A profile of differential DNA methylation in sporadic human prion disease blood: precedent, implications and clinical promise

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Neuroscience from University College London

Luke Child Dabin

MRC Prion Unit at UCL

Institute of Prion Diseases

University College London

2019

Declaration

I, Luke Child Dabin, attest that the work presented in this thesis is my own. Where information has been derived from others it has been indicated in this thesis.

Dedication

This thesis is dedicated to my family, through whose principles, encouragement and faith in me I have managed to achieve what I have achieved thus far. But it is also dedicated to Badass Cat, who after over a decade of friendship, multiple surgeries, two years living feral and subsisting on local farmers' livestock, having being hit by two cars consecutively in the space of seconds and returning a day later unharmed yet coated in something else's blood, recovering from an abscess the vet deemed inoperable on the day of a scheduled euthanasia, and repeatedly arriving home totally unflustered despite having been caught on barbed wire, finally passed from this mortal coil in 2018. Rest in peace, you magnificent bastard.

Acknowledgements

I have been tremendously fortunate to receive supervision, advice and support from my two supervisors, Dr. Emmanuelle Viré and Professor Simon Mead. Having arrived at UCL after two years in Tokyo during which I spent most of my time out of the lab and away from the literature (and with a degree in Biochemistry which focused mostly on protein biology rather than genetics, let alone epigenetics), without their tuition and patience I would not have been able to achieve what I have over the last five years. They enabled me to begin, continue and complete this work, and I cannot overstate the importance of their tutorage. It has been a true privilege.

There are many, many scientists whose advice and collaboration have been incredibly helpful over the course of this work. Dr Ankur Chakravarty, Dr Holger Hummerich and Dr. Tyler Gorrie-Stone provided me with advice and support as I learned how to use R, a statistical tool which has and will continue to prove essential to my work. Professor Stephan Beck gave me his time to discuss the technical aspects of epigenetic data and how to thoroughly analyse it, which given his authority within the field was priceless.

Professors Jonathan Mill and Katie Lunnon kindly welcomed me to their laboratory for two weeks to learn pyrosequencing, and Dr. Adam Smith and Dr. Rebecca Smith lent me their valuable time to train and advise me on proper experimental design. This technique proved essential for much of this work and I am glad to count them as friends.

Dr. Peter Sarkies, Dr. Tony Bélicard and Sheeba Singh welcomed me to their laboratory for pyrosequencing and were tremendously helpful. It is no exaggeration to say that at least half of this PhD rests on their generosity, and I cannot overstate my thanks.

Dr. Mina Ryten and Professor Sebastian Brandner examined me for my MPhil-PhD upgrade. They were interested in my work, asked difficult questions and raised points for further consideration, and I am grateful for their time and advice. Professor Brandner and Dr. Zane Jaunmuktane also helped plan further work which will lead to an investigation of brain-derived DNA. Dr. Jaunmuktane spent many hours with myself and Dr. Viré in the containment level 3 laboratory dissecting prion-infected brain tissue, which considering her incredibly busy schedule was and is tremendously appreciated.

The National Prion Clinic staff have been enormously supportive, answering my questions about the clinical aspects of prion disease and providing clinical data for a range of interesting analyses. For the latter I am very grateful to Joanna Field, Irina Neves Simoes and Rachel Knight who helped me identify patient samples and connect them with clinical metrics. Dr. Akin Nihat, Dr. Hans Odd, Professor Peter Rudge and Dr. Tze-How Mok collected patient and control blood samples for me which are, obviously, beyond value. I am also very grateful to Rachel Raybould at Cardiff and Vale Hospital who provided control DNA samples, which were essential for this work.

As well as the National Prion Clinic, I worked closely with the Genetics, Epigenetics and Bioinformatics (née Human Genetics) group of the MRC Institute of Prion Disease at UCL and as a result am indebted to several group members.

Fernando Guntoro is one of the newest members of the group and will be continuing the group's DNA methylation project. As well as being a genial and enthusiastic scientist, his advice regarding machine learning was incredibly helpful during the final months of my PhD.

James Uphill has recently left the group, but his knowledge of patient sample locations, his diligence and his humour were much appreciated during the first half of my PhD. I have very fond memories of James' antics, and his presence at Wednesday group meetings is missed.

Gary Adamson left the group relatively early on in the course of my studies, but was cheerful, helpful and unfailingly "got" my jokes. He, James and Ron had just finished cataloguing the group's inventory of over 50,000 DNA samples as I began my PhD. Without this resource my work would have been next to impossible, and Gary's sense of order and attention to detail doubtlessly contributed immensely to this titanic effort.

Penny Norsworth's work and my own have not overlapped extensively, but in the face of low morale and fatigue her cheerfulness, supply of Dilbert comics and daily coffee shop runs have fuelled no small part of this manuscript.

Ron Druyeh, master of the Ultra-Low Temperature freezers, has also moved on from the group. My work with frozen blood, serum and DNA samples stored in off-site containment level 3 laboratories would have been impossible without Ron, and his diligence and work ethic were inspirational.

Tracy Campbell has been tirelessly supportive when it comes to identifying samples and connected data. Above and beyond this, however, were her advice and patience during the early stages of my PhD as I began to extract DNA from prion patient blood. Her years of experience working in the group have benefited me when I had questions about rare aspects of the disease, and it's been a pleasure to spot her when extracting patient DNA or setting up sequencing reactions.

Andrew Thompson shares our office space and is a warm and thoughtful colleague. His clinical insight into facets of prion disease pathology and treatment has emerged in his questions and comments when I presented my work to the department and frequently led to invaluable discussions and new ideas.

I have shared space with many people and cannot thank them all, but Amir Mayahi, Dr. Azadeh Khalili-Shirazi, David Ruegg, Professor Elizabeth Fisher, Dr. Frances Wiseman, Jackie Linehan, Dr. Jonathan Wadsworth, Kevin Williams, Dr. Laszlo Hosszu, Dr. Malin Sandberg, Mark Batchelor, Nunu Anora, Professor Parmjit Jat, Samantha McCleod, Dr. Sylvia Purro, Tamsin Nazari and Ryan Peter have all helped me on many occasions during my PhD, been friendly and approachable, and exemplify the scientific ideal of very nice but clever people working together to solve hard problems. I'll miss them all.

I must also thank other PhD students for moral support but also mentorship and advice. The foremost of these is Carolin Cyran who is a fellow Wolfson PhD student. Carolin has been a pleasure to share space, time, food and conversation with, and has been a patient soundboard for all my questions (of which there have been many) no matter how stupid (as previous). I am indebted to her and am delighted to have been able to celebrate her many successes over the previous years, which have been inspiring.

I must thank several PhD students who have passed through or remained in our group, namely Emma Jones, François Kroll, Ines Whitworth, Julia Ravey and Thanos Dimitriadis. It has been exciting to meet you, share ideas and experiences and also pints in the King and Queen, and I wish you all the very best in your futures.

There is a plethora of other students and friends without whom quite a difficult journey would have been more painful. I owe particular thanks to Alan Mejia-Maza, Dr. Alexandra Philiastides, Bernie Simone, Caroline Casey, Claudia Cannovò, Dr. David Thomas, Francesca De Giorgio, Grace O'Reagan, Heather Whittaker, Dr.

Justin Tosh, Madeleine Reilly, Nora Thoeng, Dr. Prasanth Sivakumar, Dr. Ruchi Kumari Kaushik and Dr. Xun Choong. I wish to highlight Dr. Laura Pulford for her enthusiastic support and statistical advice, who was a shining star in the Institute of Neurology student office, and who died towards the end of her PhD. I deeply miss her.

I have friends who have supported me through tough times during my PhD who deserve gratitude. Caitlin Greenwood is a PhD student at Bristol with whom I have shared the woes and anxieties of the PhD experience, and who has unfailingly given me sound advice on handling problems and encouraged me to consider my health as a priority. Dr. Jonny Scaramanga has been a touchstone for PhD-related concerns and provided thoughtful comments on balancing the weights of qualitative and quantitative data. Melissa Liau has been a tremendously supportive and patient partner and has helped keep me sane when obstacles seem insurmountable. Dr. James Bonner and Dr. Karen Frudd have inspired me throughout the course of their PhD and provided much needed friendship and sympathy throughout my own studies.

I am indebted to the Leonard Wolfson Studentship Committee for funding this PhD, but also in particular to Professors Nick Fox and Nick Wood for guidance and support. Elizabeth Halton, as the coordinator for the studentship, has provided her own guidance and support and has expertly organised resources and workshops for Wolfson students such as myself. It seems strange to look back on that skype interview from Wakoshi at 7pm local time, but I was and am delighted to have been selected as a candidate and am eternally grateful for this opportunity. I must also thank Professor John Collinge, Director of the Institute for Prion Diseases, for his support and comments over the past 5 years, and the seminal work of his and others which established the Institute and drew together the talented individuals I have thanked above.

Finally, I must thank my family, which is large and wonderful, but particularly my immediate family. My parents, John and Philippa, and my sister and brother-in-law, Harriet and Matthew, as well as my recently departed and much beloved grandmother, Narin, who delighted in hearing about my studies but always insisted I should find time to cure her backache. Also, my cat. Although my dedication to Badass Cat (née Tibbles) may seem flippant, his companionship helped me combat severe depression in my late teens and he was a firm and faithful friend, who in my

absence also provided companionship to my sister and my reluctantly accepting mother and father. My whole family's encouragement, support and understanding has been truly heroic over the previous years. My only regret is that, although doing my best to balance work and life, I could not have visited more often. I hope you can take solace in the knowledge that this document will make you quite a fetching doorstop.

I would like to conclude by saying that I have struggled with depression and anxiety over the course of my studies, and that many people mentioned above have been supportive and incredibly helpful in the face of this. While mental health issues are recognised as being hugely enriched in PhD students, I do not feel any candidate can anticipate how a doctoral degree can challenge them. I am eternally grateful for the support my family, my friends, my colleagues and my supervisors have provided me with, and am delighted to finally submit this thesis.

Abstract

Sporadic Creutzfeldt-Jakob Disease (sCJD) is a rare but devastating neurodegenerative disorder characterised by misfolding, propagation and deposition of the prion protein in the brain, leading to neuronal death and rapid cognitive and functional decline. As there is no obvious genetic cause of sCJD, the epigenetic status of sCJD patients may clarify spontaneous prion disease aetiology or reveal biomarkers of the disease. Blood from patients was profiled to document genome-wide differential DNA methylation.

38 loci were identified as being differentially methylated in sCJD blood, including two which associated with disease severity as measured by the MRC Scale score. Of 7 loci considered for replication, 5 showed similar effects in a second cohort of patients, but not in patients of Alzheimer's disease, iatrogenic CJD, or inherited prion disease, suggesting these effects are specific to the sporadic form of CJD. Notably hypomethylation at a site in the promoter of *AIM2*, an inflammasome component, retained its association with disease severity.

Hypomethylation of *FKBP5*, a gene known to regulate the cellular response to cortisol, prompted further investigation which revealed that circulating cortisol is indeed elevated in sCJD patients. Profiling of frontal cortex-derived DNA showed that differential methylation observed in blood is absent from the brain methylome.

Machine learning classification of sCJD based on genome-wide methylation data was able to classify sCJD and healthy control status with an accuracy of 87.04%. This is an appreciable level of accuracy but importantly sets precedence for further classification of prion patients in more complex clinical and research settings, as well as assisting differential diagnosis of less conventional rapid dementias.

Impact Statement

Sporadic prion disease is a rare but devastating neurodegenerative disease, with 136 mortalities in the UK during 2018. This study of DNA methylation in the blood of these patients has identified one gene (*AIM2*) which appears to become progressively demethylated as the disease progresses, making it a promising biomarker for patient monitoring, particularly during therapeutic intervention.

Another gene (*FKBP5*) is also demethylated in sporadic prion disease and has established connections to psychiatric disorders and a system which communicated between the brain and the periphery, namely the hypothalamic-pituitary-adrenal axis. It possible that this system is dysregulated early in the disease process as there is no correlation with clinically presentations of disease severity. *FKBP5* has also been implicated in a number of other dementias and psychiatric disorders yet its involvement in neurodegeneration remains unclear. Further work showed that cortisol, an anti-inflammatory agent, is elevated in sporadic prion disease patient serum. This finding strengthens the case for deeper investigation into the role of *FKBP5*, a known suppressant of cortisol action, and peripheral inflammation in neurodegenerative diseases.

Machine learning classification models using our DNA methylation dataset was shown to have promising accuracy (87.04%) and could be further employed in analysis of prion disease subgroups, particularly in subclinical carriers.

At a Departmental scale, this work establishes an analytical pipeline for analysis of methylation data at the Genetics, Epigenetics and Bioinformatics group of the MRC Prion Unit at UCL, affiliated with the National Prion Clinic at UCLH NHS Foundation Trust. A blood-derived DNA extraction protocol was also optimised to complement the current Neurogenetics extraction protocol, using 1/40th of the volume yet yielding sufficient blood-derived DNA for methylation analysis. This work represents the foreground of a more detailed examination of prion and rapid dementia patients in the context of DNA methylation, which may prove clinically useful.

Table of Contents

List of Figures

List of Tables

List of Boxes

1 Introduction

Prion diseases are transmissible neurodegenerative disorders with unique aetiologies and pathologies, which affect a broad range of mammalian species. The first prion disease to be identified was scrapie, which has been endemic in parts of Europe since the 18th century (Plummer, 1946). Scrapie was known to be an infectious disease, and early experiments by Cuillé and Chelle not only demonstrated that healthy sheep could be infected through contaminants present in the paddocks previously occupied by scrapie sheep, but that the disease could also be experimentally transmitted into goats (Cuillé and Chelle, 1936). This lead to the hypothesis that scrapie was caused by a slow-acting virus, as sheep developed symptoms relatively slowly as per several other slow-acting viruses which affect the nervous system (Ter Meulen and Hall, 1978). However unlike other viruses a "scrapie agent" could not be isolated, nor cultured, nor inactivated by heat or treatment with formalin in the same way viruses could be (Parry, 1962).

In 1957 it was reported that tribespeople in the Fore region of Papua New Guinea were weathering an epidemic of a neurodegenerative disease they called Kuru, meaning "to shiver" or "to be afraid" (Gajdusek and Zigas, 1957). The disease was observed to present with ataxia and a mild tremor, followed by progressive loss of motor and cognitive functions. Brain homogenate extracted from those who died from Kuru was shown to induce similar pathology in chimpanzees if inoculated intracerebrally (Beck *et al.*, 1966), prompting a separate experiment where inoculation of homogenate from a Creutzfeldt-Jakob Disease (CJD) patient was also shown to induce neuropathology. CJD had been first described in the 1920s (Creutzfeldt, 1920) yet was still acknowledged as "an ill-defined term for a group of subacute presenile encephalopathies characterized clinically by dementia, involuntary movements (myoclonic jerks), and other less constant findings that often include ataxia" (Gibbs Jr. *et al.*, 1968). Indeed, the condition's second namesake, Alfons Jakob, is thought by some to have identified *bona fide* CJD as Hans Creutzfeldt's descriptions do not overlap with current understanding of prion disease pathology (Katscher, 1998). Transmission of prion disease pathology into chimpanzees using CJD patient brain homogenate by Gajdusek established the cause of CJD as transmissible and thus analogous to the scrapie agent.

In 1981 Prusiner *et al.* reported the purification of a hydrophobic protein component from the scrapie agent (Prusiner *et al.*, 1981). The following year Prusiner published a manuscript which considered the agent in detail and employed for the first time

the descriptor “prion”, a contraction of “proteinaceous infectious particle”, to distinguish the agent from viruses and other infectious nucleotide-based particles (Prusiner, 1982). Prions were subsequently purified from CJD patient brains and shown to contain epitopes present in the scrapie prion preparations being used in laboratories (Bockman, JM. , Kingsbury, DT., McKinley, MP., Bendheim, PE., Prusiner, 1985). To this day definitive diagnosis of CJD is only possible through detection of prion protein aggregates in brain tissue by post-mortem histological examination.

Creutzfeldt-Jakob disease is now understood to result from template-driven misfolding of the human prion protein (Griffith, 1967; Prusiner, 1991), from which follows a rapidly progressive dementia with co-existent signs such as myoclonus.

1.1 The Prion gene and its protein

Human *PRNP* is located on chromosome 20 and encodes a 253-residue protein (PrP) in two exons, of which only the second is translated. The product’s primary sequence is highly conserved across mammals (Prusiner and Scott, 1997). The mature protein is 209 amino acids long, contains an internal disulphide bond and two sites that can be variably glycosylated, and is membrane anchored (Prusiner and Stahl, 1991). PrP has a modular structure which is displayed in Figure 1. Normal genotype at codon 129 of PrP can be homozygous or heterozygous for methionine or valine and is a major disease modifying factor. Its effects on disease will be discussed in later sections of this thesis.

The relationship between PrP function and dysfunction remains unclear. While there is evidence to suggest it has multiple neurotrophic roles, *Prnp*$^{-/-}$ mice are phenotypically quite normal with some minor alterations to neuronal and immune function emerging with age (Zomosa-Signoret *et al.*, 2008). It is thought this is due to PrP’s role in maintaining axonal myelin, and decreased proteolysis or loss of glycolipid membrane anchoring produces similar effects (Bremer *et al.*, 2010). PrP has been shown to increase neurogenesis in the dentate gyrus of the hippocampus in a dose-dependent manner, yet as of itself this increased neural proliferation does not result in a greater final volume of mature neurons (Steele *et al.*, 2006). Recombinant PrP has been shown to induce synaptogenesis *in vitro*, while native PrP has been identified as a receptor for amyloid-β through which synaptic function and long-term potentiation are inhibited (Laurén *et al.*, 2009). PrP coordinates common divalent metal cations in various regions of the brain (Pushie *et al.*, 2011), in particular Cu^{2+} which induces morphological changes in full length PrP structure

towards a more beta-sheet based conformation (Thakur *et al.*, 2011) and promotes its own clathrin-mediated internalisation (Zomosa-Signoret *et al.*, 2008). The normal physiological functions of PrP have been reviewed extensively (Zomosa-Signoret *et al.*, 2008).

