Nucleotide methylation and neurocognitive phenotypes

Miles Flitton

Thesis submitted for the degree of Doctor of Philosophy at the University of Nottingham

September 2017

Dedicated with love to Betty and Winifred.

Declaration

I declare that all work presented in this thesis has been produced by myself and has not been submitted for any other degree. Every effort has been made to acknowledge collaborations and the sources of previous research have been cited throughout.

Acknowledgements

My principal thanks go to Helen Miranda Knight for her unwavering support and tireless efforts over the last three years. I have learnt a great deal under her supervision and we have achieved more over the intervening years than perhaps we would have expected. It has been my pleasure to support the development of her research group and I have the utmost faith in her ability as the group continues on their scientific journey.

Of course, I must thank Ian Macdonald for stepping into the role of second supervisor at the eleventh hour. Not only did he provide an objective lens to my work, but also a stable wisdom during the ups and downs of Helen's and my time in Nottingham.

Perhaps the most important acknowledgments go to the other members of our research group. To Maria Koromina, to Nicholas Rielly, to Braulio Martinez, to Eleanor Bellows, and to Maria Haig. There is no doubt that colleagues can make your working life effortless or they can make it a chore. Together, we have worked hard to support each other in producing the best research we can and have grown to become good friends. I hope that you continue to bring laughter into the lab and make each day more successful and enjoyable than the last.

Outside of our research group, I would like to thank Ian Mellor and his group for their support and collaboration. I would like to thank Alexey Ruzov and his group for their help with the immunohistochemistry. I would like to thank James Burston for his general support in the lab. I would like to thank the technical staff in the Medical School for their hard work every day of every year.

Outside of the University, I would like to thank my family for their support during the completion of this thesis. Particular mention goes to my brother, with whom I have had many discussions about our work and about academia in general. He is a fountain of knowledge that I will no doubt persist in tapping in the future. I think he would agree that the best part of doing a PhD is finishing it.

Finally, I would like to thank the University of Nottingham for awarding me with a PhD studentship and for providing me with the facilities to complete my research.

Abstract

Epigenetic modifications are under increasing scrutiny in research of health and disease states. The most commonly researched epigenetic modification, DNA methylation, is associated with familial forms of dementia, developmental delay syndromes, and disparate cognitive phenotypes. This thesis investigates the relationship between nucleotide methylation and brain disease using three distinct approaches.

The majority of the thesis focuses on a genetic variant located within a DNA methyltransferase, *DNMT3L* R278G, which had previously been associated with intelligence scores in a Scottish birth cohort. Data was utilised from the Oxford Project To Investigate Memory and Aging, specifically the VITACOG cohort of individuals who have mild cognitive impairment and who received B vitamin treatment over a two-year period. It was discovered that minor allele carriers of the R278G variant who received B vitamin treatment showed improved visuospatial associative memory, but diminished verbal semantic memory. These effects were reflected in rates of brain atrophy for these individuals over the course of the two-year study. *In silico* modelling suggests that the R278G variant disrupts the DNMT3A-3L complex required for *de novo* methylation.

Investigation of genetic variants in methylation genes was by examining next-generation sequencing data from the UK10K project for multiple intellectual disability and autism spectrum disorder cohorts. Thirty DNA and RNA methylation genes were investigated to identify rare coding variants these cohorts. Case-control analysis of these intellectual disability and autism cohorts compared to the TwinsUK general population controls identified four rare missense variants associated with risk for disease. As these variants were found only in individuals with intellectual disability comorbid with psychiatric disease, the findings suggest that these variants contribute to disease risk for a specific clinical population.

DNA methylation is a dynamic process where demethylation of 5mC results in the formation of further methylation marks including 5hmC and 5caC. An optimised immunohistochemical method was used to detect 5hmC and 5caC modifications in multiple regions of human brain tissue. Initial characterisation confirmed the near-ubiquitous presence of 5hmC in the human brain. Further, substantial 5caC staining was recorded in the human cerebellum that was absent in all other brain regions,

signifying an as-of-yet unidentified relationship between this methylation mark and cerebellar function.

Taken together, the different strands of investigation carried out in this thesis have substantiated the importance of nucleotide methylation in dementia progression, intellectual disability, and healthy brain function. Our use of well-curated cohort studies provides the basis for further investigation into methylation patterns and their role in the pathogenesis of multiple brain diseases.

Contents

List of figures

List of tables

Abbreviations

Δ	Change
ΔG	Free energy
ΔΔG	Change in free energy
5,10-CH_2-THF	5,10-methylenetetrahydrofolate
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
5-MTHF	5-methyltetrahydrofolate
Å	Angstrom
AD	Alzheimer's disease
ADCA-DN	Autosomal dominant cerebellar ataxia, deafness, and narcolepsy
ADHD	Attention deficit hyperactivity disorder
ADI-R	Autism diagnostic interview-revised
ADOS	Autism diagnostic observation schedule
AFR	African
AID	Activation-induced cytidine deaminase
ALKBH5	Alkylation repair homolog 5
ALU	Arthrobacter luteus elements
AMR	American
AN	Animal naming
ANK1	Ankyrin 1
APBS	Adaptive Poisson-Boltzmann Solver
APOBEC	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
APOE	Apolipoprotein
APP	Amyloid precursor protein
ARHGAP18	Rho GTPase activating protein 18
ASD	Autism spectrum disorders
ASN	Asian
Aβ	Amyloid beta
BACE	Beta-secretase
BAM	Binary sequence alignment map
BDNF	Brain-derived neurotrophic factor
BER	Base excision repair

BE-SKAT	Backward elimination sequence kernel association test
BHMT	Betaine-homocysteine S-methyltransferase
BP	Blood pressure
BPD	Bipolar disorder
C t.	C terminus
Ca^{2+}	Calcium
CaN	Calcineurin
CANTAB	Cambridge neuropsychological test automated battery
CAPI	Computer-assisted personal interview
CBS	Cystathionine beta-synthase
CF	Category fluency
Chr	Chromosome
CL	Cystathionine gamma-lyase
CMC	Combined multivariate and collapsing
CNV	Copy number variation
CpG	5'-C-phosphate-G-3'
CpH	5'-C-phosphate-A/C/T-3'
DAPI	4',6-diamidino-2-phenylindole
DCX	Doublecortin
DEL	Deletion
DHF	Dihydrofolate
DISC1	Disrupted in schizophrenia 1
dmPFC	Dorsomedial prefrontal cortex
DMR	Differentially methylated region
DMS	Delayed matching to sample
DMSO	Dimethyl sulfoxide
dMT	Demethylase
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DSSP	Dictionary of secondary structure of proteins
DUP	Duplication
DWR	Delayed word recall
EGA	European Genome-phenome Archive
ELAVL1	ELAV-like protein 1
ELP3	Elongator complex protein 3
EUR	European

ExAC	Exome aggregation consortium
FMR1	Fragile X mental retardation 1
FRET	Fluorescence resonant energy transfer
FTO	Fat mass and obesity-associated protein
GADD45	Growth arrest and DNA damage
GCL	Glutamate cysteine ligase
GDS	Geriatric depression scale
GFAP	Glial fibrillary acidic protein
GLM	General linear model
GNT	Graded naming test
GPx	Glutathione peroxidase
GR	Glutathione reductase
GSS	Glutathione synthetase
GSSG	Oxidised glutathione
GTEx	Genotype-tissue expression consortium
GWAS	Genome-wide association study
Hcy	Homocysteine
HNRNPA2B1	Heterogeneous nuclear ribonucleoproteins A2/B1
HSAN1	Hereditary Sensory and Autonomic Neuropathy type 1
HTA	Human Tissue Act
HVLT-R	Hopkins verbal learning test-revised
ID	Intellectual disability
IED	Intra-extra dimensional
IMGSAC	International Molecular Genetic Study of Autism Consortium
INDEL	Insertion/deletion
INS	Insertion
INV	Inversion
IQ	Intelligence
KASP	Kompetitive Allele Specific Polymerase Chain Reaction
kb	Kilobase
KBAC	Kernel-based adaptive cluster
kcal/mol	Kilocalorie per mole
KO	Genetic knockout
LINE	Long interspersed elements
LoF	Loss-of-function
LOFTEE	Loss-Of-Function Transcript Effect Estimator

LTD	Long term depression
LTP	Long term potentiation
LV	Left ventricle
M	Molar concentration
m^6A	N^6-methyladenosine
m^6A_m	N^6,2'-*O*-dimethyladenosine
MaCH	Markov Chain framework for genotype imputation
MAF	Minor allele frequency
MAT	Methionine adenyltransferase
MBD	Methyl-CpG-binding domain protein
MCI	Mild cognitive impairment
MDD	Major depressive disorder
MECP2	Methyl-CpG-binding domain protein 2 Rett syndrome
METTL	N^6-adenosine-methyltransferase
MGAS	Molecular Genetics of Autism Study
MHC	Major histocompatibility complex
miRNA	Micro RNA
mL	Millilitre
MMSE	Mini-mental state examination
MNP	Multiple nucleotide polymorphism
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
MS	Methionine synthase
MSP	Multisystem proteinopathy
MSR	Methionine synthase reductase
MTHFR	Methylenetetrahydrofolate reductase
MTs	Methyltransferases
N t.	N terminus
NART	National adult reading test
NCAPH2	Condensin-2 complex subunit H2
NCDS	National Child Development Study
NEIL	Endonuclease VIII-like
NeuN	Fox-3
NFE	Non-Finnish European
NGS	Next-generation sequencing
NHSB	Nottingham Health Science Biobank

NMDA	N-methyl-D-aspartate
nmol/L	Nanomole per litre
NSUN2	NOP2/Sun domain family, member2
OPTIMA	Oxford Project to Investigate Memory and Aging
OR	Odds ratio
PAL	Paired associates learning
PARSE	Parameters for solvation energy
PBE	Poisson-Boltzmann equation
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
PDB	Protein data bank
pH	Potential of hydrogen
pmol/L	Picomole per litre
PP1	Protein phosphatase 1
PQBP1	Polyglutamine-binding protein 1
PRM	Pattern recognition memory
PRRC1	Proline rich coiled coil 1
PSEN1	Presenilin 1
PSEN2	Presenilin 2
PUFA	Polyunsaturated fatty acid
Rare FIND	Rare Familial Intellectual Disability
RBFOX1	Fox-1 homolog A
RELN	Reelin
RNA	Ribonucleic acid
ROA	Rate of atrophy
ROS	Reactive oxygen species
SAH	S-adenosylhomocysteinase
SAHH	S-adenosylhomocysteinase hydrolase
SAM	S-adenosylmethionine
SCZ	Schizophrenia
SD	Standard deviation
SE	Standard error
SETD1A	SET domain containing 1A
SF	Synergy factor
SHMT	Serine hydroxymethyltransferase

SKAT	Sequence kernel association test
SLIM	School of Life Sciences Imaging
SNP	Single nucleotide polymorphism
SOP	Standard operation procedure
SSP	Spatial span
SVA	SINE/VNTR/ALU elements
SWM	Spatial working memory
TCOF1	Treacle protein
TDG	Thymine-DNA glycosylase
TET	Ten-eleven translocation methylcytosine dioxygenase
tHcy	Total homocysteine
THF	Tetrahydrofolate
THYN1	Thymocyte nuclear protein 1
TICS-M	Telephone interview for cognitive status-modified
TREM2	Triggering receptor expressed on myeloid cells 2
tRNA	Transfer RNA
TSA	Tyramide signal amplification
UHRF	Ubiquitin-like, containing PHD and RING finger domains
UTR	Untranslated region
VCF	Variant call format
VEP	Variant effect predictor
VT	Variable threshold
WDR76	WD repeating domain 76
WES	Whole-exome sequencing
WGS	Whole-genome sequencing
WR	Word recall
WTAP	Wilms tumor 1 associated protein
YTHDF	YTH N6-methyladenosine RNA binding protein
γEC	gamma-Glutamylcysteine
μm	Micrometer
μmol/L	Micromole per litre
χ2	Chi-square
Zfp57	Zinc finger protein 57 homolog

1. General introduction

1.1. Brain diseases

Brain disease is a relatively new term that refers to the myriad of neurological and psychiatric disorders associated with abnormal brain function. In many countries, the prevalence of neurodegenerative, neuropsychiatric and neurodevelopmental disorders is on the increase and presents a major burden to successive governments, with the cost of brain diseases in Europe alone estimated at nearly 800 billion euros in 2010 (DiLuca and Olesen, 2014).

Neuropsychiatric disease refers to mental conditions that have been historically associated with psychiatric research, characterised in the Diagnostic and Statistical Manual of Mental Disorders Fifth Edition (Association, 2013), rather than particular diseases of the neurological system covered by pure neurology. The most notable of these neuropsychiatric diseases include depression such as major depressive disorder (MDD), combinations of mania and depression such as bipolar disorder (BPD), or episodes of more serious psychotic symptoms such as those observed in schizophrenia (SCZ). Notable modern additions to neuropsychiatry include eating disorders such as bulimia and obsessive-compulsive disorder, a condition characterised by intrusive behavioural routines.

Neurodegenerative diseases are perhaps the most pertinent category of brain diseases being researched today owing to increasing life expectancy and thus an increasing number of elderly individuals developing neurodegenerative conditions. This growing prevalence of neurodegenerative diseases creates an obvious burden on families and social care providers. Alzheimer's disease (AD), the most common form of dementia, is characterised by problems with memory and is typically classed as a tauopathy due to the build-up of tau protein (alongside amyloid beta plaques) in the brain. Parkinson's disease (PD), the second most common form of dementia, manifests with characteristic motor problems and is typically classed as a synucleiopathy due to the accumulation of alpha-synuclein protein in the brains of those with PD. However, there is substantial biological and symptomatic crossover between these two categories of neurodegeneration, highlighting the difficulty in classifying these pathologies (Moussaud et al., 2014). Aside from these more common forms of dementia, neurodegeneration also includes diseases such as

motor neurone disease, characterised by the death of neurons associated with voluntary motor control, and prion diseases such as Creutzfeldt-Jakob disease.

Neurodevelopmental disorders describe a number of disease categories that are associated with impaired neurological and cognitive growth. Intellectual disabilities (ID) describe some forms of neurodevelopmental disease that are primarily associated with impairments in cognitive function and delays in reaching expected developmental milestones. In comparison to neuropsychiatric and neurodegenerative conditions, the cause of some forms of intellectual disability have been putatively identified through genetic studies. Examples of this include Fragile X syndrome and Down syndrome, disorders that manifests with multiple physical symptoms and varying deficits to cognitive function. In comparison, there are other forms of neurodevelopmental disorder that share characteristics with intellectual disability, but the causes of which are not well understood. One prominent example is autism spectrum disorders (ASD), a range of conditions that are often characterised by altered social function and repetitive behaviours, but also comorbidity with symptoms of intellectual disability (Matson and Shoemaker, 2009).

1.1.1. Cognition and brain diseases

Whilst there are many phenotypes associated with neurodegenerative, neuropsychiatric, and neurodevelopmental diseases that are physical in nature, the emphasis of this study will be on cognitive functions affected by disease and most principally those implicated in neurodegenerative conditions. Cognition is an umbrella term that covers all conscious and unconscious aspects of mental function associated with human behaviour. Owing to the breadth of individual traits covered by cognition, researchers have utilised endophenotypes borne out of psychology and psychiatry research to segregate mental functions related to specific diseases or aberrant behaviours, and incorporate these endophenotypes into empirical study (Flint and Munafo, 2007).

Much of our understanding of functional specialisation within the brain, namely the modulatory of neurological regions that are associated with distinct cognitive functions, began with studies of individuals who had suffered damage to particular areas of the brain. These lesion studies helped to clarify that structural deficits in specific brain regions can correspond to certain symptoms, but also that symptoms can occur without structural damage to the brain, referred to as damage to the

cognitive system (Kosslyn and Intriligator, 1992). Likewise, our developing understanding of the nuances of cognition has been aided by studies of individuals with neurodegenerative, neuropsychiatric, and neurodevelopmental disease.

Due to the biological and phenotypical spectrum associated with brain disorders, both genetic and neuroscientific research has attempted to focus on key phenotypes associated with specific brain diseases (Green et al., 2008). A key endophenotype within cognition is memory performance. Memory is commonly divided into declarative (conscious) and non-declarative (non-conscious) recollection (Graf and Schacter, 1985), with further distinctions between two forms of declarative memory; episodic memory dealing with individual events, and semantic memory dealing with forms of knowledge (Tulving, 1972). Impaired memory performance is a critical cognitive measure analysed in many brain diseases, particularly neurodegenerative disorders such as AD where the principal symptom is memory deficits (Perry et al., 2000).

In 1988, the Oxford Project to Investigate Memory and Ageing (OPTIMA) set out to improve understanding of dementia through a number of longitudinal studies principally investigating individuals with AD. Some of these studies also included individuals with mild cognitive impairment (MCI) which is thought to be a precursor or prodromal stage of dementia (Grundman et al., 2004). The findings of the OPTIMA research have consistently highlighted memory performance as a prominent indicator of MCI and AD progression respectively (De Jager et al., 2003). Research has highlighted episodic, semantic, and working memory deficits in both AD and MCI, underlining the importance of memory processing to the development of dementia (Belleville et al., 2007, Dudas et al., 2005).

Although the main focus of this thesis will be neurodegeneration and related cognitive endophenotypes such as particular facets of memory performance, there are aspects of neuropsychiatric and neurodevelopmental disease that are relevant to this work. Memory problems are reported in some neuropsychiatric and neurodevelopmental diseases, such as SCZ and specific forms of ID, but not consistently or with the same specificity as neurodegenerative memory complaints (Leyfer et al., 2006, Schuchardt et al., 2010). More stable comorbid symptoms include depressive symptoms, a hallmark of many neuropsychiatric disorders but also prevalent in individuals with dementia (Jorm, 2001). Interestingly, individuals with ID and ASD can present with psychiatric symptoms, most commonly those

associated with schizophrenia (Morgan et al., 2008, Rapoport et al., 2009). Thus, whilst our emphasis is on memory performance, one should be aware of the level of overlap between neurodegenerative, neuropsychiatric, and neurodevelopmental diseases.

1.1.2. Genetics and brain diseases

Despite the number of publications investigating the genetic mechanisms underlying neurodegenerative, neuropsychiatric, and neurodevelopmental disorders, there is still a large gap in our understanding of the genetic contribution to disease risk. In the case of neurodegenerative diseases, there have been successes in identifying genetic variants that transmit substantial risk for developing specific forms of dementia. For instance, genetic linkage studies of dementia led to the identification of key risk factors for early-onset familial AD, namely *amyloid precursor protein* (*APP*), *presenilin 1* (*PSEN1*), and *presenilin 2* (*PSEN2*). Moreover, *apolipoprotein E4* (*APOE4*) is now recognised as the largest genetic risk factor for late onset AD (Bertram and Tanzi, 2008).

These genes have all been implicated in the amyloid hypothesis (Hardy and Higgins, 1992), a prominent theory which posits that abnormal production of the protein amyloid beta (Aβ) results in neuritic plaques distinctive of AD pathology (Hardy and Selkoe, 2002). In the resulting years, a number of loci related to AD have been identified through genome-wide association studies (GWAS) (Medway and Morgan, 2014). However, it has been conceded that such genome-wide analysis has overlooked a significant portion of rare genetic variance (Ridge et al., 2013, Wellenreuther and Hansson, 2016). This raises questions as to whether established genetic approaches can identify substantial genetic variation in complex disorders such as AD (Bertram, 2011).

Genetic investigations of MCI have been relatively sparse given that the disorder only received accepted criteria around a decade ago and debate surrounding the validity of the disorder still remains (Petersen et al., 2014). Yet, promisingly, multiple studies have also highlighted *APOE4* as the most significant risk factor in MCI (Andrawis et al., 2012, Mackin et al., 2013) as well as being a risk factor in the progression from MCI to AD (Singh et al., 2012). This supports the observation that amyloid pathology is prevalent in a significant number of MCI cases (Wolk et al., 2009), providing evidence for MCI as a precursor to dementia as well as for the

amyloid hypothesis in general. However, inconsistencies have been reported in the pathology of MCI, with expected neuropathological markers identified in individuals with no signs of cognitive impairment (Stephan et al., 2012). Also, attempts to uncover genetic variants through GWAS have so far been largely unsuccessful. For instance, a recent study identified a number of loci associated with cognitive impairment in MCI, none of which reached significance without the addition of *APOE4* (Carrasquillo et al., 2015).

With regards to neuropsychiatric disorders, much research has been conducted into SCZ following the early observations that risk of this disorder was higher in relatives of affected individuals, particularly in monozygotic twins, compared to the general population (Sullivan et al., 2003). The most notable genetic risk factor identified from families a chromosomal abnormality affecting the gene *disrupted in schizophrenia 1* (*DISC1*). While questions have been raised about the validity of the *DISC1* association, it remains the clearest genetic contributor to the disease (St Clair et al., 1990, Sullivan, 2013). In the intervening years, studies have highlighted other neuropsychiatric genetic risk factors such as coding variants within the *ATP-binding cassette sub-family A member 13* (*ABCA13*) gene (Knight et al., 2009). Moreover, GWAS research has identified a haplotype in the major histocompatibility complex (MHC), a crucial region in immunity, which is associated with risk of SCZ (International Schizophrenia et al., 2009). These and other authors also note the shared genetic risk between SCZ and BPD, providing some promise of a detectable genetic component to neuropsychiatric disease (Hodgkinson et al., 2004).

The genetics of neurodevelopmental disorders exemplifies both the successes and the challenges in identifying genetic risk factors for specific diseases. As already mentioned, the cause of some forms of ID has been conclusively attributed to genetic disruption. Fragile X syndrome is caused by mutations to the *fragile X mental retardation 1* (*FMR1*) gene, most commonly causing expansion of CGG trinucleotide repeats in the 5' untranslated region (UTR) (Pieretti et al., 1991). Similarly, Down syndrome is caused by the presence of an extra full or part copy of chromosome 21 (Lejeune et al., 1959). However, the genetic contribution to other forms of neurodevelopmental disease has remained difficult to identify. For example, ASD presents with high heritability in twin studies but candidate gene studies have struggled to identify any significant risk for the disease, with the only notable success coming from studies of copy number variation (CNV) (Marshall and Scherer, 2012, Persico and Napolioni, 2013, Tick et al., 2016). GWAS results have also been

unsuccessful at replicating any identified candidate variants in large number of individuals with ASD (Torrico et al., 2017).

To tackle some of the difficulties associated with identifying or replicating variants of interest, recent genetic research into brain diseases has focused upon the contribution of rare variants. This has been made possible thanks to the growing number of next-generation sequencing (NGS) datasets being produced from cohorts across the world. Promisingly, application of this approach in family studies has already identified rare variants associated with risk for forms of ID and motor neurone disease (Karaca et al., 2015, Smith et al., 2014). Investigation of key disease genes, such as *PSEN1* and *PSEN2* for AD or *DISC1* for SCZ, has also identified rare variants linked to disease risk (Sassi et al., 2014, Teng et al., 2017). Furthermore, the examination of functional annotation has begun to influence the search for rare variants. For example, studies looking at loss-of-function (LoF) variants, those that will disrupt the production of a functioning protein, have discovered likely-pathogenic rare variants in both AD, SCZ, and ASD (Kenny et al., 2014, Steinberg et al., 2015).

1.2. Epigenetics

In most cases of brain disease, pathology is considered to be attributed to the interactive influence of genetic and environmental factors. Environmental factors might include a number of elements, from social aspects such as education or socio-economic status, to health aspects such as diet or drug use, or the impact of life events such as traumatic experiences. Despite this multitude of environmental factors, many traits in health and disease are believed to be influenced mainly by heritability (Manolio et al., 2009). This has highlighted an issue of ‘missing heritability’ in genetic research as, although many conditions are known to be highly heritable, studies have failed to identify genetic components to explain much of this heritability.

A number of possible explanations have been put forward for the inability to explain missing heritability, mostly concerning the potential for rare variants of large effect to exist that are not investigated through GWAS genotyping (Figure 1.1). However, other possibilities including the statistics used in GWAS analysis, the underestimation of the polygenic nature of genetic effects on specific traits, the difficulty in characterising measurable disease traits for genetic studies, or the idea that estimates of missing heritability could be vastly overestimated (Lee et al., 2011, Zuk et al., 2012).

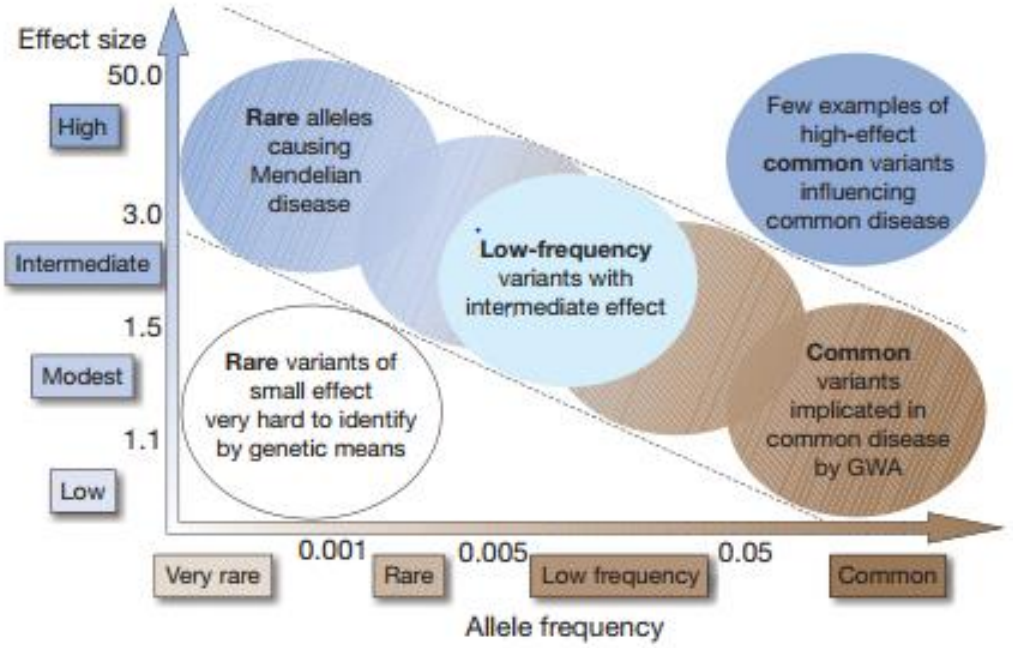


FIGURE 1.1. A model proposing the classification of genetic risk variants according to allele frequency and proposed size of effect. Rare alleles are known to cause Mendelian disease, whereas low frequency variants with intermediate effect and common variants implicated in common disease by GWAS are those that have been most well-studied. Figure taken from Manolio et al. (2009).

In an effort to explain the as-of-yet unidentified factors that act alongside known genetic variants in conferring risk of brain disease, research has turned to exploration of epigenetic mechanisms. The term epigenetics broadly refers to changes to the genome that do not directly alter the nucleotide sequence itself, stemming from the prefix “epi” which means “over” or “around”, but can influence the activity of genes and production of proteins. The list of biological mechanisms involved in carrying out epigenetic changes to the genetic landscape, as well as our understanding of them, is growing rapidly. Broadly, the major forms of epigenetic modification can be divided into those associated with DNA and those associated with RNA. Amongst the most prominent is the methylation of particular nucleotide bases in DNA and RNA, with forms of DNA methylation being the predominant focus of this study. However, there are other notable epigenetics mechanisms including various modifications of the histone proteins that package DNA into nucleosomes within cell nuclei, other chromatin remodelling mechanisms that restructure nucleosomes, and regulation of genetic regions through microRNA (miRNA) interactions.

There are a number of reasons why epigenetics has raised interest in the field of brain disease. One of the most compelling reasons is that epigenetics provides a mechanism of interaction between the environment and one’s genetic make-up. Environmental factors have been shown to act via epigenetic pathways in certain disease phenotypes, particularly cancer (Herceg, 2007). Secondly, epigenetic modifications exhibit both stable and dynamic properties. For example, epigenetic alterations are defined as changes that are heritable through cell reproduction, akin to a cellular memory (Bird, 2002), although the extent to which epigenetic alterations are heritable across generations in humans is open to debate (Slatkin, 2009). Nevertheless, despite the potential for these changes to persist during the lifetime, epigenetic modifications can also vary over time (Fraga et al., 2005) and in response to environmental factors (Alvarado et al., 2014). These elements have identified epigenetics as potentially important tool in elucidating the interplay between genetic and environmental risk factors for brain diseases (Iraola-Guzman et al., 2011).

1.2.1 DNA and RNA methylation

The most comprehensively studied of the aforementioned epigenetic processes is DNA methylation. The best-known form of DNA methylation is the addition of a methyl group to the 5^{th} carbon of cytosine nucleotide bases and hence is referred to as 5-methylcytosine (5mC). This alteration to the nucleotide sequence has most

commonly been associated with the repression of normal genetic transcription, leading to down-regulated gene expression from the methylated genetic locus. 5mC modifications were first characterised and have been heavily researched at CpG dinucleotides (Guo et al., 2011), areas of the genome where cytosine and guanine nucleotides are found beside one another.

The regulation of DNA methylation is carried out by groups of proteins that add the 5mC modification (known as “writers”), interact with the methylated site (“readers”), and assist in removing the methyl groups during demethylation (“erasers”) (Jakovcevski and Akbarian, 2012). The interaction of these proteins with CpG dinucleotides is central to the primary function of DNA methylation, namely regulating transcriptional activity. This can occur when transcriptional machinery is directly impeded by the physical presence of the methyl group or through recruitment of reader proteins, which in turn enlist protein complexes that promote a closed chromatin conformation and subsequent repression of transcriptional activity (Curradi et al., 2002).

The creation of 5mC marks is principally carried out by a group of “writer” proteins known as DNA methyltransferases (DNMTs). DNMTs recognise cytosine nucleotides within a DNA strand and facilitate the nucleophilic substitution of a methyl group to this substrate base (Smith et al., 1992). Within this family of proteins, the process of DNA methylation is further characterised as either maintenance or *de novo* methylation. Maintenance methylation, principally associated with DNMT1, describes the molecular preservation of 5mC marks through cell replication, ensuring that patterns of methylation persist in a heritable manner (Berdasco and Esteller, 2013). DNMT1 also promotes the continuation of DNA methylation by specifically targeting the unmethylated strands of hemimethylated DNA (Plass et al., 2013).

De novo methylation, associated with the function of DNMT3A, DNMT3B, and DNMT3L, describes the addition of new methyl groups to unmethylated CpG dinucleotides. In order to carry out this task, the non-enzymatic DNMT3L has been shown to form a tetrameric complex with the catalytically active DNMT3A (Figure 1.2), stabilising the conformation of DNMT3A and supporting more productive DNA methylation (Jia et al., 2007). In addition, by recognising histone H3 methylation, DNMT3L can guide the DNMT3A-3L complex to carry out methylation at specific nucleotides (Ooi et al., 2007). Although less evidence exists for the interaction

between DNMT3B and DNMT3L, one would assume that it shares functional similarities with the DNMT3A-3L complex.

Although DNA methylation has been the most heavily studied of all epigenetic mechanisms and is the primary focus of this thesis, the burgeoning field of RNA methylation will also be included in some of our analysis. 5mC modifications are found in RNA, but have been more strongly associated with bacteria, archaea, and other mammalian species as opposed to humans (Edelheit et al., 2013, Fu et al., 2014). In humans, the predominant form of RNA methylation is N^6-methyladenosine (m^6A), the addition of a methyl group at the 6^{th} nitrogen of adenosine nucleotide bases. As with DNA methylation, m^6A RNA methylation has its own group of readers, writers, and erasers that add, interact with, and remove the modification. Where interactions with DNA methylation modifications are associated with transcriptional regulation, RNA methylation is associated with regulation of post-transcriptional and translational activity (Zhao et al., 2017). Pertinently for this thesis, work using model organisms has already discovered links between proteins involved in RNA methylation and brain function (Lence et al., 2016).

1.2.2. Dynamic DNA methylation

DNA methylation and the associated proteins play an important role in healthy development and modulation or disruption of epigenetic patterns could be relevant to human disease. For instance, DNMT1 (along with other proteins associated with DNA methylation and histone modifications) is necessary in maintaining parental-specific epigenetic marks during development, termed genetic imprinting (Li and Sasaki, 2011). Whilst it is still unclear how these marks survive the genome-wide epigenetic reprogramming that occurs during preimplantation (Berdasco and Esteller, 2013), research has consistently implicated the loss of DNMT1 during this period with abnormal development (McGraw et al., 2015). Moreover, work has demonstrated that imprinting genes are predominantly expressed in the brain and are particularly important to brain function (Wilkinson et al., 2007).

The majority of research into DNA methylation has focused on the suppression of genetic transcription, particularly associated with promotor regions that contain high densities of hypomethylated CpG dinucleotides, termed CpG islands (Illingworth and Bird, 2009). These areas can be important in disease research, as hypermethylation in CpG islands has been heavily associated with the development of cancer (Herman

and Baylin, 2003). Recent work has rapidly expanded our knowledge of characteristic areas of DNA methylation (Figure 1.3). Around these CpG islands are areas of differential methylation named shores and shelves, with reports of tissue and disease specific methylation patterns in these regions (Alelu-Paz et al., 2016, Doi et al., 2009). Interestingly, specific forms of CpG islands named ravines have been identified, where low levels of methylation in the CpG island contrast with consistently high levels of methylation in the shores and shelves (Edgar et al., 2014). Large genomics regions that contain only isolated CpG sites are referred to as the open sea (Sandoval et al., 2011). Finally, DNA methylation canyons are areas of low methylation, whilst DNA methylation valleys are areas of low methylation that contain genes important in early developmental regulation (Jeong et al., 2014, Xie et al., 2013).

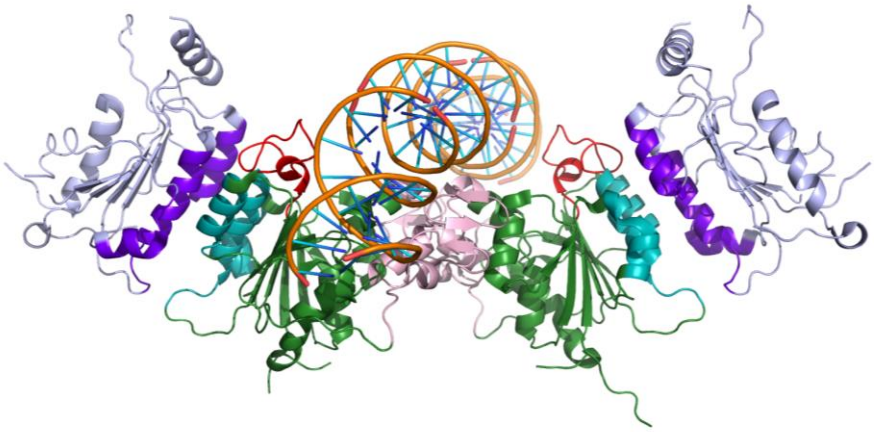


FIGURE 1.2. Model of the DNMT3A-3L tetramer complex with DNA *in situ*. The interaction sites (teal for DNMT3A, purple for DNMT3L) for this complex and the active sites in DNMT3A (red) are highlighted.

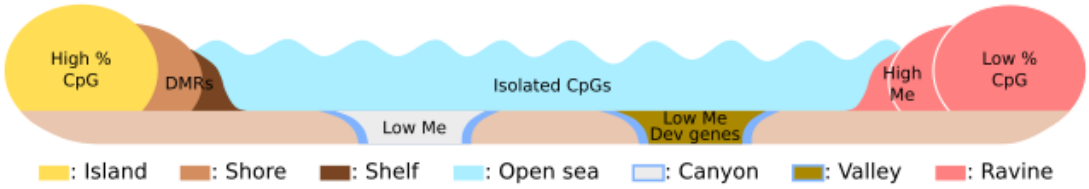


FIGURE 1.3. Depiction of the methylation environment. CpG islands were characterised as areas with a CpG content at least 60% greater than expected, often associated with promoter regions. Surrounding these islands are shores (up to 2kb) and shelves (2-4kb) where DMRs (areas of differential methylation) have been reported. Ravines are islands with a low CpG content but surrounded by highly methylated shores and shelves. The open sea describes large genomic areas with only isolated CpGs. Canyons are conserved regions of low methylation. Finally, valleys are regions mostly lacking in methylation but which harbour genes important in early developmental regulation.

Aside from these areas of CpG methylation, the epigenetic landscape has been complicated further by the identification of non-CpG or CpH methylation, where the H equals A, C, or T nucleotides. These sites could prove to be particularly important in brain disease as they are found in high levels in neurons. Indeed, CpH methylation levels increase most rapidly during synaptogenesis in the developing brain (Huttenlocher and Dabholkar, 1997). Much like CpG methylation, CpH methylation has been associated with transcriptional repression as opposed to activation and is maintained by *de novo* DNMTs (Guo et al., 2014). More detailed patterns of CpH methylation are also beginning to emerge, with tissue specific motifs identified in humans and the identification of large regions lacking CpH methylation, termed CpH deserts (Lister et al., 2013, Schultz et al., 2015).

The DNA methylation landscape is complex and our understanding of regional patterns of methylation is still developing. Another interesting feature of CpG methylation is that it is not stable throughout life. Patterns of methylation are dynamic and can be converted into other oxidised forms of DNA methylation known as 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC). These further methylation modifications are converted via active or passive DNA demethylation. Passive demethylation occurs via the absence of DNMT1 during cell replication, whilst active demethylation involves the oxidation of these further methylation marks by Ten-eleven translocation methylcytosine dioxygenase (TET) enzymes (Wu et al., 2014) (Figure 1.4). 5fC and 5caC are then converted back to unmethylated cytosine by Thymine-DNA glycosylase (TDG) through base excision repair (BER) mechanisms (Neri et al., 2015). Deamination of 5hmC to cytosine has been suggested, but is not favoured in comparison to the BER pathway (Wu and Zhang, 2010).

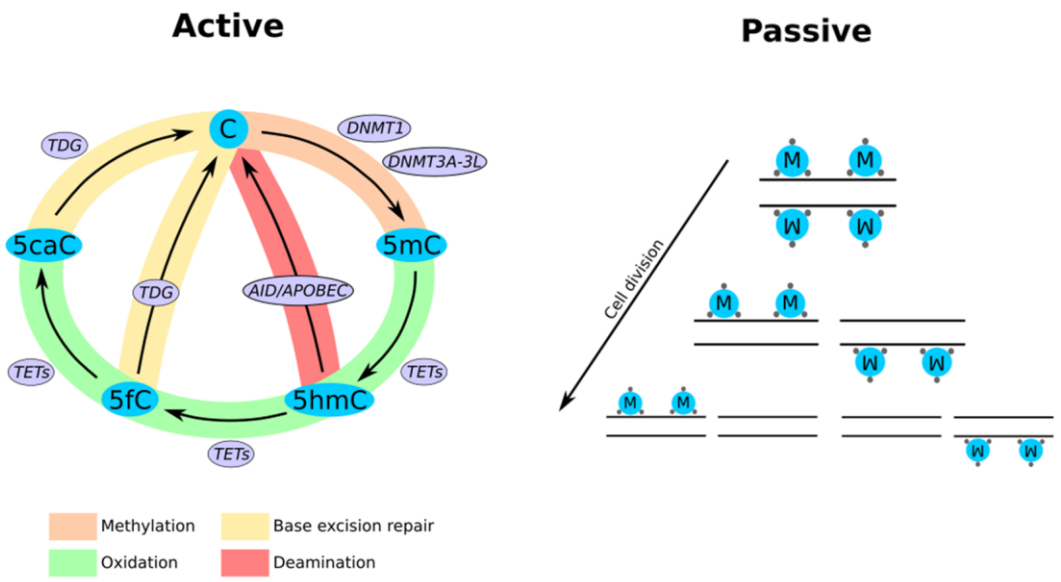


FIGURE 1.4. Pathways of active and passive DNA demethylation. The dominant view of active demethylation is that 5fC and 5caC are converted back to cytosine via TDG-mediated base excision repair (BER), but deamination of 5hmC via AID/APOBEC has also been suggested. Passive demethylation can occur – in the absence of DNMT1 – during cycles of cell division and DNA repair, with the incorporation of non-methylated strands diluting the presence of methylation as cell division continues.

The identification of these further methylation modifications has helped researchers to understand that DNA methylation is involved in more than transcriptional suppression. Indeed, it should be noted that CpG methylation in areas other than promotor regions, such as intragenic 5mC marks, has been associated with transcriptional activation as opposed to suppression (Ball et al., 2009). Similarly, 5hmC enrichment in gene bodies is associated with active transcription of these genes (Wu et al., 2011). Not only is 5hmC found abundantly in the brain, but the distribution of 5hmC within the brain correlates positively with areas pertinent to cognitive function such as the hippocampus and the cerebral cortex (Munzel et al., 2010). Moreover, levels of 5hmC have been shown to increase throughout neurodevelopment (Szulwach et al., 2011), providing further evidence that the dynamic qualities of DNA methylation play an important role in the health of the brain. There is comparatively less evidence surrounding the location and function of 5fC and 5caC, but they also appear to regulate transcription and have been successfully detected in brain tissue (Condliffe et al., 2014, Sayeed et al., 2015).

1.3. DNA methylation and cognitive phenotypes

1.3.1. Brain disease

As stated earlier, the main focus of this thesis will be on cognitive endophenotypes relevant to dementia progression. A number of studies have recently investigated whether significant hypermethylation or hypomethylation events are associated with dementia. Recent work by Lunnon et al. (2014) inspected differentially methylated regions (DMRs) in brain areas associated with AD pathology, corroborating previous reports of significant hypermethylation in the *ankyrin 1* (*ANK1*) gene within the entorhinal cortex. Further, it was shown that these methylation events regulated specific isoform expression of *ANK1* and that certain isoforms correlated strongly with pathology. The regulation of specific isoform expression by methylation is also crucial to the largest genetic risk factor, *APOE* E4. Exon 4 of the *APOE* gene, the location of the polymorphism for the E2, E3, and E4 variants, includes a hypermethylated CpG island. Moreover, the level of methylation at this locus appears dependent on the *APOE* variant one inherits (Yu et al., 2013). This means that the two encoding *APOE* E4 variants are not only key genetic risk factors, but could be potentially significant epialleles in dementia pathogenesis. Indeed, the presence of an *APOE* E4 allele has been shown to affect a downstream methylated CpG island, potentially disrupting the regulatory role of this region (Wang et al., 2008).

The discovery of important methylation events in *APOE* has been corroborated in other AD risk genes. The *amyloid precursor protein* (*APP*) gene that encodes for the amyloid protein, the most definitive sign of AD pathology, presents with aberrant methylation patterns in individuals with AD (Iwata et al., 2014). Furthermore, differential methylation within and upstream of *triggering receptor expressed on myeloid cells 2* (*TREM2*), another risk gene for AD, has been reported (Ozaki et al., 2017, Smith et al., 2016). Aside from targeted methylation, genome-wide sequencing of the methylome has also revealed changes in the genome of individuals with dementia. Methylation sequencing of the temporal cortex of AD patients has revealed hundreds of DMRs enriched in genes associated with brain function and AD pathology (Villela et al., 2016, Watson et al., 2016). Moreover, extension of this sequencing has posited that patterns of methylation change in genes associated with brain function could underlie multiple forms of neurodegeneration (Sanchez-Mut et al., 2016).

As well as changes in DNA methylation patterns, abnormal levels of certain proteins involved in the process of DNA methylation have also been reported. For instance, Mastroeni et al. (2010) investigated a selection of methylation proteins in the entorhinal cortex of AD post-mortem brain tissue, confirming that DNMT1 and the "reader" methyl-CpG-binding domain protein 2 (MBD2) were significantly depleted. Furthermore, treatment of human SH-SY5Y cells with suggested environmental risk factors for AD, such as lead, resulted in a significant decrease in the messenger RNA (mRNA) levels of *DNMT1*, *DNMT3A*, and the "reader" *methyl-CpG-binding protein 2 Rett syndrome* (*MECP2*) (Bihaqi and Zawia, 2012). Moreover, genes associated with BER, a key element of active demethylation, appear to be downregulated in AD patients (Sliwinska et al., 2017). Methylation changes have also been observed within DNA methylation genes themselves in dementia. For example, both *DNMT1* and *methylenetetrahydrofolate reductase* (*MTHFR*), which remethylates methionine necessary for methyl donation, were shown to be hypomethylated in AD brain samples (Wang et al., 2008).

Recent focus has turned to the development of dementia and DNA methylation changes that could precede a diagnosis of AD, with the most notable being associated with MCI. Assessment of global methylation in multiple brain regions from individuals with MCI discovered lower 5mC levels compared to AD and higher 5hmC levels compared to both controls and AD (Ellison et al., 2017). Methylation signatures

in MCI have also been observed for a number of genetic loci. Levels of methylation in the promotor regions of *condensin-2 complex subunit H2* (*NCAPH2*), which encodes for a protein associated with cell division and apoptosis, are significantly depleted in both AD and MCI (Kobayashi et al., 2016). Moreover, the amount of methylation in this region correlates with the rate of brain atrophy specifically in the hippocampus, providing a link between methylation changes and pathology during dementia progression (Shinagawa et al., 2016). Another notable gene is *brain-derived neurotrophic factor* (*BDNF*), which is implicated in many important brain processes such as synaptic transmission and neurogenesis as well as a number of brain diseases including AD (Patterson et al., 1996, Phillips et al., 1991, Scharfman et al., 2005). Research has shown that hypermethylation occurs in the promoter region of the *BDNF* gene in AD and MCI, and that higher levels of methylation at specific CpG dinucleotides in this region are associated with conversion from MCI to AD (Xie et al., 2017a, Xie et al., 2017b).

Aside from dementia progression, this thesis will also investigate individuals with ID and ASD. As with studies of dementia, research has indicated a key role for DNA methylation patterns and genetic variants associated within methylation proteins in these neurodevelopmental conditions. One of the most notable examples is Fragile X syndrome, where the genetic cause is known but the intellectual ability of sufferers can range from unaffected to severely disabled (Garber et al., 2008). However, early work identified DNA methylation patterns as a potential influence on phenotypic variability in this disorder. McConkie-Rosell et al. (1993) determined that levels of hypermethylation in the *FRM1* region were associated with reduced penetrance for cognitive deficits seen in Fragile X syndrome, a discovery that has recently been shown to act as a biomarker for intellectual impairment in this disorder (Godler et al., 2012). Further examples of ID have been connected to mutations in the machinery that carries out DNA methylation, such as some individuals with a rare form of overgrowth syndrome including intellectual disability, a phenotype caused by mutations in *DNMT3A* (Tatton-Brown et al., 2014). Another notable example is Rett syndrome, a neurodevelopmental disorder characterised in part by intellectual deficiencies (Neul et al., 2010), where the vast majority of cases are caused by mutations in the methylation "reader" *MECP2* (Moretti and Zoghbi, 2006).

With regards to ASD, there have also been some links between gene promoter methylation and the prevalence of this disease. For instance, aberrant methylation patterns in the promoter region of *MECP2* have been reported in individuals with

autism and appear to correlate with a reduction of MECP2 expression in ASD brains (Nagarajan et al., 2006, Nagarajan et al., 2008). However, a greater number of studies have investigated global changes to DNA methylation patterns in ASD. Multiple significant DMRs have been identified from the brains of individuals with ASD compared to controls, as well as in twin pairs discordant for ASD (Ladd-Acosta et al., 2014, Wong et al., 2014). Comparison of methylome and transcriptome data has confirmed that hypomethylation of genetic regions in ASD correlates with overexpression of these respective genes. Moreover, many of these genes were associated with key brain functions such as synaptic pruning (Nardone et al., 2014). An interesting connection has also been made with the parents of offspring with ASD. DMRs associated with ASD severity in offspring have been identified in paternal sperm, mostly in genes associated with brain development (Feinberg et al., 2015). The methylation differences seen in those with ASD has led to the formulation of a redox/methylation hypothesis of autism, where oxidative stress disrupts normal DNA methylation and subsequently impairs neural function (Deth et al., 2008). That altered methylation and antioxidant behaviour has also been observed in the parents of offspring with ASD suggests that any heritable portion of ASD may be related to disruptions in DNA methylation (James et al., 2008).

1.3.2. Learning and memory

Outside of the study of disease phenotypes, a growing body of research has investigated the association between DNA methylation and cognitive function in healthy individuals or model organisms. Moreover, much of this research has focused on learning and memory, often viewed as related cognitive functions because both processes rely on each other to realise an operational system of remembering. In neuroscientific terms, both learning and memory are closely associated with synaptic plasticity, a term which describes the modulation of synaptic strength related to activity-dependent synaptic modifications (Takeuchi et al., 2014).

A key issue in the general scientific understanding of memory is the molecular persistence of any given memory trace within the brain, commonly referred to as the engram. Crucially, it has been proposed that epigenetic mechanisms such as DNA methylation could explain how the biological perpetuation of specific memories takes place. Indeed, as highlighted in a review by Day and Sweatt (2011), the notion of DNA methylation being involved in the process of learning and memory has long been discussed (Griffith and Mahler, 1969). The combination of *de novo* and

maintenance methylation accounts for the formation and stable preservation of specific epigenetic marks, whilst the demethylation of 5hmC, 5fC, and 5caC provides a mechanism for the controlled removal of these marks. These mechanisms are necessary features of a molecular candidate for the engram (Callaghan et al., 2014).

Distinct patterns of DNA methylation have been identified during the process of memory formation and storage. As already discussed, the *BDNF* gene is important in both brain development and risk for brain diseases. It has also been strongly implicated in the consolidation of synaptic connections that underpin long term memory (Lu et al., 2008), supported by diminished BDNF activity in neurodegeneration (Mattson et al., 2004). Work utilising rodent models has discovered that an increase in *BDNF* transcription correlates with a decrease in methylation in the *BDNF* promoter region (Martinowich et al., 2003), whilst interaction with the repressive DNA methylation "reader" protein MECP2 can too influence BDNF expression (Chen et al., 2003). Associative learning studies in rodents have also shed light on patterns of methylation during memory formation and storage. Miller and Sweatt (2007) investigated DNA methylation changes in the hippocampus immediately after fear conditioning training, which revealed increased methylation of the memory inhibiting gene *protein phosphatase 1* (*PP1*) coupled with decreased methylation of the memory facilitating gene *reelin* (*RELN*). Further analysis illustrated the dynamic nature of DNA methylation, in that the methylation changes to *PP1* and *RELN* had both returned to baseline levels after 24 hours. In order to develop these findings, Miller et al. (2010) extended the examination of fear conditioning to cortical areas associated with memory storage, such as the dorsomedial prefrontal cortex (dmPFC). Following fear conditioning, an increase in promoter methylation at the memory inhibiting gene *calcineurin* (*CaN*) was observed in the dmPFC, an alteration which persisted for 30 days.

To further confirm the role of DNA methylation in learning and memory, Miller et al. (2010) reported that inhibition of DNMTs in the dmPFC resulted in disrupted memory performance, but only 48 hours after the fear conditioning training took place. In contrast, inhibition of DNMTs in the hippocampus resulted in disrupted memory formation, but had no effect on memory performance for previous conditioned memories (Han et al., 2010), highlighting the discrepancy between memory formation and storage within the brain. In support of these findings, a number of studies have emphasised the role of DNMTs and other DNA methylation proteins in learning and memory. Feng et al. (2010) illustrated that mice expressing genetic knockout (KO) of

both *DNMT1* and *DNMT3A* showed evidence of disrupted synaptic plasticity. The animals exhibited loss of long-term potentiation (LTP), the strengthening of synaptic connections associated with learning and memory, alongside greater long-term depression (LTD), the weakening of synaptic connections.

Decreasing levels of *DNMT3A* have been linked to age related memory problems and wider cognitive deficits, with experimental overexpression of DNMT3A in the mouse hippocampus shown to ameliorate these issues (Oliveira et al., 2012). This work also showed that decreased DNMT3A expression was linked to transcriptional suppression of genes involved in synaptic plasticity, exemplifying the relationship between DNA methylation and active (rather than simply repressed) transcription. Other methylation proteins have also been affiliated with learning and memory, such as members of TET protein family that have been implicated in the process of active DNA demethylation (Kohli and Zhang, 2013). For instance, Rudenko et al. (2013) found that mice expressing KO of *TET1* present specific memory deficits coupled with increased levels of LTD in the hippocampus. Therefore, there is a large body of work suggesting that disruption of the DNA methylation machinery can have significant influence upon memory function in particular.

Research has also been carried out into the association between DNA methylation patterns and intelligence (IQ), a trait that is believed to be highly heritable and could be influenced by epigenetics (Davies et al., 2011). Indeed, recent research has argued that some facets of memory, such as the manipulation of short-term information termed working memory, are in fact highly related to the intelligence scores generated from standard intelligence tests (Alexander and Smales, 1997, Colom et al., 2008). Genome-wide methylation studies have identified loci significantly associated with intelligence, but neither of the genes (*ARHGAP18* and *PRRC1*) appear to be prominently linked to cognition (Rowe et al., 2013, Yu et al., 2012a). Nevertheless, research concentrating on the methylation machinery has implicated a member of the *DNMT* family in abnormal intelligence. Haggarty et al. (2010) examined the association between chosen variants in each of the *DNMT* genes with childhood and adult intelligence scores from two Scottish birth cohorts. The minor allele of one particular variant in the *DNMT3L* gene, R278G (rs7354779), was shown to correlate with greater average intelligence in childhood and in later life, demonstrating that the function of *DNMT3L* may be closely linked to cognition.

1.3.3. B vitamin studies in dementia and aging

A separate, but equally important, facet of DNA methylation in dementia is the association with abnormally high levels of homocysteine (Hcy). Hcy is produced during the methionine pathway within the one-carbon cycle (Figure 1.5), a biochemical pathway that is integral to DNA methylation. Methionine, synthesised from Hcy, is converted into S-adenosylmethionine (SAM) and then S-adenosylhomocysteinase (SAH) before finally being converted back to Hcy (Mitchell et al., 2014). SAM is a principal methyl donor for DNA methylation, but levels of this substrate are diminished as Hcy levels accumulate in AD (Tchantchou et al., 2006). This is also coupled with increased levels of SAH, which acts as an inhibitor of DNMTs and further exacerbates the hypomethylation seen during dementia progression (Lin et al., 2014). Crucially, research has also shown that elevated SAH in AD cases correlates with poorer cognitive performance (Kennedy et al., 2004, Xu and Li, 2012).

Another feature of the methionine cycle is the involvement of a number of key vitamins, particularly vitamins B6, B12, and B9 (otherwise known as folic acid). These vitamins aid the conversion of Hcy into methionine (vitamin B12 and folic acid) or the antioxidant glutothionine (vitamin B6) respectively. Owing to the interaction between the one-carbon cycle, inhibition of methylation machinery, and the progression of cognitive decline in dementia, studies have looked to ascertain if using the aforementioned vitamins as a treatment can lower Hcy levels and reduce the risk of AD progression. Fuso et al. (2005) investigated the impact of methylation and B vitamins in regulating other substantial genetic risk factors for AD, including *APP, PSEN1*, and *β-sectretase 1* (*BACE*) in neuroblastoma cell lines. As well as revealing that these genes present hypomethylation of their promoter regions, removal of vitamin B12 and folic acid from the cell medium resulted in depleted SAM levels and increased expression of *PSEN1* and *BACE*.

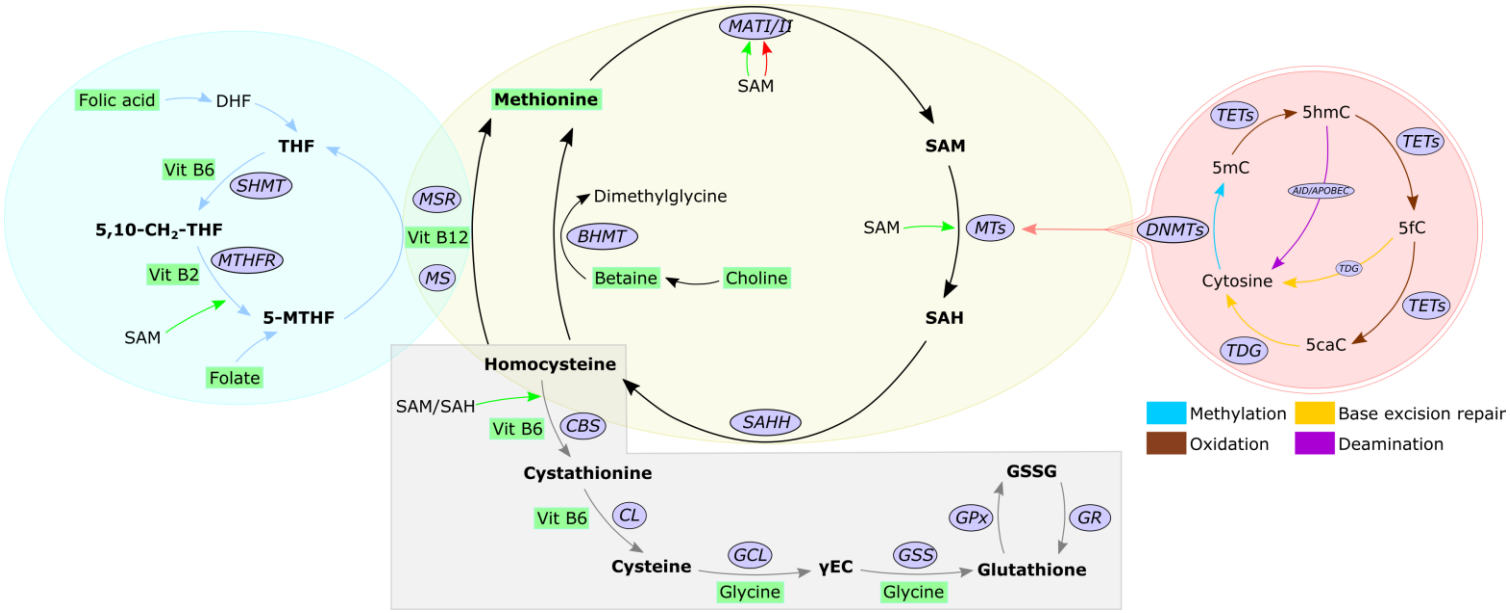


FIGURE 1.5. Depiction of the one-carbon cycle. This cycle is made up of the methionine (yellow), folate (blue), and transsulfuration (grey) pathways, along with the involvement of DNMTs and DNA methylation (red). Areas of dietary and vitamin influence are highlighted in green. Hcy is converted into methionine and then SAM, the major methyl donor for DNA methylation. Hcy can also be converted into glutathione, a key antioxidant in the brain.

In light of this, further trials have been performed to ascertain the impact of B vitamin treatment on AD progression. Sun et al. (2007) administered vitamin B6, vitamin B12, and folic acid to AD patients for 26 weeks. Whilst they saw a significant reduction in Hcy levels in the treatment group compared to the placebo group, this was not coupled with any significant change in cognitive performance. Indeed, Aisen et al. (2008) completed a similar trial for 18 months and reported comparable findings, as well as a risk of adverse psychological effects in the B vitamin treatment group.

As previously mentioned, MCI is primarily characterised as a stage of cognitive decline that lies in between cognitive performance associated with normal aging and the deficits reported in AD. Despite being of broad clinical definition, subtypes including amnestic MCI (focus on poor memory performance) and non-amnestic MCI (focus on cognitive deficits excluding memory) have been constructed to aid classification (Petersen et al., 2014). I have highlighted the small amount of research investigating DNA methylation and cognitive decline in MCI earlier in this section, but the relatively recent characterisation of MCI as a brain disease naturally limits the amount of completed work in the literature. However, despite contention over the role of Hcy in dementia progression (Nilsson et al., 2012), studies have already identified the reduction of Hcy levels through B vitamin treatment as a possible avenue of investigation in MCI.

First, it has been shown that Hcy levels are significantly higher in sufferers of MCI compared to unaffected elderly controls (Kim et al., 2007a). Furthermore, significantly lower levels of SAM and a greater ratio of SAH to SAM have been identified in MCI, with the authors asserting that this would compromise DNA methylation potential, namely the potential for methyltransferases to carry out normal methylation (Zheng et al., 2014). Second, in a study of both MCI and AD, Quadri et al. (2004) confirmed that Hcy measurements increased sequentially from normal aging, followed by MCI, with AD patients reporting the highest levels. Substantially, this examination also discovered that those MCI individuals with the highest levels of Hcy showed significantly impaired cognitive ability in comparison to those with lowest levels of Hcy, providing evidence of a relationship between Hcy and cognitive decline.

Following on from Quadri et al. (2004), the OPTIMA programme began a 2 year trial named VITACOG that set out to investigate the impact of B vitamin treatment on Hcy

levels, cognitive ability, and brain atrophy in a cohort of individuals with MCI. Smith et al. (2010) published a comprehensive analysis of the trial which demonstrated that B vitamin treatment had a considerable influence on these measurements of disease burden in MCI. The group treated with vitamin B6, B12, and folic acid displayed a slower rate of global brain atrophy compared to the placebo group, an effect which was seen most prominently in those treated individuals who had high baseline Hcy measurements. The rate of brain atrophy also correlated with poorer cognitive performance, indicating that B vitamin treatment could abate cognitive decline. Furthermore, this effect was replicated in analysis of brain regions that are specifically linked to cognitive decline in dementia (Douaud et al., 2013). In a further report on the VITACOG project, De Jager et al. (2012) analysed a broader set of cognitive and clinical measurement tools to better assess the impact of B vitamin treatment on individuals with MCI. As well as confirming that Hcy levels were significantly reduced in the treatment group compared to the placebo group, treated individuals who had high baseline Hcy measurements displayed significant protection against cognitive decline in measures of semantic memory and global cognition.

In contrast to these promising results in MCI cohorts, multiple studies investigating the influence of B vitamin treatment on normal aging have rarely shown any positive effects on cognitive performance, contributing to the view that MCI is distinct from the process of normal aging. For example, both Eussen et al. (2006) and a recent meta-analysis from Clarke et al. (2014) reported significant reductions in Hcy following B vitamin treatment in elderly cohorts, but no change in cognitive function. Hvas et al. (2004) also found no effect of B vitamin treatment on cognitive function, but did discover a significant association between raised Hcy levels and poor cognitive performance. Moreover, in contrast to the VITACOG findings, two reviews from the Cochrane Collaboration focusing on vitamin B6, vitamin B12, and folic acid (Malouf et al., 2003, Malouf and Areosa Sastre, 2003) concluded that B vitamin treatment had no notable beneficial influence on cognitive performance in healthy older individuals, those suffering with MCI, or those showing signs of dementia. Nevertheless, models of cognitive aging have shown that genetic knockout of *DNMT1* and the subsequent reduction in DNA methylation negatively influenced cognition function (Liu et al., 2011), whilst rescue of decreased DNMT3A expression was associated with recovery of age-related cognitive decline (Oliveira et al., 2012). This disparity highlights the need for further understanding regarding the role of DNA methylation in brain disease and cognitive aging.

1.4. Aims

The present study hypothesises that functionally important variants in the genes encoding for DNA methylation proteins will influence cognitive function in human brain diseases. In order to elucidate the role of DNA methylation in cognition, the present study will explore the following aims within three independent lines of study. These three lines of research and their specific objectives are:

1. Candidate gene analysis of a DNA methylation variant in MCI and controls

Chapter 1: To assess disease risk associated with the *DNMT3L* R278G variant in the complete OPTIMA cohort of AD, MCI, and aged controls.

Chapter 1: To investigate the association between the *DNMT3L* R278G variant, cognitive function, and components of the one-carbon cycle in the VITACOG cohort of individuals with MCI.

Chapter 2: To examine whether the relationship between the *DNMT3L* R278G variant and cognitive function in the VITACOG study is mirrored in measures of brain atrophy.

Chapter 2: To study the impact of the *DNMT3L* R278G variant on the function of the DNMT3L protein.

Chapter 3: To investigate the association between the *DNMT3L* R278G variant, cognitive function, and components of the one-carbon cycle in the NCDS 1958 and TwinsUK cohorts of general population controls.

2. Investigation of DNA and RNA methylation variants in ID and ASD cohorts

Chapter 4: To design a pipeline for the processing and analysis of next-generation sequencing data from previous studies.

Chapter 4: To assess disease risk associated with multiple DNA and RNA methylation genes in a combination of ID and ASD cohorts from the UK10K project.

3. Immunohistochemistry of further DNA methylation modifications in control brain tissue

Chapter 5: To optimise the immunohistochemical detection of 5hmC and 5caC in rodent brain tissue.

Chapter 5: To characterise the patterns of 5hmC and 5caC deposition in multiple regions within the human brain.

2. Methodology

2.1. Cohorts and samples

2.1.1. OPTIMA cohorts

The Oxford Project to Investigate Memory and Ageing (OPTIMA) is a programme of longitudinal studies examining multiple facets of dementia progression. These principally focused upon individuals with Alzheimer's disease (AD), but also individuals with mild cognitive impairment (MCI) as well as aged controls. Select data on a total of 578 participants was granted to the present study from the OPTIMA programme, along with DNA samples from 576 of the participants granted by David Smith, the lead researcher for the OPTIMA study. Two of the participants were omitted from the present study due to their diagnoses (no diagnosis available and Parkinson's disease respectively), leaving a total of 574 participants for further study. Demographic information for these participants is presented in Table 2.1.

The remaining 574 individuals contain participants from three strands of the OPTIMA programme. The first of these was the principal OPTIMA study, a longitudinal examination that looked to map the entire development of AD from clinical diagnosis to *post mortem*. Participants in this programme were assessed through annual cognitive tests, biochemical measurements, and magnetic resonance imaging (MRI) of changes to brain structure (Johnson et al., 2004). The next strand included was from the Foresight Challenge study, a 3 year examination of control individuals in order to identify potential cognitive or biochemical markers in those who showed greater age related cognitive decline (Budge et al., 2000). Finally, individuals were included from the VITACOG study which also utilised cognitive tests, biochemical measurements, and MRI brain imaging in order to evaluate the impact of B vitamin treatment over a two-year period on the progression of MCI (Smith et al., 2010). The total number of participants used in the present study from each OPTIMA strand is presented in Table 2.2.

TABLE 2.1. Demographic and genotype information available for all OPTIMA cohorts

Variable	AD	MCI	Control
Number	137	295	142
Sex (Female/Male)	69/68	183/112	72/70
DNMT3L R278 (A/A / G+)	73/64	167/128	77/65
APOE (no E4 / E4+)	39/98	207/88	106/36

G+ = carriers of the *DNMT3L* R278G minor G allele; E4+ = carriers of the *APOE* E4 allele.

TABLE 2.2. Number of participants within each strand of the OPTIMA programme

OPTIMA strand	Number
OPTIMA	209
VITACOG	271
Challenge	98

2.1.2. VITACOG cohort

Although the OPTIMA, VITACOG, and Challenge individuals were included in our baseline genotyping assessments, further information including cognitive data and biochemical measurements was only available for the VITACOG study (Table 2.3). Therefore, our principal investigation into the influence of *DNMT3L* R278G and *APOE* on cognitive performance was carried out using the VITACOG individuals.

As outlined in the VITACOG procedural documents (Smith et al., 2010), the following inclusion criteria were required from participants; aged 70 years or over, status of MCI measured by a Telephone Interview for Cognitive Status – Modified (TICS-M) score of 17-29, no diagnosis of dementia, the presence of an informant, possession of auditory and visual abilities necessary to complete the assessments, and a Mini-Mental State Examination (MMSE) score greater than 24. Exclusion criteria included a diagnosis of dementia, a diagnosis of cancer, an episode of stroke within the past 3 months, and involvement in other drug studies. Participants were visited once at baseline and once after the 24 month treatment period was complete, with the assessments lasting roughly 5 hours.

The participants within VITACOG were given either B vitamin treatment or a placebo for the 24 month duration of the study. The treatment consisted of 0.8mg folic acid, 0.5mg cobalamin (vitamin B12), and 20mg pyridoxine (vitamin B6) (16). Unfortunately, further dietary and nutritional information concerning the participants was not available.

2.1.2.1. VITACOG cognitive data

Multiple cognitive tests were made available within the VITACOG cohort. As exemplified in Table 2.3, only a select number of cognitive tests were used in our analyses, with the Clox, Symbol Digit Modalities, Trail Making, and Map Search tests all removed. This was due to the use of the remaining tests in previous research using the VITACOG cohort (De Jager et al., 2003), as well as work looking at cognitive measurements sensitive to the progression of dementia (Blackwell et al., 2004).

TABLE 2.3. Further baseline information available for the VITACOG cohort

Variables	VITACOG
Number	271
Sex:	
Female	169
Male	96
Age at baseline	76.1 ± 7
***DNMT3L* R278G:**	
A/A	150
G+	119
***APOE*:**	
No E4	183
E4+	87
Baseline Hcy levels (μmol/L)	11.4 ± 3.9
Vitamin treatment:	
Treatment	132
Placebo	133
Left study prematurely	6
Baseline cognitive scores:	
HVLT-R Delayed Recall	8 ± 4
CF (Fruit & Vegetables)	20 ± 7
GNT	24 ± 5
MMSE Summary	29 ± 2
PAL Total Errors	10 ± 12

HVLT-R = Hopkins Verbal Learning Test-Revised; CF = Category Fluency; GNT = Graded Naming Test; MMSE = Mini-Mental State Examination; PAL = Paired Associates Learning.
Data presented as median ± interquartile range.

In total, five tests were retained for statistical analysis within the VITACOG cohort. The Hopkins Verbal Learning Test – Revised (HVLT-R) measures verbal recognition memory by reading sets of 12 words, made up of 4 words from 3 semantic groups, to the participants. The present study specifically utilises the HVLT-R Delayed Recall score, whereby the task is to recall as many words as possible after a delay of 20 minutes. The Mini-Mental State Examination (MMSE) is a widely used global test to screen for dementia and decline in cognitive ability by questioning participants on the following functions; Orientation (to time and place), Registration (immediate recall), Attention & Calculation (concentration), Recall (delayed recall), Language (spoken and written).

Cambridge Neuropsychological Test Automated Battery (CANTAB) Paired Associates Learning (PAL) measures episodic memory visuospatially by presenting an increasing number of patterns at different positions on-screen and asking participants to correctly identify the specific location of each pattern. The present study utilised the PAL Total Errors score which summates the amount of errors made in identifying the pattern locations. The Graded Naming Test (GNT) measures object naming and semantic memory by asking participants to name black and white drawings of increasing difficulty. Finally, Category Fluency (CF) was used to measure semantic memory by asking participants to name as many items from a particular category within 1 minute. For the VITACOG visitations, the category used was “fruits and vegetables”.

2.1.2.2. VITACOG atrophy data

In order to assess brain atrophy in the VITACOG cohort, select magnetic resonance imaging (MRI) data was requested from the OPTIMA programme. MRI data along with data for key covariate variables was granted to the present study for 168 of the 271 members of the VITACOG cohort who had consented to undergo MRI scans. Demographic information for these variables is presented in Table 2.4.

TABLE 2.4. Demographics and further covariate information available for the VITACOG individuals included in the atrophy analyses

Variables	VITACOG
Number	168
Sex:	
Female	102
Male	66
Age at baseline	75.9 ± 7
***DNMT3L* R278G:**	
A/A	92
G+	75
***APOE*:**	
No E4	117
E4+	50
Baseline Hcy levels (μmol/L)	11.1 ± 3.5
Vitamin treatment:	
Treatment	85
Placebo	83
Baseline cognitive scores:	
HVLT-R Delayed Recall	8 ± 4
CF (Fruit & Vegetables)	20 ± 7
GNT	24 ± 5
MMSE Summary	29 ± 2
PAL Total Errors	10 ± 11
Brain imaging:	
Baseline brain volume (mL)	1371.6 ± 97
Yearly ROA (%)	0.8 ± 0.8
Biochemical measures:	
Folate (Vitamin B9) (nmol/L)	21.9 ± 23.5
Cobalamin (Vitamin B12) (pmol/L)	337 ± 158.5
Creatinine (nmol/L)	94 ± 18.5
Diastolic BP	80 ± 16.5
Baseline cognitive assessments:	
GDS	6 ± 6

HVLT-R = Hopkins Verbal Learning Test-Revised; CF = Category Fluency; GNT = Graded Naming Test; MMSE = Mini-Mental State Examination; PAL = Paired Associates Learning; ROA = Rate Of Atrophy; GDS = Geriatric Depression Scale. Data presented as median ± interquartile range.

As described in the primary publication regarding the VITACOG cohort (Smith et al., 2010), scans were taken at baseline and follow up 24 months later on a 1.5 Tesla MRI system with $1mm^3$ isotropic resolution at the Oxford Centre for Clinical Magnetic Resonance Research. High-resolution volumetric T1-weight images were taken over 3 repeats and averaged after alignment. SIENA analysis software was used by the OPTIMA team to derive the change in volume between the images from baseline and follow up. SIENA achieves this by automatically segmenting brain and non-brain through the use of other FMRIB Software Library tools, followed by the estimation of change between the two images for both white and grey matter that can be converted into a percentage rate of atrophy (ROA) for the whole brain. Data concerning regional brain atrophy was not made available for the present study, but has been reported in further studies using the VITACOG cohort (Douaud et al., 2013).

A number of other factors that had been measured in previous work using the VITACOG cohort were also provided along with the ROA data. Plasma biochemical levels of folate (vitamin B9), cobalamin (vitamin B12), creatinine, as well as diastolic blood pressure measurements were taken at the beginning of the trial and after 24 months. Folate and cobalamin can provide a measure of the influence of the B vitamin treatment provided during the VITACOG trial. Creatinine is a product of muscle metabolism that is removed from the blood by the kidneys but that can also act as a measure of some vitamin levels (Chen et al., 2002).

In addition, individuals were asked to complete the Geriatric Depression Scale (GDS), a self-report questionnaire of 30 yes-or-no questions that aims to estimate normal (0-9), mild depression (10-19), or severe depression (20-30) within a sample.

2.1.3. Control cohorts

To provide an accurate control comparison to the VITACOG MCI individuals, replication of our analysis of vitamins, cognitive performance, and *DNMT3L* R278G was sought in general population controls. Replication of *APOE* could not be achieved because access to this genotype was prohibited in the control cohorts. Despite this, access was granted for data from both the National Child Development Study (NCDS) 1958 birth cohort and the UK10K TwinsUK cohort.

2.1.3.1. NCDS 1958 cohort

NCDS 1958 is a longitudinal birth cohort of roughly 17,000 individuals all born within one week in 1958. These individuals have been periodically surveyed over the past 59 years to assess multiple lifestyle variables, including information about their health. For the NCDS 1958, application was arranged and approved for relevant cognitive and demographic variables for 9377 individuals through the UK Data Service. Specifically, biochemical data was gathered from the NCDS Biomedical Data survey (2002-2004) whilst cognitive data was gathered from NCDS Sweep 8 survey (2008-2009). This meant that the members of NCDS 1958 were approximately 45 years old when biochemical and health data was gathered and approximately 50 years old when cognitive testing was conducted.

The NCDS 1958 biomedical survey was a specially funded initiative organised outside of the main schedule of NCDS survey sweeps in order to obtain health measurements. Nurse visitations were organised to carry out the computer assisted personal interviews (CAPI), the physical measurements and blood taking, and the health questionnaire. Sweep 8 consisted of a computer assisted interview and series of cognitive assessments lasting roughly one hour, followed by a questionnaire. Genetic data for NCDS 1958 was sought via a separate application to the METADAC committee that curates datasets from many major UK cohort studies. Access was granted to genotyping data for 6812 individuals performed using the Illumina ImmunoBeadChip platform. After joining the genetic data with the cognitive and demographic data, 5489 individuals remained for our analyses. Demographic information for these individuals is presented in Table 2.5.

2.1.3.2. TwinsUK cohort

TwinsUK is a longitudinal registry of around 12,000 twins of all ages which began in 1992. Much like the NCDS 1958, these individuals have been consistently surveyed in the ensuing 25 years to assess multiple factors associated with health and lifestyle. For the TwinsUK cohort, application was arranged and approved for the use of relevant biochemical, cognitive, and demographic variables through the Department of Twins Research, Kings College London. Access was granted by the TwinsUK Resource Executive Committee.

TABLE 2.5. Demographic information for the NCDS 1958 cohort

Variables	**NCDS 1958**
Number	5489
Sex:	
Female	2793
Male	2696
Age	-
***DNMT3L* R278G:**	
A/A	2872
G+	2617
Vitamin intake:	
Regular	861
Irregular	186
Non-reported	4102
Baseline cognitive scores:	
WR	7 ± 2
DWR	6 ± 3
AN	22 ± 8
Ps & Ws Scanned	25 ± 4
Ps & Ws Missed	3 ± 4

WR = Word Recall; DWR = Delayed Word Recall; AN = Animal Naming.
Data presented as median ± interquartile range.

A subset of the total TwinsUK cohort was received that included no monozygotic or dizygotic twin pairs. As such, this data represented a group of unrelated individuals rather than a group of related twins. The data provided for TwinsUK is conglomerated from multiple stages of clinical assessments, from baseline assessments between 1992 and 2004 to follow up assessments continuing from 2004 onwards. As such, age of participants was provided alongside every variable as measurements were taken on multiple occasions. Age at Hcy measurement was used for demographic purposes as it was the most complete age variable available. All individuals within the TwinsUK cohort are female.

Genetic data for the TwinsUK cohort is managed by the UK10K program, a nationwide initiative to bring together genetic information from 10,000 individuals. Data access was agreed with the UK10K Project for a specific set of sequencing data for a number of cohort studies, including 1870 individuals from the TwinsUK cohort (accession numbers EGAD00001000194, EGAD00001000741, EGAD00001000790). Demographic information for these individuals is presented in Table 2.6.

2.1.3.3. Control cohort vitamin variables

In comparison to the VITACOG trial of B vitamins, the NCDS 1958 and TwinsUK cohorts are longitudinal surveys of general health variables. Therefore, they cannot provide an exact replication of the B vitamin treatment and placebo groups. However, self-reported questionnaire information was available from both cohorts regarding daily use of vitamins, including data on specific vitamins that the individuals were consuming. In NCDS 1958, access was gained to self-reported information on daily use of combined vitamins, single vitamins, and folate. In TwinsUK, access was gained to self-reported information on daily use of general vitamins as well as details on the vitamins used. From this variable, I was able to extrapolate whether an individual specifically takes B vitamins or not. From TwinsUK, access was available to biochemical measures of serum vitamin B12 and folate levels, as well as serum homocysteine (Hcy) levels. In line with our VITACOG analyses, Hcy measurements were split into high and low groups based upon the median baseline Hcy value. High and low Hcy was then split again to create middle high and middle low Hcy levels. Hcy measurements were not available from NCDS 1958.

TABLE 2.6. Demographic information for the TwinsUK cohort

Variables	TwinsUK
Number	1870
Sex	-
Age	59.7 ± 16.3
***DNMT3L* R278G:**	
A/A	996
G+	832
Hcy levels (μmol/L)	11.5 ± 5.9
Vitamin levels:	
Folate (nmol/L)	13.5 ± 12.4
Vitamin B12 (pmol/L)	572.5 ± 301
Vitamin intake:	
Regular:	1117
Took B vitamins	295
Non-reported	386
Baseline cognitive scores:	
PAL Total Errors	15 ± 17
DMS Correct	17 ± 3
DMS Mean Latency	3715.5 ± 1506
PRM Correct	21 ± 3
PRM Mean Latency	2071.5 ± 601
SSP Length	5 ± 1
SWM Within Errors	2 ± 5
5 Choice RT	354 ± 68
IED Set Shift Errors	22 ± 42

PAL = Paired Associates Learning; DMS = Delayed Matching to Sample; PRM = Pattern Recognition Memory; SSP = Spatial Span; SWM = Spatial Working Memory; RT = Reaction Time; IED = Intra-Extra Dimensional.
Data presented as median ± interquartile range.

Using the self-reported daily intake variables from NCDS 1958 and TwinsUK, new variables were derived for each cohort that described whether vitamin intake was regular, irregular (NCDS 1958 only), or non-reported. With regards to NCDS 1958, regular intake included individuals who take single vitamins, combined vitamins, or folate once a day or more, or three to six times a week. Irregular intake included individuals who take single vitamins, combined vitamins, or folate twice a week or less, or not within the last month. With regards to TwinsUK, regular intake included individuals who reported daily intake of vitamins.

2.1.3.4. Control cohort cognitive variables

NCDS 1958 participants all completed the following five cognitive tests during an interview with the aid of a CAPI program. Word recall (WR) involved the participants learning a list of 10 words read to them by the interviewer or via the CAPI program, and then recalling as many of these words as possible in two minutes. Delayed word recall (DWR) simply involved recalling as many of the words from the word recall test at the end of the interview, roughly thirty minutes after learning the list of words. Animal naming (AN) is a standard semantic fluency task which measures how many words a participant can name from a selected category, with the category tested being animals. Repetition of animals and use of character names (such as Scooby Doo) were not allowed. Finally, the Letter Cancellation test measures attentional speed and visual scanning by asking participants to mark as many “Ps” and “Ws” as possible within a document of random letters. This produced a speed score of the total number of Ps & Ws scanned, and an accuracy score of the number of Ps & Ws which were scanned but not marked (i.e. missed).

TwinsUK participants all completed a number of key tasks from the CANTAB library of cognitive tests. These are completed using a computer equipped with the CANTAB software and a touch screen. As with the VITACOG cohort, Paired Associates Learning (PAL) measures episodic memory visuospatially by presenting an increasing number of patterns at different positions on-screen and asking participants to correctly identify the specific location of each pattern. Again, the PAL Total Errors score was specifically utilised. Although the PAL test was used in both VITACOG and TwinsUK, the PAL Total Errors score was available from VITACOG and the PAL Total Errors (adjusted) was available from TwinsUK, accounting for the discrepant scores between the two cohorts. Delayed Matching to Sample (DMS) also assesses visuospatial memory by presenting a pattern to participants, followed by a

delay, before asking the participant to identify this pattern from a selection of four. The number of correct identifications and the mean latency of identification was used. Pattern Recognition Memory (PRM) involves the presentation of twelve consecutive visual patterns before asking the participant to identify a previously-presented pattern from a selection of two. The number of correct identifications and the mean latency of identification was used.

Spatial Span (SSP) again assesses visuospatial memory by presenting a series of boxes, some of which change colour, and asking participants to reproduce the order of colour changes that appeared. The longest reproduction of colour order was utilised. In the Spatial Working Memory (SWM) task, the object is to uncover hidden tokens beneath a number of boxes and move them to the side of the screen. Within-search errors, where the participant checks the same box for a token more than once, were assessed in this study. The 5 Choice Reaction Time test simply presents five circles to the participant and measures how quickly they respond to a dot appearing in one of the circles. Finally, Intra-Extra Dimensional (IED) Set Shift tests visual discrimination and attention by asking participants to identify simple stimuli. The stimuli or the rules are then changed and the participants must pass each stage. The number of errors was utilised.

2.1.4. Next-generation sequencing cohorts

Access was granted to genetic sequencing data from multiple disease cohorts within the UK10K project. The UK10K project is a large-scale collaborative study beginning in 2010 with the aim of sequencing 10,000 genomes, 4000 at the whole-genome sequencing level and 6000 at the whole-exome sequencing level. In doing this, the study aimed to better understand the contribution of rare (minor allele frequency <0.01%) and low-frequency (minor allele frequency <0.05%) coding variants to various human diseases.

Our study focused on the analysis of seven specific cohorts available from the UK10K project, four of which included individuals with autism spectrum disorders (ASD), two of which included individuals with intellectual disabilities (ID), and one which included healthy general population controls. Beginning with the ASD cohorts, the Gallagher cohort is a sample of 75 individuals from Ireland diagnosed with a specific subset of ASD due to the level of comorbid intellectual disability present in this cohort. The IMGSAC (International Molecular Genetic Study of Autism

Consortium) cohort is a study of families with children diagnosed with ASD. Although the study is international, the 114 individuals included in the UK10K IMGSAC dataset are from the UK only. MGAS (Molecular Genetics of Autism Study) are a clinical sample of 97 individuals with ASD who were collected at Maudsley Hospital, London. Skuse is a cohort of 334 children and adults with autism, Asperger's syndrome, or atypical autism. Whilst the majority were free from signs of intellectual disability, a minority had comorbid neurodevelopmental issues such as attention deficit hyperactivity disorder (ADHD). Individuals from Gallagher, IMGSAC, MGAS, and Skuse were all diagnosed using the Autism Diagnostic Interview-Revised (ADI-R) and Autism Diagnostic Observation Schedule (ADOS) assessments.

The ID cohorts assessed in this study included Rare FIND (Familial Intellectual Disability), a cohort of 124 families with intellectual impairments taken from the larger UK10K Rare Disease Sample Sets. The majority of individuals within the Rare FIND cohort have moderate to severe ID. The Muir cohort consists of 166 individuals with learning disability, many of whom also presented with comorbid schizophrenia or other psychoses, representing some of the more severe neurodevelopmental disorders.

Finally, the TwinsUK cohort of 1854 individuals, a registry of monozygotic and dizygotic twins who represent the general population in terms of disease incidence, was used as a control cohort to compare with our ASD and ID populations.

2.1.5. Brain tissue samples

To characterise the methylation modifications 5-hydroxymethylcytosine (5hmC) and 5-carboxylcytosine (5caC) in the brain, both rodent brain tissue and human brain tissue were utilised.

Whole paraffin-embedded fixed rodent brains were obtained from the University of Nottingham animal house, specifically from the sacrifice of animals where the brain was not required for further experiments. Fixation was completed by Michael Garle, a senior technician associated with the animal house. The rodents were healthy animals of no specific transgenic line. The brain sections were cut by our experienced lab technician, Maria Haig, who used the microtome facilities at the School of Life Sciences Imaging (SLIM) centre, University of Nottingham. Sections were cut for the whole brain at a thickness of 10 microns. Sections were then heated

and mounted onto standard 1mm thick microscope slides. The rodent sections were cut across the coronal plane rather than the traditional sagittal plane for rodent work as this better represents the cut from the human samples.

