Adding tables
2019-02-08 13:19:07   2.8 STATUS   STARTED Adding GO enrichment analysis results
2019-02-08 13:41:07   2.5 STATUS   COMPLETED Adding GO enrichment analysis results
2019-02-08 13:41:07   2.5 STATUS   COMPLETED Adding Region Level Information
2019-02-08 13:41:07   2.5 STATUS   COMPLETED Report Generation
2019-02-08 13:41:07   2.5 STATUS   COMPLETED Differential Methylation
2019-02-08 13:41:08   2.5   INFO   Initialized report index and saved to index.html
2019-02-08 13:41:19   2.2 STATUS   STARTED Saving RData
2019-02-08 13:41:19   2.2 WARNING   Zip not found on this Windows system, this RnBDiffMeth object will not be saved. See the instructions for installing ZIP on Windows in the FAQ section of the RnBeads website.
2019-02-08 13:42:01   2.2 STATUS   COMPLETED Saving RData
2019-02-08 13:42:12   2.2 STATUS COMPLETED RnBeads Pipeline

Appendix X: QQ plot from MethylationEPIC analysis

Quantile-Quantile plot indicates normal distribution in the dataset originating from MethylationEPIC array
Data is distributed mainly on the theoretical normal distribution line, indicating the data is normally distributed. Theoretical normal distribution line indicated by red arrow, with actual dataset line highlighted by green arrow.

Appendix XI: PANTHER analysis of mRNA expression data

Conducted February 2019.

Appendix IX: PANTHER analysis of mRNA expression data (downregulated in ALS)

81 pathways implicated, pathways matching MethylationEPIC hypermethylation are highlighted in red.

5HT1 type receptor mediated signalling pathway (P04373)	**EGF receptor signalling pathway (P00018)**	**Nicotinic acetylcholine receptor signalling pathway (P00044)**
5HT2 type receptor mediated signalling pathway (P04374)	Endothelin signalling pathway (P00019)	**Notch signalling pathway (P00045)**
5HT4 type receptor mediated signaling pathway (P04376)	FAS signalling pathway (P00020)	Oxidative stress response (P00046)
ATP synthesis (P02721)	**FGF signalling pathway (P00021)**	P53 pathway feedback loops 1 (P04392)
Adenine and hypoxanthine salvage pathway (P02723)	GABA-B receptor II signaling (P05731)	**PDGF signalling pathway (P00047)**
Adrenaline and noradrenaline biosynthesis (P00001)	General transcription by RNA polymerase I (P00022)	PI3 kinase pathway (P00048)
Alpha adrenergic receptor signaling pathway (P00002)	General transcription regulation (P00023)	Parkinson disease (P00049)
Alzheimer disease-amyloid secretase pathway (P00003)	**Gonadotropin-releasing hormone receptor pathway (P06664)**	Pentose phosphate pathway (P02762)
Alzheimer disease-presenilin pathway (P00004)	Hedgehog signalling pathway (P00025)	Pyrimidine Metabolism (P02771)
Angiogenesis (P00005)	Heme biosynthesis (P02746)	Pyruvate metabolism (P02772)
Apoptosis signalling pathway (P00006)	Heterotrimeric G-protein signalling pathway-Gi alpha and Gs alpha mediated pathway (P00026)	Ras Pathway (P04393)
Arginine biosynthesis (P02728)	Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway (P00027)	Salvage pyrimidine ribonucleotides (P02775)
Axon guidance mediated by Slit/Robo (P00008)	Histamine H1 receptor mediated signalling pathway (P04385)	Synaptic vesicle trafficking (P05734)
B cell activation (P00010)	**Huntington disease (P00029)**	T cell activation (P00053)
Beta1 adrenergic receptor signalling pathway (P04377)	Hypoxia response via HIF activation (P00030)	TCA cycle (P00051)
Beta2 adrenergic receptor signalling pathway (P04378)	**Inflammation mediated by chemokine and cytokine signalling pathway (P00031)**	TGF-beta signalling pathway (P00052)
Beta3 adrenergic receptor signalling pathway (P04379)	Insulin/IGF pathway-mitogen activated protein kinase kinase/MAP kinase cascade (P00032)	Thiamin metabolism (P02780)
CCKR signalling map (P06959)	Insulin/IGF pathway-protein kinase B signalling cascade (P00033)	Thyrotropin-releasing hormone receptor signalling pathway (P04394)
Cadherin signalling pathway (P00012)	**Integrin signalling pathway (P00034)**	Toll receptor signalling pathway (P00054)
Cell cycle (P00013)	Interferon-gamma signalling pathway (P00035)	Transcription regulation by bZIP transcription factor (P00055)
Cholesterol biosynthesis (P00014)	**Interleukin signalling pathway (P00036)**	Ubiquitin proteasome pathway (P00060)
Circadian clock system (P00015)	Ionotropic glutamate receptor pathway (P00037)	**VEGF signalling pathway (P00056)**
Cobalamin biosynthesis (P02735)	JAK/STAT signalling pathway (P00038)	**Wnt signalling pathway (P00057)**
Coenzyme A biosynthesis (P02736)	Metabotropic glutamate receptor group I pathway (P00041)	p38 MAPK pathway (P05918)
Cytoskeletal regulation by Rho GTPase (P00016)	Metabotropic glutamate receptor group II pathway (P00040)	p53 pathway by glucose deprivation (P04397)
DNA replication (P00017)	Metabotropic glutamate receptor group III pathway (P00039)	p53 pathway feedback loops 2 (P04398)
Dopamine receptor mediated signalling pathway (P05912)	Methylcitrate cycle (P02754)	p53 pathway (P00059)