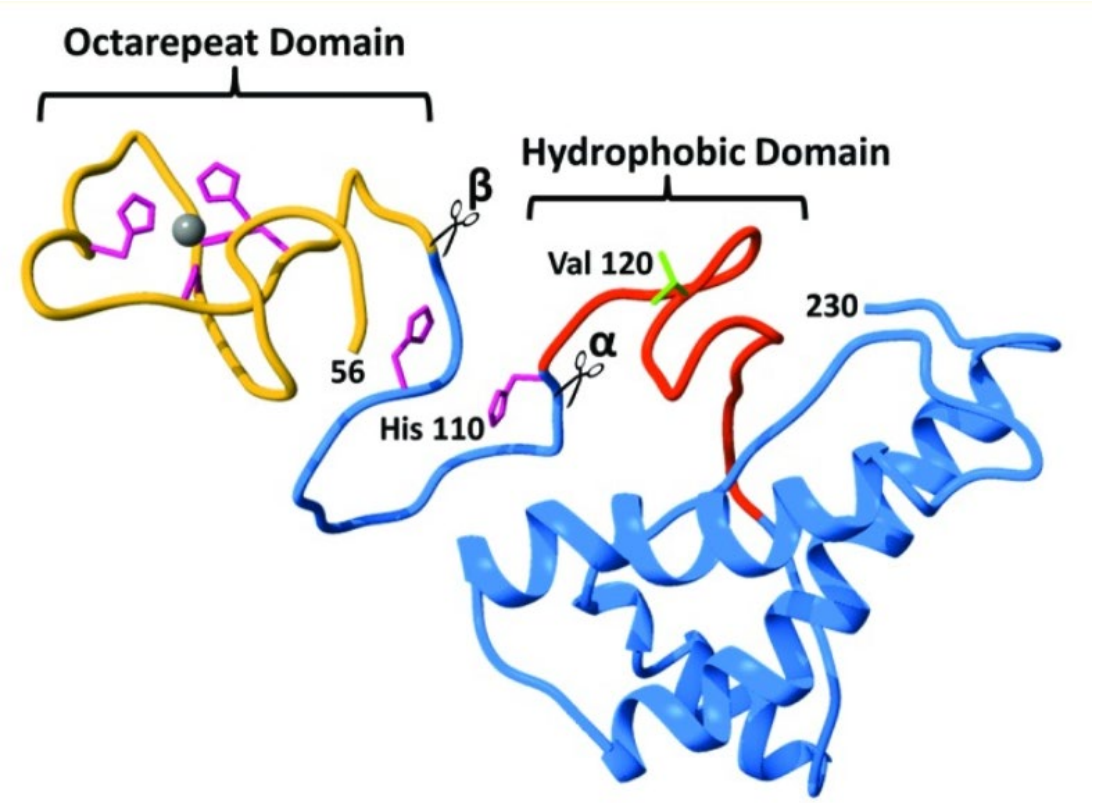


Figure 1: Tertiary structure of PrP. Histidine residues (magenta) in the N-terminal octarepeat domain coordinate Zn2+ and Cu2+ cations. N-teminal and C-terminal fragments can be produced through α or β-cleavage. Figure from McDonald and Millhauser, 2014.

The native structure of PrP is commonly termed PrP^{C} to distinguish it from the misfolded, aggregated form associated with CJD (Prusiner, 1991), which is termed PrP^{Sc} (from PrP-Scrapie). Thinking beyond the supposedly necessary characteristics of an infectious agent, this being a nucleic acid component, Griffith hypothesised that the scrapie agent could derive its infectious characteristic through molecular templating (Griffith, 1967). This proposal was vindicated through studies of *Prnp*$^{-/-}$ mice which demonstrated that native PrP^{C} is necessary for infection by PrP^{Sc}-containing inocula, which has led strength to an early model of prion disease

whereby PrP^{Sc}, arising stochastically, self-templates using PrP^{C} as a substrate. Further work has provided evidence that the label PrP^{Sc} is reductionist and perhaps unhelpful, due in part to the existence of various strains of prion particles with distinct conformations (Safar *et al.*, 1998) and phenotypic effects (Prusiner, 1991). But most strikingly recent findings suggest that the presence of PrP^{Sc} aggregates does not necessarily correlate with clinical presentation, nor spatially with neuronal loss (Sandberg *et al.*, 2014; Alibhai *et al.*, 2016). As such expansion of the "protein only" model to account for concepts such as prion strains is ongoing, with labels such as PrP^{I} and PrP^{L} becoming more common when referring to infectious and toxic (lethal) moieties. Current models of PrP^{I} suggest its structure is fibrillar in nature, and fragmentation of infectious prion fibrils into smaller assemblies through sonication has been shown to almost double a sample's infectious titre (Terry *et al.*, 2016; Vázquez-Fernández *et al.*, 2016). An alternative hypothesis, the "virino" model, postulates that PrP^{Sc} is in fact associated with a nucleotide species and that the two are necessary to infect cells and cause pathology (Kimberlin, 1982). This model also accounts for strain variation which may be influenced by specific sequences of oligonucleotides but does not explain subclinical prion infection. To date no nucleotide fraction has been successfully purified from infectious prion isolates and the model has been largely disregarded.

1.2 Creutzfeldt-Jakob Disease

CJD is the most common human prion disease and can be separated into three varieties based on aetiology: sporadic, familial and acquired. Sporadic CJD (sCJD) accounts for 85-90% of cases (Zanusso *et al.*, 2016) whilst a number of dominant alleles of the prion protein gene (*PRNP*) are known to cause inherited prion disease (IPD). Other cases can be acquired, for example through consumption of prion-contaminated food which can cause variant CJD (vCJD).

A second form of acquired prion disease, iatrogenic CJD (iCJD), mostly arises from contaminated cadaveric pituitary-derived human growth hormone inocula and contaminated dura mater transplants. The former was recognised in the UK in 1985 (Powell-Jackson, Kennedy and Whitcombe, 1985) and in total 77 British growth hormone recipients (out of 1849 total) who had all received the same preparation of growth hormone have developed iCJD. Subsequent genotyping has revealed that a polymorphism at codon 129 in the prion protein modifies the incubation period before disease onset, where possession of the valine allele has been associated with earlier disease onset (Collinge, Palmer and Dryden, 1991; Rudge *et al.*, 2015).

196 dura mater transplant recipients have developed iCJD, with the majority (123) of cases occurring in Japan where dura mater repair after neurosurgery was a common practice (Brown *et al.*, 2006). A specific brand of dura mater – LyoDura – was used in all instances where infection has been observed (Yamada *et al.*, 2009). Cases of iCJD have also been linked to neurosurgery, corneal transplant and blood transfusion from a donor carrying vCJD prions, although improved patient screening and instrument decontamination has since been effective in reducing exposure of patients to prion contaminants.

Variant CJD (vCJD) emerged in the United Kingdom in 1996 (Will *et al.*, 1996) and was recognised as distinct from "classical", sporadic CJD on account of a much younger age of onset, longer disease duration and neuropathology involving deposition of kuru-like "florid" plaques of prion protein (PrP). The first 10 victims of vCJD were identified by the CJD Surveillance Unit which was instituted in 1990 to monitor the effects of a bovine spongiform encephalopathy (BSE) epidemic in the UK and potential transmission of BSE to humans. vCJD was subsequently characterised through molecular strain analysis of PrP and murine bioassays as a strain of BSE that had indeed infected humans through the food supply (Collinge *et al.*, 1996; Bruce *et al.*, 1997; Hill *et al.*, 1997). To date, 228 cases of vCJD have been documented, 175 of which occurred in the UK (The University of Edinburgh, 2017). With one exception (Mok *et al.*, 2017), all individuals were homozygous for methionine at codon 129 of *PRNP*.

Inherited Prion Disease (IPD) was first documented in the Backer family in 1920s Germany: 70 years later a D178N mutation in *PRNP* (Kretzschmar, Neumann and Stavrou, 1995) was identified in DNA extracted from the preserved brain tissue of one of the family members. This mutation and several others (see Figure 2), as well as octapeptide repeat insertions, are now known to clinicians and used to diagnose IPD. As with iCJD, codon 129 genotype of the *PRNP* allele which encodes the mutation is known to modify some forms of inherited prion disease (Gambetti *et al.*, 2003), notably D178N which causes Fatal Familial Insomnia if the mutant *PRNP* allele encodes 129M.

Sporadic CJD (sCJD) can be distinguished from other rapidly progressive dementias by observing increased signal on T2 weighted and diffusion weighted MRI sequences in the striatum, thalamus or cortex, increased 14-3-3 protein in cerebrospinal fluid, and periodic sharp wave complexes in electroencephalograms (Zanusso *et al.*, 2016). However, a definite diagnosis according to WHO criteria can

only be made through neuropathological observation of aggregated PrP^{Sc} deposition in the brain. Subsequently there are cases where diseases such as rapidly progressive Alzheimer's Disease, Lewy Body Dementia, and viral encephalitis have been mistaken for CJD. It is also highly likely that cases of sCJD have been masked in patients with existing dementia diagnoses. As shown in Figure 3, deaths per million per year in the UK have risen from 0.5 in 1988 to 1.8 in 2017 (The National CJD Research and Surveillance Unit, 2017), possibly due to improved diagnostics and awareness but perhaps also due to a shift in median population age or other unknown factors.

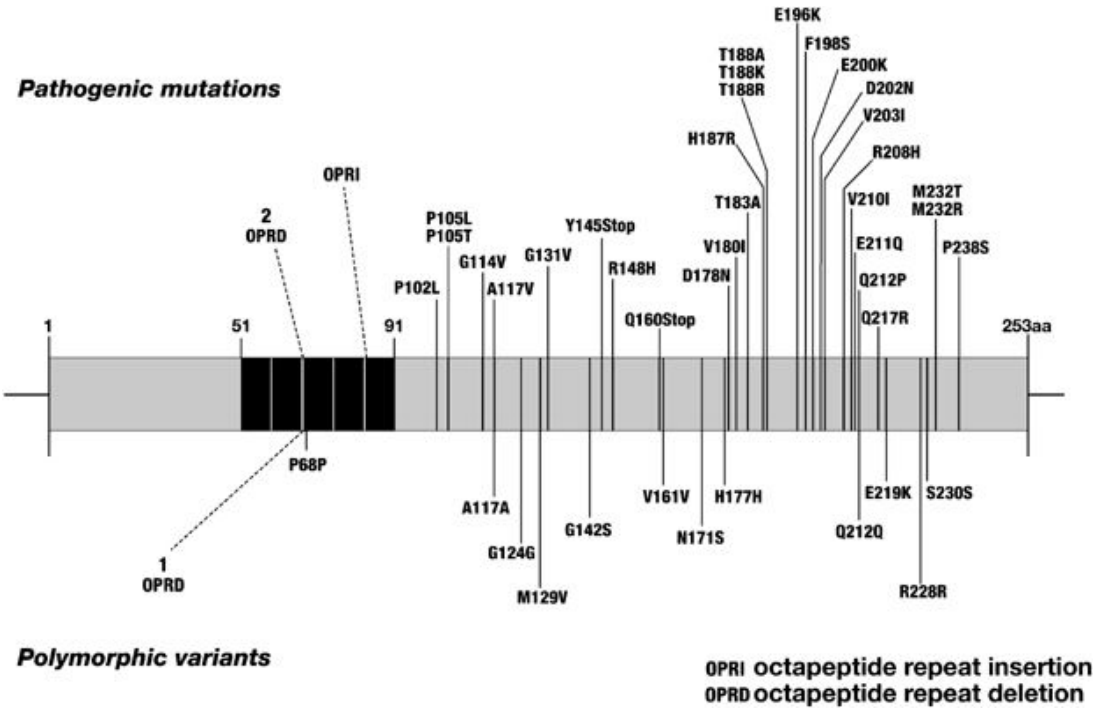


Figure 2: Known genotypes associated with familial cases of CJD (Mead, 2006). Some of the pathogenic mutations listed above have been reported in simplex cases of CJD and could reflect the chance occurrence of a rare variant in patients with sporadic CJD.

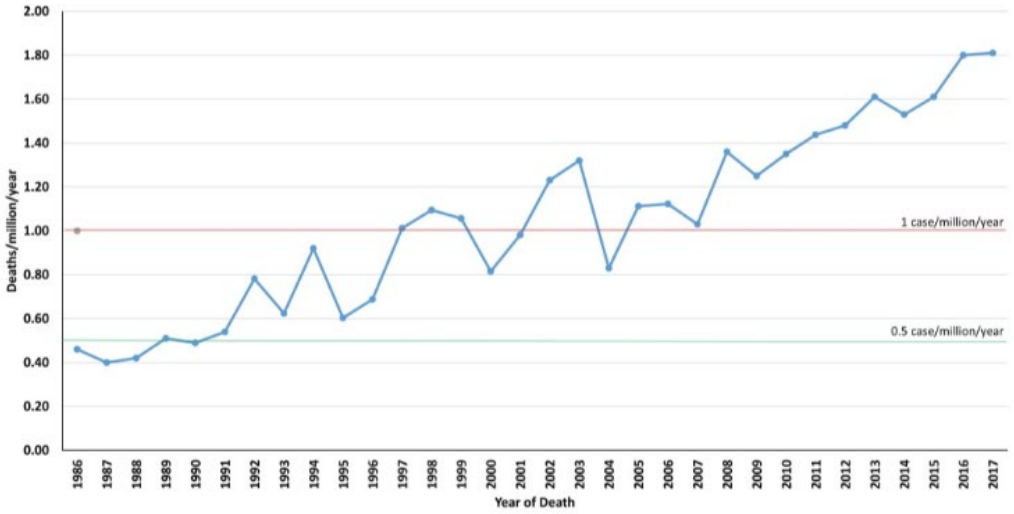


Figure 3: Deaths per million per year from sCJD between 1986 and 2017 in the UK (The National CJD Research and Surveillance Unit, 2017).

Prion diseases have a heterogeneous clinical presentation both between and within subtype. To better delineate patient progression and outcome, a scale of disease severity was published in 2013 after monitoring patients enrolled in the UK National Prion Monitoring Cohort over the course of 373 patient-years (Thompson *et al.*, 2013). The MRC Prion Disease Rating Scale (henceforth referred to as the MRC Scale) was designed such that patients could be scored retroactively based on commonly documented clinical parameters, and allows cross-sectional analysis of biomarkers and disease features across disease progression. Despite the scale proving useful for patient monitoring and triage, both important aspects of patient management, prion disease remains incurable. To facilitate rational drug design and improve patient treatment and management, deeper knowledge of the disease aetiology and pathology is needed.

1.3 Epigenetics

Epigenetics can be summarised as the study of how the action of genes can be temporally and spatially regulated, a process which has been known to direct physiological development since Conrad Waddington's coining of the term "Epigenotype" in 1942 (Waddington, 1942). Outside the original context of developmental biology epigenetics has also been studied in the context of disease, most notably cancer of which epigenetic dysregulation has become a hallmark (Brown and Strathdee, 2015), and recently such dysregulation has been convincingly linked to neurodegenerative diseases such as Alzheimer's Disease (De

Jager *et al.*, 2014; Lunnon *et al.*, 2014), Amyotrophic Lateral Sclerosis (Xi *et al.*, 2013) and Parkinson's Disease (Masliah *et al.*, 2013).

There are three canonical branches of epigenetics: the action of non-coding RNAs, modifications to histone proteins which coordinate chromatin structure, and modification of DNA, principally methylation of cytosine residues. Partly due to the chemical stability of the latter epigenetic mark and the comparative molecular fragility of protein and RNA, DNA methylation has been enthusiastically researched in many biological contexts. DNA methylation predominantly occurs at the 5'-carbon of cytosine in cytosine-guanine dinucleotides (CpGs) which are both relatively depleted and unevenly distributed across the human genome. CpGs cluster into genomic features known as CpG islands, defined as regions of >200bp containing >50% GC composition and an observed to expected CpG ratio of >60% (Bird, 1986). CpG islands are frequently located upstream, on, or downstream of gene transcription start sites and often extend into the gene body (Gardiner-Garden and Frommer, 1987). Although around 70% of all CpGs in the human genome are methylated, these islands are characterised in part by the absence of methylation. This is notable as compared to a fixed random rate of mutation of cytosine residues, methylcytosine is 42 times more likely to mutate via deamination to thymine (Cooper and Youssoufian, 1988). This suggests that CpG islands are conserved during evolution in part due to their lack of methylation (Sved and Bird, 1990; Serge Saxonov, Berg and Brutlag, 2006). As early as 1986 Adrian Bird suggested these regions may serve as binding platforms for protein factors, and in doing so prevent the binding of enzymes that introduce DNA methylation as well as providing an accessible substrate for transcription factors, chromatin modifiers or repressors. Thanks to genome-wide sequencing, we now know that 72% of all annotated promoters overlap with CpG islands (S. Saxonov, Berg and Brutlag, 2006).

1.4 DNA Methylation Machinery

1.4.1 DNA Methylation Writers

DNA methylation is introduced by DNA methyltransferases (DNMTs), a family of enzymes of which three have been identified in humans: DNMT1, DNMT3A and DMNT3B. These enzymes were originally identified in bacteria and are highly conserved, sharing the same 10 motifs between prokarya and eukarya (Kumar *et al.*, 1994). DNMT1, DNMT3A and DMNT3B take cytosine and S-adenosyl-L-methionine (SAM) as their substrates, the latter of which is the methyl-donor. In brief, the methylation reaction begins with a nucleophilic attack on the 6'-carbon of cytosine by a conserved cysteine-thiolate residue of the DNMT, after which the cytosine is everted from the double helix into the catalytic pocket of the DNMT. The now-anionic 5'-carbon of cytosine attacks the donor methyl group of SAM and S-adenosyl-L-homocysteine (SAH) is expelled. Beta-elimination separates the DNMT from the cytosine moiety and loss of a proton resolves the stabilised purine ring structure to 5-methyl-cytosine (5mC) (Bestor and Verdine, 1994). This process is allosterically stabilised by proton transfer between the 3'-nitrogen of cytosine and a proximal carboxylic amino acid and, importantly, can be competitively inhibited by the reaction's by-product SAH. A summary of the reaction is presented in Figure 4.