Human brain tissue sections were obtained from the Nottingham Health Science Biobank (NHSB), an NHS trust-led initiative to provide human biomedical samples to local translational health studies. Information on the individuals who kindly donated their brain tissue to the NHSB is presented in Table 2.7. All individuals were elderly and died as 'unaffected healthy controls', with only three of the ten brains showing signs of age-related degeneration. The average age was 80 years old (SD 9.6). Six of the individuals were male and four were female. Brains were fixed and cut across the coronal plane by the staff at the NHSB in accordance with the requirements of our tissue application. Six brain areas were requested; caudate/putamen (in an effort to include the subventricular zone), cingulate gyrus, middle frontal gyrus, hippocampus, occipital cortex, and cerebellum. Ten sections per area were received from each individual, totalling six hundred sections. These sections were paraffin-embedded and mounted to standard 1mm thick microscope slides.

Approval for the human tissue sections was gained from the NHSB under Helen Knight on 03/08/2015, biobank reference ID ACP0000101. However, in accordance with the Human Tissue Act (HTA) 2004 within the UK, a HTA licence was required to ensure that our laboratory adheres to the national HTA legislation on storage, use, and disposal of human tissue. The HTA licence for our laboratory was approved on 16/09/2016 and is currently held by Helen Knight. In order to gain the HTA licence, standard operating procedures (SOP's) for all human tissue-related protocols had to be produced, including for the use of our human tissue record masterlist. Each member of the laboratory who planned to use the human tissue also had to complete a HTA induction within the laboratory, as well as a HTA questionnaire provided by the local HTA person designate. Our human tissue samples were stored at room temperature in our HTA licenced storage room. An internal audit of the tissue was carried out every six months, whilst external audit by HTA officials was carried out once a year.

TABLE 2.7. Information on the individuals who provided brain tissue

Case ID	Diagnosis	Age	Sex
NP 19/03	Normal brain	74	Male
NP 50/03	Normal brain	83	Female
NP 17/04	Normal brain	75	Female
NP 36/04	Normal brain	78	Male
NP 42/04	Normal brain	78	Male
NP 15/06	Mild changes associated with hypoxia/ischaemic injury during cardiac arrest. Within normal limits.	76	Male
NP 22/06	Normal brain	99	Female
NP 41/06	Mild age-related changes. Within normal limits	92	Female
NP 62/06	Mild age-related changes. Within normal limits	86	Male
NP 85/07	Normal brain	63	Male

2.2. Genetic data

2.2.1. OPTIMA and VITACOG

Genotyping data for the *APOE* gene was provided by the OPTIMA project for the OPTIMA, VITACOG, and Challenge study individuals. However, genotyping data for *DNMT3L* R278G had to be generated in our research group using blood samples provided by David Smith of the OPTIMA project. The 574 eligible OPTIMA participants were genotyped for the DNMT3L R278G variant by Mariam Muse, a Master's student in our research group, using Kompetitive Allele Specific Polymerase Chain Reaction (KASP) (LGC Genomics) and confirmed through Sanger sequencing, as described in a recent methodological note (Braae et al., 2014).

The KASP assay utilises bi-allelic discrimination via two allele-specific forward primers to amplify the target sequence, coupled with a fluorescence resonant energy transfer (FRET) cassette which allows for primer-specific, and thus allele-specific, fluorescence to determine the heterozygous or homozygous nature of the genotype. The KASP assay primers were designed by LGC Genomics and are presented in Appendix 1. To validate the results of the KASP assay, a subset of 11 samples were utilised for Sanger sequencing. Standard Polymerase Chain Reaction (PCR) was carried out using primer pairs designed with the Primer3 program (Untergrasser et al., 2012). Following amplification and filtration, the samples were sequenced using an ABI automated DNA sequencer (Cambridge Biosystems). An example of the Sanger sequencing results is presented in Appendix 2.

2.2.2. NCDS 1958 and TwinsUK

For the NCDS 1958 cohort, whole-exome sequencing of 6812 of the cohort individuals had previously been performed using an Illumina ImmunoBeadChip array (accession number EGAD00001000248). Data was provided by METADAC as two sets of PED and MAP files, entitled "UVA" and "Sanger". The PED file consisted of sample and pedigree information along with allele calls for the respective variants, and the MAP file contained chromosomal and base-pair coordinates for each variant. The tool PLINK was used to convert the PED and MAP files to a binary PED (BED) file to compact the data and speed up subsequent analysis. To identify the *DNMT3L* R278G variant, PLINK was also used to cut out the *DNMT3L* region into a BED and BIM file which could be viewed in a text editor (Appendix 3).

As the hg18 genome assembly from the chip data did not align with the variant position in hg19, a surrogate marker found in both builds was used to locate the R278G variant. Two sets of PED and MAP files were received ("UVA" and "Sanger" respectively), meaning that this process was repeated for both sets of files. Assessment of the two sets of genotyping for this variant resulted in discrepant calls for 9 individuals (Appendix 4). In this case, genotypes from the dataset that had provided the majority of the calls were used ("UVA").

For the TwinsUK cohort, whole-genome sequencing data was made available via the European Genome-phenome Archive (EGA) relating to 1870 individuals from this cohort (accession number EGAD00001000741). Because the TwinsUK sequencing coverage was whole-genome as opposed to whole-exome, VCF files were available by chromosome rather than by genome. The EGA download client was used to access the genetic data via the Windows Command Prompt or the Mac/Linux Terminal. A masterlist of EGA dataset filenames was requested which allowed one to link the EGAR run names found in the TwinsUK meta-info files to the EGAF file names necessary to request, download, and decrypt the VCF files for chromosome 21 where *DNMT3L* is located (Appendix 5). The tool VCFtools was then used to splice out the *DNMT3L* gene into a smaller VCF file which could be opened in a text editor (Appendix 6).

2.2.3. Next-generation sequencing cohorts

A data access agreement with the UK10K project was agreed for genetic sequencing data from the ASD, ID, and control cohorts. This allowed for the downloading of whole-genome or whole-exome sequencing data for each cohort through the EGA. As multiple datasets containing different numbers of individuals were available for each cohort, I chose to download the dataset with the greatest number of individuals where possible. EGA accession numbers for the datasets chosen and used in our analysis are presented in Table 2.8.

TABLE 2.8. Details for each NGS cohort

Cohort	Disease	Number	EGA accession number
Rare FIND	ID	124	EGAD00001000416
Muir	ID	166	EGAD00001000443
Gallagher	ASD	75	EGAD00001000436
IMGSAC	ASD	114	EGAD00001000239
MGAS	ASD	97	EGAD00001000613
Skuse	ASD	334	EGAD00001000614
TwinsUK	Control	1854	EGAD00001000741

ID = Intellectual Disability; ASD = Autism Spectrum Disorder; IMGSAC = International Molecular Genetic Study of Autism Consortium; MGAS = Molecular Genetics of Autism Study.

The ASD and ID cohorts were all whole-exome sequenced and the TwinsUK controls were whole-genome sequenced. Sequencing was carried out for all cohorts using either the Illumina HiSeq 2000 or Illumina Genome Analyzer II platforms. Download of these sequencing datasets was completed using the EGA download streamer, a command-line client and API (application programming interface) designed to ensure secure transmission of data from the EGA servers to our local storage. In order to obtain the data access agreement, good practice for data storage and use had to be implemented within the lab. This required levels of security to be placed upon our local research group storage to ensure that only those with rights to access the data could be involved in the download and analysis of the sequencing datasets.

First the latest version of the EGA download client was downloaded, which can be accessed via the Windows Command Prompt or the Mac/Linux Terminal. The commands to download sequencing datasets from EGA can be enacted either by loading the EGA shell client within the command prompt or by calling the client using direct commands. A masterlist of EGA dataset filenames was requested which allowed one to link the EGAR run names found in the meta-info files for each dataset to the EGAF file names necessary to download individual VCF or BAM (binary sequence alignment map) files (Appendix 5). Whilst calling the EGA client and entering the username and password every time is more cumbersome than logging into the shell client, using this direct command method allowed one to decrypt multiple files simultaneously after the “-dc” switch.

Select BAM files were also downloaded for specific individuals to corroborate the sequencing quality in the VCF files if and where appropriate. The steps taken were identical to those illustrated above for the downloading of VCF files. The only alterations were the linking of EGAR run numbers to BAM file names and the selection of BAM file names within the EGA download client in place of VCF file names.

2.3. Next-generation sequencing pipeline

In order to analyse the next-generation sequencing (NGS) data from the Autism Spectrum Disorder (ASD), Intellectual Disability (ID), and control datasets, a pipeline was designed to assess datasets from multiple studies that were designed with different aims and subsequently used discrepant NGS platforms. This involved initial access of genetic data from each study, selection of genetic regions of interest for

the present study, annotation of variant positions, merging the datasets for each gene of interest, and applying genetic imputation to resolve issues associated with discrepant sequencing (Figure 2.1).

2.3.1. Gene selection

Using the EGA download client, complete whole-exome sequencing VCF files were downloaded for the ASD and ID cohorts, and whole-genome sequencing VCF files for each chromosome in the TwinsUK cohort. I had originally planned to include data from the large follow-up study to the Rare FIND cohort that included 1151 individuals. However, this dataset was excluded due to poor sequencing quality.

31 genes involved in DNA and RNA methylation were included. This list is by no means exhaustive with regards to genes involved in methylation. For example, some notable methylation-related genes were omitted that have already been investigated in the UK10K cohorts such as *SETD1A* (Singh et al., 2016). Moreover, though the focus of our previous work has mainly been on DNA methylation, recent advancements in the study of RNA methylation have associated modifications such as N^6-methyladenosine (m^6A) with brain functions such as synaptic plasticity (Hussain and Bashir, 2015). As RNA methylation has become a focus of other members in our research group, it was decided to include some prominent and recently identified RNA methylation genes.

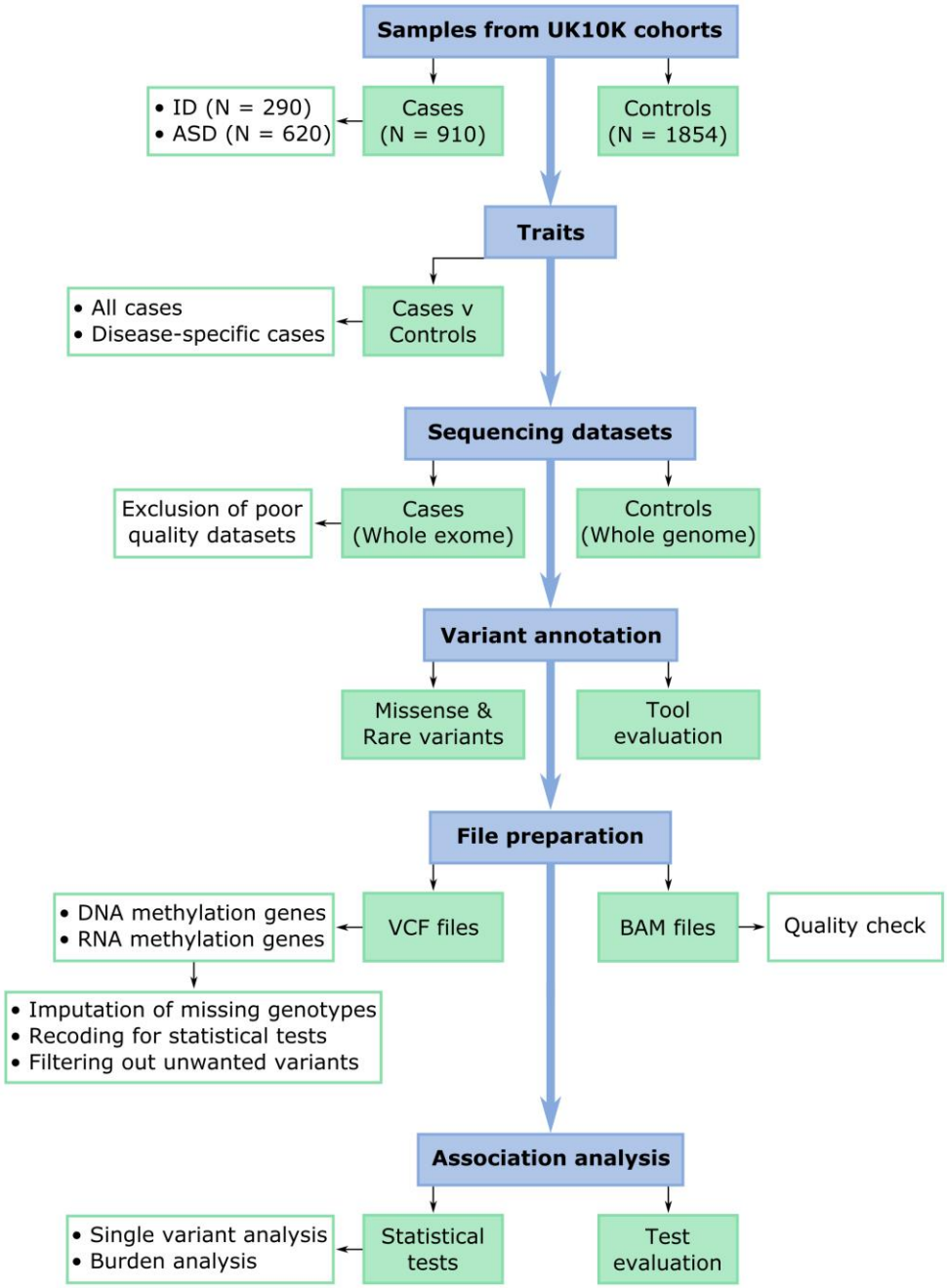


FIGURE 2.1. Next-generation sequencing pipeline used for the UK10K datasets. Data from 910 cases and 1854 controls was obtained from the UK10K project. Although further information was available for the controls, our genetic analysis was based on a case-control phenotypic outcome. Available EGA genetic datasets were downloaded for each case and control cohort. Variant annotation tools were used to identify the functional class and minor allele frequency of each variant. Our genes of interest were spliced from the VCF files, missing data was imputed, and the data was processed according to the statistical tests necessary. Single variant and burden analysis was then run for each gene of interest.

A list of the 31 DNA and RNA methylation genes along with their chromosomal position, a brief description of their function, and their broad categorisation based upon their function is presented in Table 2.9. Using this list of genes and chromosomal positions, regions were cut out of the NGS sequencing datasets for the ASD, ID, and control cohorts. A command-line tool named VCFtools was used, a package of C++ executables and PERL scripts which allow the user to access, manipulate, and analyse VCF format sequencing files. In particular, VCFtools was used to cut our genes from the larger whole-exome and chromosomal whole-genome datasets, and then convert these files from VCF format to a .tab file. The format of each variant position within the original VCF files show the reference and alternative alleles (e.g. A, T) followed by the binary genotype for each individual (e.g. 0/0, 0/1, 1/1). After the .tab conversion, each variant position shows the allelic genotypes (e.g. A/A, A/T, T/T) that allows for easier analysis downstream in our NGS pipeline (Appendix 6).

These steps were applied to each gene for each sequencing dataset from each cohort. A two kilobase (kb) window was added to each side of the chromosomal positions presented in Table 2.9 in an attempt to include any 3', 5', or splicing variants at the extremities of each genetic region. A further command in VCFtools was used to generate minor allele frequencies (MAF) for particular variants in particular cohorts, or indeed all of the cohorts combined to give a total MAF for our study population (Appendix 7).

The MAF was also calculated for each variant within each dataset during the variant annotation process. This was in order to visually assess any discrepancies in the data (e.g. poor sequencing quality) that could lead to inaccurate MAF calculation. In addition, this method was helpful in identifying interesting variants early on in the annotation process. A simple calculation was used to compute this:

- ((homozygotes * 2) + heterozygotes) / total

TABLE 2.9. DNA and RNA methylation genes included in the NGS analysis

Gene	Position	Category	Function
DNMT1	chr 19: 10,244,022 - 10,305,755	DNA methylation; Writer	Maintenance MT (Bestor et al., 1988)
DNMT3A	chr 2: 25,455,830 - 25,564,784	DNA methylation; Writer	De novo MT (Okano et al., 1999)
DNMT3B	chr 20: 31,350,191 - 31,397,162	DNA methylation; Writer	De novo MT (Okano et al., 1999)
DNMT3L	chr 21: 45,666,222 - 45,682,099	DNA methylation; Writer	Interacts with de novo MTs (Hata et al., 2002)
MBD1	chr 18: 47,795,216 - 47,808,144	DNA methylation; Reader	Methylation-mediated transcriptional repression (Fujita et al., 2003)
MBD3	chr 19: 1,576,678 - 1,592,652	DNA methylation; Reader	Binds to 5hmC (Yildirim et al., 2011)
MECP2	chr X: 153,287,264 - 153,363,188	DNA methylation; Reader	Methylation-mediated transcriptional activation & repression (Chahrour et al., 2008)
UHRF1	chr 19: 4,909,510 - 4,962,165	DNA methylation; Reader	Targets DNMT1 to hemimethylated sites (Du et al., 2015)
UHRF2	chr 9: 6,413,151 - 6,507,051	DNA methylation; Reader	Promotes TET oxidation of 5mC (Spruijt et al., 2013)
TDG	chr 12: 104,359,593 - 104,382,656	DNA methylation; Reader	Facilitates demethylation (Shen et al., 2013)
NEIL1	chr 15: 75,639,960 - 75,647,592	DNA methylation; Reader	Binds to 5hmC (Spruijt et al., 2013)

NEIL3	chr 4: 178,230,991 - 178,284,092	DNA methylation; Reader	Binds to 5hmC (Spruijt et al., 2013)
WDR76	chr 15: 44,119,112 - 44,160,617	DNA methylation; Reader	Binds to 5hmC (Spruijt et al., 2013)
THYN1	chr 11: 134,118,173 - 134,123,260	DNA methylation; Reader	Binds to 5hmC (Spruijt et al., 2013)
TET1	chr 10: 70,320,117 - 70,454,239	DNA methylation; Eraser	Oxidation dMT (Rudenko et al., 2013)
TET2	chr 4: 106,067,842 - 106,200,960	DNA methylation; Eraser	Oxidation dMT (Rudenko et al., 2013)
TET3	chr 2: 74,273,405 - 74,335,302	DNA methylation; Eraser	Oxidation dMT (Rudenko et al., 2013)
GADD45A	chr 1: 68,150,860 - 68,154,021	DNA methylation; Eraser	Promotes TDG-TET demethylation (Ma et al., 2009)
GADD45B	chr 19: 2,476,123 - 2,478,257	DNA methylation; Eraser	Promotes TDG-TET demethylation (Ma et al., 2009)
MBD2	chr 18: 51,677,971 - 51,751,158	DNA methylation; Eraser	dMT activity (Bhattacharya et al., 1999)
MBD4	chr 3: 129,149,787 - 129,159,022	DNA methylation; Eraser	Repairs cytosine deamination (Hendrich et al., 1999)
ELP3	chr 8: 27,950,584 - 28,048,669	DNA methylation; Eraser	Paternal DNA demethylation (Okada et al., 2010)
METTL3	chr 14: 21,966,282 - 21,979,457	RNA methylation; Writer	MT (Bokar et al., 1997)

METTL14	chr 4: 119,606,574 - 119,632,077	RNA methylation; Writer	MT (Liu et al., 2014)
WTAP	chr 6: 160,148,030 - 160,177,352	RNA methylation; Writer	Recruitment of METTL3/14 (Ping et al., 2014)
YTHDF2	chr 1: 29,063,133 - 29,096,287	RNA methylation; Reader	Methylation-mediated mRNA decay (Wang et al., 2014a)
YTHDF3	chr 8: 64,081,186 - 64,125,346	RNA methylation; Reader	Methylation-mediated mRNA decay (Shi et al., 2017)
HNRNPA2B1	chr 7: 26,229,556 - 26,240,413	RNA methylation; Reader	Methylation-mediated alternative splicing (Alarcon et al., 2015)
ELAVL1	chr 19: 8,023,457 - 8,070,529	RNA methylation; Reader	Methylation-mediated mRNA stability (Wang et al., 2014b)
ALKBH5	chr 17: 18,086,867 - 18,113,267	RNA methylation; Eraser	m^6A dMT (Zheng et al., 2013)
FTO	chr 16: 53,737,875 - 54,148,379	RNA methylation; Eraser	m^6A_m dMT (Mauer et al., 2017)

Chr = chromosome; MT = methyltransferase; dMT = demethylase; m^6A_m = N^6,2'-*O*-dimethyladenosine

2.3.2. Variant annotation

VCF files for each DNA and RNA methylation gene of interest from each of the ASD, ID, and control cohorts totalled 217 individual files. Initial quality control assessment of the VCF .tab files revealed an issue with the *MECP2* VCF files, in which the binary and allelic genotypes were incomplete in a large number of individuals (e.g. 0/ or A/). This effect was seen in all of the whole-exome sequencing datasets, but not the control whole-genome sequencing. As *MECP2* is the only gene of interest in our list that is found on the X chromosome, this would appear to be a ploidy problem associated with VCFtools processing of the X chromosome data. It is likely that VCFtools is the cause of the issue as no sequencing issues from the X chromosome from UK10K cohorts have been reported in the literature. This should be remedied with the VCFtools vcf-fix-ploidy PERL tool but attempts at using this tool were unsuccessful. Thus, the decision was made to continue with the remaining 30 genes and exclude *MECP2* from further analysis.

Next, every variant was annotated in each of the remaining 210 VCF files in order to provide initial characterisation of genomic variation within each gene and cohort. Principally, the focus was the specific class of variant (with particular focus on those that influence protein function such as missense variants), the specific amino acid and nucleotide change, and whether the variant has previously been identified (e.g. has an “rs” accession number). In addition, population MAFs were compared with those from our disease and control cohorts, as well as including functional prediction assessments that estimate the effect of a variant on protein function. This information can be attained manually from genome browsers such as Ensembl or pooled from databases such as the 1000 Genomes Project and the European Variant Server. In addition, although many functional annotation tools exist and prediction scores can be attained from their websites, this would be time-consuming for 210 VCF files containing tens of thousands of variants. Thus, the annotation of these VCF files were automated.

Class of variant, specific amino acid and nucleotide change, and whether the variant has previously been identified could be automated efficiently using SnpEff, a specialised variant annotation package that brings together information from multiple data sources. After downloading the SnpEff package along with the relevant human genome build (hg19/GRCh37), each variant could be annotated (Appendix 8).

In comparison, the automated annotation of MAF and functional prediction scores required more optimisation. Initially, an extra feature in SnpEff was used that allowed access to the dbNSFP (database for non-synonymous SNP's functional prediction) database. This includes MAFs from the large ExAC (Exome Aggregation Consortium) database of 60,706 individuals, alongside many notable functional prediction scores (SIFT, PolyPhen, MutationTaster, GERP++, etc.). However, this feature only successfully annotated around 5% of variant positions. Application of another package that brings together information from multiple data sources, ANNOVAR, improved this annotation to around one-third of variant positions, but most of these annotations lacked any functional prediction information.

Next, an attempt was made to automate the MAF annotation in-house by downloading the African (AFR), Asian (ASN), and European (EUR) 1000 Genomes MAF datasets from the 1000 Genomes FTP (file transfer protocol) site. Tabix was then used, a command-line tool for working with genomic files, to index these MAF datasets with specific genomic areas of interest, allowing one to match these MAF positions with the positions in our VCF files (Appendix 9). This resulted in an improved 50% MAF annotation of our variants. However, reproducing this protocol for each population MAF (AFR, ASN, EUR) would have been time-consuming.

Finally, I settled on using the Variant Effect Predictor (VEP) online tool for the GRCh37 human genome build within Ensembl. For each VCF file, the HGVS notation for each variant position (e.g. ALKBH5:c.16G>T) was deciphered from the SnpEff nucleotide change annotation. A list of notations for each VCF file could then be uploaded to VEP and the ExAC TOTAL, AFR, AMR (American), and NFE (Non-Finnish European) MAFs along with SIFT and PolyPhen functional prediction scores were returned. This provided annotation for around two-thirds of all variant positions. The annotations from VEP were also transcript-dependent, and so the canonical transcript from Ensembl was used when retaining the annotations. The ExAC database is one of the largest available, making their MAF scores preferable over the 1000 Genomes. Although VEP only returned SIFT and PolyPhen scores, these predictions perform accurately in comparison with other functional prediction tools (Dong et al., 2015). Moreover, they served only as a guide within the analysis and it is accepted that they cannot provide a definitive insight into protein function. SIFT prediction is based on the conservation of the affected variant amino acid change, whilst PolyPhen also takes into account structural information concerning the particular variant change.

As the main focus of our analysis is on rare or low-frequency variants, particularly those which are likely to impact on protein function, and owing to the large number of variants available to our study, our initial variant annotation procedure was limited to missense variants. Ideally, this would be extended to include all variants that disrupt protein function, termed loss-of-function (LoF) variants. However, this requires consensus about the class of variants that should be included as LoF. Whilst some have obvious influence on amino acid transcription, such as missense variants that alter codons or nonsense variants that result in premature stop codons, some variants are harder to interpret. For example, some splice variants have been shown to act as LoF variants (Pellanda et al., 2012). Using further annotation through the Loss-Of-Function Transcript Effect Estimator (LOFTEE) tool would allow one to assess LoF variants in future analyses. However, as this was not included in the present work, only variants that were rare or low-frequency and missense were considered interesting and retained for genetic imputation.

2.3.3. Merging datasets

To assess each gene across the ASD, ID, and control cohorts, VCF files for each DNA and RNA methylation gene for all cohorts were merged into one VCF, essentially condensing the remaining 210 files into 30. To do this, VCFtools was used and tabix. Tabix allows the user to index genome positions from files such as VCF files, and these resultant tabix index files (.tbi) are necessary for some processes where chromosomal regions are needed but not specified in the program commands (Appendix 10). This resulted in 30 VCF files, one for each DNA and RNA methylation gene, which included data from all of the disease and control cohorts.

2.3.4. Genetic imputation

As previously mentioned, one of the goals of this analysis was to design a NGS pipeline which could reliably incorporate sequencing datasets generated from multiple studies using different NGS platforms. The challenges associated with this are amplified in the case of rare and low-frequency variants as they are less likely to be included in standard array-based sequencing technologies. Although the most recent technologies are increasingly successful in reading most variant positions (Consortium et al., 2015, LaDuca et al., 2017), rare variants continue to be the most difficult to detect with a reliable depth of coverage.

In the case of the UK10K sequencing datasets, this issue means that variants of any frequency (common, low-frequency, rare) that are shared across all of the datasets within our merged VCF files can be included in the analysis. However, there may be cases where a particular variant is called in one dataset but is not called in another, leaving areas of missing calls (e.g. "./."). If this variant is particularly interesting – for example, a missense variant which is rare in one of the disease cohorts – then it would be useful to be able to compare this variant with the cohorts where it has not been called. To do this, genetic imputation can be used to estimate variant calls based on haplotypes (shared inherited genetic markers across groups of genes) from reference panels of genetic information from thousands of individuals, allowing for accurate replacement of the missing calls with allelic genotypes.

Owing to the popularity of genome-wide association studies (GWAS) and global efforts to sequence ever-larger cohorts of individuals, multiple genetic imputation tools have been developed for general academic use. Amongst the most common are Beagle, IMPUTE2, and SHAPEIT, although new cloud-based methods such as minimac3 appear to perform as well as these tools with greater efficiency and less computational load (Das et al., 2016). IMPUTE2 was used because of its use in studies of rare and low-frequency variants and its continued position as one of the leading imputation tools in the field (Kwan et al., 2016, Mitt et al., 2017). IMPUTE2 does not use population labels (cultural or geographical groups with increased genetic homogeneity) to assess the reference panels or the individuals being imputed, meaning that IMPUTE2 is not restricted to shared ancestry haplotypes. This onus on sequence identity allows IMPUTE2 to account for recent ancestral admixture and to discount divergent haplotypes.

In theory, any large-scale sequencing datasets could be used as imputation panels, providing that the correct map information (recombination rates between positions), haplotype information (record of known haplotypes to guide IMPUTE2), and legend information (details of variants in your data) are included. IMPUTE2 provides this data for the International HapMap Project and the 1000 Genomes Project. The specifics of the imputation panel, such as the number of individuals and the depth of coverage, can influence the performance of genetic imputation. Nevertheless, imputation of rare variants is still poor using established imputation panels (Zheng et al., 2015). The most up-to-date imputation panel available within IMPUTE2 is the

1000 Genomes Project Phase 3 release (October 2014) and so this was chosen for our imputation protocol. This panel contains 81,706,022 sites detailed in Table 2.10.

Our imputation protocol involves the orchestration of multiple tools in addition to IMPUTE2, including VCFtools, PLINK, GTOOL, and BCFtools. PLINK is an open-source toolkit for whole-genome sequencing analysis which can take a dataset from initial data processing through to genetic association analyses. PLINK was required to convert our VCF files into Oxford-formatted GEN and SAMPLE files for IMPUTE2 and then back into VCF files via BED file transformation for downstream analysis. PLINK version 1.9 or later is required for this latter conversion. GTOOL is a purpose-built program that can transform genotype data outputted from tools such as IMPUTE2. Finally, BCFtools is a package of executables similar to VCFtools that allows for manipulation of VCF files. Our rationale for using imputation is to remedy missing calls in specific cohorts, not to retain any variants added into our regions of interest during the haplotype processing in IMPUTE2. Thus, BCFtools was used to remove any additional variants that arise during imputation, restoring our VCF files to include only the positions present before imputation.

First PLINK was used to convert the merged VCF files for each gene into GEN and SAMPLE files needed for IMPUTE2 (Appendix 11). The map, haplotype, and legend information from the relevant chromosome of the 1000 Genomes Project Phase 3 imputation panel were downloaded and unzipped. The reason for doing this on a merged gene-by-gene basis rather than for all of the genes at the same time is that the process is computationally intensive. Therefore, breaking this process down into a gene-by-gene analysis allows one to split the process into chromosomes and subsequently improve performance.

TABLE 2.10. Positions within the 1000 Genomes Phase 3 imputation panel

Variant Type	Number of sites
Biallelic_SNP	77,818,332
Multiallelic_SNP	520,275
Biallelic_INDEL	2,982,597
Multiallelic_INDEL	324,022
Biallelic_DEL	32,306
Biallelic_DUP	5,791
Biallelic_INV	100
Biallelic_MNP	1
Multiallelic_CNV	6,210
Biallelic_INS:ME:ALU	12,491
Biallelic_INS:ME:LINE1	2,910
Biallelic_INS:ME:SVA	822
Biallelic_INS:MT	165

SNP = single nucleotide polymorphism; INDEL = insertion/deletion; DEL = deletion; DUP = duplication; INV = inversion; MNP = multiple nucleotide polymorphism; CNV = copy number variant; INS = insertion; ALU = arthrobacter luteus elements; LINE = long interspersed elements; SVA = SINE/VNTR/ALU elements

At this point, it was observed that continuing with the imputation protocol resulted in the flipping of certain reference alleles (e.g. 2, 1 as opposed to 1, 2). As a result, some MAFs were inaccurately inflated as the minor allele was being read as the reference allele by downstream tools. This occurred because some of the variants present in our merged VCF file were not found in the legend file which details each variant expected to be found in the dataset. To rectify this error, the 1000 Genomes Project Phase 3 legend file was duplicated and any variants that were present in our merged VCF file but missing in the legend file were inserted. This amended legend file was then used during imputation. Next, IMPUTE2 was executed to impute positions from the reference panel, before GTOOL and VCFtools were used to convert the output into VCF format (Appendix 12).

As previously mentioned, the imputed VCF files for our merged genes contain all of the original positions from the UK10K sequencing along with many thousands of extra variants which are added as part of the imputation process. These were removed using BCFtools. In addition, our analysis only focused on shared variants in all cohorts of any frequency as well as rare or low-frequency variants of interest (e.g. missense variants). Subsequently, any variants that were not interesting for our analysis were also removed and that were not shared between all cohorts using BCFtools.

The advantage of removing these common, less interesting (i.e. intronic) variants is that these appear to be the most susceptible to erroneous flipped genotypes, persisting even after the amendments to the legend file. All imputed merged VCF files were quality checked after imputation and any variants presenting with erroneous flipped genotypes were removed using BCFtools (Appendix 13). This resulted in 30 VCF files, one for each DNA and RNA methylation gene including all disease and control cohorts, which contain only shared positions or variants of interest. These files were assessed again for the quality of variant calls. Any variant positions containing erroneous genotype calls were removed using the commands above in BCFtools.

2.4. Statistical analysis

2.4.1. OPTIMA and VITACOG

Upon receipt of the VITACOG phenotypic data, a power calculation was performed to determine the minimum sample size required to avoid Type I (false positive) and Type II (false negative) error rates. With a confidence level of 95% (p = 0.05), a conservative estimate of the proportion of our population that have measurable attributes (50%), and confidence intervals of 0.5, our population size of 271 within VITACOG required a sample size response of 160. The sample size of each treatment group is slightly lower at 132 and 133 respectively, but these are still positive group sizes for a case-control method of analysis.

A number of differing statistical approaches were undertaken for the assessment of genotypes within each OPTIMA strand and for the in-depth analysis of genotype and cognitive performance within the VITACOG cohort. For the genotype analysis, each particular genotype was counted within each OPTIMA strand. In the case of *DNMT3L* R278G, this meant a count of individuals who had A/A, A/G, or G/G genotypes. These were then collapsed so that A/A homozygotes could be compared with individuals who possessed the minor allele (G+ or G carriers). In the case of *APOE*, this meant a count of individuals who had E2E2, E2E3, E2E4, E3E3, E3E4, or E4E4 genotypes. These were then collapsed so that those with a copy of the E4 allele (E4+ or E4 carriers) could be compared with those who did not (no E4). Minor allele frequencies (MAF) were also calculated for both *DNMT3L* R278G and *APOE* using a standard method:

- ((homozygotes * 2) + heterozygotes) / total

The genotype counts were then grouped for statistical analysis based on the disease of the individual; namely, AD, MCI, or control. In addition, the sex of the individuals was also included to assess any particular sex-specific relationships between genotype and the presence of disease. Contingency tables were created from the grouped genotype counts and chi square (χ^2) tests were used for the *DNMT3L* R278G and *APOE* genotype analyses in the combined OPTIMA strands. Chi square testing also produced accompanying odds ratio (OR) statistics, which were used to estimate the likelihood of individuals being within specific groups (i.e. those with particular genotypes being in a particular disease group).

Hardy-Weinberg equilibrium was also assessed for the OPTIMA cohorts. Hardy-Weinberg equilibrium describes the state of genetic variation within a given population. When this is not influenced by genetic or environmental factors and variation remains unchanged, there is said to be equilibrium. Natural populations very rarely find themselves in equilibrium as factors such as mutations, mating patterns and genetic drift provide evolutionary impetus. Nevertheless, assessing Hardy-Weinberg equilibrium can be of use in case-control studies to highlight potential biases in the data (Grover et al., 2010), although one must be cautious about the methods applied and the rationale for assessment (Ziegler et al., 2011, Zou and Donner, 2006).

To assess Hardy-Weinberg equilibrium, the R package HardyWeinberg (Graffelman and Camarena, 2008) was used. This package allows one to load in frequency data for a given genotype and apply a chi square test to ascertain whether the given genotype does or does not conform to Hardy-Weinberg equilibrium.

Synergy factor analysis was also applied to select combinations of genotypes for *DNMT3L* R278G and *APOE*. This method has been developed to provide a quick estimation of synergistic or antagonistic interactions between risk or protective genotypes in a case-control group comparison (Cortina-Borja et al., 2009). To calculate the synergy factor, an extended cross tabulation was produced for the two genotypes of interest in the case and control cohorts, along with OR scores for each combination (Table 2.11).

In Table 2.11, - and + refer to the absence or presence of that particular genotype. *n* refers to the number of people in the cases and controls who have the respective absence/presence of each genotype. Reference is the case-control data from which the rest of the *OR* scores can be calculated against. The calculations were performed in Microsoft Excel to transform these values into a SF score, a Z score, and a *p* value (Appendix 14).

TABLE 2.11. Data required for synergy factor analysis calculations

Genotype 1	Genotype 2	Controls	Cases	OR
-	-	*n1*	*n2*	Reference
+	-	*n3*	*n4*	*OR1*
-	+	*n5*	*n6*	*OR2*
+	+	*n7*	*n8*	*OR3*

- and + refer to the respective absence and presence of a particular genotype, *n* refers to the number of cases and controls who have the respective absence or presence of a particular genotype

For the in-depth analysis of genotype and cognitive performance within the VITACOG cohort, each cognitive test was first assessed individually for a relationship with either the *DNMT3L* R278G or *APOE* genotypes. The collapsed genotypes were utilised because they allowed for simpler comparison of two groups (e.g. A/A v G+) and because the size of the groups were more even, controlling for spurious results caused by the low number of minor allele homozygotes. Normal distribution was inspected for all variables used in this analysis using the Shapiro-Wilk and Kolmogorov-Smirnov tests. Any variables with a *p* value below 0.05 were considered not normally distributed and logarithmic transformation (log10) was applied to these variables. In addition to genotype, gender and age were included into our assessment of cognitive performance. Due to the importance of Hcy levels identified in previous work using the VITACOG cohort, the relationship between Hcy levels and our genotypes of interest was investigated, as well as the modulating influence of Hcy on the relationship between our genotypes of interest and cognitive performance. Baseline Hcy levels taken at the beginning of the VITACOG study were also examined. These baseline levels were grouped for some analyses, splitting the variable into high and low levels based upon the median baseline Hcy value. High and low Hcy was then split again to create middle high and middle low Hcy levels. This approach was used as it had been applied in previous publications using the VITACOG cohort (Smith et al., 2010).

As the individuals within VITACOG were assessed both at the beginning and the end of the 24 month treatment period, there was the option of analysing the values at each time point as well as the change between the time points. Thus, the change (Δ) in cognitive performance was derived for each cognitive test and for Hcy levels. Where the values for each time point were used, a repeated measures General Linear Model (GLM) was applied to assess the relationship between B vitamin treatment and cognitive performance. However, the majority of analyses were carried out using Δ in cognition. For this, univariate GLMs were used to analyse the relationship between genotype, B vitamin treatment group, and any relevant covariates in the VITACOG cohort, along with *post hoc* testing for multiple corrections where necessary. The student's *t*-test was applied for simple comparisons of means between groups and linear regression was used to assess the relationship between continuous variables. Statistical analysis was principally carried out using SPSS Statistics version 22.0 (IBM Corp).

Following the initial investigation into each individual cognitive test, a statistical approach akin to factor analysis was applied to evaluate any clustering of variance within our select group of cognitive tests. This approach, named Principal Component Analysis (PCA), is a transformative procedure used to identify the major sources of variance within a select number of variables. This approach was used on change (∆) in HVLT-R, MMSE, GNT, CF, and PAL. Varimax rotation, a common orthogonal method in PCA that maximises correlations between variables and components, was applied and components with an eigenvalue > 1 were retained for further analysis. Again, univariate GLMs were used to analyse the relationship between genotype, B vitamin treatment group, Hcy levels, and any relevant covariates on performance on these PCA cognitive factors.

2.4.2. VITACOG atrophy data

In line with the previous work carried out on this cohort (Smith et al., 2010), the raw ROA value was converted from a negative percentage change for the 24 months of the study to a yearly positive percentage change that was used for all of the following ROA analyses. Also in line with previous work, normalised total brain volume, measured against the shape of the skull, was converted from mm^3 to mL.

Pearson's correlation coefficients (R) were used to analyse the strength of baseline covariates with ROA. Student's *t* tests were used to analyse the baseline covariates in more detail. To assess the relationship between ROA, these key covariates, and our genotypes of interest with respect to cognitive performance, univariate GLMs were utilised in order to model the influence of covariates on ROA. This methodology was chosen over a standard multiple regression model because of its flexibility; it allowed for the application of continuous and categorical independent variables, it allowed for the addition of multiple covariates alongside independent variables, and it accounted for invariance in the independent variables which may inhibit a traditional regression model. Covariates found to be significantly associated with ROA were included in the models. The results of these models were saved as unstandardized predicted values, which provide new derived variables based upon the regression equation from each model. Final predicted models and their respective covariate results are presented in Appendix 15. The relationship between cognitive performance and these predicted ROA values was then compared with the original ROA values to account for the effect of covariates. Standard linear regressions were used to produce these comparisons.

Normal distribution was inspected for the additional covariate and brain imaging variables used in this analysis using the Shapiro-Wilk and Kolmogorov-Smirnov tests. Any variables with a *p* value below 0.05 were considered not normally distributed and logarithmic transformation (log10) was applied to these variables. Statistical analysis was principally carried out using SPSS Statistics version 22.0 (IBM Corp).

2.4.3. NCDS 1958 and TwinsUK

To begin with, the R package HardyWeinberg was used to assess Hardy-Weinberg equilibrium (Graffelman and Camarena, 2008). This package allows one to load in frequency data for a given genotype and apply a chi square test to ascertain whether the given genotype does or does not conform to Hardy-Weinberg equilibrium.

In order to accurately repeat the VITACOG analyses, univariate General Linear Models (GLMs) were again used for the analysis of biochemical, demographic and cognitive measures in the NCDS 1958 and TwinsUK cohorts, along with *post hoc* testing for multiple corrections where necessary. The student's *t*-test was applied for simple comparisons of means between groups. Statistical analysis was principally carried out using SPSS Statistics version 22.0 (IBM Corp).

Due to the implementation of PCA in the VITACOG analyses, this statistical approach was applied to a selection of variables from the cognitive tests included in NCDS 1958 and TwinsUK. Unlike VITACOG, the cognitive data from the control cohorts was only taken at one time point during the studies. From NCDS 1958, WR, DWR, AN, Ps & Ws Scanned, and Ps & Ws Missed were incorporated. From TwinsUK, DMS Total Correct, DMS Mean Latency, PRM Total Correct, PRM Mean Latency, PAL Total Errors, SWM Within Errors, SSP Length, 5 Choice Reaction Time, and IED Total Errors were incorporated. One test available within TwinsUK, the National Adult Reading Test (NART), was excluded from inclusion in the PCA as only a derived verbal IQ score was provided and it was not obvious how this variable had been derived from the original test.

More variables derived from the TwinsUK cognitive tests were available but were not included in the PCA. This was because the variables were extrapolated from the same tests and thus their inclusion in the PCA could result in a non-positive definite correlation matrix, caused by linear dependencies between these extrapolated

variables. This is in line with previous studies looking at genetic influences on cognitive performance using the CANTAB battery (Singer et al., 2006). A similar issue was encountered when applying PCA to the NCDS 1958 cognitive variables. Specifically, the Letter Cancellation Ps & Ws Missed variable was not negatively loaded into the PCA, which should be expected for negative variables (i.e. where higher scores indicate poorer performance). To combat this issue, a new Letter Cancellation variable was derived by subtracting the "missed" score from the "scanned" score to create a Ps & Ws Identified variable. This positive variable loaded properly into the PCA analysis.

In line with our previous analyses, normal distribution was inspected for all variables used in this analysis using the Shapiro-Wilk and Kolmogorov-Smirnov tests. Any variables with a *p* value below 0.05 were considered not normally distributed and logarithmic transformation (log10) was applied to these variables. The same PCA conditions from the VITACOG analyses were retained for replication in the controls. Varimax rotation was applied and components with an eigenvalue > 1 were retained for further analysis.

2.4.4. Next-generation sequencing cohorts

For simple case-control associations including all of the DNA and RNA methylation genes, the Fisher's exact test with Holm-Bonferroni multiple comparisons correction was applied in line with most modern genome-wide association study (GWAS) analysis procedures (Bush and Moore, 2012). First, each VCF was opened in Microsoft Excel and a COUNTIF function was used to count the number of wildtype (0/0) and minor allele carrier (0/1, 1/1) individuals for each variant. These were then combined to create a file with case wildtype, case minor allele, control wildtype, and control minor allele counts for each gene, and then combined to create a file for all of the genes summated together.

Fisher's exact test is an in-built statistical test within the R statistical programming language. R allows one to iterate the Fisher's test over all of the case-control variant positions, making the processing computationally efficient compared to SPSS or other graphical user interface statistical programs. The combined case-control allele count files created above were read into R to run the Fisher's exact test (Appendix 16).

In order to create Manhattan plots for the Fisher's exact test results, the external package "qqman" was used. The Fisher's test output data was first reorganised into a file that contains a unique SNP (single nucleotide polymorphism) identifier (e.g. rs1, rs2, etc), the chromosome, the chromosomal position of the variant, and the multiple comparisons corrected p value. This data was then loaded into R (Appendix 17). In line with most GWAS analysis protocols, 5×10^{-8} was taken as the threshold of statistical significance. A suggestive value of 1×10^{-5} was included for variants that could be considered statistical trends.

Upon initial investigation of the Fisher's exact test results, some extremely significant results were observed in the combined ASD and ID cases when compared to the TwinsUK controls. By reassessing the VCF files and the levels of genetic imputation, it was discovered that these inflated outlying *p* values were the result of high levels of wildtype imputation. Genetic imputation is used to rectify missing calls within sequencing data by estimating these calls using haplotype data from large reference panels including thousands of individuals. In situations where a variant has a common MAF but the sequencing quality is poor in one particular dataset, an inflated number of wildtype calls may be imputed in this dataset. This would lead to a skewed comparison with the other datasets that have the expected common MAF for the minor allele. Moreover, if the reference and alternative alleles were flipped as discussed previously, the same skewed effect would be found but with the minor allele carriers being inflated in that particular dataset.

To resolve this issue, a post-imputation quality threshold was employed. Similar techniques have been used with other genetic imputation tools such as MaCH (Markov Chain framework for genotype imputation), where a cut-off of 0.3 for imputed SNPs is applied (Pei et al., 2010). Owing to the relatively low sample size numbers in our case cohorts (<1000), a similar initial threshold of 0.25 was chosen (i.e. if 25% or more of the genotype calls in a cohort are imputed, that variant is removed from the analysis) followed by a more stringent threshold of 0.125.

2.4.4.1. Burden analysis

One of the driving reasons behind conducting large scale sequencing project such as the UK10K Project is to elucidate the influence of rare variants on human traits and diseases. The advent of cheaper sequencing technology has meant that many common variants have been comprehensively analysed in many diseases within many different phenotypic populations. However, as large sequencing studies have shown, the detection of rare variants is fraught with difficulty. Increasing sample sizes are sought in order to identify risk-associated rare variants with the requisite statistical power and any resultant rare variant findings are only likely to account for a small proportion of disease risk.

Nevertheless, rare variants can inform one about the genes or genetic functions that are involved in specific traits or diseases. This can help develop our understanding of a particular phenotype and how it arises. One popular method that has been established recently in the field of rare variant analysis is burden testing. This involves using computational strategies to aggregate rare variants within a particular genetic region of interest before assessing the effect of these combined variants on a particular phenotype, allowing researchers to detect statistically significant disease risk from rare variant data.

These computational strategies have improved over a short period of time, with a recent review by one of the leaders in the field drawing up five distinct classes of burden test (Lee et al., 2014). The earliest were simple burden tests that collapse rare variants into a single score used for case-control analysis. Adaptive burden tests were then incorporated to address some of the limitations of simple burden tests, including permutated p values and introducing the ability to estimate the direction of a variant effect (i.e. associated with risk or protection). Variance-component tests, which include some of the most popular tools used in rare variant analysis, also assess the distribution of variants rather than simply using the aggregated score. Combined variance-component and burden tests attempt to utilise the strength of variant-component tests in assessing non-causal and multi-directional variants with the strength of burden tests in assessing regions with many unidirectional causal variants. Finally, the statistical framework for a more powerful exponential component test has been designed but has not yet been implemented in a commercial or open-source tool.

An attempt was made to implement a range of these burden tests within our analysis in order to compare and contrast the results of each approach. This began with RVTESTS, a newly designed package capable of performing many rare variant analysis tests with quantitative and binary traits (Zhan et al., 2016). Moreover, RVTESTS uses VCF files as an input making it particularly useful for our NGS pipeline output. However, an initial test of rare variant analysis using one of our imputed VCF files resulted in inaccurate estimations of the MAF of our variants. As our VCF files have been through multiple tools and processing steps within our NGS pipeline, it is possible that the VCF file was being incorrectly read by the RVTESTS package. A similar issue arose when attempting to use PLINK/SEQ, a library of genetic analysis tools designed to be used in parallel with the aforementioned PLINK sequence processing tool. Some of these issues were rectified by using UNIX to edit the VCF files (Appendix 18), but ultimately the rare variant analysis steps were unsuccessful.

It then progressed to using individual packages and scripts within the R statistical programming language to run our burden tests. The R package AssotesteR was used to run the Combined Multivariate and Collapsing (CMC) simple burden test and the Variable Threshold (VT) adaptive burden test. AssotesteR has been used to run these tests in recent publications analysing whole-exome and NGS sequencing data (Hirota et al., 2016, Lescai et al., 2017). To prepare the data for these tests, a recoded genotype file was created from the imputed VCF files for each DNA and RNA methylation gene. First, the binary genotypes were transposed in Excel so that each variant was listed in the first row and all individuals listed in the first column. This data could then be imported into R for recoding (Appendix 19).

These steps take the transposed data, transform the data into a character variable to be edited, changes the wildtype homozygotes (“0/0”), minor allele heterozygotes (“0/1”) and minor allele homozygotes (“01-Jan” – Excel reads “1/1” and autocompletes it as a date) to 0, 1, and 2 respectively, and writes the data to a new file. A phenotype file was then created which annotates whether any given individual was in the control (“0”) or case (“1”) cohorts. Individuals could also be subsetted from the genotype data for analysis looking exclusively at the ASD or ID cohorts alongside the controls. With the genotype and phenotype data in place, the burden and adaptive burden tests in R were run (Appendix 20).

Another adaptive burden test was also trialled, the kernel-based adaptive cluster (KBAC) test, to compare with and validate the results of the CMC and VT tests. An available R script was used that is designed to demonstrate the KBAC methodology (Wang, 2017). Recoded genotype data files were merged with the phenotype data file to create one input file for the KBAC test, where the first column of the genotype file now contains the case-control information (Appendix 21).

Finally, the sequence kernel association test (SKAT) package was used to perform variance-component and combined variance-component and burden analysis. SKAT remains one of the most popular packages for rare variant analysis because of its relative power compared to other statistical tests and its optimised approach that combines variance-component and burden tests into one package (Lee et al., 2012). Assuming that the recoding and combining of case-control genotype data has been performed using the steps described above, the genotype and phenotype files from the CMC and VT tests can also be used for the SKAT tests (Appendix 22).

These analyses were completed for rare variants (and common variants where applicable) for all case cohorts versus controls, ASD cohorts versus controls, and ID cohorts versus controls. In line with our Fisher's exact test statistics, 5×10^{-8} was taken as the threshold of statistical significance along with a suggestive significance of 1×10^{-5} to be considered as statistical trends. In addition, a genomic imputation threshold of 0.25 was applied followed by a more stringent threshold of 0.125.

2.5. *In silico* protein modelling

In addition to the *DNMT3L* R278G variant, a number of other methylation-related genetic variants were assessed as both positive and negative controls, with the positive controls being variants that should influence protein function and the negative controls predicted to be benign changes. For *DNMT3L*, a positive control variant in the shape of R271Q was included, a nearby variant to R278G that has been reported to result in significant hypomethylation (El-Maarri et al., 2009). The negative control variant in *DNMT3L* was H313Y, a variant with no known clinical relevance. Two positive control variants were also included from the other principal DNA methyltransferases. *DNMT1* Y495C is a clinically associated variant which causes Hereditary, Sensory, and Autonomic Neuropathy 1 (HSAN1) (Klein et al., 2011), while *DNMT3A* R749C causes a rare syndromic form of intellectual disability (Tatton-Brown et al., 2014). A summary of these variants is presented in Table 2.12.

The *in silico* protein modelling performed in the present study was performed using two methods which assessed protein structure and protein surface respectively. Protein Data Bank (PDB) files for each of the proteins of interest were either accessed from the Protein Data Bank (Rose et al., 2017) which houses PDB files used in previous publications, or created by downloading the canonical amino acid sequence in FASTA format from Uniprot (Bateman et al., 2017) and submitting this to the RaptorX Structure Prediction tool (Kallberg et al., 2012). These PDB files can then be evaluated using the following modelling tools and compared accordingly. A complete list of the PDB files used in the present study is presented in Table 2.13.

PyMOL (Schrödinger) is a widely used visualisation system that allows one to produce, manipulate, and analyse molecular structures, most commonly protein structure models. Aside from its ability to produce high resolution structural images for publications, PyMOL can provide mechanical insight into protein behaviour both through native programming and through a wide range of plug-ins. Because of this, PyMOL was utilised for the protein structure and protein interaction modelling, as well as being the primary visualisation tool for any images presented. Regarding protein structure analysis, PyMOL was used to simulate mutations in the canonical amino acid sequence using the Mutagenesis wizard. Briefly, once the PDB file had been visually prepared, the amino acid position of interest was selected and the variant change was chosen. Default hydrogen, backbone, and rotamer options were retained for all analyses. This allowed for comparison of secondary structure changes such as hydrogen bond formation/disruption. For the protein surface analysis, WT and variant PDB files were prepared. The area around the residue of interest was expanded to 6 angstroms (Å) and visualised as a mesh surface representation. The surface area was then calculated for both the 6Å region and the total protein to assess surface changes in key interaction regions. Images were exported using the ‘ray’ command, normally at 2000 x 2000 resolution.

TABLE 2.12. Methylation variants included in the *in silico* analyses

Gene	Variant	Known	Clinical relevance	Publication
	R278G	rs7354779	Yes	Haggarty et al. (2010)
DNMT3L	R271Q	rs113593938	Yes	El Maarri et al. (2009)
	H313Y	-	No	-
DNMT3A	R749C	-	Yes	Tatton-Brown et al. (2014)
DNMT1	Y495C	-	Yes	Klein et al. (2011)

TABLE 2.13. PDB models used in the *in silico* analyses

Protein	PDB name	Source
DNMT3L	-	RaptorX
DNMT3A	-	RaptorX
DNMT3A-3L	4U7P	PDB
DNMT3A-3L C terminus	2QRV	PDB
DNMT3A-3L-H3	4U7T	PDB
DNMT1	-	RaptorX
DNMT1 (residues 351-1600)	4WXX	PDB
DNMT1 (replication targeting sequence)	3EPZ	PDB

In addition, two further modelling tools were used. FoldX (Schymkowitz et al., 2005) is a command-line algorithmic program that calculates the free energy of a molecular structure through the contribution of a number of key molecular interactions within a largely empirically driven force field. This free energy (ΔG) value provides a measure of protein stability, as a lower ΔG value indicates a more stable protein. The molecular interactions assessed by FoldX include Van der Waal forces, hydrogen bonds, solvation energies, water bridges, electrostatic energy of charged residues, and the entropic energy for fixing both the main backbone of the structure as well as any necessary side chains. The Van der Waal, hydrogen bond, solvation energy, and water bridge calculations are all determined by experimental data, whilst the electrostatic and entropic calculations are theoretically based (Mendes et al., 2002). FoldX provides a variety of physico-chemical properties that can be altered by the user. In our case, all of these properties were left at default values or were designated by the PDB file itself (such as in the case of water and metal bond properties). Table 2.14 details the physico-chemical properties that were standardised across all of our analyses.

Two separate processes were implemented using FoldX in the present study. First, the PDB files chosen for ΔG analysis were repaired. This step screens for and rectifies any protein structure issues intrinsic to the PDB files, such as steric clashes, reducing the total ΔG of the protein to the lowest possible state. This ensures that all subsequent calculations applied in FoldX using these PDB files are computed from a standardised position. Once the repaired files have been generated, the variants of interest can be built into the model. The number of iterations by which the variant was modelled was kept at 3 for all analyses. An average of the WT and variant model ΔG values can then be taken to calculate the change in ΔG ($\Delta\Delta G$). ΔG calculations from FoldX have a reporting accuracy (or SD) of 0.46kcal/mol, which has led publication of FoldX modelling (Rallapalli et al., 2014, Studer et al., 2014a) to bin $\Delta\Delta G$ values into distinct categories (Table 2.15) that will be applied in the present study.

TABLE 2.14. Properties used within the FoldX analysis of ΔG

Properties	Values
Temperature	298
pH	7
Ion Strength	0.050
Water	Crystal
Metal	Crystal
Van der Wall design	2
No. of runs	3

TABLE 2.15. Categories of ΔΔG estimations in FoldX applied in our analysis

Category	Estimated ΔΔG (kcal/mol)
Highly stabilising	Less than -1.84
Stabilising	Between -1.84 and -0.92
Slightly stabilising	Between -0.92 and -0.46
Neutral	Between -0.46 and 0.46
Slightly destabilising	Between 0.46 and 0.92
Destabilising	Between 0.92 and 1.84
Highly destabilising	Greater than 1.84

ΔΔG = change in free energy

Adaptive Poisson-Boltzmann Solver (APBS) is a command-line program that can be applied to solve the Poisson-Boltzmann equation (PBE). The PBE can be used to predict the electrostatic potential of solutes in solution, and is particularly useful in modelling solutes with complex shapes and molecular surfaces such as proteins. Electrostatic surface potential plays an important role in protein-protein and protein-ligand interaction, protein movement, and alterations in conformation amongst other behaviours (Sinha and Smith-Gill, 2002). The distribution of electrostatic potential is described through the dielectric constant(s) used to characterise the solution and the local electron charge, whilst the dynamics of the solution are described through the concentration of ions, the temperature, and the Boltzmann constant (quantity linking energy and temperature at the particle level) to give the PBE (Figure 2.2).

Before applying APBS to our proteins of interest, the PDB files must first be converted to PQR files (PBD-like files with occupancy and temperature features replaced with charge and radius measurements) using the PDB2PQR program, available as a command-line tool or as a web server (Dolinsky et al., 2004). This intermediary program resolves the electrostatic parameters that can otherwise be missing from standard PDB files, such as missing hydrogen atoms and atomic coordinates. A pipeline of the PDB2PQR workflow can be seen in Figure 2.3. Default settings were used for all PDB conversion, with pK_a pH set to 7 and PARSE as the parameter force field. The atomic parameters built into PARSE (parameters for solvation energy) have been validated against experimental data (Sitkoff et al., 1994) and account for the influence of implicit solvation, whereby the solution is treated as a single medium as opposed to individual molecules (Yang et al., 2009). APBS can then be applied to the output files from PDB2PQR, namely the PQR and IN files, to produce a DX file. This contains the electrostatic grid information predicted by the PBE. The DX file, along with the PQR file, can then be accessed by the APBS Tools 2.0 PyMOL plugin for visualisation and further inspection.

$$\nabla \cdot [\epsilon(r) \nabla \cdot \Phi(r)] - \epsilon(r) \kappa(r)^2 sinh[\Phi(r)] + 4\Pi p^f(r)/kT = 0$$

FIGURE 2.2. The Poisson-Boltzmann equation (PBE). This equation was adapted from Honig and Nicholls (1995), where *r* is a position vector, φ(*r*) is the electrostatic potential measured in *kT/q* (*k* is the Boltzmann constant, *T* is the absolute temperature, *q* is the proton charge), ε is the dielectric constant, ρ^f is the proton charge density. ε, φ, κ, and ρ are all functions of *r*.

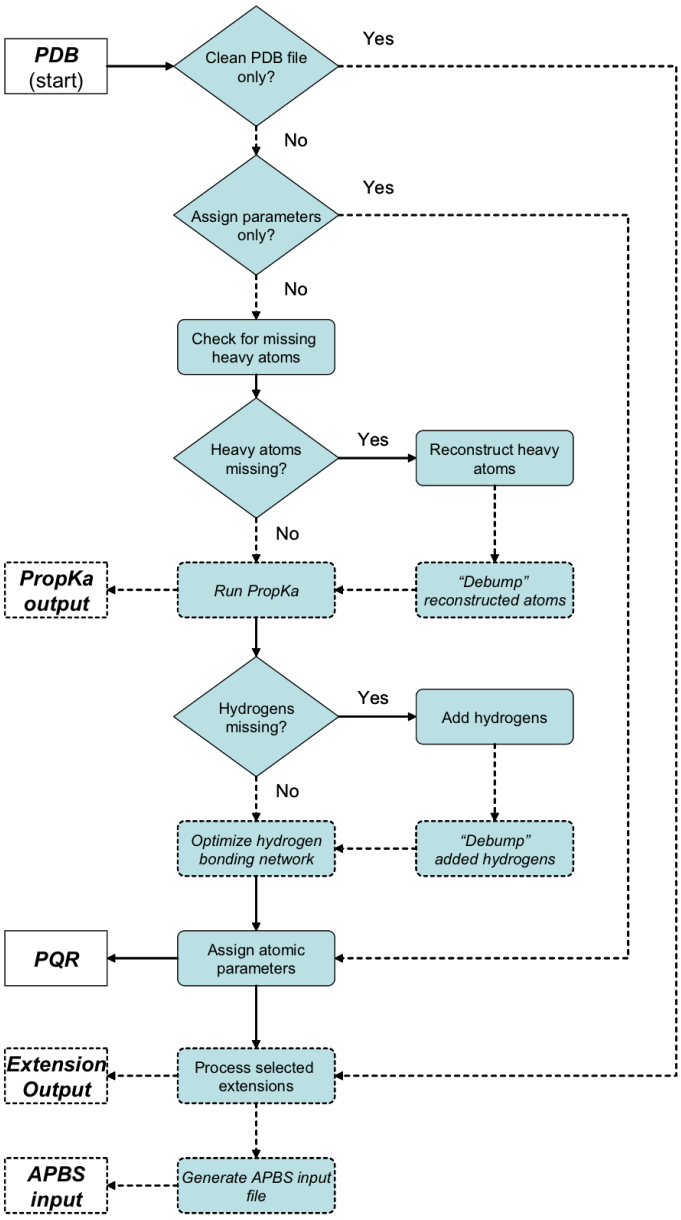


FIGURE 2.3. Flowchart of the PDB2PQR program. Figure taken from Dolinksy et al. (2007).

2.6. Immunohistochemistry

Two similar but distinct immunohistochemistry protocols were implemented in order to visualise immunoreactivity of specific proteins and DNA modifications. The standard protocol was used to assess neuronal, glial, or stem cell markers. The adapted protocol was used to examine the methylation modifications 5hmC and 5caC. Information on materials used in both protocols is presented in Appendix 23.

2.6.1. Standard protocol

Mounted brain sections were first dewaxed by removing the paraffin fixative in order to allow aqueous solutions to enter, namely allowing our antibodies to interact with their specific antigens. Sections were placed in staining racks and then into glass staining dishes. Sections were dewaxed using two 3 minute washes in xylene in a fume hood, followed by two 3 minute washes in 100% ethanol, two 3 minute washes in 70% ethanol, and one 5 minute wash in running water. The sections were transferred from the staining dish into a glass coplin jar and given one further 5 minute wash in phosphate-buffered saline (PBS) at room temperature.

The sections were removed from the coplin jar and placed into a wet chamber to ensure that the tissue or any applied solution does not dry out during incubation periods. Once placed in the wet chamber, the sections of tissue mounted on each slide were surrounded with a hydrophobic barrier using a hydrophobic PAP pen. This guarantees that the applied solutions cover the tissue as well as allowing one to use a smaller amount of solution, as opposed to covering the whole slide with solution. Blocking solution was composed of 1ml of donkey serum and 4µl of Triton X-100. 200µl of the blocking solution was applied to each section and left in the enclosed wet chamber for one hour at room temperature.

After one hour, the blocking solution was removed. The hydrophobic barrier was re-assessed to ensure that it had not been compromised during the removal of the blocking solution. Primary antibody solution was made by adding the requisite amount of primary antibody to 1ml of blocking solution. Table 2.16 presents the primary antibodies used and their concentrations. 200µl of primary antibody solution was applied to each section and left in the enclosed wet chamber overnight at room temperature.

The next day, the primary antibody solution was removed. The sections were given five 5 minute washes in PBS by pipetting 200µl of PBS onto the sections, within each of the hydrophobic barriers. Secondary antibody solution was made by adding the requisite amount of secondary antibody to 1ml of blocking solution. Table 2.16 presents the secondary antibodies used and their concentrations. 200µl of secondary antibody solution was applied to each section, within each of the hydrophobic barriers, and left in the enclosed wet chamber for two hours at room temperature.

After two hours, the secondary antibody solution was removed. The sections were given three 5 minute washes in PBS at room temperature by pipetting 200µl of PBS onto the sections, within each of the hydrophobic barriers. The area of the slides around the hydrophobic barriers was dried carefully prior to mounting. 22mm x 64mm No. 1 (0.16mm) coverslips were prepared by applying three small drops of mounting media. Either standard mounting media or mounting media including a DAPI fluorescent stain was used where appropriate. Slides were placed carefully face down upon the coverslips and secured around the edges with nail varnish. The sections were stored in the fridge for at least one hour before microscopy and were stored in the fridge when not in use.

2.6.2. Adapted protocol

The standard protocol was used as our default immunohistochemistry method when examining neuronal, glial, or stem cell markers, or when optimising a new antibody. However, for the methylation modifications 5hmC and 5caC, more sensitive techniques were required. Whilst the 5hmC antibody is robust and may even work with the standard protocol, measurement of the 5caC modification is more difficult. In addition, depurination of the tissue sections in hydrochloric acid is required to detect cytosine modifications such as 5hmC or 5caC, and this treatment is not compatible with the majority of additional antigen retrieval protocols that may otherwise be applied to enhance the 5caC signal (Abakir et al., 2016b).

TABLE 2.16. Antibodies used in the immunohistochemistry protocols

Antibody	Source	Class	Raised	Dilution
Anti-NeuN	Millipore; MAB377	Primary	Mouse	S: 1:100
Anti-GFAP	ThermoFisher; RM-2125-S	Primary	Rabbit	S: 1:50
Anti-DCX	Abcam; ab18723	Primary	Rabbit	S: 1:500
Anti-5hmC	Active Motif; 39999	Primary	Mouse	A: 1:2000
Anti-5caC	Active Motif; 61225	Primary	Rabbit	A: 1:500
Alexa Fluor 488	Abcam; ab150105	Secondary	Mouse	S: 1:200, A: 1:500
Alexa Fluor 568	Abcam; ab175471	Secondary	Rabbit	S: 1:500, A: 1:500

S = standard protocol; A = adapted protocol; NeuN = Fox-3; GFAP = Glial fibrillary acidic protein; DCX = Doublecortin; 5hmC = 5-methylhydroxycytosine; 5caC = 5-carboxylcytosine

To overcome this issue, the protocol of one of our collaborators, Alexey Ruzov, was used which specifically utilises a Tyramide Signal Amplification (TSA) kit to boost fluorescent signal. TSA kits take advantage of the reaction between horseradish peroxidase, hydrogen peroxidase, and tyramine. An enzymatic reaction caused by hydrogen peroxide results in the addition of radicals to tyramine, alongside the production of other radicals by horseradish peroxidase catalytic activity. Radicalised tyramine is then free to bind to the tissue and create an amplified signal in the area of the antigen. This method is extremely quick, has excellent spatial resolution and antibody specificity, and can be imaged with standard brightfield or confocal microscopes.

The adapted protocol follows the majority of the steps from the standard protocol with a few additional stages. Sections were placed in staining racks and then into glass staining dishes. PBX was made up by taking 50ml of PBS and adding 250µl of Triton X-100 before mixing well. Sections were dewaxed using two 10 minute washes in xylene in a fume hood, followed by two 10 minute washes in 100% ethanol outside the fume hood at room temperature. The sections were transferred from the staining dish to a glass coplin jar and given one 15 minute wash in PBX at room temperature. Following this, the sections were placed back into the fume hood and washed for 1 hour in 4M hydrochloric acid.

After one hour, the sections were removed from the 4M hydrochloric acid and given five 5 minute washes in PBS within a coplin jar at room temperature. The sections were removed from the coplin jar and placed into a wet chamber. Once inside, the sections of tissue mounted on each slide were surrounded with a hydrophobic barrier. Primary antibody solution was made by adding the requisite amount of primary antibody to 900µl of PBS and 100µl of 10% bovine serum, which is the blocking agent in the adapted protocol. Table 2 lists the primary antibodies used and their concentrations. 200µl of primary antibody solution was applied to each section and left in the enclosed wet chamber for one hour at room temperature.

After one hour, the primary antibody solution was removed and the sections were removed from the wet chamber. PBT was made up by adding 0.01% Tween 20 to PBS. The sections were given five 5 minutes washes in PBT within a coplin jar at room temperature, before being placed back into the wet chamber. Secondary antibody solution was made up by adding the requisite amount of secondary antibody to 900µl of PBS and 100µl of 10% bovine serum. Table 2.16 shows the

secondary antibodies used and their concentrations. 200µl of secondary antibody solution was applied to each section and left in the enclosed wet chamber for one hour at room temperature.

After one hour, the secondary antibody solution was removed and the sections were removed from the wet chamber. The sections were given five 5 minutes washes in PBT within a coplin jar at room temperature, before being placed back into the wet chamber. The TSA fluorescence solution was made up by mixing the TSA Plus Amplification Reagent in 60µl of dimethyl sulfoxide (DMSO). 5µl of this tyramide solution was added to 1000µl of the TSA dilution buffer. Optimisation is required to identify the right amount of tyramide solution to use and the length of incubation required. With the advice of our collaborator, Alexey Ruzov, 5µl of tyramide solution was used for exactly 2 minutes as an optimal procedure for detecting 5hmC and 5caC. Due to the time sensitive nature of the TSA amplification, TSA incubation was completed with no more than three slides at a time to ensure an accurate duration of incubation. With the slides in the wet chamber, the TSA solution was applied to each section, within each of the hydrophobic barriers, for 2 minutes exactly. After 2 minutes, the sections were placed in a coplin jar and immediately given three 5 minute washes in PBT.

After the PBT washes, the area of the slides around the hydrophobic barriers was dried carefully prior to mounting. 0.16mm coverslips were prepared by applying 3 small drops of mounting media. Either standard mounting media or mounting media including a DAPI fluorescent stain was used where appropriate. Slides were then placed carefully face down upon the coverslips and secured around the edges with nail varnish. The sections were stored in the fridge for at least one hour before microscopy and were stored in the fridge when not in use.

2.6.3. Microscopy and image analysis

Once the relevant immunohistochemistry protocol was completed, sections were visualised for characterisation of the methylation modifications and any further reference immunoreactivity using neuronal, glial, or stem cell markers. Microscopy of the sections was completed using two microscopes within the SLIM facility at the University of Nottingham. The Leica DM-IRE2 fluorescence microscope was used primarily for visualisation of the standard protocol during optimisation of neuronal antibodies. The Zeiss LSM 710 confocal microscope was used for the majority of

imaging, including optimisation and characterisation of the methylation marks 5hmC and 5caC.

For the Leica DM-IRE2, the Volocity imaging suite was used to capture images. Magnification was mostly kept at 20x, but closer analysis at 63x magnification was occasionally used and required the administration of oil to the objective. As our immunohistochemistry only ever involved the use of a rabbit or mouse antibody alongside DAPI cellular staining, the fluorescent signal was split into two channels for the rabbit (Alexa Fluor 488; green) and mouse (Alexa Fluor 568; red), with DAPI remaining as the traditional blue. For the DM-IRE2, these channels were FITC for green and TRITC for red respectively. Max exposure time was set at 100ms to avoid bleaching of the signal. Composite and single channel images were then exported for image analysis.

For the Zeiss LSM 710 confocal, the Zen imagine suite was used to capture images. Again, magnification was mostly kept at 20x but closer analysis at 40x was occasionally used and required the administration of oil to the objective. Again, only rabbit or mouse antibodies were used in this microscopy alongside DAPI staining. For the LSM 710, Zen allows one to manually edit the proportion of signal included in each channel. Thus, one could ensure that there was as little crossover as possible between the 488 green and 568 red channels. Confocal imaging necessitates more detailed adjustment of the laser and of the pinhole. As such, owing to the relative weakness of the 5caC antibody, a higher-than-average laser power was chosen. However, this can increase the risk of bleaching the signal and thus corrupting the fluorescence. For the 488 green channel, a laser power of 5% was applied. For the 568 red channel, a laser power of 10% was applied. Thus, care was taken when imaging the 5caC signal to ensure that the signal does not become bleached. 1 airy unit (the width of the pinhole which controls the confocal resolution of the imaging) was kept as the default for all of the imaging sessions. Averaging, which averages out the multiple signal intensities emitted at any given point in the tissue, was set to 4. Bit depth pixel averaging was kept at 12, which is the recommended level for imaging of fluorescence. Manual adjustment of the gain master and digital offset functions allowed modulation of saturated pixels, so that one can maximise the fluorescent signal without bleaching. Composite and single channel images were then exported for image analysis, most often at 1024 x 1024 resolution but occasionally at 512 x 512 if the image was computationally intensive.

Once the sections had been imaged, the files were imported into ImageJ for final processing. ImageJ allows one to merge the single channel or composite image files from the microscopy software to create quality output images. ImageJ was used to add scale bars to each image. The Set Scale tool is used to assess and, if necessary, amend the spatial properties included in the image file. The Scale Bar tool is then used to annotate the image with a scale bar. Some images, particularly during optimisation of the 5caC antibody, suffered from heavy non-specific background staining. In these cases, the Brightness/Contrast tool was used to confirm that brightness and contrast levels were within normal boundaries and not skewed.

3. Assessment of *DNMT3L* R278G and *APOE* in relation to cognitive performance in the OPTIMA cohorts

3.1. Preface

Genotyping of the *DNMT3L* R278G variant in the OPTIMA, VITACOG, and Challenge study cohorts was performed using a KASP genotyping assay by a former student in our research group, Mariam Muse. Genotype data for *APOE* was provided by the OPTIMA study. To build upon previous investigation of cognitive performance in the VITACOG cohort, Principal Component Analysis was applied to extrapolate and analyse dementia-sensitive cognitive factors in these individuals, alongside the *DNMT3L* R278G and *APOE* variants.

3.2. Results

3.2.1. *DNMT3L* R278G and *APOE* genotypes in the combined OPTIMA strands

The first examination was a simple assessment of the *DNMT3L* R278G and *APOE* genotype counts and allele frequencies within the Alzheimer's disease (AD), mild cognitive impairment (MCI), and control individuals aggregated from the OPTIMA, VITACOG, and Challenge cohorts. Minor allele frequencies and genotype frequencies are presented in Table 3.1 and Table 3.2 respectively.

Before assessing the allele frequencies for each disease cohort, the *DNMT3L* R278G and *APOE* genotypes were assessed for Hardy-Weinberg equilibrium. For *DNMT3L* R278G, the AD ($p = 0.18$), MCI ($p = 0.13$), and control ($p = 0.34$) populations all conformed to Hardy-Weinberg equilibrium. For *APOE*, the AD ($p < 0.01$), MCI ($p < 0.001$), and control ($p < 0.001$) populations did not conform to Hardy-Weinberg equilibrium. This is perhaps not unexpected given that the populations stem from the OPTIMA program. The aim of OPTIMA was to investigate the progress of dementia, of which *APOE* is an important factor, and it is likely that population stratification contributed to a non-random sample of *APOE* genotypes (Wittke-Thompson et al., 2005).

Previous work investigating the *DNMT3L* R278G found an association between minor allele carriers and higher intelligence scores (Haggarty et al., 2010), an association that was particularly strong when using a dominant genetic model (A/A v

A/G + G/G) compared to an additive genetic model (A/A v A/G v G/G). A dominant model asserts that one or two copies of the minor allele are required for an increase in disease risk in contrast to a recessive model where two copies of the minor allele are required for an increase in disease risk. An additive model asserts that disease risk increases from major allele homozygotes to heterozygotes to minor allele homozygotes (Clarke et al., 2011). In light of the dominant association observed with this genotype in previous work, *DNMT3L* R278G genotypes were recoded using a dominant genetic model by classifying A/G heterozygotes and G/G homozygotes as G carriers (G+) in comparison to A/A homozygotes.

Beginning with *DNMT3L* R278G, no difference in allele frequency was found between the AD, MCI, and control groups ($\chi^2 = 0.460$, $p = 0.794$). A closer inspection of the R278G G carriers alone was performed, comparing the controls with MCI and MCI with AD. Odds ratios (OR) showed that G carriers had a greater chance of being in the MCI group than the control group (OR = 1.805, 95% CI 1.003 – 3.247), and a greater chance of being in the AD group than the MCI group (OR = 1.793, 95% CI 0.991 – 3.244). This alludes to a possible dementia-risk relationship for the *DNMT3L* R278G minor allele carriers.

Sex was subsequently included in the analysis of *DNMT3L* R278G genotype. An overall effect was observed when assessing combined AD, MCI, and control groups in the analysis ($\chi^2 = 16.577$, $p = 0.011$). Splitting the analysis to look at male and female allele frequencies within each disease group showed no effect in AD ($\chi^2 = 1.076$, $p = 0.3$) or controls ($\chi^2 = 0.908$, $p = 0.341$). However, a significant effect was observed in the MCI group ($\chi^2 = 4.099$, $p = 0.043$). This appeared to be driven by female R278G minor allele carriers, where the minor allele count was 97 (expected 87) compared to 43 for the males (expected 53).

TABLE 3.1. Minor allele frequencies for *DNMT3L* R278G and *APOE*

Disease	*DNMT3L* R278G	*APOE* E4+
AD	0.25	0.45
MCI	0.24	0.15
Control	0.25	0.14

E4+ = Carrier of the *APOE* E4 allele.

TABLE 3.2. Genotype frequencies for *DNMT3L* R278G and *APOE*

Disease	***DNMT3L* R278G**			***APOE* E4**					
	A/A	**A/G**	**G/G**	**E2E2**	**E2E3**	**E2E4**	**E3E3**	**E3E4**	**E4E4**
AD	0.533	0.423	0.044	0.007	0.029	0.029	0.248	0.504	0.182
MCI	0.566	0.393	0.041	0.007	0.129	0.044	0.566	0.244	0.010
Control	0.542	0.408	0.049	0.000	0.155	0.021	0.592	0.204	0.028

TABLE 3.3. Variants and genotypes associated with *APOE* E4

***APOE* allele**	**rs429358**	**rs7412**	**Genotype**
E2E2	T/T	T/T	No E4
E2E3	T/T	T/C	No E4
E2E4	T/C	T/C	E4+
E3E3	T/T	C/C	No E4
E3E4	T/C	C/C	E4+
E4E4	C/C	C/C	E4+

E4+ = Carrier of the *APOE* E4 allele; No E4 = Does not carry the *APOE* E4 allele.

Having observed some evidence for disease risk in the minor allele carries in our analysis of *DNMT3L* R278G, analysis progressed to the examination of *APOE*. There are six genotypes associated with the *APOE* epsilon variant due to the combination of two risk variants, rs429358 and rs7412 (Table 3.2). The presence of an E4 allele is most strongly associated with risk for AD, whilst the presence of an E2 allele has been posited as being protective against AD (Wu and Zhao, 2016). Due to the prominence of the E4 allele in AD research, the *APOE* genotypes were recoded so that individuals were classed as either *APOE* E4 carriers (E4+) or *APOE* E4 non-carriers (no E4). This carrier status was used as opposed to allele frequency due to the multiple variants that make up the *APOE* genotype.

In contrast to the *DNMT3L* R278G genotype, a significant difference was found between the AD, MCI, and control groups ($\chi^2 = 82.720$, $p < 0.001$). Due to the relationship between *APOE* E4 and risk for AD, the analysis was repeated comparing AD with MCI and AD with controls. As expected, the *APOE* E4 carrier status in the AD group was significantly different from both the MCI group ($\chi^2 = 59.575$, $p < 0.001$) and the control group ($\chi^2 = 66.359$, $p < 0.001$), with *APOE* E4 carriers showing greater risk of being in the AD group than the MCI group (OR = 5.911, 95% CI 3.780 – 9.243) or control group (OR = 7.399, 95% CI 4.356 – 12.567). There was no significant difference between the MCI and control groups ($\chi^2 = 7.181$, $p = 0.066$). With regards to the influence of sex, splitting the analysis to assess male and female carrier status within each disease group showed no effect in AD ($\chi^2 = 0.797$, $p = 0.372$), MCI ($\chi^2 = 0.399$, $p = 0.527$), or controls ($\chi^2 = 3.354$, $p = 0.067$).

Finally, a synergy factor analysis approach was applied to the risk alleles for our genotypes of interest. This statistic was designed to give an estimation, albeit a limited one, of the synergistic interaction between risk or protective genetic factors (Cortina-Borja et al., 2009). This method was applied to the *DNMT3L* R278G G/G homozygotes and to the *APOE* E4 carriers, comparing the control individuals with the AD and MCI individuals in separate analyses. A synergy factor (SF) score > 1 implies a synergistic interaction between the two factors whilst an SF score < 1 implies an antagonistic interaction. Comparison of controls and MCI individuals resulted in the prediction of a non-significant antagonistic interaction (SF = 0.33, $p = 0.22$). Comparison of controls and AD resulted in the prediction of a significant antagonistic interaction between these two risk genotypes (SF = 0.07, $p = 0.04$). However, the weak SF score indicates that either the significant *p* value was driven by the strong

relationship between *APOE* E4 and AD, or that the *DNMT3L* R278G G/G genotype is not a true risk factor for disease when compared with *APOE* E4.

3.2.2. tHcy levels in the VITACOG cohort

The following set of analyses focused specifically on the 271 individuals available from the VITACOG cohort. As previously detailed, homocysteine (Hcy) levels are an important risk factor for dementia progression as well as being related to DNA methylation via the methionine pathway. Using the data available from VITACOG, our analysis confirmed the findings published by Smith et al. (2010) that individuals who received B vitamin treatment showed a significant decrease in tHcy levels following completion of the trial ($F = 147.149$, $p < 0.001$), whilst individuals who received the placebo showed a slight increase in tHcy levels. It was also confirmed that those with upper quartile baseline tHcy levels, the highest levels of baseline Hcy, showed the greatest reduction in Hcy following completion of the trial.

Next, the *DNMT3L* R278G and *APOE* E4 genotypes were included into our analysis of the change (Δ) in tHcy levels during the VITACOG trial. No difference in ΔtHcy levels was observed between the *DNMT3L* R278G A/A homozygotes or G carriers ($t = -1.582$, $p = 0.115$), and this was irrespective of B vitamin treatment or placebo group ($F = 0.754$, $p = 0.386$). Likewise, no difference in ΔtHcy levels was observed between the *APOE* E4 carriers and E4 non-carriers ($t = -0.419$, $p = 0.676$), and again this was irrespective of B vitamin treatment or placebo group ($F = 0.029$, $p = 0.866$). With regards to baseline tHcy, no difference was observed between *DNMT3L* R278G A/A homozygotes and G carriers ($t = 0.827$, $p = 0.409$). No difference in baseline tHcy was also observed between *APOE* E4 carriers and E4 non-carriers ($t = -0.418$, $p = 0.677$). A closer examination of those with low baseline tHcy revealed a slight effect of *APOE* (Figure 3.1), where E4 carriers who received the placebo showed a reduced increase in tHcy compared to E4 non-carriers ($F = 4.319$, $p = 0.040$).

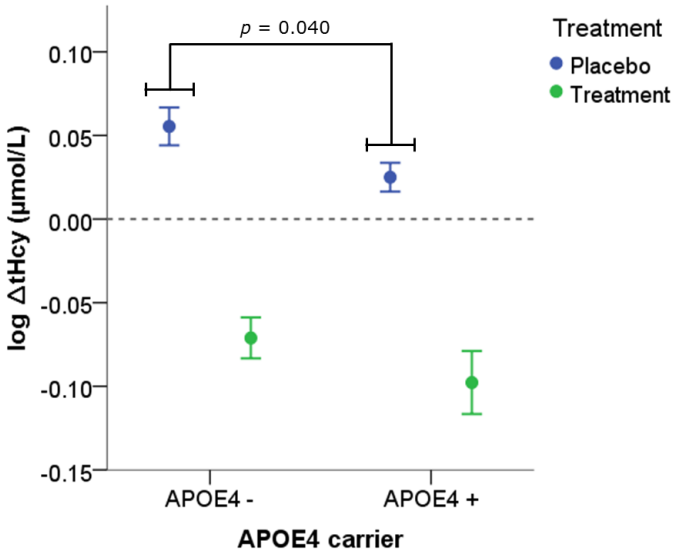


FIGURE 3.1. Inspection of ΔtHcy levels in VITACOG individuals with low baseline tHcy. Green represents B vitamin treatment and blue represents the placebo. Errors bars indicate +/- 1 SE.

3.2.3. Cognitive performance in the VITACOG cohort

Our next set of analyses focused on the relationship between our genotypes of interest and cognitive performance in the VITACOG cohort. Initial analyses were performed on change in the separate cognitive test measures chosen for the present study across the duration of the VITACOG study. This includes the Hopkins Verbal Learning Test-Revised (HVLT-R) delayed recall score, Mini-Mental State Examination (MMSE) summary score, Graded Naming Test (GNT) score, Category Fluency (CF) score, and Paired Associates Learning (PAL) total errors.

Inspection of covariates revealed no significant influence of age or sex in any of the cognitive tests, except for a significant correlation between age and ΔGNT where older individuals performed worse than younger individuals ($R = -0.248$, $p < 0.001$). The full results for these analyses are presented in Table 3.4, with bold values indicating statistically significant results. Evaluation of B vitamin treatment and each cognitive test revealed that, whilst performance on the HVLT-R, MMSE, PAL, and CF all showed significant decline over the course of the VITACOG trial, there was no effect of the B vitamin treatment on this pattern of decline. The full results for these analyses are presented in Table 3.5.

In keeping with the results investigating age, sex, and B vitamin treatment, examination of the *DNMT3L* R278G and *APOE* E4 genotypes revealed no significant association with Δ in performance for any of the cognitive tests over the duration of the VITACOG trial. However, a borderline significant association was seen for PAL total errors, where *APOE* E4 carriers showed greater decline in performance than *APOE* E4 non-carriers ($t = 1.974$, $p = 0.050$). The results for these are presented in Table 3.6.

TABLE 3.4. Influence of sex and age on individual cognitive tests in the VITACOG cohort

Cognitive test	Age	Sex
ΔHVLT-R	R = -0.076, *p* = 0.265	t = -0.207, *p* = 0.836
ΔMMSE	R = -0.123, *p* = 0.069	t = 0.294, *p* = 0.769
ΔPAL	R = 0.080, *p* = 0.245	t = -1.464, *p* = 0.145
ΔGNT	R = -0.248, *p* < 0.001**	t = 0.568, *p* = 0.570
ΔCF	R = -0.085, *p* = 0.212	t = -0.825, *p* = 0.410

ΔHVLT-R = Change in Hopkins Verbal Learning Test – Revised; ΔMMSE = Change in Mini-Mental State Examination; ΔPAL = Change in Paired Associates Learning; ΔGNT = Change in Graded Naming Test; ΔCF = Change in Category Fluency.
** = significant at the 0.01 level.