Appendix IX: PANTHER analysis of mRNA expression data (upregulated in ALS)

84 pathways implicated, pathways matching MethylationEPIC hypomethylation are highlighted in red.

5HT1 type receptor mediated signalling pathway (P04373)	Dopamine receptor mediated signalling pathway (P05912)	Metabotropic glutamate receptor group III pathway (P00039)
5HT2 type receptor mediated signalling pathway (P04374)	EGF receptor signalling pathway (P00018)	Methylcitrate cycle (P02754)
5HT3 type receptor mediated signalling pathway (P04375)	Endothelin signalling pathway (P00019)	Nicotinic acetylcholine receptor signalling pathway (P00044)
5HT4 type receptor mediated signalling pathway (P04376)	FAS signalling pathway (P00020)	Notch signalling pathway (P00045)
Acetate utilization (P02722)	FGF signalling pathway (P00021)	Oxidative stress response (P00046)
Adrenaline and noradrenaline biosynthesis (P00001)	GABA-B receptor II signalling (P05731)	PDGF signalling pathway (P00047)
Alpha adrenergic receptor signalling pathway (P00002)	Gamma-aminobutyric acid synthesis (P04384)	PI3 kinase pathway (P00048)
Alzheimer disease-amyloid secretase pathway (P00003)	General transcription regulation (P00023)	Parkinson disease (P00049)
Alzheimer disease-presenilin pathway (P00004)	Gonadotropin-releasing hormone receptor pathway (P06664)	Purine metabolism (P02769)
Angiogenesis (P00005)	Hedgehog signalling pathway (P00025)	Pyrimidine Metabolism (P02771)
Angiotensin II-stimulated signalling through G proteins and beta-arrestin (P05911)	Heme biosynthesis (P02746)	Ras Pathway (P04393)
Apoptosis signalling pathway (P00006)	Heterotrimeric G-protein signalling pathway-Gi alpha and Gs alpha mediated pathway (P00026)	Salvage pyrimidine deoxyribonucleotides (P02774)
Axon guidance mediated by Slit/Robo (P00008)	Heterotrimeric G-protein signalling pathway-Gq alpha and Go alpha mediated pathway (P00027)	Salvage pyrimidine ribonucleotides (P02775)
Axon guidance mediated by netrin (P00009)	Histamine H1 receptor mediated signalling pathway (P04385)	Succinate to proprionate conversion (P02777)
Axon guidance mediated by semaphorins (P00007)	Histamine H2 receptor mediated signalling pathway (P04386)	Synaptic vesicle trafficking (P05734)
B cell activation (P00010)	Huntington disease (P00029)	T cell activation (P00053)
Beta1 adrenergic receptor signalling pathway (P04377)	Hypoxia response via HIF activation (P00030)	TGF-beta signalling pathway (P00052)
Beta2 adrenergic receptor signalling pathway (P04378)	Inflammation mediated by chemokine and cytokine signalling pathway (P00031)	Toll receptor signalling pathway (P00054)
Beta3 adrenergic receptor signalling pathway (P04379)	Insulin/IGF pathway-mitogen activated protein kinase kinase/MAP kinase cascade (P00032)	Transcription regulation by bZIP transcription factor (P00055)
Blood coagulation (P00011)	Insulin/IGF pathway-protein kinase B signalling cascade (P00033)	Ubiquitin proteasome pathway (P00060)
CCKR signalling map (P06959)	Integrin signalling pathway (P00034)	VEGF signalling pathway (P00056)
Cadherin signalling pathway (P00012)	Interleukin signalling pathway (P00036)	Vitamin D metabolism and pathway (P04396)
Circadian clock system (P00015)	Ionotropic glutamate receptor pathway (P00037)	Wnt signalling pathway (P00057)
Cortocotropin releasing factor receptor signaling pathway (P04380)	JAK/STAT signaling pathway (P00038)	mRNA splicing (P00058)
Cytoskeletal regulation by Rho GTPase (P00016)	Metabotropic glutamate receptor group I pathway (P00041)	p53 pathway by glucose deprivation (P04397)
DNA replication (P00017)	Metabotropic glutamate receptor group II pathway (P00040)	p53 pathway feedback loops 2 (P04398)
De novo purine biosynthesis (P02738)	Muscarinic acetylcholine receptor 1 and 3 signalling pathway (P00042)	p53 pathway (P00059)
De novo pyrimidine deoxyribonucleotide biosynthesis (P02739)		