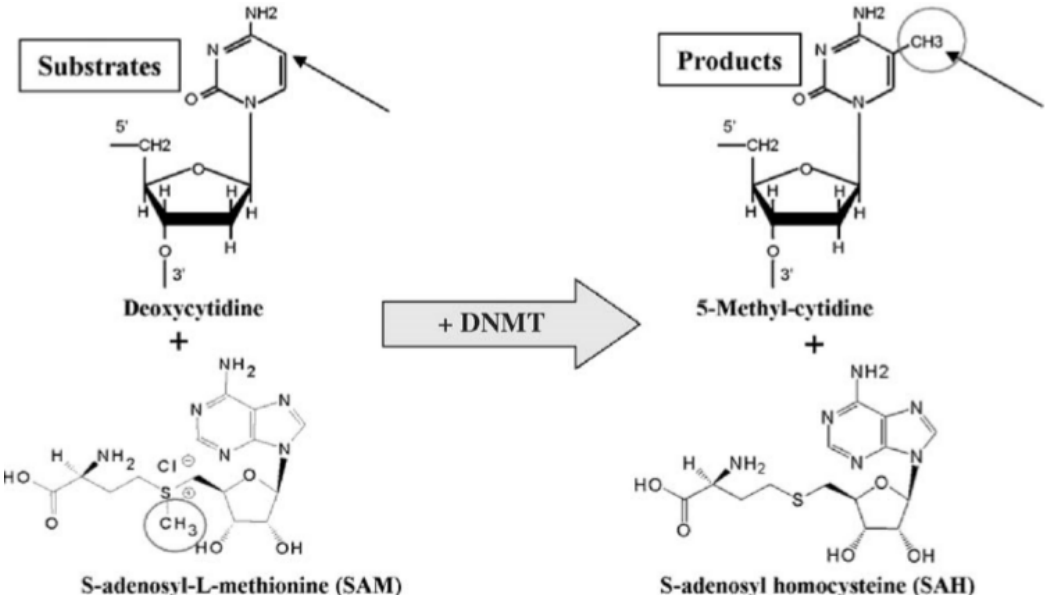


Figure 4: Chemical structures of DNMT substrates and reaction products during methylation of deoxycytidine, the nucleotide analogue of cytosine. This figure is adapted from Bestor and Verdine, 1994.

The first eukaryotic protein to be identified as having methyltransferase activity was purified from murine cell culture, after which the peptide sequence was determined by Edman sequencing and used to generate nucleotide probes to investigate a cDNA library. One cDNA product was found to encode a transcript of a gene homologous to bacterial DMNTs (Bestor *et al.*, 1988). This methyltransferase was able to symmetrically methylate hemi-methylated DNA *in vitro*, as well as restore symmetrical methylation at CpG sites in murine cells transfected with hemi-methylated DNA, while *de novo* methylation of similarly transfected unmethylated DNA was not observed. This confirmed the existence of a hypothesised maintenance DMNT which ensures symmetrical DNA methylation is restored after semi-conservative DNA replication (Stein *et al.*, 1982). In humans this is encoded by *DNMT1*. DNMT1 has been found to colocalise with replication forks during S phase of the cell cycle, as has UHRF1. Studies have shown that *Uhrf1*$^{-/-}$ mouse embryonic stem (ES) cells exhibit significant global hypomethylation and that UHRF1 has a seven times higher binding affinity for symmetrical CpG-containing sequences when hemi-methylated rather than fully methylated (Bostick *et al.*, 2007). This suggests that UHRF1 binds to hemi-methylated CpG dyads post-replication and recruits DNMT1, which restores methylation on the nascent strand. Intriguingly, further work revealed that DNA-bound UHRF1 everts 5mC out of the helix in a similar fashion to DNMT1 (Arita *et al.*, 2008; Avvakumov *et al.*, 2008; Hashimoto *et al.*, 2008).

DNMT3A and *DNMT3B* were identified in human and murine genomes by a methyltransferase motif homology search (Okano, Xie and Li, 1998) which was prompted after low levels of *de novo* DNA methylation were observed in *Dnmt1* knockdown recombinant murine stem cells (Lei *et al.*, 1996). DNMT3A and DNMT3B were found to have equal methyltransferase activity for unmethylated and hemi-methylated substrate DNA. Further research revealed that they have relaxed substrate specificity and can methylate symmetrical CpApG and GpTpC triads, though it binds preferentially to CpG dyads. The rarity of non-CpG methylation in the human genome, taken with the above, suggests that DNMT3A and DNMT3B hemi-methylate cytosines in these three contexts, before DNMT1 introduces symmetrical methylation on the unmethylated strand at CpG loci only (Ramsahoye *et al.*, 2000).

A fourth gene lacking methyltransferase activity was identified as having a cysteine-rich region homologous to similar regions in the *de novo* DNMT3s (Aapola *et al.*, 2000).This gene was named DNMT3-*Like* (*DNMT3L*) and was found to only be significantly expressed in germline cells. *Dnmt3L*^{null} mice are sterile and *Dnmt3L*$^{+/-}$ offspring of maternal *Dnmt3L*$^{-/-}$ mice are nonviable. It emerged that this was

because loss of *Dnmt3L* function lead to a failure to methylate maternally imprinted genes during oogenesis (Bourc'his *et al.*, 2001). Further work revealed that [*Dnmt3a*$^{-/-}$, *Dnmt3b*$^{+/-}$] oocytes were also unable to establish imprinting, suggesting that *DNMT3A* and *DNMT3L* in combination are responsible for initiating imprinting during development (Hata *et al.*, 2002). Immunoprecipitation studies of DNMT3L revealed that it binds specifically to histone protein H3 and that successive methylation of H3K4 prevented this interaction (Ooi *et al.*, 2007). It is thus clear that post-meiotic *de novo* DNA methylation depends at least in part on the underlying chromatin landscape being amenable to DNMT3L binding.

1.4.2 DNA Methylation Erasers

DNA methylation is removed through passive and active mechanisms. The former can rely on exclusion of DNMT1 or UHRF1 from the nucleus, leading to a dilution of 5mC across successive generations. Another form of passive removal also relies on replication-based dilution of 5mC, but in this instance 5mC is oxidised by TET1, TET2 or TET3 to 5-hydroxymethylcytosine (5hmC). DNMT1 has a considerably weaker substrate affinity for 5hmC and so DNA methylation is not maintained over successive generations and is diluted out in a similar fashion (Hashimoto *et al.*, 2012).

Suggested active mechanisms of DNA demethylation also involve TET enzymes. One proposed mechanism is that 5mC is oxidised to 5hmC which is then deaminated to 5-hydroxyuracil (5hU) by activation-induced deaminase (AID). 5hU would then be replaced by cytosine through base excision repair as several DNA glycosylases have high affinity for 5hU:G pairs. However *in vitro* work has demonstrated that bulkier adducts to the cytosine ring increasingly block AID activity, thus this mechanism is unlikely (Nabel *et al.*, 2012; Rangam *et al.*, 2012). TET enzymes are also capable of successively oxidising 5hmC to 5-formylcytosine and 5-carboxylcytosine (5caC). These moieties are substrates for the thymine DNA glycosylase TDG, and depletion of TDG in murine ES cells leads to detectable increases in 5caC (He *et al.*, 2011). Modular structures and catalytic pathways of enzymes involved in DNA methylation and demethylation are summarised in Figures 5 and 6 (Wu and Zhang, 2014a).

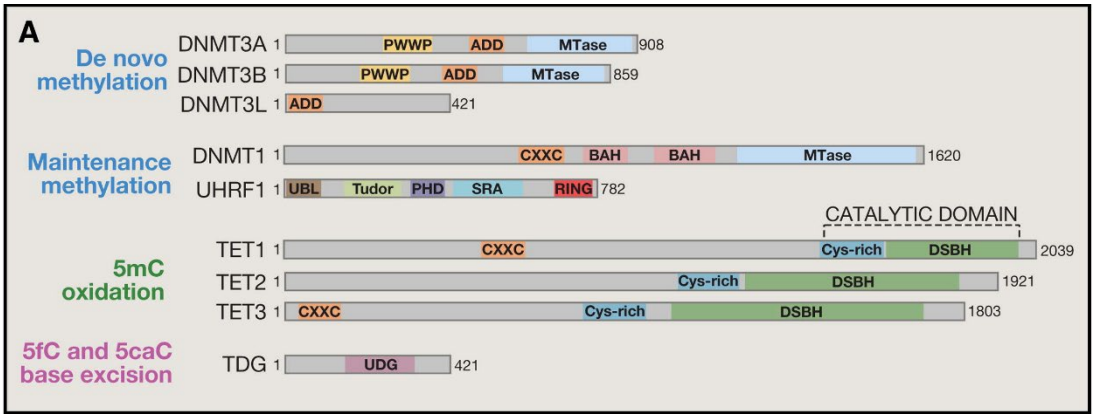


Figure 5: Writers, maintainers, and erasers of DNA methylation. Catalytic domains responsible for DNA methylation, 5mC oxidation and 5fC/5caC excision are shown in light blue, green and magenta respectively. Domain abbreviations are as follows: ADD, ATRX-DNMT3-DNMT3L; BAH, bromo-adjacent homology; CXXC, zinc-finger Cysteine-X-X-Cysteine; DSBH, Double-Stranded β-Helix; MTase, DNA methyltransferase domain; PHD, Plant homeodomain; PWWP, proline-tryptophan-tryptophan-proline motif; SRA, SET and RING associated; RING, Really interesting new gene; UBL, ubiquitin-like domain; UDG, Uracil DNA Glycosylase. Figure adapted from Wu and Zhang, 2014a

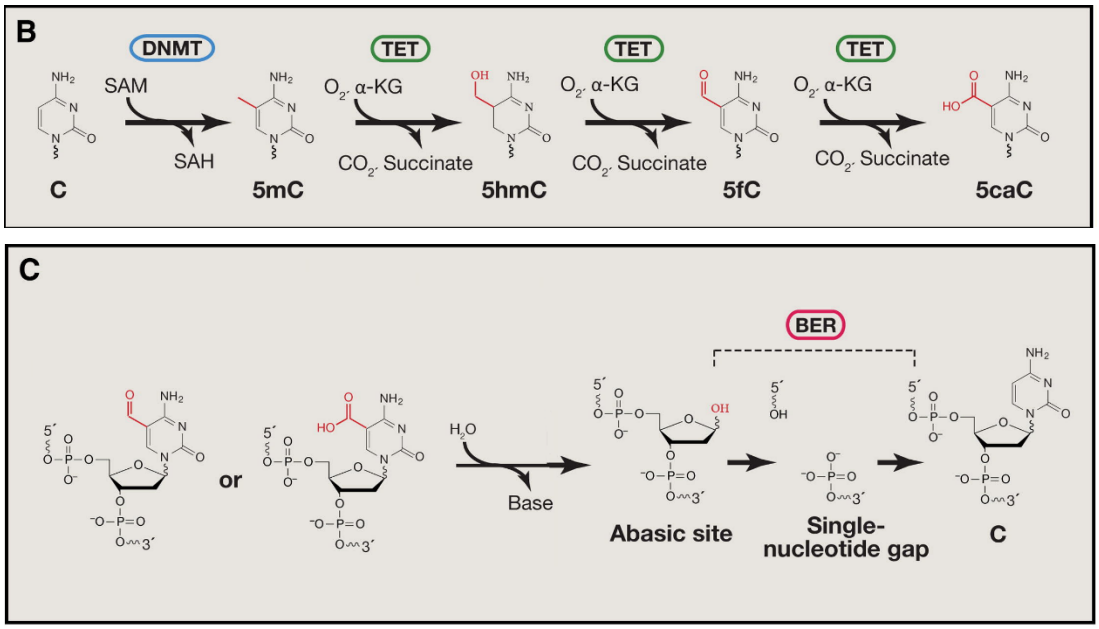


Figure 6: Catalytic pathways of passive demethylation (TET only) and active demethylation (TET & Base Excision Repair) mechanisms. Figure adapted from Wu and Zhang, 2014b

1.4.3 DNA Methylation actors

DNA methylation has a range of functions exquisitely dependent on local genomic and epigenomic context (Jones, 2012). Typically, methylation at promoters correlates with gene silencing, while active promoters are unmethylated. This means of regulation is tightly linked with chromatin structure and modifications: in an active promoter the transcription start site (TSS) is depleted of nucleosomes. Nucleosomes which flank an active TSS bear H3K4me3 marks which prevent DNMT3L binding, or may contain H2A.Z, a histone protein which also blocks *de*

novo methylation. The relationship between H2A.Z and methylation is bidirectional, as depletion of H2A.Z deposition results in global hypermethylation, while disruptive mutation of DNMTs increases H2A.Z presence (Zilberman *et al.*, 2008). Furthermore, chromatin immunoprecipitation sequencing (ChIP-Seq) analyses of TET1 distribution revealed that TET1 localises to active promoters which overlap CpG islands (Wu and Zhang, 2011), as TET1's CXXC motif binds strongly to unmethylated CpG dyads (Zhang *et al.*, 2010). TET presence at these sites therefore maintains demethylated regions by oxidising local 5mC as part of its previously described role in removal of methylation.

Conversely, silenced promoters tend to contain nucleosomes and methylated CpGs. A well-studied example is the DNA mismatch repair protein *MLH1*, which is frequently silenced in cancer. Methylation of *MLH1's* promoter directly correlates with nucleosome occupancy and in turn with decreased expression. Treatment of MLH1-silenced cells with 5-aza-2'-deoxycytidine, a cytosine analogue resistant to methylation, is sufficient to demethylate the promoter, remove nucleosomes and restore expression (Lin *et al.*, 2007). It is thus clear that DNA methylation emerges from and can remodel the underlying chromatin landscape.

An essential function of DNA methylation in humans is repression of transposon activity in somatic cells (Walsh, Chaillet and Bestor, 1998). Transposons are parasitic DNA sequences that can copy and paste themselves across the genome. In humans transposons are thought to comprise almost 60% of the total genome (de Koning *et al.*, 2011), with copies of one sequence (*Alu*) comprising up to 10% of the human genome. As well as impairing overall genomic stability in somatic cells which can contribute towards progression of cancer (Roman-Gomez *et al.*, 2008), in some instances disease can be caused by gene-specific insertions in the germline (Dombroski *et al.*, 1991). Transposons and indeed all repeat elements are enriched with CpG dinucleotides compared to the rest of the genome, albeit less so than CpG Islands, with *Alu* sequences alone containing 23% of all genomic CpGs (Luo, Lu and Xie, 2014). These sequences are highly methylated in somatic cells, and demethylation is associated with increased activity and subsequently is reflected in the global demethylation observed in many types of cancer.

Study of the difference in methylation landscapes of the active and inactive X chromosome has greatly informed our understanding of consequences of DNA methylation based on genomic context. Contrary to expectations, a ground-breaking study discovered that the active chromosome is 2.4 times more methylated than the

inactive chromosome (Hellman and Chess, 2007). Subsequent experiments confirmed that the inactive chromosome was mostly methylated at promoter regions, while the active chromosome exhibited extensive gene-body methylation, particularly in exons. As well as correlating with active transcription, gene-body DNA methylation has been linked to alternative splicing by preventing interactions between CTCF and DNA. CTCF is a scaffolding protein which binds to DNA with relatively low specificity and in doing so coordinates chromatin loops. This is thought to affect the kinetics of RNA Polymerase II during transcription, affecting the incorporation of additional exons (Shukla *et al.*, 2011; Racko *et al.*, 2018). An important example of this is to be found in activation of CD4+ T-Cells, where in naïve cells *CD45* is unmethylated at exon 5, allowing CTCF to bind. Subsequently an isoform of *CD45* containing exon 5 is expressed. However, exon 5 is methylated in activated CD4+ T-Cells, preventing CTCF binding and excluding exon 5 from the transcript (Marina *et al.*, 2015). CTCF frequently colocalises with TET1 and reporter constructs with a CTCF binding site are hypomethylated compared to analogous constructs lacking a binding site, suggesting that CTCF may be able to establish demethylation (Stadler *et al.*, 2011).

Canonically, DNA methylation has been assumed to block binding of proteins such as CTCF and transcription factors such as those which bind to cAMP-Response Elements (CREs). This model stems from early work describing how methylation of CREs in the proenkephalin promoter inhibit expression in cell culture, and a reporter gene study of promoter CRE methylation which showed similar effects on β-globin expression (Iguchi-Ariga and Schaffner, 1989; Comb and Goodman, 1990). Yet recently evidence has emerged that CpG methylation of motifs may increase the binding affinity of some transcription factors, such as CEBPβ (Zhu, Wang and Qian, 2016). This is perhaps surprising as until recently this property was associated only with proteins containing a Methyl-Binding Domain (MBD), members of a family known as MBD proteins.

The first two MBDs to be identified were Methyl-CpG Binding Proteins 1 and 2 (MeCP1 and MeCP2 respectively), although MeCP1 was later discovered to be a protein complex containing MBD1 (Meehan *et al.*, 1989; Lewis *et al.*, 1992; Cross *et al.*, 1997). Identification of the MBD allowed cDNA screening for new family members, leading to the identification of *MBD2*, *MBD3* and *MBD4*, of which MDB1, MDB2 and MBD4 were found to specifically bind methylated DNA *in vitro* and in murine cell culture (Hendrich and Bird, 2015).

Perhaps unsurprisingly, although their shared function is to bind methylated DNA, different MBDs play different roles. MBD1 interacts with chromatin assembly factor 1, and the two proteins localise to heterochromatin in the nucleus (Reese *et al.*, 2003). This suggests that MBD1 may play a role in establishing heterochromatin, or at least in its maintenance. As well as an MBD, MBD1 contains a transcriptional repression domain at its C-terminus which can silence genes without the assistance of chromatin assembly factor 1 (Ng, Jeppesen and Bird, 2002). Yet *MBD1* is also alternatively spliced into at least five isoforms which can contain either zero, two or three CXXC motifs, motifs which as described previously have high binding affinity to unmethylated CpG dyads (Saya *et al.*, 2002). Because of this, transcriptional repression and chromatin remodelling promoted by MBD1 is likely to depend on expression levels of the splice isoforms, particularly the balance between the 0/2 and 3 CXXC motif containing isoforms, as well as local density of DNA methylation in the sequences it binds to. While these four *MBD1* isoforms are promiscuous, the fifth isoform lacks a CXXC domain and so only binds to methylated DNA. It has been shown to block TET1 binding and therefore prevent oxidation of 5mC (P. Zhang *et al.*, 2017). In this sense it can be considered to have a more defined and perhaps distinct role from the other isoforms, which are less competitive and thus poorer inhibitors of TET1 but as a result allow fine tuning of TET1 activity depending on the local genomic landscape.