TABLE 3.5. Change in performance and effect of treatment for each cognitive test

Cognitive test	Δ in performance	Treatment
HVLT-R	F = 6.252, *p* = 0.013*	F = 1.245, *p* = 0.266
MMSE	F = 10.632, *p* = 0.001*	F = 0.144, *p* = 0.704
PAL	F = 54.139, *p* < 0.001**	F = 0.732, *p* = 0.393
GNT	F = 0.007, *p* = 0.932	F = 0.554, *p* = 0.458
CF	F = 4.500, *p* = 0.035*	F = 2.007, *p* = 0.158

HVLT-R = Hopkins Verbal Learning Test – Revised; MMSE = Mini-Mental State Examination; PAL = Paired Associates Learning; GNT = Graded Naming Test; CF = Category Fluency.
** = significant at the 0.01 level, * = significant at the 0.05 level

TABLE 3.6. Influence of *DNMT3L* R278G and *APOE* E4 genotypes on each cognitive test

Cognitive test	*DNMT3L* R278G	*APOE* E4
ΔHVLT-R	t = 1.127, *p* = 0.261	t = -0.418, *p* = 0.676
ΔMMSE	t = -0.532, *p* = 0.595	t = -0.582, *p* = 0.561
ΔPAL	t = 0.685, *p* = 0.494	t = 1.974, *p* = 0.050
ΔGNT	t = -1.221, *p* = 0.223	t = -0.700, *p* = 0.485
ΔCF	t = 0.595, *p* = 0.552	t = -0.1697, *p* = 0.091

ΔHVLT-R = Change in Hopkins Verbal Learning Test – Revised; ΔMMSE = Change in Mini-Mental State Examination; ΔPAL = Change in Paired Associates Learning; ΔGNT = Change in Graded Naming Test; ΔCF = Change in Category Fluency.

In addition, the association between genotype and baseline performance was investigated in each of the cognitive tests. Whilst performance in HVLT-R, MMSE, PAL, and GNT showed no difference at baseline between *DNMT3L* R278G A/A homozygotes and G carriers, or between *APOE* E4 carriers and E4 non-carriers, a difference was observed in baseline CF scores for the *DNMT3L* R278G genotype ($t = -2.214$, $p = 0.028$). Closer examination revealed that G carriers performed significantly better than A/A homozygotes in the CF test at baseline. However, differences in baseline cognitive performance do not influence scores of Δ in cognitive performance, which was used for much of our subsequent analysis. Therefore, although this result reveals an interesting baseline association with *DNMT3L* R278G, it did not influence our deeper analysis of cognitive performance and *DNMT3L* R278G in the VITACOG cohort. The results for these analyses are presented in Table 3.7.

3.2.4. Principal Component Analysis of cognitive tests in the VITACOG cohort

Apart from the association between age and ΔGNT performance and the association between the *DNMT3L* R278G genotype and baseline CF performance, no other relationships between the genotypes of interest and cognitive performance were observed. In addition, whilst B vitamin treatment clearly influenced Hcy levels in the VITACOG cohort, no relationship was observed between treatment group and cognitive performance.

In an effort to further interrogate the influence of genotype on cognitive performance, Principal Component Analysis (PCA) was applied to Δ in performance using the HVLT-R, MMSE, PAL, GNT, and CF cognitive measures in the VITACOG trial. PCA is an adaptable statistical method for identifying the main sources of variance within a set of experimental variables. Amongst its applications is the identification of related factors within cognitive test data (Lyall et al., 2016). As noted in the previous section, PCA was applied to Δ in our cognitive test data to avoid any bias associated with baseline differences in cognitive performance.

TABLE 3.7. Influence of *DNMT3L* R278G and *APOE* E4 genotypes on baseline cognitive performance

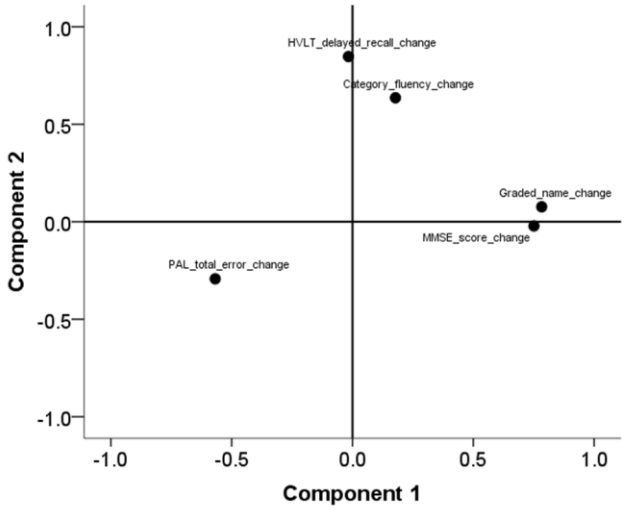
Cognitive test	***DNMT3L* R278G**	***APOE* E4**
HVLT-R	t = -1.448, *p* = 0.149	t = -1.771, *p* = 0.078
MMSE	t = 0.290, *p* = 0.772	t = -1.342, *p* = 0.181
PAL	t = 0.509, *p* = 0.611	t = 0.912, *p* = 0.363
GNT	t = 0.609, *p* = 0.543	t = -0.599, *p* = 0.550
CF	t = -2.214, *p* = 0.028*	t = -0.962, *p* = 0.337

HVLT-R = Hopkins Verbal Learning Test – Revised; MMSE = Mini-Mental State Examination; PAL = Paired Associates Learning; GNT = Graded Naming Test; CF = Category Fluency.
* = significant at the 0.05 level.

A measure of sampling adequacy (KMO & Bartlett's test) greater than 0.6 is sought for accurate PCA analysis. Our analysis produced a value of 0.631 ($p < 0.001$) and resulted in the identification of two cognitive factors; PCA Factor 1 which consisted of MMSE, PAL, and GNT, and PCA Factor 2 which consisted of HVLT-R and CF. PCA Factor 1 was interpreted to be reflective of "visuospatial associative memory" due to the segmentation of PAL and GNT, visuospatial tests that have been highlighted for their accuracy in predicting dementia progression (Blackwell et al., 2004). PCA Factor 2 was interpreted to be reflective of "verbal semantic memory" due to the segmentation of HVLT-R and CF, semantic tests that have been combined to better evaluate MCI symptoms (Duara et al., 2011). Therefore, PCA Factor 1 and PCA Factor 2 will be referred to as visuospatial associative and verbal semantic memory from here on. PCA results and correlation coefficients for the cognitive data are presented in Figure 3.2.

Having identified two distinct cognitive factors within the VITACOG cohort, the assessment of cognitive performance was continued using visuospatial associative and verbal semantic memory as our measures. A significant relationship was found between age and performance in visuospatial associative ($R = 0.182$, $p = 0.008$), where increased age was associated with risk of poorer performance. This was not observed for verbal semantic memory ($R = 0.046$, $p = 0.509$). There was no relationship between sex and performance in either visuospatial associative ($t = 1.366$, $p = 0.173$) or verbal semantic ($t = -1.543$, $p = 0.124$) memory. Analysis of treatment group revealed that there was no impact of B vitamin treatment or placebo on visuospatial associative ($t = -0.135$, $p = 0.893$) or verbal semantic ($t = 0.013$, $p = 0.990$) memory. Baseline tHcy level also showed no relationship with either visuospatial associative ($t = 1.215$, $p = 0.226$) or verbal semantic ($t = 0.447$, $p = 0.655$) memory performance, despite those individuals with upper quartile baseline tHcy showing a slight decline in visuospatial associative memory performance compared to those with lower or middle quartile tHcy levels.

A



B

	Component	
	Visuospatial associative	Verbal semantic
GNT	0.783	
MMSE	0.751	
PAL	-0.568	
HVLT-R		0.847
CF		0.636

C

Δ	HVLT-R	MMSE	PAL	GNT	CF	Visuospatial associative	Verbal semantic
HVLT-R	1						
MMSE	0.058	1					
PAL	-0.198**	-0.197**	1				
GNT	0.08	0.276**	-0.297**	1			
CF	0.186**	0.141*	-0.103	0.175**	1		
Visuospatial associative	-0.017	0.751**	-0.568**	0.783**	0.178**	1	
Verbal semantic	0.847**	-0.021	-0.293**	0.077	0.636**	0	1

FIGURE 3.2. PCA of select cognitive tests from the VITACOG cohort. This analysis resulted in the identification of two cognitive factors (A) that were named visuospatial associative and verbal semantic memory respectively (B). Correlation coefficients between the tests and the resultant cognitive factors are included (C). HVLT-R = Hopkins Verbal Learning Test - Revised; MMSE = Mini-Mental State Examination; PAL = Paired Associates Learning; GNT = Graded Naming Test; CF = Category Fluency. ** = $p < 0.01$, * = $p < 0.05$ (2 tailed, Pearson).

The association between genotype and performance on the derived cognitive factors was then examined. Beginning with *DNMT3L* R278G, it was discovered that G carriers who received B vitamin treatment showed a trend towards significant improvement in visuospatial associative memory compared to A/A homozygotes who received treatment (t = -1.569, p = 0.06; Figure 3.3A). In contrast, G carriers who received B vitamin treatment showed a significant decline in verbal semantic memory compared to A/A homozygotes who received treatment (t = 1.727, p = 0.043; Figure 3.3B). Individuals who received the placebo showed no change in either visuospatial associative or verbal semantic memory, irrespective of their genotype. These results indicate a potential pharmacogenetic interaction between B vitamin treatment and the *DNMT3L* R278G variant which influences change in cognitive performance.

In contrast, *APOE* E4 carriers who received the placebo showed a significant decline in visuospatial associative memory in comparison to *APOE* E4 non-carriers who received the placebo (t = -2.938, p = 0.004; Figure 3.3C). Individuals who received the B vitamin treatment showed no change in visuospatial associative memory. Closer inspection of the tests that make up visuospatial associative memory revealed a significant statistical interaction of the *APOE* genotype and B vitamin treatment with regards to ΔPAL (F = 4.474, p = 0.036), driven by significantly poorer performance for *APOE* E4 carriers compared to *APOE* E4 non-carriers in the placebo group (t = 3.311, p = 0.001). There was no significant relationship between *APOE* and B vitamin treatment group with regards to verbal semantic memory, despite the APOE E4 carriers displaying poorer performance in both the treatment and placebo conditions (t = -1.228, p = 0.224; Figure 3.3D).

3.2.5. Influence of baseline tHcy on cognitive performance

By assessing our genotypes of interest and by utilising our derived cognitive factors, it was possible to identify an influence of B vitamin treatment which was otherwise unobserved when examining individual cognitive tests irrespective of genotype. Due to the moderating influence of baseline Hcy levels, high and low baseline tHcy levels were incorporated into the analysis of our derived cognitive factors.

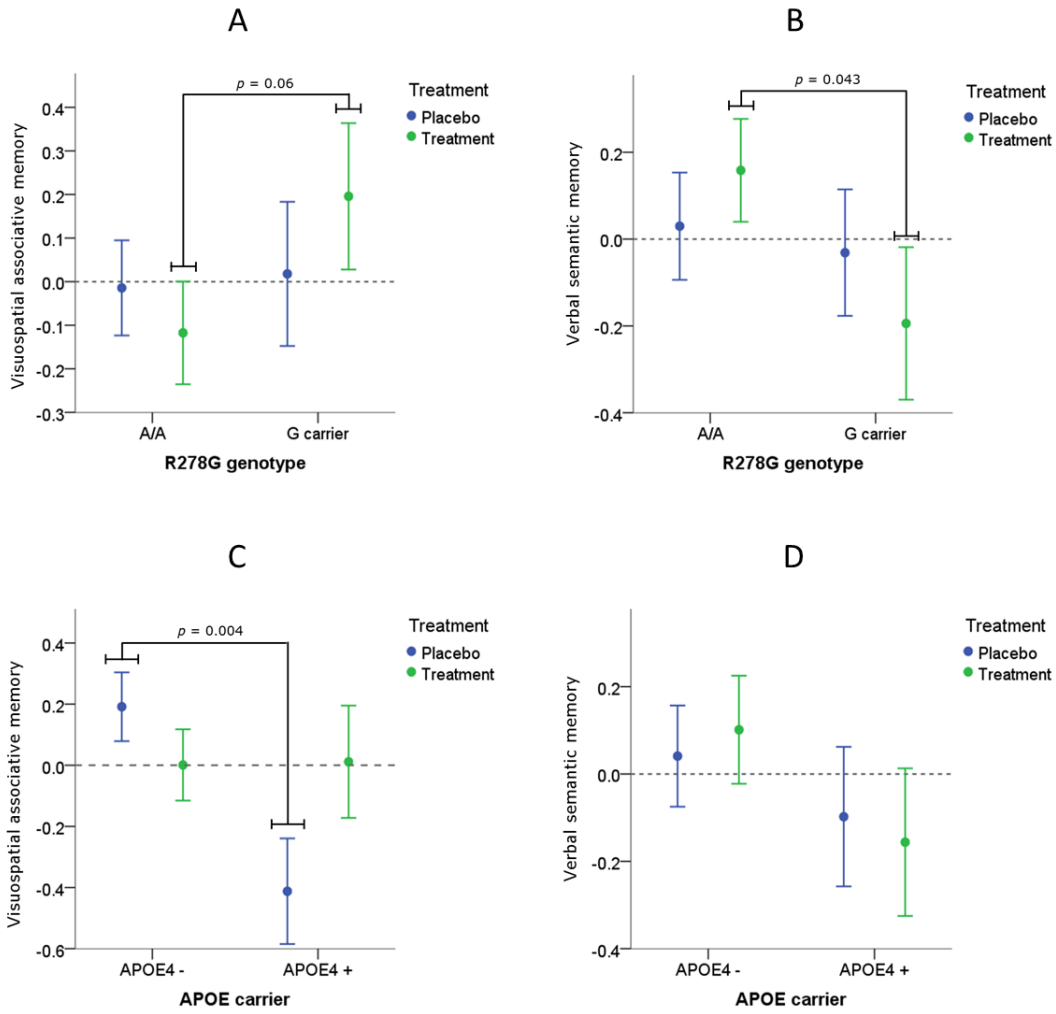


FIGURE 3.3. Investigation of the relationship between B vitamin treatment and our genotypes of interest with respect to cognitive performance. These relationships are presented for *DNMT3L* R278G and *APOE* E4 genotypes with either visuospatial associative memory (A, C) or verbal semantic memory (B, D) as the dependent variable. Errors bars indicate +/- 1 SE.

With regards to visuospatial associative memory, a significant statistical interaction of baseline tHcy, B vitamin treatment, and the *DNMT3L* R278G genotype was observed (F = 3.718, *p* = 0.012). This observation remained significant after adjustment for the influence of age as a covariate (*p* = 0.021). *Post hoc* analysis with Bonferroni correction showed that G carriers with high baseline tHcy who received B vitamin treatment showed significantly better performance than A/A homozygotes in the same conditions (*p* = 0.031; Figure 3.4A). No such statistical interaction was observed for verbal semantic memory (F = 0.506, *p* = 0.679), despite a trend towards poorer performance for G carriers with high baseline tHcy who received B vitamin treatment compared to A/A homozygotes in the same conditions (Figure 3.4B).

This approach was then replicated with upper quartile versus lower & middle baseline tHcy groups. A significant statistical interaction of baseline tHcy, treatment, and the *DNMT3L* R278G genotype on visuospatial associative memory remained (F = 3.723, *p* = 0.012), even after adjustment for the influence of age as a covariate (F = 3.434, *p* = 0.018). This appeared to be driven by the improved performance of G carriers with upper quartile baseline tHcy who received B vitamin treatment (Figure 3.4C). *Post hoc* analysis with Bonferroni correction confirmed that this observation was again significant in comparison to A/A homozygotes in the same conditions (*p* = 0.014), but not in comparison to G carriers in the other baseline tHcy and treatment conditions (*p* > 0.05). Again, no such statistical interaction was seen for verbal semantic memory (F = 0.377, p = 0.770; Figure 3.4D). These findings build upon the pharmacogenetic interaction reported earlier identifying *DNMT3L* R278G minor allele carriers, high baseline tHcy, and B vitamin treatment as influential factors relating to cognitive performance.

Analysis of the *APOE* genotype revealed a significant statistical interaction of baseline tHcy, treatment, and the *APOE* genotype for visuospatial associative memory (F = 2.380, *p* = 0.040). However, this observation did not reach statistical significance after adjustment for the influence of age as a covariate (F = 2.083, *p* = 0.104). *Post hoc* analysis with Bonferroni correction showed that *APOE* E4 carriers with low baseline tHcy in the placebo group performed significantly worse than *APOE* E4 non-carriers in the placebo group, whether they had high (*p* = 0.033) or low (*p* = 0.014) baseline tHcy (Figure 3.4E). No such statistical interaction was observed for verbal semantic memory (F = 0.347, *p* = 0.791; Figure 3.4F). Examination of upper quartile versus lower & middle baseline tHcy groups resulted in no significant statistical interaction of baseline tHcy, treatment, and the *APOE* genotype for either

visuospatial associative (F = 2.097, *p* = 0.102) or verbal semantic (F = 0.310, *p* = 0.818) memory. With regards to visuospatial associative memory, a difference was observed between *APOE* E4 carriers and non-carriers in the placebo group who had lower & middle baseline tHcy. However, *post hoc* analysis with Bonferroni correction confirmed that this was non-significant (*p* = 0.142).

The inclusion of baseline tHcy levels into the analysis developed our model of the pharmacogenetic relationship between B vitamin treatment, genotype, and cognitive function. Specifically, the improved visuospatial associative memory performance observed in *DNMT3L* R278G minor allele carriers reached greater statistical significance when assessing those with the highest levels of baseline tHcy. In contrast, the decline in verbal semantic memory performance observed in G carriers became non-significant upon inclusion of baseline tHcy levels. Finally, the significant decline in visuospatial associative memory performance observed in *APOE* E4 carriers remained significant only in those with high levels of baseline tHcy.

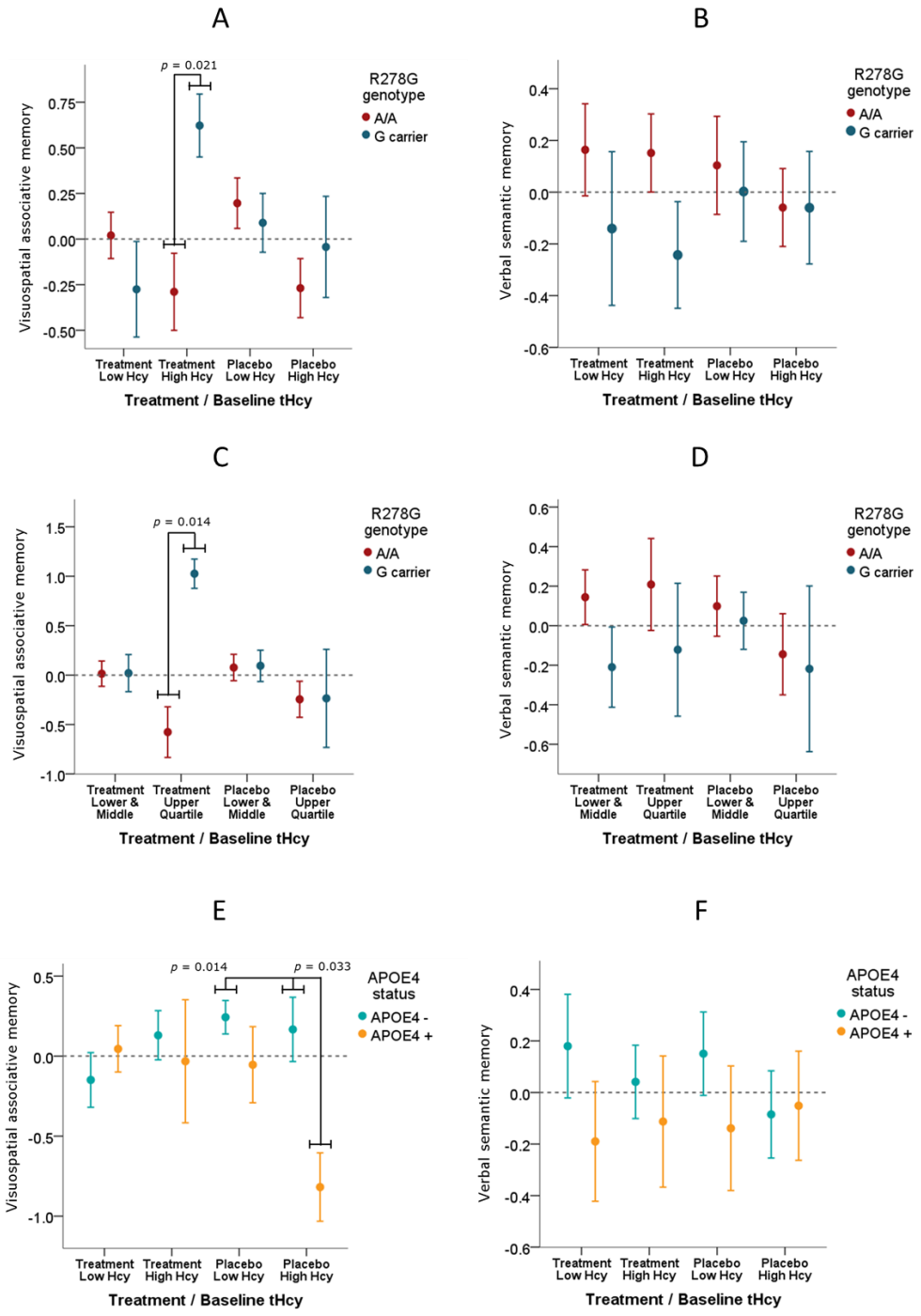


FIGURE 3.4. Interactions between B vitamin treatment, baseline tHcy levels, and genotypes with respect to cognitive performance. Analyses were grouped as B vitamin treatment and high/low Hcy levels (A, B, E, F), as well as B vitamin treatment and quartile Hcy levels for *DNMT3L* R278G (C, D). Errors bars indicate +/- 1 SE.

3.3. Discussion

An increasingly important role in normal brain function and brain disease pathogenesis has been implicated for epigenetic mechanisms such as DNA methylation. The primary focus of this chapter was to investigate the relationship between a specific coding genetic variant within one of the DNA methylation genes and performance in facets of cognitive function.

In the VITACOG cohort, a discrepant influence of the *DNMT3L* R278G minor allele was discovered on visuospatial associative memory and verbal semantic memory performance. Specifically, carriers of the minor allele who received B vitamin treatment exhibited an improvement in visuospatial associative memory performance. Furthermore, this performance was markedly improved in those individuals who had the highest levels of baseline Hcy. In contrast, carriers of the minor allele who received B vitamin treatment exhibited a decline in verbal semantic memory performance. These findings support the hypothesis that changes in cognitive function are associated with the *DNMT3L* R278G variant.

Some notable findings were also observed in relation to the *APOE* genotype in the OPTIMA cohorts and specifically within the VITACOG cohort. As expected, the presence of the *APOE* E4 genotype was higher in AD individuals compared to MCI individuals, but was also higher in MCI individuals compared to controls, supporting the notion that MCI is a prodromal risk stage in AD development. In comparison to our findings for *DNMT3L* R278G and the derived cognitive factors, carriers of the E4 allele who received the placebo exhibited a decline in visuospatial associative memory. Furthermore, there was a marked decline in those individuals who had the highest levels of baseline Hcy. This points to an influence of B vitamin treatment in the E4 carriers as they showed the same visuospatial associative memory performance as those who did not carry an E4 allele.

3.3.1. PCA and cognitive performance

There are two main arguments related to methodology that should be discussed in line with the discrepant PCA results presented in this chapter. The first is the use of PCA itself and its particular application within our study. The second focuses on the nature of the cognitive measures utilised in this PCA analysis and whether the disparate visuospatial associative memory and verbal semantic memory results

accurately highlight differences in cognitive function, as opposed to being artefacts of the methodology and our interpretation of the PCA results.

With regards to the use of PCA, as pointed out by Shlens (2014), misapplication of the assumptions of PCA could result in incorrect and incomplete analysis of the variability within the data. For instance, PCA is increasingly used in genomic studies to handle the vast amounts of multivariate data between organisms and sequences. Cases have been highlighted where the ordination process, an exploratory assessment of the data rather than a hypothesis-driven analysis, has been over-interpreted (Francois et al., 2010). However, when applying PCA to cognitive test data, the goal should be to achieve a certain level of data reduction, condensing the variables you have into representative factors, as opposed to a pure identification of latent factors within the data (Fabrigar et al., 1999).

As the comparative advantages and disadvantages of using PCA as opposed to exploratory factor analysis are contentious (Costello and Osborne, 2005), it was decided to both assess the accuracy of our derived factors using correlation matrices as well as individually analysing each cognitive test using in the PCA to determine the level of corroboration between our factors and the tests. Promisingly, our factors correlated well with their component tests and our analysis of each cognitive test showed some mirroring with the PCA results, particularly in the case of PAL Total Errors and visuospatial associative memory. Further, because there was no *a priori* model of what latent factors were present within the cognitive tests used, it seemed appropriate to favour PCA over exploratory factor analysis. Nevertheless, application of exploratory factor analysis to the cognitive tests could have been carried out, if only to provide a methodological comparison with the results of the PCA.

With regards to the cognitive measures utilised in our PCA, the verbal semantic memory grouping of HVLT-R and CF appeared congruent as both are tests of semantic memory. However, the PCA Cognitive Factor 1 grouping of MMSE, PAL, and GNT (tests of global cognition, episodic memory, and semantic memory respectively) did not share the same congruence. Another contrast between the two PCA factors is the sensory modality associated with the respective cognitive measures. PAL and GNT are both visual tasks, whilst HVLT-R and CF are verbal tasks, and this sensory specificity has been shown to be particularly prominent in tests of memory (Kane et al., 2004).

Despite some incongruence, there is evidence within the literature to support our interpretation of these two cognitive factors. Arguably the most pertinent distinction between the two PCA factors is their respective relevance to dementia. Blackwell et al. (2004) showed that the combined use of PAL and GNT within a battery of cognitive tests most accurately predicted the progression from questionable dementia to AD. Similarly, Duara et al. (2011) showed that performance on HVLT-R and CF could differentiate non-amnestic from amnestic MCI sufferers, likely due to the manner in which categorical remembering can persist during the progressive loss of other executive functions (De Jager et al., 2003). However, this distinction of the PCA factors between dementia progression and particular MCI ability faces some objection, as both Duara et al. (2011) and De Jager et al. (2003) also nominate the use of HVLT-R and CF to better identify progression from MCI to AD. Our observation of a clear genotype-dependent distinction in visuospatial associative and verbal semantic memory performance in our cohort of MCI individuals adds some support to the idea that these two combinations of cognitive tests is measuring separate facets of cognitive impairment.

As explained in the methods, MMSE, PAL, HVLT-R, CF, and GNT were included, and the Clox, Symbol Digit Modalities, Trail Making, and Map Search tests were excluded. MMSE, HVLT-R, and CF were all used in previous studies assessing changes in cognitive ability within the VITACOG study (De Jager et al., 2012). PAL and GNT were included because they had both been used by our research group in previous studies (Knight et al., 2010). Of the tests that were not included in our analyses, Symbol Digit Modalities is a simple clinical test of cognitive function where individuals are asked to pair numbers with figures. Trail Making is a test of visual attention where individuals are asked to connect a set of numbered dots. Map Search involves individuals searching for symbols within a map of a certain geographical area. Finally, Clox involves two drawing tasks; one in which individuals are asked to draw a clock with no guidance, and one in which individuals are asked to copy a clock drawn by the examiner. These tests of executive function (Symbol Digit Modalities, Clox) and visual attention (Trail Making, Map Search) appear distinct from the tests included in our analysis. However, it cannot be discounted that inclusion of these tests may result in a fuller and more accurate analysis of cognitive ability in MCI. This is particularly true in the case of the Clox test, which was found to be stabilised by B vitamin treatment in VITACOG (De Jager et al., 2012).

3.3.2. B vitamins, homocysteine, and genotype

Of particular note was the influence of certain baseline factors which emerged from the analyses. The modulating role of baseline Hcy levels, a factor predicted by previous research to be influential, appeared to be the most pertinent of these variables. Indeed, the observation that high levels of baseline Hcy confer the greatest risk for both change in Hcy levels (Smith et al., 2010) and change in cognitive function (De Jager et al., 2012), and that B vitamin treatment ameliorates this risk, was largely supported by the findings presented within this chapter. For instance, an interactive influence of *DNMT3L* R278G minor allele carriers and B vitamin treatment on visuospatial associative memory performance was initially observed. However, this relationship became particularly significant in those with the highest levels of baseline Hcy. A similar contrasting effect was observed in the *APOE* E4 carriers, where individuals with high levels of baseline Hcy who received the placebo showed the most significant decline in visuospatial associative memory performance.

What could be underlying these relationships between B vitamins (or lack thereof), Hcy levels, and risk genotypes. In the case of *APOE* E4, the findings may be easier to interpret largely due to the established risk of this genotype for dementia. To begin with, whilst raised homocysteine levels in AD are not necessarily linked to the *APOE* genotype, they are modified by the *APOE* E4 genotype in relation to cognitive performance (Elias et al., 2008, Religa et al., 2003). The *APOE* E4 genotype is associated with increased sensitivity of neuronal mechanisms to degeneration and disorder through altered amyloid and tau formation as well as impaired neuronal repair (Horsburgh et al., 2000). Indeed, an association between Hcy levels and amyloid beta formation has been reported (Irizarry et al., 2005). This appears to leave neuronal cells particularly susceptible to Hcy-induced neurotoxicity, most likely in areas related to memory such as the hippocampus (Morris, 2003).

In this scenario, it is feasible that B vitamins could influence either the *APOE* E4 genotype, Hcy levels, or both of these factors to influence cognitive performance. For example, rodent studies of the *APOE* E4 genotype have shown that deprivation of B vitamins such as folate was associated with greater oxidative damage in the brain compared to mice without the *APOE* E4 genotype (Shea et al., 2002). Moreover, treatment of these rodents with vitamins – in this case, vitamin E – led to a reduction in this oxidative damage. This draws parallels with our findings, where *APOE* E4

carriers who received B vitamin treatment did not show the same decrease in visuospatial associative memory performance as carriers who received the placebo.

It was also confirmed that B vitamin treatment reduced Hcy levels in members of the VITACOG cohort. In particular, folate and vitamin B12 play important roles in the one-carbon cycle, where folate is the major methyl donor for the methionine pathway and vitamin B12 allows conversion of Hcy to methionine by facilitating the transfer of methyl from 5-CH3-tetrahydofolate (Rush et al., 2014). Reduction of either of these B vitamins could stall the methionine pathway and cause Hcy levels to build without being converted to methionine, subsequently allowed Hcy-induced neurotoxicity to progress within the brain. This links with our finding that *APOE* E4 carriers (who may be more susceptible to neuronal damage than non-carriers) who did not receive B vitamin treatment saw a decline in visuospatial associative memory performance. Interestingly, an additional mechanism has been proposed that involves high-density lipoprotein cholesterol levels, another risk factor for dementia due to its role in amyloid beta clearance (Zuliani et al., 2010). Work has shown that Hcy can inhibit *APOE* E3, reducing the ability of APOE3 to generate high-density lipoprotein to the levels seen in *APOE* E4 carriers, and providing a common pathway for Hcy and *APOE* E4 to influence cognitive dysfunction (Minagawa et al., 2010).

The picture may be slightly more complicated for *DNMT3L* R278G than for *APOE* E4. It is initially unclear how DNMT3L, a protein almost exclusively implicated in DNA methylation and interactions with other epigenetic proteins (Deplus et al., 2002), could influence cognitive function. The *DNMT3L* R278G variant had previously been linked to improved performance on intelligence tests (Haggarty et al., 2010). Our results showed both poorer verbal semantic memory and improved visuospatial associative memory for carriers of the minor allele. Therefore, discussion will examine the biological and cognitive factors associated with this discrepant performance.

Stemming from the role of DNMT3L in facilitating DNA methylation with DNMT3A, the predicted impact of a functionally important genetic variant in *DNMT3L* would be a general loss of DNA methylation potential (Chedin et al., 2002). Indeed, this effect has been described in the case of the *DNMT3L* R271Q variant. El-Maarri et al. (2009) reported that individuals carrying this rare variant showed significant hypomethylation and confirmed *in vitro* that this resulted from disrupted interaction between DNMT3L and DNMT3A. Nevertheless, the unexpected genotype-dependent

result for visuospatial associative memory observed in our study does not conform to the classical impression of the associated hypomethylation hypothesis. This describes a scenario where genetic variants in methyltransferases cause a loss of methylation potential, the ability of methyltransferases to carry out methylation. Disrupted methylation potential then leads to transcriptional expression changes and an imbalance of the methionine cycle, namely reduced processing of methionine for methyl donation, both of which could be posited as risk factors for cognitive dysfunction (Calvaresi and Bryan, 2001).

It may be possible to tailor the general principles of the hypomethylation hypothesis to cognitive improvement rather than cognitive dysfunction. The influence of *DNMT3L* R278G on visuospatial associative memory was greatly exaggerated in individuals with the highest Hcy levels, making the reduction of Hcy by B vitamin treatment in these individuals a key feature of their cognitive improvement. Intake of B vitamins and reduction of high Hcy levels are known to improve the dynamics of the one-carbon cycle, increasing the conversion of Hcy to methionine, abolishing the deleterious effect of Hcy on SAM/SAH ratios, and allowing for more methyl donation to methyltransferases (Kruman et al., 2000). Indeed, there are multiple animal models confirming that supplementation of diet with methyl donors such as folate and vitamin B12 leads to phenotypic changes and associated methylation changes at relevant genes (Anderson et al., 2012). This includes the classic study of the *agouti* mice, where the coat colour of the mice is directly related to hypo- or hypermethylation of the *agouti* gene promoter (Michaud et al., 1994). A number of these studies of B vitamin supplementation and changes in DNA methylation patterns have been succinctly reviewed by Ulrey et al. (2005).

In the case of our finding regarding visuospatial associative memory, it could be that improved one-carbon cycle dynamics allows for differential changes in methylation dependent upon the presence or absence of the *DNMT3L* R278G minor allele. In the adult brain, this could be accomplished by DNMT3L in complex with DNMT3A, the major complex associated with *de novo* methylation patterns (Chedin et al., 2002). However, there is a question mark over the temporal specificity of DNMT3L activity. Resources such as the Allen Brain Atlas present microarray data showing that *DNMT3L* is expressed in the adult brain, but only strongly in specific areas (Appendix 24). In contrast, studies have asserted that DNMT3L is expressed at its highest levels during development, as a regulator of maternal imprinting and in the

establishment of epigenetic patterns during spermatogenesis (Aapola et al., 2002, Webster et al., 2005).

If this were the case, it would be more likely that those with the *DNMT3L* R278G minor allele have differential methylation patterns established during development compared to those with the wild-type genotype, as observed in cases of DNMT3L knockdown in embryonic stem cells (Neri et al., 2013). Moreover, in line with the case of the nearby R271Q variant (El-Maarri et al., 2009), these individuals may be phenotypically healthy during adulthood and through the course of cognitive decline. The administration of B vitamins and the subsequent changes in the one-carbon cycle may then trigger transcriptional changes that were primed by DNMT3L during development, leading to genotype-dependent discrepancies in cognitive function. In support of this, research has shown that such disease-related changes in cognitive performance that are dependent on specific genotypes can occur. For example, *APOE* E4 carriers have been shown to outperform non-carriers on certain cognitive tasks in their youth, before succumbing to poorer performance after the onset of dementia (Filippini et al., 2009). In the case of *APOE* E4, the offset of calorific benefits for cognitive decline gave rise to the theory of the thrifty allele (Corbo and Scacchi, 1999). It is possible that *DNMT3L* R278G provides a similar offset in methylation phenotypes that is so far unknown.

A further unresolved issue is that of the discrepancy between visuospatial associative memory and verbal semantic memory. Owing to the particularly specificity of PAL and GNT to dementia progression, coupled with the observation of poorer visuospatial associative memory in *APOE* E4 carriers, it would appear that a dementia-sensitive cognitive factor has been identified within the VITACOG cohort. This is further corroborated by the addition of the dementia risk factor Hcy into our analysis, which increased the significance of our visuospatial associative memory findings but had no influence on verbal semantic findings. Nevertheless, the observation that *DNMT3L* R278G A/A homozygotes who received B vitamin treatment showed improved verbal semantic memory exemplifies a clear distinction with our visuospatial associative memory findings. This could be explained by our adaptation of the hypomethylation hypothesis, as better performance would be expected in those with wild-type methylation patterns. In this case, differential methylation patterns generated by A/A homozygotes compared to G carriers may manifest in particular areas of the brain, such as the hippocampus for visuospatial associative memory and the frontal cortex for verbal semantic memory. Furthermore,

the discovery of further methylation modifications such as 5hmC, 5fC, and 5caC means that hypomethylation could in fact be demethylation. Thus, these previously unknown levels of transcriptional regulation could also underlie the discrepancy between visuospatial associative memory and verbal semantic memory performance.

3.3.3. Incidental findings, limitations, and future work

In this chapter, a number of less prominent results identified from our baseline variable and allele frequency analyses were presented. One of these was the association with sex in MCI individuals, where *DNMT3L* R278G minor allele carriers were more likely to be female. Given that the number of females was larger owing to a sex disparity in the VITACOG study, the finding could have been influenced by the unequal sample sizes. Nevertheless, the notion that females may be at higher risk has previously been shown regarding prevalence of the *APOE* E4 allele (Payami et al., 1996) and in AD (Vina and Lloret, 2010). Moreover, whilst sex differences in MCI have been less conclusive (Roberts et al., 2012, Su et al., 2014), a comprehensive study of multiple MCI subtypes by Sachdev et al. (2012) revealed a number of factors that influenced the prevalence of MCI in each sex, including cognitive function and *APOE* E4 carrier status.

It is also interesting to note that dimorphic DNA methylation patterns have been reported between males and females (McCarthy et al., 2009). In addition, genetic variants in other methyltransferase genes have been strongly linked to diseases that show greater prevalence in females, such as breast cancer (Cebrian et al., 2006, Sun et al., 2012), illustrating the potential for gender-loaded frequencies in variants associated with DNA methylation genes. Finally, the sex specific roles that *DNMT3L* plays during development had previously been highlighted. The greater number of R278G minor allele carriers in females could point to a further sex specific relationship with *DNMT3L*.

A further unexpected finding was the discrepancy between *DNMT3L* R278G A/A homozygotes and G carriers in baseline CF scores, with A/A homozygotes performing approximately one standard deviation (SD) below the mean and G carriers performing one SD above the mean. Whilst this difference remains difficult to explain, the pervading significance of baseline Hcy levels was again demonstrated. Not only was there a significant difference in baseline CF scores between those with

high and low baseline Hcy, but the interaction of baseline Hcy and *DNMT3L* R278G genotype exacerbated the discrepancy between A/A homozygotes and G carriers.

Given that our subsequent PCA analysis was performed using change (Δ) in cognitive measures, the baseline scores do not hinder our interpretation of the PCA results. Nevertheless, both A/A homozygotes and G carriers showed no significant change in CF performance over the course of the VITACOG trial, substantiating that this cognitive discrepancy was a consistent difference between the two genotype groups. It is noteworthy that previous research has highlighted the confounding question of inflated and false-negative CF results through practice effects, particularly in MCI populations (Cooper et al., 2004), and that greater diagnostic detail regarding MCI subtypes may help to elucidate the discrepant findings discovered in this chapter. Nevertheless, genotype-dependent differences in CF scores have been observed for variants in other DNA methylation genes (Benedict et al., 2011). Whilst this would appear to be coincidental, there remains the possibility that CF performance can explicate subtle cognitive changes associated with differential DNA methylation patterns.

There are some additional issues that should be raised to provide adequate context to our findings. First, the lack of comparative data from the OPTIMA and Challenge cohorts means that, whilst our analysis has uncovered some important results within the VITACOG study of MCI, there is an as-of-yet unfulfilled opportunity to replicate this analysis in the OPTIMA study of AD and the Challenge study of elderly controls. Although they did not receive B vitamins, replication of the genotype and PCA investigation would provide at least some idea about the specificity of these findings to particular disease states.

Our findings rely upon the interactive influence of Hcy levels and B vitamin treatment. Both of these variables can also be held up to a certain amount of scrutiny. As our work pertains specifically to changes in cognitive function, it could be hypothesised that reduction in Hcy levels would be occurring within the brain, particularly as Hcy has neurotoxic effects. However, Hcy levels and Hcy transport within the brain have not been well characterised. It has been confirmed that neuronal cells produce Hcy, especially in response to folate deprivation, but the exact correlation between blood plasma or serum measurements of Hcy and levels of Hcy in the brain is unclear (Ho et al., 2003, Obeid and Herrmann, 2006). Despite this, work has shown that Hcy levels in the cerebrospinal fluid correlate with those in blood serum, although the

concentrations are significantly lower, providing promising evidence that blood Hcy measurements can accurately depict brain Hcy measurements (Obeid et al., 2007).

Our findings reinforce previous reports using the VITACOG study that regular intake of B vitamins has positive effects on Hcy levels and on precise facets of cognitive performance. The VITACOG study carefully chose the amount of vitamin B6 (20mg), vitamin B12 (0.5mg), and folate (0.8mg) to provide to their treatment arm, ensuring that safety procedures were in place and that adverse effects were assessed annually. When assessing the efficacy of this treatment for the public, it would appear that a particular balance has to be struck with regards to the levels of B vitamins prescribed to individuals. The recommended daily allowance is likely to provide very little added benefit, whilst uninformed or uneducated intake of extremely high doses of B vitamins could cause neurotoxicity (Levine and Saltzman, 2004).

In conclusion, this chapter has provided evidence of altered cognitive function associated with a genetic variant within a DNA methyltransferase, as well as with a known genetic risk factor for dementia, *APOE* E4. Whilst the initial investigation into these genetic variants and their impact on cognitive function is promising, this line of enquiry requires additional work to support and validate both the findings and, crucially, the mechanisms that underlie them. The next chapter extends these analyses to include data regarding global brain atrophy and further biochemical measurements (Smith et al., 2010). *In silico* modelling of the *DNMT3L* R278G variant will also be applied in an effort to elucidate the specific impact of this amino acid substitution.

4. Functional investigation of *DNMT3L* R278G utilising brain atrophy data and *in silico* modelling

4.1. Preface

In the VITACOG cohort of individuals with MCI, significant interactions between B vitamin treatment, *DNMT3L* R278G genotype, and performance in visuospatial associative memory and verbal semantic memory were observed. An interaction was discovered between B vitamin treatment and *APOE* E4 for visuospatial associative memory. To provide additional evidence to support these findings, the relationship between genotype, cognitive performance, and brain atrophy data was analysed within the VITACOG cohort. I*n silico* modelling tools were applied to DNA methyltransferase protein models in order to characterise the functional impact of the *DNMT3L* R278G variant.

4.2. Results

4.2.1. Baseline analyses

In addition to variables such as age and sex, a number of key factors outlined in previous published work assessing brain atrophy in the VITACOG cohort were provided to our study (Smith et al., 2010). These included a number of measurements related to the B vitamin treatment provided during the VITACOG trial such as folate, cobalamin (vitamin B12), and creatinine. Diastolic blood pressure (BP) was included because hypertension is a reported risk factor for dementia from middle age onwards and therefore may manifest at the MCI stage (Kennelly et al., 2009).

Baseline total brain volume was examined to ensure that any differences in brain volume between individuals present at the beginning of the study are accounted for. Baseline tHcy was also examined because of its influence on cognitive performance in the VITACOG cohort, as detailed in the previous chapter. Finally, the Geriatric Depression Scale (GDS) was included as research has linked depressive symptoms with cognitive impairments (Burt et al., 1995), and performance on the GDS has also been implicated in predicting the conversion from MCI to AD (Modrego and Ferrandez, 2004).

First correlations between the baseline scores in these key variables and yearly rate of brain atrophy (ROA) were investigated in the placebo group alone. This revealed that age, tHcy, and creatinine at baseline all correlated positively with ROA, whilst brain volume and diastolic BP at baseline correlated negatively with ROA. Subsequently, correlations were investigated in the treatment group alone and found only a positive correlation between creatinine levels and ROA remained. This is despite the observation that, in the cohort as a whole, age, tHcy, brain creatinine, and brain volume were all significantly correlated with ROA. Correlations between baseline measurements of the aforementioned variables and ROA are presented in Table 4.1.

The absence of association between key baseline variables and ROA in the treatment group suggested that this was an effect of the B vitamin treatment. In other words, the risk associated with these significant correlations between key covariates and ROA has been mitigated in the B vitamin treatment group. To investigate this further, the baseline variables that were significantly correlated with ROA in the whole VITACOG cohort were examined more closely in order to see if they presented any notable differences between groups. The results are presented in Table 4.2.

From these results, there was no difference in baseline brain volume between the B vitamin treatment and placebo groups, between males and females, and between *DNMT3L* R278G A/A homozygotes and G carriers. No difference was observed between *APOE* E4 carriers and non-carriers in the whole VITACOG cohort. However, within the placebo group, *APOE* E4 carriers showed significantly lower baseline brain volume compared to non-carriers ($t = -2.076$, $p = 0.041$). Those with high baseline tHcy showed significantly smaller brain volume compared to those with low baseline tHcy. No difference in baseline tHcy was observed between the B vitamin treatment and placebo groups, between *DNMT3L* R278G A/A homozygotes and G carriers, or between *APOE* E4 carriers and non-carriers. Males recorded significantly higher baseline tHcy levels than females.

TABLE 4.1. Correlations between baseline scores in key variables and ROA

Variables	ROA (placebo)	ROA (treatment)	ROA (overall)
Age	R = 0.317, *p* = 0.004*	R = 0.140, *p* = 0.201	R = 0.200, *p* = 0.009*
Brain volume	R = -0.319, *p* = 0.003*	R = -0.116, *p* = 0.292	R = -0.217, *p* = 0.005*
Log tHcy	R = 0.492, *p* < 0.001**	R = -0.037, *p* = 0.735	R = 0.225, *p* = 0.003*
Log folate	R = -0.062, *p* = 0.580	R = 0.073, *p* = 0.508	R = 0.023, *p* = 0.764
Log cobalamin	R = -0.052, *p* = 0.644	R = -0.058, *p* = 0.599	R = -0.051, *p* = 0.514
Log creatinine	R = 0.258, *p* = 0.019*	R = 0.239, *p* = 0.028*	R = 0.246, *p* = 0.001*
Diastolic BP	R = -0.310, *p* = 0.004*	R = 0.095, *p* = 0.388	R = -0.108, *p* = 0.163
Log GDS	R = 0.150, *p* = 0.175	R = 0.001, *p* = 0.996	R = 0.122, *p* = 0.117

ROA = Yearly rate of brain atrophy; BP = Blood pressure; tHcy = Total homocysteine; GDS = Geriatric depression scale
** = significant at the 0.01 level, * = significant at the 0.05 level

TABLE 4.2. Results of demographic analyses for significant variables

Variables	Treatment	Sex	Baseline tHcy	*DNMT3L* R278G	*APOE*
Brain volume	t = 0.978, *p* = 0.330	t = -1.842, *p* = 0.067	t = 2.135, *p* = 0.034*	t = -1.018, *p* = 0.310	t = -1.849, *p* = 0.067
Log tHcy	t = -0.545, *p* = 0.586	t = -2.456, *p* = 0.013*	-	t = 0.827, *p* = 0.400	t = -0.418, *p* = 0.677
Log creatinine	t = -0.885, *p* = 0.377	t = -8.964, *p* < 0.001**	t = -6.528, *p* < 0.001**	t = 0.711, *p* = 0.481	t = -0.922, *p* = 0.358
Age	t = 0.295, *p* = 0.769	t = 0.214, *p* = 0.831	t = -4.199, *p* < 0.001**	t = -0.188, *p* = 0.851	t = -0.521, *p* = 0.603

tHcy = Total homocysteine

** = significant at the 0.01 level, * = significant at the 0.05 level

No difference in baseline creatinine was observed between the B vitamin treatment and placebo groups, between *DNMT3L* R278G A/A homozygotes and G carriers, and between *APOE* E4 carriers and non-carriers. Again, males recorded significantly higher baseline creatinine levels than females. In addition, those with high baseline tHcy showed significantly higher levels of baseline creatinine than those with low baseline tHcy. No difference in age was observed between the B vitamin treatment and placebo groups, between males and females, between *DNMT3L* R278G A/A homozygotes and G carriers, and between *APOE* E4 carriers and non-carriers. Those with high baseline tHcy were significantly older than those with low baseline tHcy.

Once again, these results emphasise the modulating role of Hcy, with high levels of baseline tHcy recurring as a risk factor for the other significant variables in this study. Interestingly, there were no differences at baseline between the B vitamin treatment and placebo groups for any of these key covariates. This supports the idea that the absence of association between these key covariates and ROA in the treatment group was an effect of the treatment on ROA over the course of the VITACOG trial, rather than culminating from differences between the groups at baseline. Nevertheless, there were crucial differences between the treatment groups relating to these baseline covariates. For instance, both brain volume and tHcy showed respective positive ($R = 0.410$, $p < 0.001$) and negative ($R = -0.240$, $p = 0.014$) correlations with visuospatial associative memory performance in the placebo group, but no such correlation in the treatment group.

4.2.1.1. Geriatric Depression Scale

The GDS is a self-completion questionnaire where low scores indicate a normal response, mid-range scores indicate a mild depressive response, and high scores indicate a severe depressive response (Yesavage et al., 1982). Although baseline GDS score was not correlated with ROA (Table 4.1), the potential link between depression and cognitive impairment warranted further investigation. Therefore, I looked to see if either baseline GDS scores or ΔGDS scores over the course of the trial presented any notable group differences in the VITACOG cohort. The results are summarised in Table 4.3.

TABLE 4.3. Results of demographic analyses for GDS

	Treatment	**Sex**	**Baseline tHcy**	***DNMT3L* R278G**	***APOE***
Baseline log GDS	t = -2.470, *p* = 0.014*	t = 1.197, *p* = 0.232	t = 1.101, *p* = 0.260	t = 1.023, *p* = 0.307	t = 0.180, *p* = 0.857
∆GDS	t = -0.106, *p* = 0.916	t = 0.488, *p* = 0.626	t = -0.998, *p* = 0.319	t = -1.709, *p* = 0.088	t = 0.188, *p* = 0.851

* = significant at the 0.05 level

There was no difference in baseline GDS score between males and females, between high and low baseline tHcy groups, between *DNMT3L* R278G A/A homozygotes and G carriers, and between *APOE* E4 carriers and non-carriers. Individuals in the treatment group showed significantly lower baseline GDS score in comparison to those in the placebo group. Despite this, no difference in ∆GDS was observed between the treatment groups. In addition, no correlation was observed between baseline GDS score and visuospatial associative memory (R = 0.007, *p* = 0.920) or verbal semantic memory (R = -0.033, *p* = 0.638) performance. No difference in ∆GDS was observed for sex, baseline tHcy, *DNMT3L* R278G, and *APOE*. There was, however, an effect of baseline tHcy in the placebo group, where those with high baseline tHcy levels showed an increase in GDS score over the course of the trial whilst those with low baseline tHcy showed a decrease in GDS score. This supports a clear body of work associating hyperhomocysteinemia with depressive symptoms (Bottiglieri, 2005).

Analysis of baseline values in key variables with respect to ROA has uncovered many significant relationships in the placebo group that are absent in the treatment group, pointing to an influence of the B vitamin treatment. The significant influence of baseline tHcy levels as a key modulating variable was again corroborated, where those with high levels of Hcy are at risk of smaller brain volume, higher creatinine levels, and a higher score on the GDS.

4.2.1.2. Rate of brain atrophy

Having explored the key baseline measurements, our analysis progressed to examine yearly whole brain ROA. Using the data from VITACOG, the findings of Douaud et al. (2013) were corroborated in that there was a significant difference in ROA between the B vitamin treatment and placebo groups, and that this difference was exacerbated in those with high levels of baseline tHcy.

4.2.2. ROA and cognition

In the previous chapter, it was demonstrated that a pharmacogenetic interaction of *DNMT3L* R278G and B vitamin treatment which was associated with respective improvement and decline in visuospatial associative memory and verbal semantic memory. A similar pattern was also observed for *APOE* E4 carriers with respect to visuospatial associative memory. Moreover, it was discovered that, along with B

vitamin treatment, baseline tHcy levels could have a significant moderating influence on cognitive performance.

In order to assess whether the improvement and decline in cognitive performance was reflected in reduced and accelerated ROA in the VITACOG cohort, the relationship between the *DNMT3L* R278G or *APOE* genotypes and cognitive performance with respect to ROA was investigated. In addition, the key variables identified in the whole cohort (age, baseline brain volume, baseline tHcy, and baseline creatinine) along with B vitamin treatment group were included in the statistical models. The modelled ROA was then used to examine the genotype-dependent relationships between brain atrophy and cognitive performance in the VITACOG cohort.

4.2.2.1. Influence of *DNMT3L* R278G on ROA and cognition

To begin with, the influence of the *DNMT3L* R278G variant on the relationship between cognitive performance and ROA prior to the modelling of key variables was assessed. A significant negative correlation between visuospatial associative memory performance and ROA was observed for G carriers ($R^2 = 0.276$, $p < 0.001$), but was absent in A/A homozygotes ($R^2 = 0.029$, $p = 0.116$). Conversely, a significant negative correlation between verbal semantic memory performance and ROA was observed for A/A homozygotes ($R^2 = 0.057$, $p = 0.026$), but was non-significant in G carriers ($R^2 = 0.044$, $p = 0.081$).

When modelling ROA, it was found that adjustment for covariates strengthened the negative correlation between ROA and visuospatial associative memory performance for G carriers ($R^2 = 0.420$, $p < 0.001$), whilst the non-significant correlation for A/A homozygotes became weaker ($R^2 = 0.015$, $p = 0.264$) (Figure 4.1A). Similarly, covariate adjustment strengthened the negative correlation between ROA and verbal semantic memory performance for A/A homozygotes ($R^2 = 0.294$, $p < 0.001$), whilst the non-significant correlation for G carriers also became weaker ($R^2 = 0.003$, $p = 0.652$) (Figure 4.1B).

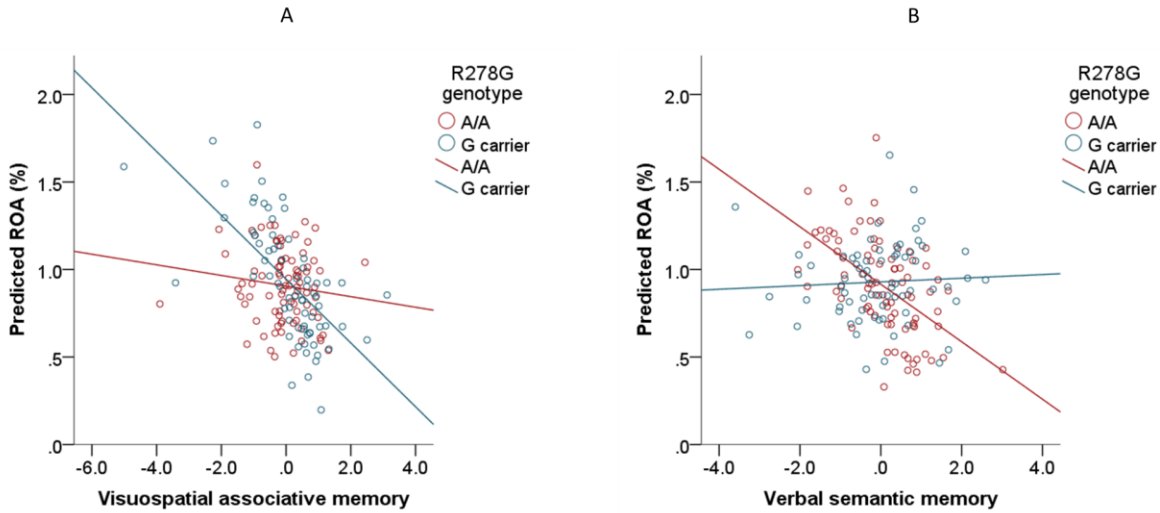


FIGURE 4.1. Influence of *DNMT3L* R278G on the relationship between cognitive performance and ROA. The predicted ROA output is adjusted for the influence of key covariates (age, baseline brain volume, baseline tHcy, baseline creatinine). Data presented for visuospatial associative memory (A) and verbal semantic memory (B). Best of fit lines included for *DNMT3L* R278G A/A homozygotes (red) and G carriers (blue).

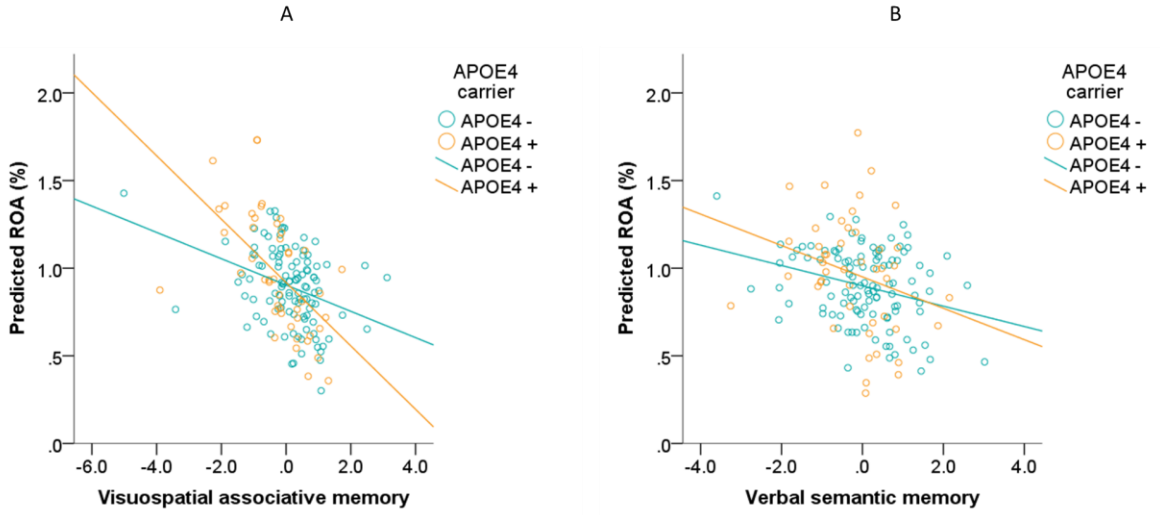


FIGURE 4.2. Influence of *APOE* E4 on the relationship between cognitive performance and ROA. The predicted ROA output is adjusted for the influence of key covariates (age, baseline brain volume, baseline tHcy, baseline creatinine). Data presented for visuospatial associative memory (A) and verbal semantic memory (B). Best of fit lines included for *APOE* E4 carriers (orange) and non-carriers (blue).

4.2.2.2. Influence of *APOE* on ROA and cognition

As with *DNMT3L* R278G, the effect of the *APOE* E4 genotype on the relationship between cognitive performance and ROA prior to the modelling of key variables was assessed. A significant negative correlation between visuospatial associative memory performance and ROA was observed for both *APOE* E4 carriers ($R^2 = 0.108$, $p = 0.026$) and non-carriers ($R^2 = 0.130$, $p < 0.001$). Further, a significant negative correlation between verbal semantic memory performance and ROA was observed for *APOE* E4 non-carriers ($R^2 = 0.052$, $p = 0.016$), but was absent in carriers ($R^2 = 0.038$, $p = 0.193$).

When modelling ROA, it was found that adjustment for covariates strengthened the negative correlation between ROA and visuospatial associative memory performance for *APOE* E4 carriers ($R^2 = 0.324$, $p < 0.001$), whilst the correlation for non-carriers became weaker ($R^2 = 0.120$, $p < 0.001$) (Figure 4.2A). Similarly, covariate adjustment strengthened the negative correlation between ROA and verbal semantic memory performance for *APOE* E4 non-carriers ($R^2 = 0.086$, $p = 0.002$), whilst the correlation for carriers remained non-significant ($R^2 = 0.072$, $p = 0.072$) (Figure 4.2B).

4.2.3. *In silico* protein modelling

A clear relationship was observed between cognitive performance and ROA that was discrepant for visuospatial associative or verbal semantic memory based upon the *DNMT3L* R278G genotype. To improve our understanding of the possible functional impact of this genetic variant, a number of *in silico* tools were utilised to specifically model features related to protein structure and protein surface dynamics. Further, the reliability of this modelling approach was tested by including other relevant DNA methyltransferase variants to act as comparative positive and negative controls, presented previously in Table 2.12. *DNMT3A* R749C and *DNMT1* Y495C are both clinically associated variants and, along with *DNMT3L* R271Q, are the positive controls. *DNMT3A* R749C has been identified as causing a syndromic form of intellectual disability (Tatton-Brown et al., 2014) and, like *DNMT3L* R278G, is located close to the DNMT3A-3L interaction sites. *DNMT1* Y495C has been reported to cause a rare form of neurodegeneration named Hereditary Sensory and Autonomic Neuropathy type 1 (HSAN1) (Klein et al., 2011) and falls within the DNA replication foci-targeting sequence of DNMT1. *DNMT3L* R271Q is not clinically associated but

has been reported to result in significant hypomethylation events in young people (El-Maarri et al., 2009) and is located close to *DNMT3L* R278G. Finally, *DNMT3L* H313Y was used as the negative control as it has no known clinical importance.

To investigate each of these genetic variants, Protein Data Bank (PDB) models were acquired for each respective methyltransferase protein, outlined previously in Table 2.13. Full length PDB models for DNMT3L, DNMT3A, and DNMT1 were produced from their amino acid sequences using the RaptorX tool. In addition, previously created models were obtained from the PDB online repository. One of the main functions of DNMT3L is to form a complex with DNMT3A to stabilise and improve the efficiency of DNA methylation deposition. A full model of the DNMT3A-3L complex was utilised, a model of the DNMT3A-3L C terminal domain which contains the DNA recognition domain and active site loops required for methylation deposition, and a model of DNMT3A-3L in complex with histone H3. DNMT3A-3L exists in an autoinhibitory complex, but histone H3 stimulates the complex to initiate methyltransferase activity (Guo et al., 2015), making this a key complex to study. In addition, two models of the N terminal domain of DNMT1 which is responsible for autoinhibitory and allosteric regulation of the protein (Zhang et al., 2015) were utilised. One model contains the majority of the N terminal domain and one contains only the replication foci targeting sequence.

4.2.3.1. Protein structure

Alterations in secondary structure were assessed using PyMOL. The *DNMT3L* R278G change resulted in the disruption of hydrogen bonds adjacent to one of the DNMT3A-3L interaction sites (Figure 4.3A). This disruption of hydrogen bonds was also seen in the clinically associated variants *DNMT3A* R749C (Figure 4.3D) and *DNMT1* Y495C (Figure 4.3E), supporting the idea that disruption in protein structure may be clinically important and associated with key changes in protein behaviour. In contrast, *DNMT3L* R271Q showed no such structural change (Figure 4.3B) whilst the negative control variant *DNMT3L* H313Y resulted in the creation of new hydrogen bonds (Figure 4.3C), potentially conferring improved protein stability.

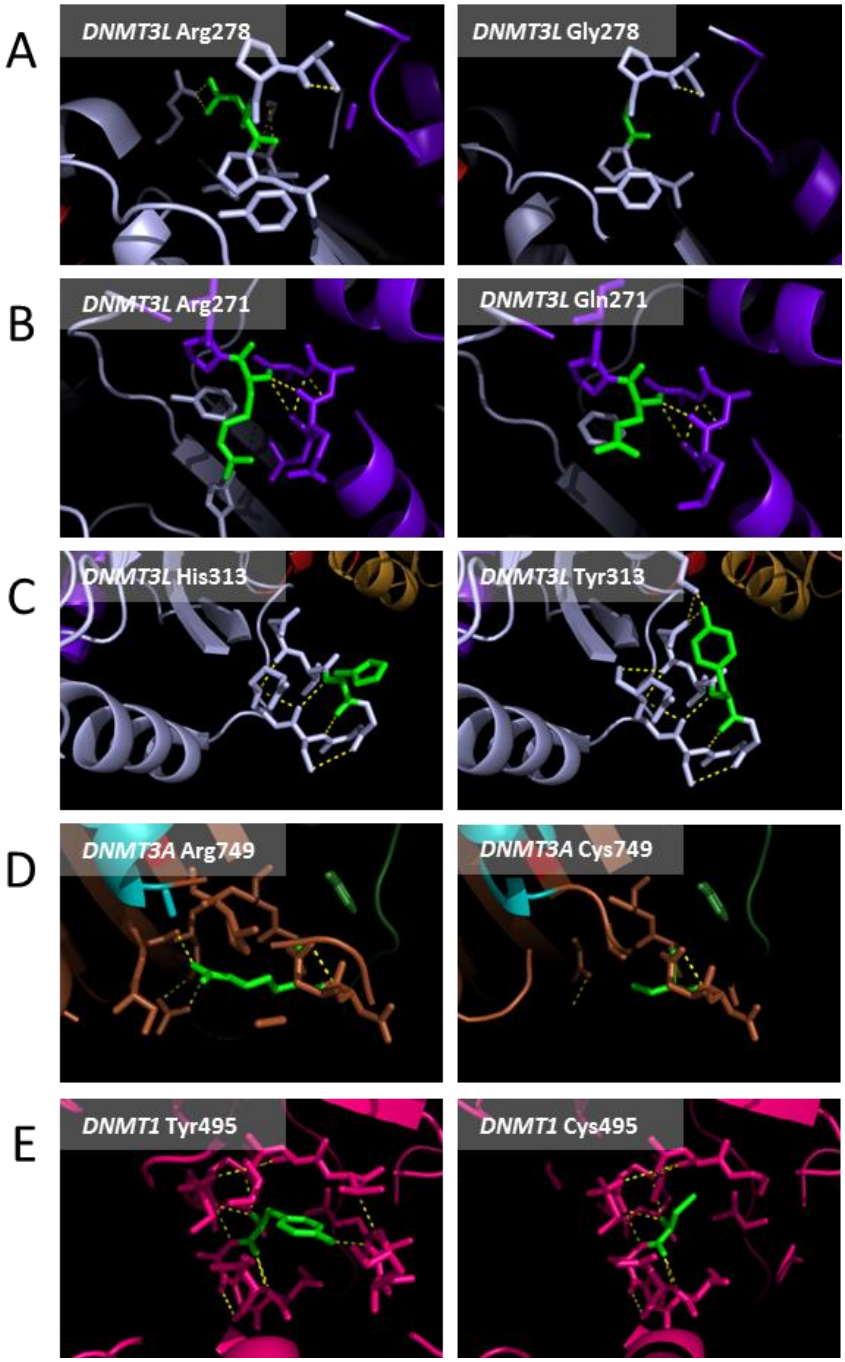


FIGURE 4.3. *In silico* comparison of wild-type and variant methyltransferase models. PyMOL mutagenesis was used to create the wild-type and variant models, where green residues indicate the variant residues and yellow dashed lines indicate hydrogen bonds. Models presented are *DNMT3L* R278G (A), the nearby *DNMT3L* R271Q (B), a negative control *DNMT* variant (C), and two clinically associated *DNMT* variants (D, E).

FoldX was subsequently implemented to provide a quantitative element to the protein structure modelling by estimating changes in free energy (ΔG) related to the variant amino acid substitutions, where a lower ΔG value indicates a more stable protein. Change in ΔG (ΔΔG) was calculated by comparing WT models with the mutated variant models. Both clinically associated variants presented destabilising increases in ΔG, with *DNMT1* Y495C highly destabilising across all protein models and *DNMT3A* R749C highly destabilising in both the auto-inhibitory and active DNMT3A-3L-H3 complex models. *DNMT3L* R278G and R271Q resulted in slight to moderate destabilising ΔΔG in the native models and auto-inhibitory DNMT3A-3L complex models, but were highly destabilising in the active DNMT3A-3L-H3 complex model. The destabilising ΔΔG in the active complex model further suggests that the *DNMT3L* R278G variant has a functional impact akin to those with clinical importance or those which result in changes to DNA methylation patterns. In contrast, the negative control variant *DNMT3L* H313Y showed no ΔΔG in the native models or active DNMT3A-3L-H3 complex models. However, it also presented a highly destabilising ΔΔG in the auto-inhibitory DNMT3A-3L complex models, perhaps illustrating that the H313Y variant is not functionally benign despite this behaviour not being of clinical importance. The results are presented in Figure 4.4.

4.2.3.2. Protein surface

Using PyMOL, changes in elementary structure and surface area were assessed for the regions immediately surrounding the variants of interest. Modelling revealed that the *DNMT3L* R278G variant resulted in a loss of surface area more substantial than all of the other methyltransferase variants included in this analysis, as depicted by the lack of structure (yellow) for the Gly278 variant found adjacent to a DNMT3A-3L interaction site (purple) in Figure 4.5A. The clinically associated *DNMT3A* R749C (Figure 4.5D) and *DNMT1* Y495C (Figure 4.5E) variants both resulted in similar reductions in surface area, whilst a smaller reduction in surface area was observed for *DNMT3L* R271Q (Figure 4.5B). In comparison, the negative control variant *DNMT3L* H313Y resulted in the only observed increase in surface area, consistent with the results from the protein structure analysis (Figure 4.5C). The surface area calculations of whole protein surface area and the 6Å area around the variant of interest are presented in Table 4.4.

		Average ΔΔG scores (runs = 3)					
Gene	Variant	DNMT3L	3A-3L	3A-3L C t. (x1)	3A-3L C t. (x2)	3A-3L-H3 (x1)	3A-3L-H3 (x2)
DNMT3L	R278G	1.057	0.902	0.576	0.527	1.627	3.729
DNMT3L	R271Q	0.635	0.598	0.414	1.18	1.79	3.242
DNMT3L	H313Y	-0.156	-0.328	1.956	3.897	0.44	-0.035

Gene	Variant	DNMT3A	3A-3L	3A-3L C t. (x1)	3A-3L C t. (x2)	3A-3L-H3 (x1)	3A-3L-H3 (x2)
DNMT3A	R749C	0.844	1.571	1.913	3.349	1.13	2.671

Gene	Variant	DNMT1	N t. (x1)	N t. (x2)	RFTS (x1)	RFTS (x2)
DNMT1	Y495C	4.166	5.136	9.771	4.569	8.82

= highly stabilising
= stabilising
= slightly stabilising
= neutral
= slightly destabilising
= destabilising
= highly destabilising

FIGURE 4.4. Comparison of change in free energy quantifications between wild-type and variant methyltransferase models. x1 refers to models with 1 WT and 1 variant model in a dimer complex, x2 refers to models with 2 variant models in a dimer complex, 3A-3L refers to the DNMT3A-3L complex, C t. and N t. refer to the C and N terminal domains, H3 refers to histone H3, and RFTS refers to the replication foci targeting sequence. All numbers indicate the change in free energy from wild-type to variant. Models included are *DNMT3L* R278G, the nearby *DNMT3L* R271Q, a negative control *DNMT3L* H313Y variant, and two clinically associated variants in *DNMT3A* and *DNMT1*.

TABLE 4.4. Surface area changes associated with methyltransferase variants

Gene	Variant	Δ whole area (Å^2)	Δ 6Å area (Å^2)
	R278G	-92.055	-93.402
DNMT3L	R271Q	-20.024	-24.968
	H313Y	24.265	24.651
DNMT3A	R749C	-58.168	-59.649
DNMT1	Y495C	-53.505	-55.121

Å = Angstrom

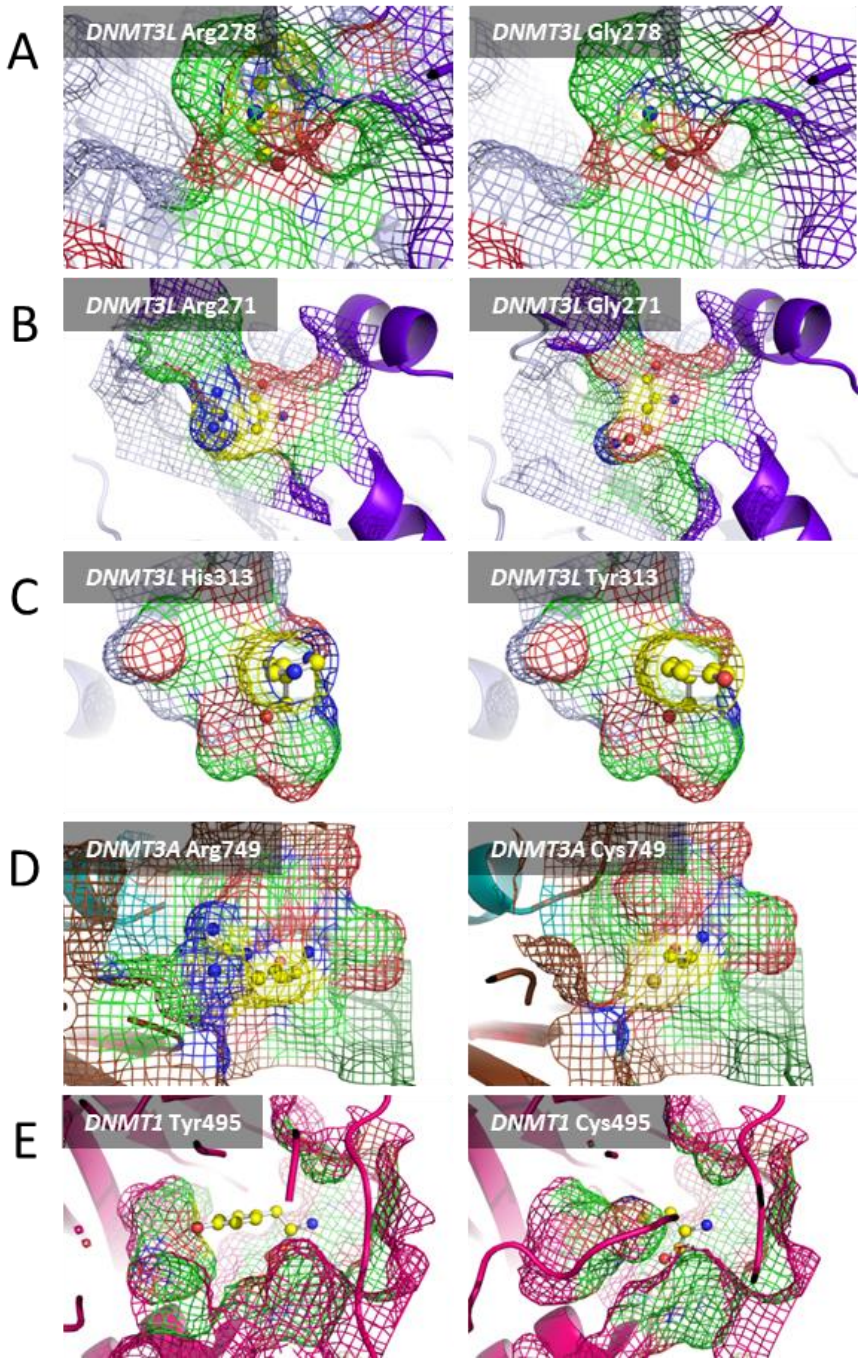


FIGURE 4.5. *In silico* surface area assessment of wild-type and variant methyltransferase models. Yellow areas indicate variant residues, green refers to carbon element areas, grey refers to hydrogen element areas, blue refers to nitrogen element areas, and red refers to oxygen element areas. Models presented are *DNMT3L* R278G (A), the nearby *DNMT3L* R271Q (B), a negative control *DNMT* variant (C), and two clinically associated *DNMT* variants (D, E).

The Adaptive Poisson-Boltzmann Solver (ABPS) was then utilised to investigate whether the observed changes to protein surface associated with these variants were having a functional impact on protein dynamics crucial to behaviours such as protein-protein binding. The *DNMT3L* R278G variant resulted in a distinct transition from positive to negative electrostatic surface potential stretching over the DNMT3A-3L interaction sites (Figure 4.6A). Whilst far less pronounced, the *DNMT3L* R271Q variant also resulted in greater negative electrostatic surface potential around the DNMT3A-3L interaction sites (Figure 4.6B), suggesting that changes in surface dynamics around the DNMT3A-3L binding regions may influence the establishment of DNA methylation patterns. Similar transitions from positive to negative electrostatic surface potential were also observed for the clinically associated variants, *DNMT3A* R749C (Figure 4.6D) and *DNMT1* Y495C (Figure 4.6E). In contrast, no observable change in electrostatic surface potential was found for the *DNMT3L* H313Y negative control variant (Figure 4.6C).

Taken together, our *in silico* analyses suggest that both structural and electrostatic perturbations are putatively caused by the *DNMT3A* R278G variant. Moreover, the proximity of these disruptions to the DNMT3A-3L interaction sites suggests a potential impact on this protein complex. As the nearby *DNMT3L* R271Q variant has been associated with changes in global methylation patterns in healthy individuals, one would predict that the R278G variant could have a similar effect on the DNA methylation landscape.

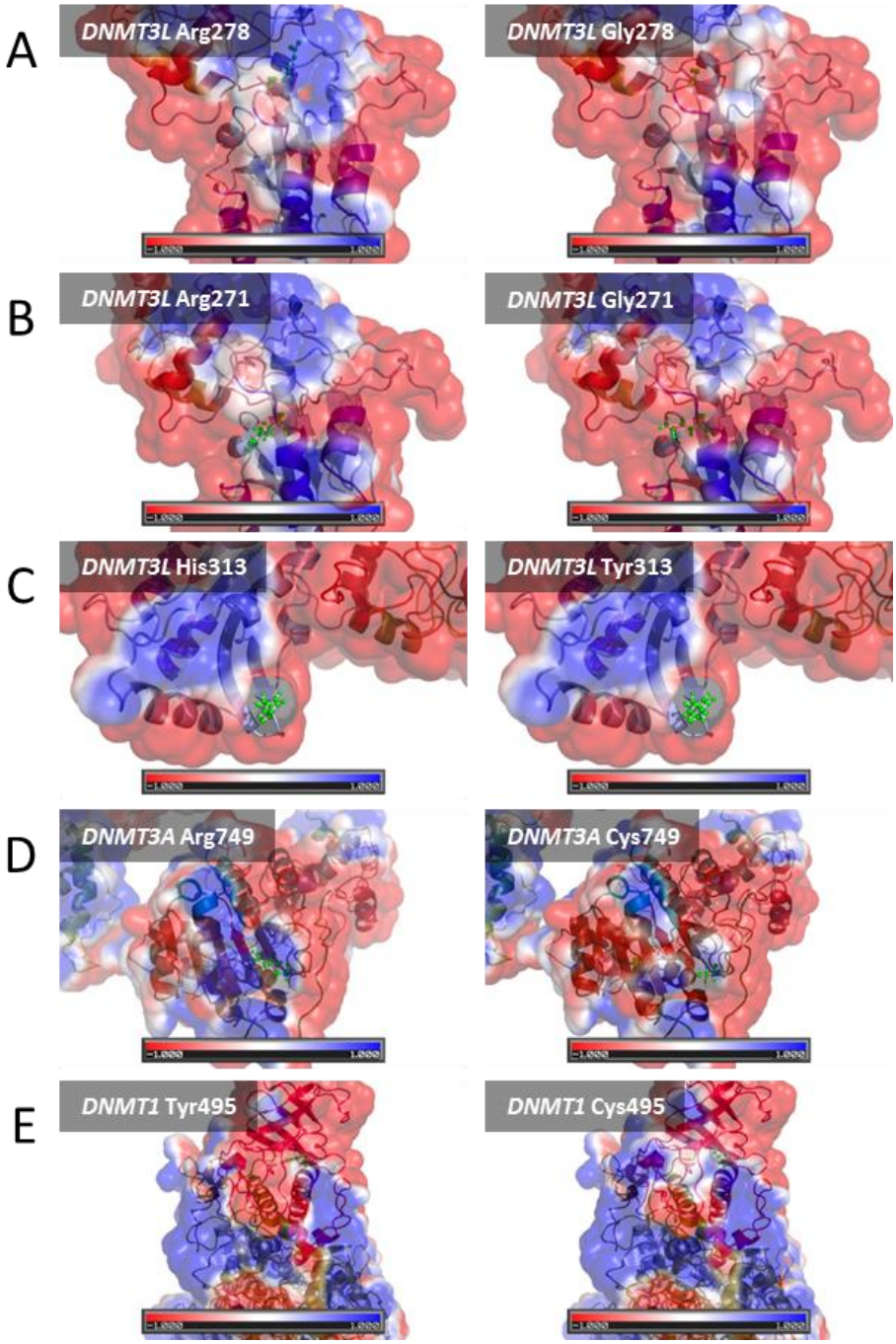


FIGURE 4.6. *In silico* electrostatic surface potential estimations for wild-type and variant methyltransferase models. APBS was used to estimate the surface potential, where green ball and stick figures denote variant residues, blue areas refer to positive electrostatic potential, and red areas refer to negative electrostatic potential. Models presented are *DNMT3L* R278G (A), the nearby *DNMT3L* R271Q (B), a negative control *DNMT* variant (C), and two clinically associated *DNMT* variants (D, E).

4.3. Discussion

In chapter three, a relationship between B vitamin treatment, homocysteine levels, and the *DNMT3L* R278G and *APOE* E4 genotypes was demonstrated with respect to particular facets of cognitive performance. The focus of this chapter was to further investigate these relationships using brain atrophy data from the VITACOG cohort. In addition, one aim was to better understand the functional impact of the *DNMT3L* R278G variant and subsequently provide evidence for altered DNA methylation associated with this variant.

Following assessment of brain atrophy within the VITACOG cohort, the previously observed relationships between cognitive performance and our genotypes of interest, *DNMT3L* R278G and *APOE* E4, were corroborated. A significant negative correlation was discovered between visuospatial associative memory and ROA that was present in *DNMT3L* R278G G carriers but absent in A/A homozygotes. Conversely, a significant negative correlation was discovered between verbal semantic memory and ROA that was present in *DNMT3L* R278G A/A homozygotes but absent in G carriers. A significant negative correlation was also discovered between visuospatial associative memory and ROA that was present in *APOE* E4 carriers but absent in non-carriers.

Owing to the contrasting findings regarding *DNMT3L* R278G and the comparative lack of literature surrounding the function of DNMT3L compared to APOE, the influence of the R278G variant was assessed using *in silico* modelling tools. Using models for multiple methyltransferase proteins, structural changes associated with the *DNMT3L* R278G variant were observed that were also seen in clinically associated methyltransferase variants. This resulted in moderate destabilising effects in the DNMT3A-3L active complex, changes that were not observed for the benign control variant. Due to the role of DNMT3L in forming a complex with DNMT3A, protein surface dynamics were also examined and found that the R278G negatively influenced electrostatic potential at the interaction sites between DNMT3A and DNMT3L. This provides some evidence for a disruption of the DNMT3A-3L complex associated with the R278G variant that may result in altered deposition of DNA methylation patterns.

4.3.1. Atrophy data and *in silico* techniques

As detailed earlier, access was granted to rate of whole brain atrophy data generated from the duration of the VITACOG study that was then processed to create a yearly percentage rate of whole brain atrophy. It is perhaps surprising that conclusive genotype-dependent relationships between ROA and cognitive performance were observed using data from the whole brain. This is because published research examining ROA in the VITACOG cohort had specifically measured regions of grey matter pertinent to the progression of dementia, such as medial temporal areas including the hippocampus, entorhinal cortex, and parahippocampal gyrus (Douaud et al., 2013).

Previous studies had led one to hypothesise that visuospatial associative memory changes in *DNMT3L* R278G G carriers would manifest in these medial temporal areas, whilst changes in verbal semantic memory in A/A homozygotes would involve both frontal and medial temporal areas associated with semantic processing. Such changes are commonly observed in the frontotemporal dementia subtype called semantic dementia (Hodges et al., 1999). For example, work using individuals from the OPTIMA cohort found that visuospatial associative deficits were most strongly associated with atrophy in the medial temporal lobes (Zamboni et al., 2013). Similar findings using the PAL test have been observed in individuals with MCI concerning specific hippocampal atrophy (de Rover et al., 2011). In comparison, application of a battery of semantic memory tests in MCI individuals showed that impairments were correlated with the medial temporal regions (McDonald et al., 2012). Studies which included individuals with MCI and AD showed that particular semantic memory deficits were associated with atrophy of the anterior temporal lobes and the inferior prefrontal cortex, regions that are both linked to semantic cognition (Joubert et al., 2010).

Nevertheless, it would appear that brain atrophy changes associated with both cognitive factors are likely to be more complicated and certainly a matter for debate. Previous work assessing ROA in the VITACOG cohort showed that poorer performance in the HVLT-R and CF tests was associated with greater grey matter loss in the left hippocampus and entorhinal cortex, but with no mention of the frontal cortex (Douaud et al., 2013). In contrast, poorer performance in the MMSE test was associated with grey matter loss bilaterally in the amygdalohippocampal complex and entorhinal complex, whilst the PAL and GNT tests were not assessed. These findings

suggest that the medial temporal lobe areas are most relevant to both visuospatial associative and verbal semantic memory performance, and that the distinction between both of these factors may not be linked to interactions with other brain regions but could be related to more minute alterations within the medial temporal areas. For example, studies of semantic memory in MCI and AD have consistently highlighted the perirhinal and entorhinal cortices (Barbeau et al., 2012, Davies et al., 2004, Du et al., 2004), whilst visuospatial associative memory ability has been more robustly associated with the hippocampus alone (Golomb et al., 1994, Johnson et al., 2008, Longoni et al., 2015). However, this pattern is by no means definitive as other brain regions have been implicated in both forms of memory (Blacker et al., 2006, Gardini et al., 2011). This underlines the complexity in understanding the brain regions relevant to each cognitive factor, and highlights the need for deeper analysis of these brain regions in this cohort using our cognitive factors.

The second arm of this chapter was the functional analysis of the *DNMT3L* R278G variant using *in silico* modelling tools. Short of carrying out molecular assays of DNMT3L function using *in vitro* systems, these *in silico* approaches offer an opportunity to interrogate the function of DNMT3L using well-characterised tools. An advantage of the *in silico* approach utilised in this chapter was the availability of multiple PDB files, both within each protein and between proteins to provide greater depth to our comparative analysis. Our use of clinically associated variants in other methyltransferases, particular DNMT3A due to its interaction with DNMT3L, added a level of credence to our interpretations that would be difficult to assert otherwise.