Appendix XII: Indexes used for BS-NGS

Details of the indexes used in the Zymo Pico Methyl-Seq™ library formation in preparation for BS-NGS		
Index	Illumina index	Sequence
A	2	CGATGT
B	4	TGACCA
C	5	ACAGTG
D	6	GCCAAT
E	7	CAGATC
F	12	CTTGTA

Appendix XIII: FastQC summary table for BS-NGS samples

Table depicting the quality control conducted using FastQC on Galaxy on the BS-NGS samples

FastQC report sample IDs	39/11	53/96	35/96	34/05	05/07	94/06	39/97	41/04	85/07	87/92	98/02	200/97
Total sequences	30743866	23591384	16102774	26378198	12409865	29241301	37836915	43782417	30855190	26541915	37623635	30000208
Sequences flagged as poor quality	0	0	0	0	0	0	0	0	0	0	0	0
47Sequence length	66	66	66	66	66	66	66	66	66	66	66	66
Percent GQ	45	45	41	40	38	51	40	37	48	38	47	46
Per base sequence quality	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Per tile sequence quality	Caution	Pass	Pass	Pass	Caution	Caution	Pass	Caution	Caution	Caution	Pass	Pass
Per sequence quality scores	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Per base sequence content	Caution	Fail	Fail	Caution	Caution	Fail	Fail	Caution	Fail	Fail	Fail	Fail
Per sequence GQ content	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail
Per base N content	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Sequence length distribution	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Sequence duplication levels	Fail	Fail	Fail	Caution	Caution	Fail	Fail	Fail	Fail	Fail	Fail	Fail
Overrepresented sequences	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail
Adapter content	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail

Appendix XIV: Permissions for reuse of figures/tables

Order Completed

Thank you for your order.

This Agreement between University of Sheffield -- Charlie Appleby-Mallinder ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

printable details

License Number	4664700047152
License date	Sep 09, 2019
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Neuropathology and Applied Neurobiology
Licensed Content Title	Loss of nuclear TDP-43 in amyotrophic lateral sclerosis (ALS) causes altered expression of splicing machinery and widespread dysregulation of RNA splicing in motor neurones
Licensed Content Author	Paul G. Ince, Pamela J. Shaw, Stephen B. Wharton, et al
Licensed Content Date	Sep 25, 2014
Licensed Content Volume	40
Licensed Content Issue	6
Licensed Content Pages	16
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Table 3
Will you be translating?	No
Title of your thesis / dissertation	DNA methylation in Amyotrophic Lateral Sclerosis
Expected completion date	Oct 2019
Expected size (number of pages)	250
Requestor Location	University of Sheffield SITraN Glossop Road
	Sheffield, S10 2HQ United Kingdom Attn: University of Sheffield
Publisher Tax ID	EU826007151
Total	0.00 GBP

Order Completed

Thank you for your order.

This Agreement between University of Sheffield -- Charlie Appleby-Mallinder ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

printable details

License Number	4664700300346
License date	Sep 09, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Neuroscience
Licensed Content Title	The expanding biology of the C9orf72 nucleotide repeat expansion in neurodegenerative disease
Licensed Content Author	Aaron R. Haeusler, Christopher J. Donnelly, Jeffrey D. Rothstein
Licensed Content Date	May 6, 2016
Licensed Content Volume	17
Licensed Content Issue	6
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	DNA methylation in Amyotrophic Lateral Sclerosis
Institution name	n/a
Expected presentation date	Oct 2019
Portions	figure 1
Requestor Location	University of Sheffield SITraN Glossop Road
	Sheffield, S10 2HQ United Kingdom Attn: University of Sheffield
Total	0.00 GBP

Appendix XV: Evidence of Publication

Appleby-Mallinder, C., Schaber, E., Kirby, J., Shaw, P.J., Cooper-Knock, J, Heath, P.R and Highley, J.R., (2020). TDP43 proteinopathy is associated with aberrant DNA methylation in Human Amyotrophic Lateral Sclerosis. *Neuropathology and Applied Neurobiology* (in press).