MBD2 has been shown to recruit the MeCP1 complex, and in doing so associates with histone deacetylases which facilitate gene silencing (Ng *et al.*, 1999). However, *MBD2* also has two isoforms, of which *MBD2a* has an N-terminal domain which allows it to interact with RNA helicase a, a factor in CRE-binding (CREB) complexes. In doing so, *MBD2a* can actually activate expression of genes governed by CREB through binding to methylated, and thus canonically silenced, CREs (Fujita *et al.*, 2003).

MeCP1 is now also known as the NuRD complex. NuRD on its own cannot bind to methylated DNA, relying instead on MBD2 to recruit it. Interestingly, alongside MBD1, the closely-related MBD3 was also identified as part of the NuRD complex (Zhang *et al.*, 1999). This prompted knockout experiments in mice where *Mbd3*$^{-/-}$ was found to be embryonically lethal, while *Mbd2*$^{-/-}$ mice were viable and could produce offspring (Hendrich *et al.*, 2001). This suggested both differences in function and potential redundancy between the two genes and so NuRD complexes containing tagged MBD2 and MBD3 were purified and analysed. Strikingly the two MBD-NuRD complexes did not copurify and tagged complexes were found to

contain different subunits (Le Guezennec *et al.*, 2005). Le Guezennec *et al*. hypothesise that MBD2/NuRD complexes bind to demethylated DNA and establish histone deacetylation, which favours binding of MBD3/NuRD complexes to deacetylated nucleosomes rather than to DNA. Thus deacetylation and transcriptional suppression can be established by MBD2 and maintained by MBD3.

To further complicate matters, two paralogues of *MBD3* have been identified with mutually opposing functions. MBD3L1 replaces MBD3 in the NuRD complex and can interact directly with MBD2, enhancing silencing of methylated genes without the need to bind to deacetylated nucleosomes (Jiang, Jin and Pfeifer, 2004). Conversely, MBD3L2 can compete with MBD2 for binding of the NuRD complex but lacks an MBD domain. In doing so, it can displace the NuRD complex, potentially clearing the way for transcriptional reactivation (Jiang *et al.*, 2005).

Less complex are the known roles of *MBD4-6*. MBD4 is a thymine glycolase with high affinity for symmetrically methylated CpGs where one methylated cytosine has undergone deamination to thymine (rather than uracil, the deamination product of unmethylated cytosine). In doing so it excises the thymine and plays a role in mismatch repair (Hendrich *et al.*, 1999). The recently discovered proteins MBD5 and MBD6, despite containing MBDs, do not show any ability to bind methylated DNA *in vitro*. However, they do localise to heterochromatin, suggesting a potential role in remodelling or maintenance (Sasai *et al.*, 2010).

It is clear from the above that *MBD1-3* and *MBD3L1-2* play complex functions depending on their splice isoform ratio, their relative stoichiometries, their subcellular localisation, and their interactions with underlying the chromatin and methylation landscape.

Unlike MeCP1, MeCP2 is a single protein. Situated on the X chromosome, *MeCP2* is notorious for developing germline mutations which can cause the neurodevelopmental disorder known as Rett's Syndrome, almost exclusively in females. Rett's syndrome is characterised by early loss of speech, coordination, microcephaly, ataxia and autism, although it is typically non-lethal (Amir *et al.*, 1999). Like *MBD1*, *MeCP2* contains both an MBD and a transcriptional repression domain which coordinates histone deacetylases. In fitting with the post-natal emergence of Rett's Syndrome, *MeCP2* is most highly expressed in mature neurons rather than developing precursors or other brain cell types such as astrocytes or glia (Kishi and Macklis, 2004). Here, incredibly, the absolute level of MeCP2 seems to correlate tightly with the level of CpG methylation, with high throughput ChIP-

sequencing revealing that ~70% of CpGs (the proportion of genome-wide CpGs that are typically methylated in human brain (Ehrlich *et al.*, 1982)) associate with MeCP2 (Skene *et al.*, 2010). This blanketing of neuronal DNA suggests a generic epigenetic control of transcriptional regulation, and indeed its role is more complex than simple transcription repression. Counter-intuitively, while disruption or upregulation of *MeCP2* dysregulates thousands of genes 85% of those genes are upregulated in both cases (Chahrour *et al.*, 2008). Like MBP2, MeCP2 associates with CREB1 at CREs near actively transcribed genes, as well as covering regions of methylated DNA and preventing TET-induced demethylation.

As discussed above, DNA methylation has a complex bidirectional relationship with both underlying chromatin modifications and associated proteins which in turn are differentially enriched across genomic features. It is also clear that this relationship is extremely dynamic but also partitioned into several stages between DNA methylation and demethylation, and chromatin condensation and relaxation, allowing for subtle and adaptive control of gene expression. It is therefore not surprising that DNA methylation can be influenced by a cell's environment, be this during differentiation in response to intercellular signalling, over the course of a disease or even as a lasting response after tobacco smoke exposure (Joehanes *et al.*, 2016). With the recent development of new high-throughput technologies to measure changes in DNA methylation, epigenetic research has become increasingly useful in uncovering novel biological mechanisms and providing new insight into disease processes.

1.5 Measuring DNA Methylation

There are two approaches to measuring DNA methylation (global and specific) and methodology varies between them. Early research focused on global levels of DNA methylation, and so techniques such as high-performance liquid chromatography UV (HPLC-UV) (Kuo *et al.*, 1980) or liquid chromatography mass spectrometry (LCMS) (Song *et al.*, 2005) were developed to hydrolyse DNA, separate cytosine and 5mC through chromatography and then calculate the global 5mC/C ratio. Antibodies against 5mC can be used in ELISAs to measure global methylation with greater ease but a trade off in accuracy. These methods are useful when large changes in methylation are expected, such as in analysis of tumour-derived DNA. Due to their genomic abundance and enrichment of CpG sites, methylation levels at transposons such as LINE-1 are also studied as a surrogate marker of global DNA methylation levels (Yang *et al.*, 2004).

A subsequent quantitative development in the study of global DNA methylation is the luminometric methylation assay (LUMA) (Karimi *et al.*, 2006), which incorporates the restriction endonucleases HpaII and MspI. HpaII only cuts DNA at unmethylated CCGG sequences, while MspI recognises the same sequence but is insensitive to methylation. Each DNA sample is digested in two separate reactions using EcoRI and each respective enzyme, after which the EcoRI 5'-AATT and HpaII/MspI 5'-CG overhangs are filled in using a DNA Polymerase extension assay. This is done in a pyrosequencer, which uses predefined stepwise addition of nucleotides, incorporation of which is coupled with a luciferase reaction that produces light, to sequence DNA. The ratio between the light signal during the incorporation of guanosine and cytosine in paired MspI and HpaII-digested samples is equivalent to the global proportion of 5mC.

The second approach to studying DNA methylation is site-specific and early methods also made use of restriction enzymes. DNA can be fragmented and then digested with the methylation-sensitive HpaII enzyme which only cuts at unmethylated sites, before PCR using primers which anneal around the CpG of interest. The amount of product generated from intact fragments, normalised to a separate control amplicon from the same sample which does not contain CCGG, can be used to quantitate differences in methylation between samples within the amplicon, albeit with low resolution. Another approach using next generation sequencing is to digest DNA using HpaII, sequence the fragmented products and align them to a reference genome: the number of sequencing reads at each intact HpaII site corresponds to the methylation level at that site (Kurdyukov and Bullock, 2016).

The current gold standard approach to detecting DNA methylation is to begin with bisulphite conversion (Frommer *et al.*, 1992). This chemical process results in the deamination of unlabelled cytosine to uracil, which is typically followed by PCR resulting in a conversion to thymine. Methylated cytosines are highly resistant to conversion and so remain unaffected. After this treatment CpG sites can be assayed for cytosine or thymine using standard genotyping techniques, allowing for a ratio of C:T and thus a ratio of methylated:unmethylated cytosine at the site to be calculated. However, the product of bisulphite conversion is fragmented, single stranded DNA with increased sequence redundancy, which has consequences for long-term storage and assay design. After conversion a simple melt curve qPCR or the difference in Ct between two qPCR reactions with primers specific to the

methylated and unmethylated site can be used to calculate relative methylation levels.

In the absence of a prior hypothesis, measuring DNA methylation at hundreds of thousands of loci is required. Whole Genome Bisulphite Sequencing (WGBS) is the most expensive means of detecting methylation on this scale in terms of sample input, cost and analysis. Because most of the data collected does not relate to cytosine methylation, Reduced Representation Bisulphite Sequencing (RRBS) is a more cost-effective approach which involves digesting genomic DNA using MspI to enrich the sample for fragments terminating in a digested CpG before bisulphite treatment and sequencing.

Illumina Infinium arrays are cheaper per sample than the previous two techniques but are limited to measuring methylation at approximately 27,000, 450,000 or 850,000 sites, and the distribution of probes on the former two arrays is mostly restricted to CpG islands and shores. Moreover, the arrays can currently only be used to investigate human DNA preventing study of methylation in mouse models of disease. Yet the cost-efficiency of the array has permitted the analysis of sufficient numbers of samples to power studies of genome-wide DNA methylation, and as a result the array has become widely used. The technology underpinning the Infinium array is described in more detail in the next section.

1.5.1 The 450K Array

The 450K Infinium beadchip array contains probes targeting 485,764 CpG sites genome-wide, covering 99% of RefSeq genes with an average of 7 probes per gene (Illumina, 2012). Each probe contains over 100 nucleotides, of which two at the centre correspond to the CpG site. There are two types of probe chemistries and the first (Type I) is inherited from the 450K's 27K precursor array. These consist of a pair of probes at each site, one of which recognises unmethylated CpGs and contains a CA corresponding to GT sequences where unmethylated cystosines are converted to thymines, and one which recognises methylated CpGs and contains a CG corresponding to an unconverted guanine-cytosine pair. Because DNA methylation is thought to generally correlate across short sections of DNA (~50bp), CpGs underlying the probe hybridisation site are assumed to be methylated or unmethylated per probe type and the probe sequence is designed to reflect this (Bibikova *et al.*, 2006; Eckhardt *et al.*, 2011). The second type (Type II) uses one probe per site and at the 3' end is a cytosine to complement the guanine of a CpG pair, allowing either adenine or guanine to be incorporated upon hybridisation to

unmethylated or methylated DNA. In doing so probes per site decreases from 2 to 1, and underlying CpG sites are accommodated for by including variable “R” nucleotides, preventing bias towards different hybridisation efficiencies of methylated or unmethylated DNA amplicons. Probes lose binding efficiency after inclusion of more than 3 R nucleotides, and so while allowing greater target capacity, Type II probes can only reliably target sites in regions of low CpG density (Bibikova, Barnes, *et al.*, 2011).

Once DNA is hybridised to probes, labelled nucleotides are incorporated in a single-nucleotide extension step. With methylated DNA cytosine or guanine are incorporated and if DNA is unmethylated adenine or thymine are incorporated. The former nucleotides are labelled with biotin and the latter with 2,4-dinitrophenol. Immunochemistry can then be used to identify methylated and unmethylated loci based on green or red fluorescence respectively. Site-specific signal intensity is recorded and output per sample in intensity data (iDAT) files. The process of hybridisation and extension based on probe type is illustrated in Figure 7.

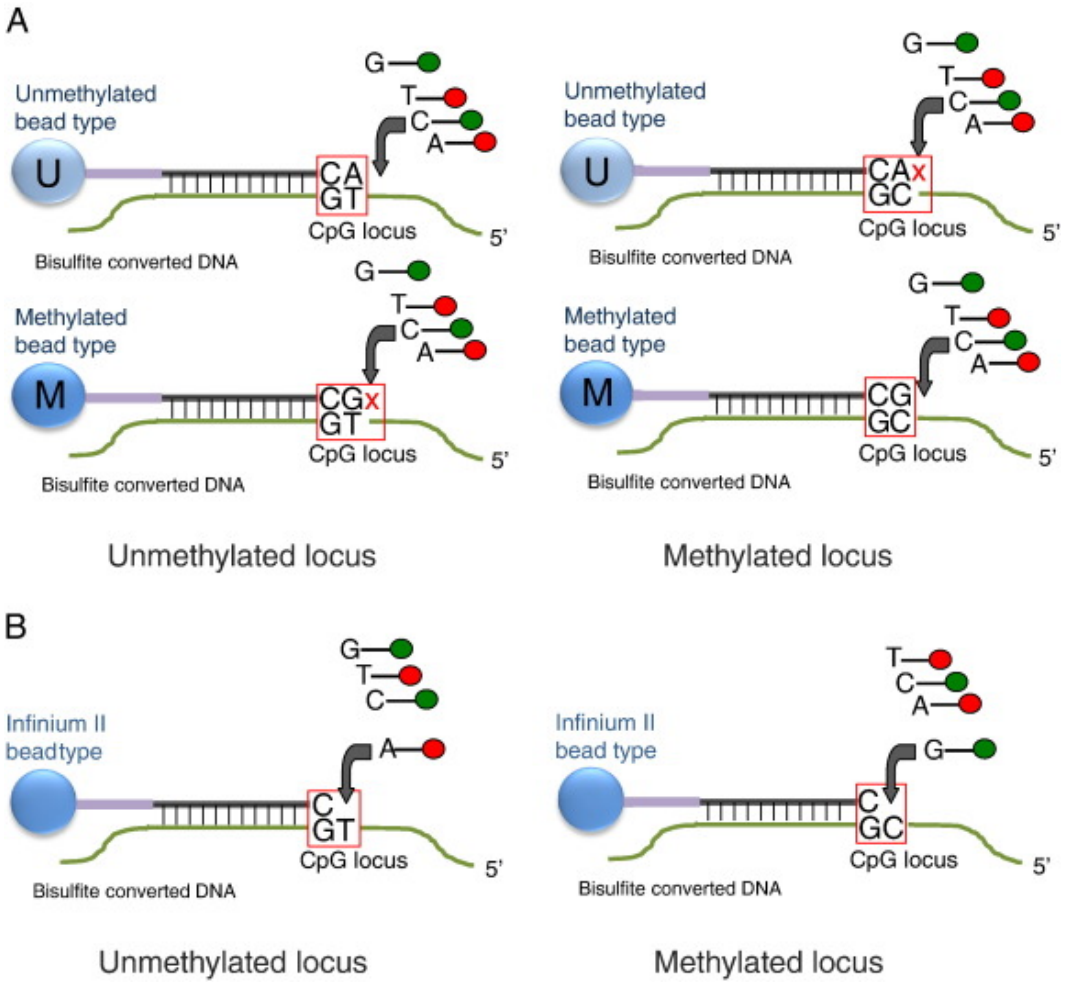


Figure 7: Type I (A) and II (B) 450K probe chemistries. Type I probes are complementary to either methylated or unmethylated DNA, hybridisation of DNA to either probe is linked to a single nucleotide extension at the site immediately adjacent to the CpG target which is identified by immunofluorescence. Type II probes allow incorporation of either adenine or guanine at the site complementary to the target cytosine, meaning the same probe can report either a methylated (red) or unmethylated (green) signal (Figure from Bibikova, Le, et al. 2011).

1.5.2 Addressing technical limitations of the 450K Array

The technical limitations of the 450K array have been extensively discussed in the research community and reviewed in the literature. The intensities in iDAT files are used to calculate a ratio called "β" ($\beta = \frac{Methylated\ fraction}{Methylated\ fraction + Unmethylated\ fraction + \alpha}$), where α is a corrective factor. β varies from 0 (totally unmethylated) to 1 (totally methylated) and so is an intuitive measure of DNA methylation, but in 2010 Du *et al.* showed that a logit-transformed variant of β, "M", leads to less bias during statistical manipulation of extreme β values (Du *et al.*, 2010).

In 2013 Chen *et al.* published a list of cross-reactive probes included on the array (Y. A. Chen *et al.*, 2013), suggesting these probes could not be considered reliable.

In the same year Triche *et al*. devised a pipeline (normal-exponential out-of-band, or "noob") for correcting background fluorescence (Triche *et al.*, 2013). By that time it was also known that the two different probe chemistries present on the array lead to heteroscedastic bias between type I and II probes, possibly because Type II probes had higher background signal and so β values are compressed away from 0 and 1, and towards the median of β = 0.5 (Dedeurwaerder *et al.*, 2011; Pidsley *et al.*, 2013). A number of normalisation strategies were subsequently published to correct this bias and thus increase Type II assay sensitivity (Maksimovic, Gordon and Oshlack, 2012; Pidsley *et al.*, 2013; Teschendorff *et al.*, 2013).

Illumina's GenomeStudio software allows basic manipulation of 450K data, but not the adjustments detailed above. Researchers began to develop and expand upon multiple diverse packages using the statistical platform R for data processing, quality control and complex analysis. Wrappers such as RnBeads and ChAMP began to gather these functions into relatively user-friendly programming steps in 2014 (Assenov *et al.*, 2014; Morris *et al.*, 2014). Interest also surrounded the issue of cellular heterogeneity and how imbalances of cell composition between and within sample groups could contaminate data. Houseman *et al*. published a reference-based method for blood and later a reference-free method for other tissues to gauge cell composition from methylation data, which researchers adopted as a tool to regress out the effects of such changes from EWAS (Houseman *et al.*, 2012; Houseman, Molitor and Marsit, 2014). Such corrections are now regarded as "critical" for epigenetic analysis of heterogeneous tissues (Jaffe and Irizarry, 2014). In 2016 a list of probes overlapping single nucleotide polymorphisms with mean allele frequencies of 5% and above was published and while these may have utility, for example in imputing ethnic background of the sample donor, many studies routinely exclude these probes from analysis (Zhou, Laird and Shen, 2016). Various permutations of these corrections are regularly applied in any study which incorporates 450K array data, and discussions of the technical constraints of the Illumina BeadChip platforms and how best to process data continue.

1.6 DNA Methylation in Neurodegeneration

Many studies of altered DNA methylation in disease have a major limitation in that it is difficult to establish whether the observed effects are causal or a consequence of pathogenesis. This is particularly difficult to resolve in studies of central nervous diseases, where the affected tissue can only be sampled post mortem and therefore pathology is usually advanced. Study of blood-derived DNA as a proxy tissue is a useful means of identifying methylation sites which may become increasingly altered across the course of the disease, and in some instances such changes have been found to reflect similar changes in the brain (Masliah *et al.*, 2013). Although familial neurodegenerative diseases have been studied, it is perhaps sporadic disease where study of DNA methylation might prove most useful, as these forms of disease are the most prevalent and lack Mendelian genetic determinants. Instead, alongside disease modifying polymorphisms, risk has been linked to lifestyle factors such as exercise (Sofi *et al.*, 2011), diet and Body Mass Index (Luchsinger and Gustafson, 2009), years of education (Larsson *et al.*, 2017), smoking history (Wang *et al.*, 2011), sleep disturbance (Clark and Warren, 2013) and late-life depression (Singh-Manoux *et al.*, 2017). Study of differential methylation in disease may reveal how these factors modify disease risk and pathogenesis, as well expand upon understanding of the disease mechanism itself. Yet dementia and other forms of neurodegeneration are frequently heterogeneous in clinical presentation, age of onset, disease duration and neuropathological profile, confounding dissection of nuanced pathological mechanisms.