Whilst the breadth of PDB models available to our analysis provides some advantages, it also raises questions over the reproducibility of our findings. Where possible, both publically available (created during other studies and uploaded to the PDB database) and in-house (created from the FASTA sequence using online prediction tools) PDB files were incorporated for the native protein models. However, it was observed that differences can occur between these publically available and in-house files. For instance, the level of hydrogen bond disruption associated with the *DNMT3L* R278G was greater in the publically available model compared to the in-house model. In this case, the result from the publically available model was chosen as it has been used in previous studies. However, it is possible that this model contains specific features (solvent accessibility, binding sites, contact map estimations) that differ from the templates used in our in-house models (Peng and Xu, 2011).

There are certain properties within the *in silico* modelling tools utilised in this chapter that were generally kept at default settings. However, more could be done to improve the accuracy of our modelling by amending these default properties. Hydrogen bond estimations within PyMOL are calculated using the standard Dictionary of Secondary Structure of Proteins (DSSP) algorithm and cannot be altered within this tool, although adapted methods have been presented that appear to provide more accurate estimations of hydrogen bonds (Andersen et al., 2002). To quantify the effects seen in PyMOL, FoldX was utilised to estimate changes in free energy across the protein. FoldX was initially used to remedy any issues within our PDB models, providing a standardised starting point before assessing changes in free energy. However, as explained in the methods, the default options were applied for the standardising process and these options use information from the PDB file itself, such as the estimation of water and metal bonds. As the lab-made and publically available PDB files can contain different features, it is possible that these differences could influence the FoldX standardisation process. Rather than overlooking these default options, more expertise may be required to alter these options and ensure that the standardisation process has a uniform effect on all of the PDB models utilised.

FoldX estimates changes in free energy associated with the stability or instability of genetic mutations. The program was designed and relies upon empirical data regarding the stability of particular mutations and has been independently verified on 548 mutations in 22 genes (Tokuriki et al., 2008). In comparison to FoldX, there are more sophisticated molecular dynamics tools such as NAMD that can provide complex simulations of the free energy changes associated with a genetic variant and animate precise features of protein stability. In support of our analysis, the predictive accuracy of FoldX has been exemplified in multiple studies of genetic mutations (Rallapalli et al., 2014, Studer et al., 2014b). Moreover, the contrasting optimisation of dimer models containing either one or two instances of a genetic mutation of interest has also been used to good effect (Studer et al., 2014a).

With regards to protein surface dynamics, the Adaptive Poisson-Boltzmann Solver (APBS) was applied to estimate electrostatic surface potential measurements for our PDB files. Again, there are features within the application of APBS that can be scrutinised. To begin with, the process of generating electrostatic information for a PDB model requires the use of the PDB2PQR program. Whilst this tool is increasingly robust, the authors admit that inconsistencies in the quality of PDB files

can hinder the accurate estimation of electrostatics (Dolinsky et al., 2007). The application of the Poisson-Boltzmann equation (PBE) to estimate electrostatics in biomolecular systems has also been dissected, with suggestions that the PBE applies well to molecules with low charge density or in simple salt solutions but not to molecules with high charge density or in multivalent solutions (Baker, 2004). Finally, there is a body of research dedicated to the assessment of force fields used within such experiments. The PARSE (parameters for solvation energy) default force field that treats the solution as a single medium was applied (Chung et al., 2010). However, multiple force fields exist that are tailored towards specific features of protein secondary structure, providing further opportunity to improve the accuracy of our electrostatic estimations (Beauchamp et al., 2012, Cino et al., 2012).

4.3.2. Mechanisms of action

In the previous chapter, the biological mechanisms that could explain how B vitamin treatment, high Hcy levels, and the *DNMT3L* R278G genotype in particular resulted in distinct changes in cognitive performance were discussed. Key processes were outlined that are likely to be involved in the observed cognitive changes, including the reduction of Hcy neurotoxicity in the brain and improved dynamics within the one-carbon cycle that could allow for genotype-dependent methylation changes to be actualised.

The relationship between the *DNMT3L* R278G genotype and cognitive performance has been substantiated by outlining associated changes in ROA. Taking into account the importance of B vitamins and Hcy levels, there are a number of possible mechanisms that could explain the observed ROA changes. The neurotoxic effects of Hcy within the brain have previously been highlighted. Work has consistently shown that Hcy exerts its excitotoxic influence through activation of N-methyl-D-aspartate (NMDA) receptors, one of the major excitatory glutamate receptors within the brain. NMDA receptor stimulation leads to calcium (Ca^{2+}) influx and it is excessive Ca^{2+} influx that can trigger cell death. Studies have found multiple pathways associated with Hcy-induced excitotoxicity, including activation of the apoptotic signaller caspase-3, generation of damaging reactive oxygen species (ROS), phosphorylation of the ERK-MAP kinase, and induction of lipid peroxidation (Jara-Prado et al., 2003, Kim et al., 2007b, Lipton et al., 1997, Poddar and Paul, 2009). In comparison to the reduction of Hcy levels through B vitamin, research has also demonstrated that treatment with NMDA receptor antagonists or antioxidants

can protect cells against the Hcy toxicity (Flott-Rahmel et al., 1998, Jara-Prado et al., 2003, Zieminska et al., 2006). Thus, it is likely that high levels of Hcy have a direct influence on brain atrophy, as exemplified in studies of individuals with hyperhomocysteinemia (Bleich et al., 2003).

In addition to the influence of Hcy, there are other features of the one-carbon cycle that could contribute to ROA changes. Disturbances in the one-carbon cycle that lead to increasing levels of Hcy are likely to disrupt the transsulfuration pathway, part of the one-carbon cycle that deals with the conversion of Hcy into cysteine and glutathione. Glutathione is an antioxidant that is essential in the clearance of ROS, and amelioration of glutathione levels has been shown to contribute to neuronal damage (Vitvitsky et al., 2006). As with Hcy, a number of pathways have been identified by which glutathione depletion could contribute to cell death, including sensitivity to nitric oxide levels, susceptibility to mitochondrial dysfunction, clearance of lipid peroxidation, and disposal of multiple ROS (Butterfield et al., 2006, Chen et al., 2005, Dringen, 2000, Sasaki et al., 2001). A large amount of work has also investigated glutathione levels in AD. Reduced glutathione levels have been reported in AD individuals and transgenic rodent models, although discrepant results have also been published (Butterfield et al., 2002, Liu et al., 2005, Zhang et al., 2012a). Moreover, genetic studies have identified specific polymorphisms in the glutathione-S-transferase (*GST*) gene that, like *APOE* E4, increase risk for AD (Pinhel et al., 2008). These observations support a role for glutathione alongside Hcy in explaining ROA changes, particularly the relationship between improved one-carbon cycle dynamics such as increased SAM levels and increased glutathione activity (Cavallaro et al., 2010). Interestingly, research has demonstrated that direct injection of glutathione into the hippocampus of rodents blocked neuronal cell death, exemplifying the potential role of glutathione in key brain regions such as the hippocampus (Saija et al., 1994).

Our assumption is that disruption of the one-carbon cycle, exemplified by high Hcy levels and reduced glutathione levels, impacts upon production of methyl groups and thus influences the ability of methyltransferases to carry out normal methylation. Whilst methylation levels may not directly influence ROA, changes to gene transcription related to methylation patterns could influence the functional state of neuron cells in key brain regions. For example, base-excision repair (BER) is the primary mechanism initiated by cells to deal with DNA damage. It is also the process

by which the further methylation derivatives 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) are converted back into unmethylated cytosine bases. Research has shown significant BER deficiency in AD but not limited to areas of DNA damage, suggesting that deficient demethylation may also be implicated in AD pathogenesis (Weissman et al., 2007).

BER activity has been found to decrease with increasing age, allowing for age-related methylation changes in the promoters of AD risk genes (Xu et al., 2008). As an example, hypomethylation of the *presenilin 1* gene promoter can lead to increased amyloid beta production, a key risk factor for the disease (Wang et al., 2008). However, elevation of SAM levels silences this effect, highlighting an important relationship between one-carbon cycle dynamics, changes in DNA methylation, and subsequent impact on disease phenotypes (Scarpa et al., 2003). That an identical relationship has been reported in the *β-secretase* gene, another risk factor for dementia, provides further support for the role of methylation in dementia pathogenesis (Fuso et al., 2005). With respect to cognitive decline rather than dementia, research has highlighted that a decline in neurogenesis with age coincides with increasing levels of DNA hypomethylation, providing a potential link between change in methylation patterns and neuronal function (Liu et al., 2009). Deletion of the methyltransferase *DNMT1* in neuronal precursor cells led to the development of highly demethylated neurons that were quickly eliminated via apoptosis (Fan et al., 2001). Methyltransferases may also have an indirect effect on neuronal survival through the altered methylation of genes associated with neuronal health and plasticity, such as *BDNF* (Chen et al., 2003).

This involvement of methyltransferases could provide an explanation for the *DNMT3L* R278G genotype-dependent changes observed in our study, whether that be in areas of neurogenesis in the adult brain such as the hippocampus or across more widespread brain areas during development, depending on the respective presence of DNMT3L in the adult and developing brain. Without additional detailed experimental data on the function of DNMT3L or the pathology of individuals within the VITACOG cohort, it would be remiss to favour one of these explanations over another. Given the close interactions between B vitamins, Hcy, glutathione, and methylation in the one-carbon cycle, it is likely that all of these factors contributed to the changes in cognitive performance and ROA described in this chapter.

4.3.3. Data considerations

In addition to the brain atrophy data, information was granted on key variables such as folate levels, cobalamin (vitamin B12) levels, diastolic blood pressure, and creatinine levels. In the previous chapter, the influence of B vitamin treatment on cognitive performance was assessed by examining the placebo and treatment groups respectively. With the addition of the folate and cobalamin data, there remains an opportunity to replicate our analyses of cognitive performance and ROA by taking into account those who have higher B vitamin levels and those who have lower B vitamin levels.

The original VITACOG study accounted for the effect of these measurements by classifying individuals within the treatment and placebo groups as being compliant or non-compliant. For the individuals who received B vitamin treatment, compliance was defined as an increase in folate of >10 nmol/L and an increase of vitamin B12 of >150 pmol/L. For the individuals who received the placebo, compliance was defined as an increase in folate of <10 nmol/L and an increase of vitamin B12 of <150 pmol/L. Of the 167 individuals included in the atrophy analysis, they characterised 136 individuals as being biologically compliant. When filtering only those who are compliant, they found that the effect of B vitamin treatment on ROA was slightly improved (Smith et al., 2010). Therefore, it is possible that applying biological compliance to our analyses could strengthen the observed genotype-dependent relationship between B vitamin treatment and cognitive performance.

In addition, amongst the covariate data supplied by VITACOG was the Geriatric Depression Scale (GDS). On the GDS, low scores indicate a normal response, mid-range scores indicate a mild depressive response, and high scores indicate a severe depressive response. The initial reports characterising MCI as a discrete disease showed that individuals with MCI had lower GDS scores than individuals with AD, providing a possible clinical distinction between the prodromal and developed stages of dementia (Petersen et al., 1999). The only observation regarding GDS performance was that individuals who received B vitamin treatment showed lower scores on the GDS, but no change compared to the placebo group throughout the trial. This baseline difference mirrors another coincidental finding seen in the previous chapter, where *DNMT3L* R278G G carriers performed better on the baseline CF test compared to A/A homozygotes.

The reliability and validity of the GDS has been corroborated in recent studies of depressed and non-depressed individuals, making it unlikely that the result is spurious (Sivrioglu et al., 2009). However, there is no obvious reason why the treatment group should be less depressed than the placebo group before the trial has begun. A significant association between depressive symptoms and memory performance has been reliably demonstrated in a meta-analysis of 147 studies into depression and cognition (Burt et al., 1995). However, the lack of correlation between baseline GDS scores and visuospatial associative memory, verbal semantic memory, or ROA diminishes the likelihood that our results were influenced by this baseline effect.

5. Replication of *DNMT3L* R278G analyses concerning cognitive performance in the NCDS 1958 and TwinsUK control cohorts

5.1. Preface

In the VITACOG cohort of individuals with MCI, altered visuospatial associative memory and verbal semantic memory performance was observed dependent upon the *DNMT3L* R278G variant. Corresponding changes in rate of brain atrophy were discovered that were also dependent upon genotype. To replicate our previous findings, data was used from two general population control cohorts to analyse vitamin variables, cognitive performance, and *DNMT3L* R278G.

5.2. Results

5.2.1. Cognitive factors in NCDS 1958 and TwinsUK

With the aim of replicating our analysis of vitamins, Hcy, *DNMT3L* R278G, and cognitive performance data conducted using the VITACOG cohort, Principal Component Analysis was applied to selected cognitive variables from the NCDS 1958 and TwinsUK general population control cohorts.

PCA was initially applied to word recall (WR), delayed word recall (DWR), animal naming (AN), Ps & Ws Scanned, and Ps & Ws Missed from the NCDS 1958 cohort. This resulted in two cognitive factors; PCA Factor 1 which consisted of WR, DWR, and AN, and PCA Factor 2 which consisted of Ps & Ws Scanned and Ps & Ws Missed. However, Ps & Ws Missed was not negatively loaded into the PCA as would be expected for a variable which inversely records performance (i.e. higher scores indicate poorer performance). Therefore, this PCA analysis was repeated but replacing Ps & Ws Missed with Ps & Ws Identified, a new variable derived by subtracted Ps & Ws missed from Ps & Ws Scanned. The repeated PCA resulted in the same two cognitive factors, but with the expected positive loading of Ps & Ws Identified.

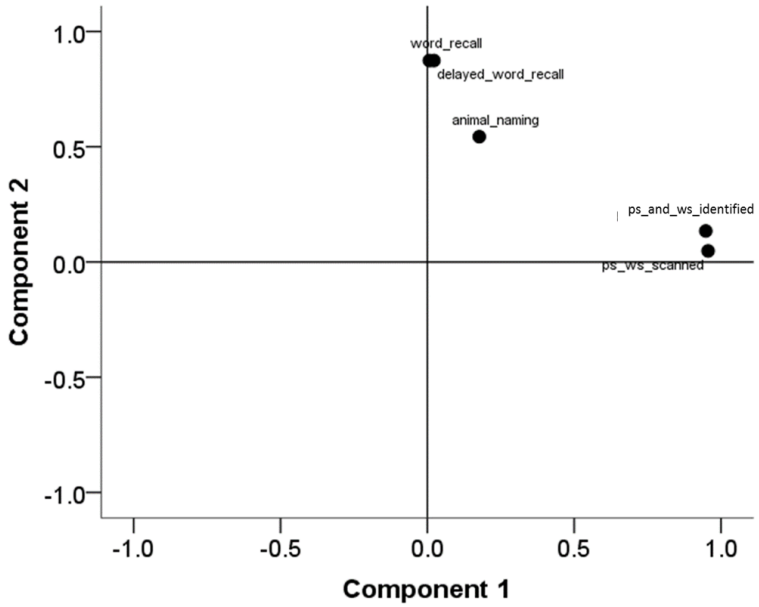
Although sampling adequacy (KMO & Bartlett's test) greater than 0.6 is generally expected for PCA analysis, our analysis resulted in a value of 0.550 ($p < 0.001$). I did not feel that this was discrepant enough to prohibit further analysis using these cognitive factors. PCA Factor 1 was interpreted to be reflective of verbal semantic

memory due to the segmentation of WR, DWR, and AN, semantic memory tests which have all been used in the evaluation of MCI progression (Coen et al., 1997, Cooper et al., 2004, Purser et al., 2005). In contrast, PCA Factor 2 was interpreted to be reflective of visual scanning due to the inclusion of both letter cancellation tasks (Uttl and Pilkenton-Taylor, 2001). PCA results and correlation coefficients for the NCDS 1958 cognitive data are presented in Figure 5.1.

In the TwinsUK cohort, PCA was applied to Delayed matching to sample (DMS) Total Correct and Mean Latency scores, Pattern recognition memory (PRM) Total Correct and Mean Latency scores, Paired associates learning (PAL) Total Errors, Spatial span (SSP) Length, Spatial working memory (SWM) Within Errors, Five Choice Reaction Time, and Intra-extra dimensional (IED) Set Shift Total Errors. As explained in the methods, these particular variables were chosen in order to avoid linear dependencies within each cognitive test. Sampling adequacy (KMO & Bartlett's test) reached an appropriate value of 0.697 ($p < 0.001$). This PCA analysis resulted in three cognitive factors; PCA Factor 1 which consisted of DMS Total Correct, PRM Total Correct, PAL Total Errors, and SSP Length, PCA Factor 2 which consisted of DMS Mean Latency and PRM Mean Latency, and PCA Factor 3 which consisted of SWM Within Errors, Five Choice Reaction Time, and IED Set Shift Total Errors.

The grouping of PAL, DMS, and PRM performance, all visuospatial tests that have applied to the study of dementia progression (Fowler et al., 1995, Galloway et al., 1992), led to the interpretation of this factor as being reflective of visuospatial associative memory. In comparison, due to the segmentation of both latency variables from DMS and PRM respectively, this factor was interpreted to be reflective of visual scanning performance. The final factor containing SWM Errors, Five Choice Reaction Time, and IED Errors appeared to be reflective of visual discrimination performance. However, because this factor was not identified in VITACOG or NCDS 1958, it was not included in our analysis. Instead, the focus remained on the comparison between the VITACOG and TwinsUK visuospatial associative memory factors, the VITACOG and NCDS 1958 verbal semantic memory factors, and the NCDS 1958 and TwinsUK visual scanning factors. PCA results and correlation coefficients for the TwinsUK cognitive data are presented in Figure 5.2.

A



B

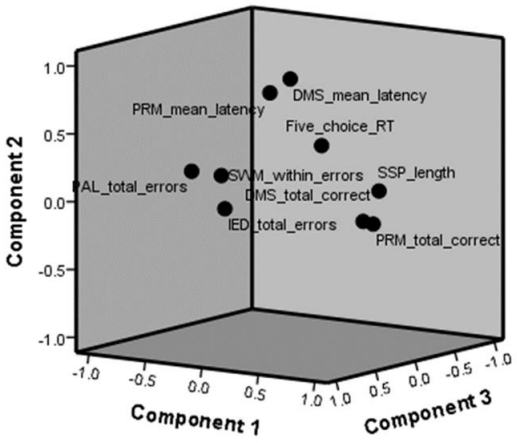
	Component	
	Visual scanning	Verbal semantic
Ps & Ws scanned	0.956	
Ps & Ws identified	0.948	
Word recall		0.874
Delayed word recall		0.874
Animal naming		0.543

C

	Word recall	Delayed word recall	Animal naming	Ps & Ws scanned	Ps & Ws identified	Visual scanning	Verbal semantic
Word recall	1						
Delayed word recall	0.648**	1					
Animal naming	0.275**	0.272**	1				
Ps & Ws scanned	0.078**	0.083**	0.129**	1			
Ps & Ws identified	0.136**	0.152**	0.188**	0.840**	1		
Visual scanning	0.007	0.022	0.177**	0.956**	0.948**	1	
Verbal semantic	0.874**	0.874**	0.543**	0.048**	0.134**	<0.001	1

FIGURE 5.1. PCA of select cognitive tests from the NCDS 1958 cohort. This analysis resulted in the identification of two cognitive factors (A) that were reflective of visual scanning and verbal semantic memory respectively (B). Correlation coefficients between the tests and the resultant factors are presented (C). ** = $p < 0.01$, * = $p < 0.05$ (2 tailed, Pearson).

A



B

	Component	
	Visuospatial associative	Visual scanning
DMS correct	0.738	
PRM correct	0.714	
PAL errors	-0.685	
SSP length	0.671	
DMS latency		0.869
PRM latency		0.714

C

	DMS correct	PRM correct	PAL errors	SSP length	DMS latency	PRM latency	Visuospatial associative	Visual scanning
DMS correct	1							
PRM correct	0.380**	1						
PAL errors	-0.346**	-0.423**	1					
SSP length	0.353**	0.284**	-0.341**	1				
DMS latency	-0.132	-0.76	0.102	< 0.001	1			
PRM latency	-0.165*	-0.290**	0.184**	-0.139*	0.367**	1		
Visuospatial associative	0.738**	0.714**	-0.685**	0.671**	0.045	-0.249**	1	
Visual scanning	-0.101	-0.146*	0.145*	0.073	0.869**	0.714**	< 0.001	1

FIGURE 5.2. PCA of select cognitive tests from the TwinsUK cohort. This analysis resulted in the identification of three cognitive factors, one of which was not included in subsequent analysis (A). The remaining two factors were reflective of visuospatial associative memory and visual scanning (B). Correlation coefficients between the tests and the retained factors are presented (C). DMS = Delayed Matching to Sample; PRM = Pattern Recognition Memory; PAL = Paired Associates Learning; SSP = Spatial Span. ** = $p < 0.01$, * = $p < 0.05$ (2 tailed, Pearson).

5.2.2. Baseline analyses in NCDS 1958 and TwinsUK

As with the cohorts from the OPTIMA program, the *DNMT3L* R278G genotype was assessed for Hardy-Weinberg equilibrium in the NCDS 1958 and TwinsUK cohorts. For this genotype, the NCDS 1958 ($p = 0.63$) and TwinsUK ($p = 0.54$) cohorts both conformed to Hardy-Weinberg equilibrium.

Having identified key cognitive factors in our general population control cohorts, the influence of baseline variables on cognitive performance was assessed. For NCDS 1958, all of the participants were the same age and so no effect of age was analysed. When investigating the influence of sex, it was found that females performed significantly better than males on both verbal semantic memory ($t = -6.859$, $p < 0.001$) and visual scanning ($t = -8.521$, $p < 0.001$). Conversely, for the TwinsUK cohort, all of the participants were female and so no effect of sex was analysed. As the TwinsUK cohort have been assessed for different variables at multiple times, age was examined when and where necessary for those respective variables. However, homocysteine (Hcy) levels from serum were also taken in the TwinsUK cohort and the age variable from this measurement was the most complete of the age variables provided. Using this age variable, an association was found between age and both visuospatial associative memory ($R = 0.208$, $p = 0.002$) and visual scanning ($R = 0.137$, $p = 0.046$), with older individuals showing poorer performance on both factors. There was no association between Hcy levels and either visuospatial associative memory ($t = -0.395$, $p = 0.693$) or visual scanning ($t = 0.937$, $p = 0.350$).

5.2.3. Vitamin influence in NCDS 1958 and TwinsUK

As the members of NCDS 1958 and TwinsUK were not subjected to any interventions in the same manner as the B vitamin treatment undertaken in the VITACOG cohort, vitamin levels recorded from blood serum in TwinsUK and self-reported vitamin supplement intake from both NCDS 1958 and TwinsUK were investigated instead.

For the NCDS 1958 cohort, self-reported qualitative data on daily single, combined, and folate vitamin use was provided. These variables were summated to ascertain regular, irregular, and non-reported vitamin supplement intake. Interestingly, a

significant relationship was observed between vitamin supplement intake and both verbal semantic memory (F = 26.932, $p < 0.001$) and visual scanning (F = 4.314, $p = 0.013$) performance (Figure 5.3A & B). Upon further investigation, it was found that single, combined, and folate vitamin use all significantly influenced verbal semantic memory performance, but only folate was associated with better visual scanning performance. Removal of those individuals whose vitamin intake was non-reported led to a loss of statistical significance in both models. *Post hoc* analysis with Bonferroni correction revealed that those who recorded regular vitamin intake performed significantly better than those whose vitamin intake was non-reported for both verbal semantic memory ($p < 0.001$) and visual scanning ($p = 0.010$). In the case of verbal semantic memory, those with regular vitamin intake also showed a trend towards significantly better performance than those with irregular vitamin intake ($p = 0.083$).

For the TwinsUK cohort, self-reported qualitative data on overall daily vitamin use was provided. By reviewing the precise vitamins taken by participants, a further variable classifying specific B vitamin intake was extrapolated. Vitamin B12 and folate were also measured from blood serum to provide biochemical vitamin levels. Promisingly, individuals with regular vitamin intake had higher levels of vitamin B12 and folate, although this was only significant in the case of folate levels (t = -3.938, $p < 0.001$). Furthermore, individuals who specifically took B vitamins had higher levels of vitamin B12 and folate, but neither of these observations reached statistical significance.

Beginning with vitamin intake, individuals who regularly took vitamins had significantly lower Hcy levels than individuals who had non-reported vitamin intake (t = 4.462, $p < 0.001$). Moreover, those who specifically took B vitamins had lower Hcy than those who took other vitamins, but this was not statistically significant (t = 1.251, $p = 0.211$). Examining serum vitamin B12 and folate levels also showed a significant effect on Hcy levels (F = 6.364, $p = 0.002$). *Post hoc* analysis with Bonferroni correction revealed that individuals with low levels of both vitamin B12 and folate had significantly higher Hcy levels than individuals with high levels of both B vitamins ($p = 0.003$) or individuals with high levels of one B vitamin and low levels of the other ($p = 0.018$).

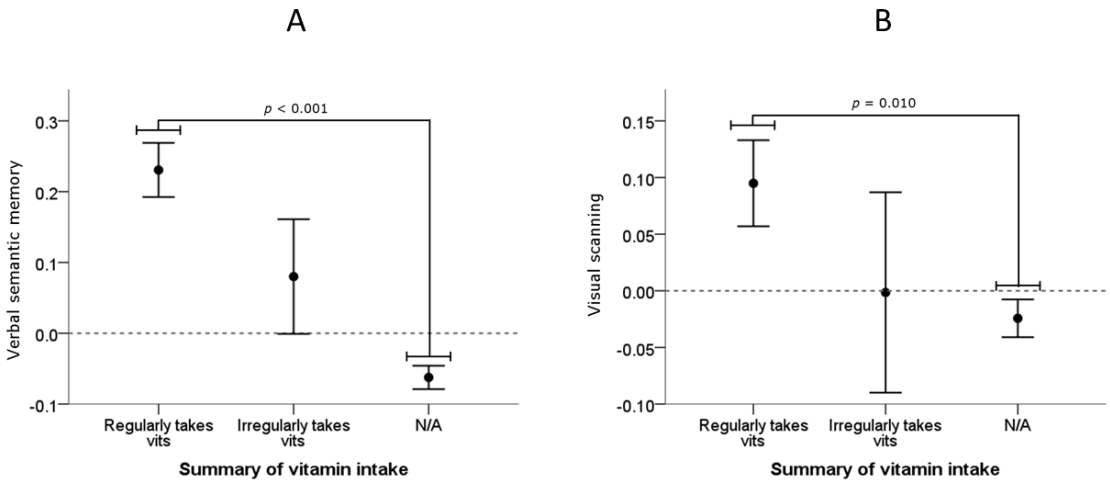


FIGURE 5.3. Relationship between vitamin intake and cognitive performance in the NCDS 1958 cohort. Data presented for verbal semantic memory (A) and visual scanning (B). Errors bars indicate +/- 1 SE.

Assessment of the TwinsUK cognitive factors revealed no association between serum vitamin B12 or folate levels and performance in visuospatial associative memory ($F = 1.221$, $p = 0.297$) or visual scanning ($F = 0.954$, $p = 0.387$) (Figure 5.4A & B). When investigating vitamin supplement intake, there was also no association between regular vitamin use and performance on visual scanning ($t = 0.245$, $p = 0.806$) (Figure 5.4D). However, regular vitamin intake was associated with significantly better performance on visuospatial associative memory ($t = -3.931$, $p < 0.001$) (Figure 5.4C). Moreover, within those regularly taking vitamins, specific use of B vitamins was associated with significantly better visuospatial associative memory performance ($t = -2.462$, $p = 0.015$).

In both general population control cohorts, a notable influence of vitamin intake on cognitive performance was observed, but no conclusive influence of vitamin levels. With respect to replicating the results seen in the VITACOG cohort, our analysis of Hcy in the TwinsUK cohort supports the findings that B vitamins can help to reduce Hcy levels. That only visuospatial associative memory performance was influenced by B vitamin use in the TwinsUK cohort provides direct support for the sensitivity of this cognitive factor, as well as for the benefit of B vitamins on tests linked to early cognitive decline. That folate intake was the most significant predictor of cognitive benefit amongst the vitamin variables within NCDS 1958 cohort further also supports the importance of B vitamins.

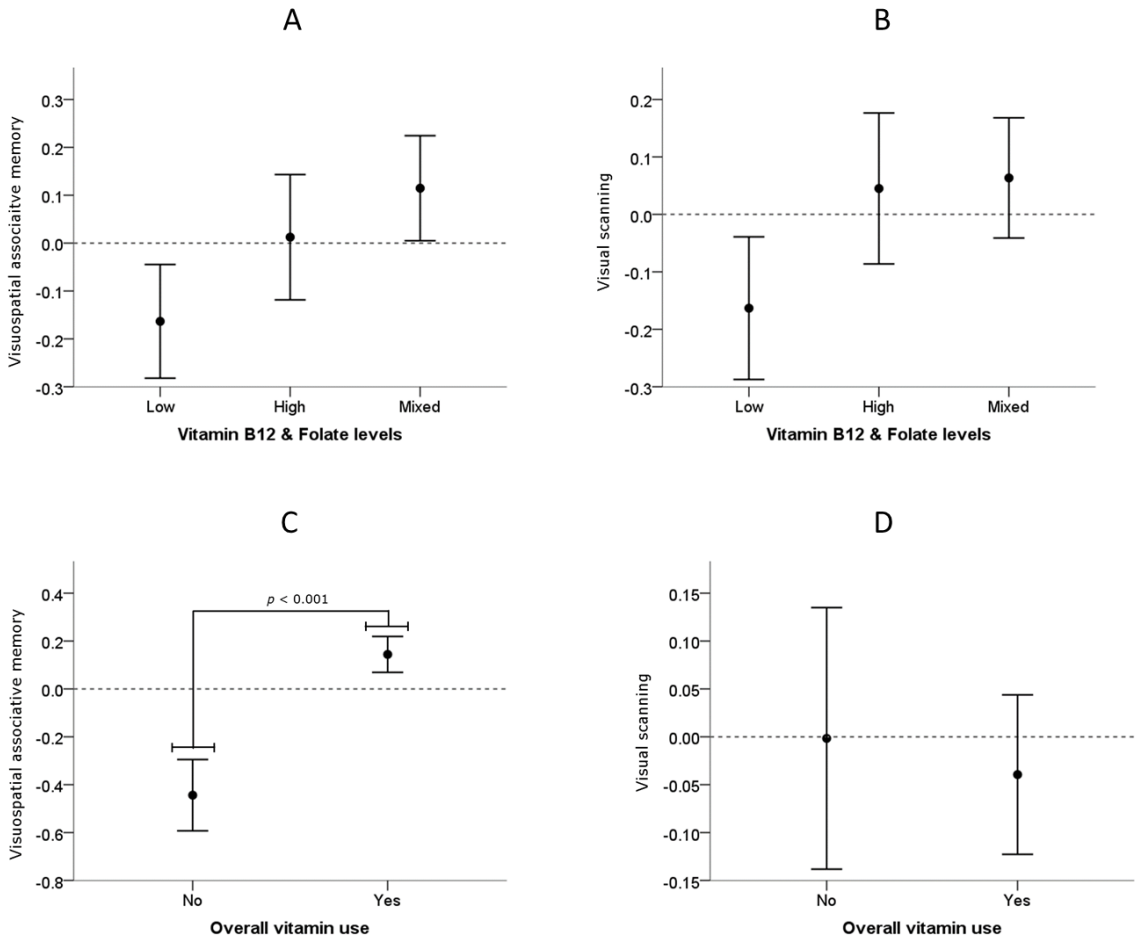


FIGURE 5.4. Relationship between vitamin intake or levels and cognitive performance in the TwinsUK cohort. Data is presented for vitamin B12 & folate levels (A, B) and for self-reported vitamin intake (C, D). Errors bars indicate +/- 1 SE.

5.2.4. Baseline genotype analyses in NCDS 1958 and TwinsUK

Next, our replication analysis turned to the *DNMT3L* R278G genotype which had an allele-specific influence on cognitive performance in the VITACOG cohort. As with the VITACOG analyses, genotypes for this variant have been grouped as either major allele A/A homozygotes or minor allele G carriers (made up of A/G heterozygotes and G/G homozygotes). Minor allele frequencies and genotype frequencies for the *DNMT3L* R278G genotype for both the NCDS 1958 and TwinsUK cohorts can be found in Table 5.1 and Table 5.2 respectively.

For the NCDS 1958, no relationship between the *DNMT3L* R278G variant and either verbal semantic memory (t = 0.269, $p = 0.788$) or visual scanning (t = -1.248, $p = 0.212$) was observed. For the TwinsUK cohort, there was no relationship between the *DNMT3L* R278G variant and Hcy levels (t = -0.238, $p = 0.812$), vitamin B12 levels (t = 1.428, $p = 0.154$), or folate levels (t = 0.950, $p = 0.343$). With regards to cognition, no relationship between the *DNMT3L* R278G variant and either visuospatial associative memory (t = 0.744, $p = 0.457$) or visual scanning (t = 1.512, $p = 0.132$) was observed.

5.2.5. Interaction of genotype and vitamins on cognitive performance

As the NCDS 1958 cohort lacks information on Hcy measurements, the opportunity for complete replication in these individuals is lacking. However, it still allows for assessment of the *DNMT3L* R278G genotype and its relationship with cognitive performance in a general population control cohort. The *DNMT3L* R278G genotype and vitamin supplement intake was first examined in this cohort.

TABLE 5.1. *DNMT3L* R278G minor allele frequencies for NCDS 1958 and TwinsUK

Cohort	*DNMT3L* R278G
NCDS 1958	0.18
TwinsUK	0.16

TABLE 5.2. *DNMT3L* R278G genotype frequencies for NCDS 1958 and TwinsUK

Cohort	*DNMT3L* R278G		
	A/A	**A/G**	**G/G**
NCDS 1958	0.523	0.398	0.078
TwinsUK	0.541	0.386	0.064

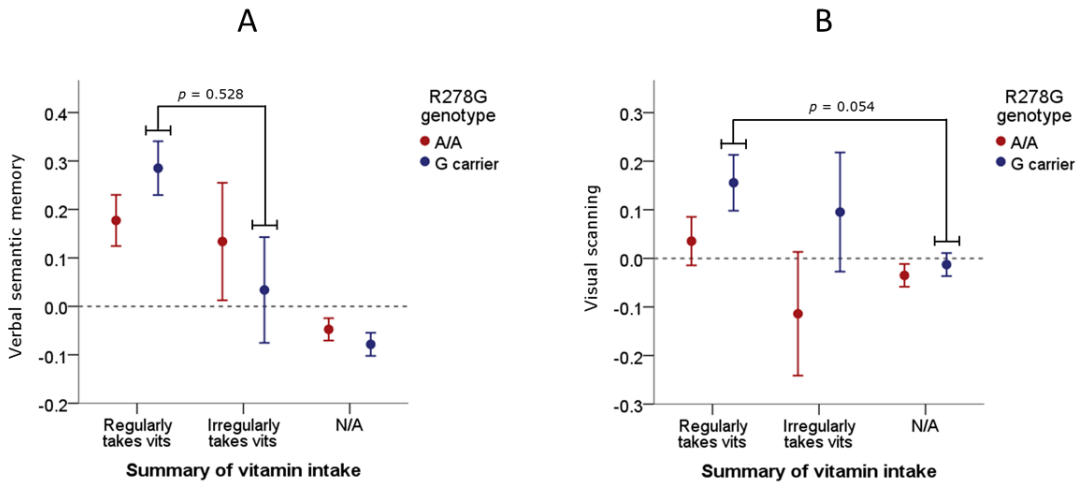


FIGURE 5.5. Further investigation of the relationship between vitamin intake and cognitive performance in the NCDS 1958 cohort. Data presented for verbal semantic memory (A) and visual scanning (B), and grouped by *DNMT3L* R278G A/A homozygotes (red) and G carriers (blue). Errors bars indicate +/- 1 SE.

Beginning with verbal semantic memory, there was no significant statistical interaction between *DNMT3L* R278G and vitamin intake (F = 1.641, p = 0.194) (Figure 5A). Although G carriers who regularly took vitamins performed better than G carriers who irregularly took vitamins, *post hoc* analysis with Bonferroni correction showed that this pattern was not significant (p = 0.528) (Figure 5.5A). With regards to visual scanning, there was again no significant statistical interaction between *DNMT3L* R278G and vitamin intake (F = 1.282, p = 0.277). *Post hoc* analysis with Bonferroni correction showed that G carriers who regularly took vitamins showed a trend towards significantly better performance compared with G carriers whose vitamin intake was non-recorded (p = 0.054), but not G carriers who irregularly took vitamins (p = 1.000) (Figure 5.5B). In contrast, both of these comparisons were non-significant for A/A homozygotes. Finally, sex had a significant baseline influence on cognitive performance in NCDS 1958, but addition of this as a covariate within these models did not alter the result for either verbal semantic memory (F = 1.646, p = 0.193) or visual scanning (F = 1.324, p = 0.266).

For the TwinsUK cohort, our examination began by looking at the influence of the *DNMT3L* R278G genotype on the relationship between vitamins and cognitive performance. Beginning with serum vitamin measurements, *DNMT3L* R278G had no influence on the non-significant relationships between vitamin B12 and folate levels and performance on visuospatial associative memory (F = 0.906, p = 0.406) or visual scanning (F = 0.276, p = 0.759). Having observed no association with vitamin levels, self-reported vitamin intake was subsequently analysed.

With regards to visuospatial associative memory, where regular vitamin intake and specifically B vitamin intake was associated with better performance, these relationships only remained significant in A/A homozygotes and not in G carriers. A/A homozygotes who regularly took vitamins performed significantly better than those with non-reported vitamin intake (t = -3.471, p = 0.001), an association that was non-significant for G carriers (t = -1.794, p = 0.076) (Figure 5.6A). Moreover, A/A homozygotes who specifically took B vitamins performed significantly better than those who took other vitamins (t = -2.516, p = 0.014), an association that was again non-significant for G carriers (t = -0.856, p = 0.398) (Figure 5.6B). With regards to visual scanning performance, A/A homozygotes with non-reported vitamin intake showed worse performance (higher values indicate greater latency of scanning) than G carriers (t = -2.298, p = 0.025) (Figure 5.6C). However, no similar pattern was observed when assessing the specific use of B vitamins (Figure 5.6D).

When incorporating Hcy levels into the TwinsUK analysis, no significant association was found between Hcy levels and vitamin B12 levels on visuospatial associative memory ($F = 1.229$, $p = 0.300$) or visual scanning performance ($F = 1.881$, $p = 0.134$), or between Hcy levels and folate levels on visuospatial associative memory ($F = 0.867$, $p = 0.459$) or visual scanning performance ($F = 0.358$, $p = 0.783$). Furthermore, the *DNMT3L* R278G genotype did not influence any of these relationships (or the lack thereof). Nevertheless, it was observed that A/A homozygotes with high Hcy and high B vitamin levels performed better in visuospatial associative memory, whilst G carriers with low Hcy and low B vitamin levels performed worse (Figure 5.7A). This comparison did not reach statistical significance in our *post hoc* analysis with Bonferroni correction due to high levels of variance ($p = 1.000$). Conversely, it was observed that A/A homozygotes with high Hcy and high B vitamin levels performed worse in visual scanning, whilst G carriers with high Hcy and low B vitamin levels performed better (Figure 5.7B). Whilst this observation is at odds with our expectations, this comparison again did not reach statistical significance in *post hoc* analysis with Bonferroni correction ($p = 1.000$).

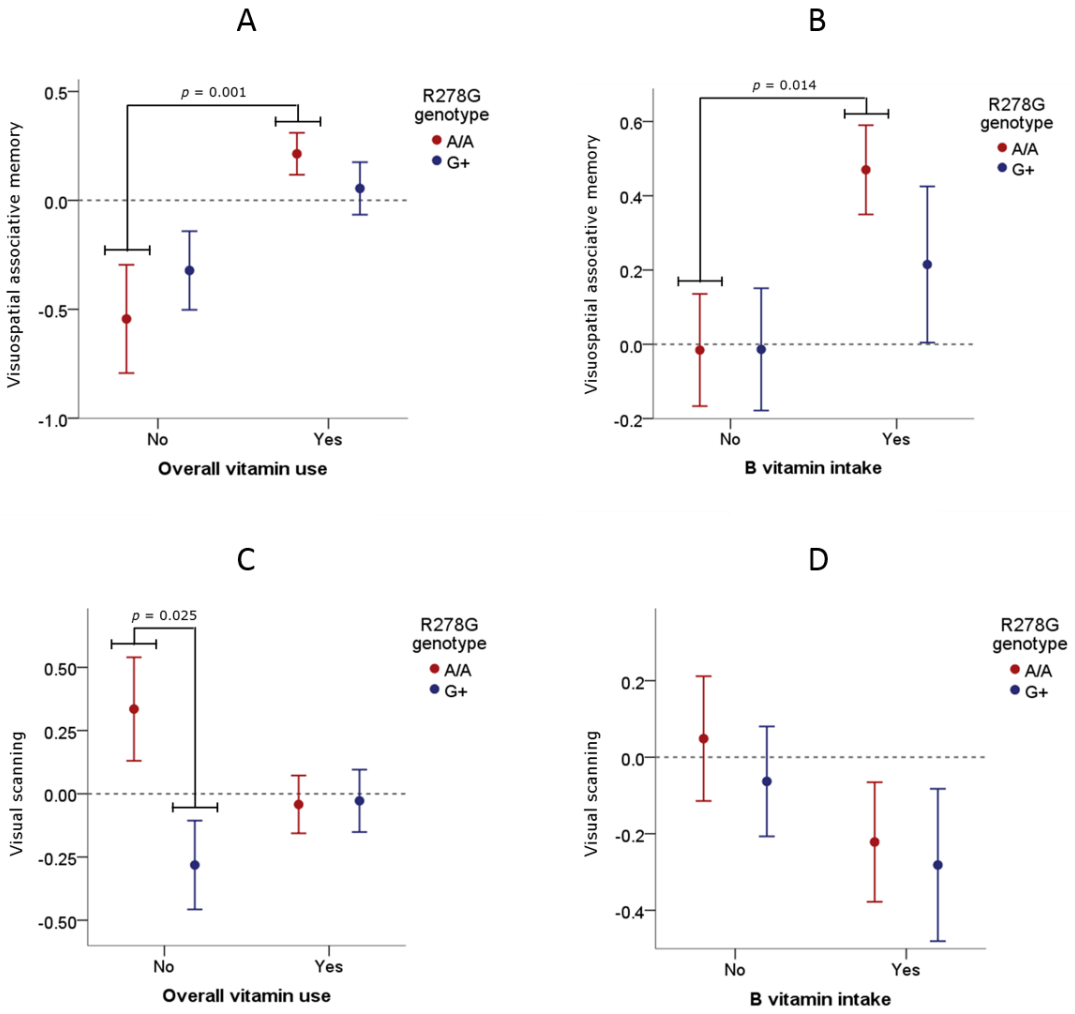


FIGURE 5.6. Further investigation of vitamin intake and cognitive performance in the TwinsUK cohort. Data presented for visuospatial associative memory (A, B) and visual scanning (C, D), and grouped by *DNMT3L* R278G A/A homozygotes (red) and G carriers (blue). Errors bars indicate +/- 1 SE.

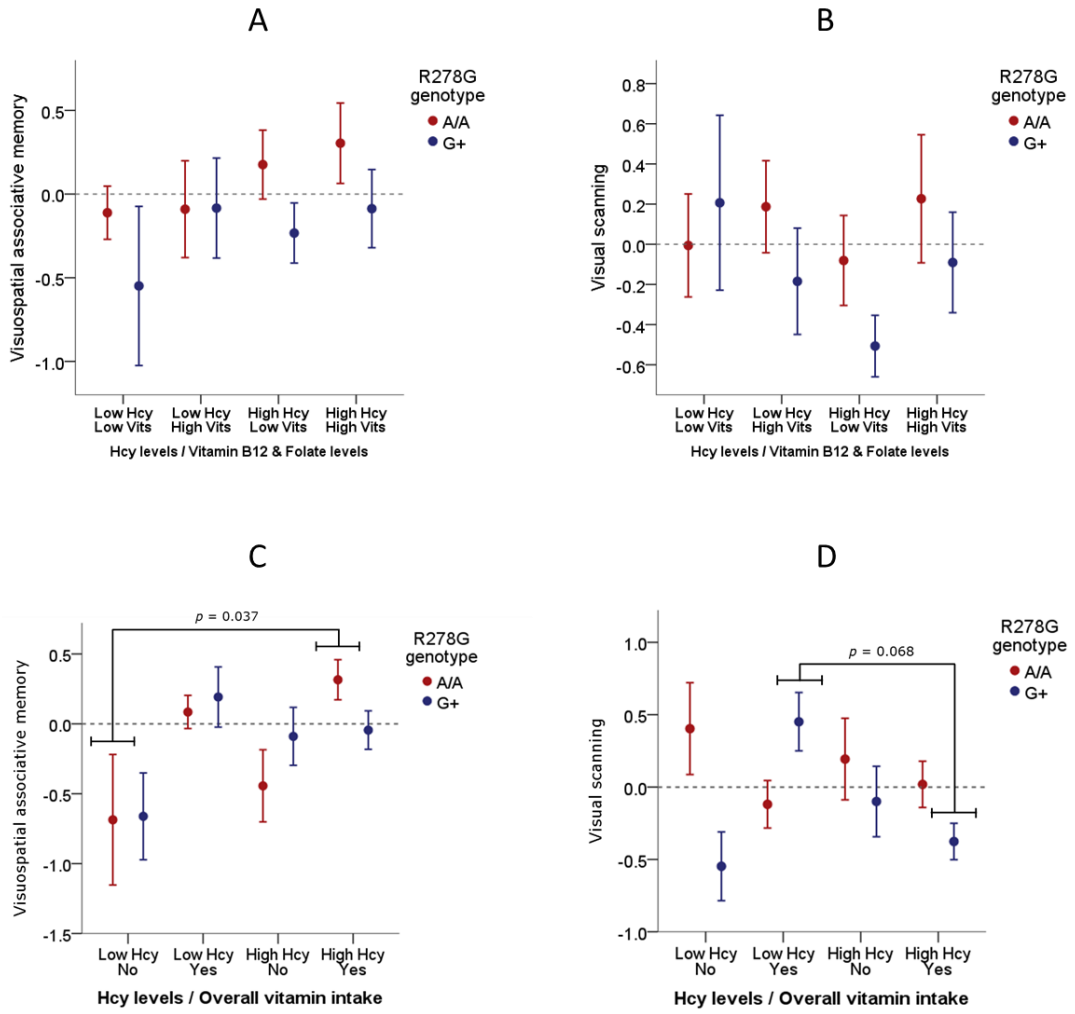


FIGURE 5.7. Inclusion of Hcy levels into analysis of vitamin intake or levels and cognitive performance in the TwinsUK cohort. Data presented for vitamin levels (A, B) and vitamin intake (C, D) alongside Hcy levels, and grouped by *DNMT3L* R278G A/A homozygotes (red) and G carriers (blue). Errors bars indicate +/- 1 SE.

Having observed no significant association with vitamin levels, self-reported vitamin intake was assessed. With regards to visuospatial associative memory, there was a significant effect of Hcy levels and vitamin intake on performance (F = 6.014, *p* = 0.001). *Post hoc* analysis with Bonferroni correction confirmed that individuals with low Hcy & non-reported vitamin intake performed significantly worse than individuals with low Hcy & regular vitamin intake (*p* = 0.004) or individuals with high Hcy & regular vitamin intake (*p* = 0.001). Addition of the *DNMT3L* R278G genotype to this model did not have a significant influence (F = 0.041, *p* = 0.839), nor a significant statistical interaction with Hcy levels and vitamin intake (F = 1.126, *p* = 0.298). However, *post hoc* analysis with Bonferroni correction revealed that the best performing group, A/A homozygotes with high Hcy & regular vitamin intake, performed significantly better than A/A homozygotes with low Hcy & non-reported vitamin intake (*p* = 0.037) (Figure 5.7C).

Age had a significant baseline influence on cognitive performance in TwinsUK, but the addition of age as a covariate within these models did not alter the statistical interaction of Hcy levels and vitamin intake (F = 7.990, *p* < 0.001) or the inclusion of *DNMT3L* R278G (F = 0.804, *p* = 0.493). Replication of these analyses with quartile Hcy levels corroborated the results, whereby the effect of Hcy levels and vitamin intake was significant (F = 5.459, *p* = 0.001), the statistical interaction with *DNMT3L* R278G was non-significant (F = 1.754, *p* = 0.158), and *post hoc* analysis with Bonferroni correction showed that the A/A homozygotes with upper quartile Hcy & regular vitamin intake performed significantly better than those with low and middle Hcy & non-reported vitamin intake (*p* = 0.045). Finally, investigation of B vitamin intake as opposed to overall vitamin intake resulted in no effect of B vitamin intake and Hcy levels on cognitive performance (F = 1.588, *p* = 0.196), as well as no statistical interaction with *DNMT3L* R278G (F = 1.170, *p* = 0.324). However, B vitamin intake and quartile Hcy levels did result in a statistical interaction with *DNMT3L* R278G (F = 3.329, *p* = 0.022), but *post hoc* analysis with Bonferroni correction could not reveal any specific differences that were driving this result.

With regards to visual scanning, although there was no significant effect of Hcy levels and vitamin intake (F = 1.429, *p* = 0.236) or the *DNMT3L* R278G genotype (F = 3.005, *p* = 0.085), there was a statistical interaction between Hcy levels, vitamin intake, and *DNMT3L* R278G (F = 4.502, *p* = 0.004). Addition of age as a covariate did not alter this statistical interaction (F = 4.088, *p* = 0.008). As shown in Figure 5.7D, whilst the A/A homozygotes show reasonably consistent performance across

the conditions, G carriers showed better performance with high Hcy & regular vitamin intake (p = 0.068) or low Hcy & non-reported vitamin intake (p = 0.117) compared to low Hcy & regular vitamin intake. However, neither of these comparisons reached statistical significance in a *post hoc* analysis with Bonferroni correction.

Replication of these analyses with quartile Hcy levels corroborated the results for visual scanning, whereby the effect of Hcy levels and vitamin intake was non-significant (F = 1.266, p = 0.287), the statistical interaction with *DNMT3L* R278G was significant (F = 3.142, p = 0.027), and there were no significant results following *post hoc* analysis with Bonferroni correction. Furthermore, investigation of B vitamin intake as opposed to overall vitamin intake resulted in no effect of B vitamin intake and Hcy levels on cognitive performance (F = 1.808, p = 0.150), as well as no statistical interaction with *DNMT3L* R278G (F = 1.379, p = 0.253). Replication of this analysis with quartile Hcy levels did not influence the results.

To summarise, in the NCDS 1958 cohort there was little influence of the *DNMT3L* R278G genotype, despite a trend towards improved performance in G carriers for both visual scanning and verbal semantic memory. In the TwinsUK cohort, G carriers also performed better on visual scanning, including when the influence of Hcy was taken into account. For visuospatial associative memory, it was the A/A homozygotes that appeared to show more consistent performance, in contrast to the improved performance of G carriers in the VITACOG analyses.

5.3. Discussion

In chapter three, a relationship was reported between B vitamin treatment, homocysteine levels, and the *DNMT3L* R278G genotype with respect to particular facets of cognitive performance. In chapter four, these relationships were substantiated by uncovering genotype-dependent relationships between cognitive performance and brain atrophy, before providing evidence for the functional impact of the R278G variant on the DNMT3A-3L methyltransferase complex. The focus of this chapter was to replicate the analyses of chapter three as closely and as accurately as possible in two general population control cohorts, providing a comparison with our findings in the VITACOG MCI cohort.

The NCDS 1958 and TwinsUK control cohorts contained self-reported vitamin intake information that was utilised as a comparison with the B vitamin treatment carried out in VITACOG. It was observed that regular self-reported vitamin intake was associated with improved visuospatial associative memory (TwinsUK), verbal semantic memory (NCDS 1958), and visual scanning (NCDS 1958 only) performance. B vitamin intake was derived from the self-reported vitamin intake information in TwinsUK and visuospatial associative memory benefits were discovered in individuals taking B vitamins compared to other vitamins. The TwinsUK cohort had biochemical measurements of B vitamin levels taken from blood, but no association between B vitamin levels and visuospatial associative memory or visual scanning were observed. However, both regular vitamin intake and higher B vitamin levels were associated with lower Hcy levels in the TwinsUK cohort. These findings confirm the beneficial effects of B vitamin intake on Hcy levels and specific cognitive domains previously drawn from the VITACOG study.

The *DNMT3L* R278G genotype was then included in the analyses. In the NCDS 1958 cohort, no influence of the R278G variant was observed with regards to the significant relationship between regular vitamin intake and verbal semantic memory or visual scanning performance. Similarly in TwinsUK, no influence of the R278G variant was observed on the non-significant relationship between B vitamin levels and visuospatial associative memory or visual scanning performance. However, it was discovered that A/A homozygotes who regularly took vitamins showed better visuospatial associative memory performance in TwinsUK. Furthermore, A/A homozygotes with high Hcy levels who regularly took vitamins showed better visuospatial associative memory performance than those with low Hcy whose vitamin

intake was non-reported. Here, a genotype-dependent relationship between vitamin intake and visuospatial associative memory was again observed, but with the A/A homozygotes within the TwinsUK controls compared to the G carriers within the VITACOG MCI individuals, illustrating an intriguing allele-specific difference between the two cohorts.

5.3.1. Regular vitamin intake in the control cohorts

Having sought to replicate our analysis of B vitamins, Hcy levels, and the *DNMT3L* R278G genotype on particular facets of cognitive performance, two significant but possibly unexpected findings were observed in our two control cohorts. The first was that self-reported regular vitamin intake, but not biochemical measurements of vitamin levels, was associated with better visuospatial associative memory and verbal semantic memory performance in the general population controls. The second was that A/A homozygotes who had high Hcy and regularly took vitamins performed better on visuospatial associative memory, contrasting with the same relationship observed in G carriers in VITACOG.

Concerning the first finding, no influence of B vitamin treatment alone on either of our derived cognitive factors was previously observed in the VITACOG cohort. That said, the VITACOG cohort had previously reported a vitamin effect on cognition when assessing specific cognitive tests, such as a stabilising of Clox test performance (De Jager et al., 2012). More striking is the observation that those who self-reported as taking B vitamins showed particular visuospatial associative memory benefit over those who took other vitamins. Although a number of studies have assessed the influence of particular vitamins in the NCDS 1958 and TwinsUK studies, very few of them appear to have focused on the relationship with cognitive performance. In NCDS 1958, one study investigated the link between genetic variants that influence specific nutrients such as vitamin B12 and cognitive ability, but found no associations (Alfred et al., 2013).

Much of the literature examining the relationship between vitamins and cognition is, understandably, carried out using supplementation protocols such as the one used in VITACOG or through biochemical measurement of vitamin levels. However, whilst the research regarding vitamin levels is more consistent in identifying a relationship with cognitive function, supplement studies have been mixed at best in reporting effects on cognitive performance. For example, compared to the isolated effects

seen in the VITACOG study of MCI, B vitamin supplementation has been largely ineffective in studies of AD and controls (Aisen et al., 2008, Sun et al., 2007, van Uffelen et al., 2008). Further, B vitamin supplementation in healthy individuals has been shown to both reduce Hcy levels and lead to beneficial effects on multiple cognitive domains, but also reduce Hcy levels with no influence on cognition (Clarke et al., 2014, Durga et al., 2007). When looking at biochemical measurements of B vitamins levels, high folate levels are associated with reduced risk for MCI and general cognitive impairment (Agnew-Blais et al., 2015, Doets et al., 2014). In addition, higher B vitamin levels have been associated with improved performance on specific tests of memory, hippocampal learning, semantic fluency, and general executive function (Kobe et al., 2016, Moorthy et al., 2012, Tucker et al., 2005, Vogiatzoglou et al., 2013). Therefore, it is perhaps surprising that such a relationship between B vitamin levels and cognitive performance was not observed in the TwinsUK controls.

Studies of self-reported B vitamin intake are rare in comparison to those examining supplementation or biochemical measurements. However, studies that have looked at self-reported B vitamin intake suggest a positive association with cognition, particularly memory function and protection against cognitive decline (Bryan and Calvaresi, 2004, Morris et al., 2005). Therefore, it is promising to see a similar relationship between self-reported vitamin intake and cognitive performance in our control cohorts. Despite this, there remains a discrepancy between the significant relationship between vitamin intake and visuospatial associative memory in TwinsUK that was absent when examining B vitamin levels in the same cohort. One would assume that vitamin levels should be the more accurate measure of this relationship, calling into question the validity of self-reported questionnaire data. It is possible that the self-reported data, rather than highlighting a causal relationship between vitamin intake and cognition, has instead acted as a proxy for other environmental factors. Indeed, individuals who report regular vitamin intake may be more health-conscious and more likely to exercise regularly, behaviour that is generally agreed to benefit cognition. As further nutritional and lifestyle information was not available from NCDS 1958 or TwinsUK, one can only speculate as to such underlying relationships.

In a similar manner, those that regularly took vitamins may be benefiting from the influence of a supplement other than B vitamins. Considering that individuals who self-reported having regular vitamin intake may take some form of multivitamin, the literature was searched for studies of multivitamin supplementation. However, the

majority of these showed no relationship with cognition, despite individual effects reported on contextual memory and Stroop test performance respectively (Cockle et al., 2000, Grodstein et al., 2013, Harris et al., 2012, Pipingas et al., 2014, Wolters et al., 2005). Whilst comparatively more studies have been carried out into vitamins D and E, they appear to show the same pattern as that described for B vitamins, where biochemical measurements correlate with cognitive function but supplementation is inefficacious. For instance, whilst vitamin D and E levels have been correlated with certain measures of cognitive, supplementation of either vitamin D or E has proven largely ineffective (Annweiler et al., 2012, Cherubini et al., 2005, Littlejohns et al., 2014, Lloret et al., 2009). Promisingly, however, is the body of work showing that polyunsaturated fatty acid (PUFAs) levels and supplementation can positively influence cognitive performance and risk of dementia (Chiu et al., 2008, Conquer et al., 2000, Eriksdotter et al., 2015, Kalmijn et al., 2004). A recent study using the VITACOG cohort showed an interactive influence of PUFAs and B vitamins, where the effect of B vitamins on cognitive decline was only observed in those with high PUFA levels (Oulhaj et al., 2016). This would suggest that future work examining self-report data and biochemical measurements for both B vitamins and PUFAs may provide further insight into the findings presented in this chapter.

5.3.2. Discrepant *DNMT3L* R278G findings in MCI and controls

The second notable finding discovered in our control replication was that A/A homozygotes who regularly took vitamins performed better on visuospatial associative memory. Moreover, A/A homozygotes who had high Hcy and regularly took vitamins performed better than those who had non-reported vitamin intake. This contrasts with the finding in our VITACOG MCI individuals, where G carriers who had high Hcy and underwent B vitamin treatment performed better on visuospatial associative memory. Taken together, these results suggest a model in which healthy middle age *DNMT3L* R278G A/A homozygotes who regularly take vitamins demonstrate better visuospatial associative memory performance. However, once individuals decline to MCI levels, B vitamins appear to boost the visuospatial associative memory performance of G carriers (Figure 5.8).

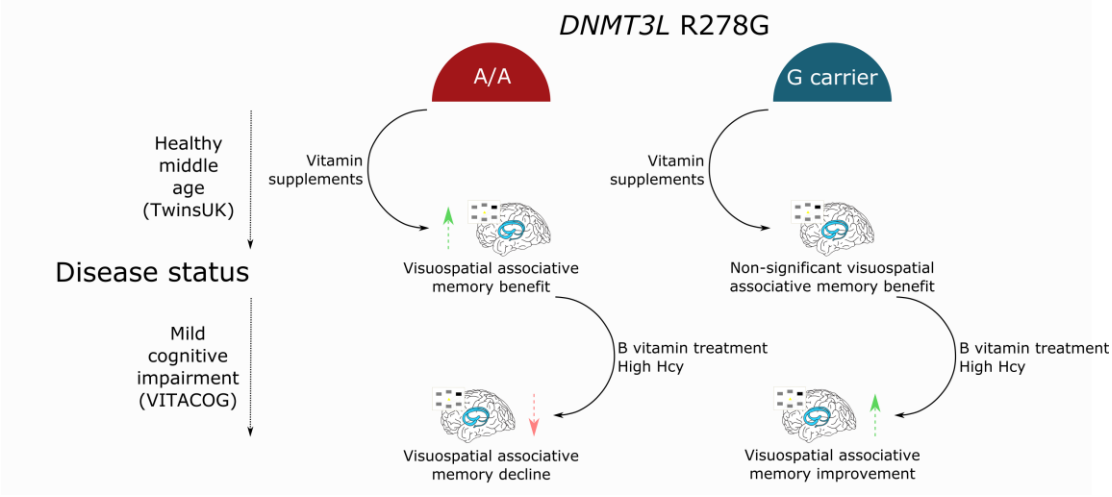


FIGURE 5.8. Illustrated model of the relationship between disease status, methionine pathway components, and the *DNMT3L* R278G variant with respect to cognitive performance. Beneficial (green arrow) or detrimental (red arrow) cognitive outcomes in visuospatial associative memory are associated with interactions between vitamin intake, Hcy risk, and the *DNMT3L* R278G genotype.

This model provides further support for the idea that alterations in the relationship between Hcy levels, B vitamins, and DNA methylation within the one-carbon cycle can influence specific measures of cognitive function. In the previous chapters, the mechanisms by which this could occur were discussed, initially focusing on the reduction of Hcy neurotoxicity and restoration of glutathione levels stimulated by improved one-carbon cycle dynamics. These factors almost certainly have a part to play in the overall picture of B vitamin supplementation and its influence on both cognitive performance and brain atrophy. However, the discrepant genotype-dependent differences in visuospatial associative memory discovered in MCI and controls respectively call for a mechanistic explanation centred on the differences in methylation patterns related to the *DNMT3L* R278G genotype.

As previously mentioned, the hypomethylation hypothesis provides a general explanation for the impairment of normal DNA methylation patterns caused by disruption of the one-carbon cycle. Whilst this hypothesis is normally applied to cognitive dysfunction, the principles can be applied to cognitive improvement by emphasising that lowering of Hcy levels and restoration of one-carbon cycle function can allow beneficial DNA methylation activity to occur (i.e. the silencing of risk-related gene expression or the promotion of beneficial gene expression). In this study, the measurement of visuospatial associative memory appears to have highlighted a particular change in cognitive function that would otherwise have gone undetected. Discussion in the previous chapter exemplified the sensitivity of visuospatial associative memory in measuring medial temporal lobe function, particularly that of the hippocampus. Changes to DNA methylation patterns are known to occur in the hippocampus during memory formation and consolidation, and these mnemonic processes can be disrupted through inhibition of methyltransferase and demethylase proteins (Feng et al., 2010, Miller and Sweatt, 2007, Rudenko et al., 2013). Thus, it is in the hippocampus that one would expect the *DNMT3L* R278G genotype-dependent changes in DNA methylation patterns to occur.

There are certain additional factors that must be taken into account when reviewing this model. To begin with, our findings using control cohorts included regular vitamin intake, not B vitamin intake alone. This suggests the involvement of other vitamins, or indeed other pathways, that can act as the trigger for changes in DNA methylation patterns. A notable example would be the potential involvement of PUFAs within the pathways described above. Previous evidence suggests that the relationship between B vitamins and cognitive performance may only occur in the presence of

high PUFA levels (Oulhaj et al., 2016). Further, the authors suggest that the formation of particular fatty acids such as phosphatidylcholine is facilitated by the presence of B vitamins, as increased Hcy has been correlated with decreased phosphatidylcholine in AD, cementing the relationship between B vitamins and PUFAs (Selley, 2007).

Separate to the interactions with the one-carbon cycle, research has also shown that PUFAs can have a direct impact on brain function. For example, the key synaptic processes long-term potentiation (LTP) and long-term depression (LTD), as well as the release of key neurotransmitters such as serotonin, are inhibited in the hippocampal regions of mice on PUFA-deficient diets (Chalon, 2006, Delpech et al., 2015, Thomazeau et al., 2017). All of these deficits, as well as the effects on neurogenesis and memory performance, could be rescued with PUFA treatment (Crupi et al., 2012, Kavraal et al., 2012, Plamondon and Roberge, 2008, Vancassel et al., 2008).

Many other vitamins commonly included in multivitamin supplements have also been investigated in relation to DNA methylation events. Beginning with vitamin D, studies of maternal vitamin D deprivation in rodents found that the offspring showed some small changes in methylation, but no change in gene expression (Xue et al., 2016). However, vitamin D supplementation in human mothers was found to have a significant influence on differential methylation patterns in the respective infants (Anderson et al., 2016). Moreover, a recent randomized control trial in African-Americans found that vitamin D supplementation resulted in global increases in DNA methylation (Zhu et al., 2016). There is less research into the influence of vitamin D and methylation in disease compared to B vitamins, but vitamin D has been shown to play a beneficial role in models of autoimmune disease, seemingly through corresponding genome-wide changes in methylation patterns (Zeitelhofer et al., 2017). However, unlike PUFAs, work has illustrated that vitamin D appears to influence DNA methylation independently of one-carbon cycle dynamics, pointing to an as-of-yet unknown mechanistic relationship with methylation (Beckett et al., 2016).

Aside from vitamin D, studies of vitamin C supplementation in maternal smokers have presented positive effects on infant health associated with differential methylation patterns (Shorey-Kendrick et al., 2017). Interestingly, a recent study has discovered that both vitamins A and C can stimulate differential methylation patterns

in vitro by modulating TET-mediated demethylation activity (Hore et al., 2016). This outlines a possible pathway independent of one-carbon cycle methyl donation through which vitamins could influence methylation patterns through demethylation dynamics, drawing parallels with the association between increased BER activity and age-related DNA methylation changes in dementia discussed in the previous chapter. However, as *DNMT3L* R278G is putatively a methyltransferase rather than a demethylase, the findings observed in this study are likely associated with the deposition of methyl marks rather than their removal.

As observed with the nearby R271Q variant in *DNMT3L*, genetic variants in methyltransferase genes have been shown to cause changes in DNA methylation patterns (El-Maarri et al., 2009). Similar influences on the deposition of methylation patterns have been linked to splice variants within the maintenance and *de novo* DNA methyltransferases (Bonfils et al., 2000, Weisenberger et al., 2004). Furthermore, variants in methyltransferases have been associated with methylation changes observed in diseases such as schizophrenia and specific cancers (Saradalekshmi et al., 2014, Tahara et al., 2009). With regards to the potential methylation changes associated with *DNMT3L* R278G stemming from our findings, there are multiple factors that could act as the trigger that differentiates our control and MCI groups. For example, the burden of high levels of Hcy coupled with B vitamin treatment could be the only necessary factors to stimulate genotype-dependent change in methylation patterns in MCI. The difference in age between the MCI (76 years old) and control (NCDS 1958 = 45-50 years old; TwinsUK = 57 years old) could play a role, as methylation levels are known to alter with aging (Bollati et al., 2009).

Disease-specific methylation changes have been observed in MCI compared to AD and control individuals (Ellison et al., 2017). This could result in a situation where the control and MCI individuals are primed to exhibit disease-specific behavior stemming from differential methylation patterns, with the genotype-dependent activity of DNMT3L simply exacerbating these differences. In a similar vein, *DNMT3L* R278G genotype-dependent change in methylation patterns could be acting upon other risk genotypes to enact disease-specific behavior. For example, *APOE* contains multiple methylations sites with research suggesting that the E4 allele alters methylation levels both within *APOE* and downstream of the gene (Foraker et al., 2015, Wang et al., 2008, Yu et al., 2013). *APOE* methylation levels have been shown to change significantly with age but not with disease (Ma et al., 2015, Tannorella et al., 2015).

Thus, it is possible that methylation-dependent change in *APOE* transcription can account for some of the difference between our control and MCI cohorts.

Finally, there could be other disease-associated changes that influence DNMT3L function. DNMT3L is known to interact with histone H3 proteins during the deposition of DNA methylation. However, histone H3 can be disrupted in dementia, potentially leading to further disruption of normal methylation patterns (Lithner et al., 2013, Ogawa et al., 2003). In addition, genetic variants have been also been shown to disrupt the binding between DNMT3L and histone H3, providing another mechanism that could influence differential methylation (Ooi et al., 2007). Thus, whilst the influence of Hcy and vitamins in the one-carbon cycle is a favorable model, I have demonstrated that multiple biological mechanisms may underline the visuospatial associative memory differences between MCI and controls in our study.

5.3.3. Methodological questions

Understandably, attempting to carry out an exact replication of the VITACOG analyses would have been difficult given the number of variables included in the study and the length of the B vitamin treatment trial. Our combination of the NCDS 1958 and TwinsUK longitudinal general population cohorts has succeeded in replicating the main analyses from VITACOG, with information on brain atrophy and the *APOE* genotype much harder to come across. However, there are a number of methodological issues that could be raised with our replicative analysis.

A number of these issues have been raised already. One of these is the difference in age between our MCI and control cohorts. Whilst this may have contributed to the discrepant findings between MCI and controls, the identification of a genotype-dependent change in middle age that contrasted with the change in MCI is arguably more interesting than a straight comparison with age-matched controls. Another was the comparison of self-reported and biochemical levels of vitamins, which has already been discussed within this chapter, as well as the comparison of these two measures with the vitamin treatment methodology used in VITACOG. There are other B vitamin trials carried out in healthy participants that could have been used as our control cohort (Bryan et al., 2002, Eussen et al., 2006, Hvas et al., 2004, Kennedy et al., 2010). However, many of these other trials did not include DNA sample collection and genetic analysis within their design and therefore sequencing

could only have been carried out if blood samples were available from these studies, incurring further financial and time costs.

There are questions of accuracy regarding the variables used within our replication analyses. To begin with, the nature of serial assessment within these longitudinal control cohorts means that data is provided for the same individuals from different time points. This retrospective analysis of longitudinal data requires the integration of isolated data from individuals irrespective of the age at which it was taken. Difficulty in controlling for the relationships between these variables means that temporal bias (one cannot estimate causality because the temporal information for each variable is absent or omitted) influences the control observations to a greater extent than those from VITACOG. Moreover, the TwinsUK may suffer from selection bias as cognitive data was only available for a minority of individuals.

For the self-reported vitamin intake in NCDS 1958, new variables were derived that summated vitamin intake. This may also suffer from a selection bias as the parameters of the new variable were decided by myself. One could also question the inclusion of individuals whose vitamin intake data is non-reported, as one cannot be entirely sure that their vitamin intake is reduced compared to those who provided self-report data. Indeed, removal of the non-reported individuals resulted in both verbal semantic and visual scanning results becoming non-significant. Nevertheless, assessment of the original vitamin intake variables from NCDS 1958 showed that single vitamin and folate intake also showed the same association with verbal semantic memory performance, whilst folate intake showed an association with visual scanning (Appendix 25), providing support for the use of our derived vitamin intake variables.

The sensitivity of our analysis to changes in cognitive performance is dependent upon the accuracy of the PCA factors. It would not be feasible to locate and gain access to cohorts from different studies that have data from identical cognitive tests. Instead, shared cognitive tests were identified between the VITACOG study and our control cohorts to ensure that our derived PCA factors could be compared accurately. The presence of PAL in both the VITACOG and TwinsUK visuospatial associative memory factors is key, whilst the inclusion of the visuospatial tests DMS, PRM, and SSP add further support to the similarity of these factors. However, these abstract recognition tests were not coupled with a test of semantic memory such as GNT, the key difference between the VITACOG and TwinsUK factors. With regards

to verbal semantic memory, there was no such inclusion of identical tests between VITACOG and NCDS 1958. Nevertheless, the measures seem relate well, as both VITACOG and NCDS 1958 contained measures of delayed recall (HVLT-R and DWR) and semantic fluency (CF and AN). Thus, whilst the tests within the respective cognitive factors from each cohort are not identical, one can be confident that they are reflective of similar underlying cognitive processes.

The visual scanning factor was retained although it was not present in the VITACOG cohort in order to provide a "control" measure of cognitive performance in both of our control cohorts. PCA in the TwinsUK cohort also resulted in a visual discrimination factor that was removed from further analysis, as a "control" cognitive factor was already included and there was no comparable factor from VITACOG or NCDS 1958. As outlined in the methods, some data had to be omitted from the PCA analysis due to issues with linear dependencies between extrapolated variables from the same cognitive tests.

The National Adult Reading Test (NART) verbal IQ score also had to be omitted as the scale of performance could not be ascertained or how the verbal IQ had been calculated within the original NART test. However, further analysis of these cohorts may wish to include this test, as this more general indicator of premorbid intelligence could influence the segregation of the PCA factors. Longitudinal studies have shown that premorbid intelligence is associated with disease risk (Koenen et al., 2009, Zammit et al., 2004). Moreover, differences in cognitive performance between individuals with a brain disease are associated with their premorbid intelligence (Wells et al., 2015). In the VITACOG cohort, one could offer some control for individual differences by using change in cognition rather than cognitive performance at one time point. This was not possible in either of the control cohorts, making it more likely that normal variation in intelligence could influence the results and underlining the need to include a more general intelligence test such as NART.

6. Examination of rare variants within DNA and RNA methylation genes in learning disability and comorbid psychiatric disease cohorts

6.1. Preface

Our analysis of one specific methyltransferase variant, *DNMT3L* R278G, had discovered previously unknown relationships with cognitive performance. Following this analysis, genetic variation in additional genes involved in DNA and RNA methylation were investigated. To accomplish this, next-generation sequencing (NGS) data was used from individuals with intellectual disability (ID) and ID comorbid with psychiatric conditions, autism spectrum disorder (ASD), and general population controls.

6.2. Results

6.2.1. Descriptive summary of variants

Genetic variation was investigated within a selection of genes associated with DNA and RNA methylation, specifically assessing the relationship between rare and low-frequency missense variants and specific brain disorders. These methylation genes can be divided into different categories based on the function of the encoded proteins; namely, readers, writers, and erasers of the methyl modification (Table 6.1). Missense variants were chosen because of their influence of protein function. Other variants that similarly disrupt protein function, such as loss-of-function (LoF) variants, will be included in future work. Rare variants are those that have a minor allele frequency (MAF) < 0.01 and low-frequency variants are those that have a MAF < 0.05.

The TwinsUK cohort of 1854 individuals was utilised as a general population control cohort. The ID group comprised 290 individuals from the Rare FIND (Familial Intellectual Disability) and Muir cohorts. The Rare FIND cohort includes individuals with moderate to severe ID while the Muir cohort includes individuals with learning disabilities, many of whom also presented with comorbid psychiatric disease. Finally, the ASD group comprised 620 individuals from the Gallagher, IMGSAC, MGAS, and Skuse cohorts. In total, 57,787 variants were identified across 30 DNA and RNA methylation genes. *MECP2* was initially included in the list but was omitted from the analysis due to problems with ploidy during the NGS processing pipeline.

TABLE 6.1. DNA and RNA methylation genes and their functional categories

DNA methylation			RNA methylation		
Writers	**Readers**	**Erasers**	**Writers**	**Readers**	**Erasers**
DNMT1	*MBD1*	*TET1*	*METTL3*	*YTHDF2*	*ALKBH5*
DNMT3A	*MBD3*	*TET2*	*METTL14*	*YTHDF3*	*FTO*
DNMT3B	*UHRF1*	*TET3*	*WTAP*	*HNRNPA2B1*	
DNMT3L	*UHRF2*	*GADD45A*		*ELAVL1*	
	TDG	*GADD45B*			
	NEIL1	*MBD2*			
	NEIL3	*MBD4*			
	WDR76	*ELP3*			
	THYN1				

6.2.1.1. Characterising missense variants

In total, 855 missense variants were identified in all of the methylation genes within our cohorts. 274 missense variants were found in the controls individuals, 220 in the ID cohorts, and 361 in the ASD cohorts. The majority of these were within DNA methylation genes (736) compared to RNA methylation genes (119), particularly within DNA methylation eraser category (377).

Each variant position was annotated where possible with a reference MAF from the ExAC consortium database, a large aggregated catalogue of sequencing data covering 60,706 individuals from healthy and disease populations, thus providing a large general population reference MAF. Following this annotation, 52 missense variants were discovered with a MAF in the control or case cohorts which differed from the ExAC reference MAF. The majority of these variants (46) were found in the DNA methylation genes. Two-thirds of the missense variants (35) were found at a higher MAF in our cohorts than the ExAC reference. SIFT and PolyPhen functional annotation for each variant left 24 of the 52 missense variants that were predicted to have a damaging effect on the protein. Full details for these variants can be found in Appendix 26.

114 missense variants were also highlighted as missing an ExAC reference MAF. This could either be because the variant is especially rare, as appeared to be the case in the majority of these variants (106), or because our automated annotation process using ExAC has failed to pick up that particular variant position. The majority of these variants (97) did not appear to have previously been identified (i.e. lacking an “rs” accession number), but this could in part be due to gaps in the annotation pipeline as this process is dependent upon up-to-date external databases. In addition to these rare missense variants, 6 low-frequency and common missense variants were identified that had been annotated as functionally damaging (Appendix 27). Of interest, *DNMT3L* R278G was included in this group, possibly providing other targets similar to R278G which could be investigated further for their relationship with cognitive function.

6.2.1.2. Rare and low-frequency missense variants

The percentage of rare and low-frequency missense variants within the total number of variants were assessed. The number of rare and low-frequency variants which were predicted to be functionally damaging using the SIFT and PolyPhen functional annotation tools were also taken into account. Both rare and low-frequency missense variants were found more frequently in the case cohorts than in controls, a pattern that was largely retained for those variants that were annotated as functionally damaging. Of note was the comparatively greater number of rare missense variants in the ID cohorts compared to controls (Figure 6.1A).

The percentage of rare and low-frequency missense variants within the total number of missense variants were then considered. The majority of missense variants in the controls were rare, but this could be due to the whole-genome sequencing used in the TwinsUK cohort picking up a greater proportion of rare missense variants compared to the whole-exome sequencing using in the case cohorts. Despite this, the proportion of rare damaging variants did not differ between cases (0.36%) and controls (ID = 0.57%, ASD = 0.42%). Low-frequency missense variants were again found more frequently in the case cohorts than controls, with no difference in the proportion of predicted damaging variants between cases and controls (Figure 6.1B).

6.2.1.3. Missense variants within DNA and RNA methylation functional categories

To acquire a more detailed picture of these specific methylation genes, our assessment of these rare and low-frequency missense variants was divided by DNA and RNA methylation genes and by functional categories (writers, readers, and erasers). All categories of DNA methylation genes showed a higher proportion of rare missense variants in the controls. In most categories for both DNA and RNA methylation genes, the proportion of rare missense variants was higher in ID than ASD cohorts. However, for the RNA methylation eraser category, 100% of all missense variants were rare in controls and ID whilst over 90% were rare in ASD (Figure 6.1C). The lack of common or low-frequency missense variants in the RNA methylation erasers could indicate that this category is comparatively intolerant to genetic disruption and therefore a biologically important category.

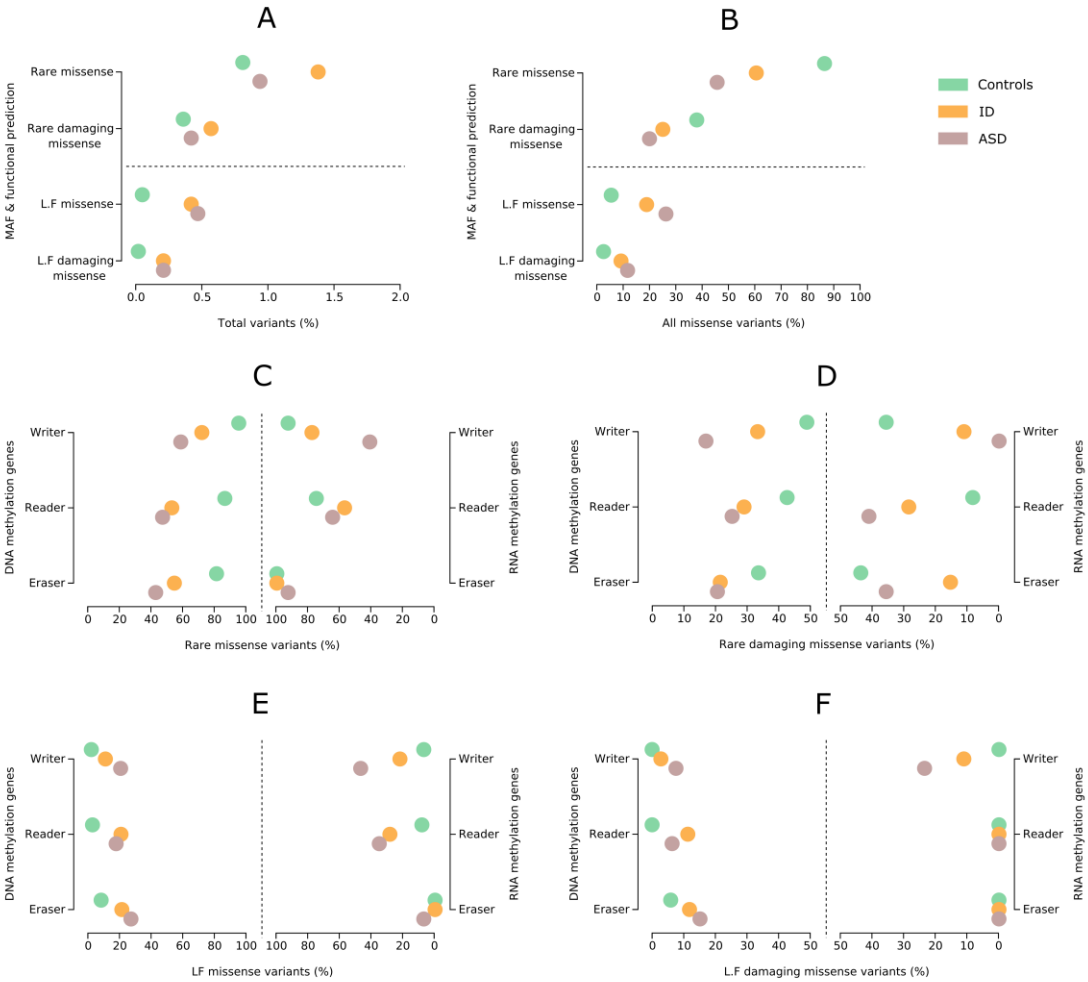


FIGURE 6.1. Rare and low-frequency variants in DNA and RNA methylation genes from the ASD, ID, and control cohorts. Green represents the control cohorts, orange represents the ID cohorts, and purple represents the ASD cohorts. Rare and low-frequency variants are presented as a proportion of the total number of variants (A) and the total number of missense variants (B). The proportion of rare (C, D) and low-frequency (E, F) missense variants are compared between DNA and RNA methylation genes.

The number of rare missense variants that were predicted to be functionally damaging was then taken into account. Most functional categories for DNA and RNA methylation genes again showed a higher proportion of rare damaging missense variants in the controls and a lower proportion in ASD. However, the RNA methylation readers showed the reverse picture with a higher proportion of rare damaging missense variants in both case populations and lower in controls (Figure 6.1D). This contrasts with the previous observations that greater numbers of rare damaging missense variants in controls compared to cases may be caused by the use of whole-genome sequencing in the controls. Moreover, it suggests that the RNA methylation readers could be a functionally important category for genetic risk in our ASD and ID cohorts.

Low-frequency missense variants (Figure 6.1E) and damaging low-frequency missense variants were also assessed (Figure 6.1F) in each of our DNA and RNA methylation categories. However, the number of these low-frequency variants was comparatively low, making it difficult to draw any clear patterns from this data. Nevertheless, a greater proportion of low-frequency missense variants in the ASD cohorts was observed compared to controls, and this pattern was retained when looking only at functionally damaging low-frequency missense variants.

In summary, our initial assessment of variant counts has identified key patterns associated with ID and ASD in each of the RNA methylation gene categories. Namely, a greater number of rare missense variants were found in ID compared to controls, a contrasting proportional increase in rare damaging missense variants were found in both ASD and ID compared to controls, and a proportional increase in low-frequency damaging variants in ASD compared to controls.

6.2.2. Case-control analysis of variants

In an effort to identify any variants associated with risk or protection for ASD and ID, a candidate gene set (30 DNA and RNA methylation genes) case-control analysis was performed akin to the methodologies used in GWAS statistical analysis. A total of 32,987 variants across all methylation genes remained after removal of non-shared variants (variants that contained missing data in at least one cohort) and erroneous genotypes (such as cases where the reference and alternative allele have been flipped during genetic imputation), as well as the removal of any supplementary variants which were added from the imputation panel. For the statistical analysis,

variants were retained that were present in all of the cohort datasets (shared variants) and missense variants, totalling 1422 variants. Table 6.2 presents a breakdown of these 1422 variants across the methylation genes.

The case-control analysis was run using the Fisher's exact test. The test was applied at two thresholds of genetic imputation and with two methods of aggregating the variants. The two thresholds of imputation were 0.25 (i.e. if 25% or more of the genotype calls in a cohort are imputed, that variant is removed from the analysis) followed by a more stringent threshold of 0.125. The two methods of aggregation were gene-wide (test run on each individual gene before aggregating the results) followed by a more stringent all-gene analysis (test run on all of the genes merged together). The analysis was run using all cases (combined ASD and ID) versus controls before splitting the case cohorts for more detailed disease-specific results.

The case-control analysis began by using the 0.25 threshold and gene-wide aggregation. This resulted in the identification of 14 significant variants, 9 that reached the GWAS level of significance and 5 that reached the suggestive level of significance (Figure 6.2A; red and blue lines respectively). This included 6 missense variants, highlighted in green in the Manhattan plots. This analysis was repeated retaining the 0.25 threshold but applying the all-gene aggregation method. This resulted in the identification of 13 significant variants, 7 that reached the GWAS level of significance and 6 that reached the suggestive level of significance (Figure 6.2B). Both gene-wide and all-gene analyses identified the same group of significant variants, including a missense variant in *UHRF2* which did not reach significance in the latter analysis.

Next, the imputation threshold was increased to 0.125 and carried out the case-control analysis beginning with gene-wide aggregation. This resulted in the identification of 10 significant variants, 6 that reached the GWAS level of significance and 4 that reached the suggestive level of significance (Figure 6.3A). This included 4 missense variants, again highlighted in green in the Manhattan plots. This analysis was repeated retaining the 0.125 threshold but applying the all-gene aggregation method. All 10 variants identified in the gene-level analysis were also identified in the all-gene analysis, with only one of the variants dropping from GWAS to suggestive significance (Figure 6.3B).

TABLE 6.2. Breakdown of variants included in case-control analysis

Gene	Total variants	Missense variants	% missense variants	Missense and shared variants	% missense and shared variants
DNMT1	1409	35	2.48%	143	10.15%
DNMT3A	1857	13	0.70%	66	3.55%
DNMT3B	1114	20	1.80%	87	7.81%
DNMT3L	473	14	2.96%	60	12.68%
MBD1	443	24	5.42%	65	14.67%
MBD3	475	4	0.84%	19	4.00%
UHRF1	1190	0	0.00%	66	5.55%
UHRF2	2104	14	0.67%	42	2.00%
TDG	538	10	1.86%	35	6.51%
NEIL1	331	12	3.63%	31	9.37%
NEIL3	1054	24	2.28%	59	5.60%
WDR76	728	12	1.65%	41	5.63%
THYN1	229	6	2.62%	25	10.92%
TET1	2291	60	2.62%	98	4.28%
TET2	2210	48	2.17%	92	4.16%
TET3	1227	33	2.69%	81	6.60%
GADD45A	137	0	0.00%	0	0.00%
GADD45B	122	4	3.28%	7	5.74%
MBD2	1259	4	0.32%	22	1.75%
MBD4	276	14	5.07%	28	10.14%
ELP3	1758	13	0.74%	47	2.67%
METTL3	357	5	1.40%	38	10.64%
METTL14	468	8	1.71%	36	7.69%
WTAP	686	6	0.87%	29	4.23%
YTHDF2	644	9	1.40%	23	3.57%
YTHDF3	763	8	1.05%	30	3.93%
HNRNAP2B1	574	2	0.35%	55	9.58%
EVAVL1	909	3	0.33%	20	2.20%
ALKBH5	626	5	0.80%	20	3.19%
FTO	6735	24	0.36%	57	0.85%
TOTAL:	32,987	434	-	1422	-

Shared variants = variants that are present in all of the cohort datasets

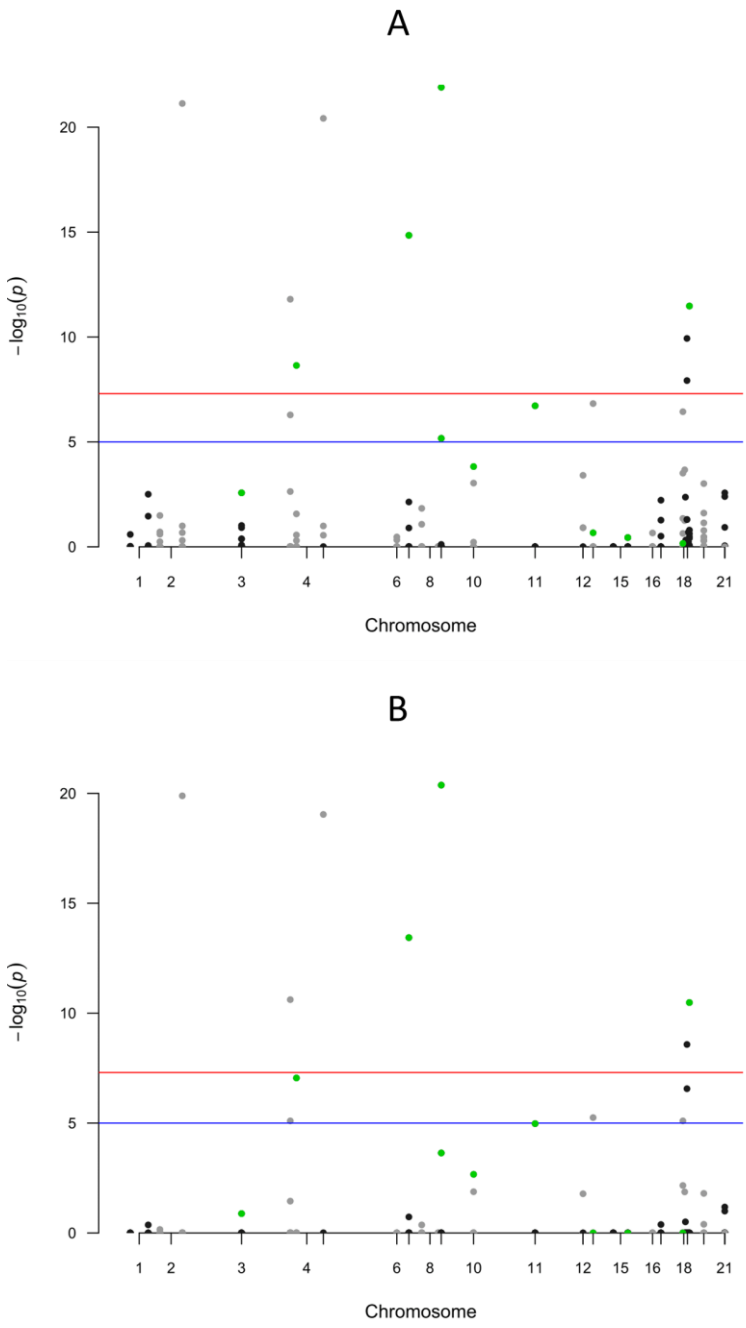


FIGURE 6.2. Manhattan plots for case-control analysis with 0.25 threshold. The red lines indicate GWAS significance and the blue lines indicate suggestive significance. The green dots represent missense variants. Data presented for gene-wide (A) and all-gene (B) aggregation of variants.

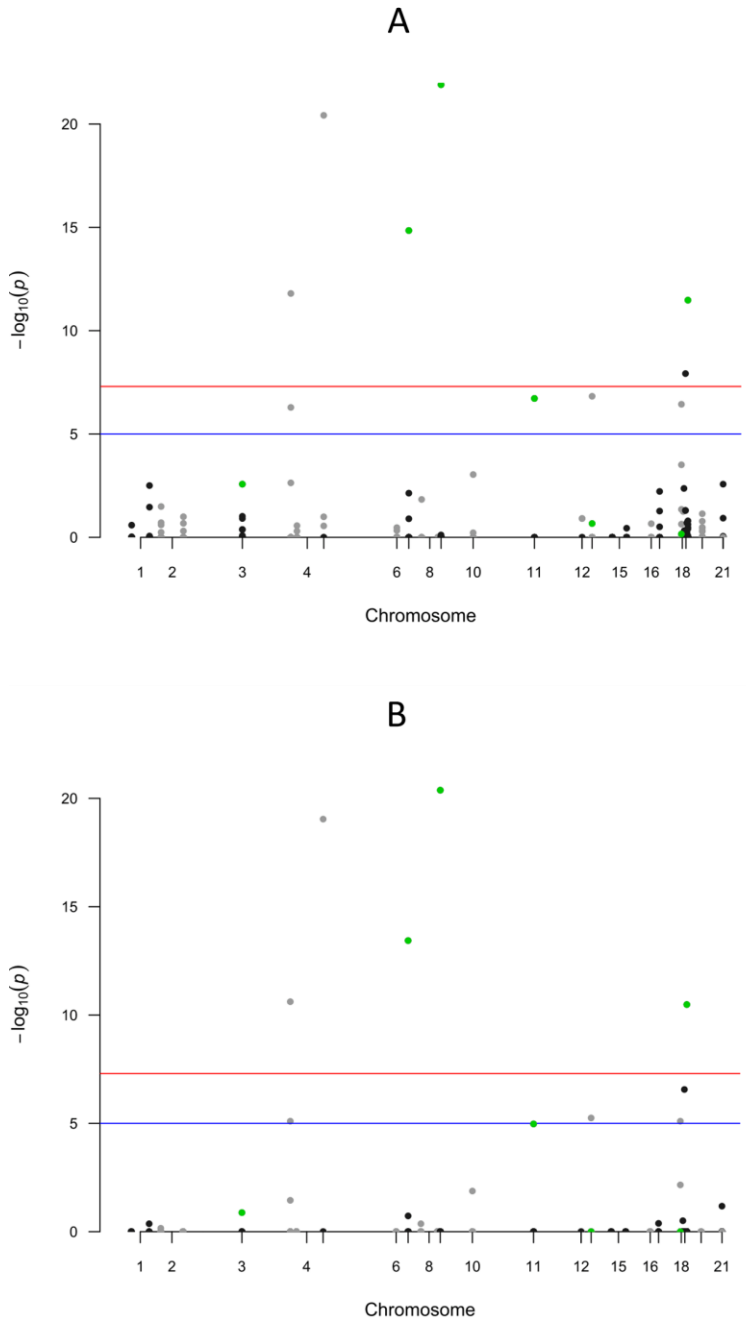


FIGURE 6.3. Manhattan plots for case-control analysis with 0.125 threshold. The red lines indicate GWAS significance and the blue lines indicate suggestive significance. The green dots represent missense variants. Data presented for gene-wide (A) and all-gene (B) aggregation of variants.

The results of the most stringent analysis were retained, namely the application of the 0.125 threshold and all-gene aggregation, which are presented in Table 6.3. Of the 10 variants that were significant in this case-control analysis, 8 were found more in the disease cohorts (risk) and 2 were found more in the control cohort (protective). 8 were found in the DNA methylation genes and 2 in the RNA methylation genes. The majority of these were methylation readers (6), with the remainder made up of 2 writers and two variants within the DNA eraser *TET2*. Alongside the 4 missense variants, an intronic duplication was identified, a substitution which could influence exon splicing, a non-coding exonic duplication, a substitution in a gene-flanking region, a synonymous variant, and a deletion. Assessment of the genetic imputation quality showed that 3 of the 10 variants (in *UHRF2, HNRNPA2B1,* and *DNMT1*) had undergone 100% control imputation. However, all 3 were found only in the ID cohorts and not in the ASD cohorts, suggesting that these variants are associated with specific risk for ID.

The 4 variants that were significant at the 0.25 threshold (bottom of Table 6.3) were also assessed. This list included 3 DNA methylation variants and 1 RNA methylation variant. The DNA readers *UHRF1* and *UHRF2* were again present in this list, alongside a DNA eraser and RNA writer. The variant in *TET3* was found more in the controls (protective) with the remainder being found more in the cases (risk). In addition to the 2 missense variants, an insertion and a non-coding substitution was identified. However, investigation of the genetic imputation quality showed that both missense variants had undergone 100% control imputation. As these missense variants were found in both the ID and ASD controls, a case cannot be made for these variants being disease-specific. This increases the likelihood that they would also be present in the controls had the data not required imputation, and therefore these variants were disregarded from the analysis.

SIFT and PolyPhen annotation were subsequently used to further characterise the influence of these variants on protein function. 3 of the 4 missense variants that were significant at the 0.125 threshold were annotated as functionally damaging by either SIFT or PolyPhen (Table 6.4). All of these variants had previously been identified as evidenced by their rs accession numbers, although the change in *THYN1* is more commonly an Arginine to Threonine substitution as opposed to the Arginine to Proline substitution observed in our cohorts. Interestingly, 3 of the 4 variants were Arginine substitutions. Arginine residues are involved in creating hydrogen bonds, key to protein structural stability, and have been shown to be concentrated in

particular genes associated with particular phenotypes (Luleyap et al., 2006). More importantly, 3 of these variants were previously found only in the ID cohorts. Assessment of the 4th variant revealed that it was also found only in the ID cohorts and not in the ASD cohorts, again suggesting the identification of a group of variants associated with specific risk for ID. Due to this, the distribution of these variants was assessed within our two ID cohorts and the majority were present in the Muir cohort compared to the Rare FIND, with the *HNRNPA2B1* variant present only in Muir (Appendix 28). This suggests that these variants could be more closely associated with individuals who have learning disabilities and comorbid psychiatric disease. However, the information needed to confirm the precise clinical characteristics of the individuals carrying these variants was not available.

6.2.3. Burden analysis of variants

Having performed a case-control analysis for each of the 1422 variants across the 30 DNA and RNA methylation genes, the burden of variants was examined within each gene and the relationship of this burden with disease. As with our previous analysis, the various burden tests were performed at the 0.25 threshold of genetic imputation before applying the more stringent 0.125 threshold.

Simple burden and adaptive burden tests were conducted at the 0.25 threshold, but the results of these tests were difficult to validate as they did not return detailed *p* values (Appendix 29). Only the KBAC adaptive burden test of rare missense variants returned accurate estimates of genetic burden.

TABLE 6.3. Significant variants identified in case-control analysis

Gene	Category	Class	Change	*p* value	Direction
UHRF2	DNA reader	Missense	c.1925G>T	4.19E-21	Risk
NEIL3	DNA reader	Intronic	c.413+10dupA	9.06E-20	Protective
HNRNPA2B1	RNA reader	Missense	c.1048A>G	3.65E-14	Risk
TET2	DNA eraser	Sequence feature	c.4182+119C>A	2.42E-11	Risk
DNMT1	DNA writer	Missense	c.1064G>C	3.28E-11	Risk
UHRF1	DNA reader	Non-coding exonic	n.2552dupC	2.73E-07	Risk
METTL3	RNA writer	Intronic	c.*3749G>A	5.63E-06	Protective
TET2	DNA eraser	Synonymous	c.4203T>C	7.94E-06	Risk
MBD1	DNA reader	Sequence feature	c.111-125delG	7.94E-06	Risk
THYN1	DNA reader	Missense	c.296G>C	1.06E-05	Risk
TET3	DNA eraser	Upstream variant	c.-67_-66insC	7.37E-22	Protective
UHRF1	DNA reader	Intronic	n.2394+56C>G	1.17E-10	Risk
METTL14	RNA writer	Missense	c.511C>A	2.27E-09	Risk
UHRF2	DNA reader	Missense	c.2222A>C	6.74E-06	Risk

TABLE 6.4. Further details regarding significant missense variants

Gene	Change	AA change	Known	SIFT	PolyPhen
UHRF2	c.1925G>T	p.Cys432Phe	rs79300900	Deleterious	Prob. dam.
HNRNPA2B1	c.1048A>G	p.Arg350Gly	rs117917826	Tolerated	Benign
DNMT1	c.1064G>C	p.Arg355Pro	rs201945078	Tolerated	Prob. dam.
THYN1	c.296G>C	p.Arg99Pro	rs372232379	Deleterious	Prob. dam.

6.2.3.1. Variance-component and combined tests

Variance-component analysis of rare missense variants was then conducted at the 0.25 threshold. In comparison to simple and adaptive burden tests, variance-component tests also take into account the distribution of variants within a genetic region. Of the 28 DNA and RNA methylation genes that remained for burden analysis (*UHRF1* and *GADD45A* contained no missense variants following quality control), 5 reached GWAS significance (Table 6.5). These were *DNMT1, UHRF2, THYN1, HNRNPA2B1,* and *METTL14*. Interestingly, all 5 of these genes contained at least one rare missense variant that was highly significant in the previous case-control analysis (*METTL14* contained a variant discarded after the 0.25 threshold analysis for having 100% control imputation). In addition, the level of significance correlated between the variance-component tests and the case-control analysis, with *UHRF2* being most significant and *THYN1* being least significant. Application of combined variance-component and burden tests, which combine the power of burden tests with the sensitivity of variance-component tests, resulted in the same 5 genes reaching GWAS significance (Table 6.5). These findings suggest that the burden analysis could be driven by single variants with large effects as opposed to the accumulation of many variants with small effects.

Extension of the variance-component analysis to all rare variants again returned the same 5 genes at either GWAS or suggestive significance, along with *MBD1* and *TET2*. Further inclusion of common variants alongside rare missense variants also returned the same 5 genes (minus *DNMT1*), along with *MBD1, TET2,* and *TET3*. These three additional genes also contained at least one variant that was highly significant in the case-control analysis, providing further evidence that the burden analysis appears to be driven by individual rare variants with large genetic effects.

TABLE 6.5. *p* values from combined variance-component and burden tests

	0.25 threshold		0.125 threshold	
Gene	Variance-component	Combined variance-component and burden	Variance-component	Combined variance-component and burden
DNMT1	6.06E-14**	2.66E-13**	6.2E-14**	2.91E-13**
DNMT3A	0.112813	0.033345	0.110171	0.010206
DNMT3B	0.037565	0.047009	0.038654	0.05348
DNMT3L	0.194045	0.245947	0.208274	0.312856
MBD1	0.098537	0.001002	0.241051	0.016593
MBD3	0.10618	0.023286	0.10618	0.023286
UHRF1	-	-	-	-
UHRF2	4.04E-28**	2.83E-27**	2.55E-26**	1.79E-25**
TDG	0.011238	0.018081	0.011071	0.017323
NEIL1	0.061138	0.084281	0.170481	0.272229
NEIL3	0.368139	0.529283	0.364966	0.524317
WDR76	0.141014	0.107906	0.141014	0.107906
THYN1	9.55E-09**	1.17E-09**	9.55E-09**	1.17E-09**
TET1	0.001402	5.17E-04	0.11196	0.047281
TET2	0.497043	0.369823	0.493132	0.33176
TET3	0.024871	9.25E-04	0.025822	0.001592
GADD45A	-	-	-	-
GADD45B	0.071128	0.036502	0.071128	0.036502
MBD2	0.060373	0.087608	0.060373	0.087608
MBD4	0.006033	0.004109	0.005906	0.002774
ELP3	0.031954	0.058997	0.028763	0.044704
METTL3	0.002676	0.001331	0.002676	0.001331
METTL14	2.43E-11**	1.70E-10**	2.43E-11**	1.70E-10**
WTAP	0.918271	1	0.918271	1
YTHDF2	0.011999	4.10E-05	0.011999	4.10E-05
YTHDF3	0.043605	0.037129	0.043605	0.037129
HNRNAP2B1	5.32E-17**	1.38E-16**	5.32E-17**	1.38E-16**
EVAVL1	-	-	-	-
ALKBH5	0.305042	0.131406	0.305042	0.131406
FTO	0.105337	0.179362	0.104058	0.17696

** = significant at GWAS significance levels ($p < 5 \times 10^{-8}$)

Finally, the common and rare variant analysis were re-run but without the application of the 0.25 threshold, thus including all flagged variants included in our data. Of the 28 remaining DNA and RNA methylation genes, 19 reached GWAS or suggestive significance. Notably, *GADD45B* reached the plainly spurious *p* value of 1.1E-111 despite having not reached significance in any of the previous burden or case-control analyses. This analysis demonstrates the importance of applying stringent quality control measures within an NGS processing protocol, as well as revisiting the data to ensure that erroneous variants are flagged before statistical analysis is applied. The results for all of these variance-component and combined analyses are presented in Appendix 30.

Next, the variance-component and combined variance-component and burden analysis of rare missense variants was repeated, but with the application of the more stringent 0.125 threshold. As with the analysis at the 0.25 threshold, *DNMT1, UHRF2, HNRNPA2B1, THYN1,* and *METTL14* all reached GWAS significance (Table 5). Extension to all rare variants again returned these 5 genes (minus *METTL14*), along with *MBD1* and *TET2*. Finally, the addition of common variants to the rare missense variants returned these 5 genes (minus *DNMT1*), along with *TET2*. The loss of *METTL14* in the all rare tests, as well as *MBD1* and *TET3* in the common and rare tests, could be because the driving effect of any single significant variants had been masked by the addition of further non-significant rare or common variants.

6.2.3.2. Case cohort specific analysis

To assess the specific relationships between genetic burden in these DNA and RNA methylation genes and our disease cohorts, the variance-component and combined variance-component and burden analysis were repeated at the 0.125 threshold in the ID and ASD cohorts separately. Analysis of the rare missense variants revealed that *DNMT1*, *UHRF2, THYN1,* and *HNRNPA2B1* all reached GWAS significance in the ID cohorts but not in the ASD cohorts. Similarly, *METTL14* was the only variant which reached significance for the ASD cohorts but did not reach significance in the ID cohorts. This corroborates our earlier discovery of disease-specific variants in the case-control analysis.

In the ID cohorts, extension of the variance-component analysis to all rare variants identified the same 4 genes, with the addition of *MBD1*. Moreover, the inclusion of common variants alongside rare missense variants again identified these 4 genes.

For the ASD cohorts, extension to all rare variants resulted in *DNMT3B* and *TET2* reaching suggestive significance, whilst inclusion of common variants alongside rare missense variants identified *METTL14* and *TET2*. These results provide strong evidence that genetic variation in methylation genes confers disease-specific risk. In addition, the complete burden analyses suggest that burden of rare variants in DNA and RNA methylation genes is particularly associated with risk for ID in comparison to ASD.

6.3. Discussion

In the last three experimental chapters of this study, a number of relationships have been demonstrated between a genetic variant in a DNA methyltransferase, *DNMT3L* R278G, and cognitive performance in both MCI and control cohorts. Our findings show that concentrating on particular measures of cognitive performance can elucidate the contribution of otherwise-overlooked variants. The focus of this chapter was to characterise the presence of genetic variants within a candidate gene set of DNA and RNA methylation genes using next-generation sequencing (NGS) data from individuals with autism-spectrum disorder (ASD), individuals with moderate to severe intellectual disability (ID) and ID comorbid with psychiatric disease, as well as general population controls.

A NGS pipeline was designed to retrieve, annotate, process, and analyse data concerning nearly 60,000 variants from 30 DNA and RNA methylation genes, with the emphasis of our analysis being on rare damaging missense variants that may not have been identified using classic GWAS array approaches. Upon first assessment of rare and low-frequency variants across all 30 methylation genes, a pattern of increased rare damaging missense variants was discovered in the RNA readers in our ID and ASD cohorts. This contrasted with the higher proportion of rare damaging variants observed in the control cohort in all other methylation categories, highlighting a potentially important class of RNA methylation genes with respect to risk for ASD and ID.

The number of variants was subsequently filtered down to leave only the missense variants and variants that were shared across our control, ASD, and ID cohorts, which resulted in a total of 1422 variant positions. Single variants analysis on controls versus all cases revealed 10 variants that reached suggestive or GWAS significance thresholds. This included 4 missense variants that were found in the ID cohorts but not in the ASD cohorts. Application of burden analysis techniques also resulted in the identification of the same 4 genes that harboured these significant missense variants. Moreover, case-specific analysis confirmed that these 4 genes were significantly associated with ID but not ASD. Thus, a pattern of genetic risk has been identified in specific DNA and RNA methylation genes that is associated with rare moderate to severe ID and ID comorbid with psychiatric disease as opposed to ASD.

6.3.1. RNA readers in ID and ASD

Our examination of rare and low-frequency variants in controls, ID, and ASD showed a higher proportion of rare missense variants were identified in controls compared to cases. In contrast, a higher proportion of low-frequency missense variants were identified in cases than controls. Further, the number of rare variants within the total number of variants was not higher in controls, reducing the likelihood that this is a biased observation. This pattern could be an example of the increased efficiency of whole-genome sequencing (WGS) used in the control cohorts compared with whole-exome sequencing (WES) used in the case cohorts. Studies of rare variants using WGS and WES have shown that, whilst both techniques identify the vast majority of exonic mutations, WGS could detect further coding positions that were missed in WES (Belkadi et al., 2015). This could be related to the sensitivity of WES to reduced read depth (Meynert et al., 2014).

After examining the functional categories (readers, writers, erasers) for DNA and RNA methylation genes, it was discovered that most categories showed a greater proportion of rare and rare damaging missense variants from the total number of missense variants in controls than the cases. However, RNA readers showed the opposite pattern, with a greater proportion of rare damaging missense variants in the cases compared to controls. This functional category includes the genes *YTHDF2, YTHDF3, HNRNPA2B1,* and *ELAVL1*.. *YTHDF2* and *YTHDF3* have been implicated in N^6-methyladenosine (m^6A) methylation-mediated mRNA decay (Shi et al., 2017, Wang et al., 2014a). *HNRNPA2B1* has been associated with m^6A methylation-mediated alternative splicing and *ELAVL1* has been associated with m^6A methylation mediated mRNA stability (Alarcon et al., 2015, Wang et al., 2014b).

The study of RNA methylation and cognition is in its infancy compared to that of DNA methylation. Few studies have been carried out into the individual roles of RNA readers and there have been no investigations into mutations within these genes for neurodevelopmental or neuropsychiatric disease. With regards to *YTHDF2*, the few studies of mutations in this gene have provided evidence for an association with cancer risk. For example, work has shown that mutations and copy number variations in a number of RNA methylation genes (including *YTHDF2*) were significantly linked to risk of acute myeloid leukemia, whilst another study has discovered chromosomal translocations affecting *YTHDF2* present in this disease (Kwok et al., 2017, Nguyen et al., 2006). Research has also shown that increased

homozygosity of a particular polymorphic region (QM376-400) in *YTHDF2* was associated with longer lifespan (Cardelli et al., 2006).

In comparison, *YTHDF3* has only been investigated in relation to multiple sclerosis and no polymorphisms showed an association with this disease (Varade et al., 2012). Both YTHDF2 and YTDHF3 have been shown to bind exclusively to the m6A modification (Dominissini et al., 2012). *In silico* structural techniques have elucidated the key residues involved in m^6A RNA binding for these proteins, describing the creation of an "aromatic cage" (made up of three aromatic amino acids) around the m^6A nucleotide to facilitate direct interaction with this mark (Li et al., 2014, Zhu et al., 2014). This highlights a potential mechanism where genetic variants in *YTHDF2* or *YTHDF3* could disrupt the aromatic cage required for m^6A binding and subsequently influence RNA methylation patterns in disease.

There has been greater interest in the disease risk associated with *HNRNPA2B1* due to its involvement in cancer and elements of multisystem proteinopathy (MSP), which involves body myopathy, Paget disease of bone, frontotemporal dementia, and amyotrophic lateral sclerosis. Beginning with cancer, HNRNPA2B1 facilitates epithelial-mesenchymal transition in pancreatic cancer, key in the initiation of the cancer invasion-metastasis cascade, whilst down-regulation of HNRNPA2B1 promotes the apoptotic removal of pancreatic carcinoma cells (Barcelo et al., 2014, Chen et al., 2011, Dai et al., 2017). Moreover, HNRNPA2B1 overexpression has been demonstrated in breast and lung cancer tissue (Dowling et al., 2015, Hu et al., 2017). With regards to MSP, patient population studies have struggled to identify disease-associated mutations in *HNRNPA2B1*, despite their presence in family studies of MSP, with the authors affirming that these mutations were rare and required greater sample sizes (Benatar et al., 2013, Le Ber et al., 2014, Seelen et al., 2014). Promisingly, mutations such as *HNRNPA2B1* P310L have been reported in patients with the MSP-constituent diseases Paget disease of bone and amyotrophic lateral sclerosis (Qi et al., 2017, Soong et al., 2014). Aside from its interactions in key disease pathways such as those highlighted in the cancer literature, HNRNPA2B1 contains a prion-like domain that is prone to pathological behaviour such as stress granule formation when mutated (Kim et al., 2013). Finally, the HNRNPA2B1 can bind to m^6A and regulate alternative splicing. Recent work has linked this to disease pathogenesis, with mutations in *HNRNPA2B1* shown to splice out an exon of the D-amino acid oxidase gene associated with ALS (Martinez et al., 2016).

As exemplified by its name, *ELAVL1* (also known as *HuR*) or embryonic lethal, abnormal vision, drosophila-like 1, is necessary for embryonic development due to its role in the posttranscriptional regulation of gene expression associated with morphogenesis and ontogeny (Katsanou et al., 2009). Polymorphisms in *ELAVL1* have linked to risk for subtypes of skin cancer but not breast cancer (Su et al., 2017, Upadhyay et al., 2013). Antagonisation of particular mi-RNA binding patterns and alternative splicing of key genes by ELAVL1 has been associated with pathological angiogenesis (Chang et al., 2014, Lu et al., 2014). Similarly, ELAVL1 appears to stabilise and promote expression of mRNA transcripts linked to the development of acute respiratory distress syndrome (Hoffman et al., 2017). Further studies have raised questions over the role of ELAVL1 in alternative splicing, but *in vitro* research has confirmed its role in regulating translation alongside HNRNPA2B1 (Uren et al., 2011, Zhang et al., 2017). Interestingly, ELAVL1 is necessary for spermatogenesis as it controls the translation of spermatogenic regulators, and a relationship with spermatogenesis has also been documented for DNMT3L (Chi et al., 2011). Moreover, folate deficiency has been shown to reduce the expression of *ELAVL1* and correlated with both azoospermia and low sperm count, providing further evidence for a link between B vitamins and methylation genes (Yuan et al., 2017).

Evidence has been presented for the involvement of RNA reader proteins in a number of diseases, but can any relationships between RNA methylation and mRNA processing in ID or ASD be deciphered? Beginning with ID, mutations within a tRNA methyltransferase gene, *NSUN2*, have been reported to cause autosomal-recessive forms of ID (Khan et al., 2012). Molecular work has shown that *NSUN2* mutations lead to a loss of tRNA methylation and subsequent abnormal RNA processing, whilst models of *NSUN2*-related ID have rescued some phenotypic deficits by increasing expression of the protein (Abbasi-Moheb et al., 2012, Hussain et al., 2013). Mutations in other genes associated with RNA methylation such as *TCOF1* can also result in particular forms of ID (Gonzales et al., 2005). As discussed, RNA readers are associated with methylation-mediated alternative splicing of mRNA or regulation of mRNA stability. Research has illustrated that impaired function of mRNA alternative splicing effectors such as PQBP1 can cause X-linked mental retardation (Kalscheuer et al., 2003). Indeed, further investigation has confirmed that the loss of PQBP1 is associated with diminished dendritic growth (Wang et al., 2013). Furthermore, dysregulation of mRNA stability has been outlined in multiple forms of ID (Bassell and Warren, 2008, Kuechler et al., 2015).

For ASD, there has been no research concerning RNA methylation and associated genes despite work showing differential DNA methylation patterns in ASD compared to controls (Ladd-Acosta et al., 2014). However, work outside of RNA methylation has alluded to the involvement of mRNA alternative splicing and stability regulation. For example, disruption of the RNA binding protein RBFOX1 is observed in ASD and this protein has been associated with alternative splicing events in various ASD risk genes (Ray et al., 2013, Weyn-Vanhentenryck et al., 2014). The primary genetic risk factor associated with the intellectual disability Rett syndrome, MECP2, appears to influence alternative splicing and thus a role in ASD has been posited (Young et al., 2005, Zappella et al., 2003). Similarly, the causative gene *FMRP* for the intellectual disability Fragile-X syndrome is also associated with risk for ASD, with disrupted FMRP shown to result in dysregulated mRNA translation and subsequent deficits in synaptic function (Steinberg and Webber, 2013, Zalfa et al., 2003).

It could be proposed that RNA methylation and regulation of mRNA translation is important for cognitive function and brain disease and may relate to the synaptic tagging hypothesis. This hypothesis concerns the role of local protein synthesis during LTP or LTD in synapses and how these processes are organised in these tagged synapses (Frey and Morris, 1997). Work has shown that local protein synthesis is necessary for the long-term strengthening of synapses and that translational control of RNA is key to the dynamics of this (Martin et al., 1997). Needless to say, there are far too many proteins involved to cover here but the RNA translation aspects have been well reviewed by Richter and Klann (2009). However, the discovery of NSUN2 and FMRP localisation in neuronal dendrites is of particular note to our findings (Hussain and Bashir, 2015). With regards to RNA methylation, m^6A levels have been associated with memory formation in rodent studies (Widagdo et al., 2016). Thus, it is possible that RNA readers could influence the course of ID and ASD through alterations in mRNA translation at the synaptic level and beyond.

6.3.2. DNA methylation missense variants and intellectual disability

From our single variant, variance-component, and combined variance-component burden analysis, 4 genes each containing a rare missense mutation were identified as significantly associated with ID. These genes were *UHRF2*, *HNRNPA2B1*, *DNMT1*, and *THYN1*. *HNRNPA2B1*, the only RNA methylation gene, has already been discussed and so the ensuing discussion will concentrate on the three remaining DNA methylation genes.

Beginning with *DNMT1*, the main function of this gene is well known as it encodes the maintenance methyltransferase within humans (Bestor et al., 1988). Although this role has been established for a long time, insights into the details of DNMT1 activity are still being made. For instance, DNMT1 has been shown to strongly favour hemimethylated CpG sites over unmethylated when carrying out maintenance methylation, with autoinhibitory mechanisms in specific domains of DNMT1 ensuring that unmethylated CpGs are obstructed from interacting with the target recognition domain (Hermann et al., 2004, Song et al., 2011). As outlined in our *in silico* modelling presented in chapter four, mutations in *DNMT1* have been shown to cause a neurodegenerative condition, Hereditary Sensory and Autonomic Neuropathy type 1 (HSAN1) (Klein et al., 2011, Smets et al., 2017). Mutations in *DNMT1* also cause other neurological conditions such as autosomal dominant Cerebellar Ataxia, Deafness, and Narcolepsy (ADCA-DN) (Pedroso et al., 2013, Winkelmann et al., 2012). Interestingly, DNMT1 methylation contributes to pathological gene silencing events in cancer cells independently of the other DNA methyltransferases, highlighting the possible disease risks associated with maintenance versus *de novo* methylation (Robert et al., 2003).

UHRF2 encodes for a ubiquitin-protein ligase, a family most commonly associated with protein degradation. UHRF2 has been shown to bind to 5hmC as well as promote TET oxidation of 5mC into 5hmC (Spruijt et al., 2013, Zhou et al., 2014). UHRF2 also plays a role in linking DNA methylation and histone dynamics, interacting with histone deacetylases, DNA methyltransferases, methylated histone H3, and methylated DNA (Mori et al., 2012, Pichler et al., 2011). This protein also has a number of functions aside from its role in DNA methylation. UHRF2 is involved in the facilitation of DNA damage repair as well as acting as a sumo ligase, a post-translational modification involved in a range of cellular functions (Luo et al., 2013, Oh and Chung, 2013). The majority of studies investigating *UHRF2* in disease have been specifically in cancer, where disruption of UHRF2 function influences cancer cell growth (Lu and Hallstrom, 2013, Wu et al., 2012). However, key mutations in Parkinson's disease have been associated with *UHRF2* dysregulation (Reinhardt et al., 2013). Meanwhile, mice carrying a *UHRF2* genetic knockout demonstrate seizures, abnormal brain activity, and alterated 5mC patterns (Liu et al., 2017).

Finally, *THYN1* (also known as *THYN28*) has been identified as a reader of 5hmC methylation in a proteomic study of DNA methylation binding partners (Spruijt et al., 2013). However, little further work has been carried out involving this gene. Analysis

of the crystal structure of THYN1 revealed domains normally associated with RNA binding (Yu et al., 2009). Findings from disease related studies have highlighted unusual *THYN1* expression in cancer patients and a deletion of chromosome 11 (including *THYN1*) in two cases of Jacobsen syndrome, a rare developmental disability that can include symptoms of ID (Papasotiriou et al., 2017, Sheth et al., 2014).

How might differences in 5mC levels associated with disruption in *DNMT1* or altered reading of 5hmC levels associated with disruption to *UHRF2* and *THYN1* relate to ID pathology? To begin with, the pathogenesis of a number of important ID subtypes involves changes in DNA methylation. Fragile-X syndrome is caused by a lack of FMRP protein, needed for normal development, due to a CGG repeat expansion in the *FMR1* gene. However, an established body of work has shown that abnormal hypermethylation of the *FMR1* promoter is responsible for the resultant lack of gene transcription due to the loss of a DNA methylation boundary upstream of *FMR1* (Naumann et al., 2009, Sutcliffe et al., 1992). Not only can this abnormal methylation identify Fragile-X syndrome mutants before birth, but the degree of impairment associated with the disease is correlated with the level of methylation present (Devys et al., 1992, McConkie-Rosell et al., 1993). Another two examples are Angelman and Prader-Willi syndromes, both associated with disruption of chromosome 15q11-q13 and both disorders of genetic imprinting, as a cluster of imprinting genes is located in this chromosomal region (Perk et al., 2002). Angelman and Prader-Willi syndromes show distinct patterns of DNA methylation in this region, posited as imprinting regulation (Driscoll et al., 1992, Zeschnigk et al., 1997). As DNMT1 is necessary for genetic imprinting with mutations in *DNMT1* interfering with imprinting patterns, this provides a tangible link between genetic variants in *DNMT1* and changes in DNA methylation associated with ID (Hirasawa et al., 2008, Howell et al., 2001).

There is comparatively less work looking into the role of 5hmC and any associated readers of this modification in ID. However, some insights can be drawn from studies of Fragile-X syndrome and Rett syndrome. Beginning with Fragile-X syndrome, alongside the aforementioned changes in 5mC in *FMR1*, increases in 5hmC levels has been observed in the *FMR1* promoter as well as neuron-specific gains across the *FMR1* gene (Esanov et al., 2016, Yao et al., 2014). In turn, differentially hydroxymethylated regions observed during neuronal development are enriched in ID risk genes regulated by the FMRP protein (Wang et al., 2012). Rett syndrome is a rare form of ID that can be caused by mutations in the DNA methylation reader

MECP2. MECP2 has been shown to bind to both 5mC and 5hmC with the latter being enriched in actively transcribed genes. Moreover, disease mutations in *MECP2* can specifically inhibit binding with 5hmC (Mellen et al., 2012). MECP2 deficits are correlated with increasing levels of 5hmC, with MECP2 able to halt the conversion of 5mC to 5hmC by TET proteins (Szulwach et al., 2011). As MECP2 is involved in transcriptional activation and repression, it could be that many downstream processes following 5hmC binding are involved in ID pathology, drawing particular parallels with the multifunctional UHRF2 identified in our analysis.

An interesting additional point linking the work from this chapter with the previous three experimental chapters can be found in studies of another neurodevelopmental disease, Down's syndrome. Differential methylation patterns have been observed in Down's syndrome individuals compared to controls, particularly in studies of lymphocytes (Kerkel et al., 2010). Moreover, polymorphisms in *MTHFR*, which encodes an enzyme involved in the remethylation of Hcy to methionine, have been consistently associated with risk for this disease (Hobbs et al., 2000, Scala et al., 2006). As one may expect, Down's syndrome individuals with these polymorphisms and differential methylation patterns also presented with decreased levels of Hcy, methionine, SAM, SAH, and glutathione, whilst treatment with folinic acid and methyl-B12 improved the health of cultured cells from these individuals (Al-Gazali et al., 2001, Pogribna et al., 2001). Thus, the relationship between one-carbon cycle components such as B vitamins and Hcy, DNA methylation, and genetic variants associated with methylation predicted to underline our findings in MCI and control individuals has also been observed in ID.

6.3.3. Pipeline of next-generation sequencing analysis

The development of a NGS pipeline to process and analyse data from discrepant sequencing technologies and multiple cohorts was a particular challenge, but not a unique one. Recent large-scale analyses of sequencing data for over 200,000 individuals from different sequencing facilities using different platforms have highlighted similar downstream analysis issues (Auer et al., 2016). There are undoubtedly more established automated NGS pipelines that could eliminate some of issues with the processing steps outlined in our methods. However, the innate biases within some of the most popular commercial and open-source NGS pipeline packages raise questions over the validity of inputting data into a "black-box" pipeline, where the user may not understand the processes involved and the issues

to be considered when interpreting any subsequent genetic analysis (Hwang et al., 2015). With this in mind, I will look to discuss some of the issues associated with our NGS pipeline and how these could inform similar analyses in the future.

In our search for an effective variant annotation tool, the use of the Variant Effect Predictor (VEP) from Ensembl was decided upon to provide functional prediction scores and minor allele frequency (MAF) information from the Exome Aggregation Consortium (ExAC) database. When inputting our genetic variant data, the cDNA change associated with the canonical gene transcript was used. However, given the focus of this study is on cognitive function and brain disease, it may be prudent to also consider analysing brain-specific gene transcripts (McCarthy et al., 2014). This could be carried out using data from the Genotype-Tissue Expression (GTEx) consortium that has collected data from 54 areas in the human body (Consortium, 2015).

Regarding MAF annotation, the size of the ExAC database (>60,000 individuals) and breadth of variant coverage from this database was a positive. Questions could be raised over our application of SIFT and PolyPhen scores to categorise variants as being functionally damaging. As outlined in similar NGS analysis, there is no gold standard in the field of functional prediction (Auer et al., 2016). There have been criticisms of the specificity of these programs in dealing with gain-of-function mutations in particular (Flanagan et al., 2010). Other tools such as MutationTaster and FATHMM appear to offer more accurate annotation, but their gene or genome-wide data cannot be downloaded for in-house annotation and is not included in generic annotation tools such as VEP (Schwarz et al., 2014, Shihab et al., 2013). Moreover, uploading our VCF files to tools such as MutationTaster was prohibited. Therefore, inclusion of more accurate functional prediction scores into our study would require a considerable time commitment.

The most contentious element of our NGS pipeline was the use of genetic imputation. The merging of multiple cohort datasets from differing sequencing platforms meant that gaps in the data for particular variants in particular cohorts were inevitable. To limit the effect of this, our analysis (and imputation) focused only on rare and low-frequency missense variants. Although imputation successfully filled in these gaps, there appeared to be issues in estimating homozygotes as only wild-type calls were properly imputed. Whilst this is not a major issue for the interpretation of rare variants, it would be more difficult to assess common variants using this method.

As outlined in the methods, the flipping of the reference alleles in the VCF files was resolved by inputting our variant details into the required legend files. It was suspected that this could influence imputation accuracy, but re-running imputation with an original legend file did not alter the pattern of wild-type estimations.

A possible resolution may lie in the choice of imputation tool and imputation panel utilised in this chapter, as other studies have reported heterozygotic imputation using other programs and multiple reference panels, increasing the pool of haplotype information available (Kreiner-Moller et al., 2015). Nevertheless, it is also possible that further optimisation of these genetic imputation tools is required within our NGS pipeline. Thresholds were applied to the imputation to reduce the likelihood of false positive signals emerging from our analysis. However, a few significant variants were retained that were entirely imputed in the controls because they were present in only one of our case cohorts. Whilst this supports the identification of disease-specific genetic variance in our data, one must also accept that replication in a control cohort with adequate sequencing coverage would be necessary to validate these findings.

To comprehensively examine the effect of genetic variation in our DNA and RNA methylation genes, single variant statistical testing and burden analysis approaches were applied. Whilst the single variant testing was successful, the inclusion of burden analysis tools proved to be more difficult. RVTESTS and PLINK/SEQ could not properly parse imputed VCF files and the results from the CMC, VT, and KBAC tests were unreliable. SKAT, arguably the most popular burden analysis package in the field at present and previously used in similar analyses using cohort data, appeared to produce more accurate results (Consortium et al., 2015). However, SKAT identified significant burden of genetic variants in genes that included a highly significant variant identified in the single variant testing. It is possible that our sample size lacked the statistical power necessary to assess true genetic burden, meaning that only highly significant variants were identified by the SKAT tests. Our combined case-control population was under 3000 individuals, likely to afford low statistical power when investigating rare variants (Moutsianas et al., 2015). Adapted methods such as backward elimination SKAT (BE-SKAT) could be used to confirm whether our highly significant variants were driving our burden analysis results (Lin, 2016).

7. Distribution of DNA methylation modifications in rodent and human brain tissue

7.1. Preface

In order to develop our research beyond the analysis of genetic and cognitive data, immunohistochemical fluorescence techniques were used to visualise and characterise intermediate methylation modifications, 5-hydroxymethylcytosine (5hmC) and 5-carboxylcytosine (5caC), in brain tissue. Rodent brain tissue was initially used to optimise staining with antibodies for the methylation modifications as well as optimisation of neuronal, glial, and neurogenesis markers. Human brain sections from elderly control individuals were then assessed for the distribution of methylation modifications in key regions of the brain.

7.2. Results

7.2.1. Rodent neuronal staining

In order to characterise the methylation marks 5hmC and 5caC, a neuronal cell marker was included during our optimisation. This was to enable one to identify neuronal-specific methylation staining patterns as it is difficult to identify specific cell populations from the nuclear fluorescent signal alone. In addition, neuronal marker reference staining allowed one to identify any inaccurate or unexpected staining patterns.

NeuN was chosen as our neuronal reference marker because of its wide use in the literature, making it arguably the most common fluorescence biomarker of mature adult neurons in a variety of species and tissues. Optimisation of the neuronal reference staining was first performed in rodent tissue obtained from the University of Nottingham animal house. Rodent tissue was used for the optimisation because of the comparative worth and scarcity of our human tissue samples. Initial staining with NeuN using our standard immunohistochemistry protocol was unsuccessful, as evidenced by the staining of the rodent midbrain in Figure 7.1. DAPI staining confirmed the presence of cells within the midbrain tissue, but no NeuN signal was picked up. Increasing the magnification to 63x only showed dim background signal, but no neuronal specific staining.

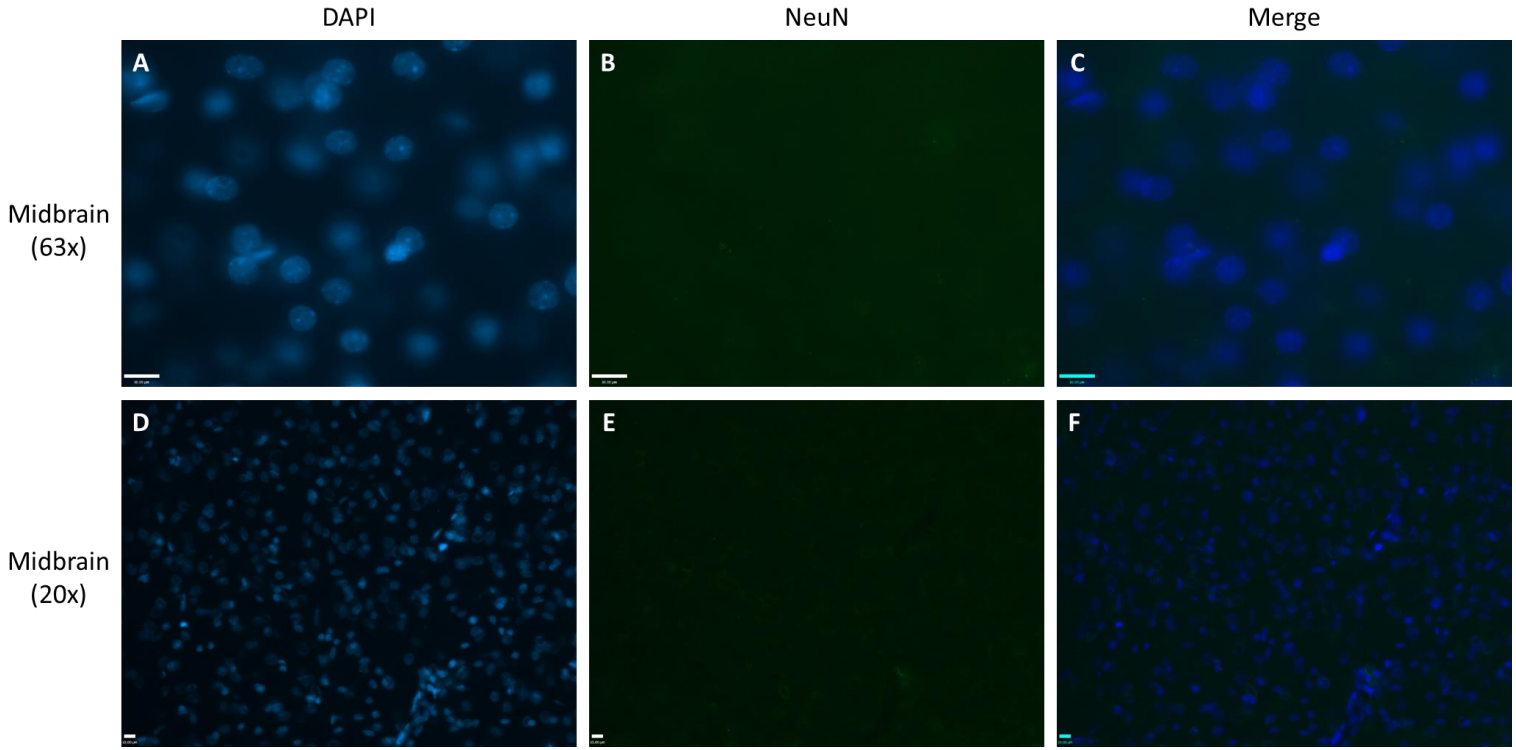


FIGURE 7.1. DAPI and NeuN staining in rodent midbrain at 63x and 20x magnification on the DM-IRE2. Scale bars at both 63x and 20x magnification indicate 10μm in length. DAPI is represented by blue and NeuN by green. Whilst DAPI appeared clear at both resolutions (A & D), no NeuN staining could be seen (B & E).

Subsequent attempts were made to optimise the standard protocol and rectify the lack of NeuN staining. Our previous attempts had not included Triton X-100 in the blocking buffer and had included the use of rockers to keep the blocking, primary antibody, and secondary antibody solutions moving during each incubation step. The latter could have caused issues during the incubation steps as the hydrophobic barriers holding the solutions over the tissue are sensitive and could have been compromised during rocking. The addition of Triton X-100 to the blocking solution and removal of rocking during incubations resulted in successful staining of NeuN in the rodent tissue (Figure 7.2). The staining of neuronal cells appeared to be specific as evidenced by the lack of NeuN staining in the white matter tracts alongside the left ventricle and in the areas surrounding the hippocampal dentate gyrus and CA regions. Some background staining was evident, but did not compromise the specificity of the neuronal staining.

7.2.2. Rodent 5hmC and 5caC

With the standard protocol consistently providing immunoreactivity for neuronal cells, an adapted protocol was used to perform immunofluorescence using antibodies specific for the 5hmC and 5caC methylation modifications. With the adapted protocol, a form of antigen retrieval is required in the form of hydrochloric acid treatment, along with a signal amplification step after the secondary antibody has been applied. This is because the detection of 5caC had proved difficult with the available antibodies. However, the work of our colleague, Alexey Ruzov, had identified this adapted protocol as a successful method of detecting 5caC signal in brain tissue. The 5hmC and 5caC antibodies were chosen because of their previous inclusion in the work of our colleague, giving confidence that the two marks could be efficiently co-stained with these antibodies.

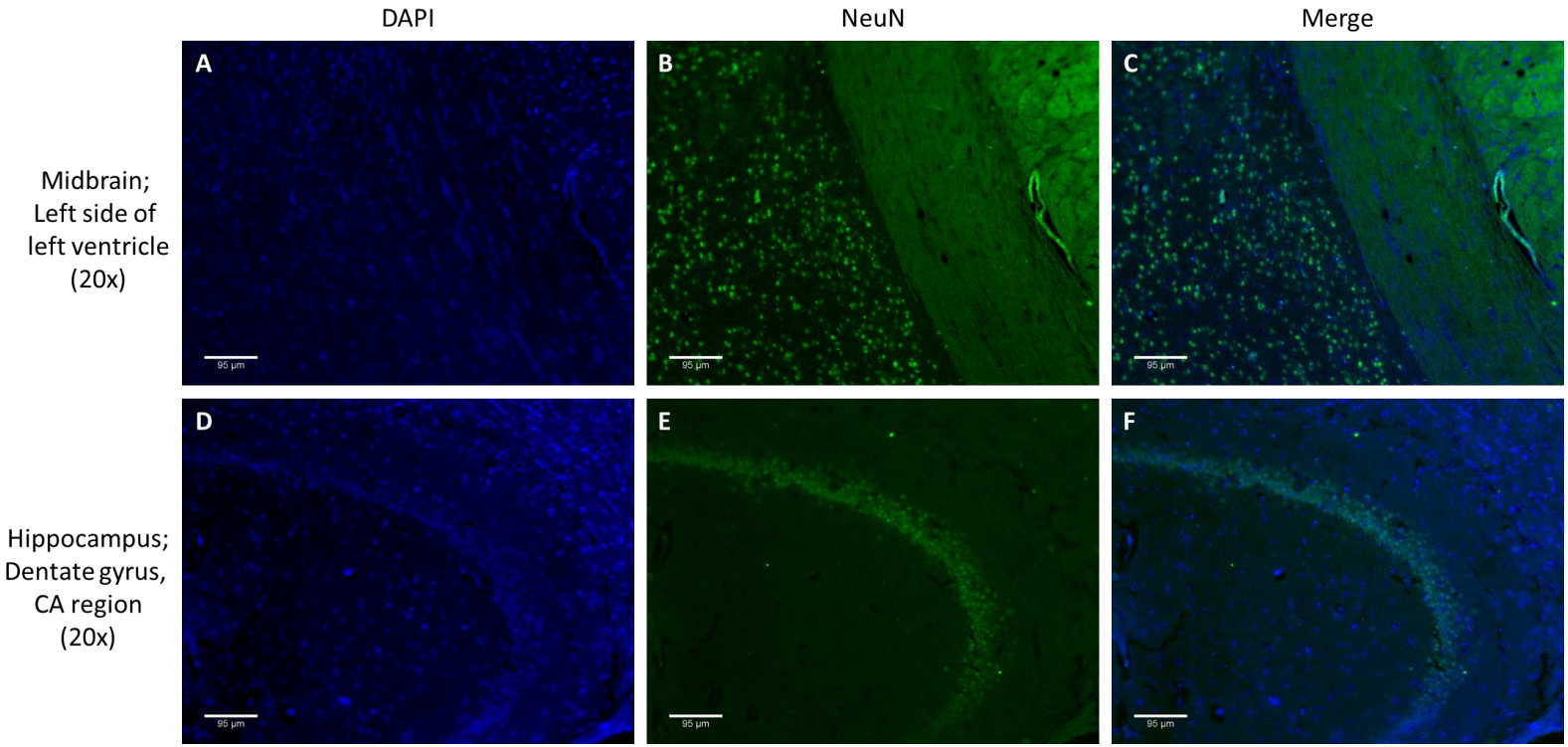


FIGURE 7.2. DAPI and NeuN staining in rodent midbrain and hippocampal regions at 20x magnification on the DM-IRE2. Scale bars indicate 95µm in length. DAPI is represented by blue and NeuN by green. Clear NeuN staining of neuronal cells can be seen in both brain regions (B & E).

Initial optimisation of this adapted protocol focused on optimising the concentration of hydrochloric acid used. The application of hydrochloric acid is required to ensure that the 5caC antibody can efficiently penetrate the tissue but a high molarity of hydrochloric acid also interferes with the DAPI fluorescence included in the mounting media, which could be useful in identifying cell populations during microscopy. Our preliminary efforts used 2M hydrochloric acid in an attempt to balance the 5caC and DAPI staining. However, these efforts were unsuccessful as both signals were not present. This, the original adapted protocol was used which uses 4M hydrochloric acid, concentrating our efforts on the successful staining of 5caC and ignoring the potential staining of DAPI.

Previous studies assessing the levels of 5hmC and 5caC in rodent brain tissue using immuno-staining techniques have shown staining of 5hmC across most developing and adult brain regions (Munzel et al., 2010). As a general conclusion, levels of 5hmC in previous work appeared to be higher in neocortical regions and lower in hindbrain regions. 5caC staining has been reported in a wide range of developing cell and tissue types in rodent but there has been less evidence of 5caC staining in adult rodent brain tissue (Wheldon et al., 2014). Our initial staining of 5hmC and 5caC in adult rodent brain both supported and contrasted with previous findings. In line with earlier work, clear 5hmC staining was observed in the nuclei of cells in all brain regions examined. This is exemplified in layers 1 to 3 of the parietal cortex and in the hippocampus (Figure 7.3). In comparison, weak staining of 5caC was also seen in some areas of the adult rodent brain. As presented in Figure 7.3B & E, 5caC abundance is low but distinguishable from the background signal. This contrasts with earlier studies that failed to successfully stain 5caC in adult rodent brain, confirming that the presence of this modification can be characterised in adult tissue using immunofluorescence staining.

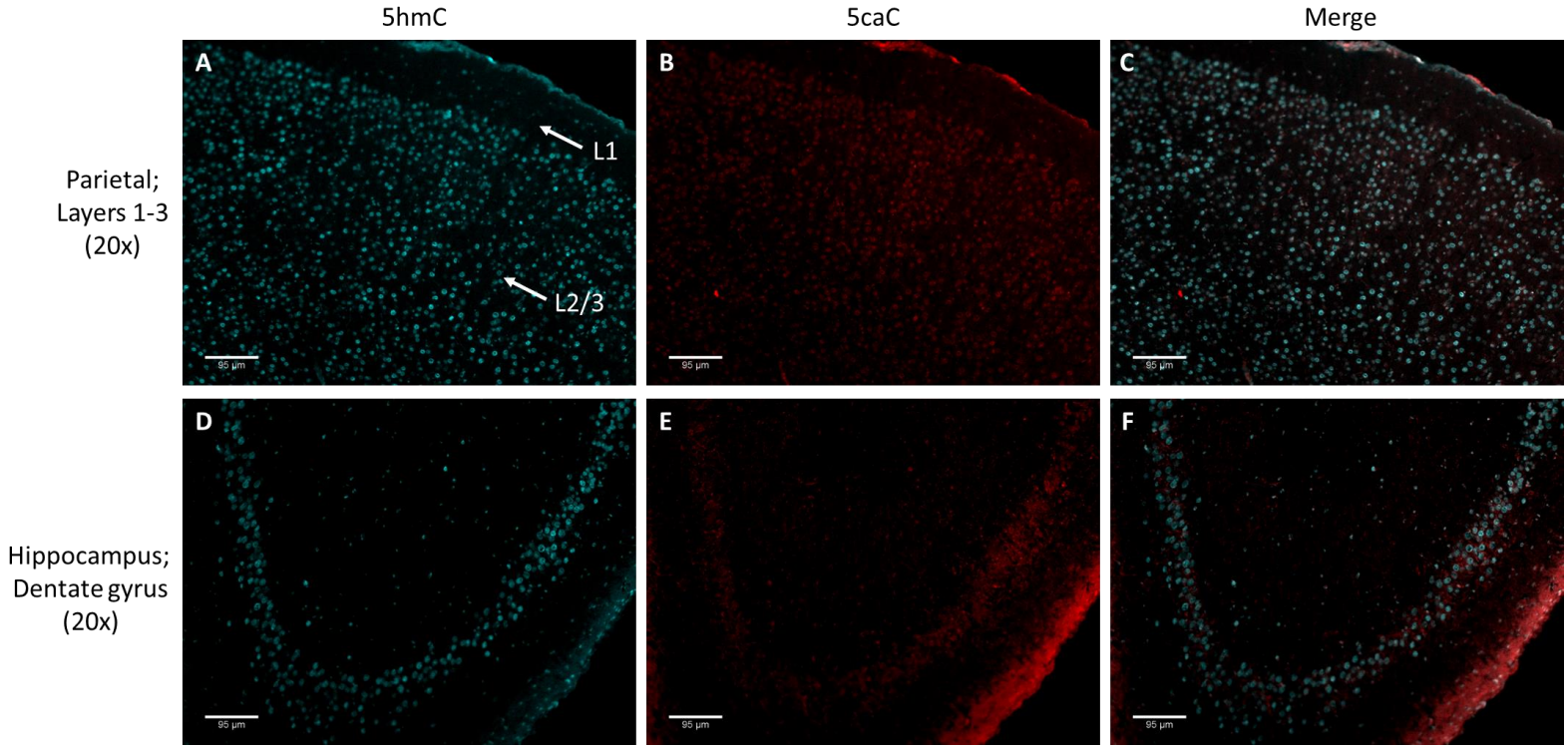


FIGURE 7.3. 5hmC and 5caC staining in rodent parietal cortex and hippocampus at 20x magnification on the DM-IRE2. Scale bars indicate 95µm in length. 5hmC is represented by teal and 5caC by red. Clear staining of 5hmC can be seen in layers 1 to 3 of the parietal cortex and the dentate gyrus of the hippocampus (A & D), while faint 5caC immunoreactivity is also evident (B & E).

Following the success of the adapted protocol on the fluorescence microscope, the adapted protocol was repeated for image analysis on the confocal microscope. For this analysis, serial sections were also included from the same animal to compare NeuN immunostaining with methylation modification distribution. As with the previous analysis using the fluorescence microscope, 5hmC signal was observed clearly in all adult rodent brain regions assessed using the confocal microscope (Figure 7.4). This staining was particularly pronounced in the hippocampal dentate gyrus, a region associated with adult neurogenesis. Indeed, the 5caC signal was also most pronounced in this region, albeit at a much weaker signal intensity. In comparison, no 5caC signal was observed in layers 1 and 2 of the temporal cortex, suggesting that 5caC displays a tissue specific pattern of abundance in comparison to the almost ubiquitous 5hmC staining. Finally, NeuN neuronal reference staining indicates the specificity of 5hmC and 5caC to the granule cells of the dentate gyrus. The comparison within the temporal cortex was more difficult to interpret, appearing to show 5hmC staining in neuronal cells but also in other surrounding cell types. 5hmC has been identified in glial cells as well as neuronal cell types, making this a likely explanation for the additional 5hmC staining observed in the non-neuronal layer 1 of the temporal cortex (Lister et al., 2013).

7.2.3. Rodent GFAP and DCX

The optimisation of immunofluorescence in rodent tissue resulted in successful staining of the methylation modifications 5hmC and 5caC, as well as neuronal NeuN reference staining. When assessing 5hmC and 5caC abundance, there was evidence of non-neuronal 5hmC deposition which was presumed to be evidence of 5hmC staining in glial cells. There was also evidence of staining in the dentate gyrus of the hippocampus, a key area of neurogenesis in the adult brain. This is consistent with previous work using adult rodent tissue which has reported 5caC staining in the subventricular zone, a second area of neurogenesis.

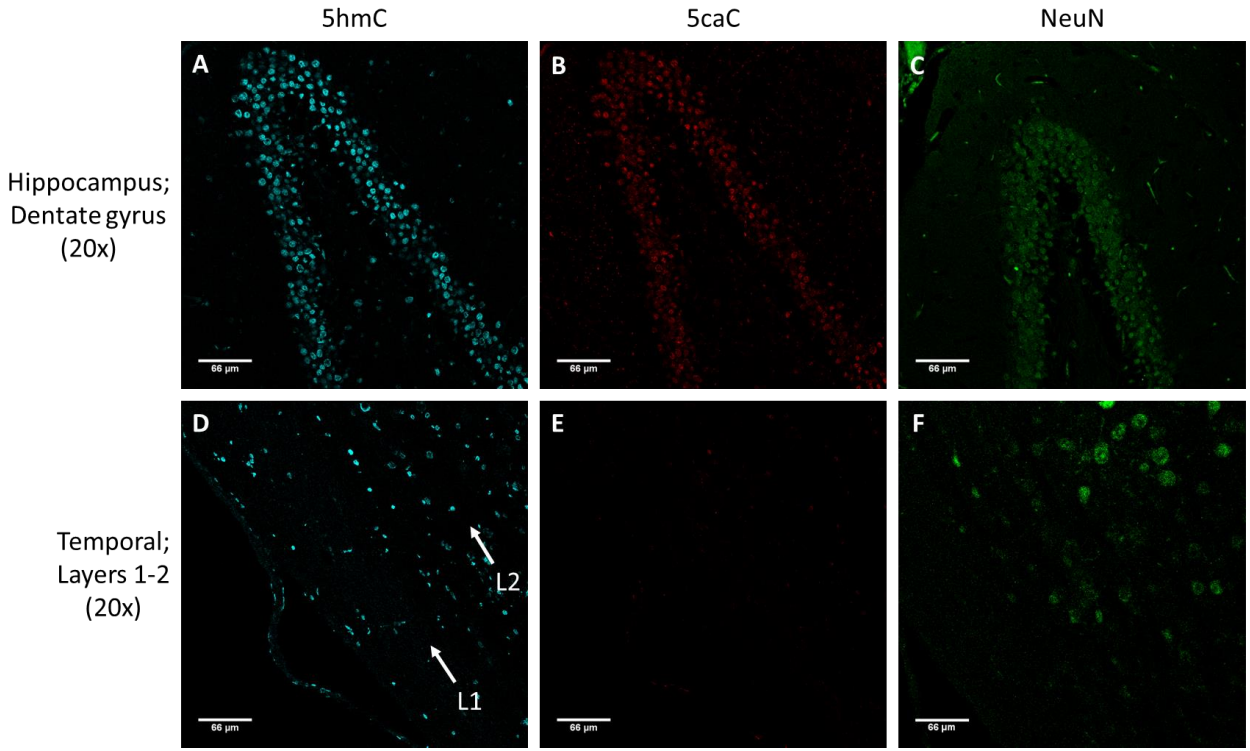


FIGURE 7.4. 5hmC, 5caC, and NeuN staining in rodent hippocampus and temporal cortex at 20x magnification on the LSM-710. Scale bars indicate 66µm in length. 5hmC is represented by teal, 5caC by red, and NeuN by green. Clear staining of 5hmC can be seen in both regions, particularly in the dentate gyrus of the hippocampus (A). Some staining of 5caC also present in hippocampus (B), but not apparent in layers 1 and 2 of the temporal cortex (D). Staining appears to be mostly neuronal specific when comparing with the NeuN staining in the hippocampus (C), but there is additional staining of 5hmC in layer 1 of the temporal cortex, most likely of glial cells.

To investigate whether glial or developing neuronal cells could be co-stained to assess alongside our methylation modifications, the standard immunohistochemistry protocol was used to stain adult rodent brain tissue sections with the GFAP and doublecortin (DCX) antibodies. GFAP was chosen because, much like NeuN, it is the gold standard for assessment of a specific type of glial cell, astrocytes, in the nervous system. Astrocytes are also abundant in the dentate gyrus and can be used as a marker of radial glial stem cells (Figure 7.5). In addition, where NeuN is a marker for mature neurons, DCX is a marker for precursor or immature neuronal cells as well as later stage cell differentiation during adult neurogenesis (Couillard-Despres et al., 2005).

The results indicated a similar pattern in abundance for the methylation modifications 5hmC and 5caC, where 5hmC was clearly observed but 5caC was barely identifiable. Assessment of GFAP showed successful staining of astrocytes cells, as observed by the characteristic star-shaped astrocytic cells (Figure 7.6). In contrast, DCX did not show immunoreactivity in any brain regions, including in areas of potential neurogenesis such as the hippocampus and the subventricular zone (located in the olfactory bulb in the rodent). The DCX antibody thus required further optimisation in rodent brain tissue before it could be successfully implemented in the human tissue. Unfortunately, this lay beyond the scope of the present project and will be continued by subsequent members of the research group.

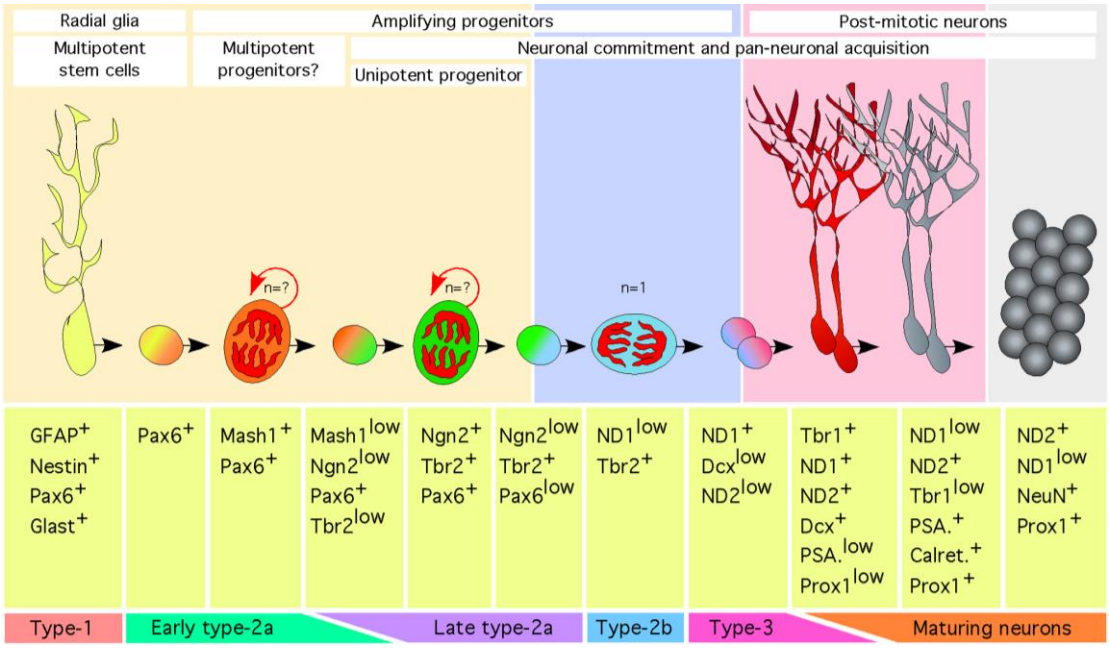


FIGURE 7.5. Depiction of antibodies during the different stages of neurogenesis. NeuN was applied to identify adult neuronal cells, DCX to identify precursor neuronal cells, and GFAP to identify astrocytic glial cells but also as a marker for radial glia stem cells. This image was adapted from Roybon et al. (2009).

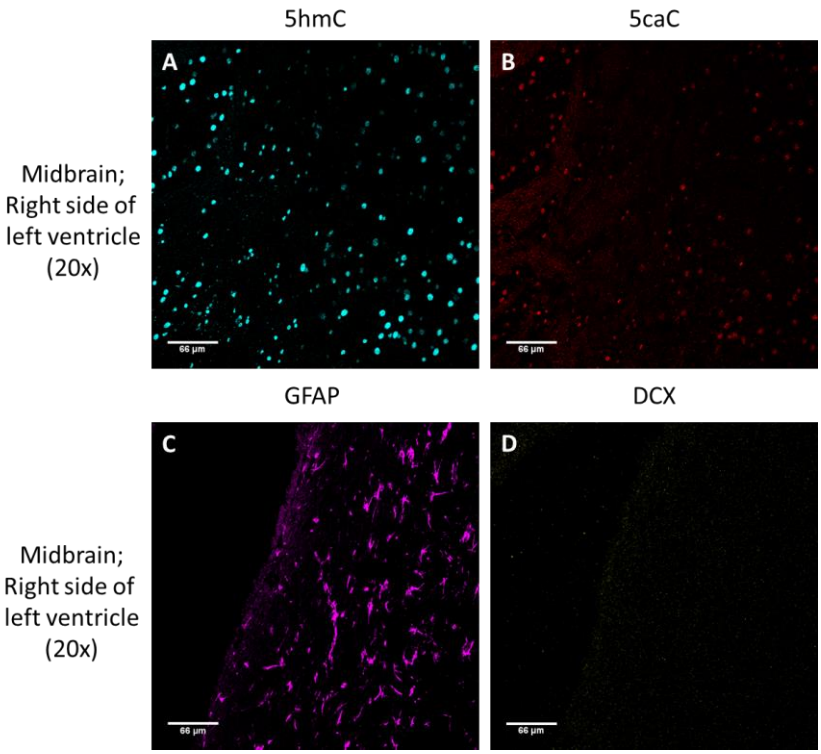


FIGURE 7.6. 5hmC, 5caC, GFAP, and DCX staining in rodent midbrain at 20x magnification on the LSM-710. Scale bars indicate 66µm in length. 5hmC is represented by teal, 5caC by red, GFAP by purple, and DCX by yellow. 5hmC (A) and GFAP (C) signals were clear in this region, whilst 5caC (B) signal was weak and DCX (D) was absent.

5hmC and 5caC abundance in rodent tissue provided a few key insights. Evidence was found to support previous accounts suggesting that 5hmC is widespread in the adult rodent brain, as well as showing particularly high abundance in cells in specific areas such as the hippocampus. Novel insight was provided into the detection of 5caC staining in the rodent brain, albeit with only weak signal intensity in regions such as the hippocampus, which disproves previous studies that failed to find any 5caC signal in the adult brain. Initial characterisation of neuronal specific patterns of 5hmC and 5caC was carried out, providing evidence of brain regions that exhibit differentially lesser and greater non-neuronal abundance. Finally, attempts at co-staining with astrocytic and developing neuronal stem cell antibodies were initiated in order to provide a stronger panel of reference staining to compare with the methylation modifications.

7.2.4. Human 5hmC and 5caC

Following the optimisation studies in adult rodent brain tissue, focus turned to human brain tissue. This tissue was from ten healthy elderly individuals acquired through the Nottingham Health Science Biobank, with six brain regions provided from each individual. As with the literature concerning 5hmC and 5caC patterns in rodent brain, 5hmC presence has been reported throughout the adult human brain. However, in contrast with the rodent studies, 5hmC has been observed at increased levels during ageing in the cerebellum, where 5hmC was reported to be at low levels in the rodent (Kraus et al., 2015). Compared to 5hmC, 5caC has received much less attention in the methylation literature and this extends to immunofluorescence staining studies in human brain. Studies of AD have assessed levels of 5caC, showing them to be comparable to 5mC levels in controls in the entorhinal cortex and cerebellum (Condliffe et al., 2014). However, they did not characterise patterns of abundance in particular cell types or further brain regions. Thus, our study aims to characterise 5hmC and 5caC in human brain at a cellular and regional level.

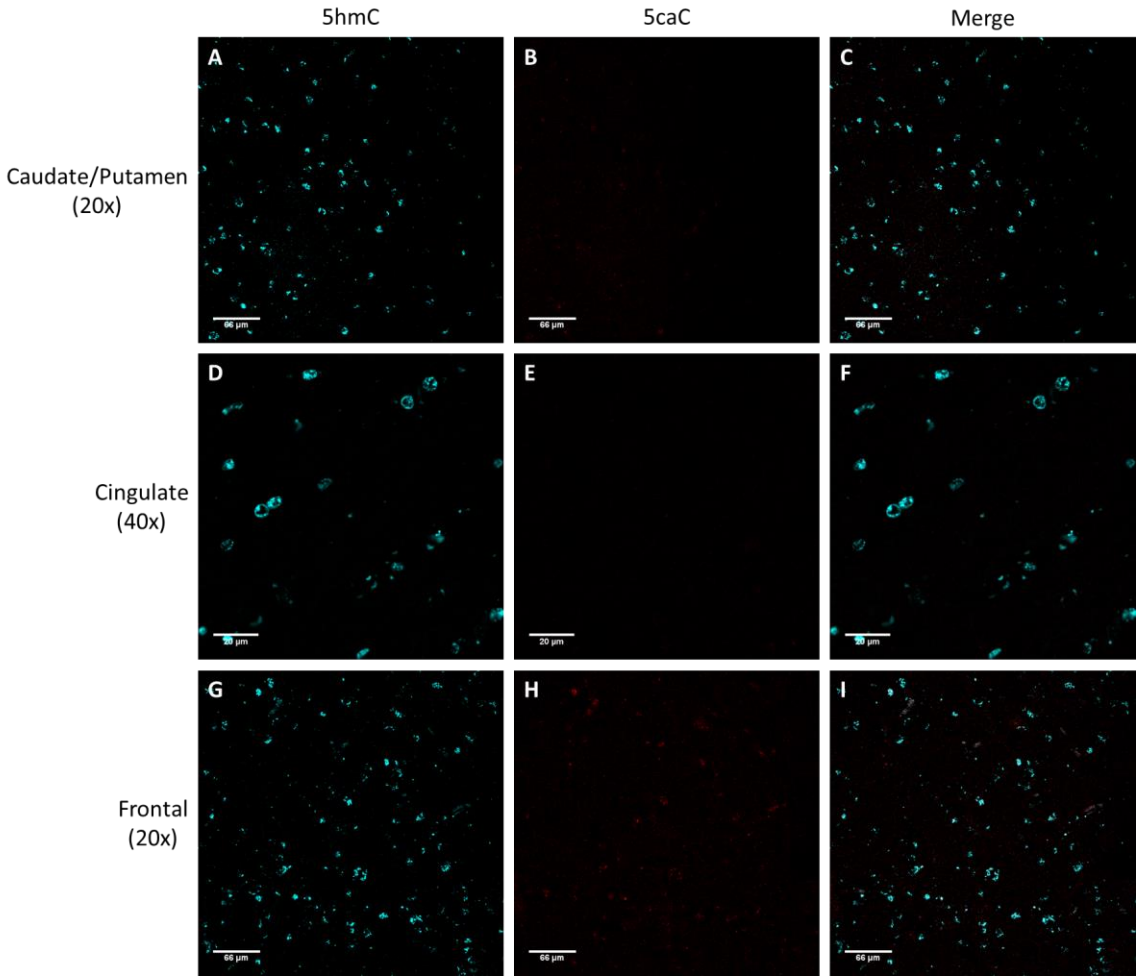


FIGURE 7.7. 5hmC and 5caC staining in all available human brain regions at 20x and 40x magnification on the LSM-710. Scale bars indicate 66µm in length. 5hmC is represented by teal and 5caC by red. In line with the rodent work, 5hmC staining was seen clearly in all brain regions (A, D, G, J, M, P). 5caC was seen faintly in the hippocampus and clearly in the cerebellum (K & Q). The staining in the hippocampus could be of large neuronal or glial cells that are 5caC specific, but further investigation using other reference markers would be required to validate this.

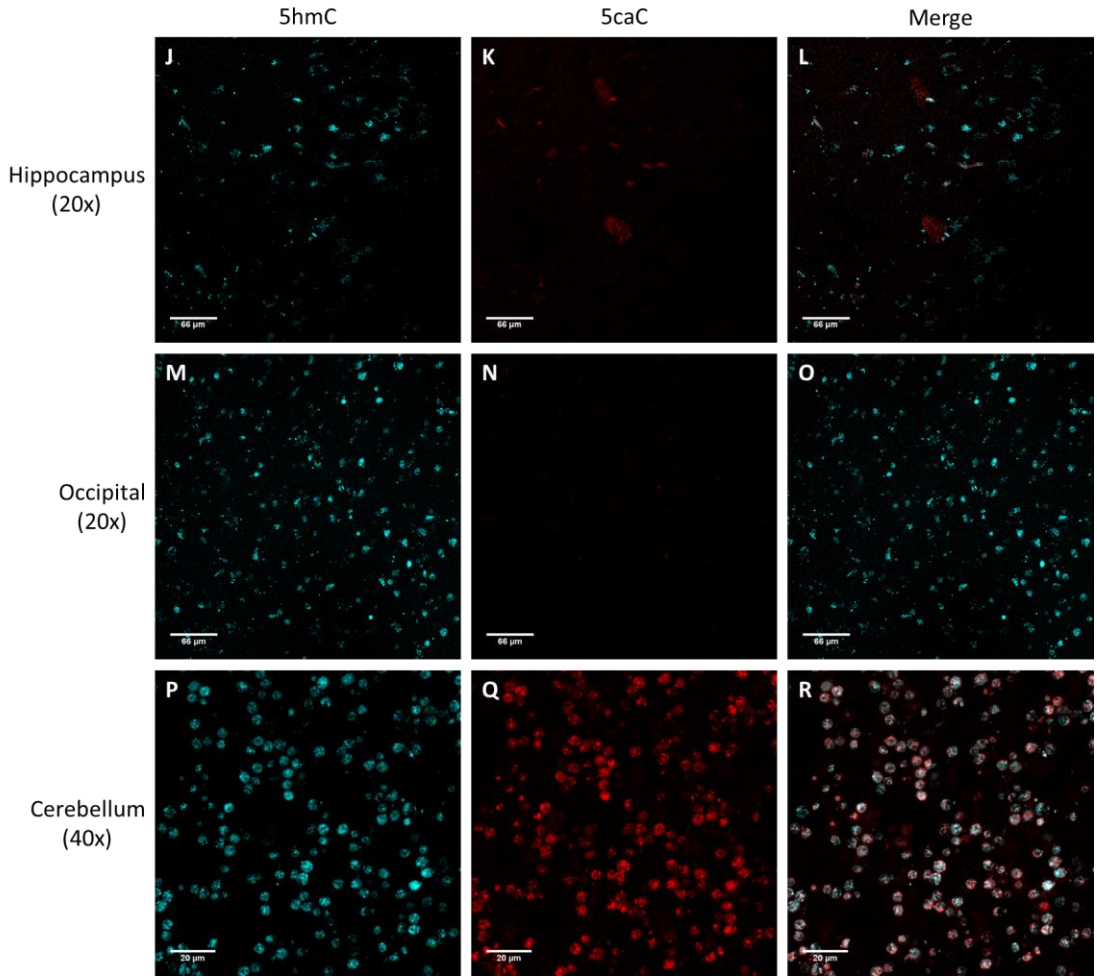


FIGURE 7.7 (continued from page 203).

The adapted immunohistochemistry protocol was performed on tissue from each of the six brain regions provided for the elderly human brain tissue. Assessment of 5hmC staining revealed results similar to those seen in the adult rodent brain tissue and in keeping with published work on 5hmC levels in human brain. Namely, 5hmC signal was observed clearly in the caudate/putamen (which includes the subventricular zone), cingulate gyrus, middle frontal gyrus, hippocampus, occipital cortex, and cerebellum (Figure 7.7). This staining appeared to be particularly strong in the neocortical regions, the middle frontal gyrus and the occipital cortex, and most apparent in the cerebellar granule cells.

Assessment of 5caC staining revealed a striking contrast in abundance between the brain regions. 5caC was virtually absent in the caudate/putamen, cingulate gyrus, middle frontal gyrus, and occipital cortex. In the hippocampus, faint staining of 5caC was observed that could indicate either neuronal or glial cells that are 5caC specific. However, without further reference markers such as NeuN or GFAP, one cannot be sure that these are not faint aberrant signals. The patterns in these five brain areas contrast starkly with the results in the cerebellum, which showed the strongest 5caC signal of any brain region in rodent or human throughout the analyses (Figure 7.7).

7.2.5. Human cerebellum and caudate/putamen

In light of the discovery of clear 5caC staining in the human cerebellum tissue sections, it was decided to perform further staining to confirm and further characterise the patterns of 5hmC and 5caC in this region. As shown in Figure 7.8, immunoreactivity of both 5hmC and 5caC is clearly seen in the granule cells of the cerebellum, with 5hmC staining appeared stronger than 5caC in line with our previous experiments. Interestingly, there appears to be two notable cells stained strongly for 5caC but absent for 5hmC (Figure 7.8B). Their location suggests that these could be Purkinje cells, large neurons that make wide and deep projections into the cerebellar cortex and exert inhibitory signals to modulate cerebellar activity. Without confirmation with a Purkinje cell marker such as PCP4, one cannot definitively confirm the true location of this 5caC signal. Nevertheless, this provides possible evidence of 5caC specific staining of neuronal cells in the human brain.

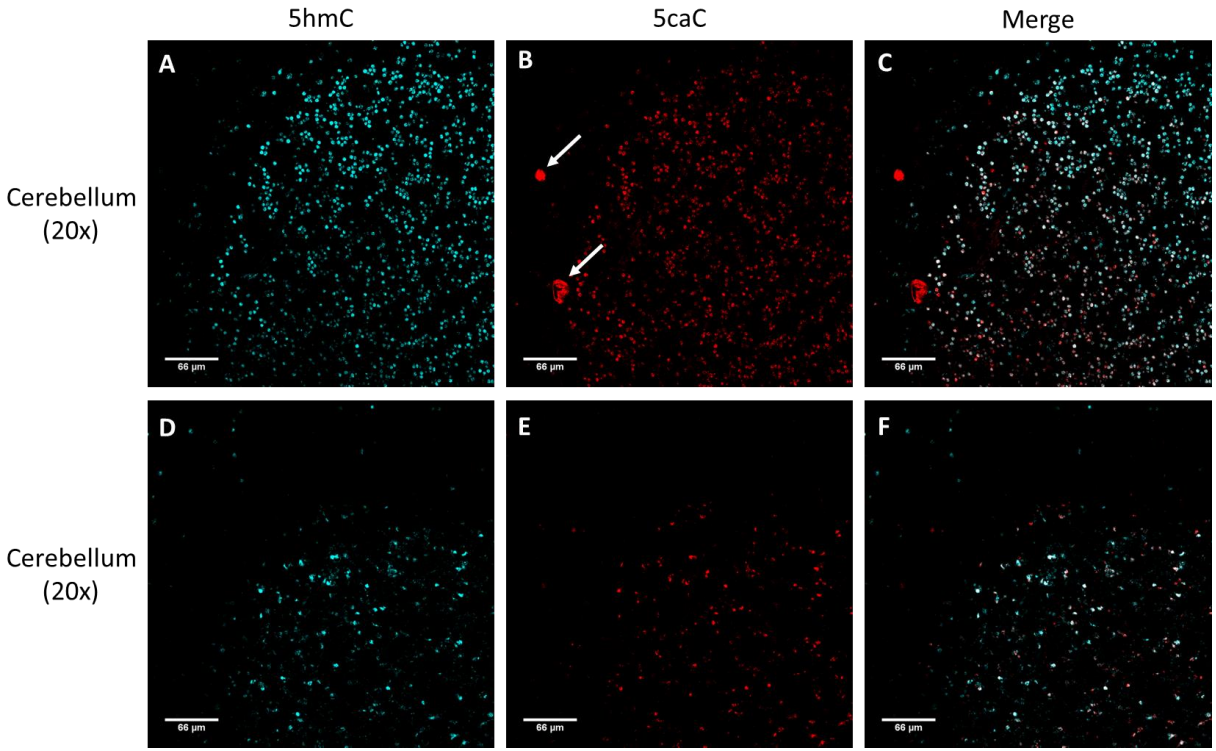


FIGURE 7.8. 5hmC and 5caC staining in human cerebellum at 20x magnification on the LSM-710. Scale bars indicate 66µm in length. 5hmC is represented by teal and 5caC by red. Clear 5hmC and 5caC staining can be observed. Possible staining of Purkinje cells can also be seen for 5caC (B).

Although evidence of clear 5caC staining in the caudate/putamen was not seen during our initial immunohistochemistry, this region is of particular interest as it contain the subventricular zone, one of the few regions in the adult human brain that retains the ability to develop new neuronal cells. Without a properly optimised marker for developing neurons such as DCX, it is impossible to draw accurate conclusions about the location of 5hmC and 5caC staining. In spite of this, repeated immunohistochemistry of the caudate/putamen did reveal areas of 5caC staining which mirrored the already clear 5hmC staining (Figure 7.9). This provides further evidence that 5caC staining can be identified in the human brain, setting the scene for further work to better characterise the location and abundance of 5hmC and 5caC in key brain regions.