1.6.1 DNA Methylation and Amyotrophic Lateral Sclerosis

Amyotrophic Lateral Sclerosis (ALS) is a complex and variable neurodegenerative disorder characterised by the progressive death of motor neurons (Zarei *et al.*, 2015). As with many neurodegenerative diseases, almost 95% of cases are sporadic and median age of onset is 64. ALS is incurable and disease duration is typically 2-3 years, although some individuals such as the late Stephen Hawking have been known to survive for much longer after initial diagnosis.

Several genes are associated with both familial and sporadic ALS and of these *C9orf72* is the most common (Dejesus-hernandez *et al.*, 2011; Renton *et al.*, 2011). A hexanucleotide expansion (GGGGCC) in this region is associated with disease development, and profiling of blood-derived DNA revealed an association with expansion length, shorter disease duration, and hypermethylation of a CpG island adjacent to the expansion (Xi *et al.*, 2013). This is remarkably similar to

observations that in peripheral blood mononuclear cell-derived DNA the length of a GAA expansion in *FXN*, the most common cause of Friedreich's Ataxia (FA), positively correlates with upstream hypermethylation and downstream hypomethylation. Upstream hypermethylation in turn inversely correlated with age of onset and *FXN* expression, while downstream hypomethylation correlated positively with earlier age of onset (Evans-Galea *et al.*, 2012). Because of the difficulty of accurately sequencing the length of such repeat expansions, altered DNA methylation at the expansion boundary was thus a useful prognostic marker in FA and is potentially useful in ALS. Similarly, Nuclear Magnetic Resonance Imaging of ALS patient brains identified an association with increased *C9orf72* promoter methylation in blood and reduced grey matter loss over time (McMillan *et al.*, 2015).

As DNA methylation varies between cell types, studies of leukocyte-specific methylation in whole blood have also been used as a proxy measure of leukocyte proportions in monozygotic twins discordant for ALS. Researchers found that over the disease course regulatory T-cells are gradually depleted and macrophage populations increase, which in combination leads to increased production of the inflammatory cytokines IL-1, IL-6 and TNF-α (Lam *et al.*, 2016). This reduction of regulatory T-cells mirrors findings from ALS model mouse research, and further investigation showed that the demethylated region specific to regulatory T-cells in the first intron of FOXP3 was hypermethylated in this cell population in rapidly progressive ALS patients (Beers *et al.*, 2017). This dysregulation of cytokine production in the periphery is particularly notable as motor neuron axons are outside the blood-brain barrier, and thus particularly vulnerable to aberrant and cytotoxic leukocyte activity. Findings such as these raise the possibility of epigenetic therapy, and follow-up work by Beers *et al*. showed that *in vitro* expansion of regulatory T-cell populations could restore normal inhibition of effector T-cells, suggesting extraction, expansion and re-transfusion of ALS patient regulatory T-cells may have a therapeutically effect (Beers *et al.*, 2017).

1.6.2 DNA Methylation and Parkinson's Disease

Parkinson's Disease (PD) is the second most common neurodegenerative disease and is clinically characterised by rigidity and tremor, followed by psychological symptoms and cognitive decline. The median age of onset is 60 and risk increases with age, while median disease duration is 15 years (Lee and Gilbert, 2016). Males are between 1.5 and 4 times more likely to develop PD, potentially in part due to the neuroprotective role of oestrogen in dopaminergic neurons which are particularly

vulnerable in PD (Gillies *et al.*, 2014). There are multiple subtypes of PD, each with a wide range of clinical presentations and trajectories, making management of the disease difficult.

PD is also associated with several genes, although mutations in these genes are not completely penetrant and so a complex array of factors are thought to underlie pathogenesis. SNCA*,* also known as alpha-synuclein, is heavily deposited in post-mortem PD brain and so has been a primary focus of research in molecular and genetic studies of PD. Aggregated deposits such as these “Lewy Bodies” are observed in many dementias and proteopathic seeding and protein dysmetabolism are thought to underlie many neurodegenerative diseases. *SNCA* expression was found to be elevated in the substantia nigra of PD patient brains (Gründemann *et al.*, 2008), and a CpG island overlapping the promoter was found to be hypomethylated in PD patient substantia nigra, putamen and cortex-derived DNA (Jowaed *et al.*, 2010). It is possible that increased expression of a protein prone to aggregation and deposition could over time be sufficient to initiate pathogenesis, in which case demethylation of the promoter region may be a necessary step during the early stage of disease. Demethylation and upregulation of *SNCA* in HEK 293 cells was shown to be induced by dopamine (Matsumoto *et al.*, 2010). While there are obvious differences between HEK 293 cells and neurons, these data do suggest an explanation for how dopaminergic neurons in the substantia nigra may be particularly vulnerable during PD pathogenesis. Increased expression of *SCNA* initiates a positive feedback loop: SCNA sequesters DNMT1 in the cytoplasm, excluding it from the nucleus and resulting in a 30% global reduction of DNA methylation in PD frontal cortex-derived DNA (Desplats *et al.*, 2011). Given that neurons are post-mitotic, it is likely that this decrease in methylation is specific to a non-neuronal cell type which loses DNA methylation over successive mitotic generation.

Methylation potential has been proposed as a prognostic and diagnostic marker in PD. The plasma SAM/SAH ratio is a measure of the physiological capacity to regulate DNA methylation, with higher values indicating greater capacity. This ratio correlates positively with DemTect score, a measure of cognitive function (Kalbe *et al.*, 2004), and negatively with platelet SCNA levels (Obeid *et al.*, 2009). This is particularly relevant within the context of PD as a common prodrug therapeutic, L-Dopa, must be methylated to become active. This process depletes cerebral SAM by between 36-76% in PD patients receiving L-Dopa. Consequently, these has been speculation that supplementing PD patient diets with folate and vitamin B12 may

prevent hyperhomocysteinaemia and potentially delay disease progression (Müller, Woitalla and Kuhn, 2003). Whether there is a connection between SAM/SAH dysmetabolism and disease severity or rate of progression remains uninvestigated and altered DNA methylation profiles in patients being treated with L-Dopa may be worth studying.

1.6.3 DNA Methylation and Alzheimer's Disease

Alzheimer's Disease (AD) is the most common form of dementia. Risk increases with age, and while a number of dominant causative mutations can cause familial forms of the disease, over 95% of cases are late onset Alzheimer's Disease (LOAD) (Lu *et al.*, 2013). Some cases of LOAD may not be sporadic in the purest sense as a common and potent genetic risk factor has been identified, namely the apolipoprotein 4 allele *APOE4* which is enriched more than two-fold in LOAD patients compared to the general population (Saunders *et al.*, 1993; van der Flier *et al.*, 2011). Study of familial AD has revealed mutations in genes which handle metabolism of the amyloid precursor protein (APP). Neuropathological hallmarks of Alzheimer's disease include deposition of aggregated forms of amyloid-β, a metabolite of APP, and deposition of hyperphosphorylated tau, a microtubule associated protein, into structures respectively named plaques and tangles. These two proteins have been studied extensively and while multiple biological processes are known to be involved in AD pathology, pathogenesis remains incompletely understood as non-demented individuals can exhibit AD-like pathology without developing clinical symptoms (Lue *et al.*, 1996).

Early studies of DNA methylation in AD showed decreased immunoreactivity of 5mC in neurons and glia of the post-mortem entorhinal cortex, an area of the brain affected at the start of AD pathogenesis, while neuronal DNMT1 immunoreactivity was also found to be reduced (Mastrianni *et al.*, 2003). Interestingly this decrease in 5mC correlates with the abundance of neurofibrillary tangles, deposition of which correlates well with the development of clinical symptoms. Hyperphosphorylation of tau was also shown to be dependent on the availability of SAM, as the tau phosphotase PP2A is activated by methylation at L309, thus a decreased SAM/SAH ratio leads to tau hyperphosphorylation in N2A cells and mouse brain (Sontag *et al.*, 2007). A decreased SAM/SAH ratio has also been found to result in *PSEN1* and *BACE1* promoter demethylation and upregulation, which in turn shifts APP metabolism towards producing greater fractions of amyloid-β (Fuso *et al.*, 2005).

SAM is a methyl donor in many biological reactions, so to what extent dysregulation of DNA methylation contributes to this imbalance is unclear.

DNA derived from frontal cortex and hippocampal tissue is not differentially methylated at genes known to harbour causative mutations in AD, such as *APP* and *PSEN1* (Barrachina and Ferrer, 2009). However, a *DNMT3B* haplotype (rs998382, rs2424913) was found to increase risk of developing AD to an odds ratio of 3.03 (Pezzi *et al.*, 2014), and subsequently shown to interact with the *APOE4* allele to synergistically increase risk to an odds ratio of 11.13, almost double the expected ratio of both haplotypes combined (de Bem *et al.*, 2016). The precise relationship between the two haplotypes and potential of increased AD risk is unclear.

As well as differential methylation at genes known to cause or predispose towards AD, with the advent of genome-wide methylome profiling dozens of genes have also been shown to exhibit differential methylation in AD in various tissues, as reviewed by Qazi *et al*. (Qazi *et al.*, 2018). Reflective of the broad range of cellular dysfunction associated with AD, these genes relate to pathways such as inflammation and oxidative stress (*COX-2, NGB, NF-κB, UQCRC1*), cytoskeletal and cellular morphology (*SYP, SORBS3, SPTBN4, ANK1, CDH23*), and neurotrophic activity (*BDNF, CREB, CREB5, TBXAR2*). Of these, *ANK1* has been reproducibly identified as becoming hypermethylated and indeed dysregulated as Braak stage increases in AD cortex across multiple cohorts of patients and in several studies, the latest of which implicate this gene in other neurodegenerative diseases such as Huntington's Disease and Parkinson's Disease (De Jager *et al.*, 2014; Lunnon *et al.*, 2014; Mastroeni *et al.*, 2017; Smith *et al.*, 2019). However, while the biological functions of these genes make them good candidates for further research in the absence of functional studies their roles in disease pathogenesis and progression remain incompletely understood.

1.7 Clinical aspects of DNA methylation

It is important to note that the changes in DNA methylation observed in the neurodegenerative diseases discussed above are relatively modest, and rarely exceed 10%. However, due in part to the large changes in DNA methylation characteristic of the diseases, DNA methylation has been most successfully connected to clinical research in the context of cancer. While genome-wide alterations in DNA methylation are a hallmark of cancer, specific changes such as hypermethylation of the *MLH1* promoter is almost diagnostic of sporadic endometrial cancer (Simpkins *et al.*, 2002). Furthermore, measuring methylation at

the *MLH1* promoter can triage patients who may have developed endometrial cancer as a consequence of Lynch syndrome, where *MLH1* hypermethylation is not involved, rather than sporadic cancer more efficiently than genetic sequencing, with obvious benefits to counselling and disease management (Bruegl *et al.*, 2014).

Site-specific altered methylation was considered as a diagnostic tool in the context of breast cancer as early as 1997 and then considered as a prognostic tool in 2003 (Huang *et al.*, 1997; Chen *et al.*, 2003). With the advent of genome-wide technologies, profiling DNA methylation in different tissues from various types of cancer patients has become easier and more rewarding. Oncologists have recently used the 450K array to establish a reference dataset of 9 control neurological tissues and 82 brain tumour classes, which were able to classify sample type using a simple random forest machine learning tool (Capper *et al.*, 2018). Sensitivity and specificity of the machine learning model was 0.989 and 0.999 respectively, and in a replication cohort 977/1,104 patient samples could be matched to a tumour class, of which 838/977 classifications matched decisions made by histopathologists. The remaining 139 patient samples were further profiled in terms of DNA copy number and targeted sequencing, leading to diagnostic revisions in 129 of 139 cases. Because of the genetic heterogeneity of brain tumours, methylation analysis provides a cost-effective way to not only inform diagnosis but also potentially identify rare novel subclasses of cancer.

The extent of lymphocyte invasion into a tumour is recognised to be prognostic of patient outcome, with greater infiltration associating with improved survival (Fridman *et al.*, 2017). Using a support vector regression model, tumour-derived DNA methylation data can be deconvoluted to infer the proportions of immune cells present in a tumour biopsy, thus potentially improving patient stratification (Chakravarthy *et al.*, 2018). Given the above, it is likely that DNA methylation data will increasingly inform cancer diagnosis and prognosis, with translational potential for studies of neurodegenerative disease.

While brain biopsies are invasive and technically complex, liquid biopsies such as cerebral-spinal fluid and peripheral blood are more comfortable for the patient and easier to perform. A recent study found that, due to the large changes in DNA methylation found in cancer, methylation-dependent immunoprecipitation (MeDIP) of circulating tumour DNA from plasma has potential for cancer diagnosis (Shen *et al.*, 2018). Because of the low abundance of normal cell-free DNA, let alone that derived from tumours, bisulphite conversion is likely to destroy extracted material

and thus be refractory to analysis. MeDIP-sequencing bypasses this problem as sequences are enriched for by differences in regional methylation between case and control. Shen *et al*. found 14,716 DMRs between 24 pancreatic cancer patients and 24 matched controls, the majority of which (9,931) were hypermethylated in patients. Further work shows that this technique can be used to classify cancer type (pancreatic, breast, lung, renal, bladder, AML). Because of the uniquely dramatic changes to the methylome observed in cancer it may be that this technique is not tractable for profiling neurodegenerative diseases, but nevertheless this work is illustrative of the huge translational potential of studying DNA methylation.

The prodrug 5-azacytidine, a cytosine analogue which cannot be methylated by DNMTs, was approved by the American Food and Drug Administration in 2004 for the treatment of myelodysplastic syndromes (MDS) , a group of blood cancers, and acute myeloid leukaemia (AML) (KAMINSKAS, FARRELL and WANG, YONG-CHENG SRIDHARA, RAJESHWARI PAZDUR, 2005). Like many chemotherapeutics, 5-azacytidine interferes with DNA replication and does so by irreversibly inhibiting the action of DNMT1 (Friedman, 1981). As with any treatment which non-specifically disrupts DNA replication and thus cell division, there are many side effects from treatment with 5-azacytidine, including fever, gastrointestinal effects, headaches, dizziness, weakness and insomnia. Another analogue, 5-aza-2'-deoxycytidine, was subsequently approved in 2006 by the FDA and by the European Commission for treatment of AML in 2012 (Agency, 2013). This analogue differs from 5-azacytidine in that it specifically inhibits DNMT1, whereas 5-azacytidine can also be incorporated into RNA and inhibit RNA methyltransferases. Curiously there are to date no trials directly comparing the clinical efficacy of these drugs in AML, and a meta-analysis of both treatments revealed no significant difference between the two (Almasri *et al.*, 2015). A comparative study in AML cell lines revealed disparate effects on cell function and viability, possibly as 5-azacytidine also affects RNA processing and in fact is preferentially incorporated into RNA (Hollenbach *et al.*, 2010). The effect of these drugs on genome-wide DNA methylation is relatively unknown as studies of DNA methylation in response to drug application in vitro has focused on genes previously associated with AML, and changes where observed did not tend to correlate with altered gene expression (Estey, 2013). Cultured bone marrow samples from MDS and AML patients which were profiled on the 450K array after 5-azacytidine treatment showed widespread differential methylation at 65, 769 probes, of which 65,664 were demethylated, albeit at a modest level (median $\Delta\beta$ = -0.028) after 24 hours. However, these effects

were found to cluster strongly with patient identity rather than treatment status, suggesting the data may be confounded by non-experimental factors (Tobiasson *et al.*, 2017).

1.8 Study Rationale and Aims

As described above, study of DNA methylation in neurodegenerative disorders can not only uncover new associations between the disease and candidate genes, but also reveal new biomarkers, prognostic markers and targets for therapeutic intervention. Nuanced studies of specific genes have also correlated DNA methylation with known risk factors for the disease, such as an association with *BDNF* promoter methylation with blood C-Reactive Protein and homocysteine levels, and with *APOE4* status in AD patients, suggesting a use of *BDNF* methylation as a proxy for disease risk (Chang *et al.*, 2014). In the tauopathy Peripheral Supranuclear Palsy, a known risk haplotype at 17q21.31 acts as a methylation quantitative trait locus, with increased dosage of the risk allele correlating with degree of hypomethylation at the *MAPT* locus in both blood and brain (Li *et al.*, 2014). In addition, association of magnitude of differential methylation with disease severity has been established in AD, where Braak staging was found to correlate with degree of hypermethylation at the ANK1 locus in the entorhinal cortex (De Jager *et al.*, 2014; Lunnon *et al.*, 2014). Methylation as a proxy measure of anticipated disease severity and risk is thus also potentially useful for patient counselling and monitoring.

Except for codon 129 genotype, there are few routinely employed prognostic or risk markers for sCJD. PrP is the main candidate protein for targeted intervention, and treatments address general disease symptoms rather than specific biological pathways. Prion disease research and prion patient treatment could therefore potentially benefit from a study of the sCJD methylome. For example, changes to DNA methylation which associate with disease severity or rate of progression could serve as biomarkers for study of novel therapeutics, while changes which seem to be established before clinical decline may reflect pathogenic events.

Study of DNA methylation in sCJD blood is a more practical and useful approach than study of brain tissue for several reasons: blood is much more accessible from a clinical perspective and so has promise as a source of disease biomarkers. Moverover, sCJD is a rare disease and not all patients have post mortem examination, whereas nearly all patients donate research blood when blood is taken for genetic testing to exclude mutations in *PRNP,* resulting in a larger sample

resource for study and greater statistical power. Blood can be sampled longitudinally, allowing correlation of the methylome at various stages of disease with clinical phenotype such as disease severity and rate of decline. Additionally, while prion-infected brain tissue is an Advisory Committee for Dangerous Pathogens (ACDP) category level 3 biohazard, blood from patients diagnosed with prion disease can be handled at ACDP containment level 2 meaning it can be processed in laboratories and hospitals without the need for specialised facilities or training. Finally, blood subtypes have been extensively studied and phenotyped for decades, and bioinformatic tools to deconvolute the effects of cellular heterogeneity in blood are significantly more advanced than the equivalent for brain tissue.