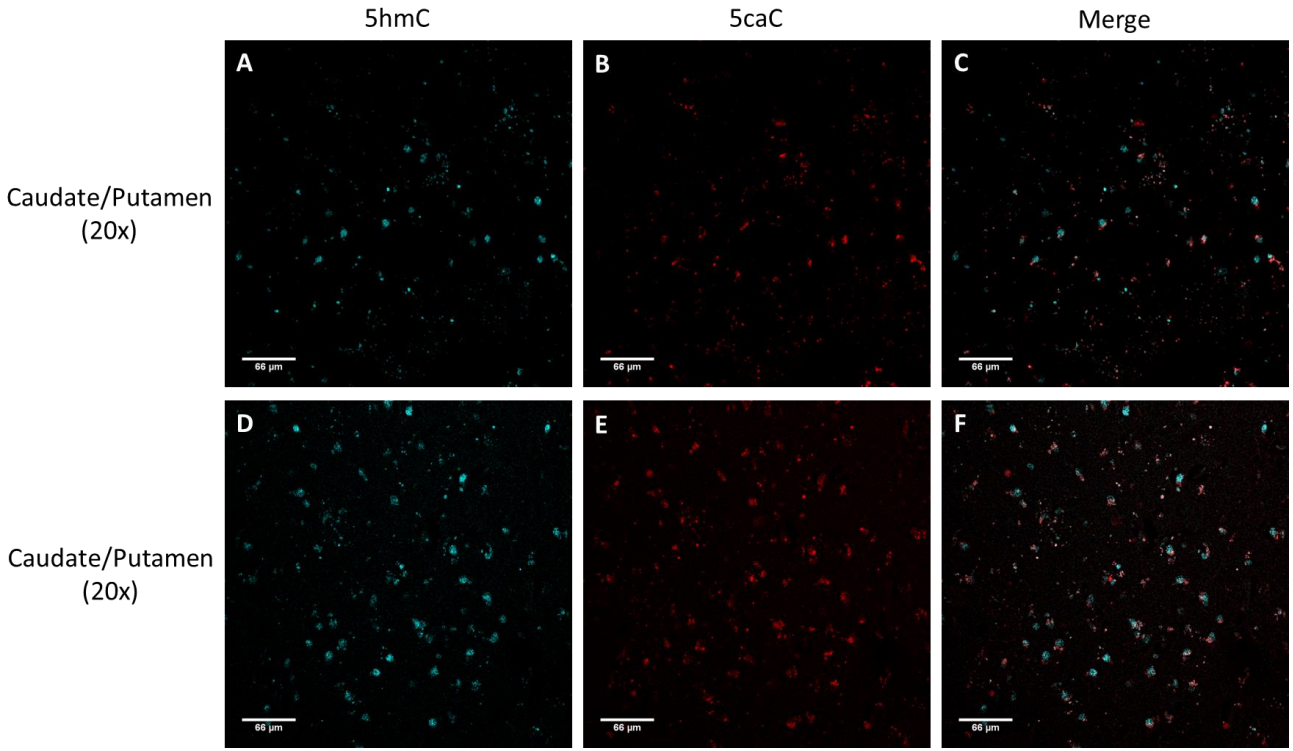


FIGURE 7.9. 5hmC and 5caC staining of the caudate/putamen in human brain at 20x magnification on the LSM-710. Scale bars indicate 66µm in length. 5hmC is represented by teal and 5caC by red. In contrast to our initial staining of this region, some 5caC signal can be observed (B & E).

7.3. Discussion

This thesis has demonstrated a relationship between a missense variant in a DNA methyltransferase, *DNMT3L* R278G, and cognitive performance in MCI and controls, before identifying further methylation genes and missense variants associated with risk for ID and comorbid psychiatric disease in multiple NGS cohorts. The focus of this chapter was to characterise 5hmC and 5caC in human brain tissue and lay the foundations for further assessment in individuals with brain disease. Although this chapter was mainly concerned with the optimisation and characterisation of these modifications, the future potential and utility of this line of work will also be discussed.

Rodent brain tissue was initially used to optimise our two immunohistochemistry protocols. The standard protocol was used to examine neuronal, glial, and stem cell markers in brain tissue as reference stains for our methylation marks. The adapted protocol was used to examine 5hmC and 5caC staining, due to the sensitivity of the 5caC antibody in particular. As expected, consistent 5hmC signal was observed in all of the brain areas stained from both the human and rodent samples. In comparison, little reliable 5caC signal was observed in the majority of rodent and human brain sections. However, faint staining was observed in the hippocampus in both rodent and human, as well as strong 5caC staining in the cerebellum of our human samples. Thus, the immunohistochemistry protocols can successfully characterise methylation modifications in brain tissue and these observations can be combined with key neuronal, glial, and stem cell markers.

7.3.1. Methylation modifications in brain tissue

Consistent 5hmC staining was identified in cell nuclei across rodent and human brain sections, as well as particular colocalisation with neuronal markers and possible colocalisation with astrocytic markers. Although the majority of previous studies looking at 5hmC levels have used next-generation sequencing methodologies rather than immunofluorescent staining, there remains a burgeoning literature about the localisation and function of this modification. For example, research has reported that this methylation modification is enriched in the brain. Studies have estimated that 40% of the DNA methylation in the brain is 5hmC (Irier et al., 2014). When comparing the different brain regions in rodents, 5hmC was found consistently in all of these areas, corroborating our results in both rodent and human brain tissue (Munzel et al., 2010). Further, Munzel et al. (2010) also discovered some interesting

between-region patterns, with 5hmC levels found to be in higher abundance in neocortical regions and the hippocampus but lower in the cerebellum or retinal areas.

Compared with other tissue types, 5hmC is 10 times more likely to be found in neuronal tissue, an enrichment that appears to be conserved between species such as amphibians and mammals (Almeida et al., 2012, Sun et al., 2014). Notable studies have investigated cell-type specific 5hmC distribution in the cerebellum and observed patterns of 5hmC in Purkinje cells, granular cells, and Bergmann glia that differed from 5mC or DAPI signal. Specifically, high levels of 5hmC were observed prominently in Purkinje cells, to a lesser extent in granular cells, and lesser still Bergmann glia (Mellen et al., 2012, Szulwach et al., 2011). The authors suggested that these high 5hmC levels in the cerebellum could indicate the presence of 5caC, but that they could not accurately detect this mark. Our results would substantiate their claims, as comparable levels of 5caC were observed in the cerebellum to that of 5hmC, positing a tissue-specific role for this further demethylation mark. In addition, the observation of 5hmC in glial cell types could offer support for our observation of non-neuronal 5hmC staining in human brain sections.

Further research has aimed to understand more about the functions of 5hmC and how the deposition of this mark may change under certain healthy and pathological conditions. 5hmC levels have been shown to increase in the hippocampus with normal aging, but also in response to certain stress conditions (Chouliaras et al., 2013, Li et al., 2015). With regards to pathology, one of the most studied diseases for 5hmC levels is AD. Nevertheless, the results from these studies are mixed at best. Decreased levels of 5hmC across the AD brain have been observed, supported by a negative correlation between 5hmC deposition and the Braak score of neurofibrillary tangle pathology (Chouliaras et al., 2013, Condliffe et al., 2014). However, contrasting increased levels of 5hmC in the frontal and temporal areas of the AD brain have also been reported (Coppieters et al., 2014). Finally, as one may expect, studies have also shown no change in 5mC or 5hmC levels in AD (Lashley et al., 2015). Some of this discrepancy is likely down to the different techniques used by each study to measure 5hmC levels, whether that be immunofluorescent staining or sequencing, highlighting the caution one must take when interpreting findings regarding differential methylation.

Sequencing studies in non-diseased individuals have reported differential hydroxymethylation patterns in ID and ASD risk genes, whilst decreased levels of the

Rett syndrome risk protein MECP2 correlate with a loss of 5hmC levels (Cheng et al., 2015). However, as reported in the dementia literature, research has countered this evidence by showing no enrichment of 5hmC levels in ASD risk genes (Wang et al., 2012). The complexity of 5hmC patterns is once again exemplified in cancer tissue, where 5hmC depletion has been observed in malignant glioma and breast cancer tissue but 5hmC increase has been observed in low-grade tumours (Eleftheriou et al., 2015, Orr et al., 2012). In our control human brain tissue, it would be difficult to estimate increased or decreased levels of 5hmC between the brain regions and would require high resolution quantification of staining intensity. However, the ubiquitous nature of 5hmC in our brain sections would suggest that areas where consistent increase of 5hmC has been reported, such as the hippocampus, are of key importance in understanding the function of this modification.

Our previous experimental chapters have highlighted the hippocampus as a potential area where the influence of genetic variants in methylation genes could impact upon cognitive function. Rather than a loss or gain of differential methylation patterns, it could be that this influence is due to a change from one functional methylation modification to another. TET-mediated demethylation has been documented in the hippocampus, with genetic knockout of *TET1* and the subsequent reduction in 5hmC levels in this area shown to correlate with poorer memory performance in rodent studies (Feng et al., 2015, Rudenko et al., 2013). As exemplified in the previous chapter, there are also a myriad of 5hmC binding proteins that can affect transcriptional activity directly or affect the binding of transcription factors (Zhubi et al., 2014). Interestingly, certain vitamins have also been shown to associate with 5hmC. For example, vitamin C can induce TET-mediated DNA demethylation and is thus associated with enhanced 5hmC levels (Dao et al., 2014, Santiago et al., 2014). Rodent work has shown that B vitamin imbalance can lead to hypomethylation, but these studies only assessed 5mC levels and not 5hmC (Fernandez-Roig et al., 2012). Therefore, rather than the loss of methylation patterns, this vitamin imbalance could actually result in increased levels of 5hmC or further demethylation modifications, underlining the complexity of possible methylation changes associated with our previous findings.

In comparison to the increasing amount of literature investigating 5hmC patterns and the function of this mark, 5caC and a third oxidative derivative 5-formylcytosine (5fC) have been understudied and the role of these demethylation marks has been overlooked. Little reliable 5caC signal was observed in most of the rodent and human

brain regions examined. Nevertheless, the unexpected finding of robust 5caC signal in the human cerebellum, along with some fainter staining in the hippocampus, illustrates that this methylation modification is present in the brain and likely has a functional role in line with the other methylation marks.

Immunohistochemistry studies carried out by our collaborators had previously confirmed the presence of 5caC signal in a range of embryonic mouse tissues, as well as in the adult mouse subventricular zone (Alioui et al., 2012, Wheldon et al., 2014). They also observed no 5caC signal in the mouse neocortex or hippocampus. Although faint 5caC signal was observed in the rodent hippocampus, this staining was not substantial in comparison to the 5hmC signal. Genome-wide sequencing of 5caC has indicated that this mark is found more frequently in intragenic regions of highly expressed genes and in the promoters of lowly expressed genes, whilst single-base resolution sequencing of 5caC observed the presence of this mark in hypomethylated promoters of highly expressed genes, positing a similar role to 5hmC in transcriptional activation as opposed to repression (Neri et al., 2015, Shen et al., 2013). In the few disease studies carried out using immunohistochemistry that included 5caC, the picture is similar to that of 5hmC. For example, research has shown both a decrease of 5caC in AD individuals as well as no change in 5caC levels in AD (Bradley-Whitman and Lovell, 2013, Condliffe et al., 2014). Some evidence has also been provided for the 5caC presence in glioma and breast cancer tissue compared to the absence of 5caC in normal tissue (Eleftheriou et al., 2015).

5caC is converted from 5hmC by the DNA glycosylase thymine-DNA glycosylase or TDG. Research has shown that genetic knockout of *TDG* influences both 5caC and 5fC patterns in mouse embryonic stem cells (Shen et al., 2013). Electrophoretic assays of 5caC-TDG binding have shown that 5caC is the preferential binding partner to TDG in comparison to 5fC, likely due to the specific recognition of the carboxyl moiety of 5caC (Zhang et al., 2012b). The main function of 5caC, like the other methylation modifications, appears to be a relationship with transcription factors and modulation of transcriptional activity. 5caC shows some contrast with 5hmC, as 5caC was found to increase transcription factor binding whilst 5hmC inhibited transcription factor binding (Sayeed et al., 2015). Interestingly, the same authors also showed that 5caC decreased the thermal stability of the DNA whilst 5hmC increased thermal stability. Further work has not only corroborated this effect on thermal stability, but also that binding of certain transcription factors was more stable in the presence of 5caC (Golla et al., 2014).

Given that many more DNA methylation readers have been associated with 5caC binding than 5hmC, this mark could have a substantial but as-of-yet unknown influence on transcriptional activity (Spruijt et al., 2013). Of note is the observation that DNMT1 binds to 5caC, as 5hmC is known to block DNMT1 binding (Tahiliani et al., 2009). Although DNMT1 itself does not appear to act as a transcription factor, it does interact with transcription factors and this interaction is predicted to guide the inheritance of some DNA methylation patterns (Chin et al., 2016, Hervouet et al., 2010). Therefore, 5caC could play a role in the direction of maintenance methylation deposition. Alternatively, members of the DNMT family have been shown to act as both demethylases and dehydroxymethylases, under certain calcium and redox states, outlining a further association between 5caC and demethylation activity (Chen et al., 2012, Chen et al., 2013).

Could the presence of 5caC highlight alternative mechanisms that explain changes in cognitive function? Aside from the many potential downstream effects of multiple transcription factors binding to 5caC, this methylation modification does have a unique effect upon DNA polymerase activity. Work has shown that DNA polymerases stall in the presence of 5caC, as G:5caC pairings seem to be read as mismatches and subsequently attract mismatch repair proteins, a process that could lead to the activation of cell death pathways (Hickman and Samson, 1999, Shibutani et al., 2014). With regards to variants in methylation genes, research has discovered cases of mutations in transcription factors such as zinc finger protein 57 homolog (Zfp57) that lead to altered 5caC binding (Liu et al., 2013). This is comparable to the Rett syndrome mutation in *MECP2*, which inhibits the binding of MECP2 to 5hmC (Mellen et al., 2012). In the current study, particular staining of 5caC and 5hmC was observed in the cerebellum and to a lesser extent in the hippocampus. Both of these areas share transcription factor expression profiles during glutamatergic neurogenesis (Hevner et al., 2006). Therefore, altered 5hmC and 5caC profiles in these key brain regions could influence the development of neurons and glutamatergic signalling, and this could subsequently underlie discrepancies in cognitive function.

7.3.2. Immunohistochemistry of brain tissue

As already highlighted in the discussion of this chapter, other research groups have favoured the application of methylation sequencing technologies over immunohistochemistry. Although the goal of our research group is to combine both

methodologies along with traditional genotyping, the aim of this chapter was to confirm that the immunohistochemical method is viable and that it can offer an experimental opportunity to characterise these methylation marks in biological tissue, rather than estimating their position using bisulfite sequencing or methylation sequencing techniques.

With regards to our choice of investigating 5hmC and 5caC over 5mC and 5fC, the reason was principally two-fold. In discussions with our collaborator, Alexey Ruzov, it was clear that the antibodies for 5mC, 5hmC, and 5caC were more reliable than those for 5fC at the start of this study, although our collaborators have since successfully applied a 5fC antibody in immunohistochemistry (Abakir et al., 2016a). More pertinent, however, was the biology of the marks themselves. It has been noted a number of times in this discussion that research reporting hypomethylation may not have been able to accurately study further demethylation modifications such as 5hmC and 5caC, and therefore the interpretation of their findings can be called into question as our study has illustrated the multiple roles these further modifications play. 5hmC is increasingly prominent in the field of DNA methylation, making it an inevitable choice. Apart from the antibody concerns between 5fC and 5caC, the role of 5caC as the last demethylation step before restoration of the cytosine base made it an intriguing candidate for study.

There remained some issues with our application of this methodology. One of these issues was the time needed to optimise the protocols for human brain tissue. However, other issues were related to the approach itself and would require greater optimisation of the protocols. To begin with, the rodent sections used were cut across the coronal plane, as is common for human tissue but not for rodent tissue. Had this chapter been focused on the characterisation of 5hmC and 5caC in rodents, as opposed to using this tissue simply to optimise the techniques, then the use of coronal over sagittal tissue could cause difficulties in interpretation as previous studies were more likely to use sagittal cuts. A more pressing concern regarding our work in the human tissue was the inability to accurately locate the subventricular zone in our caudate/putamen tissue sections. This specific region could not be identified using images from previous immunohistochemistry studies and so turned to the inclusion of a reference stain for stem cell markers, as the subventricular zone is a principal area of adult neurogenesis in the human brain. However, our DCX staining proved unsuccessful and therefore further work would be required to optimise a stem cell reference stain to identify this brain region. Considering previous

work had observed 5caC signal in the subventricular zone of adult mice, this would be an important investigation of 5caC presence in the adult human.

With regards to the methodology applied here, there are some notable elements of the standard and adapted protocols that are important to consider, particularly when attempting to replicate this technique. During the optimisation of the standard protocol, the antibody solution was applied to the whole of the slides rather than drawing around the tissue sections with the hydrophobic barrier pen. However, this approach did not yield successful staining and also used considerably more antibody solution than the hydrophobic barrier approach. Whilst no noticeable issues were observed with the application of the pen, certain complaints have been raised whereby the hydrophobic solution may leach into the sections and interfere with the fluorescent signal. To improve the clarity of our fluorescent signal and reduce the background noise, particularly in the case of 5caC, the obvious solution would be to apply an antigen retrieval step. However, our collaborators had found that antigen retrieval interfered with the hydrochloric acid step necessary to the adapted protocol (Abakir et al., 2016a). As the 5caC staining suffered from higher background signal than 5hmC, likely due to the weakness of the antibody, further optimisation of this stain is required. One possibility is that longer application of the TSA amplification kit could enhance the weak 5caC signal, but this may also increase background signal. Another is the use of Sudan Black B, a commercial dye that has been consistently shown to reduce background auto-fluorescence in tissue sections. Our research group has begun to use this dye to good effect, but the implementation of this step into our adapted protocol may be detrimental and would require more optimisation to ensure that Sudan Black does not interfere with the hydrochloric acid step.

Finally, any research that involves histochemical or fluorescent signal depends upon the accurate use of microscopy and imaging software. Our microscopy was principally carried out using a confocal microscope and Zen imaging software. Although confocal imaging is easy to use and allows accurate imaging of most areas within a cell or tissue, it is particularly temporally sensitive. Work has highlighted the risks associated with photobleaching following long exposure times in confocal microscopy (St Croix et al., 2005). Indeed, one can obtain quality wide-field images without the same risk of bleaching, particularly at low magnification. Nevertheless, the pressure within the scientific community to obtain ever-greater resolution subsumes these arguments. Whilst quantification was not carried out using Pearson correlation coefficients for colocalisation using Zen or other software such as ImageJ,

the latter program was used to process the image files for this chapter. ImageJ was utilised to add scale bars and to ensure that each fluorescent channel was appropriately coloured. However, studies have emphasised the use of such software to manipulate image files, even to small degrees, to ensure that images are of publication quality (Cromey, 2010). Therefore, caution should always be advised when analysing microscopy images to ensure that processing of the image files does not become manipulation of the image data.

8. General discussion

8.1. Review of findings from the experimental chapters

In this thesis, a number of experimental approaches have been used to investigate the relationship between nucleotide methylation and neurocognitive phenotypes. The main focus of the investigation centred on the study of one particular coding variant in one of the DNA methyltransferase genes, *DNMT3L* R278G. Analysis of MCI and control datasets was employed to examine the relationship between this variant genotype, key components of the one-carbon cycle, and cognitive performance. To extend our investigation of nucleotide methylation, next-generation sequencing (NGS) datasets from intellectual disability (ID), ID comorbid with psychiatric disease, and autism spectrum disorder (ASD) were examined to assess disease risk associated with DNA and RNA methylation genes. Finally, an adapted immunohistochemistry method was utilised to perform preliminary characterisation of further DNA methylation modifications in rodent and human brain tissue.

In the first chapter, the combined OPTIMA cohorts including individuals with Alzheimer's disease (AD), mild cognitive impairment (MCI), and aged controls were assessed. However, compared to a known risk genotype for AD (*APOE* E4), the *DNMT3L* R278G was not conclusively associated with the progression of dementia. Deeper analysis was then performed of specific dementia-sensitive cognitive factors, visuospatial associative memory and verbal semantic memory, in the VITACOG B vitamin treatment study of MCI. It was discovered that carriers of the *DNMT3L* R278G minor allele who received B vitamin treatment showed an improvement in visuospatial associative memory, markedly so in those who had high levels of baseline homocysteine (Hcy). Conversely, the major allele homozygotes who received B vitamin treatment showed an improvement in verbal semantic memory.

In the second chapter, this genotype-dependent relationship with cognitive function was substantiated by interrogating rate of brain atrophy (ROA) data from the VITACOG cohort. A negative correlation was discovered between ROA and visuospatial associative memory for *DNMT3L* R278G minor allele carriers, and a negative correlation between ROA and verbal semantic memory in major allele homozygotes, mirroring the genotype-dependent effects observed in the first chapter. To further explore the functional impact of this variant on the DNMT3L protein, multiple *in silico* modelling techniques were applied. These tools predicted that the

R278G variant would have a disruptive effect on the structure of DNMT3L and could compromise the binding efficiency between DNMT3L and DNMT3A, a key complex in the deposition of *de novo* methylation patterns.

In the third chapter, our analysis of the *DNMT3L* R278G variant was replicated in two general population control cohorts, namely the NCDS 1958 and TwinsUK cohorts. Neither of these studies had the potential to directly replicate the B vitamin treatment performed in the longitudinal VITACOG study, but these cohorts did include self-reported vitamin intake information and serum measurements of B vitamin levels. Baseline analyses revealed that regular vitamin intake, particularly B vitamin intake, was associated with better visuospatial associative memory performance in controls. In addition, regular vitamin intake and high B vitamin levels were associated with lower Hcy levels, supporting the findings from the VITACOG study. However, in contrast to our results in VITACOG, it was discovered that major allele homozygotes who regularly took vitamins performed better in visuospatial associative memory, not minor allele carriers. This illustrated a genotype-dependent relationship with visuospatial associative memory that manifested differently in MCI and controls.

In the fourth chapter, our investigation of methylation genes and cognitive phenotypes was extended beyond the analysis of *DNMT3L* R278G and MCI. NGS data from multiple ID and ASD cohorts was examined within the UK10K project alongside general population control data from the TwinsUK cohort. A sequencing pipeline was designed to process the combination of disparate NGS datasets into one stream of analysis. Genetics variants, particularly rare missense variants, in thirty DNA and RNA methylation genes were assessed for disease risk in ID and ASD respectively. Evidence was discovered for enrichment of rare damaging missense variants in RNA methylation “readers” in the ID and ASD cohorts compared to controls. Furthermore, single variant and burden analysis revealed that a small number of variants and their respective genes were associated specifically with risk for ID. These findings highlight a potentially important relationship between genetic variation in methylation genes and risk for ID.

In the fifth chapter, attempts were made to accurately measure the location of further methylation marks in human brain tissue, with a view to recording more detailed characterisation of these marks whilst taking into account genotype information for key variants such as *DNMT3L* R278G. A standard immunohistochemistry protocol was utilised in rodent and human brain tissue to identify reference markers for

neuronal and glial cells, allowing validation of the nuclear presence of further methylation marks in key cell types. An adapted immunohistochemistry protocol was then implemented to identify 5hmC and 5caC staining within human brain tissue. Initial characterisation corroborated and built upon previous reports by showing widespread staining of 5hmC in the human brain, as well as localised 5caC staining within the human cerebellum. The latter observation could help to shed light on the function of this relatively unknown methylation mark.

8.2. How do these findings develop our understanding of methylation and cognition?

8.2.1. Methylation genotypes and the one-carbon cycle

The analysis from the first three chapters illustrates a clear influence of the DNA methyltransferase variant *DNMT3L* R278G on cognitive performance. This finding adds to the growing body of literature linking genetic variants in methylation genes or disruption of methylation proteins with the pathology of brain diseases. Some of these variants have been causal mutations, such as those reported in *DNMT1* and *DNMT3A* for a syndromic form of early neurodegeneration and a rare form of intellectual disability respectively (Klein et al., 2011, Tatton-Brown et al., 2014). Others have been observations noting the reduced expression of DNA methyltransferases, DNA methylation “readers”, and genes associated with active demethylation in dementia (Bihaqi and Zawia, 2012, Sliwinska et al., 2017).

A core factor in our results associating the *DNMT3L* R278G genotype with cognitive function involve particular components of the one-carbon cycle. Our findings in both MCI and general population controls were dependent upon interactions with vitamins, particularly B vitamins, and Hcy levels. The relationship between the one-carbon cycle and DNA methylation has been well documented, as abnormal dynamics within this cycle has been consistently shown to influence methylation activity (Ulrey et al., 2005). Indeed, our observation of an interaction between the one-carbon cycle and DNA methylation (in the form of the *DNMT3L* R278G variant) has been hypothesised previously in multiple diseases. The hypomethylation hypothesis associated with cognitive dysfunction, where loss of methylation potential affects methylation patterns, and the redox/methylation hypothesis highlighted in ASD, where oxidative stress disrupts normal methylation, are both forms of this observation (Calvaresi and Bryan, 2001, Deth et al., 2008).

Our analysis discovered that the influence of *DNMT3L* R278G on cognitive performance was associated with different genotypes in MCI compared to general population controls. A model was presented explaining how the interaction of one-carbon cycle components and the R278G genotype may associate with dementia progression. In healthy middle age, major allele homozygotes who took vitamins showed better visuospatial associative memory performance. Once individuals decline to MCI levels, minor allele carriers who take B vitamins show improved visuospatial associative memory performance. Our prediction is that differential methylation patterns associated with the respective R278G alleles primes the potential for changes in cognition and that these changes are then triggered by dynamics of the one-carbon cycle such as elevated vitamin intake or lowered Hcy levels. This could be seen as an adapted form of the hypomethylation or redox/methylation hypotheses. Whilst our findings are important in supporting the link between methylation and the one-carbon cycle, assessment of a wider range of cognitively impaired individuals including MCI and AD is needed to support our particular model of genotype-dependent cognitive change.

An important caveat to our findings regarding *DNMT3L* R278G in both MCI and general population controls is the particular importance of the cognitive factors utilised in this study. No association was observed between this genotype and performance on any individual test in VITACOG, bar a baseline difference in Category Fluency. Similarly, no association was observed between this genotype and performance on any individual test in NCDS 1958 or TwinsUK. Therefore, the application of PCA to decipher cognitive factors from the test data available was key to identifying subtle changes in cognition that would have otherwise gone unobserved. Indeed, the use of PCA has been successfully applied in other genetic studies looking at human intelligence (Davies et al., 2011). Furthermore, the results of our study in individuals with MCI adds further credence to the sensitivity of visuospatial associative memory in detecting cognitive changes associated with the progression of dementia (Blackwell et al., 2004).

8.2.2. Genetic risk for ID in methylation genes

Considering that our observations of a relationship between *DNMT3L* R278G and cognition were striking and that a relationship between this variant and intelligence had previously been identified (Haggarty et al., 2010), this variant appears to have been largely overlooked in genetic analyses. This variant was chosen because it was relatively common, providing more statistical power to perform analyses on the minor allele carriers, and because it was predicted to be functionally damaging in a number of protein prediction tools. However, the advent of genome-wide association studies (GWAS) and the subsequent failure to identify large sources of heritability from common variants has made variants such as *DNMT3L* R278G more unattractive as a candidate.

As a consequence, assessment continued into coding variants that were predicted to have a functional impact on protein expression. However, focus was placed on the rare missense variants that would not be identified by GWAS array screening. In doing so, four rare missense variants were identified in *DNMT1*, *UHRF2*, *HNRNPA2B1*, and *THYN1* respectively that were associated with specific risk for ID. The DNA methylation "writer" *DNMT1* has established functions that can be linked to the development of ID, such as the disruption of normal maintenance methylation patterns. The involvement of DNA methylation "readers" *UHRF2* and *THYN1* suggest that transcriptional regulatory processes may also be involved in forms of ID, supported by the Rett syndrome risk gene *MECP2* which is heavily involved in transcriptional regulation (Szulwach et al., 2011). Perhaps the most interesting result is the identification of the RNA methylation "reader" *HNRNPA2B1*, as aberrant alternative splicing caused by mutations in this gene have recently been identified in other brain diseases (Martinez et al., 2016). The precise mechanisms associating such diverse transcriptional and translational processes with the development of ID are not yet clear, but this finding provides further evidence for an association between less prominent epigenetic modifications such as RNA methylation and risk for brain disease.

Previous work has identified variants in DNA methylation genes, such as *DNMT3A* and *MECP2*, which carry risk for certain forms of ID (Moretti and Zoghbi, 2006, Tatton-Brown et al., 2014). Along with the findings from our research, this suggests a clear link between the disruption of normal methylation function and normal intellectual development. It is clear that genomic imprinting, the silencing of parental

alleles via methylation during early development, has a substantial impact on the growth of the brain and the differentiation of neural regions (Keverne et al., 1996). Altered imprinting has been identified in forms of ID such as Angelman and Prader-Willi syndromes (Perk et al., 2002), and our identification of variation in *DNMT1* suggests that similar changes could be occurring in rare familial forms of ID and ID comorbid with psychiatric disease. More pertinent for future research is the association between variation in "readers" of both DNA and RNA methylation and risk for ID, as these regulators of transcriptional and translational activity could yet underlie the neurological changes associated with ID (Zoghbi and Bear, 2012).

8.2.3. Localised presence of further methylation marks in the brain

The final experimental chapter within this study used an immunohistochemical fluorescence approach to detecting further intermediate methylation modifications in the human brain. The recent advances in multiple forms of methylation sequencing, ranging from oxidative bisulfite sequencing or methylation assays for 5mC and 5hmC to methylase-assisted sequencing for 5fC and 5caC, are making single-base detection of these modifications at a genome-wide level attainable. However, immunostaining techniques still provide a practical means of examining these methylation marks at cellular resolution in tissue samples.

Our work confirmed a near ubiquitous presence of 5hmC in rodent and human brain tissue. This corroborates research asserting that 40% of DNA methylation present in the brain is 5hmC (Irier et al., 2014). As previously described, 5hmC plays a dynamic role within DNA methylation as it is associated with transcriptional activation rather than the transcriptional suppression associated with 5mC. Our results support the idea that 5hmC is a regulatory modification in its own right, rather than just an intermediary mark in the demethylation process. This view is in keeping with the developing landscape of methylation, one which will lead to re-evaluation of earlier work assessing DNA methylation on the basis of 5mC levels alone (Jin et al., 2010). Sequencing studies are helping to develop a deeper understanding of the distribution and function of many CpG and CpH modifications. This is exemplified by 5hmC which has been shown to display bias towards G-rich sequences and low CpG regions (Yu et al., 2012b). As our findings showed ubiquitous regional 5hmC presence in the human brain, the goal for higher resolution immunofluorescence will be characterise the cell type specific distribution of 5hmC in healthy and disease brain tissues.

In comparison to 5mC and 5hmC, little work has been carried out into the function or abundance of 5fC and 5caC. 5caC was investigated because of its position as the final methylation modification before conversion back to unmethylated cytosine and because of the availability of a 5caC antibody previously utilised by a colleague (Alioui et al., 2012). An unexpected and striking specificity of 5caC staining was observed in the cerebellum of the human brain, whilst 5caC staining was almost absent in all other brain areas included in our experiments. 5caC has been associated with regulation of transcriptional activity through the binding of multiple transcription factors (Spruijt et al., 2013). It was posited that this highlights the presence of certain transcriptional profiles occurring within the cerebellum, such as those recorded during glutamatergic neurogenesis (Hevner et al., 2006). Indeed, our findings support other work outlining evidence of a cerebellar-specific transcriptome (Schuller et al., 2006). Potential 5caC specific staining of Purkinje cells, large inhibitory neurons that project into the cerebellum, was also identified. Recent work has shown that mature Purkinje cells are capable of large-scale demethylation and remethylation events, behaviour that was only thought to occur during embryonic development (Zhou et al., 2016). The authors note that this coincided with dendritosynaptic growth, namely the formation of connections between different neurons, highlighting a potential association between methylation dynamics and synaptic behaviour. Nevertheless, the relevance of the 5caC mark in particular to cerebellar transcriptomics remains unclear and necessitates future research to corroborate and build upon these results.

8.3. Limitations and future research

In this thesis, I have taken advantage of well-curated datasets from well-designed cohort studies. However, this is not to say that there are not limitations associated with these datasets and with our analyses that could be rectified in future research. For example, much of our analysis focused upon individuals with MCI from the VITACOG cohort, part of the wider OPTIMA program. The characterisation of MCI as a discrete disease has been relatively recent and research is continuing to develop our understanding of this condition. One important aspect is the heterogeneity within the disease, both between amnestic and non-amnestic subtypes and also between those in the early stages of the disease and those in the latter stages. Work has shown that medial temporal lobe activation of individuals with MCI is similar to that of aged controls or individuals with AD, dependent upon the stage of MCI development (Dickerson et al., 2005). Without further clinical information pertaining to the

VITACOG cohort, it would be impossible to study this facet of performance. However, access to the OPTIMA cohort of individuals with AD and the Challenge cohort of aged controls could allow one to replicate our analyses and subsequently discover whether the VITACOG individuals match those with AD or the controls. Moreover, one could then ascertain the association between *DNMT3L* R278G and cognitive function in AD, developing our model of disease risk associated with this variant.

There are also two corollaries to our analysis of individuals with MCI. The first surrounds the use of visuospatial associative memory and verbal semantic memory. Visuospatial associative memory is particularly sensitive to dementia progression whilst verbal semantic memory should provide more nuanced assessment of MCI itself. Replication of our analyses in individuals with AD would allow one to confirm the specificity of visuospatial associative memory, while further information on the condition of the VITACOG individuals would allow one to investigate the sensitivity of verbal semantic memory. The second relates to the analysis of brain atrophy data. Information pertaining to whole brain atrophy was granted from the VITACOG cohort. However, previous work using this cohort has shown that the medial temporal lobes, particularly the hippocampus, are the areas most likely to be affected by MCI and AD. Therefore, further atrophy and clinical data on the MCI individuals would allow one to provide evidence for our hypothesis that visuospatial associative memory deficits associate with hippocampal atrophy and to provide support for changes in brain atrophy during the development of MCI.

Our analysis of the cohort datasets could be further developed. One example would be the application of Mendelian randomisation, a method whereby the causal influence on a certain trait can be inferred using a genetic factor that has a known effect on the independent variable. In the case of our relationship between B vitamins, Hcy, and cognitive performance, one could focus on multiple markers associated with B vitamin and Hcy levels (Tanaka et al., 2009) or a single marker such as *MTHFR* C677T. This genotype has consistently been associated with low folate and high Hcy levels (de Bree et al., 2003). Using this genotype, one could more robustly assess the influence of vitamins and Hcy on cognitive performance in our case and control cohorts. For this method to work, the genotype must be independent of the dependent variable and any confounding variables (VanderWeele et al., 2014). Fortunately, evidence suggests that *MTHFR* C677T does not influence cognition outside of the relationship with B vitamins and Hcy (Moorthy et al., 2012).

Perhaps the most interesting line of research for the future will focus on the *DNMT3L* R278G variant. For instance, one would wish to perform DNA methylation sequencing of the VITACOG cohort members to provide initial evidence that the R278G variant does indeed influence methylation patterns. An array-based approach may limit our ability to assess further methylation modifications but would allow one to interrogate any regions of the genome where significant differentially methylated regions are present. Promisingly, the TwinsUK general population cohort has made sequencing data for 5mC available for 100 members of the cohort. Providing that an agreement to link these individuals with our genotyping data could be arranged, this may provide a first-pass assessment of the R278G variant and methylation patterns in controls.

In addition, members of our research group have recently confirmed that the R278G variant creates an epiallele, a position in the genome where the level of methylation differs between individuals. It is not clear whether this would allow for further regulation of *DNMT3L* itself and hence influence the methylation potential of the DNMT3A-3L complex. Assuming that measurable methylation levels occur at this epiallele site, *in vitro* co-expression of DNMT3A and DNMT3L would allow one to assess whether regulation of this epiallele influences this complex. A number of recent publications have utilised the cas9 DNA endonuclease enzyme, famed for its use in CRISPR/cas9 gene editing technologies (Zhang et al., 2014), to induce methylation changes by targeting methyltransferase or demethylase constructs at areas of the genome, effectively creating epi-CRISPR/cas9 (Lei et al., 2017, Stepper et al., 2017, Xiong et al., 2017). Should measureable methylation levels at the epiallele site be difficult to identify, these techniques could be used to introduce DNA methylation at this site and subsequently assess the influence on the DNMT3A-3L complex.

This thesis hypothesised that the *DNMT3L* R278G variant would disrupt the binding between DNMT3L and DNMT3A, leading to altered methylation patterns and transcriptional activity associated with cognitive function. Nevertheless, further experiments are necessary to elucidate the function of the DNMT3L protein. To begin with, focus was on the DNMT3A-3L complex because DNMT3A is the most prominent *de novo* methyltransferase and *in silico* modelling of this complex has been carried out previously (Jia et al., 2007). Yet, DNMT3L also forms a complex with DNMT3B to perform DNA methylation (Pacaud et al., 2014, Suetake et al., 2004). Thus, our *in silico* modelling could be extended to create a model of

DNMT3B-3L and to assess the impact of the R278G variant on this complex. *In vitro* examination could provide further insight into the DNMT3L protein. For example, *in vitro* investigation of DNMT3A and DNMT3B has shown that both are capable of demethylation of 5hmC to unmethylated cytosine, particularly in cases of oxidative stress (Chen et al., 2012). This is of particular interest considering the role of oxidative stress in the redox/methylation hypothesis and the production of the antioxidant glutathione in the one-carbon cycle. Notable experiments have utilised B vitamin culture mediums to evaluate methylation changes and expression levels of key genes (Fuso et al., 2005). Extension of this approach would allow one to both confirm the modulatory interaction between B vitamins and *DNMT3L* R278G as well as confirm the disruption of the DNMT3A-3L complex.

Having completed our analysis of *DNMT3L* R278G, disease risk associated with variants in thirty DNA and RNA methylation genes was assessed. Four rare missense variants were discovered in these genes linked to particular risk for ID. There were limitations associated with our development of a NGS pipeline and with the cohorts available for this analysis. To begin with, it was recognised that missing data could influence the case-control analysis, prompting application of thresholds of imputation. Whilst this allowed more accurate interpretation of our findings, cases remained where the control data remained heavily imputed. In these cases, variants were only retained depending upon their presence in the case cohorts. To ensure that the variants identified are truly associated with disease risk, replication of the NGS analysis in datasets with wide coverage and sufficient read depth could be implemented. This would also allow one to identify further ID datasets containing greater phenotypic information.

Rare missense variants of interest were found primarily but not exclusively in one of the two ID datasets, namely the Muir cohort. Whilst the phenotypic information provided for all of the NGS cohort datasets was limited, the Muir cohort contains individuals with comorbid learning disabilities and psychosis compared to Rare FIND cohort which contains individuals with rare familial ID. This raises a question as to whether these rare missense variants are more reliably associated with general ID or forms of ID that include comorbid psychosis. As mentioned in the methods, one gene that was not included in our analyses was the histone methyltransferase *SETD1A*. Loss-of-function (LoF) mutations in *SETD1A* were initially associated with risk for schizophrenia but closer inspection showed that the majority of these individuals also had forms of ID, drawing comparisons with our findings in the Muir cohort (Singh et

al., 2016). Extension of this work to larger cohorts confirmed that rare damaging variants such as LoF mutations are enriched in individuals with comorbid ID and psychosis as well as those who have schizophrenia alone (Singh et al., 2017, Takata et al., 2014). These studies and our findings exemplify that ID comorbid with psychosis appears to be of particular interest in rare variant analysis. Therefore, extension of our work would require cohorts with psychosis alone to compare with our current results. In addition, the inclusion of LoF variants in addition to missense variants would be necessary to comprehensively characterise the risk of rare damaging variants in these brain diseases.

The success of our *DNMT3L* R278G analyses raises the possibility of replicating this approach using candidates from our NGS analysis. Whilst the missense variants may be too rare to afford statistical power, cohort studies of requisite size (n > 10,000) such as the UK Biobank could allow for some assessment of cognitive function (Sniekers et al., 2017). This may be particularly important for RNA methylation and cognition, which was not the main focus of this study and which is still burgeoning as a field. One of the most common RNA methylation marks, *N*6-methyladenosine or m^6A, is currently being studied by other members of the research group due to its role in brain development (Jia et al., 2011, Meyer et al., 2012). Inhibition of m^6A methyltransferases results in markedly altered gene expression and splicing (Dominissini et al., 2012), whilst KO of an "eraser" of m^6A marks led to increased methylation of mRNAs related to neuronal signalling (Hess et al., 2013). Interestingly, a complete lack of low-frequency or common missense variants were observed in the RNA methylation "eraser" family, suggesting that this group of genes is of particular functional importance and that deeper analysis of any variants in this family should be carried out (Gibson, 2012).

Finally, the immunohistochemistry of the further methylation modifications 5hmC and 5caC provides the most obvious opportunity for future research. Some interesting preliminary results were obtained with the limited amount of staining performed but further replications are required to ensure the validity of our findings, particularly relating to 5caC due to the comparative weakness of the respective antibody. I attempted to stain for 5hmC and 5caC in one session before staining for reference markers in sequential sections of tissue. This is certainly not an ideal method and, given enough time to optimise the correct antibodies, a triple-staining approach using 5hmC, 5caC, and a reference marker would be used. The short-to-medium term aim of this immunofluorescent approach is to characterise and quantify the presence of

these methylation marks in both control and diseased human brain tissue. The medium-to-long term aim would be to combine this approach with genotyping for a particular variant, such as *DNMT3L* R278G, to provide evidence of genotype-dependent methylation changes occurring in specific regions of the brain.

8.4. Conclusions

The aim of this thesis was to further our understanding of the relationship between nucleotide methylation and neurocognitive phenotypes. In studying the *DNMT3L* R278G variant and cognitive endophenotypes associated with dementia progression, a genotype was identified that has a clear influence on cognitive performance. Using *in silico* approaches and brain atrophy data, further evidence of a mechanism by which this genotype may influence methylation patterns and cognition was provided. The examination of nucleotide methylation genes in brain disease cohorts provided further evidence of a link between genetic variation associated with methylation and cognition. Finally, successfully characterisation of methylation modifications was achieved in human brain tissue.

These three novel strands of evidence emphasis the relationship between nucleotide methylation and neurocognitive phenotypes. The involvement of multiple DNA and RNA methylation modifications will no doubt predicate the characterisation of further forms of nucleotide methylation and their subsequent associations with cognitive function. Furthermore, the use of multiple neurodegenerative, neurodevelopmental, and neuropsychiatric cohorts emphasise the importance of methylation in brain disease, providing ample opportunity for further research to target these epigenetic mechanisms for therapeutic benefit.

References

AAPOLA, U., LIIV, I. & PETERSON, P. 2002. Imprinting regulator DNMT3L is a transcriptional repressor associated with histone deacetylase activity. *Nucleic Acids Res,* 30**,** 3602-8.

ABAKIR, A., WHELDON, L., JOHNSON, A. D., LAURENT, P. & RUZOV, A. 2016a. Detection of Modified Forms of Cytosine Using Sensitive Immunohistochemistry. *J Vis Exp*.

ABAKIR, A., WHELDON, L. M. & RUZOV, A. 2016b. Immunohistochemical Detection of Oxidized Forms of 5-Methylcytosine in Embryonic and Adult Brain Tissue. *In:* KARPOVA, N. (ed.) *Epigenetic Methods in Neuroscience Research.* NY, USA: Humana Press.

ABBASI-MOHEB, L., MERTEL, S., GONSIOR, M., NOURI-VAHID, L., KAHRIZI, K., CIRAK, S., WIECZOREK, D., MOTAZACKER, M. M., ESMAEELI-NIEH, S., CREMER, K., WEISSMANN, R., TZSCHACH, A., GARSHASBI, M., ABEDINI, S. S., NAJMABADI, H., ROPERS, H. H., SIGRIST, S. J. & KUSS, A. W. 2012. Mutations in NSUN2 cause autosomal-recessive intellectual disability. *Am J Hum Genet,* 90**,** 847-55.

AGNEW-BLAIS, J. C., WASSERTHEIL-SMOLLER, S., KANG, J. H., HOGAN, P. E., COKER, L. H., SNETSELAAR, L. G. & SMOLLER, J. W. 2015. Folate, vitamin B-6, and vitamin B-12 intake and mild cognitive impairment and probable dementia in the Women's Health Initiative Memory Study. *J Acad Nutr Diet,* 115**,** 231-41.

AISEN, P. S., SCHNEIDER, L. S., SANO, M., DIAZ-ARRASTIA, R., VAN DYCK, C. H., WEINER, M. F., BOTTIGLIERI, T., JIN, S., STOKES, K. T., THOMAS, R. G., THAL, L. J. & ALZHEIMER DISEASE COOPERATIVE, S. 2008. High-dose B vitamin supplementation and cognitive decline in Alzheimer disease: a randomized controlled trial. *JAMA,* 300**,** 1774-83.

AL-GAZALI, L. I., PADMANABHAN, R., MELNYK, S., YI, P., POGRIBNY, I. P., POGRIBNA, M., BAKIR, M., HAMID, Z. A., ABDULRAZZAQ, Y., DAWODU, A. & JAMES, S. J. 2001. Abnormal folate metabolism and genetic polymorphism of the folate pathway in a child with Down syndrome and neural tube defect. *Am J Med Genet,* 103**,** 128-32.

ALARCON, C. R., GOODARZI, H., LEE, H., LIU, X., TAVAZOIE, S. & TAVAZOIE, S. F. 2015. HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events. *Cell,* 162**,** 1299-308.

ALELU-PAZ, R., CARMONA, F. J., SANCHEZ-MUT, J. V., CARIAGA-MARTINEZ, A., GONZALEZ-CORPAS, A., ASHOUR, N., OREA, M. J., ESCANILLA, A., MONJE, A., GUERRERO MARQUEZ, C., SAIZ-RUIZ, J., ESTELLER, M. & ROPERO, S. 2016. Epigenetics in Schizophrenia: A Pilot Study of Global DNA Methylation in Different Brain Regions Associated with Higher Cognitive Functions. *Front Psychol,* 7**,** 1496.

ALEXANDER, J. R. M. & SMALES, S. 1997. Intelligence, learning and long-term memory. *Personality and Individual Differences,* 23**,** 815-825.

ALFRED, T., BEN-SHLOMO, Y., COOPER, R., HARDY, R., DEARY, I. J., ELLIOTT, J., HARRIS, S. E., HYPPONEN, E., KIVIMAKI, M., KUMARI, M., MADDOCK, J., POWER, C., STARR, J. M., KUH, D., DAY, I. N. M. & TEAM, H. S. 2013. Genetic

Variants Influencing Biomarkers of Nutrition Are Not Associated with Cognitive Capability in Middle-Aged and Older Adults. *Journal of Nutrition,* 143**,** 606-612.
ALIOUI, A., WHELDON, L. M., ABAKIR, A., FERJENTSIK, Z., JOHNSON, A. D. & RUZOV, A. 2012. 5-Carboxylcytosine is localized to euchromatic regions in the nuclei of follicular cells in axolotl ovary. *Nucleus,* 3**,** 565-9.
ALMEIDA, R. D., SOTTILE, V., LOOSE, M., DE SOUSA, P. A., JOHNSON, A. D. & RUZOV, A. 2012. Semi-quantitative immunohistochemical detection of 5-hydroxymethyl-cytosine reveals conservation of its tissue distribution between amphibians and mammals. *Epigenetics,* 7**,** 137-40.
ALVARADO, S., FERNALD, R. D., STOREY, K. B. & SZYF, M. 2014. The dynamic nature of DNA methylation: a role in response to social and seasonal variation. *Integr Comp Biol,* 54**,** 68-76.
ANDERSEN, C. A., PALMER, A. G., BRUNAK, S. & ROST, B. 2002. Continuum secondary structure captures protein flexibility. *Structure,* 10**,** 175-84.
ANDERSON, C. M., THIELE, D. K., RALPH, J. L., PERLEY, D. & OHM, J. E. 2016. Vitamin D Supplementation and DNA Methylation Patterns during Pregnancy and Lactation in Mothers and Infants. *The Official Journal of the Federation of American Socities for Experimental Biology,* 30.
ANDERSON, O. S., SANT, K. E. & DOLINOY, D. C. 2012. Nutrition and epigenetics: an interplay of dietary methyl donors, one-carbon metabolism and DNA methylation. *J Nutr Biochem,* 23**,** 853-9.
ANDRAWIS, J. P., HWANG, K. S., GREEN, A. E., KOTLERMAN, J., ELASHOFF, D., MORRA, J. H., CUMMINGS, J. L., TOGA, A. W., THOMPSON, P. M. & APOSTOLOVA, L. G. 2012. Effects of ApoE4 and maternal history of dementia on hippocampal atrophy. *Neurobiol Aging,* 33**,** 856-66.
ANNWEILER, C., HERRMANN, F. R., FANTINO, B., BRUGG, B. & BEAUCHET, O. 2012. Effectiveness of the combination of memantine plus vitamin D on cognition in patients with Alzheimer disease: a pre-post pilot study. *Cogn Behav Neurol,* 25**,** 121-7.
ASSOCIATION, A. P. 2013. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition,* Arlington, VI, American Psychiatric Publishing.
AUER, P. L., REINER, A. P., WANG, G., KANG, H. M., ABECASIS, G. R., ALTSHULER, D., BAMSHAD, M. J., NICKERSON, D. A., TRACY, R. P., RICH, S. S., PROJECT, N. G. E. S. & LEAL, S. M. 2016. Guidelines for Large-Scale Sequence-Based Complex Trait Association Studies: Lessons Learned from the NHLBI Exome Sequencing Project. *Am J Hum Genet,* 99**,** 791-801.
BAKER, N. A. 2004. Poisson-Boltzmann methods for biomolecular electrostatics. *Methods Enzymol,* 383**,** 94-118.
BALL, M. P., LI, J. B., GAO, Y., LEE, J. H., LEPROUST, E. M., PARK, I. H., XIE, B., DALEY, G. Q. & CHURCH, G. M. 2009. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol,* 27**,** 361-8.
BARBEAU, E. J., DIDIC, M., JOUBERT, S., GUEDJ, E., KORIC, L., FELICIAN, O., RANJEVA, J. P., COZZONE, P. & CECCALDI, M. 2012. Extent and neural basis of semantic memory impairment in mild cognitive impairment. *J Alzheimers Dis,* 28**,** 823-37.
BARCELO, C., ETCHIN, J., MANSOUR, M. R., SANDA, T., GINESTA, M. M., SANCHEZ-AREVALO LOBO, V. J., REAL, F. X., CAPELLA, G., ESTANYOL, J. M., JAUMOT,

M., LOOK, A. T. & AGELL, N. 2014. Ribonucleoprotein HNRNPA2B1 interacts with and regulates oncogenic KRAS in pancreatic ductal adenocarcinoma cells. *Gastroenterology,* 147**,** 882-892 e8.
BASSELL, G. J. & WARREN, S. T. 2008. Fragile X syndrome: loss of local mRNA regulation alters synaptic development and function. *Neuron,* 60**,** 201-14.
BATEMAN, A., MARTIN, M. J., O'DONOVAN, C., MAGRANE, M., ALPI, E., ANTUNES, R., BELY, B., BINGLEY, M., BONILLA, C., BRITTO, R., BURSTEINAS, B., BYE-A-JEE, H., COWLEY, A., DA SILVA, A., DE GIORGI, M., DOGAN, T., FAZZINI, F., CASTRO, L. G., FIGUEIRA, L., GARMIRI, P., GEORGHIOU, G., GONZALEZ, D., HATTON-ELLIS, E., LI, W. Z., LIU, W. D., LOPEZ, R., LUO, J., LUSSI, Y., MACDOUGALL, A., NIGHTINGALE, A., PALKA, B., PICHLER, K., POGGIOLI, D., PUNDIR, S., PUREZA, L., QI, G. Y., ROSANOFF, S., SAIDI, R., SAWFORD, T., SHYPITSYNA, A., SPERETTA, E., TURNER, E., TYAGI, N., VOLYNKIN, V., WARDELL, T., WARNER, K., WATKINS, X., ZARU, R., ZELLNER, H., XENARIOS, I., BOUGUELERET, L., BRIDGE, A., POUX, S., REDASCHI, N., AIMO, L., ARGOUD-PUY, G., AUCHINCLOSS, A., AXELSEN, K., BANSAL, P., BARATIN, D., BLATTER, M. C., BOECKMANN, B., BOLLEMAN, J., BOUTET, E., BREUZA, L., CASAL-CASAS, C., DE CASTRO, E., COUDERT, E., CUCHE, B., DOCHE, M., DORNEVIL, D., DUVAUD, S., ESTREICHER, A., FAMIGLIETTI, L., FEUERMANN, M., GASTEIGER, E., GEHANT, S., GERRITSEN, V., GOS, A., GRUAZ-GUMOWSKI, N., HINZ, U., HULO, C., JUNGO, F., KELLER, G., LARA, V., LEMERCIER, P., LIEBERHERR, D., LOMBARDOT, T., MARTIN, X., MASSON, P., MORGAT, A., NETO, T., NOUSPIKEL, N., PAESANO, S., PEDRUZZI, I., PILBOUT, S., POZZATO, M., PRUESS, M., RIVOIRE, C., ROECHERT, B., et al. 2017. UniProt: the universal protein knowledgebase. *Nucleic Acids Research,* 45**,** D158-D169.
BEAUCHAMP, K. A., LIN, Y. S., DAS, R. & PANDE, V. S. 2012. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J Chem Theory Comput,* 8**,** 1409-1414.
BECKETT, E. L., DUESING, K., MARTIN, C., JONES, P., FURST, J., KING, K., NIBLETT, S., YATES, Z., VEYSEY, M. & LUCOCK, M. 2016. Relationship between methylation status of vitamin D-related genes, vitamin D levels, and methyl-donor biochemistry. *Journal of Nutrition & Intermediary Metabolism,* 6**,** 8-15.
BELKADI, A., BOLZE, A., ITAN, Y., COBAT, A., VINCENT, Q. B., ANTIPENKO, A., SHANG, L., BOISSON, B., CASANOVA, J. L. & ABEL, L. 2015. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A,* 112**,** 5473-8.
BELLEVILLE, S., CHERTKOW, H. & GAUTHIER, S. 2007. Working memory and control of attention in persons with Alzheimer's disease and mild cognitive impairment. *Neuropsychology,* 21**,** 458-69.
BENATAR, M., WUU, J., FERNANDEZ, C., WEIHL, C. C., KATZEN, H., STEELE, J., OSKARSSON, B. & TAYLOR, J. P. 2013. Motor neuron involvement in multisystem proteinopathy: implications for ALS. *Neurology,* 80**,** 1874-80.
BENEDICT, C., JACOBSSON, J. A., RONNEMAA, E., SALLMAN-ALMEN, M., BROOKS, S., SCHULTES, B., FREDRIKSSON, R., LANNFELT, L., KILANDER, L. & SCHIOTH, H. B. 2011. The fat mass and obesity gene is linked to reduced verbal fluency in overweight and obese elderly men. *Neurobiol Aging,* 32**,** 1159 e1-5.

BERDASCO, M. & ESTELLER, M. 2013. Genetic syndromes caused by mutations in epigenetic genes. *Hum Genet,* 132**,** 359-83.
BERTRAM, L. 2011. Alzheimer's genetics in the GWAS era: a continuing story of 'replications and refutations'. *Curr Neurol Neurosci Rep,* 11**,** 246-53.
BERTRAM, L. & TANZI, R. E. 2008. Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. *Nat Rev Neurosci,* 9**,** 768-78.
BESTOR, T., LAUDANO, A., MATTALIANO, R. & INGRAM, V. 1988. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J Mol Biol,* 203**,** 971-83.
BHATTACHARYA, S. K., RAMCHANDANI, S., CERVONI, N. & SZYF, M. 1999. A mammalian protein with specific demethylase activity for mCpG DNA. *Nature,* 397**,** 579-83.
BIHAQI, S. W. & ZAWIA, N. H. 2012. Alzheimer's disease biomarkers and epigenetic intermediates following exposure to Pb in vitro. *Curr Alzheimer Res,* 9**,** 555-62.
BIRD, A. 2002. DNA methylation patterns and epigenetic memory. *Genes Dev,* 16**,** 6-21.
BLACKER, D., BYRNES, M. L., MASTAGLIA, F. L. & THICKBROOM, G. W. 2006. Differential activation of frontal lobe areas by lexical and semantic language tasks: a functional magnetic resonance imaging study. *J Clin Neurosci,* 13**,** 91-5.
BLACKWELL, A. D., SAHAKIAN, B. J., VESEY, R., SEMPLE, J. M., ROBBINS, T. W. & HODGES, J. R. 2004. Detecting dementia: Novel neuropsychological markers of preclinical Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders,* 17**,** 42-48.
BLEICH, S., BANDELOW, B., JAVAHERIPOUR, K., MULLER, A., DEGNER, D., WILHELM, J., HAVEMANN-REINECKE, U., SPERLING, W., RUTHER, E. & KORNHUBER, J. 2003. Hyperhomocysteinemia as a new risk factor for brain shrinkage in patients with alcoholism. *Neurosci Lett,* 335**,** 179-82.
BOKAR, J. A., SHAMBAUGH, M. E., POLAYES, D., MATERA, A. G. & ROTTMAN, F. M. 1997. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA,* 3**,** 1233-47.
BOLLATI, V., SCHWARTZ, J., WRIGHT, R., LITONJUA, A., TARANTINI, L., SUH, H., SPARROW, D., VOKONAS, P. & BACCARELLI, A. 2009. Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech Ageing Dev,* 130**,** 234-9.
BONFILS, C., BEAULIEU, N., CHAN, E., COTTON-MONTPETIT, J. & MACLEOD, A. R. 2000. Characterization of the human DNA methyltransferase splice variant Dnmt1b. *J Biol Chem,* 275**,** 10754-60.
BOTTIGLIERI, T. 2005. Homocysteine and folate metabolism in depression. *Prog Neuropsychopharmacol Biol Psychiatry,* 29**,** 1103-12.
BRAAE, A., THOMPSON, C. E. & MORGAN, K. 2014. Comparison of custom designed KASP and TaqMan genotyping assays for a rare genetic variant identified through resequencing GWAS loci. *LGC Application note,* GAPP-0003.
BRADLEY-WHITMAN, M. A. & LOVELL, M. A. 2013. Epigenetic changes in the progression of Alzheimer's disease. *Mech Ageing Dev,* 134**,** 486-95.

BRYAN, J. & CALVARESI, E. 2004. Associations between dietary intake of folate and vitamins B-12 and B-6 and self-reported cognitive function and psychological well-being in Australian men and women in midlife. *J Nutr Health Aging,* 8**,** 226-32.

BRYAN, J., CALVARESI, E. & HUGHES, D. 2002. Short-term folate, vitamin B-12 or vitamin B-6 supplementation slightly affects memory performance but not mood in women of various ages. *J Nutr,* 132**,** 1345-56.

BUDGE, M., JOHNSTON, C., HOGERVORST, E., DE JAGER, C., MILWAIN, E., IVERSEN, S. D., BARNETSON, L., KING, E. & SMITH, A. D. 2000. Plasma total homocysteine and cognitive performance in a volunteer elderly population. *Ann N Y Acad Sci,* 903**,** 407-10.

BURT, D. B., ZEMBAR, M. J. & NIEDEREHE, G. 1995. Depression and memory impairment: a meta-analysis of the association, its pattern, and specificity. *Psychol Bull,* 117**,** 285-305.

BUSH, W. S. & MOORE, J. H. 2012. Chapter 11: Genome-wide association studies. *PLoS Comput Biol,* 8**,** e1002822.

BUTTERFIELD, D. A., POCERNICH, C. B. & DRAKE, J. 2002. Elevated glutathione as a therapeutic strategy in Alzheimer's disease. *Drug Development Research,* 56**,** 428-437.

BUTTERFIELD, D. A., REED, T., PERLUIGI, M., DE MARCO, C., COCCIA, R., CINI, C. & SULTANA, R. 2006. Elevated protein-bound levels of the lipid peroxidation product, 4-hydroxy-2-nonenal, in brain from persons with mild cognitive impairment. *Neurosci Lett,* 397**,** 170-3.

CALLAGHAN, B. L., LI, S. & RICHARDSON, R. 2014. The elusive engram: what can infantile amnesia tell us about memory? *Trends Neurosci,* 37**,** 47-53.

CALVARESI, E. & BRYAN, J. 2001. B vitamins, cognition, and aging: a review. *J Gerontol B Psychol Sci Soc Sci,* 56**,** P327-39.

CAMBRIDGE BIOSYSTEMS. Cambridge, UK.

CARDELLI, M., MARCHEGIANI, F., CAVALLONE, L., OLIVIERI, F., GIOVAGNETTI, S., MUGIANESI, E., MORESI, R., LISA, R. & FRANCESCHI, C. 2006. A polymorphism of the YTHDF2 gene (1p35) located in an Alu-rich genomic domain is associated with human longevity. *J Gerontol A Biol Sci Med Sci,* 61**,** 547-56.

CARRASQUILLO, M. M., CROOK, J. E., PEDRAZA, O., THOMAS, C. S., PANKRATZ, V. S., ALLEN, M., NGUYEN, T., MALPHRUS, K. G., MA, L., BISCEGLIO, G. D., ROBERTS, R. O., LUCAS, J. A., SMITH, G. E., IVNIK, R. J., MACHULDA, M. M., GRAFF-RADFORD, N. R., PETERSEN, R. C., YOUNKIN, S. G. & ERTEKIN-TANER, N. 2015. Late-onset Alzheimer's risk variants in memory decline, incident mild cognitive impairment, and Alzheimer's disease. *Neurobiol Aging,* 36**,** 60-7.

CAVALLARO, R. A., FUSO, A., NICOLIA, V. & SCARPA, S. 2010. S-adenosylmethionine prevents oxidative stress and modulates glutathione metabolism in TgCRND8 mice fed a B-vitamin deficient diet. *J Alzheimers Dis,* 20**,** 997-1002.

CEBRIAN, A., PHAROAH, P. D., AHMED, S., ROPERO, S., FRAGA, M. F., SMITH, P. L., CONROY, D., LUBEN, R., PERKINS, B., EASTON, D. F., DUNNING, A. M., ESTELLER, M. & PONDER, B. A. 2006. Genetic variants in epigenetic genes and breast cancer risk. *Carcinogenesis,* 27**,** 1661-9.

CHAHROUR, M., JUNG, S. Y., SHAW, C., ZHOU, X., WONG, S. T., QIN, J. & ZOGHBI, H. Y. 2008. MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science,* 320**,** 1224-9.
CHALON, S. 2006. Omega-3 fatty acids and monoamine neurotransmission. *Prostaglandins Leukot Essent Fatty Acids,* 75**,** 259-69.
CHANG, S. H., ELEMENTO, O., ZHANG, J., ZHUANG, Z. W., SIMONS, M. & HLA, T. 2014. ELAVL1 regulates alternative splicing of eIF4E transporter to promote postnatal angiogenesis. *Proc Natl Acad Sci U S A,* 111**,** 18309-14.
CHEDIN, F., LIEBER, M. R. & HSIEH, C. L. 2002. The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proc Natl Acad Sci U S A,* 99**,** 16916-21.
CHEN, C. C., WANG, K. Y. & SHEN, C. K. 2012. The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases. *J Biol Chem,* 287**,** 33116-21.
CHEN, C. C., WANG, K. Y. & SHEN, C. K. 2013. DNA 5-methylcytosine demethylation activities of the mammalian DNA methyltransferases. *J Biol Chem,* 288**,** 9084-91.
CHEN, J., HE, J., OGDEN, L. G., BATUMAN, V. & WHELTON, P. K. 2002. Relationship of serum antioxidant vitamins to serum creatinine in the US population. *Am J Kidney Dis,* 39**,** 460-8.
CHEN, T., PEARCE, L. L., PETERSON, J., STOYANOVSKY, D. & BILLIAR, T. R. 2005. Glutathione depletion renders rat hepatocytes sensitive to nitric oxide donor-mediated toxicity. *Hepatology,* 42**,** 598-607.
CHEN, W. G., CHANG, Q., LIN, Y., MEISSNER, A., WEST, A. E., GRIFFITH, E. C., JAENISCH, R. & GREENBERG, M. E. 2003. Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. *Science,* 302**,** 885-9.
CHEN, Z. Y., CAI, L., ZHU, J., CHEN, M., CHEN, J., LI, Z. H., LIU, X. D., WANG, S. G., BIE, P., JIANG, P., DONG, J. H. & LI, X. W. 2011. Fyn requires HnRNPA2B1 and Sam68 to synergistically regulate apoptosis in pancreatic cancer. *Carcinogenesis,* 32**,** 1419-26.
CHENG, Y., BERNSTEIN, A., CHEN, D. & JIN, P. 2015. 5-Hydroxymethylcytosine: A new player in brain disorders? *Exp Neurol,* 268**,** 3-9.
CHERUBINI, A., MARTIN, A., ANDRES-LACUEVA, C., DI IORIO, A., LAMPONI, M., MECOCCI, P., BARTALI, B., CORSI, A., SENIN, U. & FERRUCCI, L. 2005. Vitamin E levels, cognitive impairment and dementia in older persons: the InCHIANTI study. *Neurobiol Aging,* 26**,** 987-94.
CHI, M. N., AURIOL, J., JEGOU, B., KONTOYIANNIS, D. L., TURNER, J. M., DE ROOIJ, D. G. & MORELLO, D. 2011. The RNA-binding protein ELAVL1/HuR is essential for mouse spermatogenesis, acting both at meiotic and postmeiotic stages. *Mol Biol Cell,* 22**,** 2875-85.
CHIN, H. G., PONNALURI, V. K., ZHANG, G., ESTEVE, P. O., SCHAUS, S. E., HANSEN, U. & PRADHAN, S. 2016. Transcription factor LSF-DNMT1 complex dissociation by FQI1 leads to aberrant DNA methylation and gene expression. *Oncotarget,* 7**,** 83627-83640.
CHIU, C. C., SU, K. P., CHENG, T. C., LIU, H. C., CHANG, C. J., DEWEY, M. E., STEWART, R. & HUANG, S. Y. 2008. The effects of omega-3 fatty acids monotherapy in Alzheimer's disease and mild cognitive impairment: a

preliminary randomized double-blind placebo-controlled study. *Prog Neuropsychopharmacol Biol Psychiatry,* 32**,** 1538-44.
CHOULIARAS, L., MASTROENI, D., DELVAUX, E., GROVER, A., KENIS, G., HOF, P. R., STEINBUSCH, H. W., COLEMAN, P. D., RUTTEN, B. P. & VAN DEN HOVE, D. L. 2013. Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiol Aging,* 34**,** 2091-9.
CHUNG, W. K., EVANS, S. T., FREED, A. S., KEBA, J. J., BAER, Z. C., REGE, K. & CRAMER, S. M. 2010. Utilization of Lysozyme Charge Ladders to Examine the Effects of Protein Surface Charge Distribution on Binding Affinity in Ion Exchange Systems. *Langmuir,* 26**,** 759-768.
CINO, E. A., CHOY, W. Y. & KARTTUNEN, M. 2012. Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation,* 8**,** 2725-2740.
CLARKE, G. M., ANDERSON, C. A., PETTERSSON, F. H., CARDON, L. R., MORRIS, A. P. & ZONDERVAN, K. T. 2011. Basic statistical analysis in genetic case-control studies. *Nat Protoc,* 6**,** 121-33.
CLARKE, R., BENNETT, D., PARISH, S., LEWINGTON, S., SKEAFF, M., EUSSEN, S. J., LEWERIN, C., STOTT, D. J., ARMITAGE, J., HANKEY, G. J., LONN, E., SPENCE, J. D., GALAN, P., DE GROOT, L. C., HALSEY, J., DANGOUR, A. D., COLLINS, R., GRODSTEIN, F. & COLLABORATION, B. V. T. T. 2014. Effects of homocysteine lowering with B vitamins on cognitive aging: meta-analysis of 11 trials with cognitive data on 22,000 individuals. *Am J Clin Nutr,* 100**,** 657-66.
COCKLE, S. M., HALLER, J., KIMBER, S., DAWE, R. A. & HINDMARCH, I. 2000. The influence of multivitamins on cognitive function and mood in the elderly. *Aging and Mental Health,* 4**,** 339-353.
COEN, R. F., KIRBY, M., SWANWICK, G. R., MAGUIRE, C. P., WALSH, J. B., COAKLEY, D., O'NEILL, D. & LAWLOR, B. A. 1997. Distinguishing between patients with depression or very mild Alzheimer's disease using the Delayed-Word-Recall test. *Dement Geriatr Cogn Disord,* 8**,** 244-7.
COLOM, R., ABAD, F. J., QUIROGA, M. A., SHIH, P. C. & FLORES-MENDOZA, C. 2008. Working memory and intelligence are highly related constructs, but why? *Intelligence,* 36**,** 584-606.
CONDLIFFE, D., WONG, A., TROAKES, C., PROITSI, P., PATEL, Y., CHOULIARAS, L., FERNANDES, C., COOPER, J., LOVESTONE, S., SCHALKWYK, L., MILL, J. & LUNNON, K. 2014. Cross-region reduction in 5-hydroxymethylcytosine in Alzheimer's disease brain. *Neurobiol Aging,* 35**,** 1850-4.
CONQUER, J. A., TIERNEY, M. C., ZECEVIC, J., BETTGER, W. J. & FISHER, R. H. 2000. Fatty acid analysis of blood plasma of patients with Alzheimer's disease, other types of dementia, and cognitive impairment. *Lipids,* 35**,** 1305-12.
CONSORTIUM, G. T. 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science,* 348**,** 648-60.
CONSORTIUM, U. K., WALTER, K., MIN, J. L., HUANG, J., CROOKS, L., MEMARI, Y., MCCARTHY, S., PERRY, J. R., XU, C., FUTEMA, M., LAWSON, D., IOTCHKOVA, V., SCHIFFELS, S., HENDRICKS, A. E., DANECEK, P., LI, R., FLOYD, J., WAIN, L. V., BARROSO, I., HUMPHRIES, S. E., HURLES, M. E.,

ZEGGINI, E., BARRETT, J. C., PLAGNOL, V., RICHARDS, J. B., GREENWOOD, C. M., TIMPSON, N. J., DURBIN, R. & SORANZO, N. 2015. The UK10K project identifies rare variants in health and disease. *Nature,* 526**,** 82-90.

COOPER, D. B., LACRITZ, L. H., WEINER, M. F., ROSENBERG, R. N. & CULLUM, C. M. 2004. Category fluency in mild cognitive impairment: reduced effect of practice in test-retest conditions. *Alzheimer Dis Assoc Disord,* 18**,** 120-2.

COPPIETERS, N., DIERIKS, B. V., LILL, C., FAULL, R. L., CURTIS, M. A. & DRAGUNOW, M. 2014. Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. *Neurobiol Aging,* 35**,** 1334-44.

CORBO, R. M. & SCACCHI, R. 1999. Apolipoprotein E (APOE) allele distribution in the world. Is APOE*4 a 'thrifty' allele? *Ann Hum Genet,* 63**,** 301-10.

CORTINA-BORJA, M., SMITH, A. D., COMBARROS, O. & LEHMANN, D. J. 2009. The synergy factor: a statistic to measure interactions in complex diseases. *BMC Res Notes,* 2**,** 105.

COSTELLO, A. B. & OSBORNE, J. W. 2005. Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment, Research & Evaluation,* 10.

COUILLARD-DESPRES, S., WINNER, B., SCHAUBECK, S., AIGNER, R., VROEMEN, M., WEIDNER, N., BOGDAHN, U., WINKLER, J., KUHN, H. G. & AIGNER, L. 2005. Doublecortin expression levels in adult brain reflect neurogenesis. *Eur J Neurosci,* 21**,** 1-14.

CROMEY, D. W. 2010. Avoiding twisted pixels: ethical guidelines for the appropriate use and manipulation of scientific digital images. *Sci Eng Ethics,* 16**,** 639-67.

CRUPI, R., CAMBIAGHI, M., DECKELBAUM, R., HANSEN, I., MINDES, J., SPINA, E. & BATTAGLIA, F. 2012. n-3 fatty acids prevent impairment of neurogenesis and synaptic plasticity in B-cell activating factor (BAFF) transgenic mice. *Prev Med,* 54 Suppl**,** S103-8.

CURRADI, M., IZZO, A., BADARACCO, G. & LANDSBERGER, N. 2002. Molecular mechanisms of gene silencing mediated by DNA methylation. *Mol Cell Biol,* 22**,** 3157-73.

DAI, S., ZHANG, J., HUANG, S., LOU, B., FANG, B., YE, T., HUANG, X., CHEN, B. & ZHOU, M. 2017. HNRNPA2B1 regulates the epithelial-mesenchymal transition in pancreatic cancer cells through the ERK/snail signalling pathway. *Cancer Cell Int,* 17**,** 12.

DAO, T., CHENG, R. Y., REVELO, M. P., MITZNER, W. & TANG, W. 2014. Hydroxymethylation as a Novel Environmental Biosensor. *Curr Environ Health Rep,* 1**,** 1-10.

DAS, S., FORER, L., SCHONHERR, S., SIDORE, C., LOCKE, A. E., KWONG, A., VRIEZE, S. I., CHEW, E. Y., LEVY, S., MCGUE, M., SCHLESSINGER, D., STAMBOLIAN, D., LOH, P. R., IACONO, W. G., SWAROOP, A., SCOTT, L. J., CUCCA, F., KRONENBERG, F., BOEHNKE, M., ABECASIS, G. R. & FUCHSBERGER, C. 2016. Next-generation genotype imputation service and methods. *Nat Genet,* 48**,** 1284-1287.

DAVIES, G., TENESA, A., PAYTON, A., YANG, J., HARRIS, S. E., LIEWALD, D., KE, X., LE HELLARD, S., CHRISTOFOROU, A., LUCIANO, M., MCGHEE, K., LOPEZ, L., GOW, A. J., CORLEY, J., REDMOND, P., FOX, H. C., HAGGARTY, P., WHALLEY, L. J., MCNEILL, G., GODDARD, M. E., ESPESETH, T., LUNDERVOLD, A. J.,

REINVANG, I., PICKLES, A., STEEN, V. M., OLLIER, W., PORTEOUS, D. J., HORAN, M., STARR, J. M., PENDLETON, N., VISSCHER, P. M. & DEARY, I. J. 2011. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry,* 16**,** 996-1005.
DAVIES, R. R., GRAHAM, K. S., XUEREB, J. H., WILLIAMS, G. B. & HODGES, J. R. 2004. The human perirhinal cortex and semantic memory. *Eur J Neurosci,* 20**,** 2441-6.
DAY, J. J. & SWEATT, J. D. 2011. Cognitive neuroepigenetics: a role for epigenetic mechanisms in learning and memory. *Neurobiol Learn Mem,* 96**,** 2-12.
DE BREE, A., VERSCHUREN, W. M., BJORKE-MONSEN, A. L., VAN DER PUT, N. M., HEIL, S. G., TRIJBELS, F. J. & BLOM, H. J. 2003. Effect of the methylenetetrahydrofolate reductase 677C-->T mutation on the relations among folate intake and plasma folate and homocysteine concentrations in a general population sample. *Am J Clin Nutr,* 77**,** 687-93.
DE JAGER, C. A., HOGERVORST, E., COMBRINCK, M. & BUDGE, M. M. 2003. Sensitivity and specificity of neuropsychological tests for mild cognitive impairment, vascular cognitive impairment and Alzheimer's disease. *Psychol Med,* 33**,** 1039-50.
DE JAGER, C. A., OULHAJ, A., JACOBY, R., REFSUM, H. & SMITH, A. D. 2012. Cognitive and clinical outcomes of homocysteine-lowering B-vitamin treatment in mild cognitive impairment: a randomized controlled trial. *Int J Geriatr Psychiatry,* 27**,** 592-600.
DE ROVER, M., PIRONTI, V. A., MCCABE, J. A., ACOSTA-CABRONERO, J., ARANA, F. S., MOREIN-ZAMIR, S., HODGES, J. R., ROBBINS, T. W., FLETCHER, P. C., NESTOR, P. J. & SAHAKIAN, B. J. 2011. Hippocampal dysfunction in patients with mild cognitive impairment: a functional neuroimaging study of a visuospatial paired associates learning task. *Neuropsychologia,* 49**,** 2060-70.
DELPECH, J. C., THOMAZEAU, A., MADORE, C., BOSCH-BOUJU, C., LARRIEU, T., LACABANNE, C., REMUS-BOREL, J., AUBERT, A., JOFFRE, C., NADJAR, A. & LAYE, S. 2015. Dietary n-3 PUFAs Deficiency Increases Vulnerability to Inflammation-Induced Spatial Memory Impairment. *Neuropsychopharmacology,* 40**,** 2774-87.
DEPLUS, R., BRENNER, C., BURGERS, W. A., PUTMANS, P., KOUZARIDES, T., DE LAUNOIT, Y. & FUKS, F. 2002. Dnmt3L is a transcriptional repressor that recruits histone deacetylase. *Nucleic Acids Res,* 30**,** 3831-8.
DETH, R., MURATORE, C., BENZECRY, J., POWER-CHARNITSKY, V. A. & WALY, M. 2008. How environmental and genetic factors combine to cause autism: A redox/methylation hypothesis. *Neurotoxicology,* 29**,** 190-201.
DEVYS, D., BIANCALANA, V., ROUSSEAU, F., BOUE, J., MANDEL, J. L. & OBERLE, I. 1992. Analysis of full fragile X mutations in fetal tissues and monozygotic twins indicate that abnormal methylation and somatic heterogeneity are established early in development. *Am J Med Genet,* 43**,** 208-16.
DICKERSON, B. C., SALAT, D. H., GREVE, D. N., CHUA, E. F., RAND-GIOVANNETTI, E., RENTZ, D. M., BERTRAM, L., MULLIN, K., TANZI, R. E., BLACKER, D., ALBERT, M. S. & SPERLING, R. A. 2005. Increased hippocampal activation in mild cognitive impairment compared to normal aging and AD. *Neurology,* 65**,** 404-11.

DILUCA, M. & OLESEN, J. 2014. The cost of brain diseases: a burden or a challenge? *Neuron,* 82**,** 1205-8.
DOETS, E. L., UELAND, P. M., TELL, G. S., VOLLSET, S. E., NYGARD, O. K., VAN'T VEER, P., DE GROOT, L. C., NURK, E., REFSUM, H., SMITH, A. D. & EUSSEN, S. J. 2014. Interactions between plasma concentrations of folate and markers of vitamin B(12) status with cognitive performance in elderly people not exposed to folic acid fortification: the Hordaland Health Study. *Br J Nutr,* 111**,** 1085-95.
DOI, A., PARK, I. H., WEN, B., MURAKAMI, P., ARYEE, M. J., IRIZARRY, R., HERB, B., LADD-ACOSTA, C., RHO, J., LOEWER, S., MILLER, J., SCHLAEGER, T., DALEY, G. Q. & FEINBERG, A. P. 2009. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet,* 41**,** 1350-3.
DOLINSKY, T. J., CZODROWSKI, P., LI, H., NIELSEN, J. E., JENSEN, J. H., KLEBE, G. & BAKER, N. A. 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Research,* 35**,** W522-W525.
DOLINSKY, T. J., NIELSEN, J. E., MCCAMMON, J. A. & BAKER, N. A. 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research,* 32**,** W665-W667.
DOMINISSINI, D., MOSHITCH-MOSHKOVITZ, S., SCHWARTZ, S., SALMON-DIVON, M., UNGAR, L., OSENBERG, S., CESARKAS, K., JACOB-HIRSCH, J., AMARIGLIO, N., KUPIEC, M., SOREK, R. & RECHAVI, G. 2012. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature,* 485**,** 201-6.
DONG, C., WEI, P., JIAN, X., GIBBS, R., BOERWINKLE, E., WANG, K. & LIU, X. 2015. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet,* 24**,** 2125-37.
DOUAUD, G., REFSUM, H., DE JAGER, C. A., JACOBY, R., NICHOLS, T. E., SMITH, S. M. & SMITH, A. D. 2013. Preventing Alzheimer's disease-related gray matter atrophy by B-vitamin treatment. *Proc Natl Acad Sci U S A,* 110**,** 9523-8.
DOWLING, P., POLLARD, D., LARKIN, A., HENRY, M., MELEADY, P., GATELY, K., O'BYRNE, K., BARR, M. P., LYNCH, V., BALLOT, J., CROWN, J., MORIARTY, M., O'BRIEN, E., MORGAN, R. & CLYNES, M. 2015. Abnormal levels of heterogeneous nuclear ribonucleoprotein A2B1 (hnRNPA2B1) in tumour tissue and blood samples from patients diagnosed with lung cancer. *Mol Biosyst,* 11**,** 743-52.
DRINGEN, R. 2000. Metabolism and functions of glutathione in brain. *Prog Neurobiol,* 62**,** 649-71.
DRISCOLL, D. J., WATERS, M. F., WILLIAMS, C. A., ZORI, R. T., GLENN, C. C., AVIDANO, K. M. & NICHOLLS, R. D. 1992. A DNA methylation imprint, determined by the sex of the parent, distinguishes the Angelman and Prader-Willi syndromes. *Genomics,* 13**,** 917-24.
DU, A. T., SCHUFF, N., KRAMER, J. H., GANZER, S., ZHU, X. P., JAGUST, W. J., MILLER, B. L., REED, B. R., MUNGAS, D., YAFFE, K., CHUI, H. C. & WEINER,

M. W. 2004. Higher atrophy rate of entorhinal cortex than hippocampus in AD. *Neurology,* 62**,** 422-7.

DU, J., JOHNSON, L. M., JACOBSEN, S. E. & PATEL, D. J. 2015. DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol,* 16**,** 519-32.

DUARA, R., LOEWENSTEIN, D. A., GREIG, M. T., POTTER, E., BARKER, W., RAJ, A., SCHINKA, J., BORENSTEIN, A., SCHOENBERG, M., WU, Y., BANKO, J. & POTTER, H. 2011. Pre-MCI and MCI: neuropsychological, clinical, and imaging features and progression rates. *Am J Geriatr Psychiatry,* 19**,** 951-60.

DUDAS, R. B., CLAGUE, F., THOMPSON, S. A., GRAHAM, K. S. & HODGES, J. R. 2005. Episodic and semantic memory in mild cognitive impairment. *Neuropsychologia,* 43**,** 1266-76.

DURGA, J., VAN BOXTEL, M. P., SCHOUTEN, E. G., KOK, F. J., JOLLES, J., KATAN, M. B. & VERHOEF, P. 2007. Effect of 3-year folic acid supplementation on cognitive function in older adults in the FACIT trial: a randomised, double blind, controlled trial. *Lancet,* 369**,** 208-16.

EDELHEIT, S., SCHWARTZ, S., MUMBACH, M. R., WURTZEL, O. & SOREK, R. 2013. Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genet,* 9**,** e1003602.

EDGAR, R., TAN, P. P., PORTALES-CASAMAR, E. & PAVLIDIS, P. 2014. Meta-analysis of human methylomes reveals stably methylated sequences surrounding CpG islands associated with high gene expression. *Epigenetics Chromatin,* 7**,** 28.

EL-MAARRI, O., KARETA, M. S., MIKESKA, T., BECKER, T., DIAZ-LACAVA, A., JUNEN, J., NUSGEN, N., BEHNE, F., WIENKER, T., WAHA, A., OLDENBURG, J. & CHEDIN, F. 2009. A systematic search for DNA methyltransferase polymorphisms reveals a rare DNMT3L variant associated with subtelomeric hypomethylation. *Hum Mol Genet,* 18**,** 1755-68.

ELEFTHERIOU, M., PASCUAL, A. J., WHELDON, L. M., PERRY, C., ABAKIR, A., ARORA, A., JOHNSON, A. D., AUER, D. T., ELLIS, I. O., MADHUSUDAN, S. & RUZOV, A. 2015. 5-Carboxylcytosine levels are elevated in human breast cancers and gliomas. *Clin Epigenetics,* 7**,** 88.

ELIAS, M. F., ROBBINS, M. A., BUDGE, M. M., ELIAS, P. K., DORE, G. A., BRENNAN, S. L., JOHNSTON, C. & NAGY, Z. 2008. Homocysteine and cognitive performance: modification by the ApoE genotype. *Neurosci Lett,* 430**,** 64-9.

ELLISON, E. M., ABNER, E. L. & LOVELL, M. A. 2017. Multiregional analysis of global 5-methylcytosine and 5-hydroxymethylcytosine throughout the progression of Alzheimer's disease. *J Neurochem,* 140**,** 383-394.

ERIKSDOTTER, M., VEDIN, I., FALAHATI, F., FREUND-LEVI, Y., HJORTH, E., FAXEN-IRVING, G., WAHLUND, L. O., SCHULTZBERG, M., BASUN, H., CEDERHOLM, T. & PALMBLAD, J. 2015. Plasma Fatty Acid Profiles in Relation to Cognition and Gender in Alzheimer's Disease Patients During Oral Omega-3 Fatty Acid Supplementation: The OmegAD Study. *J Alzheimers Dis,* 48**,** 805-12.

ESANOV, R., ANDRADE, N. S., BENNISON, S., WAHLESTEDT, C. & ZEIER, Z. 2016. The FMR1 promoter is selectively hydroxymethylated in primary neurons of fragile X syndrome patients. *Hum Mol Genet,* 25**,** 4870-4880.

EUSSEN, S. J., DE GROOT, L. C., JOOSTEN, L. W., BLOO, R. J., CLARKE, R., UELAND, P. M., SCHNEEDE, J., BLOM, H. J., HOEFNAGELS, W. H. & VAN STAVEREN, W. A. 2006. Effect of oral vitamin B-12 with or without folic acid on cognitive function in older people with mild vitamin B-12 deficiency: a randomized, placebo-controlled trial. *Am J Clin Nutr,* 84**,** 361-70.

FABRIGAR, L. R., WEGENER, D. T., MACCALLUM, R. C. & STRAHAN, E. J. 1999. Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods,* 4**,** 272-299.

FAN, G., BEARD, C., CHEN, R. Z., CSANKOVSZKI, G., SUN, Y., SINIAIA, M., BINISZKIEWICZ, D., BATES, B., LEE, P. P., KUHN, R., TRUMPP, A., POON, C., WILSON, C. B. & JAENISCH, R. 2001. DNA hypomethylation perturbs the function and survival of CNS neurons in postnatal animals. *J Neurosci,* 21**,** 788-97.

FEINBERG, J. I., BAKULSKI, K. M., JAFFE, A. E., TRYGGVADOTTIR, R., BROWN, S. C., GOLDMAN, L. R., CROEN, L. A., HERTZ-PICCIOTTO, I., NEWSCHAFFER, C. J., FALLIN, M. D. & FEINBERG, A. P. 2015. Paternal sperm DNA methylation associated with early signs of autism risk in an autism-enriched cohort. *Int J Epidemiol,* 44**,** 1199-210.

FENG, J., SHAO, N., SZULWACH, K. E., VIALOU, V., HUYNH, J., ZHONG, C., LE, T., FERGUSON, D., CAHILL, M. E., LI, Y., KOO, J. W., RIBEIRO, E., LABONTE, B., LAITMAN, B. M., ESTEY, D., STOCKMAN, V., KENNEDY, P., COUROUSSE, T., MENSAH, I., TURECKI, G., FAULL, K. F., MING, G. L., SONG, H., FAN, G., CASACCIA, P., SHEN, L., JIN, P. & NESTLER, E. J. 2015. Role of Tet1 and 5-hydroxymethylcytosine in cocaine action. *Nat Neurosci,* 18**,** 536-44.

FENG, J., ZHOU, Y., CAMPBELL, S. L., LE, T., LI, E., SWEATT, J. D., SILVA, A. J. & FAN, G. 2010. Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat Neurosci,* 13**,** 423-30.

FERNANDEZ-ROIG, S., LAI, S. C., MURPHY, M. M., FERNANDEZ-BALLART, J. & QUADROS, E. V. 2012. Vitamin B12 deficiency in the brain leads to DNA hypomethylation in the TCblR/CD320 knockout mouse. *Nutr Metab (Lond),* 9**,** 41.

FILIPPINI, N., MACINTOSH, B. J., HOUGH, M. G., GOODWIN, G. M., FRISONI, G. B., SMITH, S. M., MATTHEWS, P. M., BECKMANN, C. F. & MACKAY, C. E. 2009. Distinct patterns of brain activity in young carriers of the APOE-epsilon 4 allele. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 7209-7214.

FLANAGAN, S. E., PATCH, A. M. & ELLARD, S. 2010. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet Test Mol Biomarkers,* 14**,** 533-7.

FLINT, J. & MUNAFO, M. R. 2007. The endophenotype concept in psychiatric genetics. *Psychol Med,* 37**,** 163-80.

FLOTT-RAHMEL, B., SCHURMANN, M., SCHLUFF, P., FINGERHUT, R., MUSSHOFF, U., FOWLER, B. & ULLRICH, K. 1998. Homocysteic and homocysteine sulphinic acid exhibit excitotoxicity in organotypic cultures from rat brain. *Eur J Pediatr,* 157 Suppl 2**,** S112-7.

FORAKER, J., MILLARD, S. P., LEONG, L., THOMSON, Z., CHEN, S., KEENE, C. D., BEKRIS, L. M. & YU, C. E. 2015. The APOE Gene is Differentially Methylated in Alzheimer's Disease. *J Alzheimers Dis,* 48**,** 745-55.

FOWLER, K. S., SALING, M. M., CONWAY, E. L., SEMPLE, J. M. & LOUIS, W. J. 1995. Computerized delayed matching to sample and paired associate performance in the early detection of dementia. *Appl Neuropsychol,* 2**,** 72-8.

FRAGA, M. F., BALLESTAR, E., PAZ, M. F., ROPERO, S., SETIEN, F., BALLESTAR, M. L., HEINE-SUNER, D., CIGUDOSA, J. C., URIOSTE, M., BENITEZ, J., BOIX-CHORNET, M., SANCHEZ-AGUILERA, A., LING, C., CARLSSON, E., POULSEN, P., VAAG, A., STEPHAN, Z., SPECTOR, T. D., WU, Y. Z., PLASS, C. & ESTELLER, M. 2005. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A,* 102**,** 10604-9.

FRANCOIS, O., CURRAT, M., RAY, N., HAN, E., EXCOFFIER, L. & NOVEMBRE, J. 2010. Principal component analysis under population genetic models of range expansion and admixture. *Mol Biol Evol,* 27**,** 1257-68.

FREY, U. & MORRIS, R. G. 1997. Synaptic tagging and long-term potentiation. *Nature,* 385**,** 533-6.

FU, L., GUERRERO, C. R., ZHONG, N., AMATO, N. J., LIU, Y., LIU, S., CAI, Q., JI, D., JIN, S. G., NIEDERNHOFER, L. J., PFEIFER, G. P., XU, G. L. & WANG, Y. 2014. Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J Am Chem Soc,* 136**,** 11582-5.

FUJITA, N., WATANABE, S., ICHIMURA, T., TSURUZOE, S., SHINKAI, Y., TACHIBANA, M., CHIBA, T. & NAKAO, M. 2003. Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. *J Biol Chem,* 278**,** 24132-8.

FUSO, A., SEMINARA, L., CAVALLARO, R. A., D'ANSELMI, F. & SCARPA, S. 2005. S-adenosylmethionine/homocysteine cycle alterations modify DNA methylation status with consequent deregulation of PS1 and BACE and beta-amyloid production. *Mol Cell Neurosci,* 28**,** 195-204.

GALLOWAY, P. H., SAHGAL, A., MCKEITH, I. G., LLOYD, S., COOK, J. H., FERRIER, I. N. & EDWARDSON, J. A. 1992. Visual-Pattern Recognition Memory and Learning-Deficits in Senile Dementias of Alzheimer and Lewy Body Types. *Dementia,* 3**,** 101-107.

GARBER, K. B., VISOOTSAK, J. & WARREN, S. T. 2008. Fragile X syndrome. *Eur J Hum Genet,* 16**,** 666-72.

GARDINI, S., CONCARI, L., PAGLIARA, S., GHETTI, C., VENNERI, A. & CAFFARRA, P. 2011. Visuo-spatial imagery impairment in posterior cortical atrophy: a cognitive and SPECT study. *Behav Neurol,* 24**,** 123-32.

GIBSON, G. 2012. Rare and common variants: twenty arguments. *Nat Rev Genet,* 13**,** 135-45.

GODLER, D. E., SLATER, H. R., BUI, Q. M., STOREY, E., ONO, M. Y., GEHLING, F., INABA, Y., FRANCIS, D., HOPPER, J. L., KINSELLA, G., AMOR, D. J., HAGERMAN, R. J. & LOESCH, D. Z. 2012. Fragile X mental retardation 1 (FMR1) intron 1 methylation in blood predicts verbal cognitive impairment in female carriers of expanded FMR1 alleles: evidence from a pilot study. *Clin Chem,* 58**,** 590-8.

GOLLA, J. P., ZHAO, J., MANN, I. K., SAYEED, S. K., MANDAL, A., ROSE, R. B. & VINSON, C. 2014. Carboxylation of cytosine (5caC) in the CG dinucleotide in the E-box motif (CGCAG|GTG) increases binding of the Tcf3|Ascl1 helix-loop-helix heterodimer 10-fold. *Biochem Biophys Res Commun,* 449**,** 248-55.

GOLOMB, J., KLUGER, A., DE LEON, M. J., FERRIS, S. H., CONVIT, A., MITTELMAN, M. S., COHEN, J., RUSINEK, H., DE SANTI, S. & GEORGE, A. E. 1994. Hippocampal formation size in normal human aging: a correlate of delayed secondary memory performance. *Learn Mem,* 1**,** 45-54.

GONZALES, B., HENNING, D., SO, R. B., DIXON, J., DIXON, M. J. & VALDEZ, B. C. 2005. The Treacher Collins syndrome (TCOF1) gene product is involved in pre-rRNA methylation. *Hum Mol Genet,* 14**,** 2035-43.

GRAF, P. & SCHACTER, D. L. 1985. Implicit and explicit memory for new associations in normal and amnesic subjects. *J Exp Psychol Learn Mem Cogn,* 11**,** 501-18.

GRAFFELMAN, J. & CAMARENA, J. M. 2008. Graphical tests for Hardy-Weinberg equilibrium based on the ternary plot. *Hum Hered,* 65**,** 77-84.

GREEN, A. E., MUNAFO, M. R., DEYOUNG, C. G., FOSSELLA, J. A., FAN, J. & GRAY, J. R. 2008. Using genetic data in cognitive neuroscience: from growing pains to genuine insights. *Nat Rev Neurosci,* 9**,** 710-20.

GRIFFITH, J. S. & MAHLER, H. R. 1969. DNA ticketing theory of memory. *Nature,* 223**,** 580-2.

GRODSTEIN, F., O'BRIEN, J., KANG, J. H., DUSHKES, R., COOK, N. R., OKEREKE, O., MANSON, J. E., GLYNN, R. J., BURING, J. E., GAZIANO, M. & SESSO, H. D. 2013. Long-term multivitamin supplementation and cognitive function in men: a randomized trial. *Ann Intern Med,* 159**,** 806-14.

GROVER, V. K., COLE, D. E. & HAMILTON, D. C. 2010. Attributing Hardy-Weinberg disequilibrium to population stratification and genetic association in case-control studies. *Ann Hum Genet,* 74**,** 77-87.

GRUNDMAN, M., PETERSEN, R. C., FERRIS, S. H., THOMAS, R. G., AISEN, P. S., BENNETT, D. A., FOSTER, N. L., JACK, C. R., JR., GALASKO, D. R., DOODY, R., KAYE, J., SANO, M., MOHS, R., GAUTHIER, S., KIM, H. T., JIN, S., SCHULTZ, A. N., SCHAFER, K., MULNARD, R., VAN DYCK, C. H., MINTZER, J., ZAMRINI, E. Y., CAHN-WEINER, D., THAL, L. J. & ALZHEIMER'S DISEASE COOPERATIVE, S. 2004. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Arch Neurol,* 61**,** 59-66.

GUO, J. U., MA, D. K., MO, H., BALL, M. P., JANG, M. H., BONAGUIDI, M. A., BALAZER, J. A., EAVES, H. L., XIE, B., FORD, E., ZHANG, K., MING, G. L., GAO, Y. & SONG, H. 2011. Neuronal activity modifies the DNA methylation landscape in the adult brain. *Nat Neurosci,* 14**,** 1345-51.

GUO, J. U., SU, Y., SHIN, J. H., SHIN, J., LI, H., XIE, B., ZHONG, C., HU, S., LE, T., FAN, G., ZHU, H., CHANG, Q., GAO, Y., MING, G. L. & SONG, H. 2014. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat Neurosci,* 17**,** 215-22.

GUO, X., WANG, L., LI, J., DING, Z., XIAO, J., YIN, X., HE, S., SHI, P., DONG, L., LI, G., TIAN, C., WANG, J., CONG, Y. & XU, Y. 2015. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature,* 517**,** 640-4.

HAGGARTY, P., HOAD, G., HARRIS, S. E., STARR, J. M., FOX, H. C., DEARY, I. J. & WHALLEY, L. J. 2010. Human intelligence and polymorphisms in the DNA methyltransferase genes involved in epigenetic marking. *PLoS One,* 5**,** e11329.

HAN, J., LI, Y., WANG, D., WEI, C., YANG, X. & SUI, N. 2010. Effect of 5-aza-2-deoxycytidine microinjecting into hippocampus and prelimbic cortex on acquisition and retrieval of cocaine-induced place preference in C57BL/6 mice. *Eur J Pharmacol,* 642**,** 93-8.

HARDY, J. & SELKOE, D. J. 2002. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science,* 297**,** 353-6.

HARDY, J. A. & HIGGINS, G. A. 1992. Alzheimer's disease: the amyloid cascade hypothesis. *Science,* 256**,** 184-5.

HARRIS, E., MACPHERSON, H., VITETTA, L., KIRK, J., SALI, A. & PIPINGAS, A. 2012. Effects of a multivitamin, mineral and herbal supplement on cognition and blood biomarkers in older men: a randomised, placebo-controlled trial. *Hum Psychopharmacol,* 27**,** 370-7.

HATA, K., OKANO, M., LEI, H. & LI, E. 2002. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development,* 129**,** 1983-93.

HENDRICH, B., HARDELAND, U., NG, H. H., JIRICNY, J. & BIRD, A. 1999. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature,* 401**,** 301-4.

HERCEG, Z. 2007. Epigenetics and cancer: towards an evaluation of the impact of environmental and dietary factors. *Mutagenesis,* 22**,** 91-103.

HERMAN, J. G. & BAYLIN, S. B. 2003. Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med,* 349**,** 2042-54.

HERMANN, A., GOYAL, R. & JELTSCH, A. 2004. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem,* 279**,** 48350-9.

HERVOUET, E., VALLETTE, F. M. & CARTRON, P. F. 2010. Dnmt1/Transcription factor interactions: an alternative mechanism of DNA methylation inheritance. *Genes Cancer,* 1**,** 434-43.

HESS, M. E., HESS, S., MEYER, K. D., VERHAGEN, L. A., KOCH, L., BRONNEKE, H. S., DIETRICH, M. O., JORDAN, S. D., SALETORE, Y., ELEMENTO, O., BELGARDT, B. F., FRANZ, T., HORVATH, T. L., RUTHER, U., JAFFREY, S. R., KLOPPENBURG, P. & BRUNING, J. C. 2013. The fat mass and obesity associated gene (Fto) regulates activity of the dopaminergic midbrain circuitry. *Nat Neurosci,* 16**,** 1042-8.

HEVNER, R. F., HODGE, R. D., DAZA, R. A. & ENGLUND, C. 2006. Transcription factors in glutamatergic neurogenesis: conserved programs in neocortex, cerebellum, and adult hippocampus. *Neurosci Res,* 55**,** 223-33.

HICKMAN, M. J. & SAMSON, L. D. 1999. Role of DNA mismatch repair and p53 in signaling induction of apoptosis by alkylating agents. *Proc Natl Acad Sci U S A,* 96**,** 10764-9.

HIRASAWA, R., CHIBA, H., KANEDA, M., TAJIMA, S., LI, E., JAENISCH, R. & SASAKI, H. 2008. Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes Dev,* 22**,** 1607-16.

HIROTA, K., AKAGAWA, H., ONDA, H., YONEYAMA, T., KAWAMATA, T. & KASUYA, H. 2016. Association of Rare Nonsynonymous Variants in PKD1 and PKD2 with Familial Intracranial Aneurysms in a Japanese Population. *J Stroke Cerebrovasc Dis,* 25**,** 2900-2906.

HO, P. I., ASHLINE, D., DHITAVAT, S., ORTIZ, D., COLLINS, S. C., SHEA, T. B. & ROGERS, E. 2003. Folate deprivation induces neurodegeneration: roles of oxidative stress and increased homocysteine. *Neurobiol Dis,* 14**,** 32-42.

HOBBS, C. A., SHERMAN, S. L., YI, P., HOPKINS, S. E., TORFS, C. P., HINE, R. J., POGRIBNA, M., ROZEN, R. & JAMES, S. J. 2000. Polymorphisms in genes involved in folate metabolism as maternal risk factors for Down syndrome. *Am J Hum Genet,* 67**,** 623-30.

HODGES, J. R., PATTERSON, K., WARD, R., GARRARD, P., BAK, T., PERRY, R. & GREGORY, C. 1999. The differentiation of semantic dementia and frontal lobe dementia (temporal and frontal variants of frontotemporal dementia) from early Alzheimer's disease: a comparative neuropsychological study. *Neuropsychology,* 13**,** 31-40.

HODGKINSON, C. A., GOLDMAN, D., JAEGER, J., PERSAUD, S., KANE, J. M., LIPSKY, R. H. & MALHOTRA, A. K. 2004. Disrupted in schizophrenia 1 (DISC1): association with schizophrenia, schizoaffective disorder, and bipolar disorder. *Am J Hum Genet,* 75**,** 862-72.

HOFFMAN, O., BURNS, N., VADASZ, I., ELTZSCHIG, H. K., EDWARDS, M. G. & VOHWINKEL, C. U. 2017. Detrimental ELAVL-1/HuR-dependent GSK3beta mRNA stabilization impairs resolution in acute respiratory distress syndrome. *PLoS One,* 12**,** e0172116.

HONIG, B. & NICHOLLS, A. 1995. Classical electrostatics in biology and chemistry. *Science,* 268**,** 1144-9.

HORE, T. A., VON MEYENN, F., RAVICHANDRAN, M., BACHMAN, M., FICZ, G., OXLEY, D., SANTOS, F., BALASUBRAMANIAN, S., JURKOWSKI, T. P. & REIK, W. 2016. Retinol and ascorbate drive erasure of epigenetic memory and enhance reprogramming to naive pluripotency by complementary mechanisms. *Proc Natl Acad Sci U S A,* 113**,** 12202-12207.

HORSBURGH, K., MCCARRON, M. O., WHITE, F. & NICOLL, J. A. 2000. The role of apolipoprotein E in Alzheimer's disease, acute brain injury and cerebrovascular disease: evidence of common mechanisms and utility of animal models. *Neurobiol Aging,* 21**,** 245-55.

HOWELL, C. Y., BESTOR, T. H., DING, F., LATHAM, K. E., MERTINEIT, C., TRASLER, J. M. & CHAILLET, J. R. 2001. Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell,* 104**,** 829-38.

HU, Y., SUN, Z., DENG, J., HU, B., YAN, W., WEI, H. & JIANG, J. 2017. Splicing factor hnRNPA2B1 contributes to tumorigenic potential of breast cancer cells through STAT3 and ERK1/2 signaling pathway. *Tumour Biol,* 39**,** 1010428317694318.

HUSSAIN, S. & BASHIR, Z. I. 2015. The epitranscriptome in modulating spatiotemporal RNA translation in neuronal post-synaptic function. *Front Cell Neurosci,* 9**,** 420.

HUSSAIN, S., SAJINI, A. A., BLANCO, S., DIETMANN, S., LOMBARD, P., SUGIMOTO, Y., PARAMOR, M., GLEESON, J. G., ODOM, D. T., ULE, J. & FRYE, M. 2013. NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell Rep,* 4**,** 255-61.

HUTTENLOCHER, P. R. & DABHOLKAR, A. S. 1997. Regional differences in synaptogenesis in human cerebral cortex. *J Comp Neurol,* 387**,** 167-78.
HVAS, A. M., JUUL, S., LAURITZEN, L., NEXO, E. & ELLEGAARD, J. 2004. No effect of vitamin B-12 treatment on cognitive function and depression: a randomized placebo controlled study. *J Affect Disord,* 81**,** 269-73.
HWANG, S., KIM, E., LEE, I. & MARCOTTE, E. M. 2015. Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Sci Rep,* 5**,** 17875.
IBM CORP. Armonk, NY.
ILLINGWORTH, R. S. & BIRD, A. P. 2009. CpG islands--'a rough guide'. *FEBS Lett,* 583**,** 1713-20.
INTERNATIONAL SCHIZOPHRENIA, C., PURCELL, S. M., WRAY, N. R., STONE, J. L., VISSCHER, P. M., O'DONOVAN, M. C., SULLIVAN, P. F. & SKLAR, P. 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature,* 460**,** 748-52.
IRAOLA-GUZMAN, S., ESTIVILL, X. & RABIONET, R. 2011. DNA methylation in neurodegenerative disorders: a missing link between genome and environment? *Clin Genet,* 80**,** 1-14.
IRIER, H., STREET, R. C., DAVE, R., LIN, L., CAI, C., DAVIS, T. H., YAO, B., CHENG, Y. & JIN, P. 2014. Environmental enrichment modulates 5-hydroxymethylcytosine dynamics in hippocampus. *Genomics,* 104**,** 376-82.
IRIZARRY, M. C., GUROL, M. E., RAJU, S., DIAZ-ARRASTIA, R., LOCASCIO, J. J., TENNIS, M., HYMAN, B. T., GROWDON, J. H., GREENBERG, S. M. & BOTTIGLIERI, T. 2005. Association of homocysteine with plasma amyloid beta protein in aging and neurodegenerative disease. *Neurology,* 65**,** 1402-8.
IWATA, A., NAGATA, K., HATSUTA, H., TAKUMA, H., BUNDO, M., IWAMOTO, K., TAMAOKA, A., MURAYAMA, S., SAIDO, T. & TSUJI, S. 2014. Altered CpG methylation in sporadic Alzheimer's disease is associated with APP and MAPT dysregulation. *Hum Mol Genet,* 23**,** 648-56.
JAKOVCEVSKI, M. & AKBARIAN, S. 2012. Epigenetic mechanisms in neurological disease. *Nat Med,* 18**,** 1194-204.
JAMES, S. J., MELNYK, S., JERNIGAN, S., HUBANKS, A., ROSE, S. & GAYLOR, D. W. 2008. Abnormal transmethylation/transsulfuration metabolism and DNA hypomethylation among parents of children with autism. *J Autism Dev Disord,* 38**,** 1966-75.
JARA-PRADO, A., ORTEGA-VAZQUEZ, A., MARTINEZ-RUANO, L., RIOS, C. & SANTAMARIA, A. 2003. Homocysteine-induced brain lipid peroxidation: effects of NMDA receptor blockade, antioxidant treatment, and nitric oxide synthase inhibition. *Neurotox Res,* 5**,** 237-43.
JEONG, M., SUN, D., LUO, M., HUANG, Y., CHALLEN, G. A., RODRIGUEZ, B., ZHANG, X., CHAVEZ, L., WANG, H., HANNAH, R., KIM, S. B., YANG, L., KO, M., CHEN, R., GOTTGENS, B., LEE, J. S., GUNARATNE, P., GODLEY, L. A., DARLINGTON, G. J., RAO, A., LI, W. & GOODELL, M. A. 2014. Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat Genet,* 46**,** 17-23.
JIA, D., JURKOWSKA, R. Z., ZHANG, X., JELTSCH, A. & CHENG, X. 2007. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature,* 449**,** 248-51.

JIA, G., FU, Y., ZHAO, X., DAI, Q., ZHENG, G., YANG, Y., YI, C., LINDAHL, T., PAN, T., YANG, Y. G. & HE, C. 2011. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol,* 7**,** 885-7.

JIN, S. G., KADAM, S. & PFEIFER, G. P. 2010. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res,* 38**,** e125.

JOHNSON, N., BELL, P., JONOVSKA, V., BUDGE, M. & SIM, E. 2004. NAT gene polymorphisms and susceptibility to Alzheimer's disease: identification of a novel NAT1 allelic variant. *BMC Med Genet,* 5**,** 6.

JOHNSON, S. C., SCHMITZ, T. W., ASTHANA, S., GLUCK, M. A. & MYERS, C. 2008. Associative learning over trials activates the hippocampus in healthy elderly but not mild cognitive impairment. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn,* 15**,** 129-45.

JORM, A. F. 2001. History of depression as a risk factor for dementia: an updated review. *Aust N Z J Psychiatry,* 35**,** 776-81.

JOUBERT, S., BRAMBATI, S. M., ANSADO, J., BARBEAU, E. J., FELICIAN, O., DIDIC, M., LACOMBE, J., GOLDSTEIN, R., CHAYER, C. & KERGOAT, M. J. 2010. The cognitive and neural expression of semantic memory impairment in mild cognitive impairment and early Alzheimer's disease. *Neuropsychologia,* 48**,** 978-88.

KALLBERG, M., WANG, H., WANG, S., PENG, J., WANG, Z., LU, H. & XU, J. 2012. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc,* 7**,** 1511-22.

KALMIJN, S., VAN BOXTEL, M. P., OCKE, M., VERSCHUREN, W. M., KROMHOUT, D. & LAUNER, L. J. 2004. Dietary intake of fatty acids and fish in relation to cognitive performance at middle age. *Neurology,* 62**,** 275-80.

KALSCHEUER, V. M., FREUDE, K., MUSANTE, L., JENSEN, L. R., YNTEMA, H. G., GECZ, J., SEFIANI, A., HOFFMANN, K., MOSER, B., HAAS, S., GUROK, U., HAESLER, S., ARANDA, B., NSHEDJAN, A., TZSCHACH, A., HARTMANN, N., ROLOFF, T. C., SHOICHET, S., HAGENS, O., TAO, J., VAN BOKHOVEN, H., TURNER, G., CHELLY, J., MORAINE, C., FRYNS, J. P., NUBER, U., HOELTZENBEIN, M., SCHARFF, C., SCHERTHAN, H., LENZNER, S., HAMEL, B. C., SCHWEIGER, S. & ROPERS, H. H. 2003. Mutations in the polyglutamine binding protein 1 gene cause X-linked mental retardation. *Nat Genet,* 35**,** 313-5.

KANE, M. J., HAMBRICK, D. Z., TUHOLSKI, S. W., WILHELM, O., PAYNE, T. W. & ENGLE, R. W. 2004. The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *J Exp Psychol Gen,* 133**,** 189-217.

KARACA, E., HAREL, T., PEHLIVAN, D., JHANGIANI, S. N., GAMBIN, T., COBAN AKDEMIR, Z., GONZAGA-JAUREGUI, C., ERDIN, S., BAYRAM, Y., CAMPBELL, I. M., HUNTER, J. V., ATIK, M. M., VAN ESCH, H., YUAN, B., WISZNIEWSKI, W., ISIKAY, S., YESIL, G., YUREGIR, O. O., TUG BOZDOGAN, S., ASLAN, H., AYDIN, H., TOS, T., AKSOY, A., DE VIVO, D. C., JAIN, P., GECKINLI, B. B., SEZER, O., GUL, D., DURMAZ, B., COGULU, O., OZKINAY, F., TOPCU, V., CANDAN, S., CEBI, A. H., IKBAL, M., YILMAZ GULEC, E., GEZDIRICI, A., KOPARIR, E., EKICI, F., COSKUN, S., CICEK, S., KARAER, K., KOPARIR, A., DUZ, M. B., KIRAT, E., FENERCIOGLU, E., ULUCAN, H., SEVEN, M., GURAN, T., ELCIOGLU, N., YILDIRIM, M. S., AKTAS, D., ALIKASIFOGLU, M., TURE, M.,

YAKUT, T., OVERTON, J. D., YUKSEL, A., OZEN, M., MUZNY, D. M., ADAMS, D. R., BOERWINKLE, E., CHUNG, W. K., GIBBS, R. A. & LUPSKI, J. R. 2015. Genes that Affect Brain Structure and Function Identified by Rare Variant Analyses of Mendelian Neurologic Disease. *Neuron,* 88**,** 499-513.

KATSANOU, V., MILATOS, S., YIAKOUVAKI, A., SGANTZIS, N., KOTSONI, A., ALEXIOU, M., HAROKOPOS, V., AIDINIS, V., HEMBERGER, M. & KONTOYIANNIS, D. L. 2009. The RNA-binding protein Elavl1/HuR is essential for placental branching morphogenesis and embryonic development. *Mol Cell Biol,* 29**,** 2762-76.

KAVRAAL, S., ONCU, S. K., BITIKTAS, S., ARTIS, A. S., DOLU, N., GUNES, T. & SUER, C. 2012. Maternal intake of Omega-3 essential fatty acids improves long term potentiation in the dentate gyrus and Morris water maze performance in rats. *Brain Res,* 1482**,** 32-9.

KENNEDY, B. P., BOTTIGLIERI, T., ARNING, E., ZIEGLER, M. G., HANSEN, L. A. & MASLIAH, E. 2004. Elevated S-adenosylhomocysteine in Alzheimer brain: influence on methyltransferases and cognitive function. *J Neural Transm,* 111**,** 547-67.

KENNEDY, D. O., VEASEY, R., WATSON, A., DODD, F., JONES, E., MAGGINI, S. & HASKELL, C. F. 2010. Effects of high-dose B vitamin complex with vitamin C and minerals on subjective mood and performance in healthy males. *Psychopharmacology (Berl),* 211**,** 55-68.

KENNELLY, S. P., LAWLOR, B. A. & KENNY, R. A. 2009. Blood pressure and dementia - a comprehensive review. *Ther Adv Neurol Disord,* 2**,** 241-60.

KENNY, E. M., CORMICAN, P., FURLONG, S., HERON, E., KENNY, G., FAHEY, C., KELLEHER, E., ENNIS, S., TROPEA, D., ANNEY, R., CORVIN, A. P., DONOHOE, G., GALLAGHER, L., GILL, M. & MORRIS, D. W. 2014. Excess of rare novel loss-of-function variants in synaptic genes in schizophrenia and autism spectrum disorders. *Mol Psychiatry,* 19**,** 872-9.

KERKEL, K., SCHUPF, N., HATTA, K., PANG, D., SALAS, M., KRATZ, A., MINDEN, M., MURTY, V., ZIGMAN, W. B., MAYEUX, R. P., JENKINS, E. C., TORKAMANI, A., SCHORK, N. J., SILVERMAN, W., CROY, B. A. & TYCKO, B. 2010. Altered DNA methylation in leukocytes with trisomy 21. *PLoS Genet,* 6**,** e1001212.

KEVERNE, E. B., FUNDELE, R., NARASIMHA, M., BARTON, S. C. & SURANI, M. A. 1996. Genomic imprinting and the differential roles of parental genomes in brain development. *Brain Res Dev Brain Res,* 92**,** 91-100.

KHAN, M. A., RAFIQ, M. A., NOOR, A., HUSSAIN, S., FLORES, J. V., RUPP, V., VINCENT, A. K., MALLI, R., ALI, G., KHAN, F. S., ISHAK, G. E., DOHERTY, D., WEKSBERG, R., AYUB, M., WINDPASSINGER, C., IBRAHIM, S., FRYE, M., ANSAR, M. & VINCENT, J. B. 2012. Mutation in NSUN2, which encodes an RNA methyltransferase, causes autosomal-recessive intellectual disability. *Am J Hum Genet,* 90**,** 856-63.

KIM, H. J., KIM, N. C., WANG, Y. D., SCARBOROUGH, E. A., MOORE, J., DIAZ, Z., MACLEA, K. S., FREIBAUM, B., LI, S., MOLLIEX, A., KANAGARAJ, A. P., CARTER, R., BOYLAN, K. B., WOJTAS, A. M., RADEMAKERS, R., PINKUS, J. L., GREENBERG, S. A., TROJANOWSKI, J. Q., TRAYNOR, B. J., SMITH, B. N., TOPP, S., GKAZI, A. S., MILLER, J., SHAW, C. E., KOTTLORS, M., KIRSCHNER, J., PESTRONK, A., LI, Y. R., FORD, A. F., GITLER, A. D., BENATAR, M., KING, O. D., KIMONIS, V. E., ROSS, E. D., WEIHL, C. C., SHORTER, J. & TAYLOR, J. P.

2013. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature,* 495**,** 467-73.
KIM, J., PARK, M. H., KIM, E., HAN, C., JO, S. A. & JO, I. 2007a. Plasma homocysteine is associated with the risk of mild cognitive impairment in an elderly Korean population. *J Nutr,* 137**,** 2093-7.
KIM, J. H., CHO, S. Y., LEE, J. H., JEONG, S. M., YOON, I. S., LEE, B. H., LEE, J. H., PYO, M. K., LEE, S. M., CHUNG, J. M., KIM, S., RHIM, H., OH, J. W. & NAH, S. Y. 2007b. Neuroprotective effects of ginsenoside Rg3 against homocysteine-induced excitotoxicity in rat hippocampus. *Brain Res,* 1136**,** 190-9.
KLEIN, C. J., BOTUYAN, M. V., WU, Y. H., WARD, C. J., NICHOLSON, G. A., HAMMANS, S., HOJO, K., YAMANISHI, H., KARPF, A. R., WALLACE, D. C., SIMON, M., LANDER, C., BOARDMAN, L. A., CUNNINGHAM, J. M., SMITH, G. E., LITCHY, W. J., BOES, B., ATKINSON, E. J., MIDDHA, S., DYCK, P. J. B., PARISI, J. E., MER, G., SMITH, D. I. & DYCK, P. J. 2011. Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. *Nature Genetics,* 43**,** 595-U140.
KNIGHT, H. M., BROWN, M., DODGEON, B., MAUGHAN, B., RICHARDS, M., ELLIOT, J., SAHAKIAN, B. J. & ROBBINS, T. W. 2010. Investigating individual differences in memory and cognition in the National Child Development Study cohort members using a life course approach. . London: Centre for Longitudinal Studies.
KNIGHT, H. M., PICKARD, B. S., MACLEAN, A., MALLOY, M. P., SOARES, D. C., MCRAE, A. F., CONDIE, A., WHITE, A., HAWKINS, W., MCGHEE, K., VAN BECK, M., MACINTYRE, D. J., STARR, J. M., DEARY, I. J., VISSCHER, P. M., PORTEOUS, D. J., CANNON, R. E., ST CLAIR, D., MUIR, W. J. & BLACKWOOD, D. H. 2009. A cytogenetic abnormality and rare coding variants identify ABCA13 as a candidate gene in schizophrenia, bipolar disorder, and depression. *Am J Hum Genet,* 85**,** 833-46.
KOBAYASHI, N., SHINAGAWA, S., NAGATA, T., SHIMADA, K., SHIBATA, N., OHNUMA, T., KASANUKI, K., ARAI, H., YAMADA, H., NAKAYAMA, K. & KONDO, K. 2016. Development of Biomarkers Based on DNA Methylation in the NCAPH2/LMF2 Promoter Region for Diagnosis of Alzheimer's Disease and Amnesic Mild Cognitive Impairment. *Plos One,* 11.
KOBE, T., WITTE, A. V., SCHNELLE, A., GRITTNER, U., TESKY, V. A., PANTEL, J., SCHUCHARDT, J. P., HAHN, A., BOHLKEN, J., RUJESCU, D. & FLOEL, A. 2016. Vitamin B-12 concentration, memory performance, and hippocampal structure in patients with mild cognitive impairment. *Am J Clin Nutr,* 103**,** 1045-54.
KOENEN, K. C., MOFFITT, T. E., ROBERTS, A. L., MARTIN, L. T., KUBZANSKY, L., HARRINGTON, H., POULTON, R. & CASPI, A. 2009. Childhood IQ and adult mental disorders: a test of the cognitive reserve hypothesis. *Am J Psychiatry,* 166**,** 50-7.
KOHLI, R. M. & ZHANG, Y. 2013. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature,* 502**,** 472-9.
KOSSLYN, S. M. & INTRILIGATOR, J. M. 1992. Is cognitive neuropsychology plausible? The perils of sitting on a one-legged stool. *J Cogn Neurosci,* 4**,** 96-105.

KRAUS, T. F., GUIBOURT, V. & KRETZSCHMAR, H. A. 2015. 5-Hydroxymethylcytosine, the "Sixth Base", during brain development and ageing. *J Neural Transm (Vienna),* 122**,** 1035-43.
KREINER-MOLLER, E., MEDINA-GOMEZ, C., UITTERLINDEN, A. G., RIVADENEIRA, F. & ESTRADA, K. 2015. Improving accuracy of rare variant imputation with a two-step imputation approach. *Eur J Hum Genet,* 23**,** 395-400.
KRUMAN, II, CULMSEE, C., CHAN, S. L., KRUMAN, Y., GUO, Z., PENIX, L. & MATTSON, M. P. 2000. Homocysteine elicits a DNA damage response in neurons that promotes apoptosis and hypersensitivity to excitotoxicity. *J Neurosci,* 20**,** 6920-6.
KUECHLER, A., ZINK, A. M., WIELAND, T., LUDECKE, H. J., CREMER, K., SALVIATI, L., MAGINI, P., NAJAFI, K., ZWEIER, C., CZESCHIK, J. C., ARETZ, S., ENDELE, S., TAMBURRINO, F., PINATO, C., CLEMENTI, M., GUNDLACH, J., MAYLAHN, C., MAZZANTI, L., WOHLLEBER, E., SCHWARZMAYR, T., KARIMINEJAD, R., SCHLESSINGER, A., WIECZOREK, D., STROM, T. M., NOVARINO, G. & ENGELS, H. 2015. Loss-of-function variants of SETD5 cause intellectual disability and the core phenotype of microdeletion 3p25.3 syndrome. *Eur J Hum Genet,* 23**,** 753-60.
KWAN, J. S., LI, M. X., DENG, J. E. & SHAM, P. C. 2016. FAPI: Fast and accurate P-value Imputation for genome-wide association study. *Eur J Hum Genet,* 24**,** 761-6.
KWOK, C. T., MARSHALL, A. D., RASKO, J. E. & WONG, J. J. 2017. Genetic alterations of m6A regulators predict poorer survival in acute myeloid leukemia. *J Hematol Oncol,* 10**,** 39.
LADD-ACOSTA, C., HANSEN, K. D., BRIEM, E., FALLIN, M. D., KAUFMANN, W. E. & FEINBERG, A. P. 2014. Common DNA methylation alterations in multiple brain regions in autism. *Mol Psychiatry,* 19**,** 862-71.
LADUCA, H., FARWELL, K. D., VUONG, H., LU, H. M., MU, W., SHAHMIRZADI, L., TANG, S., CHEN, J., BHIDE, S. & CHAO, E. C. 2017. Exome sequencing covers >98% of mutations identified on targeted next generation sequencing panels. *PLoS One,* 12**,** e0170843.
LASHLEY, T., GAMI, P., VALIZADEH, N., LI, A., REVESZ, T. & BALAZS, R. 2015. Alterations in global DNA methylation and hydroxymethylation are not detected in Alzheimer's disease. *Neuropathol Appl Neurobiol,* 41**,** 497-506.
LE BER, I., VAN BORTEL, I., NICOLAS, G., BOUYA-AHMED, K., CAMUZAT, A., WALLON, D., DE SEPTENVILLE, A., LATOUCHE, M., LATTANTE, S., KABASHI, E., JORNEA, L., HANNEQUIN, D., BRICE, A. & FRENCH RESEARCH NETWORK ON, F. F.-A. 2014. hnRNPA2B1 and hnRNPA1 mutations are rare in patients with "multisystem proteinopathy" and frontotemporal lobar degeneration phenotypes. *Neurobiol Aging,* 35**,** 934 e5-6.
LEE, S., ABECASIS, G. R., BOEHNKE, M. & LIN, X. 2014. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet,* 95**,** 5-23.
LEE, S., EMOND, M. J., BAMSHAD, M. J., BARNES, K. C., RIEDER, M. J., NICKERSON, D. A., TEAM, N. G. E. S. P.-E. L. P., CHRISTIANI, D. C., WURFEL, M. M. & LIN, X. 2012. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet,* 91**,** 224-37.

LEE, S. H., WRAY, N. R., GODDARD, M. E. & VISSCHER, P. M. 2011. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet,* 88**,** 294-305.
LEI, Y., ZHANG, X., SU, J., JEONG, M., GUNDRY, M. C., HUANG, Y. H., ZHOU, Y., LI, W. & GOODELL, M. A. 2017. Targeted DNA methylation in vivo using an engineered dCas9-MQ1 fusion protein. *Nat Commun,* 8**,** 16026.
LEJEUNE, J., GAUTHIER, M. & TURPIN, R. 1959. Les Chromosomes Humains En Culture De Tissus. *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences,* 248**,** 602-603.
LENCE, T., AKHTAR, J., BAYER, M., SCHMID, K., SPINDLER, L., HO, C. H., KREIM, N., ANDRADE-NAVARRO, M. A., POECK, B., HELM, M. & ROIGNANT, J. Y. 2016. m6A modulates neuronal functions and sex determination in Drosophila. *Nature,* 540**,** 242-247.
LESCAI, F., ALS, T. D., LI, Q., NYEGAARD, M., ANDORSDOTTIR, G., BISKOPSTO, M., HEDEMAND, A., FIORENTINO, A., O'BRIEN, N., JARRAM, A., LIANG, J., GROVE, J., PALLESEN, J., EICKHARDT, E., MATTHEISEN, M., BOLUND, L., DEMONTIS, D., WANG, A. G., MCQUILLIN, A., MORS, O., WANG, J. & BORGLUM, A. D. 2017. Whole-exome sequencing of individuals from an isolated population implicates rare risk variants in bipolar disorder. *Transl Psychiatry,* 7**,** e1034.
LEVINE, S. & SALTZMAN, A. 2004. Pyridoxine (vitamin B6) neurotoxicity: enhancement by protein-deficient diet. *J Appl Toxicol,* 24**,** 497-500.
LEYFER, O. T., FOLSTEIN, S. E., BACALMAN, S., DAVIS, N. O., DINH, E., MORGAN, J., TAGER-FLUSBERG, H. & LAINHART, J. E. 2006. Comorbid psychiatric disorders in children with autism: interview development and rates of disorders. *J Autism Dev Disord,* 36**,** 849-61.
LGC GENOMICS. Hoddesdon, UK.
LI, F., ZHAO, D., WU, J. & SHI, Y. 2014. Structure of the YTH domain of human YTHDF2 in complex with an m(6)A mononucleotide reveals an aromatic cage for m(6)A recognition. *Cell Res,* 24**,** 1490-2.
LI, S., PAPALE, L. A., KINTNER, D. B., SABAT, G., BARRETT-WILT, G. A., CENGIZ, P. & ALISCH, R. S. 2015. Hippocampal increase of 5-hmC in the glucocorticoid receptor gene following acute stress. *Behav Brain Res,* 286**,** 236-240.
LI, Y. & SASAKI, H. 2011. Genomic imprinting in mammals: its life cycle, molecular mechanisms and reprogramming. *Cell Res,* 21**,** 466-73.
LIN, N., QIN, S., LUO, S., CUI, S., HUANG, G. & ZHANG, X. 2014. Homocysteine induces cytotoxicity and proliferation inhibition in neural stem cells via DNA methylation in vitro. *FEBS J,* 281**,** 2088-96.
LIN, W. Y. 2016. Beyond Rare-Variant Association Testing: Pinpointing Rare Causal Variants in Case-Control Sequencing Study. *Sci Rep,* 6**,** 21824.
LIPTON, S. A., KIM, W. K., CHOI, Y. B., KUMAR, S., D'EMILIA, D. M., RAYUDU, P. V., ARNELLE, D. R. & STAMLER, J. S. 1997. Neurotoxicity associated with dual actions of homocysteine at the N-methyl-D-aspartate receptor. *Proc Natl Acad Sci U S A,* 94**,** 5923-8.
LISTER, R., MUKAMEL, E. A., NERY, J. R., URICH, M., PUDDIFOOT, C. A., JOHNSON, N. D., LUCERO, J., HUANG, Y., DWORK, A. J., SCHULTZ, M. D., YU, M., TONTI-FILIPPINI, J., HEYN, H., HU, S., WU, J. C., RAO, A., ESTELLER, M., HE, C., HAGHIGHI, F. G., SEJNOWSKI, T. J., BEHRENS, M. M. & ECKER, J. R. 2013.

Global epigenomic reconfiguration during mammalian brain development. *Science,* 341**,** 1237905.
LITHNER, C. U., LACOR, P. N., ZHAO, W. Q., MUSTAFIZ, T., KLEIN, W. L., SWEATT, J. D. & HERNANDEZ, C. M. 2013. Disruption of neocortical histone H3 homeostasis by soluble Abeta: implications for Alzheimer's disease. *Neurobiol Aging,* 34**,** 2081-90.
LITTLEJOHNS, T. J., HENLEY, W. E., LANG, I. A., ANNWEILER, C., BEAUCHET, O., CHAVES, P. H., FRIED, L., KESTENBAUM, B. R., KULLER, L. H., LANGA, K. M., LOPEZ, O. L., KOS, K., SONI, M. & LLEWELLYN, D. J. 2014. Vitamin D and the risk of dementia and Alzheimer disease. *Neurology,* 83**,** 920-8.
LIU, H., HARRELL, L. E., SHENVI, S., HAGEN, T. & LIU, R. M. 2005. Gender differences in glutathione metabolism in Alzheimer's disease. *J Neurosci Res,* 79**,** 861-7.
LIU, J., YUE, Y., HAN, D., WANG, X., FU, Y., ZHANG, L., JIA, G., YU, M., LU, Z., DENG, X., DAI, Q., CHEN, W. & HE, C. 2014. A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat Chem Biol,* 10**,** 93-5.
LIU, L., VAN GROEN, T., KADISH, I., LI, Y., WANG, D., JAMES, S. R., KARPF, A. R. & TOLLEFSBOL, T. O. 2011. Insufficient DNA methylation affects healthy aging and promotes age-related health problems. *Clin Epigenetics,* 2**,** 349-60.
LIU, L., VAN GROEN, T., KADISH, I. & TOLLEFSBOL, T. O. 2009. DNA methylation impacts on learning and memory in aging. *Neurobiol Aging,* 30**,** 549-60.
LIU, Y., OLANREWAJU, Y. O., ZHANG, X. & CHENG, X. 2013. DNA recognition of 5-carboxylcytosine by a Zfp57 mutant at an atomic resolution of 0.97 A. *Biochemistry,* 52**,** 9310-7.
LIU, Y., ZHANG, B., MENG, X., KORN, M. J., PARENT, J. M., LU, L. Y. & YU, X. 2017. UHRF2 regulates local 5-methylcytosine and suppresses spontaneous seizures. *Epigenetics***,** 1-10.
LLORET, A., BADIA, M. C., MORA, N. J., PALLARDO, F. V., ALONSO, M. D. & VINA, J. 2009. Vitamin E paradox in Alzheimer's disease: it does not prevent loss of cognition and may even be detrimental. *J Alzheimers Dis,* 17**,** 143-9.
LONGONI, G., ROCCA, M. A., PAGANI, E., RICCITELLI, G. C., COLOMBO, B., RODEGHER, M., FALINI, A., COMI, G. & FILIPPI, M. 2015. Deficits in memory and visuospatial learning correlate with regional hippocampal atrophy in MS. *Brain Structure & Function,* 220**,** 435-444.
LU, H. & HALLSTROM, T. C. 2013. The nuclear protein UHRF2 is a direct target of the transcription factor E2F1 in the induction of apoptosis. *J Biol Chem,* 288**,** 23833-43.
LU, Y., CHRISTIAN, K. & LU, B. 2008. BDNF: a key regulator for protein synthesis-dependent LTP and long-term memory? *Neurobiol Learn Mem,* 89**,** 312-23.
LU, Y. C., CHANG, S. H., HAFNER, M., LI, X., TUSCHL, T., ELEMENTO, O. & HLA, T. 2014. ELAVL1 modulates transcriptome-wide miRNA binding in murine macrophages. *Cell Rep,* 9**,** 2330-43.
LULEYAP, H. U., ALPTEKIN, D., PAZARBASI, A., KASAP, M., KASAP, H., DEMIRHINDI, H., MUNGAN, N., OZER, G. & FROSTER, U. G. 2006. The importance of arginine mutation for the evolutionary structure and function of phenylalanine hydroxylase gene. *Mutat Res,* 601**,** 39-45.

LUNNON, K., SMITH, R., HANNON, E., DE JAGER, P. L., SRIVASTAVA, G., VOLTA, M., TROAKES, C., AL-SARRAJ, S., BURRAGE, J., MACDONALD, R., CONDLIFFE, D., HARRIES, L. W., KATSEL, P., HAROUTUNIAN, V., KAMINSKY, Z., JOACHIM, C., POWELL, J., LOVESTONE, S., BENNETT, D. A., SCHALKWYK, L. C. & MILL, J. 2014. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat Neurosci,* 17**,** 1164-70.

LUO, T., CUI, S., BIAN, C. & YU, X. 2013. Uhrf2 is important for DNA damage response in vascular smooth muscle cells. *Biochem Biophys Res Commun,* 441**,** 65-70.

LYALL, D. M., CULLEN, B., ALLERHAND, M., SMITH, D. J., MACKAY, D., EVANS, J., ANDERSON, J., FAWNS-RITCHIE, C., MCINTOSH, A. M., DEARY, I. J. & PELL, J. P. 2016. Cognitive Test Scores in UK Biobank: Data Reduction in 480,416 Participants and Longitudinal Stability in 20,346 Participants. *PLoS One,* 11**,** e0154222.

MA, D. K., GUO, J. U., MING, G. L. & SONG, H. 2009. DNA excision repair proteins and Gadd45 as molecular players for active DNA demethylation. *Cell Cycle,* 8**,** 1526-31.

MA, Y., SMITH, C. E., LAI, C. Q., IRVIN, M. R., PARNELL, L. D., LEE, Y. C., PHAM, L., ASLIBEKYAN, S., CLAAS, S. A., TSAI, M. Y., BORECKI, I. B., KABAGAMBE, E. K., BERCIANO, S., ORDOVAS, J. M., ABSHER, D. M. & ARNETT, D. K. 2015. Genetic variants modify the effect of age on APOE methylation in the Genetics of Lipid Lowering Drugs and Diet Network study. *Aging Cell,* 14**,** 49-59.

MACKIN, R. S., INSEL, P., TOSUN, D., MUELLER, S. G., SCHUFF, N., TRURAN-SACREY, D., RAPTENTSETSANG, S. T., LEE, J. Y., JACK, C. R., JR., AISEN, P. S., PETERSEN, R. C., WEINER, M. W. & ALZHEIMER'S DISEASE NEUROIMAGING, I. 2013. The effect of subsyndromal symptoms of depression and white matter lesions on disability for individuals with mild cognitive impairment. *Am J Geriatr Psychiatry,* 21**,** 906-14.

MALOUF, M., GRIMLEY, E. J. & AREOSA, S. A. 2003. Folic acid with or without vitamin B12 for cognition and dementia. *Cochrane Database Syst Rev***,** CD004514.

MALOUF, R. & AREOSA SASTRE, A. 2003. Vitamin B12 for cognition. *Cochrane Database Syst Rev***,** CD004326.

MANOLIO, T. A., COLLINS, F. S., COX, N. J., GOLDSTEIN, D. B., HINDORFF, L. A., HUNTER, D. J., MCCARTHY, M. I., RAMOS, E. M., CARDON, L. R., CHAKRAVARTI, A., CHO, J. H., GUTTMACHER, A. E., KONG, A., KRUGLYAK, L., MARDIS, E., ROTIMI, C. N., SLATKIN, M., VALLE, D., WHITTEMORE, A. S., BOEHNKE, M., CLARK, A. G., EICHLER, E. E., GIBSON, G., HAINES, J. L., MACKAY, T. F., MCCARROLL, S. A. & VISSCHER, P. M. 2009. Finding the missing heritability of complex diseases. *Nature,* 461**,** 747-53.

MARSHALL, C. R. & SCHERER, S. W. 2012. Detection and characterization of copy number variation in autism spectrum disorder. *Methods Mol Biol,* 838**,** 115-35.

MARTIN, K. C., CASADIO, A., ZHU, H., YAPING, E., ROSE, J. C., CHEN, M., BAILEY, C. H. & KANDEL, E. R. 1997. Synapse-specific, long-term facilitation of aplysia sensory to motor synapses: a function for local protein synthesis in memory storage. *Cell,* 91**,** 927-38.

MARTINEZ, F. J., PRATT, G. A., VAN NOSTRAND, E. L., BATRA, R., HUELGA, S. C., KAPELI, K., FREESE, P., CHUN, S. J., LING, K., GELBOIN-BURKHART, C., FIJANY, L., WANG, H. C., NUSSBACHER, J. K., BROSKI, S. M., KIM, H. J., LARDELLI, R., SUNDARARAMAN, B., DONOHUE, J. P., JAVAHERIAN, A., LYKKE-ANDERSEN, J., FINKBEINER, S., BENNETT, C. F., ARES, M., JR., BURGE, C. B., TAYLOR, J. P., RIGO, F. & YEO, G. W. 2016. Protein-RNA Networks Regulated by Normal and ALS-Associated Mutant HNRNPA2B1 in the Nervous System. *Neuron,* 92**,** 780-795.

MARTINOWICH, K., HATTORI, D., WU, H., FOUSE, S., HE, F., HU, Y., FAN, G. & SUN, Y. E. 2003. DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science,* 302**,** 890-3.

MASTROENI, D., GROVER, A., DELVAUX, E., WHITESIDE, C., COLEMAN, P. D. & ROGERS, J. 2010. Epigenetic changes in Alzheimer's disease: decrements in DNA methylation. *Neurobiol Aging,* 31**,** 2025-37.

MATSON, J. L. & SHOEMAKER, M. 2009. Intellectual disability and its relationship to autism spectrum disorders. *Res Dev Disabil,* 30**,** 1107-14.

MATTSON, M. P., MAUDSLEY, S. & MARTIN, B. 2004. BDNF and 5-HT: a dynamic duo in age-related neuronal plasticity and neurodegenerative disorders. *Trends Neurosci,* 27**,** 589-94.

MAUER, J., LUO, X., BLANJOIE, A., JIAO, X., GROZHIK, A. V., PATIL, D. P., LINDER, B., PICKERING, B. F., VASSEUR, J. J., CHEN, Q., GROSS, S. S., ELEMENTO, O., DEBART, F., KILEDJIAN, M. & JAFFREY, S. R. 2017. Reversible methylation of m6Am in the 5' cap controls mRNA stability. *Nature,* 541**,** 371-375.

MCCARTHY, D. J., HUMBURG, P., KANAPIN, A., RIVAS, M. A., GAULTON, K., CAZIER, J. B. & DONNELLY, P. 2014. Choice of transcripts and software has a large effect on variant annotation. *Genome Med,* 6**,** 26.

MCCARTHY, M. M., AUGER, A. P., BALE, T. L., DE VRIES, G. J., DUNN, G. A., FORGER, N. G., MURRAY, E. K., NUGENT, B. M., SCHWARZ, J. M. & WILSON, M. E. 2009. The epigenetics of sex differences in the brain. *J Neurosci,* 29**,** 12815-23.

MCCONKIE-ROSELL, A., LACHIEWICZ, A. M., SPIRIDIGLIOZZI, G. A., TARLETON, J., SCHOENWALD, S., PHELAN, M. C., GOONEWARDENA, P., DING, X. & BROWN, W. T. 1993. Evidence that methylation of the FMR-I locus is responsible for variable phenotypic expression of the fragile X syndrome. *Am J Hum Genet,* 53**,** 800-9.

MCDONALD, C. R., GHARAPETIAN, L., MCEVOY, L. K., FENNEMA-NOTESTINE, C., HAGLER, D. J., JR., HOLLAND, D., DALE, A. M. & ALZHEIMER'S DISEASE NEUROIMAGING, I. 2012. Relationship between regional atrophy rates and cognitive decline in mild cognitive impairment. *Neurobiol Aging,* 33**,** 242-53.

MCGRAW, S., ZHANG, J. X., FARAG, M., CHAN, D., CARON, M., KONERMANN, C., OAKES, C. C., MOHAN, K. N., PLASS, C., PASTINEN, T., BOURQUE, G., CHAILLET, J. R. & TRASLER, J. M. 2015. Transient DNMT1 suppression reveals hidden heritable marks in the genome. *Nucleic Acids Res,* 43**,** 1485-97.

MEDWAY, C. & MORGAN, K. 2014. Review: The genetics of Alzheimer's disease; putting flesh on the bones. *Neuropathol Appl Neurobiol,* 40**,** 97-105.

MELLEN, M., AYATA, P., DEWELL, S., KRIAUCIONIS, S. & HEINTZ, N. 2012. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell,* 151**,** 1417-30.
MENDES, J., GUEROIS, R. & SERRANO, L. 2002. Energy estimation in protein design. *Curr Opin Struct Biol,* 12**,** 441-6.
MEYER, K. D., SALETORE, Y., ZUMBO, P., ELEMENTO, O., MASON, C. E. & JAFFREY, S. R. 2012. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell,* 149**,** 1635-46.
MEYNERT, A. M., ANSARI, M., FITZPATRICK, D. R. & TAYLOR, M. S. 2014. Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC Bioinformatics,* 15**,** 247.
MICHAUD, E. J., VAN VUGT, M. J., BULTMAN, S. J., SWEET, H. O., DAVISSON, M. T. & WOYCHIK, R. P. 1994. Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. *Genes Dev,* 8**,** 1463-72.
MILLER, C. A., GAVIN, C. F., WHITE, J. A., PARRISH, R. R., HONASOGE, A., YANCEY, C. R., RIVERA, I. M., RUBIO, M. D., RUMBAUGH, G. & SWEATT, J. D. 2010. Cortical DNA methylation maintains remote memory. *Nat Neurosci,* 13**,** 664-6.
MILLER, C. A. & SWEATT, J. D. 2007. Covalent modification of DNA regulates memory formation. *Neuron,* 53**,** 857-69.
MINAGAWA, H., WATANABE, A., AKATSU, H., ADACHI, K., OHTSUKA, C., TERAYAMA, Y., HOSONO, T., TAKAHASHI, S., WAKITA, H., JUNG, C. G., KOMANO, H. & MICHIKAWA, M. 2010. Homocysteine, another risk factor for Alzheimer disease, impairs apolipoprotein E3 function. *J Biol Chem,* 285**,** 38382-8.
MITCHELL, E. S., CONUS, N. & KAPUT, J. 2014. B vitamin polymorphisms and behavior: evidence of associations with neurodevelopment, depression, schizophrenia, bipolar disorder and cognitive decline. *Neurosci Biobehav Rev,* 47**,** 307-20.
MITT, M., KALS, M., PARN, K., GABRIEL, S. B., LANDER, E. S., PALOTIE, A., RIPATTI, S., MORRIS, A. P., METSPALU, A., ESKO, T., MAGI, R. & PALTA, P. 2017. Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur J Hum Genet,* 25**,** 869-876.
MODREGO, P. J. & FERRANDEZ, J. 2004. Depression in patients with mild cognitive impairment increases the risk of developing dementia of Alzheimer type: a prospective cohort study. *Arch Neurol,* 61**,** 1290-3.
MOORTHY, D., PETER, I., SCOTT, T. M., PARNELL, L. D., LAI, C. Q., CROTT, J. W., ORDOVAS, J. M., SELHUB, J., GRIFFITH, J., ROSENBERG, I. H., TUCKER, K. L. & TROEN, A. M. 2012. Status of vitamins B-12 and B-6 but not of folate, homocysteine, and the methylenetetrahydrofolate reductase C677T polymorphism are associated with impaired cognition and depression in adults. *J Nutr,* 142**,** 1554-60.
MORETTI, P. & ZOGHBI, H. Y. 2006. MeCP2 dysfunction in Rett syndrome and related disorders. *Curr Opin Genet Dev,* 16**,** 276-81.
MORGAN, V. A., LEONARD, H., BOURKE, J. & JABLENSKY, A. 2008. Intellectual disability co-occurring with schizophrenia and other psychiatric illness: population-based study. *Br J Psychiatry,* 193**,** 364-72.

MORI, T., IKEDA, D. D., YAMAGUCHI, Y., UNOKI, M. & PROJECT, N. 2012. NIRF/UHRF2 occupies a central position in the cell cycle network and allows coupling with the epigenetic landscape. *FEBS Lett,* 586**,** 1570-83.
MORRIS, M. C., EVANS, D. A., BIENIAS, J. L., TANGNEY, C. C., HEBERT, L. E., SCHERR, P. A. & SCHNEIDER, J. A. 2005. Dietary folate and vitamin B12 intake and cognitive decline among community-dwelling older persons. *Arch Neurol,* 62**,** 641-5.
MORRIS, M. S. 2003. Homocysteine and Alzheimer's disease. *Lancet Neurol,* 2**,** 425-8.
MOUSSAUD, S., JONES, D. R., MOUSSAUD-LAMODIERE, E. L., DELENCLOS, M., ROSS, O. A. & MCLEAN, P. J. 2014. Alpha-synuclein and tau: teammates in neurodegeneration? *Mol Neurodegener,* 9**,** 43.
MOUTSIANAS, L., AGARWALA, V., FUCHSBERGER, C., FLANNICK, J., RIVAS, M. A., GAULTON, K. J., ALBERS, P. K., GO, T. D. C., MCVEAN, G., BOEHNKE, M., ALTSHULER, D. & MCCARTHY, M. I. 2015. The power of gene-based rare variant methods to detect disease-associated variation and test hypotheses about complex disease. *PLoS Genet,* 11**,** e1005165.
MUNZEL, M., GLOBISCH, D., BRUCKL, T., WAGNER, M., WELZMILLER, V., MICHALAKIS, S., MULLER, M., BIEL, M. & CARELL, T. 2010. Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew Chem Int Ed Engl,* 49**,** 5375-7.
NAGARAJAN, R. P., HOGART, A. R., GWYE, Y., MARTIN, M. R. & LASALLE, J. M. 2006. Reduced MeCP2 expression is frequent in autism frontal cortex and correlates with aberrant MECP2 promoter methylation. *Epigenetics,* 1**,** e1-11.
NAGARAJAN, R. P., PATZEL, K. A., MARTIN, M., YASUI, D. H., SWANBERG, S. E., HERTZ-PICCIOTTO, I., HANSEN, R. L., VAN DE WATER, J., PESSAH, I. N., JIANG, R., ROBINSON, W. P. & LASALLE, J. M. 2008. MECP2 promoter methylation and X chromosome inactivation in autism. *Autism Res,* 1**,** 169-78.
NARDONE, S., SAMS, D. S., REUVENI, E., GETSELTER, D., ORON, O., KARPUJ, M. & ELLIOTT, E. 2014. DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways. *Transl Psychiatry,* 4**,** e433.
NAUMANN, A., HOCHSTEIN, N., WEBER, S., FANNING, E. & DOERFLER, W. 2009. A distinct DNA-methylation boundary in the 5'- upstream sequence of the FMR1 promoter binds nuclear proteins and is lost in fragile X syndrome. *Am J Hum Genet,* 85**,** 606-16.
NERI, F., INCARNATO, D. & OLIVIERO, S. 2015. DNA methylation and demethylation dynamics. *Oncotarget,* 6**,** 34049-50.
NERI, F., KREPELOVA, A., INCARNATO, D., MALDOTTI, M., PARLATO, C., GALVAGNI, F., MATARESE, F., STUNNENBERG, H. G. & OLIVIERO, S. 2013. Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. *Cell,* 155**,** 121-34.
NEUL, J. L., KAUFMANN, W. E., GLAZE, D. G., CHRISTODOULOU, J., CLARKE, A. J., BAHI-BUISSON, N., LEONARD, H., BAILEY, M. E., SCHANEN, N. C., ZAPPELLA, M., RENIERI, A., HUPPKE, P., PERCY, A. K. & RETTSEARCH, C. 2010. Rett syndrome: revised diagnostic criteria and nomenclature. *Ann Neurol,* 68**,** 944-50.

NGUYEN, T. T., MA, L. N., SLOVAK, M. L., BANGS, C. D., CHERRY, A. M. & ARBER, D. A. 2006. Identification of novel Runx1 (AML1) translocation partner genes SH3D19, YTHDf2, and ZNF687 in acute myeloid leukemia. *Genes Chromosomes Cancer,* 45**,** 918-32.
NILSSON, K., GUSTAFSON, L. & HULTBERG, B. 2012. Elevated plasma homocysteine level is not primarily related to Alzheimer's disease. *Dement Geriatr Cogn Disord,* 34**,** 121-7.
OBEID, R. & HERRMANN, W. 2006. Mechanisms of homocysteine neurotoxicity in neurodegenerative diseases with special reference to dementia. *FEBS Lett,* 580**,** 2994-3005.
OBEID, R., KOSTOPOULOS, P., KNAPP, J. P., KASOHA, M., BECKER, G., FASSBENDER, K. & HERRMANN, W. 2007. Biomarkers of folate and vitamin B12 are related in blood and cerebrospinal fluid. *Clin Chem,* 53**,** 326-33.
OGAWA, O., ZHU, X., LEE, H. G., RAINA, A., OBRENOVICH, M. E., BOWSER, R., GHANBARI, H. A., CASTELLANI, R. J., PERRY, G. & SMITH, M. A. 2003. Ectopic localization of phosphorylated histone H3 in Alzheimer's disease: a mitotic catastrophe? *Acta Neuropathol,* 105**,** 524-8.
OH, Y. & CHUNG, K. C. 2013. UHRF2, a ubiquitin E3 ligase, acts as a small ubiquitin-like modifier E3 ligase for zinc finger protein 131. *J Biol Chem,* 288**,** 9102-11.
OKADA, Y., YAMAGATA, K., HONG, K., WAKAYAMA, T. & ZHANG, Y. 2010. A role for the elongator complex in zygotic paternal genome demethylation. *Nature,* 463**,** 554-8.
OKANO, M., BELL, D. W., HABER, D. A. & LI, E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell,* 99**,** 247-57.
OLIVEIRA, A. M., HEMSTEDT, T. J. & BADING, H. 2012. Rescue of aging-associated decline in Dnmt3a2 expression restores cognitive abilities. *Nat Neurosci,* 15**,** 1111-3.
OOI, S. K., QIU, C., BERNSTEIN, E., LI, K., JIA, D., YANG, Z., ERDJUMENT-BROMAGE, H., TEMPST, P., LIN, S. P., ALLIS, C. D., CHENG, X. & BESTOR, T. H. 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature,* 448**,** 714-7.
ORR, B. A., HAFFNER, M. C., NELSON, W. G., YEGNASUBRAMANIAN, S. & EBERHART, C. G. 2012. Decreased 5-hydroxymethylcytosine is associated with neural progenitor phenotype in normal brain and shorter survival in malignant glioma. *PLoS One,* 7**,** e41036.
OULHAJ, A., JERNEREN, F., REFSUM, H., SMITH, A. D. & DE JAGER, C. A. 2016. Omega-3 Fatty Acid Status Enhances the Prevention of Cognitive Decline by B Vitamins in Mild Cognitive Impairment. *Journal of Alzheimers Disease,* 50**,** 547-557.
OZAKI, Y., YOSHINO, Y., YAMAZAKI, K., SAO, T., MORI, Y., OCHI, S., YOSHIDA, T., MORI, T., IGA, J. I. & UENO, S. I. 2017. DNA methylation changes at TREM2 intron 1 and TREM2 mRNA expression in patients with Alzheimer's disease. *J Psychiatr Res,* 92**,** 74-80.
PACAUD, R., SERY, Q., OLIVER, L., VALLETTE, F. M., TOST, J. & CARTRON, P. F. 2014. DNMT3L interacts with transcription factors to target DNMT3L/DNMT3B to specific DNA sequences: role of the

DNMT3L/DNMT3B/p65-NFkappaB complex in the (de-)methylation of TRAF1. *Biochimie,* 104**,** 36-49.
PAPASOTIRIOU, I., NTANOVASILIS, D. & APOSTOLOU, P. 2017. Comparison of genomic and gene expression profile between cancer and healthy samples. *Journal of Clinical Oncology,* 35.
PATTERSON, S. L., ABEL, T., DEUEL, T. A., MARTIN, K. C., ROSE, J. C. & KANDEL, E. R. 1996. Recombinant BDNF rescues deficits in basal synaptic transmission and hippocampal LTP in BDNF knockout mice. *Neuron,* 16**,** 1137-45.
PAYAMI, H., ZAREPARSI, S., MONTEE, K. R., SEXTON, G. J., KAYE, J. A., BIRD, T. D., YU, C. E., WIJSMAN, E. M., HESTON, L. L., LITT, M. & SCHELLENBERG, G. D. 1996. Gender difference in apolipoprotein E-associated risk for familial Alzheimer disease: a possible clue to the higher incidence of Alzheimer disease in women. *Am J Hum Genet,* 58**,** 803-11.
PEDROSO, J. L., POVOAS BARSOTTINI, O. G., LIN, L., MELBERG, A., OLIVEIRA, A. S. & MIGNOT, E. 2013. A novel de novo exon 21 DNMT1 mutation causes cerebellar ataxia, deafness, and narcolepsy in a Brazilian patient. *Sleep,* 36**,** 1257-9, 1259A.
PEI, Y. F., ZHANG, L., LI, J. & DENG, H. W. 2010. Analyses and comparison of imputation-based association methods. *PLoS One,* 5**,** e10827.
PELLANDA, H., NAMOUR, F., FOFOU-CAILLIEREZ, M., BRESSENOT, A., ALBERTO, J. M., CHERY, C., AYAV, A., BRONOWICKI, J. P., GUEANT, J. L. & FORGES, T. 2012. A splicing variant leads to complete loss of function of betaine-homocysteine methyltransferase (BHMT) gene in hepatocellular carcinoma. *Int J Biochem Cell Biol,* 44**,** 385-92.
PENG, J. & XU, J. 2011. A multiple-template approach to protein threading. *Proteins,* 79**,** 1930-9.
PERK, J., MAKEDONSKI, K., LANDE, L., CEDAR, H., RAZIN, A. & SHEMER, R. 2002. The imprinting mechanism of the Prader-Willi/Angelman regional control center. *EMBO J,* 21**,** 5807-14.
PERRY, R. J., WATSON, P. & HODGES, J. R. 2000. The nature and staging of attention dysfunction in early (minimal and mild) Alzheimer's disease: relationship to episodic and semantic memory impairment. *Neuropsychologia,* 38**,** 252-71.
PERSICO, A. M. & NAPOLIONI, V. 2013. Autism genetics. *Behav Brain Res,* 251**,** 95-112.
PETERSEN, R. C., CARACCIOLO, B., BRAYNE, C., GAUTHIER, S., JELIC, V. & FRATIGLIONI, L. 2014. Mild cognitive impairment: a concept in evolution. *J Intern Med,* 275**,** 214-28.
PETERSEN, R. C., SMITH, G. E., WARING, S. C., IVNIK, R. J., TANGALOS, E. G. & KOKMEN, E. 1999. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol,* 56**,** 303-8.
PHILLIPS, H. S., HAINS, J. M., ARMANINI, M., LARAMEE, G. R., JOHNSON, S. A. & WINSLOW, J. W. 1991. BDNF mRNA is decreased in the hippocampus of individuals with Alzheimer's disease. *Neuron,* 7**,** 695-702.
PICHLER, G., WOLF, P., SCHMIDT, C. S., MEILINGER, D., SCHNEIDER, K., FRAUER, C., FELLINGER, K., ROTTACH, A. & LEONHARDT, H. 2011. Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J Cell Biochem,* 112**,** 2585-93.

PIERETTI, M., ZHANG, F. P., FU, Y. H., WARREN, S. T., OOSTRA, B. A., CASKEY, C. T. & NELSON, D. L. 1991. Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell,* 66**,** 817-22.

PING, X. L., SUN, B. F., WANG, L., XIAO, W., YANG, X., WANG, W. J., ADHIKARI, S., SHI, Y., LV, Y., CHEN, Y. S., ZHAO, X., LI, A., YANG, Y., DAHAL, U., LOU, X. M., LIU, X., HUANG, J., YUAN, W. P., ZHU, X. F., CHENG, T., ZHAO, Y. L., WANG, X., RENDTLEW DANIELSEN, J. M., LIU, F. & YANG, Y. G. 2014. Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res,* 24**,** 177-89.

PINHEL, M. A., NAKAZONE, M. A., CACAO, J. C., PITERI, R. C., DANTAS, R. T., GODOY, M. F., GODOY, M. R., TOGNOLA, W. A., CONFORTI-FROES, N. D. & SOUZA, D. 2008. Glutathione S-transferase variants increase susceptibility for late-onset Alzheimer's disease: association study and relationship with apolipoprotein E epsilon4 allele. *Clin Chem Lab Med,* 46**,** 439-45.

PIPINGAS, A., CAMFIELD, D. A., STOUGH, C., SCHOLEY, A. B., COX, K. H., WHITE, D., SARRIS, J., SALI, A. & MACPHERSON, H. 2014. Effects of multivitamin, mineral and herbal supplement on cognition in younger adults and the contribution of B group vitamins. *Hum Psychopharmacol,* 29**,** 73-82.

PLAMONDON, H. & ROBERGE, M. C. 2008. Dietary PUFA supplements reduce memory deficits but not CA1 ischemic injury in rats. *Physiol Behav,* 95**,** 492-500.

PLASS, C., PFISTER, S. M., LINDROTH, A. M., BOGATYROVA, O., CLAUS, R. & LICHTER, P. 2013. Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. *Nat Rev Genet,* 14**,** 765-80.

PODDAR, R. & PAUL, S. 2009. Homocysteine-NMDA receptor-mediated activation of extracellular signal-regulated kinase leads to neuronal cell death. *J Neurochem,* 110**,** 1095-106.

POGRIBNA, M., MELNYK, S., POGRIBNY, I., CHANGO, A., YI, P. & JAMES, S. J. 2001. Homocysteine metabolism in children with Down syndrome: in vitro modulation. *Am J Hum Genet,* 69**,** 88-95.

PURSER, J. L., FILLENBAUM, G. G., PIEPER, C. F. & WALLACE, R. B. 2005. Mild cognitive impairment and 10-year trajectories of disability in the Iowa Established Populations for Epidemiologic Studies of the Elderly cohort. *J Am Geriatr Soc,* 53**,** 1966-72.

QI, X., PANG, Q., WANG, J., ZHAO, Z., WANG, O., XU, L., MAO, J., JIANG, Y., LI, M., XING, X., YU, W., ASAN & XIA, W. 2017. Familial Early-Onset Paget's Disease of Bone Associated with a Novel hnRNPA2B1 Mutation. *Calcif Tissue Int,* 101**,** 159-169.

QUADRI, P., FRAGIACOMO, C., PEZZATI, R., ZANDA, E., FORLONI, G., TETTAMANTI, M. & LUCCA, U. 2004. Homocysteine, folate, and vitamin B-12 in mild cognitive impairment, Alzheimer disease, and vascular dementia. *Am J Clin Nutr,* 80**,** 114-22.

RALLAPALLI, P. M., ORENGO, C. A., STUDER, R. A. & PERKINS, S. J. 2014. Positive selection during the evolution of the blood coagulation factors in the context of their disease-causing mutations. *Mol Biol Evol,* 31**,** 3040-56.

RAPOPORT, J., CHAVEZ, A., GREENSTEIN, D., ADDINGTON, A. & GOGTAY, N. 2009. Autism spectrum disorders and childhood-onset schizophrenia: clinical

and biological contributions to a relation revisited. *J Am Acad Child Adolesc Psychiatry,* 48**,** 10-8.
RAY, D., KAZAN, H., COOK, K. B., WEIRAUCH, M. T., NAJAFABADI, H. S., LI, X., GUEROUSSOV, S., ALBU, M., ZHENG, H., YANG, A., NA, H., IRIMIA, M., MATZAT, L. H., DALE, R. K., SMITH, S. A., YAROSH, C. A., KELLY, S. M., NABET, B., MECENAS, D., LI, W., LAISHRAM, R. S., QIAO, M., LIPSHITZ, H. D., PIANO, F., CORBETT, A. H., CARSTENS, R. P., FREY, B. J., ANDERSON, R. A., LYNCH, K. W., PENALVA, L. O., LEI, E. P., FRASER, A. G., BLENCOWE, B. J., MORRIS, Q. D. & HUGHES, T. R. 2013. A compendium of RNA-binding motifs for decoding gene regulation. *Nature,* 499**,** 172-7.
REINHARDT, P., SCHMID, B., BURBULLA, L. F., SCHONDORF, D. C., WAGNER, L., GLATZA, M., HOING, S., HARGUS, G., HECK, S. A., DHINGRA, A., WU, G., MULLER, S., BROCKMANN, K., KLUBA, T., MAISEL, M., KRUGER, R., BERG, D., TSYTSYURA, Y., THIEL, C. S., PSATHAKI, O. E., KLINGAUF, J., KUHLMANN, T., KLEWIN, M., MULLER, H., GASSER, T., SCHOLER, H. R. & STERNECKERT, J. 2013. Genetic correction of a LRRK2 mutation in human iPSCs links parkinsonian neurodegeneration to ERK-dependent changes in gene expression. *Cell Stem Cell,* 12**,** 354-67.
RELIGA, D., STYCZYNSKA, M., PEPLONSKA, B., GABRYELEWICZ, T., PFEFFER, A., CHODAKOWSKA, M., LUCZYWEK, E., WASIAK, B., STEPIEN, K., GOLEBIOWSKI, M., WINBLAD, B. & BARCIKOWSKA, M. 2003. Homocysteine, apolipoproteine E and methylenetetrahydrofolate reductase in Alzheimer's disease and mild cognitive impairment. *Dement Geriatr Cogn Disord,* 16**,** 64-70.
RICHTER, J. D. & KLANN, E. 2009. Making synaptic plasticity and memory last: mechanisms of translational regulation. *Genes Dev,* 23**,** 1-11.
RIDGE, P. G., MUKHERJEE, S., CRANE, P. K., KAUWE, J. S. & ALZHEIMER'S DISEASE GENETICS, C. 2013. Alzheimer's disease: analyzing the missing heritability. *PLoS One,* 8**,** e79771.
ROBERT, M. F., MORIN, S., BEAULIEU, N., GAUTHIER, F., CHUTE, I. C., BARSALOU, A. & MACLEOD, A. R. 2003. DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells. *Nat Genet,* 33**,** 61-5.
ROBERTS, R. O., GEDA, Y. E., KNOPMAN, D. S., CHA, R. H., PANKRATZ, V. S., BOEVE, B. F., TANGALOS, E. G., IVNIK, R. J., ROCCA, W. A. & PETERSEN, R. C. 2012. The incidence of MCI differs by subtype and is higher in men: the Mayo Clinic Study of Aging. *Neurology,* 78**,** 342-51.
ROSE, P. W., PRLIC, A., ALTUNKAYA, A., BI, C., BRADLEY, A. R., CHRISTIE, C. H., COSTANZO, L. D., DUARTE, J. M., DUTTA, S., FENG, Z., GREEN, R. K., GOODSELL, D. S., HUDSON, B., KALRO, T., LOWE, R., PEISACH, E., RANDLE, C., ROSE, A. S., SHAO, C., TAO, Y. P., VALASATAVA, Y., VOIGT, M., WESTBROOK, J. D., WOO, J., YANG, H., YOUNG, J. Y., ZARDECKI, C., BERMAN, H. M. & BURLEY, S. K. 2017. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res,* 45**,** D271-D281.
ROWE, S. J., ROWLATT, A., DAVIES, G., HARRIS, S. E., PORTEOUS, D. J., LIEWALD, D. C., MCNEILL, G., STARR, J. M., DEARY, I. J. & TENESA, A. 2013. Complex variation in measures of general intelligence and cognitive change. *PLoS One,* 8**,** e81189.

ROYBON, L., HJALT, T., STOTT, S., GUILLEMOT, F., LI, J. Y. & BRUNDIN, P. 2009. Neurogenin2 directs granule neuroblast production and amplification while NeuroD1 specifies neuronal fate during hippocampal neurogenesis. *PLoS One,* 4**,** e4779.

RUDENKO, A., DAWLATY, M. M., SEO, J., CHENG, A. W., MENG, J., LE, T., FAULL, K. F., JAENISCH, R. & TSAI, L. H. 2013. Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron,* 79**,** 1109-22.

RUSH, E. C., KATRE, P. & YAJNIK, C. S. 2014. Vitamin B12: one carbon metabolism, fetal growth and programming for chronic disease. *Eur J Clin Nutr,* 68**,** 2-7.

SACHDEV, P. S., LIPNICKI, D. M., CRAWFORD, J., REPPERMUND, S., KOCHAN, N. A., TROLLOR, J. N., DRAPER, B., SLAVIN, M. J., KANG, K., LUX, O., MATHER, K. A., BRODATY, H., MEMORY & AGEING STUDY, T. 2012. Risk profiles of subtypes of mild cognitive impairment: the sydney memory and ageing study. *J Am Geriatr Soc,* 60**,** 24-33.

SAIJA, A., PRINCI, P., PISANI, A., LANZA, M., SCALESE, M., ARAMNEJAD, E., CESERANI, R. & COSTA, G. 1994. Protective effect of glutathione on kainic acid-induced neuropathological changes in the rat brain. *Gen Pharmacol,* 25**,** 97-102.

SANCHEZ-MUT, J. V., HEYN, H., VIDAL, E., MORAN, S., SAYOLS, S., DELGADO-MORALES, R., SCHULTZ, M. D., ANSOLEAGA, B., GARCIA-ESPARCIA, P., PONS-ESPINAL, M., DE LAGRAN, M. M., DOPAZO, J., RABANO, A., AVILA, J., DIERSSEN, M., LOTT, I., FERRER, I., ECKER, J. R. & ESTELLER, M. 2016. Human DNA methylomes of neurodegenerative diseases show common epigenomic patterns. *Transl Psychiatry,* 6**,** e718.

SANDOVAL, J., HEYN, H., MORAN, S., SERRA-MUSACH, J., PUJANA, M. A., BIBIKOVA, M. & ESTELLER, M. 2011. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics,* 6**,** 692-702.

SANTIAGO, M., ANTUNES, C., GUEDES, M., SOUSA, N. & MARQUES, C. J. 2014. TET enzymes and DNA hydroxymethylation in neural development and function - how critical are they? *Genomics,* 104**,** 334-40.

SARADALEKSHMI, K. R., NEETHA, N. V., SATHYAN, S., NAIR, I. V., NAIR, C. M. & BANERJEE, M. 2014. DNA Methyl Transferase (DNMT) Gene Polymorphisms Could Be a Primary Event in Epigenetic Susceptibility to Schizophrenia. *Plos One,* 9.

SASAKI, T., SENDA, M., KIM, S., KOJIMA, S. & KUBODERA, A. 2001. Age-related changes of glutathione content, glucose transport and metabolism, and mitochondrial electron transfer function in mouse brain. *Nucl Med Biol,* 28**,** 25-31.

SASSI, C., GUERREIRO, R., GIBBS, R., DING, J., LUPTON, M. K., TROAKES, C., AL-SARRAJ, S., NIBLOCK, M., GALLO, J. M., ADNAN, J., KILLICK, R., BROWN, K. S., MEDWAY, C., LORD, J., TURTON, J., BRAS, J., ALZHEIMER'S RESEARCH, U. K. C., MORGAN, K., POWELL, J. F., SINGLETON, A. & HARDY, J. 2014. Investigating the role of rare coding variability in Mendelian dementia genes (APP, PSEN1, PSEN2, GRN, MAPT, and PRNP) in late-onset Alzheimer's disease. *Neurobiol Aging,* 35**,** 2881 e1-6.

SAYEED, S. K., ZHAO, J., SATHYANARAYANA, B. K., GOLLA, J. P. & VINSON, C. 2015. C/EBPbeta (CEBPB) protein binding to the C/EBP|CRE DNA 8-mer

TTGC|GTCA is inhibited by 5hmC and enhanced by 5mC, 5fC, and 5caC in the CG dinucleotide. *Biochim Biophys Acta,* 1849**,** 583-9.
SCALA, I., GRANESE, B., SELLITTO, M., SALOME, S., SAMMARTINO, A., PEPE, A., MASTROIACOVO, P., SEBASTIO, G. & ANDRIA, G. 2006. Analysis of seven maternal polymorphisms of genes involved in homocysteine/folate metabolism and risk of Down syndrome offspring. *Genet Med,* 8**,** 409-16.
SCARPA, S., FUSO, A., D'ANSELMI, F. & CAVALLARO, R. A. 2003. Presenilin 1 gene silencing by S-adenosylmethionine: a treatment for Alzheimer disease? *FEBS Lett,* 541**,** 145-8.
SCHARFMAN, H., GOODMAN, J., MACLEOD, A., PHANI, S., ANTONELLI, C. & CROLL, S. 2005. Increased neurogenesis and the ectopic granule cells after intrahippocampal BDNF infusion in adult rats. *Exp Neurol,* 192**,** 348-56.
SCHRÖDINGER, L. The PYMOL Molecular Graphics System. 1.3 ed.
SCHUCHARDT, K., GEBHARDT, M. & MAEHLER, C. 2010. Working memory functions in children with different degrees of intellectual disability. *J Intellect Disabil Res,* 54**,** 346-53.
SCHULLER, U., KHO, A. T., ZHAO, Q., MA, Q. & ROWITCH, D. H. 2006. Cerebellar 'transcriptome' reveals cell-type and stage-specific expression during postnatal development and tumorigenesis. *Mol Cell Neurosci,* 33**,** 247-59.
SCHULTZ, M. D., HE, Y., WHITAKER, J. W., HARIHARAN, M., MUKAMEL, E. A., LEUNG, D., RAJAGOPAL, N., NERY, J. R., URICH, M. A., CHEN, H., LIN, S., LIN, Y., JUNG, I., SCHMITT, A. D., SELVARAJ, S., REN, B., SEJNOWSKI, T. J., WANG, W. & ECKER, J. R. 2015. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature,* 523**,** 212-6.
SCHWARZ, J. M., COOPER, D. N., SCHUELKE, M. & SEELOW, D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods,* 11**,** 361-2.
SCHYMKOWITZ, J., BORG, J., STRICHER, F., NYS, R., ROUSSEAU, F. & SERRANO, L. 2005. The FoldX web server: an online force field. *Nucleic Acids Res,* 33**,** W382-8.
SEELEN, M., VISSER, A. E., OVERSTE, D. J., KIM, H. J., PALUD, A., WONG, T. H., VAN SWIETEN, J. C., SCHELTENS, P., VOERMANS, N. C., BAAS, F., DE JONG, J. M., VAN DER KOOI, A. J., DE VISSER, M., VELDINK, J. H., TAYLOR, J. P., VAN ES, M. A. & VAN DEN BERG, L. H. 2014. No mutations in hnRNPA1 and hnRNPA2B1 in Dutch patients with amyotrophic lateral sclerosis, frontotemporal dementia, and inclusion body myopathy. *Neurobiol Aging,* 35**,** 1956 e9-1956 e11.
SELLEY, M. L. 2007. A metabolic link between S-adenosylhomocysteine and polyunsaturated fatty acid metabolism in Alzheimer's disease. *Neurobiol Aging,* 28**,** 1834-9.
SHEA, T. B., ROGERS, E., ASHLINE, D., ORTIZ, D. & SHEU, M. S. 2002. Apolipoprotein E deficiency promotes increased oxidative stress and compensatory increases in antioxidants in brain tissue. *Free Radic Biol Med,* 33**,** 1115-20.
SHEN, L., WU, H., DIEP, D., YAMAGUCHI, S., D'ALESSIO, A. C., FUNG, H. L., ZHANG, K. & ZHANG, Y. 2013. Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell,* 153**,** 692-706.
SHETH, F. J., DATAR, C., ANDRIEUX, J., PANDIT, A., NAYAK, D., RAHMAN, M. & SHETH, J. J. 2014. Distal Deletion of Chromosome 11q Encompassing

Jacobsen Syndrome without Platelet Abnormality. *Clin Med Insights Pediatr,* 8**,** 45-9.
SHI, H., WANG, X., LU, Z., ZHAO, B. S., MA, H., HSU, P. J., LIU, C. & HE, C. 2017. YTHDF3 facilitates translation and decay of N6-methyladenosine-modified RNA. *Cell Res,* 27**,** 315-328.
SHIBUTANI, T., ITO, S., TODA, M., KANAO, R., COLLINS, L. B., SHIBATA, M., URABE, M., KOSEKI, H., MASUDA, Y., SWENBERG, J. A., MASUTANI, C., HANAOKA, F., IWAI, S. & KURAOKA, I. 2014. Guanine- 5-carboxylcytosine base pairs mimic mismatches during DNA replication. *Sci Rep,* 4**,** 5220.
SHIHAB, H. A., GOUGH, J., COOPER, D. N., DAY, I. N. & GAUNT, T. R. 2013. Predicting the functional consequences of cancer-associated amino acid substitutions. *Bioinformatics,* 29**,** 1504-10.
SHINAGAWA, S., KOBAYASHI, N., NAGATA, T., KUSAKA, A., YAMADA, H., KONDO, K. & NAKAYAMA, K. 2016. DNA methylation in the NCAPH2/LMF2 promoter region is associated with hippocampal atrophy in Alzheimer's disease and amnesic mild cognitive impairment patients. *Neuroscience Letters,* 629**,** 33-37.
SHLENS, J. 2014. A tutorial on Principal Component Analysis. *arXiv preprint arXiv:1404.1100*.
SHOREY-KENDRICK, L. E., MCEVOY, C. T., FERGUSON, B., BURCHARD, J., PARK, B. S., GAO, L., VUYLSTEKE, B. H., MILNER, K. F., MORRIS, C. D. & SPINDEL, E. R. 2017. Vitamin C Prevents Offspring DNA Methylation Changes Associated with Maternal Smoking in Pregnancy. *Am J Respir Crit Care Med*.
SINGER, J. J., MACGREGOR, A. J., CHERKAS, L. F. & SPECTOR, T. D. 2006. Genetic influences on cognitive function using the Cambridge neuropsychological test automated battery. *Intelligence,* 34**,** 421-428.
SINGH, N., WANG, A. Y., SANKARANARAYANAN, P., FLETCHER, P. T., JOSHI, S. & ALZHEIMER'S DISEASE NEUROIMAGING, I. 2012. Genetic, structural and functional imaging biomarkers for early detection of conversion from MCI to AD. *Med Image Comput Comput Assist Interv,* 15**,** 132-40.
SINGH, T., KURKI, M. I., CURTIS, D., PURCELL, S. M., CROOKS, L., MCRAE, J., SUVISAARI, J., CHHEDA, H., BLACKWOOD, D., BREEN, G., PIETILAINEN, O., GERETY, S. S., AYUB, M., BLYTH, M., COLE, T., COLLIER, D., COOMBER, E. L., CRADDOCK, N., DALY, M. J., DANESH, J., DIFORTI, M., FOSTER, A., FREIMER, N. B., GESCHWIND, D., JOHNSTONE, M., JOSS, S., KIROV, G., KORKKO, J., KUISMIN, O., HOLMANS, P., HULTMAN, C. M., IYEGBE, C., LONNQVIST, J., MANNIKKO, M., MCCARROLL, S. A., MCGUFFIN, P., MCINTOSH, A. M., MCQUILLIN, A., MOILANEN, J. S., MOORE, C., MURRAY, R. M., NEWBURY-ECOB, R., OUWEHAND, W., PAUNIO, T., PRIGMORE, E., REES, E., ROBERTS, D., SAMBROOK, J., SKLAR, P., ST CLAIR, D., VEIJOLA, J., WALTERS, J. T., WILLIAMS, H., SWEDISH SCHIZOPHRENIA, S., STUDY, I., STUDY, D. D. D., CONSORTIUM, U. K., SULLIVAN, P. F., HURLES, M. E., O'DONOVAN, M. C., PALOTIE, A., OWEN, M. J. & BARRETT, J. C. 2016. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat Neurosci,* 19**,** 571-7.
SINGH, T., WALTERS, J. T. R., JOHNSTONE, M., CURTIS, D., SUVISAARI, J., TORNIAINEN, M., REES, E., IYEGBE, C., BLACKWOOD, D., MCINTOSH, A. M., KIROV, G., GESCHWIND, D., MURRAY, R. M., DI FORTI, M., BRAMON, E.,

GANDAL, M., HULTMAN, C. M., SKLAR, P., STUDY, I., CONSORTIUM, U. K., PALOTIE, A., SULLIVAN, P. F., O'DONOVAN, M. C., OWEN, M. J. & BARRETT, J. C. 2017. The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nat Genet,* 49**,** 1167-1173.