There is a dearth of non-*PRNP* risk factors in sCJD and so a genome-wide approach is necessary to identify novel targets. Of the genome-wide methods considered earlier, the most cost-effective technology which permits study of enough samples to power a genome-wide study is the Illumina Infinium 450K array. This platform also benefits from extensive use in the literature, an abundance of open-access software to process and manipulate data, and good correlation with other validated technologies such as pyrosequencing (Pearson's R^2=0.967, (Dedeurwaerder *et al.*, 2011)).

The initial aim of the study was to profile genome wide methylation using 450K arrays in a control and sCJD cohort and investigate differential methylation between both groups, correlate methylation values in the sCJD group with clinical metadata, and use an array of bioinformatic tools to infer whether differential methylation might affect particular biological pathways, originate from specific leukocyte classes, overlap functional genomic features or prove useful as a diagnostic or prognostic tool as described above in the context of cancer.

In the results and discussion chapters I will describe the results of this initial stage of the research, how findings were validated and replicated using pyrosequencing and how replicated findings were tested for specificity in DNA derived from other prion disease and patients with different neurodegenerative dementias. I will then explore the tissue specificity of observed epigenetic dysregulation and its effects on physiological function. Finally the the utility of 450K data as a diagnostic tool in predicting sCJD status will be considered.

2 Methods

2.1 Discovery Study using Illumina Infinium 450K Beadchip

2.1.1 Sample Selection

Sporadic CJD samples were obtained from patients referred to the National Prion Clinic (London, UK) or the National CJD Research and Surveillance Unit (Edinburgh, UK) under a national referral system. Patients were enrolled into the National Prion Monitoring Cohort Study which entailed a visit to hospitals, hospices or patient homes by National Prion Clinic staff followed by a review of diagnosis of probable sCJD according to published World Health Organisation criteria (World Health Organisation, 2003) by senior clinicians (J. Collinge, S. Mead and/or P. Rudge). The Cohort Study allowed longitudinal monitoring of patients via further visits and telephone assessments, and the connection of each patient's genetic, clinical, and neuropathological data over the course of the disease. This eventually lead to the development of the Medical Research Council Prion Disease Scale, a measure of disease severity between 20 (healthy) and 0 (moribund) based on clinical and care milestones, which in many cases could be retrospectively applied to patient samples taken before the development of the Scale based on clinical records (Thompson *et al.*, 2013). Diagnosis of definite sCJD was confirmed postmortem by the Neuropathology Division of the National Hospital for Neurology and Neurosurgery (University College London Hospitals NHS Foundation Trust), according to the World Health Organisation criteria.

For the discovery study 114 patients enrolled in the Cohort Study with pathologically confirmed definite sCJD were selected. 112 control samples were from healthy, non-cognitively impaired elderly volunteers recruited by the Medical Research Council Prion Unit from the National Blood Service (31 samples), or from healthy non-cognitively impaired elderly volunteers recruited by Cardiff and Vale NHS Hospital (75 samples). All donors were from the UK and of white British ethnic origin. Sample demographics for this section of the work and all subsequent sections are displayed below in Table 1.

Study Stage	Group	Number of individuals	Average Age (range)	Sex (% F)	Codon 129 (%) MM:MV:VV	Average MRC Scale Score (range)
Exploratory	sCJD	114	68.0 (49-85)	50.9	46:22:32	6 (0-20)
	Control	106	69.0 (41-83)	55.7	Unknown	20
Replication	sCJD	72	59.9 (26-86)	58.3	54:23:23	4.5 (0-20)
	Control	114	78.2(61-93)	64.9	Unknown	20
Specificity	Control	114	78.2(61-93)	64.9	Unknown	20
	AD	59	71.8 (58-87)	47.5	Unknown	NA
	iCJD	18	46.4 (41-53)	11.1	27:73:00	11.7 (1-18)
	IPD	11	48.5 (38-68)	72.7	44:44:12	18.6 (13-20)
Brain	sCJD	58	68.8 (40-87)	42.9	65:25:10	0
	Control	33	74.0 (41-89)	54.2	Unknown	0

Table 1: Demographics of patient and control samples used in experiments.

The clinical and laboratory studies were approved by the local research ethics committee of the National Hospital for Neurology and Neurosurgery (University College London Hospitals NHS Foundation Trust) and Cardiff and Value Hospital (Cardiff and Vale University Health Board), and the Institute of Neurology (Department of Brain Sciences, University College London) respectively. All patient samples were obtained with written consent from patients, next of kin, carers or Independent Mental Capacity, with ethical approval provided from the Scotland A Research Ethics Committee.

Samples were selected from these two groups based on age and sex matching between control and sCJD groups. As control samples were less plentiful than sCJD samples, 56 male and 56 female donors within the typical age range of sCJD age of onset (55-85 years of age) were selected and matched with sCJD patients of equivalent age and sex. Inclusion preference was given to patients enrolled in the Cohort Study from whom multiple tissue samples had been taken (specifically brain tissue, and RNA-stabilisation tube-stored blood for study of RNA). Where available clinical data such as *PRNP* codon 129 genotype, MRC Scale score, MRC Scale slope, and age of onset were linked to patient samples for associative analysis of these data with DNA methylation. In total 114 sCJD and 106 control samples were selected for study. Group demographics are displayed in Figure 8.

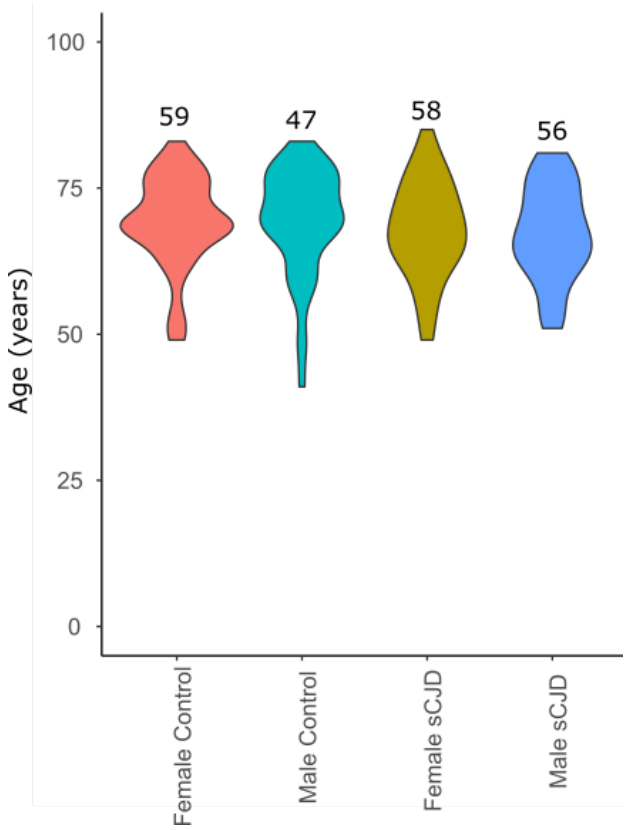


Figure 8: Violin plots of age density per sample group in the 450K array discovery study. Numbers per group are displayed above each plot.

2.1.2 Extraction of genomic DNA from blood

Extraction of genomic DNA from blood took place in a Class II MSC hood, with all used/unneeded materials being discarded into a container of 2 M NaOH and left to sterilise overnight. In order to optimise DNA extraction from whole blood, several kits and conditions were tested. These are listed below:

Zymo Quick gDNA MiniPrep Kit

Several conditions were trialled using this kit: the manufacturer's protocol (a), the manufacturer's protocol with proteinase K digestion steps (b, c), and the protocol with a proteinase digestion step and a subsequent leukocyte pelleting and wash step.

a) 200 µl blood was mixed with 800 µl genomic lysis buffer, mixed by vortexing and left to stand for 5 minutes at room temperature. This was then transferred to a purification column which was centrifuged in a benchtop microfuge for 30 seconds at 10,000 *g* before the column was washed with 200 µl DNA pre-wash buffer, centrifuged again under the same settings and

then washed with 500 µl DNA wash buffer. After another centrifugation step 50 µl DNA elution buffer (preheated to 65ºC) was incubated on the column for 5 minutes at room temperature before DNA was eluted by centrifugation at top speed for 30 seconds.

b) As (a) with additional protease K (200 µl of 2x Digestion Buffer, 20 µl protease K) treatment for 10 min at 65ºC before mixing with lysis buffer.
c) As (a) with additional protease K treatment overnight.
d) As (b) with centrifugation at maximum speed for 30 seconds, discarding of supernatant and resuspension of the pellet in 200 µl dH_20 before mixing with lysis buffer.

ZymoBeads

a) 50 µl blood was mixed with 200 µl genomic lysis buffer, before mixing with 10 µl ZymoBeads. After a 5 minute incubation at room temperature the tube was centrifuged at 1,500 *g* for 1 minute and the supernatant aspirated. The pellet was resuspended in 200 µl genomic lysis buffer before another centrifugation and aspiration. The pellet was resuspended in 200 µl DNA pre-wash buffer and transferred to a fresh tube before centrifugation and aspiration of the supernatant. The pellet was then resuspended in 500 µl g-DNA wash buffer before a final centrifugation and aspiration. Elution buffer was preheated to 65 ºC, 50 µl were mixed with the pellet and incubated for 5 min before centrifugation at 10,000 *g* for 1 minute. The supernatant containing purified DNA was collected.
b) As (a) but volume of blood increased to 200 µl and genomic lysis buffer to 800 µl.

PAXgene Blood DNA Kit

Samples were processed according to Qiagen protocol but all volumes reduced by 1/5 to conserve sample volume. 1 ml buffer BG was mixed with 10 µl PreAnalytiX protease. 1.7 ml blood was added to 5 ml buffer BG1 and mixed by inversion. Samples were centrifuged for 5 minutes at 2,500 *g*, after which the supernatant was discarded and 1 ml buffer BG2 was added and mixed by vortexing for 5 seconds. After centrifugation at 2,500 *g* for 2 minutes the supernatant was discarded and 1 ml buffer BG3/PreAnalytiX protease was mixed with the pellet by vortexing for 20 seconds. Samples were incubated at 65ºC for 10 minutes before a brief vortex and addition of 1 ml 100% isopropanol. Samples were inverted 20 times before centrifugation at 2,500 *g* for 3 minutes. The supernatant was discarded and residual

isopropanol was mopped up using a cotton bud. The pellet was washed with 1 ml 70% ethanol before centrifugation at 2,500 *g* for 3 minutes. The supernatant was discarded and tubes were left inverted for 10 minutes at room temperature, before precipitated DNA was dissolved in 200 µl buffer BG4 at 65ºC for one hour, followed by overnight incubation at room temperature.

Illustra Nucleon BACC3 Genomic DNA Extraction Kit

8 ml of blood was poured into a 50 ml falcon tube containing solution A (30 ml dH_20, 10 ml reagent A) and mixed for 4 minutes. The tube was centrifuged at 1300 *g* for 5 minutes and supernatant discarded. The pellet was resuspended in 2 ml reagent B and transferred to a 15 ml falcon tube to which 500 µl Sodium perchlorate was added and mixed by shaking, followed by addition of 2 ml chloroform and more mixing by shaking. 300 µl resin beads were added to the suspension without shaking which was the centrifuged at 1300 *g* for 3 minutes. The supernatant was decanted into a fresh 15 ml falcon tube and DNA precipitated by addition of 2 ml ethanol. DNA was spooled onto the tip of a heat-sealed glass Pasteur pipette and left to air dry before resuspension in 500 µl Tris-EDTA buffer.

2.1.3 DNA Quality Control

Concentration and integrity of extracted genomic DNA was measured by TapeStation (Agilent), or a QuBit 2.0 fluorometer (Thermo Fisher) and agarose gel electrophoresis. A Clean and Concentrator kit (Zymo) was used to clean and/or concentrate fragmented (TapeStation DIN value <0.7) or dilute samples ([DNA] < 25 $ngµl^{-1}$). Sample concentrations were adjusted to 25 $ngµl^{-1}$.

2.1.4 Bisulphite Conversion

500 ng of DNA in 20 µl was bisulphite converted in a Veriti thermocycler (Thermo Fisher) overnight using an EZ DNA Methylation-Gold kit (Zymo) according to the manufacturer's protocol. Briefly, 130 µl CT conversion reagent was mixed with 20 µl DNA and incubated in a thermal cycler for 10 min at 98ºC, 150 min at 64 ºC, then maintained overnight at 4ºC.The following day converted samples were loaded into spin columns, washed and then desulphonated. After two more washes column-bound bisulphite DNA was eluted in two volumes of 15 µl elution buffer then stored at -80ºC.

2.1.5 Infinium Beadchip Hybridisation

Bisulphite DNA was sent to UCL genomics for array hybridisation in accordance with the Infinium HD Assay 15019519_B, 2011 (Illumina, 2015). Briefly, bisulphite DNA was amplified, enzymatically fragmented, precipitated and resuspended before hybridisation. Samples were hybridised to Illumina Infinium 450K beadchips in two batches of 12 (July 2015, August 2015) and three batches of 96 (August 2015, December 2015, February 2016). Each batch of samples was spatially randomised across a 96-well plate to prevent segregation of CJD and control samples between separate beadchips, which might predispose the data to batch effects. A total of 112 controls, 116 sCJD samples and 20 vCJD samples were hybridised, with each batch of 12 or 96 samples containing a methylated, unmethylated and commercial leukocyte DNA standard (Zymo, Zymo, AMSBIO). Beadchips were then washed, stained and dried before imaging using an iScan (Illumina). IDAT files were generated and delivered from UCL Genomics via dropbox.

2.2 Data Analysis of Illumina Infinium 450K Beadchip Study

2.2.1 Case-Control Study Using RnBeads

RnBeads version 1.6.1 (Assenov *et al.*, 2014) and dependent packages (RnBeads.hg38, biomaRt, GenomicRanges, GEOquery, ggbio, GOstats, Gviz, IlluminaHumanMethylation450kmanifest, methylumi, rtracklayer, sva, RefFreeEWAS, MCLUST, hexbin) were downloaded from *https://bioconductor.org/biocLite.R* and loaded into R version 3.2.4 (Team, 2016). IDAT files were deposited into a working directory subfolder named “data” and a second empty subfolder was created titled “analysis”: these were assigned to the variables “idat.dir” and “report.dir” respectively. The manifest .csv file containing sample IDs, technical data and patient metadata was placed in the working directory and assigned to the variable “sample annotation”. Case-control analysis settings were configured by Dr. Holger Hummerich and performed using the hg38 build of the human genome, default preprocessing, QC, normalisation and exploratory functions, and RefFreeEWAS to correct for cellular heterogeneity (Houseman, Molitor and Marsit, 2014), as well as inclusion of age and sex as covariates. Multiple testing was compensated for using Benjamini-Hochberg correction. The code used to define these parameters is displayed below in Box 1. The pipeline was run on an independent server using 20 core parallel-processing and completed analysis after 30 hours and 18 minutes.

```
library(RnBeads)


rnb.options(assembly = "hg38",
filtering.sex.chromosomes.removal=TRUE,
identifiers.column = "Sample_ID", import = TRUE,
preprocessing = TRUE, qc = TRUE, exploratory = TRUE,
export.to.bed = FALSE, export.to.trackhub = NULL,
exploratory.region.profiles = character(0),
differential = TRUE,
differential.comparison.columns = "Sample_Group",
differential.site.test.method="refFreeEWAS",
differential.adjustment.sva = TRUE,
covariate.adjustment.columns = c("Gender","Age"),
filtering.cross.reactive = TRUE,
inference.sva.num.method = "leek",
differential.adjustment.celltype = TRUE, inference =
TRUE, inference.targets.sva = c("Sample_Group"),
differential.enrichment = TRUE, logging.disk = TRUE,
enforce.memory.management=TRUE)



rnb.run.analysis(dir.reports = report.dir,
sample.sheet = sample.annotation, data.dir =
idat.dir, data.type = "infinium.idat.dir",
save.rdata = TRUE)
```

Box 1: rnb.options used to define parameters for the rnb.run.analysis function. Report and data directories were defined as "report.dir" and "idat.dir", while sample sheet location was defined as "sample.annotation".

2.2.2 Case-Control Study Using ChAMP

ChAMP version 2.10.2 (Morris *et al.*, 2014) and limma version 3.36.5 (Ritchie *et al.*, 2015) were downloaded from *https://bioconductor.org/biocLite.R* and loaded into R version 3.3.4. IDAT files and manifest .csv file were placed directly into the working directory. These were processed and corrected as per documentation (Data import, QC, BMIQ normalisation, SVD, batch correction using ComBat (Johnson, Li and Rabinovic, 2007), Houseman or EpiDISH (version 1.2.0) reference-based celltype correction (Houseman, Molitor and Marsit, 2014; Teschendorff and Zheng, 2017)) on a local dual-core processor machine with 32Gb RAM. Limma was then used to fit the corrected data to the following linear regression model:

Corrected Beta ~ Sample_Group + Age + Sex

The code used to complete this pipeline is displayed below in Box 2.

```
library(ChAMP)
library(limma)
library(EpiDISH)
myLoad<-champ.load()
champ.QC()
myNorm<-champ.norm()
champ.SVD()
myCombat<-champ.runCombat(batchname=c("Slide","Array")
champ.SVD()

#Houseman Correction
myRefbase<-champ.refbase(beta=myCombat)
CorrectedBeta<-myRefbase$CorrectedBeta*100

#EpiDISH Correction
ref.m <- centDHSbloodDMC.m[,1:6]
out.l <- epidish(myCombat, ref.m, method = "RPC")
message("Mean value for each estimated Cell Proportion:")
print(colMeans(out.l$estF))cellFrac<-as.matrix(out.l$estF)
message(names(which.min(colMeans(cellFrac))),
        " has the smallest cell proportion, all other cell
proportions will be corrected by linear regression.")
lm.o <- lm(t(myCombat) ~ cellFrac[,-
1*which.min(colMeans(cellFrac))])
EpiBeta <- t(lm.o$res)+rowMeans(beta)
CorrectedBeta<-EpiBeta*100;
message("Correction for cellular heterogeneity between
samples is complete.\n")

#CorrectedBeta used in final linear model
Mod <- model.matrix(~0 + Sample_Group + Age + Sex,
data = myLoad$pd)
colnames(Mod)[1:2] <-
levels(factor(myLoad$pd$Sample_Group))
Contrast.matrix <- makeContrasts(sCJD - CTRL,
levels = Mod)
Fit <- lmFit(CorrectedBeta, Mod)
Fit <- contrasts.fit(Fit, Contrast.matrix)
Fit <- eBayes(Fit)
BonferroniLimmaLong <- topTable(Fit,
number = nrow(CorrectedBeta), coef = 1,
adjust.method="bonferroni",p.value = 1.00 )
```

Box 2: ChAMP pipeline run in the working directory containing sample manifest.csv and .idat files. Either Houseman or EpiDISH correction is used depending on the analysis

Differential methylation (Δβ) values at CpG sites, *p* values and Bonferroni adjusted *p* values were extracted, and association between corrected beta values at these sites with MRC Scale Score and Slope (Thompson *et al.*, 2013; Mead *et al.*, 2016) and *PRNP* codon 129 genotype was tested for using bivariate Pearson's correlation in SPSS version 25 (Corp., 2017).