SINHA, N. & SMITH-GILL, S. J. 2002. Electrostatics in protein binding and function. *Curr Protein Pept Sci,* 3**,** 601-14.

SITKOFF, D., LOCKHART, D. J., SHARP, K. A. & HONIG, B. 1994. Calculation of electrostatic effects at the amino terminus of an alpha helix. *Biophys J,* 67**,** 2251-60.

SIVRIOGLU, E. Y., SIVRIOGLU, K., ERTAN, T., ERTAN, F. S., CANKURTARAN, E., AKI, O., ULUDUZ, D., INCE, B. & KIRLI, S. 2009. Reliability and validity of the Geriatric Depression Scale in detection of poststroke minor depression. *J Clin Exp Neuropsychol,* 31**,** 999-1006.

SLATKIN, M. 2009. Epigenetic inheritance and the missing heritability problem. *Genetics,* 182**,** 845-50.

SLIWINSKA, A., SITAREK, P., TOMA, M., CZARNY, P., SYNOWIEC, E., KRUPA, R., WIGNER, P., BIALEK, K., KWIATKOWSKI, D., KORYCINSKA, A., MAJSTEREK, I., SZEMRAJ, J., GALECKI, P. & SLIWINSKI, T. 2017. Decreased expression level of BER genes in Alzheimer's disease patients is not derivative of their DNA methylation status. *Prog Neuropsychopharmacol Biol Psychiatry,* 79**,** 311-316.

SMETS, M., LINK, S., WOLF, P., SCHNEIDER, K., SOLIS, V., RYAN, J., MEILINGER, D., QIN, W. & LEONHARDT, H. 2017. DNMT1 mutations found in HSANIE patients affect interaction with UHRF1 and neuronal differentiation. *Hum Mol Genet,* 26**,** 1522-1534.

SMITH, A. D., SMITH, S. M., DE JAGER, C. A., WHITBREAD, P., JOHNSTON, C., AGACINSKI, G., OULHAJ, A., BRADLEY, K. M., JACOBY, R. & REFSUM, H. 2010. Homocysteine-lowering by B vitamins slows the rate of accelerated brain atrophy in mild cognitive impairment: a randomized controlled trial. *PLoS One,* 5**,** e12244.

SMITH, A. R., SMITH, R. G., CONDLIFFE, D., HANNON, E., SCHALKWYK, L., MILL, J. & LUNNON, K. 2016. Increased DNA methylation near TREM2 is consistently seen in the superior temporal gyrus in Alzheimer's disease brain. *Neurobiol Aging,* 47**,** 35-40.

SMITH, B. N., TICOZZI, N., FALLINI, C., GKAZI, A. S., TOPP, S., KENNA, K. P., SCOTTER, E. L., KOST, J., KEAGLE, P., MILLER, J. W., CALINI, D., VANCE, C., DANIELSON, E. W., TROAKES, C., TILOCA, C., AL-SARRAJ, S., LEWIS, E. A., KING, A., COLOMBRITA, C., PENSATO, V., CASTELLOTTI, B., DE BELLEROCHE, J., BAAS, F., TEN ASBROEK, A. L., SAPP, P. C., MCKENNA-YASEK, D., MCLAUGHLIN, R. L., POLAK, M., ASRESS, S., ESTEBAN-PEREZ, J., MUNOZ-BLANCO, J. L., SIMPSON, M., CONSORTIUM, S., VAN RHEENEN, W., DIEKSTRA, F. P., LAURIA, G., DUGA, S., CORTI, S., CEREDA, C., CORRADO, L., SORARU, G., MORRISON, K. E., WILLIAMS, K. L., NICHOLSON, G. A., BLAIR, I. P., DION, P. A., LEBLOND, C. S., ROULEAU, G. A., HARDIMAN, O., VELDINK, J. H., VAN DEN BERG, L. H., AL-CHALABI, A., PALL, H., SHAW, P. J., TURNER, M. R., TALBOT, K., TARONI, F., GARCIA-REDONDO, A., WU, Z., GLASS, J. D., GELLERA, C., RATTI, A., BROWN, R. H., JR., SILANI, V., SHAW,

C. E. & LANDERS, J. E. 2014. Exome-wide rare variant analysis identifies TUBA4A mutations associated with familial ALS. *Neuron,* 84**,** 324-31.
SMITH, S. S., KAPLAN, B. E., SOWERS, L. C. & NEWMAN, E. M. 1992. Mechanism of human methyl-directed DNA methyltransferase and the fidelity of cytosine methylation. *Proc Natl Acad Sci U S A,* 89**,** 4744-8.
SNIEKERS, S., STRINGER, S., WATANABE, K., JANSEN, P. R., COLEMAN, J. R. I., KRAPOHL, E., TASKESEN, E., HAMMERSCHLAG, A. R., OKBAY, A., ZABANEH, D., AMIN, N., BREEN, G., CESARINI, D., CHABRIS, C. F., IACONO, W. G., IKRAM, M. A., JOHANNESSON, M., KOELLINGER, P., LEE, J. J., MAGNUSSON, P. K. E., MCGUE, M., MILLER, M. B., OLLIER, W. E. R., PAYTON, A., PENDLETON, N., PLOMIN, R., RIETVELD, C. A., TIEMEIER, H., VAN DUIJN, C. M. & POSTHUMA, D. 2017. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat Genet,* 49**,** 1107-1112.
SONG, J., RECHKOBLIT, O., BESTOR, T. H. & PATEL, D. J. 2011. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science,* 331**,** 1036-40.
SOONG, B. W., LIN, K. P., GUO, Y. C., LIN, C. C., TSAI, P. C., LIAO, Y. C., LU, Y. C., WANG, S. J., TSAI, C. P. & LEE, Y. C. 2014. Extensive molecular genetic survey of Taiwanese patients with amyotrophic lateral sclerosis. *Neurobiol Aging,* 35**,** 2423 e1-6.
SPRUIJT, C. G., GNERLICH, F., SMITS, A. H., PFAFFENEDER, T., JANSEN, P. W., BAUER, C., MUNZEL, M., WAGNER, M., MULLER, M., KHAN, F., EBERL, H. C., MENSINGA, A., BRINKMAN, A. B., LEPHIKOV, K., MULLER, U., WALTER, J., BOELENS, R., VAN INGEN, H., LEONHARDT, H., CARELL, T. & VERMEULEN, M. 2013. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell,* 152**,** 1146-59.
ST CLAIR, D., BLACKWOOD, D., MUIR, W., CAROTHERS, A., WALKER, M., SPOWART, G., GOSDEN, C. & EVANS, H. J. 1990. Association within a family of a balanced autosomal translocation with major mental illness. *Lancet,* 336**,** 13-6.
ST CROIX, C. M., SHAND, S. H. & WATKINS, S. C. 2005. Confocal microscopy: comparisons, applications, and problems. *Biotechniques,* 39**,** S2-5.
STEINBERG, J. & WEBBER, C. 2013. The roles of FMRP-regulated genes in autism spectrum disorder: single- and multiple-hit genetic etiologies. *Am J Hum Genet,* 93**,** 825-39.
STEINBERG, S., STEFANSSON, H., JONSSON, T., JOHANNSDOTTIR, H., INGASON, A., HELGASON, H., SULEM, P., MAGNUSSON, O. T., GUDJONSSON, S. A., UNNSTEINSDOTTIR, U., KONG, A., HELISALMI, S., SOININEN, H., LAH, J. J., DEMGENE, AARSLAND, D., FLADBY, T., ULSTEIN, I. D., DJUROVIC, S., SANDO, S. B., WHITE, L. R., KNUDSEN, G. P., WESTLYE, L. T., SELBAEK, G., GIEGLING, I., HAMPEL, H., HILTUNEN, M., LEVEY, A. I., ANDREASSEN, O. A., RUJESCU, D., JONSSON, P. V., BJORNSSON, S., SNAEDAL, J. & STEFANSSON, K. 2015. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. *Nat Genet,* 47**,** 445-7.
STEPHAN, B. C., HUNTER, S., HARRIS, D., LLEWELLYN, D. J., SIERVO, M., MATTHEWS, F. E. & BRAYNE, C. 2012. The neuropathological profile of mild cognitive impairment (MCI): a systematic review. *Mol Psychiatry,* 17**,** 1056-76.

STEPPER, P., KUNGULOVSKI, G., JURKOWSKA, R. Z., CHANDRA, T., KRUEGER, F., REINHARDT, R., REIK, W., JELTSCH, A. & JURKOWSKI, T. P. 2017. Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. *Nucleic Acids Res,* 45**,** 1703-1713.

STUDER, R. A., CHRISTIN, P. A., WILLIAMS, M. A. & ORENGO, C. A. 2014a. Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proc Natl Acad Sci U S A,* 111**,** 2223-8.

STUDER, R. A., OPPERDOES, F. R., NICOLAES, G. A., MULDER, A. B. & MULDER, R. 2014b. Understanding the functional difference between growth arrest-specific protein 6 and protein S: an evolutionary approach. *Open Biol,* 4.

SU, S. C., LIN, C. W., LIU, Y. F., FAN, W. L., CHEN, M. K., YU, C. P., YANG, W. E., SU, C. W., CHUANG, C. Y., LI, W. H., CHUNG, W. H. & YANG, S. F. 2017. Exome Sequencing of Oral Squamous Cell Carcinoma Reveals Molecular Subgroups and Novel Therapeutic Opportunities. *Theranostics,* 7**,** 1088-1099.

SU, X., SHANG, L., XU, Q., LI, N., CHEN, J., ZHANG, L., ZHANG, L. & HUA, Q. 2014. Prevalence and predictors of mild cognitive impairment in Xi'an: a community-based study among the elders. *PLoS One,* 9**,** e83217.

SUETAKE, I., SHINOZAKI, F., MIYAGAWA, J., TAKESHIMA, H. & TAJIMA, S. 2004. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J Biol Chem,* 279**,** 27816-23.

SULLIVAN, P. F. 2013. Questions about DISC1 as a genetic risk factor for schizophrenia. *Mol Psychiatry,* 18**,** 1050-2.

SULLIVAN, P. F., KENDLER, K. S. & NEALE, M. C. 2003. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry,* 60**,** 1187-92.

SUN, M. Y., YANG, X. X., XU, W. W., YAO, G. Y., PAN, H. Z. & LI, M. 2012. Association of DNMT1 and DNMT3B polymorphisms with breast cancer risk in Han Chinese women from South China. *Genet Mol Res,* 11**,** 4330-41.

SUN, W., ZANG, L., SHU, Q. & LI, X. 2014. From development to diseases: the role of 5hmC in brain. *Genomics,* 104**,** 347-51.

SUN, Y., LU, C. J., CHIEN, K. L., CHEN, S. T. & CHEN, R. C. 2007. Efficacy of multivitamin supplementation containing vitamins B6 and B12 and folic acid as adjunctive treatment with a cholinesterase inhibitor in Alzheimer's disease: a 26-week, randomized, double-blind, placebo-controlled study in Taiwanese patients. *Clin Ther,* 29**,** 2204-14.

SUTCLIFFE, J. S., NELSON, D. L., ZHANG, F., PIERETTI, M., CASKEY, C. T., SAXE, D. & WARREN, S. T. 1992. DNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum Mol Genet,* 1**,** 397-400.

SZULWACH, K. E., LI, X., LI, Y., SONG, C. X., WU, H., DAI, Q., IRIER, H., UPADHYAY, A. K., GEARING, M., LEVEY, A. I., VASANTHAKUMAR, A., GODLEY, L. A., CHANG, Q., CHENG, X., HE, C. & JIN, P. 2011. 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat Neurosci,* 14**,** 1607-16.

TAHARA, T., SHIBATA, T., ARISAWA, T., NAKAMURA, M., YAMASHITA, H., YOSHIOKA, D., OKUBO, M., MARUYAMA, N., KAMANO, T., KAMIYA, Y., FUJITA, H., NAGASAKA, M., IWATA, M., TAKAHAMA, K., WATANABE, M. & HIRATA, I. 2009. Impact of catechol-O-methyltransferase (COMT) gene

polymorphism on promoter methylation status in gastric mucosa. *Anticancer Res,* 29**,** 2857-61.
TAHILIANI, M., KOH, K. P., SHEN, Y., PASTOR, W. A., BANDUKWALA, H., BRUDNO, Y., AGARWAL, S., IYER, L. M., LIU, D. R., ARAVIND, L. & RAO, A. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science,* 324**,** 930-5.
TAKATA, A., XU, B., IONITA-LAZA, I., ROOS, J. L., GOGOS, J. A. & KARAYIORGOU, M. 2014. Loss-of-function variants in schizophrenia risk and SETD1A as a candidate susceptibility gene. *Neuron,* 82**,** 773-80.
TAKEUCHI, T., DUSZKIEWICZ, A. J. & MORRIS, R. G. 2014. The synaptic plasticity and memory hypothesis: encoding, storage and persistence. *Philos Trans R Soc Lond B Biol Sci,* 369**,** 20130288.
TANAKA, T., SCHEET, P., GIUSTI, B., BANDINELLI, S., PIRAS, M. G., USALA, G., LAI, S., MULAS, A., CORSI, A. M., VESTRINI, A., SOFI, F., GORI, A. M., ABBATE, R., GURALNIK, J., SINGLETON, A., ABECASIS, G. R., SCHLESSINGER, D., UDA, M. & FERRUCCI, L. 2009. Genome-wide association study of vitamin B6, vitamin B12, folate, and homocysteine blood concentrations. *Am J Hum Genet,* 84**,** 477-82.
TANNORELLA, P., STOCCORO, A., TOGNONI, G., PETROZZI, L., SALLUZZO, M. G., RAGALMUTO, A., SICILIANO, G., HASLBERGER, A., BOSCO, P., BONUCCELLI, U., MIGLIORE, L. & COPPEDE, F. 2015. Methylation analysis of multiple genes in blood DNA of Alzheimer's disease and healthy individuals. *Neurosci Lett,* 600**,** 143-7.
TATTON-BROWN, K., SEAL, S., RUARK, E., HARMER, J., RAMSAY, E., DEL VECCHIO DUARTE, S., ZACHARIOU, A., HANKS, S., O'BRIEN, E., AKSGLAEDE, L., BARALLE, D., DABIR, T., GENER, B., GOUDIE, D., HOMFRAY, T., KUMAR, A., PILZ, D. T., SELICORNI, A., TEMPLE, I. K., VAN MALDERGEM, L., YACHELEVICH, N., CHILDHOOD OVERGROWTH, C., VAN MONTFORT, R. & RAHMAN, N. 2014. Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. *Nat Genet,* 46**,** 385-8.
TCHANTCHOU, F., GRAVES, M., ORTIZ, D., CHAN, A., ROGERS, E. & SHEA, T. B. 2006. S-adenosyl methionine: A connection between nutritional and genetic risk factors for neurodegeneration in Alzheimer's disease. *J Nutr Health Aging,* 10**,** 541-4.
TENG, S., THOMSON, P. A., MCCARTHY, S., KRAMER, M., MULLER, S., LIHM, J., MORRIS, S., SOARES, D. C., HENNAH, W., HARRIS, S., CAMARGO, L. M., MALKOV, V., MCINTOSH, A. M., MILLAR, J. K., BLACKWOOD, D. H., EVANS, K. L., DEARY, I. J., PORTEOUS, D. J. & MCCOMBIE, W. R. 2017. Rare disruptive variants in the DISC1 Interactome and Regulome: association with cognitive ability and schizophrenia. *Mol Psychiatry*.
THOMAZEAU, A., BOSCH-BOUJU, C., MANZONI, O. & LAYE, S. 2017. Nutritional n-3 PUFA Deficiency Abolishes Endocannabinoid Gating of Hippocampal Long-Term Potentiation. *Cereb Cortex,* 27**,** 2571-2579.
TICK, B., BOLTON, P., HAPPE, F., RUTTER, M. & RIJSDIJK, F. 2016. Heritability of autism spectrum disorders: a meta-analysis of twin studies. *J Child Psychol Psychiatry,* 57**,** 585-95.
TOKURIKI, N., STRICHER, F., SERRANO, L. & TAWFIK, D. S. 2008. How protein stability and new functions trade off. *PLoS Comput Biol,* 4**,** e1000002.

TORRICO, B., CHIOCCHETTI, A. G., BACCHELLI, E., TRABETTI, E., HERVAS, A., FRANKE, B., BUITELAAR, J. K., ROMMELSE, N., YOUSAF, A., DUKETIS, E., FREITAG, C. M., CABALLERO-ANDALUZ, R., MARTINEZ-MIR, A., SCHOLL, F. G., RIBASES, M., ITAN, BATTAGLIA, A., MALERBA, G., DELORME, R., BENABOU, M., MAESTRINI, E., BOURGERON, T., CORMAND, B. & TOMA, C. 2017. Lack of replication of previous autism spectrum disorder GWAS hits in European populations. *Autism Res,* 10**,** 202-211.

TUCKER, K. L., QIAO, N., SCOTT, T., ROSENBERG, I. & SPIRO, A., 3RD 2005. High homocysteine and low B vitamins predict cognitive decline in aging men: the Veterans Affairs Normative Aging Study. *Am J Clin Nutr,* 82**,** 627-35.

TULVING, E. 1972. Episodic and semantic memory. *In:* TULVING, E. & DONALDSON, W. (eds.) *Organisation of Memory.* New York: Academic Press.

ULREY, C. L., LIU, L., ANDREWS, L. G. & TOLLEFSBOL, T. O. 2005. The impact of metabolism on DNA methylation. *Hum Mol Genet,* 14 Spec No 1**,** R139-47.

UNTERGRASSER, A., CUTCUTACHE, I., KORESSAAR, T., YE, J., FAIRCLOTH, B. C., REMM, M. & ROZEN, S. G. 2012. Primer3 - new capabilities and interfaces. *Nucleic Acids Res,* 40**,** e115.

UPADHYAY, R., SANDUJA, S., KAZA, V. & DIXON, D. A. 2013. Genetic polymorphisms in RNA binding proteins contribute to breast cancer survival. *Int J Cancer,* 132**,** E128-38.

UREN, P. J., BURNS, S. C., RUAN, J., SINGH, K. K., SMITH, A. D. & PENALVA, L. O. 2011. Genomic analyses of the RNA-binding protein Hu antigen R (HuR) identify a complex network of target genes and novel characteristics of its binding sites. *J Biol Chem,* 286**,** 37063-6.

UTTL, B. & PILKENTON-TAYLOR, C. 2001. Letter cancellation performance across the adult life span. *Clin Neuropsychol,* 15**,** 521-30.

VAN UFFELEN, J. G., CHINAPAW, M. J., VAN MECHELEN, W. & HOPMAN-ROCK, M. 2008. Walking or vitamin B for cognition in older adults with mild cognitive impairment? A randomised controlled trial. *Br J Sports Med,* 42**,** 344-51.

VANCASSEL, S., LEMAN, S., HANONICK, L., DENIS, S., ROGER, J., NOLLET, M., BODARD, S., KOUSIGNIAN, I., BELZUNG, C. & CHALON, S. 2008. n-3 polyunsaturated fatty acid supplementation reverses stress-induced modifications on brain monoamine levels in mice. *J Lipid Res,* 49**,** 340-8.

VANDERWEELE, T. J., TCHETGEN TCHETGEN, E. J., CORNELIS, M. & KRAFT, P. 2014. Methodological challenges in mendelian randomization. *Epidemiology,* 25**,** 427-35.

VARADE, J., COMABELLA, M., ORTIZ, M. A., ARROYO, R., FERNANDEZ, O., PINTO-MEDEL, M. J., FEDETZ, M., IZQUIERDO, G., LUCAS, M., GOMEZ, C. L., RABASA, A. C., ALCINA, A., MATESANZ, F., ALLOZA, I., ANTIGUEDAD, A., GARCIA-BARCINA, M., OTAEGUI, D., OLASCOAGA, J., SAIZ, A., BLANCO, Y., MONTALBAN, X., VANDENBROECK, K. & URCELAY, E. 2012. Replication study of 10 genes showing evidence for association with multiple sclerosis: validation of TMEM39A, IL12B and CBLB [correction of CLBL] genes. *Mult Scler,* 18**,** 959-65.

VILLELA, D., RAMALHO, R. F., SILVA, A. R., BRENTANI, H., SUEMOTO, C. K., PASQUALUCCI, C. A., GRINBERG, L. T., KREPISCHI, A. C. & ROSENBERG, C. 2016. Differential DNA Methylation of MicroRNA Genes in Temporal

Cortex from Alzheimer's Disease Individuals. *Neural Plast,* 2016**,** 2584940.
VINA, J. & LLORET, A. 2010. Why women have more Alzheimer's disease than men: gender and mitochondrial toxicity of amyloid-beta peptide. *J Alzheimers Dis,* 20 Suppl 2**,** S527-33.
VITVITSKY, V., THOMAS, M., GHORPADE, A., GENDELMAN, H. E. & BANERJEE, R. 2006. A functional transsulfuration pathway in the brain links to glutathione homeostasis. *J Biol Chem,* 281**,** 35785-93.
VOGIATZOGLOU, A., SMITH, A. D., NURK, E., DREVON, C. A., UELAND, P. M., VOLLSET, S. E., NYGAARD, H. A., ENGEDAL, K., TELL, G. S. & REFSUM, H. 2013. Cognitive function in an elderly population: interaction between vitamin B12 status, depression, and apolipoprotein E epsilon4: the Hordaland Homocysteine Study. *Psychosom Med,* 75**,** 20-9.
WANG, G. T. 2017. *KBAC Statistic Implementation* [Online]. Available: http://tigerwang.org/software/kbac.
WANG, Q., MOORE, M. J., ADELMANT, G., MARTO, J. A. & SILVER, P. A. 2013. PQBP1, a factor linked to intellectual disability, affects alternative splicing associated with neurite outgrowth. *Genes Dev,* 27**,** 615-26.
WANG, S. C., OELZE, B. & SCHUMACHER, A. 2008. Age-specific epigenetic drift in late-onset Alzheimer's disease. *PLoS One,* 3**,** e2698.
WANG, T., PAN, Q., LIN, L., SZULWACH, K. E., SONG, C. X., HE, C., WU, H., WARREN, S. T., JIN, P., DUAN, R. & LI, X. 2012. Genome-wide DNA hydroxymethylation changes are associated with neurodevelopmental genes in the developing human cerebellum. *Hum Mol Genet,* 21**,** 5500-10.
WANG, X., LU, Z., GOMEZ, A., HON, G. C., YUE, Y., HAN, D., FU, Y., PARISIEN, M., DAI, Q., JIA, G., REN, B., PAN, T. & HE, C. 2014a. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature,* 505**,** 117-20.
WANG, Y., LI, Y., TOTH, J. I., PETROSKI, M. D., ZHANG, Z. & ZHAO, J. C. 2014b. N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat Cell Biol,* 16**,** 191-8.
WATSON, C. T., ROUSSOS, P., GARG, P., HO, D. J., AZAM, N., KATSEL, P. L., HAROUTUNIAN, V. & SHARP, A. J. 2016. Genome-wide DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease. *Genome Med,* 8**,** 5.
WEBSTER, K. E., O'BRYAN, M. K., FLETCHER, S., CREWTHER, P. E., AAPOLA, U., CRAIG, J., HARRISON, D. K., AUNG, H., PHUTIKANIT, N., LYLE, R., MEACHEM, S. J., ANTONARAKIS, S. E., DE KRETSER, D. M., HEDGER, M. P., PETERSON, P., CARROLL, B. J. & SCOTT, H. S. 2005. Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proc Natl Acad Sci U S A,* 102**,** 4068-73.
WEISENBERGER, D. J., VELICESCU, M., CHENG, J. C., GONZALES, F. A., LIANG, G. & JONES, P. A. 2004. Role of the DNA methyltransferase variant DNMT3b3 in DNA methylation. *Mol Cancer Res,* 2**,** 62-72.
WEISSMAN, L., JO, D. G., SORENSEN, M. M., DE SOUZA-PINTO, N. C., MARKESBERY, W. R., MATTSON, M. P. & BOHR, V. A. 2007. Defective DNA base excision repair in brain from individuals with Alzheimer's disease and amnestic mild cognitive impairment. *Nucleic Acids Res,* 35**,** 5545-55.
WELLENREUTHER, M. & HANSSON, B. 2016. Detecting Polygenic Evolution: Problems, Pitfalls, and Promises. *Trends Genet,* 32**,** 155-64.

WELLS, R., SWAMINATHAN, V., SUNDRAM, S., WEINBERG, D., BRUGGEMANN, J., JACOMB, I., CROPLEY, V., LENROOT, R., PEREIRA, A. M., ZALESKY, A., BOUSMAN, C., PANTELIS, C., WEICKERT, C. S. & WEICKERT, T. W. 2015. The impact of premorbid and current intellect in schizophrenia: cognitive, symptom, and functional outcomes. *NPJ Schizophr,* 1**,** 15043.

WEYN-VANHENTENRYCK, S. M., MELE, A., YAN, Q., SUN, S., FARNY, N., ZHANG, Z., XUE, C., HERRE, M., SILVER, P. A., ZHANG, M. Q., KRAINER, A. R., DARNELL, R. B. & ZHANG, C. 2014. HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep,* 6**,** 1139-52.

WHELDON, L. M., ABAKIR, A., FERJENTSIK, Z., DUDNAKOVA, T., STROHBUECKER, S., CHRISTIE, D., DAI, N., GUAN, S., FOSTER, J. M., CORREA, I. R., JR., LOOSE, M., DIXON, J. E., SOTTILE, V., JOHNSON, A. D. & RUZOV, A. 2014. Transient accumulation of 5-carboxylcytosine indicates involvement of active demethylation in lineage specification of neural stem cells. *Cell Rep,* 7**,** 1353-61.

WIDAGDO, J., ZHAO, Q. Y., KEMPEN, M. J., TAN, M. C., RATNU, V. S., WEI, W., LEIGHTON, L., SPADARO, P. A., EDSON, J., ANGGONO, V. & BREDY, T. W. 2016. Experience-Dependent Accumulation of N6-Methyladenosine in the Prefrontal Cortex Is Associated with Memory Processes in Mice. *J Neurosci,* 36**,** 6771-7.

WILKINSON, L. S., DAVIES, W. & ISLES, A. R. 2007. Genomic imprinting effects on brain development and function. *Nat Rev Neurosci,* 8**,** 832-43.

WINKELMANN, J., LIN, L., SCHORMAIR, B., KORNUM, B. R., FARACO, J., PLAZZI, G., MELBERG, A., CORNELIO, F., URBAN, A. E., PIZZA, F., POLI, F., GRUBERT, F., WIELAND, T., GRAF, E., HALLMAYER, J., STROM, T. M. & MIGNOT, E. 2012. Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. *Hum Mol Genet,* 21**,** 2205-10.

WITTKE-THOMPSON, J. K., PLUZHNIKOV, A. & COX, N. J. 2005. Rational inferences about departures from Hardy-Weinberg equilibrium. *Am J Hum Genet,* 76**,** 967-86.

WOLK, D. A., PRICE, J. C., SAXTON, J. A., SNITZ, B. E., JAMES, J. A., LOPEZ, O. L., AIZENSTEIN, H. J., COHEN, A. D., WEISSFELD, L. A., MATHIS, C. A., KLUNK, W. E. & DE-KOSKY, S. T. 2009. Amyloid imaging in mild cognitive impairment subtypes. *Ann Neurol,* 65**,** 557-68.

WOLTERS, M., HICKSTEIN, M., FLINTERMANN, A., TEWES, U. & HAHN, A. 2005. Cognitive performance in relation to vitamin status in healthy elderly German women-the effect of 6-month multivitamin supplementation. *Prev Med,* 41**,** 253-9.

WONG, C. C., MEABURN, E. L., RONALD, A., PRICE, T. S., JEFFRIES, A. R., SCHALKWYK, L. C., PLOMIN, R. & MILL, J. 2014. Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Mol Psychiatry,* 19**,** 495-503.

WU, H., D'ALESSIO, A. C., ITO, S., WANG, Z. B., CUI, K. R., ZHAO, K. J., SUN, Y. E. & ZHANG, Y. 2011. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes & Development,* 25**,** 679-684.

WU, H., WU, X., SHEN, L. & ZHANG, Y. 2014. Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nat Biotechnol,* 32**,** 1231-40.
WU, J., LIU, S., LIU, G., DOMBKOWSKI, A., ABRAMS, J., MARTIN-TREVINO, R., WICHA, M. S., ETHIER, S. P. & YANG, Z. Q. 2012. Identification and functional analysis of 9p24 amplified genes in human breast cancer. *Oncogene,* 31**,** 333-41.
WU, L. & ZHAO, L. 2016. ApoE2 and Alzheimer's disease: time to take a closer look. *Neural Regen Res,* 11**,** 412-3.
WU, S. C. & ZHANG, Y. 2010. Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol,* 11**,** 607-20.
XIE, B., LIU, Z., LIU, W., JIANG, L., ZHANG, R., CUI, D., ZHANG, Q. & XU, S. 2017a. DNA Methylation and Tag SNPs of the BDNF Gene in Conversion of Amnestic Mild Cognitive Impairment into Alzheimer's Disease: A Cross-Sectional Cohort Study. *J Alzheimers Dis,* 58**,** 263-274.
XIE, B., XU, Y., LIU, Z., LIU, W., JIANG, L., ZHANG, R., CUI, D., ZHANG, Q. & XU, S. 2017b. Elevation of Peripheral BDNF Promoter Methylation Predicts Conversion from Amnestic Mild Cognitive Impairment to Alzheimer's Disease: A 5-Year Longitudinal Study. *J Alzheimers Dis,* 56**,** 391-401.
XIE, W., SCHULTZ, M. D., LISTER, R., HOU, Z., RAJAGOPAL, N., RAY, P., WHITAKER, J. W., TIAN, S., HAWKINS, R. D., LEUNG, D., YANG, H., WANG, T., LEE, A. Y., SWANSON, S. A., ZHANG, J., ZHU, Y., KIM, A., NERY, J. R., URICH, M. A., KUAN, S., YEN, C. A., KLUGMAN, S., YU, P., SUKNUNTHA, K., PROPSON, N. E., CHEN, H., EDSALL, L. E., WAGNER, U., LI, Y., YE, Z., KULKARNI, A., XUAN, Z., CHUNG, W. Y., CHI, N. C., ANTOSIEWICZ-BOURGET, J. E., SLUKVIN, I., STEWART, R., ZHANG, M. Q., WANG, W., THOMSON, J. A., ECKER, J. R. & REN, B. 2013. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell,* 153**,** 1134-48.
XIONG, T., MEISTER, G. E., WORKMAN, R. E., KATO, N. C., SPELLBERG, M. J., TURKER, F., TIMP, W., OSTERMEIER, M. & NOVINA, C. D. 2017. Targeted DNA methylation in human cells using engineered dCas9-methyltransferases. *Sci Rep,* 7**,** 6732.
XU, G., HERZIG, M., ROTREKL, V. & WALTER, C. A. 2008. Base excision repair, aging and health span. *Mech Ageing Dev,* 129**,** 366-82.
XU, Z. & LI, X. 2012. DNA Methylation in Neurodegenerative Disorders. *Curr Transl Geriatr Exp Gerontol Rep,* 1**,** 199-205.
XUE, J., SCHOENROCK, S. A., VALDAR, W., TARANTINO, L. M. & IDERAABDULLAH, F. Y. 2016. Maternal vitamin D depletion alters DNA methylation at imprinted loci in multiple generations. *Clin Epigenetics,* 8**,** 107.
YANG, N. P., LEE, Y. H., LIN, C. H., CHUNG, Y. C., CHEN, W. J. & CHOU, P. 2009. Utilization of and direct expenditure for emergency medical care in Taiwan: a population-based descriptive study. *J Epidemiol,* 19**,** 41-8.
YAO, B., LIN, L., STREET, R. C., ZALEWSKI, Z. A., GALLOWAY, J. N., WU, H., NELSON, D. L. & JIN, P. 2014. Genome-wide alteration of 5-hydroxymethylcytosine in a mouse model of fragile X-associated tremor/ataxia syndrome. *Hum Mol Genet,* 23**,** 1095-107.
YESAVAGE, J. A., BRINK, T. L., ROSE, T. L., LUM, O., HUANG, V., ADEY, M. & LEIRER, V. O. 1982. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res,* 17**,** 37-49.

YILDIRIM, O., LI, R., HUNG, J. H., CHEN, P. B., DONG, X., EE, L. S., WENG, Z., RANDO, O. J. & FAZZIO, T. G. 2011. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell,* 147**,** 1498-510.

YOUNG, J. I., HONG, E. P., CASTLE, J. C., CRESPO-BARRETO, J., BOWMAN, A. B., ROSE, M. F., KANG, D., RICHMAN, R., JOHNSON, J. M., BERGET, S. & ZOGHBI, H. Y. 2005. Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proc Natl Acad Sci U S A,* 102**,** 17551-8.

YU, C. C., FURUKAWA, M., KOBAYASHI, K., SHIKISHIMA, C., CHA, P. C., SESE, J., SUGAWARA, H., IWAMOTO, K., KATO, T., ANDO, J. & TODA, T. 2012a. Genome-wide DNA methylation and gene expression analyses of monozygotic twins discordant for intelligence levels. *PLoS One,* 7**,** e47081.

YU, C. E., CUDABACK, E., FORAKER, J., THOMSON, Z., LEONG, L., LUTZ, F., GILL, J. A., SAXTON, A., KRAEMER, B., NAVAS, P., KEENE, C. D., MONTINE, T. & BEKRIS, L. M. 2013. Epigenetic signature and enhancer activity of the human APOE gene. *Hum Mol Genet,* 22**,** 5036-47.

YU, F., SONG, A., XU, C., SUN, L., LI, J., TANG, L., YU, M., YEATES, T. O., HU, H. & HE, J. 2009. Determining the DUF55-domain structure of human thymocyte nuclear protein 1 from crystals partially twinned by tetartohedry. *Acta Crystallogr D Biol Crystallogr,* 65**,** 212-9.

YU, M., HON, G. C., SZULWACH, K. E., SONG, C. X., ZHANG, L., KIM, A., LI, X., DAI, Q., SHEN, Y., PARK, B., MIN, J. H., JIN, P., REN, B. & HE, C. 2012b. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell,* 149**,** 1368-80.

YUAN, H. F., ZHAO, K., ZANG, Y., LIU, C. Y., HU, Z. Y., WEI, J. J., ZHOU, T., LI, Y. & ZHANG, H. P. 2017. Effect of folate deficiency on promoter methylation and gene expression of Esr1, Cav1, and Elavl1, and its influence on spermatogenesis. *Oncotarget,* 8**,** 24130-24141.

ZALFA, F., GIORGI, M., PRIMERANO, B., MORO, A., DI PENTA, A., REIS, S., OOSTRA, B. & BAGNI, C. 2003. The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses. *Cell,* 112**,** 317-27.

ZAMBONI, G., DE JAGER, C. A., DRAZICH, E., DOUAUD, G., JENKINSON, M., SMITH, A. D., TRACEY, I. & WILCOCK, G. K. 2013. Structural and functional bases of visuospatial associative memory in older adults. *Neurobiol Aging,* 34**,** 961-72.

ZAMMIT, S., ALLEBECK, P., DAVID, A. S., DALMAN, C., HEMMINGSSON, T., LUNDBERG, I. & LEWIS, G. 2004. A longitudinal study of premorbid IQ Score and risk of developing schizophrenia, bipolar disorder, severe depression, and other nonaffective psychoses. *Arch Gen Psychiatry,* 61**,** 354-60.

ZAPPELLA, M., MELONI, I., LONGO, I., CANITANO, R., HAYEK, G., ROSAIA, L., MARI, F. & RENIERI, A. 2003. Study of MECP2 gene in Rett syndrome variants and autistic girls. *Am J Med Genet B Neuropsychiatr Genet,* 119B**,** 102-7.

ZEITELHOFER, M., ADZEMOVIC, M. Z., GOMEZ-CABRERO, D., BERGMAN, P., HOCHMEISTER, S., N'DIAYE, M., PAULSON, A., RUHRMANN, S., ALMGREN, M., TEGNER, J. N., EKSTROM, T. J., GUERREIRO-CACAIS, A. O. & JAGODIC,

M. 2017. Functional genomics analysis of vitamin D effects on CD4+ T cells in vivo in experimental autoimmune encephalomyelitis. *Proc Natl Acad Sci U S A,* 114**,** E1678-E1687.
ZESCHNIGK, M., LICH, C., BUITING, K., DOERFLER, W. & HORSTHEMKE, B. 1997. A single-tube PCR test for the diagnosis of Angelman and Prader-Willi syndrome based on allelic methylation differences at the SNRPN locus. *Eur J Hum Genet,* 5**,** 94-8.
ZHAN, X., HU, Y., LI, B., ABECASIS, G. R. & LIU, D. J. 2016. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics,* 32**,** 1423-6.
ZHANG, C., RODRIGUEZ, C., SPAULDING, J., AW, T. Y. & FENG, J. 2012a. Age-dependent and tissue-related glutathione redox status in a mouse model of Alzheimer's disease. *J Alzheimers Dis,* 28**,** 655-66.
ZHANG, F., WEN, Y. & GUO, X. 2014. CRISPR/Cas9 for genome editing: progress, implications and challenges. *Hum Mol Genet,* 23**,** R40-6.
ZHANG, L., LU, X., LU, J., LIANG, H., DAI, Q., XU, G. L., LUO, C., JIANG, H. & HE, C. 2012b. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat Chem Biol,* 8**,** 328-30.
ZHANG, Z., HUANG, A., ZHANG, A. & ZHOU, C. 2017. HuR promotes breast cancer cell proliferation and survival via binding to CDK3 mRNA. *Biomed Pharmacother,* 91**,** 788-795.
ZHANG, Z. M., LIU, S., LIN, K., LUO, Y., PERRY, J. J., WANG, Y. & SONG, J. 2015. Crystal Structure of Human DNA Methyltransferase 1. *J Mol Biol,* 427**,** 2520-31.
ZHAO, B. S., ROUNDTREE, I. A. & HE, C. 2017. Post-transcriptional gene regulation by mRNA modifications. *Nat Rev Mol Cell Biol,* 18**,** 31-42.
ZHENG, G., DAHL, J. A., NIU, Y., FEDORCSAK, P., HUANG, C. M., LI, C. J., VAGBO, C. B., SHI, Y., WANG, W. L., SONG, S. H., LU, Z., BOSMANS, R. P., DAI, Q., HAO, Y. J., YANG, X., ZHAO, W. M., TONG, W. M., WANG, X. J., BOGDAN, F., FURU, K., FU, Y., JIA, G., ZHAO, X., LIU, J., KROKAN, H. E., KLUNGLAND, A., YANG, Y. G. & HE, C. 2013. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell,* 49**,** 18-29.
ZHENG, H. F., RONG, J. J., LIU, M., HAN, F., ZHANG, X. W., RICHARDS, J. B. & WANG, L. 2015. Performance of genotype imputation for low frequency and rare variants from the 1000 genomes. *PLoS One,* 10**,** e0116487.
ZHENG, M., ZHANG, M., YANG, J., ZHAO, S., QIN, S., CHEN, H., GAO, Y. & HUANG, G. 2014. Relationship between blood levels of methyl donor and folate and mild cognitive impairment in Chinese patients with type 2 diabetes: a case-control study. *J Clin Biochem Nutr,* 54**,** 122-8.
ZHOU, F. C., RESENDIZ, M., LO, C. L. & CHEN, Y. 2016. Cell-Wide DNA De-Methylation and Re-Methylation of Purkinje Neurons in the Developing Cerebellum. *PLoS One,* 11**,** e0162063.
ZHOU, T., XIONG, J., WANG, M., YANG, N., WONG, J., ZHU, B. & XU, R. M. 2014. Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. *Mol Cell,* 54**,** 879-86.
ZHU, H., BHAGATWALA, J., HUANG, Y., POLLOCK, N. K., PARIKH, S., RAED, A., GUTIN, B., HARSHFIELD, G. A. & DONG, Y. 2016. Race/Ethnicity-Specific Association of Vitamin D and Global DNA Methylation: Cross-Sectional and Interventional Findings. *PLoS One,* 11**,** e0152849.

ZHU, T., ROUNDTREE, I. A., WANG, P., WANG, X., WANG, L., SUN, C., TIAN, Y., LI, J., HE, C. & XU, Y. 2014. Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine. *Cell Res,* 24**,** 1493-6.

ZHUBI, A., CHEN, Y., DONG, E., COOK, E. H., GUIDOTTI, A. & GRAYSON, D. R. 2014. Increased binding of MeCP2 to the GAD1 and RELN promoters may be mediated by an enrichment of 5-hmC in autism spectrum disorder (ASD) cerebellum. *Transl Psychiatry,* 4**,** e349.

ZIEGLER, A., VAN STEEN, K. & WELLEK, S. 2011. Investigating Hardy-Weinberg equilibrium in case-control or cohort studies or meta-analysis. *Breast Cancer Res Treat,* 128**,** 197-201.

ZIEMINSKA, E., MATYJA, E., KOZLOWSKA, H., STAFIEJ, A. & LAZAREWICZ, J. W. 2006. Excitotoxic neuronal injury in acute homocysteine neurotoxicity: role of calcium and mitochondrial alterations. *Neurochem Int,* 48**,** 491-7.

ZOGHBI, H. Y. & BEAR, M. F. 2012. Synaptic dysfunction in neurodevelopmental disorders associated with autism and intellectual disabilities. *Cold Spring Harb Perspect Biol,* 4.

ZOU, G. Y. & DONNER, A. 2006. The merits of testing Hardy-Weinberg equilibrium in the analysis of unmatched case-control data: a cautionary note. *Ann Hum Genet,* 70**,** 923-33.

ZUK, O., HECHTER, E., SUNYAEV, S. R. & LANDER, E. S. 2012. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A,* 109**,** 1193-8.

ZULIANI, G., CAVALIERI, M., GALVANI, M., VOLPATO, S., CHERUBINI, A., BANDINELLI, S., CORSI, A. M., LAURETANI, F., GURALNIK, J. M., FELLIN, R. & FERRUCCI, L. 2010. Relationship between low levels of high-density lipoprotein cholesterol and dementia in the elderly. The InChianti study. *J Gerontol A Biol Sci Med Sci,* 65**,** 559-64.

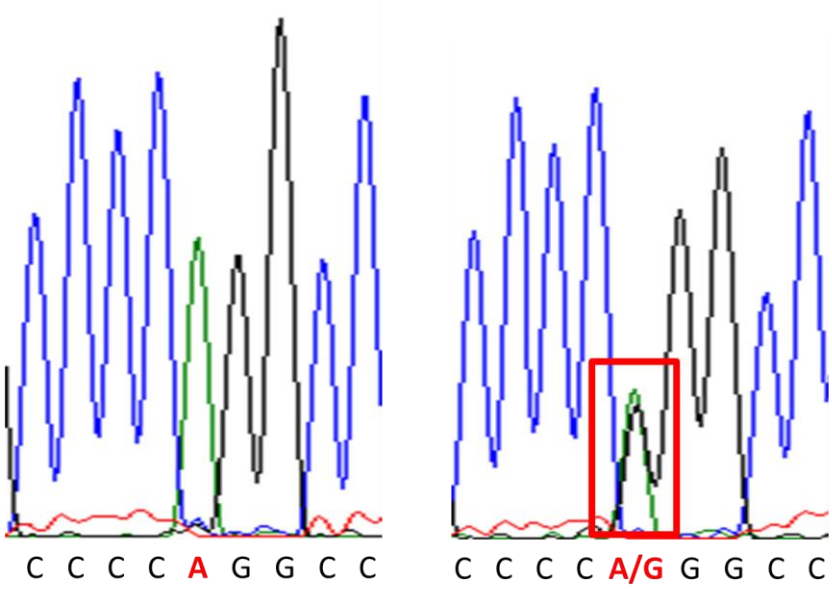
Appendices

Appendix 1. Primers for the *DNMT3L* R278G designed by LGC Genomics

Primers	Sequence	Tm	GC content
Forward	GGGTTGCCTGTCTGTCTCTT	59.30°C	55%
Reverse	GAGCCTGGGGTGTAGACTGA	60.26°C	60%

Tm = temperature

Appendix 2. Example of Sanger sequencing confirmation of the *DNMT3L* R278G variant



Right hand chromatogram shows the dual peaks for a heterozygous carrier of the *DNMT3L* R278G variant.

Appendix 3. PLINK commands to cut BED files

- cd /path/to/PLINK

- ./plink --file /path/to/file.ped and file.map --extract /path/to/myrange.txt --range --out /save/as/file.bed --make-bed
  # Where “myrange.txt” is a text file which contains the chromosome, start base pair position, end base pair position, and identifier for your region of interest

Appendix 4. Discrepant genotyping from the NCDS 1958 UVA and Sanger datasets

ID	UVA	Sanger
KN17453W	A/A	T/C
KN20330R	A/A	T/C
KN21362Y	G/G	T/C
KN21400F	A/A	T/C
KN22764E	A/A	T/C
KN24119Q	G/G	T/T
KN24463N	A/G	T/T
KN25486B	A/A	T/C
KN26109M	A/A	C/C

Appendix 5. Commands for the EGA download client

- java –jar EgaDemoClient.jar –p username password –rf filename –re decryptionkey –label request_filename
  # Where “filename” is the EGAF file name, “decryptionkey” is the unique Identifier you use for secure encryption/decryption.

- java –jar EgaDemoClient.jar –p username password –dr request_filename

- java –jar EgaDemoClient.jar –p username password –dc path/to/file –dck decryptionkey

Appendix 6. VCFtools commands to splice out genetic regions

- cd /path/to/VCFtools

- ./bin/VCFtools --vcf /path/to/whole-genome or whole-exome VCF file --chr x --from-bp a --to-bp b --out output.vcf --recode
  # Where “x” is the chromosome your gene is located on, “a” is the first base position of the region of interest, “b” is the last base position of the region of interest.
  # The “--recode" flag is required to write the region of interest to a new VCF file.

- gzip /path/to/output.vcf

- export PERL5LIB=/path/to/VCFtools/lib/perl5/site_perl/
  # PERL5LIB environment variable must be set to include Vcf.pm module in order to use the VCFtools PERL scripts.

- zcat < /path/to/output.vcf.gz | ./perl/vcf-to-tab > output.tab

Appendix 7. VCFtools commands calculate MAFs

- ./bin/vcftools --vcf /path/to/output.vcf --freq --out output.frq

Appendix 8. Commands to utilise SnpEff annotation

- cd /path/to/SnpEff

- java –Xmx4g –jar snpEff.jar GRCh37.75 /path/to/output.vcf > /save/as/output_ann.vcf

Appendix 9. tabix commands to download specific regions of 1000 Genomes MAF data

tabix –f –h
ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20100804/supporting/AFR.2of4intersection_allele_freq.20100804.genotypes.vcf.gz 21:45662253-45686069 > /Users/msxmf2/Desktop/afr.vcf
“AFR” in the ftp link refers to the African allele frequency. Change this to ASN for Asian and EUR for European.
“21:45662253-45686069” refers to the area of the genome that you require MAF data for.
Running this command will also download the relevant population allele frequency reference file to your tabix folder. You could reference this file in subsequent commands, but it is probably safer to use the ftp link in case of updates.

Appendix 10. VCFtools commands to merge multiple VCF files

VCF files for each gene were first zipped:

- bgzip /path/to/output.vcf

Tabix was then used to create the .tbi files:

- cd /path/to/tabix

- tabix –p vcf output.vcf.gz

VCFtools was used to merge all DNA and RNA methylation genes from each cohort:

- cd /path/to/VCFtools

- export PERL5LIB=/path/to/VCFtools/lib/perl5/site_perl/

- export PATH=$(Aapola et al.):/path/to/tabix
  # Adds the tabix directory to your profile PATH so that it can be called during the vcf-merge command.

- ./perl/vcf-merge output1.vcf.gz output2.vcf.gz… | gzip –c > merge.vcf.gz
  # Where “output1.vcf.gz output2.vcf.gz…” is a list of VCF files for a specific gene from each cohort.

Appendix 11. Commands for Oxford format recoding

- cd /path/to/PLINK

- ./plink --vcf /path/to/merge.vcf --recode oxford --out merge_oxford

Appendix 12. Commands for genetic imputation using IMPUTE2, GTOOL, and VCFtools

- cd /path/to/IMPUTE2

- ./impute2 -m /path/to/1000 Genomes Phase 3 map.txt -h /path/to/1000 Genomes Phase 3 hap.gz -l /path/to/amended legend.gz -g /path/to/merge_oxford.gen -int x y -Ne 20000 -o /save/as/gene.impute2
  # Where “x” and “y” are the start and end chromosomal positions of your merged gene of interest.
  # “Ne” is the effective size of the population. IMPUTE2 suggests using 20,000 for most analyses.

GTOOL can be used to convert the output .impute2 file into PED and MAP format:

- cd /path/to/GTOOL

- ./gtool –G --g /path/to/gene.impute2 --s /path/to/merge_oxford.sample --ped /save/as/gene.ped --map /save/as/gene.map --phenotype phenotype_1 --threshold 0.3
  # Where “phenotype” is the column in the SAMPLE file which needs to be outputted to the PED file.
  # “Threshold” describes the point at which the maximum of the 3 probabilities that make up a binary genotype call must reach in order to be retained (and not annotated as missing). After running multiple imputations with varying thresholds, I settled on 0.3 as it appeared to retain calls in the majority of variants.

The resultant PED and MAP files can then be converted back to a VCF file using VCFtools:

- cd /path/to/PLINK

- ./plink --file /path/to/gene.ped and gene.map --make-bed --out /save/as/bedgene

- ./plink --bfile /path/to/bedgene --recode vcf --out /save/as/imputed_gene.vcf

Appendix 13. BCFtools commands to remove variant positions from VCF files

- bgzip /path/to/imputed_gene.vcf

- cd /path/to/tabix

- tabix –p vcf imputed_gene.vcf.gz

BCFtools could then be used to remove any unwanted variants:

- cd /path/to/BCFtools

- ./bcftools filter -R /path/to/positions.tab /path/to/imputed_gene.vcf.gz | bgzip –c > /save/as/final_gene.vcf.gz
  # Where “positions.tab” is a tab-delimited file containing all of the variant base-pair positions to be removed

Appendix 14. Synergy factor analysis commands for Excel

- x = *OR1* * *OR2*

- SF = *OR3* * x

- a = LN(SF)
  # LN is a function within Excel for calculating the natural logarithm of a number.

- b = SQRT((1/*n1*)+(1/*n2*)+(1/*n3*)+(1/*n4*)+(1/*n5*)+(1/*n6*)+(1/*n7*)+(1/*n8*))
  # SQRT is a function within Excel for calculating the square root of a number.

- Z score = a/b

- *p* value = NORM.S.DIST(Z score, FALSE)

  # NORM.S.DIST is a function within Excel to return the normal distribution, equivalent to a *p* value. FALSE is used if the Z score is negative, TRUE if positive.

Appendix 15. General linear models used to produce predicted regression variables for ROA data

Visuospatial associative memory, ROA, *DNMT3L* R278G

Between subjects effects

Treatment	F = 6.204, *p* = 0.014
Baseline tHcy	F = 1.898, *p* = 0.170
Baseline brain vol	F = 2.459, *p* = 0.119
Baseline creatinine	F = 3.088, *p* = 0.081
PCA1	F = 6.926, *p* = 0.009
G carrier	F = 0.109, *p* = 0.742
PCA1 * G carrier	F = 4.430, *p* = 0.037

Corrected model: R Squared = 0.168 (adjusted = 0.129)

Verbal semantic memory, ROA, *DNMT3L* R278G

Between subjects effects

Treatment	F = 7.069, *p* = 0.009
Baseline tHcy	F = 1.242, *p* = 0.267
Baseline brain vol	F = 0.377, *p* = 0.540
Baseline creatinine	F = 3.711, *p* = 0.056
Age	F = 1.480, *p* = 0.226
PCA2	F = 2.654, *p* = 0.106
G carrier	F = 0.051, *p* = 0.821
PCA2 * G carrier	F = 1.505, *p* = 0.222

Corrected model: R Squared = 0.153 (adjusted = 0.107)

Visuospatial associative memory, ROA, *APOE* E4

Between subjects effects

Treatment	F = 7.139, *p* = 0.008
Baseline brain vol	F = 2.924, *p* = 0.089
Baseline creatinine	F = 5.394, *p* = 0.022
PCA1	F = 6.018, *p* = 0.015
APOE4	F = 0.129, *p* = 0.720
PCA1 * APOE4	F = 0.405, *p* = 0.526

Corrected model: R Squared = 0.150 (adjusted = 0.116)

Verbal semantic memory, ROA, *APOE* E4

Between subjects effects

Treatment	F = 7.772, *p* = 0.006
Baseline tHcy	F = 1.004, *p* = 0.318
Baseline brain vol	F = 0.658, *p* = 0.418
Baseline creatinine	F = 3.631, *p* = 0.059
Age	F = 1.601, *p* = 0.208
PCA2	F = 3.751, *p* = 0.055
APOE4	F = 0.005, *p* = 0.946
PCA2 * APOE4	F = 0.857, *p* = 0.356

Corrected model: R Squared = 0.150, (adjusted = 0.103)

Appendix 16. R commands for iterated Fishers test

- data <- read.csv("/path/to/allele_counts.csv")

- p.val <- apply(data, 1, function(x) fisher.test(matrix(x, nr=2))$p.value)
  # Turns each line of the data variable into a case-control contingency table and returns the p value for each variant.

- OR <- apply(data, 1, function(x) fisher.test(matrix(x, nr=2))$estimate)
  # Turns each line of the data variable into a case-control contingency table and returns the p value for each variant.

- conf <- apply(data, 1, function(x) fisher.test(matrix(x, nr=2))$conf.int
  # Turns each line of the data variable into a case-control contingency table and returns the p value for each variant.

- conf <- t(conf)
  # Transposes the confidence interval data so that it can be exported easily.

The in-built R statistical package "p.adjust" was used to apply the Holm-Bonferroni multiple comparisons correction to the Fisher's test results:

- adj.p.value <- p.value(p.val, "holm")

- out <- cbind(OR, p.val, adj.p.val, conf)

- write.csv(out, file = "path/to/Fishers_test_results.csv")

Appendix 17. R commands to produce Manhattan plots

- library(qqman)

- data <- read.csv("/path/to/manhattan_data.csv")

- missense <- c("rs1", "rs2",…)
  # Where "missense" is a character vector made up of SNP identifiers of interest which will be highlighted on the Manhattan plot; in this case, all missense variants.

- manhattan(data, cex.axis = 0.8, suggestiveline = T, highlight = missense)
  # Where "cex.axis" is the size of the x axis labels.

Appendix 18. UNIX commands to edit components of VCF files

```
sed ‘s/VCFv4\.2/VCFv4.1/’ < /path/to/out.vcf > /path/to/out2.vcf
## Changes the VCF header (format) from v4.2 to 4.1 (needed for PLINK)

tr –d ‘\r’ < file.txt > file2.txt
## Clears CRLF line break formatting from a text file (phenotypic file)

tr ‘\r’ ‘\n’ < file.txt > file2.txt
## Clears CR line break formatting from a text file (phenotypic file)

cat new.vcf | awk ‘BEGIN { OFS = ” \t ” } { if ($1 == “0”) ($1 = “21”); print }’ > new2.vcf
## Changes all “0”s in the chromosome column to “21”
## OFS = “\t” ensures that the output file is tab delimited and thus remains in VCF
format
```

Appendix 19. R commands to recode genotype data

- data <- read.csv(“/path/to/transposed_vcf.csv”)
- data[] <- lapply(data, as.character)
- data[data == “0/0”] <- “0”
- data[data == “0/1”] <- “1”
- data[data == “01-Jan”] <- “2”
- data[data == “./.”] <- “9”
- write.csv(data, file = “/save/as/genotype_data.csv”)

Appendix 20. R commands to run AssotesteR burden analysis

- condata <- data[1:1854]
- IDdata <- data[1855:2144]
- ASDdata <- data[2145:2764
- IDandcon <- rbind(condata, IDdata)
- ASDandcon <- rbind(condata, ASDdata)

With the genotype and phenotype data in place, I then ran the burden and adaptive burden tests in R:

- library(assotesteR)
- g <- as.matrix(read.csv("/path/to/genotype_data.csv")
- p <- as.matrix(read.csv("/path/to/phenotype_data.csv")
- p_edit <- p[,1]
  # May be required to select the first column of the phenotype data if is not properly read into R.
- myCMC = CMC(p_edit, g, maf = 0.01, perm = 100)
  # Runs the CMC test with a rare MAF cut-off and a p value permutation of 100 (the default value within AssotesteR).
- myVT = VT(p_edit, g, maf = 0.01, perm = 100)
  # Runs the VT test with a rare MAF cut-off and a p value permutation of 100 (the default value within AssotesteR).

Appendix 21. R commands to run KBAC burden analysis

- library(KBAC)
- new.dat <- read.csv("/path/to/merged_phenotype_genotype.csv")
- alpha <- 0.05
- num.perm <- 3000
- quiet <- 1
- alternative <- 1
- maf.upper <- 0.01
- kbac.pvalue <- KbacTest(new.dat, alpha, num.perm, maf.upper, alternative)
  # Runs the KBAC test on new.dat.
  # Alpha (in other words, the significance level) and number of permutations are kept at the default. Quiet is set to show the results matrix for the test. Alternative accepts the alternative hypothesis when true.
- print(kbac.pvalue)

Appendix 22. R commands to run SKAT burden analysis

- g <- as.matrix(read.csv("/path/to/genotype_data.csv")

- p <- as.matrix(read.csv("/path/to/phenotype_data.csv")

- g2 <- subset(g, select = -X)
  # May be required to remove the first column, as inputting into R sometimes creates a new first column for row numbers.

- obj <- SKAT_Null_Model(p ~ 1, out_type="D")
  # "SKAT_Null_Model" estimates parameters and obtains residuals under a null model of no association. The SKAT test proper can then be applied.
  # Where "p ~ 1" denotes that no covariate information is being used and "out_type="D"" denotes that a dichotomous phenotype variable is being used as opposed to a continuous variable.

- SKAT(g2, obj)$p.value
  # Runs the SKAT variance-component test for rare variants.

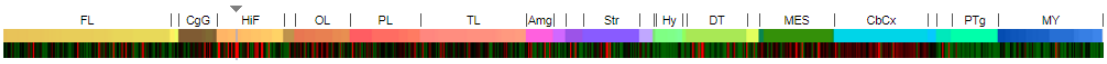
- SKAT_CommonRare(g2, obj)$p.value
  # Runs the SKAT variance-component test to assess the combined effect of common and rare variants.

- SKAT(g2, obj, method = "optimal.adj")$p.value
  # Runs the optimized combined variance-component and burden test for rare
  # variants.

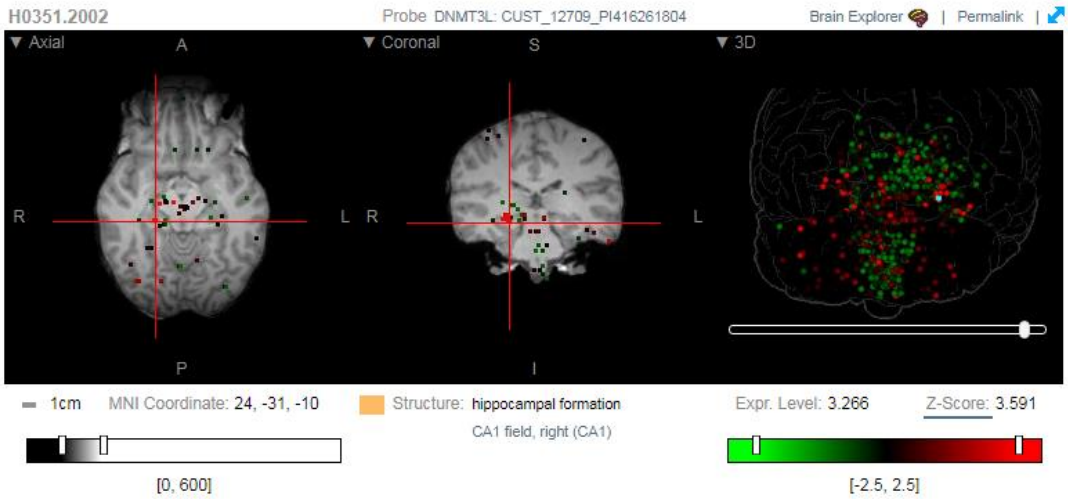
Appendix 23. Materials for immunohistochemistry protocols

Step	Product	Company	Product no.
Dewaxing	Xylene	Fisher Scientific	X/0250/PB17
Dewaxing	Ethanol	Honeywell	458600
Dewaxing	HCl acid	Honeywell	07102
Washes	PBS	Oxoid	BR0014G
Washes	Tween 20	Sigma-Aldrich	X100-100ML
Tissue preparation	PAP pen	Zytomed	ZUC064
Blocking	Donkey serum	Sigma-Aldrich	D9663-10ML
Blocking	Bovine serum	Sigma-Aldrich	F9665-100ML
Blocking	Triton X-100	Fisher Scientific	BP151-100
Antibody amplification	TSA kit	Perkin Elmer	NEL741001KT
Antibody amplification	DMSO	Sigma-Aldrich	276855-100ML

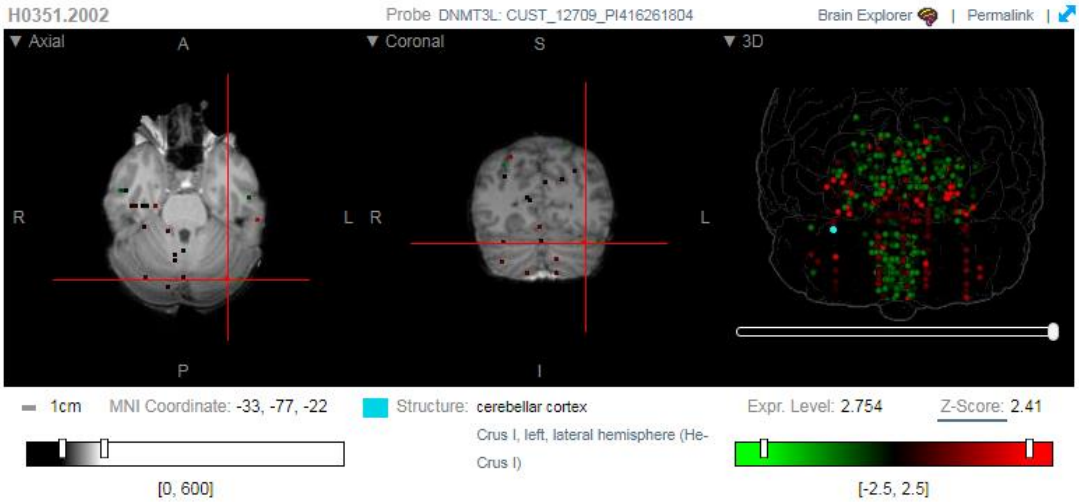
Appendix 24. Allen Brain Atlas Human expression data for *DNMT3L* R278G



Expression profile across brain areas shows multiple peaks of *DNMT3L* expression (red) throughout the brain

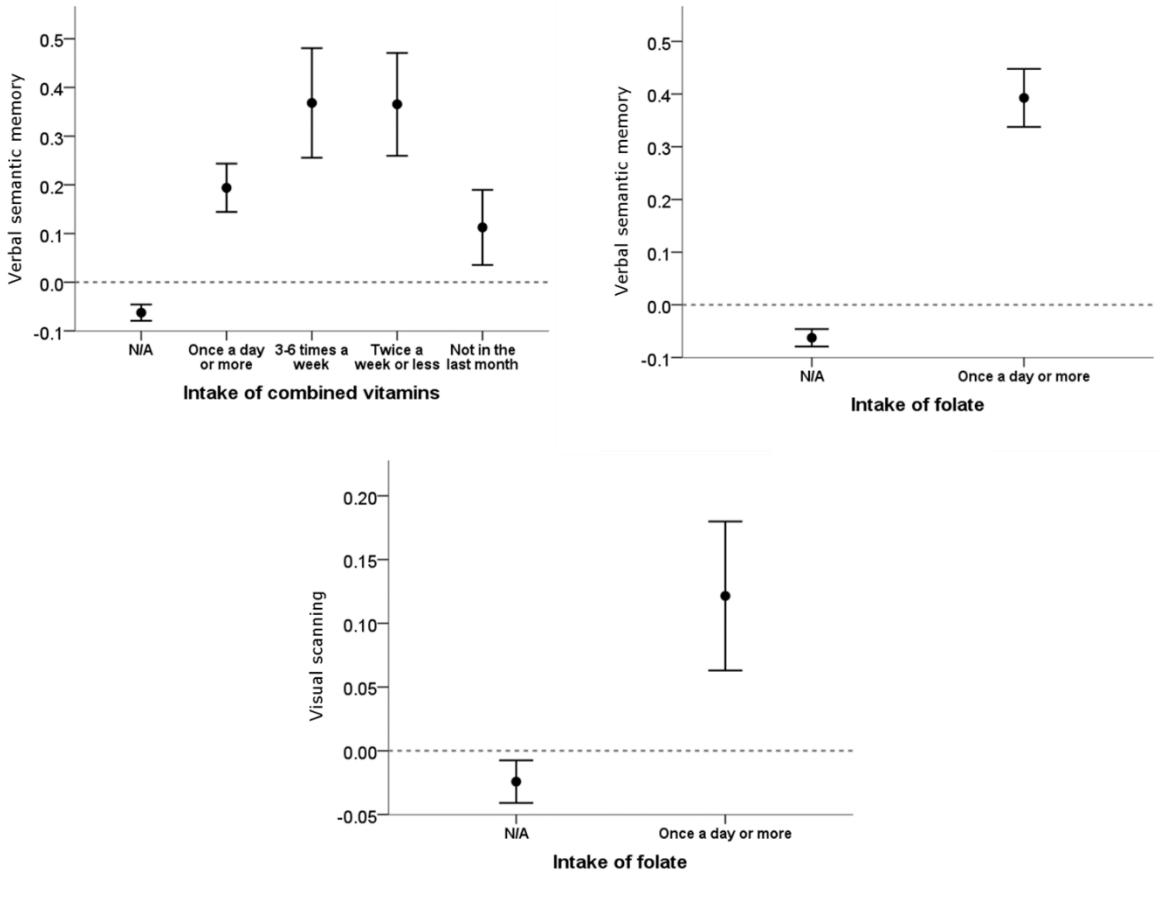


Particularly high levels of *DNMT3L* expression can be seen in the hippocampal formation.



Particularly high levels of *DNMT3L* expression can also be seen in the cerebellar cortex.

Appendix 25. NCDS 1958 individual vitamin results for verbal semantic memory and visual scanning



Regular and irregular combined vitamin intake is associated with better verbal semantic memory performance. Similarly, regular intake of folate is associated with better verbal semantic and visual scanning performance.

Appendix 26. Missense variants with differing MAFs compared to ExAC

Gene	Variant	Known	RareFinds	TwinsUK	Gallagher	Muir	IMGSAC	MGAS	Skuse	ExAC	SIFT	PolyPhen
DNMT1	p.Arg355Pro	rs201945078	0.028226		0.006667	0.054217			0.002994	0.006107	tol (0.28)	pos dama (0.449)
	p.Thr349Met	Novel						0.010309		1.66E-05	tol (0.16)	ben (0.013)
	p.Leu309Val	rs61758430						0.010309		0.000972	tol LC (0.23)	ben (0.014)
	p.His97Arg	rs16999593		0.00027						0.022	tol LC (0.53)	pos dama (0.635)
	p.Arg69His	rs61750053	0.028226		0.013333	0.009036	0.004386	0.005155	0.010479	0.009077	tol LC (1)	ben (0.002)
DNMT3A	p.Val480Gly	rs200099128	0.052419		0.106667	0.033133	0.017544	0.07732	0.109281	0.036	dele (0)	prob dama (0.956)
	p.Glu30Ala	rs143730975		0.001348		0.006024	0.004386	0.015464		0.002916	dele LC(0.03)	ben (0.292)

MAF < 0.01 | MAF < 0.05 | MAF > 0.05 | Benign | Damaging

Gene	Variant	Known	RareFinds	TwinsUK	Gallagher	Muir	IMGSAC	MGAS	Skuse	ExAC	SIFT	PolyPhen
DNMT3B	p.Gly25Arg	Novel	0.004032	0.000809				0.010309	0.001497	0.00075	dele LC (0.01)	prob dama (0.999)
	p.Pro203Leu	rs147945634		0.000539			0.004386	0.010309		0.000593	tol (0.09)	ben (0.125)
	p.Ala384Thr	rs150682895	0.016129	0.012945	0.013333	0.012048	0.013158	0.020619	0.002994	0.007923	tol (1)	ben (0.007)
MBD1	p.Ile545Met	rs183864846	0.004032	0.009709	0.006667	0.012048	0.017544	0.015464	0.005988	0.007796	-	-
	p.Leu437Val	Novel				0.012048				0.000157	dele (0)	pos dama (0.483)
	p.Asp33Gly	Novel	0.012097			0.006024				0.000313	dele (0)	prob dama (0.943)
UHRF2	p.Cys432Phe	rs79300900	0.032258			0.120482				0.005387	dele (0)	prob dama (0.98)
	p.Tyr741Ser	Novel	0.028226			0.021084				0.007191	dele (0.01)	pos dama (0.506)

Gene	Variant	Known	RareFinds	TwinsUK	Gallagher	Muir	IMGSAC	MGAS	Skuse	ExAC	SIFT	PolyPhen
NEIL1	p.Pro131Ala	Novel				0.012048				0.001081	tol (0.13)	ben (0.001)
	p.Asp252Asn	rs5745926					0.004386	0.015464		0.001927	tol (0.28)	pos dama (0.57)
NEIL3	p.Arg38Cys	rs34007209	0.032258	0.019148	0.006667		0.02193	0.06701	0.022455	0.041	tol (0.06)	ben (0.35)
	p.Gln172His	rs17064658	0.012097	0.009978		0.003012	0.013158		0.016467	0.014	tol (0.1)	ben (0.327)
	p.Ile346Val	rs17064676		0.00027					0.001497	0.015	tol (1)	ben (0.004)
WDR76	p.Ala80Thr	rs34623235	0.008065	0.002967	0.006667	0.006024	0.013158	0.005155	0.004491	0.00224	dele LC (0.01)	ben (0.118)
	p.Leu535Phe	rs150494621	0.012097	0.005394	0.013333			0.015464	0.001497	0.004604	dele (0.01)	pos dama (0.635)
THYN1	p.Arg99Pro	Novel				0.01506				0.000404	dele (0)	prob dama (0.996)

Gene	Variant	Known	RareFinds	TwinsUK	Gallagher	Muir	IMGSAC	MGAS	Skuse	ExAC	SIFT	PolyPhen
TET1	p.Val435Ile	rs73276466		0.00027				0.020619	0.002994	0.005	tol (0.6)	ben (0.001)
	p.Gln683Glu	rs139785845	0.012097	0.002157		0.006024			0.005988	0.000881	tol (0.11)	ben (0.146)
	p.Ser701Asn	rs117273115	0.008065	0.009709		0.003012	0.013158	0.005155	0.008982	0.007874	tol (0.05)	ben (0.012)
	p.Pro1879His	rs201665550		0.00027				0.010309		1.65E-05	dele (0.01)	prob dama (0.968)
	p.Thr2042Pro	Novel			0.026667			0.005155	0.010479	0.000461	dele (0.01)	prob dama (0.917)
TET2	p.Pro50Arg	rs12498609	0.008065	0.014024			0.008772	0.025773	0.013473	0.06	dele (0)	pos dama (0.646)
	p.Leu55Phe	rs111948941	0.016129	0.018069	0.013333	0.009036	0.013158	0.005155	0.02994	0.015	dele (0.01)	ben (0.384)
	p.Val239Met	rs6843141	0.020161	0.018069	0.013333	0.01506	0.013158	0.051546	0.032934	0.055	tol (1)	ben (0.001)

Gene	Variant	Known	RareFinds	TwinsUK	Gallagher	Muir	IMGSAC	MGAS	Skuse	ExAC	SIFT	PolyPhen
TET2	p.Gly376Asp	rs61744960	0.060484	0.041532	0.046667	0.027108	0.04386	0.046392	0.053892	0.027	dele (0.01)	prob dama (0.953)
	p.Pro384Leu	rs17253672	0.052419	0.050971	0.06	0.048193	0.052632	0.051546	0.067365	0.048	dele (0.03)	ben (0.118)
	p.Tyr888His	rs144386291	0.020161	0.012945	0.006667	0.012048	0.008772	0.025773	0.01497	0.006894	dele (0)	prob dama (0.99)
	p.Met1722Ile	rs62623390	0.004032	0.006472			0.008772	0.010309	0.002994	0.003439	tol (0.11)	ben (0.015)
	p.Val1739Leu	rs142312318	0.004032	0.009709		0.006024	0.02193		0.016467	0.002662	tol (0.35)	ben (0.006)
	p.Pro1744Ser	rs146348065	0.020161	0.012675	0.006667	0.012048	0.008772	0.025773	0.01497	0.005322	tol (0.31)	ben (0.095)
	p.His1799Arg	rs62621450	0.044355	0.031284	0.02	0.021084	0.02193	0.056701	0.047904	0.045	dele (0.01)	pos dama (0.787)

Gene	Variant	Known	RareFinds	TwinsUK	Gallagher	Muir	IMGSAC	MGAS	Skuse	ExAC	SIFT	PolyPhen
TET3	p.Glu152Gly	Novel	0.004032							0.022	dele (0)	prob dama (0.997)
	p.Pro294Ser	rs61741171	0.028226	0.063107	0.04	0.042169	0.039474	0.061856	0.058383	0.043	tol (0.06)	ben (0.184)
	p.Val455Met	rs72816199		0.002157		0.006024		0.020619	0.001497	0.005319	tol LC (0.07)	ben (0.012)
	p.Arg577Gln	rs57955681		0.00027				0.010309		0.000621	tol (0.18)	ben (0.04)
MBD4	p.Asp568His	rs2307293	0.004032	0.007282	0.006667	0.003012	0.017544		0.005988	0.005387	dele (0.01)	prob dama (1)
	p.Ile358Thr	rs2307298	0.016129	0.011327	0.02	0.003012	0.013158	0.010309	0.016467	0.006952	tol (0.29)	ben (0.006)
	p.Glu346Lys	rs140693	0.024194	0.002157				0.010309	0.008982	0.051	tol (0.55)	ben (0.001)
	p.Ser342Pro	rs2307289	0.008065	0.002967	0.006667			0.020619	0.008982	0.029	tol (0.06)	ben (0.024)

Gene	Variant	Known	RareFinds	TwinsUK	Gallagher	Muir	IMGSAC	MGAS	Skuse	ExAC	SIFT	PolyPhen
METTL3	p.Arg36Trp	rs139460746	0.012097	0.003506	0.02		0.013158	0.010309	0.010479	0.00266	dele (0.02)	ben (0.118)
METTL14	p.Gln171Lys	rs78906144	0.008065		0.006667	0.003012		0.020619	0.019461	0.003921	tol (1)	ben (0.003)
YTHDF2	p.Val221Met	rs61787570	0.004032	0.003776	0.013333	0.009036	0.004386	0.025773	0.004491	0.004269	tol (0.25)	ben (0.318)
HNRNPA2B1	p.Arg350Gly	rs117917826				0.10241				0.014	tol LC (0.52)	ben (0.432)
ELAVL1	p.Ala88Val	rs35986520		0.072816						0.01	-	-
FTO	p.Ala163Thr	rs145884431		0.002157		0.003012		0.010309	0.005988	0.002076	tol (0.54)	ben (0.004)

Appendix 27. Common and low-frequency damaging missense variants

Gene	Variant	Known	RareFinds	TwinsUK	Gallagher	Muir	IMGSAC	MGAS	Skuse	ExAC	SIFT	PolyPhen
DNMT3L	p.Arg278Gly	rs7354779	0.225806	0.259709	0.2	0.210843	0.324561	0.298969	0.270958	0.214	dele (0)	ben (0.427)
NEIL3	p.Pro117Arg	rs7689099	0.125	0.118662	0.093333	0.13253	0.096491	0.113402	0.103293	0.103	dele (0.03)	ben (0.327)
TET1	p.Val128Phe	rs142008363	0.016129	0.020227	0.02	0.021084	0.017544	0.025773	0.010479	0.014	dele (0.01)	pos dama (0.721)
TET1	p.Asp162Gly	rs10823229	0.439516	0.389159	0.34	0.394578	0.346491	0.365979	0.401198	0.329	dele (0.04)	ben (0.219)
TET1	p.Lys1019Glu	rs72799515	0.032258	0.020227	0.033333	0.024096	0.013158	0.015464	0.020958	0.023	dele (0.04)	ben (0.23)
TET2	p.Leu1742Trp	rs34402524	0.125	0.12028	0.133333	0.138554	0.135965	0.097938	0.121257	0.118	dele (0)	pos dama (0.764)

Appendix 28. Presence of significant missense variants in specific ID cohorts

	UHRF2 C432F	*HNRNPA2B1* R350G	*DNMT1* R355P	*THYN1* R99P
Rare FIND	6.45%	0%	5.64%	0.81%
Muir	24.09%	20.48%	10.84%	12.05%

Table indicates the percentage of individuals carrying a minor allele in each cohort.

Appendix 29. Simple burden and adaptive burden results at 0.25 threshold

			Burden	Adaptive burden	
			Rare missense & all common/LF	All rare	Rare missense
DNA	Writers	DNMT1	< 0.001	< 0.001	3.33E-04
		DNMT3A	< 0.001	0.3	0.020327
		DNMT3B	< 0.001	< 0.001	0.006664
		DNMT3L	< 0.001	< 0.001	0.04965
	Readers	MBD1	< 0.001	< 0.001	0.005665
		MBD3	< 0.001	0.04	0.036654
		MeCP2	N/A	N/A	N/A
		UHRF1	N/A	N/A	N/A
		UHRF2	< 0.001	< 0.001	3.33E-04
		TDG	< 0.001	< 0.001	0.009997
		NEIL1	0.21	< 0.001	0.024325
		NEIL3	< 0.001	0.04	0.500167
		WDR76	< 0.001	0.15	0.036988
		THYN1	< 0.001	< 0.001	3.33E-04
	Erasers	TET1	< 0.001	< 0.001	6.66E-04
		TET2	< 0.001	< 0.001	0.067311
		TET3	< 0.001	< 0.001	6.66E-04
		GADD45A	N/A	N/A	N/A
		GADD45B	N/A	0.04	0.044985
		MBD2	0.64	0.02	0.117294
		MBD4	< 0.001	< 0.001	0.002666
		ELP3	< 0.001	0.04	0.01966
RNA	Writers	METTL3	< 0.001	0.41	0.002666
		METTL14	< 0.001	< 0.001	3.33E-04
		WTAP	< 0.001	< 0.001	0.663779
	Readers	YTHDF2	< 0.001	< 0.001	3.33E-04
		YTHDF3	0.17	0.14	0.027991
		HNRNPA2B1	< 0.001	< 0.001	3.33E-04
		ELAVL1	N/A	0.42	N/A
	Erasers	ALKBH5	< 0.001	0.1	0.083972
		FTO	< 0.001	< 0.001	0.043652

Appendix 30. Variance-component and combined case-control results at 0.25 threshold

			Variance component				Combined
			Rare missense	Rare missense & all common/LF	All rare	All variants	Rare missense
DNA	Writers	DNMT1	6.055E-14	1.43943E-05	8.952E-11	2.917E-05	2.66E-13
		DNMT3A	0.1128134	0.002174744	0.0045551	2.43E-51	0.03334479
		DNMT3B	0.0375651	0.002008587	2.402E-05	1.165E-10	0.04700939
		DNMT3L	0.1940454	8.80213E-05	0.100296	5.267E-13	0.2459469
	Readers	MBD1	0.0985374	4.42623E-12	1.578E-10	5.9E-22	0.00100153
		MBD3	0.10618	0.000212669	0.4125494	0.000609	0.02328643
		MeCP2	N/A	N/A	N/A	N/A	N/A
		UHRF1	N/A	N/A	N/A	N/A	N/A
		UHRF2	4.042E-28	6.98712E-26	1.702E-26	6.037E-22	2.83E-27
		TDG	0.0112376	0.000678472	0.0036894	5.228E-48	0.01808069
		NEIL1	0.0611377	0.280242	0.0562318	0.2535955	0.08428117
		NEIL3	0.368139	0.002308186	0.6384164	0.0009199	0.5292834
		WDR76	0.1410143	0.28931	0.1320259	2.412E-10	0.1079055
		THYN1	9.546E-09	3.55678E-08	5.227E-08	7.784E-06	1.17E-09
		TET1	0.0014023	0.01231756	1.764E-05	1.151E-05	5.17E-04
		TET2	0.4970431	8.1397E-07	1.776E-06	9.845E-09	0.369823
		TET3	0.0248712	2.6297E-10	0.0087251	5.157E-13	9.25E-04
	Erasers	GADD45A	N/A	N/A	N/A	N/A	N/A
		GADD45B	0.0711278	N/A	0.1999522	1.06E-111	0.03650229
		MBD2	0.0603732	0.1806395	0.3276901	0.0004747	0.08760751
		MBD4	0.0060329	0.000187064	0.0002825	1.837E-05	0.00410883
		ELP3	0.0319539	0.202766	0.0175407	2.532E-10	0.05899665
RNA	Writers	METTL3	0.0026758	0.000491448	0.0029425	1.369E-89	0.00133072
		METTL14	2.427E-11	9.29506E-07	2.49E-07	1.182E-06	1.70E-10
		WTAP	0.9182705	0.2441599	0.2501299	0.0771827	1
	Readers	YTHDF2	0.0119986	0.0459544	0.0215383	6.789E-40	4.10E-05
		YTHDF3	0.0436048	0.5482764	0.0842767	8.861E-44	0.03712862
		HNRNPA2B1	5.319E-17	8.68148E-12	1.961E-13	7.899E-08	1.38E-16
		ELAVL1	N/A	N/A	0.2612195	2.606E-15	N/A
	Erasers	ALKBH5	0.3050423	4.91919E-05	0.6208074	7.573E-06	0.131406
		FTO	0.1053368	0.1106032	0.0019553	0.006241	0.1793618

Suggestive significance

GWAS significance