Bumphunter (Jaffe *et al.*, 2012) was used within ChAMP to identify Differentially Methylated Regions (DMRs) from corrected data using default configuration.

2.2.3 Inference of Signal Origin Using eFORGE

To infer possible cell or tissue type specificity of results, the list of cg codes (eg. cg10636246) corresponding to significantly Differentially Methylated Positions (DMPs) were uploaded to eFORGE version 1.2 (Breeze *et al.*, 2016). Site-specific associations with functional elements and modifications were investigated by associating the list with DNAse1 hypersensitive sites (DHSs) and Histone 3 (H3) marks using default parameters.

2.2.4 Study of Differentially Variable Methylation Sites Using iEVORA

To assess differential variability of methylation at CpG sites between sCJD and control samples, the iEVORA algorithm (Teschendorff *et al.*, 2012, 2016) was downloaded from Teschendorff *et al*. 2016 (Supplementary Software 1) and applied to 450K corrected beta data in R version 3.3.4.

2.2.5 Study of Genomic and CpG Feature Distribution Using GOLDMINE

ChIP-Seq data from T-Cells, B-Cells, Monocytes, Neurotrophils and Natural Killer Cells (E04_15, E031_15, E062_15, E030_15, E046_15 respectively) were downloaded from the Roadmap resource (Roadmap Epigenomics Consortium *et al.*, 2015) (http://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/coreMarks/jointModel/final/). These were aligned to the hg19 build of the human genome, which was labelled with coding gene regions, Dnase Hypersensitive Sites, Transcription Factor Binding Sites and Enhancer Sites. DMRs and DMPs were aligned to the genome and genomic and CpG feature distribution of each group was called.

2.2.6 Study of Common Motifs Amongst Differentially Methylated Regions and Positions

DMR sequences were input to MEME-Suite in FASTA format (Bailey *et al.*, 2009). Identified motifs were probed for common transcription factor binding using MEME-Suite's TOMTOM tool under default settings.

2.2.7 Enrichment Analysis of Differentially Methylated Transcription Factor binding sites and Differentially Methylated Genes

Transcription factors found to be associated with DMRs (TOMTOM) and DMPs (Goldmine), and genes overlapping DMRS and DMPs respectively were uploaded separately to Enrichr (E. Y. Chen *et al.*, 2013; Kuleshov *et al.*, 2016) which associated uploaded genes with gene ontology and biological pathways using publicly available repositories.

2.2.8 Machine Learning Classification of 450K Dataset

CaRET version 6.0-81 (Kuhn, 2008) was used as a wrapper for machine classification of disease phenotype in a blinded subset of 450K array data. The 500 and 5000 most significantly altered probes were screened for pairwise correlation and redundant probes were discarded. Beta values at remaining loci from 105 sCJD patient and 105 controls were united, randomised and partitioned into a training set (75% of samples) and a test set (25% of samples). 10-fold 10-repeat Recursive Feature Elimination (RFE) was used to select features for model training from the top 500 probes (ranked by significance of difference between sCJD and control groups) with iterative removal of one probe at a time.

RFE-selected probes, probes corresponding to the 38 DMPs and the top 5000 probes as ranked by significance were all used to train stochastic gradient boosting machine (gbm) models and averaged neural network models (avNNet) using 10-fold 10-repeat crossvalidation in all instances. gbm models were tuned by adjusting the number of boosting iterations and tree depth, while avNNet models were tuned by adjusting number of hidden neuronal layers and weight decay value.

Finally a stack of random forest (ranger), gradient boosting machine (gbm), multivariate adaptive regression splines (earth), support vector machine (svmRadial), and averaged neural network (avNNet) models were generated, compared for accuracy and then combined using generalise linear regression to form a meta-model using the caretStack function.

2.3 Validation and Replication Using Pyrosequencing

2.3.1 Sample Selection

A second age and sex-matched cohort of 72 sCJD and 114 control samples was assembled according to the same methods and criteria as described in Section 2.1.1. Sample demographics are displayed below in Figure 9.

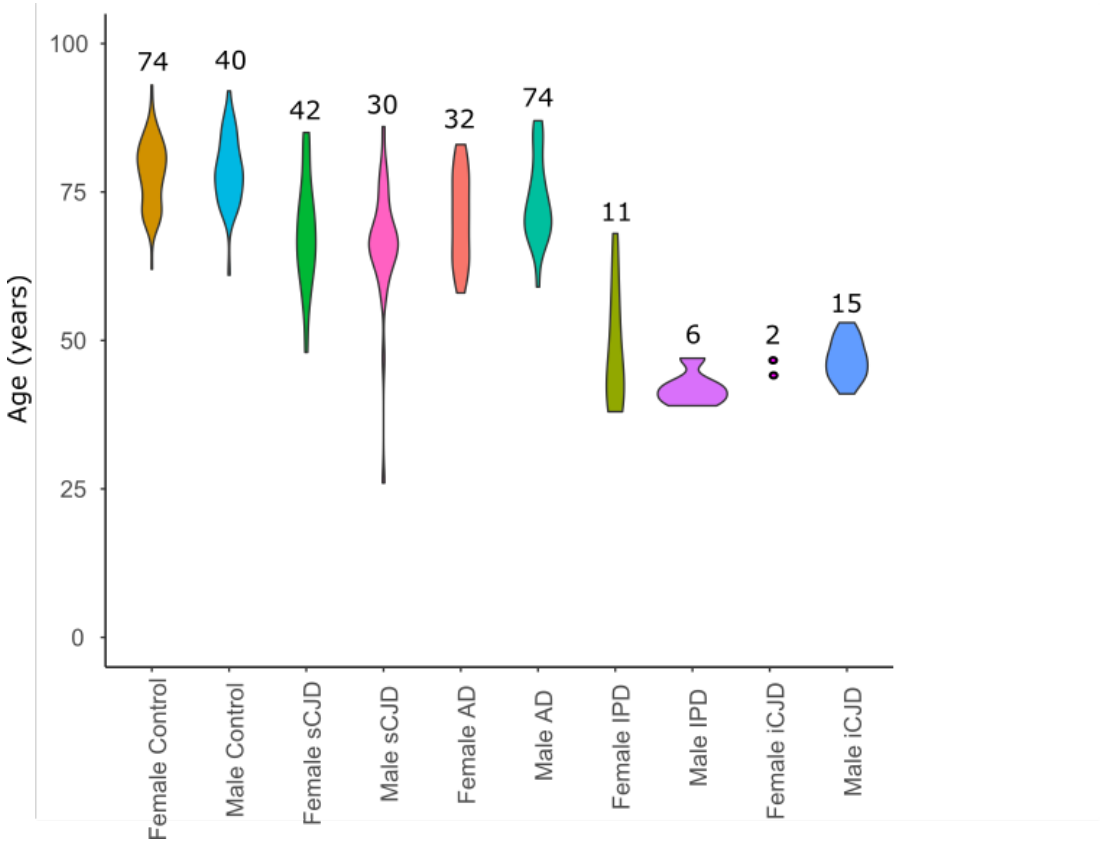


Figure 9: Violin plots of age density per sample group involved in replication and test for disease specificity. Numbers per group are displayed above each plot

2.3.2 Bisulphite PCR for Pyrosequencing

Sites for validation and replication were selected based on statistical significance, effect size and biological function. Pyrosequencing assays overlapping cg03546163 (*FKBP5*), cg00052684 (*FKBP5*), cg25114611 (*FKBP5*), cg17714703 (*UHRF1*), cg02481950 (*METTL9*), cg02448796 (*KCNAB2*), cg04286737 (*PRNP*), cg05001044 (*MIR1977*), cg11823178 (*ANK1*), cg14008593 (*TTLL10*), cg10636246 (*AIM2*), and cg17515347 (*AIM2*) were automatically designed using PyroMark Assay Design version 2.0 (Qiagen), and manually adjusted to fit the following criteria:

i) Primers do not overlap CpG sites
ii) Forward and Reverse primer length does not exceed 25 nucleotides
iii) Sequencing primer's annealing temperature is 40^{o}C

Primers were then checked for specificity using BiSearch (Arányi *et al.*, 2006). Forward, reverse and sequencing primers were ordered through EuroFins.

Samples were bisulphite converted using an EZ-96 DNA Methylation-Gold kit (Zymo) according to manufacturer's instructions. Assay PCR reactions were optimised using bisulphite converted DNA and a thermocycler configuration with a variable extension step temperature (Table 2, Table 3). Temperatures which produced the greatest amount of specific product were used in subsequent reactions. PCR was performed using HOTFIREPol kits (Solis Biodyne).

Bisulphite PCR

Mastermix	Per 2 µl DNA
10x Buffer B1	2 µl
10 mM dNTPs	0.2 µl
25 mM $MgCL_2$	1.32 µl
H_2O	13.28 µl
10 µM Primers	1 µl
HOT FIRE pol	0.2 µl

Table 2: Reaction mixture for bisulphite PCR using HOTFIREPol reagents

Stage	Temperature	Duration	Number of cycles
Stage I	95°C	15 min	X 1
Stage II	95°C	45 s	
	52-62°C	45 s	X 35
	72°C	3 min	
Stage III	72 °C	10 min	X 1
Total		3h 3 min	

Table 3: Thermocycler settings for bisulphite PCR featuring a 52-62°C variable extension step in increments of 2 °C across 6 sets of 2 columns of a 96 well plate.

Optimum PCR elongation conditions of 56°C were determined for cg03546163, cg00052684, cg25114611, cg17714703, cg17515347, cg11823178, 58 °C for cg02481950, cg02448796, cg04286737 and 62°C determined for cg10636246, cg05001044, cg02481950. Primer sequences are detailed below in Table XX.

CpG	Gene	Forward	Reverse	Sequencing	PCR Extension Temperature (°C)
cg02481950	METTL9	GTGATTTAGAGTAGTGGGG	Biotin-CTAAATACCATCAACCCTATTATCA	GTAATGGTAGTGAGATT	56
cg10626346	AIM2	AAGTGTTGTTAGTAGGGTAGT	Biotin-AAATACCTACACTCCTCCCTTAAA	GTATAAGAGAATTTGAAAAT	58
cg05001044	MIR1977	TGGATATATAGGTATGGTTTGAGTTATG	Biotin-AATCCTAAAACTCAAAACACTAC	TATTAATTGGTTTTTTAGGG	56
cg03546163	FKBP5	TGGTTTTTTGATGTTTAGTGTGG	Biotin-CCAATTCCTTTCAACTATTTACATA	GATAGGTTGAATAATAATTT	56
cg04286737	PRNP	TTTGGGTTATGTTTATGTGAGTGT	Biotin-CCTCCCTATTACACTAAATCAACT	GGTAATTAGGAGGTTTGGAA	60
cg11823178	ANK1	AGAAGAGAGATTGGAGAGAGAGTTTATA	Biotin-CTCCTTCCCTTCCTTTCCTAAATC	GGGTTTTTTTATAGTTATTTT	60
cg17515347	AIM2	GGAGGGGTTTTTTTGGGAAATG	Biotin-AAAAAATTCCCACTCCCATCATAC	TTATTTTTTTTTTGATTAAAGATG	60
cg02448796	KCNAB2	TTTGGGTGTTTTTTGTTTGTATTATAA	Biotin-CACTACCTTTCCCTATAACTAT	TTTGGGGATAGAGTA	56
cg00052684	FKBP5	GAGAATAGGTAGTGTTTAGGGATTGAG	Biotin-ACTCATTCTACTTAACATCACCACAA	TATTAGAGATAATGTATGTGAA	60
cg25114611	FKBP5	AGGGTTATTTTTGGGGGTGG	Biotin-CTCCCCTAAACCCACTCTCA	TTTGGAGAGTGTTAG	60
cg17714703	UHRF1	Biotin-TGGAGAATTTGGGGAGTT	CTCTTTATAAAATAAAACCTTCTATACT	CCTACCTCTTCTCTA	58

Table 4: Primer sequences and annealing temperatures for bisulphite PCR reactions used in replication assays

2.3.3 Pyrosequencing

A PyroMark Q24 was used for result validation and a PyroMark Q96 was used for replication (Qiagen). 20 µl of each PCR reaction product was added to 60 µl sepharose bead solution (2 µl sepharose beads, 40 µl pyrosequencing binding buffer, 18 µl H_2O) in a 96 well plate. The plate was sealed and vortexed at 1500 rpm using an IKA MS 3 vortexer with a 96-well plate adaptor (IKA, UK) for at least 5 minutes and no more than 10 minutes. 25 µl 0.3 µM sequencing primer was added to each well of a PyroMark Q24 or Q96 plate and placed on the vacuum workstation. Bead-immobilised DNA was then passaged through a vacuum workstation (Qiagen) before mixing with sequencing primer in the PyroMark plate for 10 seconds. The plate was heated at 80ºC for 2 minutes, allowed to cool for 5 minutes and then inserted into the pyrosequencer. A PyroMark cartridge was loaded with PyroMark reagents (Qiagen) according to volumes computed by PyroMark assay design and inserted into the pyrosequencer. Nucleotide dispensation order was generated by inputting the target sequence before and after bisulphite conversion. Where possible a non-CpG cytosine was selected as a bisulphite conversion control. 93 samples were assayed per pyrosequencing run as well as methylated and unmethylated standards and a zero-template control.

2.4 Testing Disease Specificity of Replicated Differentially Methylated Sites Using Pyrosequencing

A third cohort of 59 Alzheimer's Disease (AD), 18 Iatrogenic CJD (iCJD) and 11 Inherited Prion Disease (IPD) patients was assembled from the National Prion Clinic and Dementia Research Centre's register of samples according to the same criteria as described in Section 2.1.1. AD patient samples were collected with consent by the Dementia Research Centre in compliance with the local research ethics committee of the National Hospital for Neurology and Neurosurgery (University College London Hospitals NHS Foundation Trust).

Due to differences in age of onset, disease duration and prevalence per sex in iCJD and IPD, in this instance age and sex matching was less exact between disease and control groups. Sample demographics are displayed above in Figure 9. Samples were assayed for methylation at cg03546163 (*FKBP5*), cg17515347 (*AIM2*), cg10636246 (*AIM2*), cg17714703 (*UHRF1*), cg02481950 (*METTL9*) and cg11823178 (*ANK1*).

2.5 Testing Tissue Specificity of Replicated Differentially Methylated Sites Using Pyrosequencing

A fourth cohort of frontal-cortex derived DNA samples from 33 non-prion and 58 sCJD patients was assembled from the National Prion Clinic's register of samples as described in Section 2.2.1. Non-sCJD patient frontal cortex tissue was obtained from Cambridge Brain Bank (Cambridge, UK), while sCJD patient frontal cortex tissue was obtained from the Neuropathology Division of the National Hospital for Neurology and Neurosurgery (University College London Hospitals NHS Foundation Trust). Samples were assayed for methylation at cg03546163 (*FKBP5*), cg17515347 (*AIM2*), cg10636246 (*AIM2*), cg04286737 (*PRNP*), cg17714703 (*UHRF1*), cg02481950 (*METTL9*), and cg11823178 (*ANK1*).

2.6 Testing Leukocyte Specificity of Replicated Differentially Methylated Sites Using Magnet Assisted Cell Sorting and Pyrosequencing

2.6.1 Sample Selection

Fresh whole blood was collected prospectively from 7 sCJD patients and 13 healthy relatives/spouses of patients during home visits to patients by National Prion Clinic staff (T. H. Mok, A. Nihat, H. Odd, P. Rudge,), with research and ethical consent being obtained as described in Section 2.1.1. Sample demographics are shown in Figure 10 below.

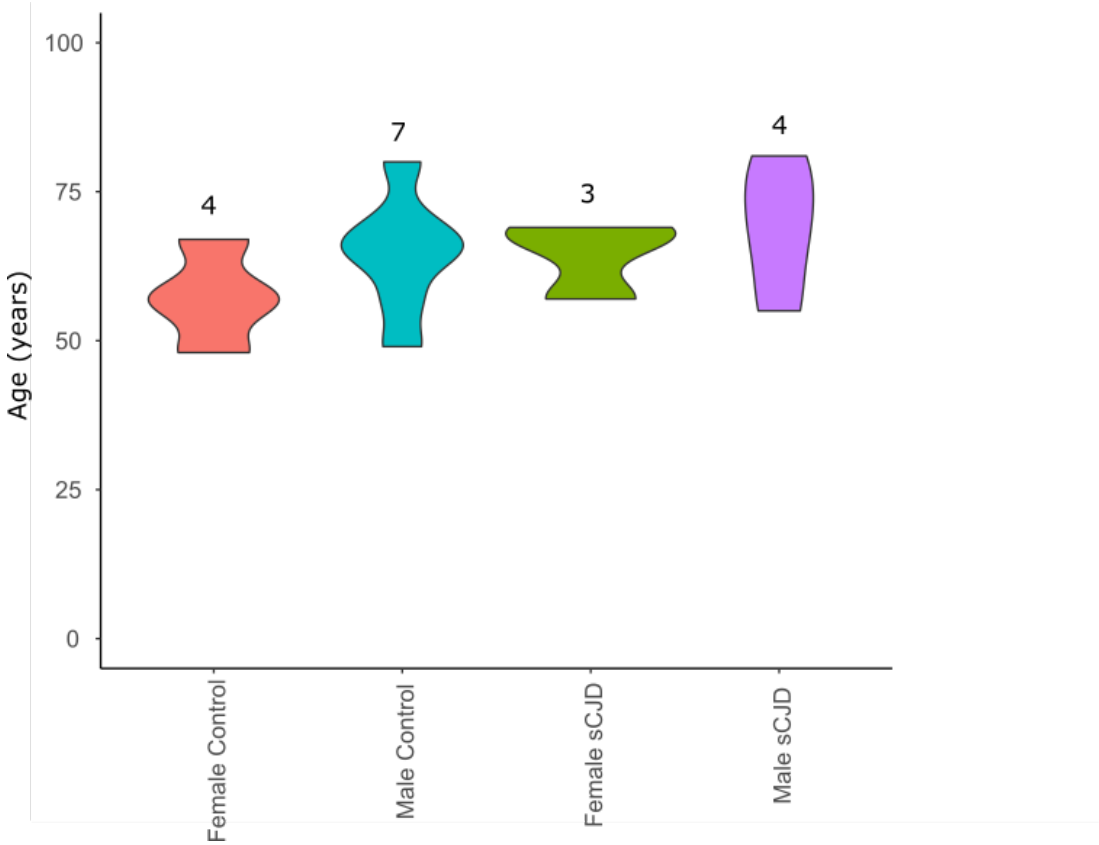


Figure 10: Violin plots of age density per sample group from the Magnet-Assisted Cell Sorting experiment cohort. Numbers per group are displayed above each plot.

2.6.2 Magnet Assisted Cell Sorting

MACS was used to enrich cell fractions for T-Cells, B-Cells, Neutrophils and Monocytes using antibodies raised against CD4+, CD15+, CD19+ and CD14+ respectively. The MACS (Miltenyi) protocol was optimized using blood samples donated from non-MRC staff and students at UCL's Institute of Neurology, with consent provided according to the criteria of the research ethics committee of the National Hospital for Neurology and Neurosurgery (University College London Hospitals NHS Foundation Trust). This allowed comparison of the effects of filtering blood prior to MACS and washing micro-bead conjugated cell pellet prior to column loading on final DNA yield.

Sorting took place in a Class II MSC hood, with all used/unneeded materials being discarded into a container of 2 M NaOH and left to sterilise overnight. According to the manufacturer's protocol, Whole Blood Column Elution Buffer was warmed to ambient temperature before starting MACS. Depending on blood sample volume, 1 to 1.5 ml aliquots of blood were passed through 30 µm filters into 15 ml falcon

tubes. 75 µL antibody-conjugated Microbeads per 1.5ml whole blood was added to the tube, mixed gently and incubated for 15 min on ice. The sealed tubes containing cell mixtures were then removed from the MSC hood and centrifuged at 445 *g* for 10 minutes at room temperature. Tubes were returned to the MSC hood and after aspiration of the supernatant the cell pellet was resuspended in 1 ml ice-cold separation buffer (PBS, 0.5% BSA, 2mM EDTA, pH 7.2). Labelled cell mixtures were loaded onto Whole Blood Columns set in a QuadroMACS magnetic separator and flowthrough was collected in a 15 ml falcon tube. Each column was washed three times with 3 ml separation buffer before the column was removed from the QuadroMACS separator and eluted with 5 ml Whole Blood Column Elution Buffer. Sealed tubes containing eluted cells were removed from the MSC hood and centrifuged at 445 *g* for 10 minutes at room temperature. Tubes were returned to the MSC hood and 4.7 ml supernatant was aspirated. The pellet was resuspended in the residual supernatant and 100 µl of cell suspension was added to 300 µl TriZol reagent (Zymo) and stored at -20^{o}C for downstream RNA analysis. DNA was immediately extracted from the remaining 200 µl of suspension using a Quick gDNA MiniPrep Kit (Zymo) with a 10 min proteinase K incubation step as described in Section 2.1.2.

2.7 Statistics and Graphics

Pyrosequencing data were analysed in SPSS version 25 (Corp., 2017), which was also used to correlate corrected beta values with patient metadata. All other analyses were performed in R (Team, 2016). Genome-wide inflation values (λ) were calculated by dividing the median chi-squared statistic of probe *p* values by the expected median (Devlin and Roeder, 1999; Hinrichs *et al.*, 2012). Plots were generated using ggplot2 version 3.0.0 (Wickham, 2016), or qqman version 0.1.4 (Turner, 2018) and edited in Inkscape version 0.91 (Harrington, 2005). Where shown, box plots adhere to Tukey representation (Mcgill, Tukey and Larsen, 1978), where upper and lower hinges of the box correspond to the interquartile range, the median line divides the box and two whiskers protrude to the largest value up to 1.5 times the interquartile range. Outlying points are plotted individually. Where there are significant differences between experimental groups, a black bar with asterisks denotes the Bonferroni-adjusted significance level (* = $p< 0.05$, ** = $p< 0.01$, *** = $p<0.001$).

3 Using Illumina's 450K Infinium Beadchip to investigate genome-wide DNA methylation in whole blood from sporadic prion disease patients

The principle aim of this project was to investigate differences in DNA methylation in blood both between sCJD patients and control volunteers, as well as within the sCJD patient group, to uncover molecular markers of the disease. In order to accomplish this, a new blood-derived DNA extraction protocol and bisulphite conversion protocol had to be established, as routinely used blood-derived DNA extraction protocols were high yield but low throughput. Moreover, an analysis pipeline for 450K Beadchip data would have to be selected and different settings tested to remove unreliable data and improve quality control metrics such as technical or batch variability and genome-wide significance inflation. Finally, I aimed to perform exploratory analysis and linear modelling of the data to examine the genomic distribution and potential biological consequences of differential methylation, as well as test for associations with disease traits.

3.1 Relevant Methods and Sample Demographics

Group	Number	Average Age (range)	Sex (% F)	Codon 129 (%) MM:MV:VV	Average MRC Scale Score (range)
sCJD	114	68.0 (49-85)	50.9	46:23:32	6 (0-20)
Control	106	69.0 (41-83)	55.7	Unknown	20

Table 5: Demographics of patients and controls whose samples were used in the initial discovery cohort.

3.2 Sample Preparation

3.2.1 DNA Extraction from Whole Blood

As this project would involve multiple extractions of DNA from EDTA-stored blood, I initially compared several extraction protocols as detailed in Section 2.1.2. While the Prion Unit Neurogenetics Division has an established protocol (illustra Nucleon BACC3 Genomic DNA Extraction Kit) which far exceeded any other method in terms of DNA yield as well as yield as a function of input volume, batches of samples were limited to six at a time. The PAXgene protocol had similar batch constraints and performed the worst out of all methods. Batches for all other protocols were limited by the capacity of a benchtop microfuge, allowing 24 samples to be processed per batch. DNA yield was low in general, which was the second most important limiting factor during optimisation as bisulphite conversion kits recommend 500 ng input DNA for optimal conversion and fragment recovery. During optimisation 3 EDTA-stored blood samples were processed using each method and the average yields are presented in Table 6.

Method	Proteinase K	Starting Volume (µl)	Final Volume (µl)	[DNA] ($ng\mu l^{-1}$)	DNA (ng)	DNA per starting volume ($ng\mu l^{-1}$)
ZymoBeads	No	50	35	0.3	10.5	0.21
ZymoBeads	No	200	35	0.6	21	0.105
gDNA	10 min	200	50	11.6	580	2.9
gDNA	Overnight	200	50	1	50	0.25
gDNA with pellet wash step	10 min	200	50	3.4	170	0.85
PAXgene	No	1700	200	0.05	10	0.05
Prion Unit Protocol	No	4000	500	413	206500	51.265

Table 6: Performance of ZymoBeads and gDNA kits (Zymo), PAXgene (Qiagen) and the Prion Unit's in house DNA extraction protocol.

As a function of input sample volume, protocol duration and complexity, throughput, and output quantity the Zymo Quick gDNA MiniPrep Kit with a 10-minute proteinase K digestion step was best suited for extraction of >500 ng DNA from small volumes of blood and was used for all subsequent extractions of genomic DNA from blood.

3.2.2 Bisulphite Conversion

As the quality and consistency of bisulphite conversion is essential in studies of DNA methylation I compared two conversion kits (Zymo and Qiagen), the main difference between which was bisulphite reaction incubation length (150 and 300 minutes respectively). For each method, 5 patient DNA samples, a methylated control and unmethylated control (Zymo) and a commercially-sourced leukocyte DNA sample (AMSBIO) were converted and hybridised to two Illumina 450K beadchip arrays.

Preliminary analysis in GenomeStudio showed very poor detection of CpG sites (based on signal intensity) in samples converted using the Qiagen kit (Figure 11), while Zymo bisulphite conversion kits produced DNA which was successfully probed at almost half a million CpG sites according to GenomeStudio's detection algorithm (unpublished). I therefore used the Zymo conversion kit for all subsequent experiments.

As bisulphite-converted DNA resembles RNA in terms of high uracil content and single stranded conformation, I analysed the integrity of Zymo bisulphite-converted DNA using a high-sensitivity Agilent RNA ScreenTape. Bisulphite converted DNA was found to be present in fragments between 25-2000 nucleotides long, as shown in Figure 12. As 450K array probes are approximately 100 nucleotides in length, this distribution of fragment length after conversion suggests appreciable probe target availability.

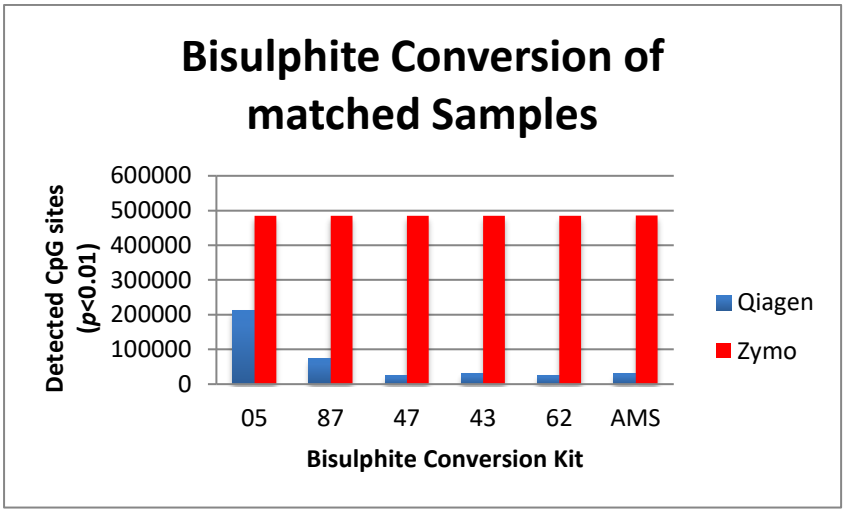


Figure 11: Detection rate of CpGs in bisulphite-converted DNA derived from Qiagen and Zymo protocols. Although the exact method is unpublished, detection p is likely calculated from iDAT file signal intensity per probe by GenomeStudio software compared to a reference null signal. X-axis labels refer to the penultimate two digits of a sample code, or commercially-sourced leukocyte DNA (AMS).

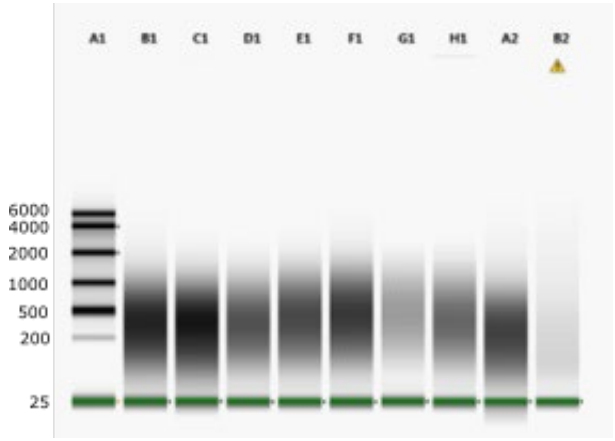


Figure 12: High-Sensitivity RNA Screentape of bisulphite conversion products of genomic DNA. Samples were derived from patients (B1-F1), unmethylated and methylated control DNA (G1 & H1), a commercially-sourced leukocyte control DNA sample (A2) and DPEC-treated water (B2).

3.3 Analysing 450K data using RnBeads

3.3.1 Pre-processing and Quality Control

RnBeads is a one-step, powerful analytical pipeline for the analysis of 450K data capable of automatically generating thousands of tables, high-resolution plots and web browser-based reports. I initially chose RnBeads for analysis because of its relatively simple start-to-finish design: options were defined as described in Section 2.2.1 and the entire analysis was run as one customised function on a remote server using 20 cores. Input files were .iDAT files from the 450K array and a .csv file containing Sample_ID, Sentrix_ID, Sentix_Position, Sample_Pool, Sample_Well, Sample_Group, Gender, Age, Smoking history, MRC Scale Score and Codon129 genotype. Analysis took 1 d 9 h 23 min 35 s and generated 9.00 Gb data. Due to the inflexibility and long runtime of the pipeline, use of RnBeads was eventually discontinued and replaced with ChAMP, a more modular analytical pipeline described in Section 3.4 Analysing 450K data using ChAMP

RnBeads initially excluded 29928 array probes which are cross-reactive with multiple genomic loci (Y. A. Chen *et al.*, 2013) and 10119 probes which overlap commonly occurring SNPs (Zhou, Laird and Shen, 2016). The inbuilt Greedycut algorithm iteratively removed unreliable probes and samples based on a probe detection *p* value threshold of 0.05: in other words, probe-wide and sample-wide detection significances were optimised. 4954 probes were discarded during this process over the course of 5000 iterations. Rnbeads automatically maps the array-wide density of beta values (a metric between 0 (unmethylated) and 1 (methylated) per probe) after this process, as displayed in Figure 13i.

Beta values were normalised using the SWAN method (Maksimovic, Gordon and Oshlack, 2012), the results of which are visualised in Figure 14. 65 technical probes and 1270 non-CpG probes were removed, as were 10160 probes on sex chromosomes. Array-wide beta density is once again automatically mapped after this process and is displayed in Figure 13ii.

Principle Component Analysis identified 217 components, the first two of which explain 13.79% and 11.70% of variance respectively: these are compared in Figure 15. Association of principle components with metadata was then performed using parametric (Pearson) tests where metadata is numeric and non-parametric (Wilcoxon or Kruskal-Wallis) tests where data is categorical and has two values, or more, respectively. This analysis showed significant associations of Sentrix ID

(beadchip number), Sample Group, Codon 129 Genotype and MRC Scale Score with Principle components 1 and 2, as shown in Figure 16.

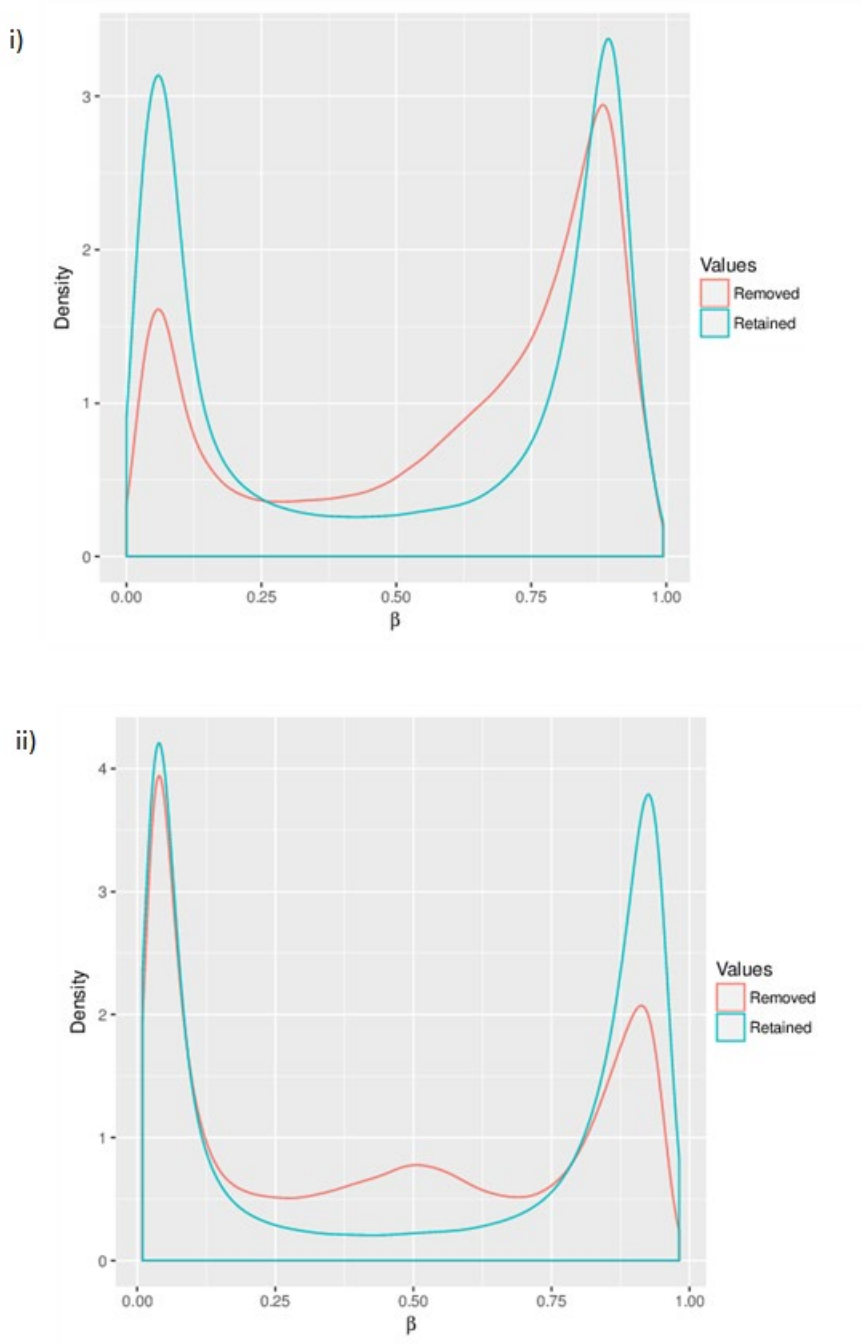


Figure 13: Density plots of beta values after RnBeads pre-processing. i) Density of beta values at probes removed and retained after excluding cross-reactive probes, probes that overlie commonly occurring SNPs, and probes identified as unreliable by Greedycut. ii) Density of beta values at probes removed and retained after excluding technical, non-CpG and sex chromosome probes.

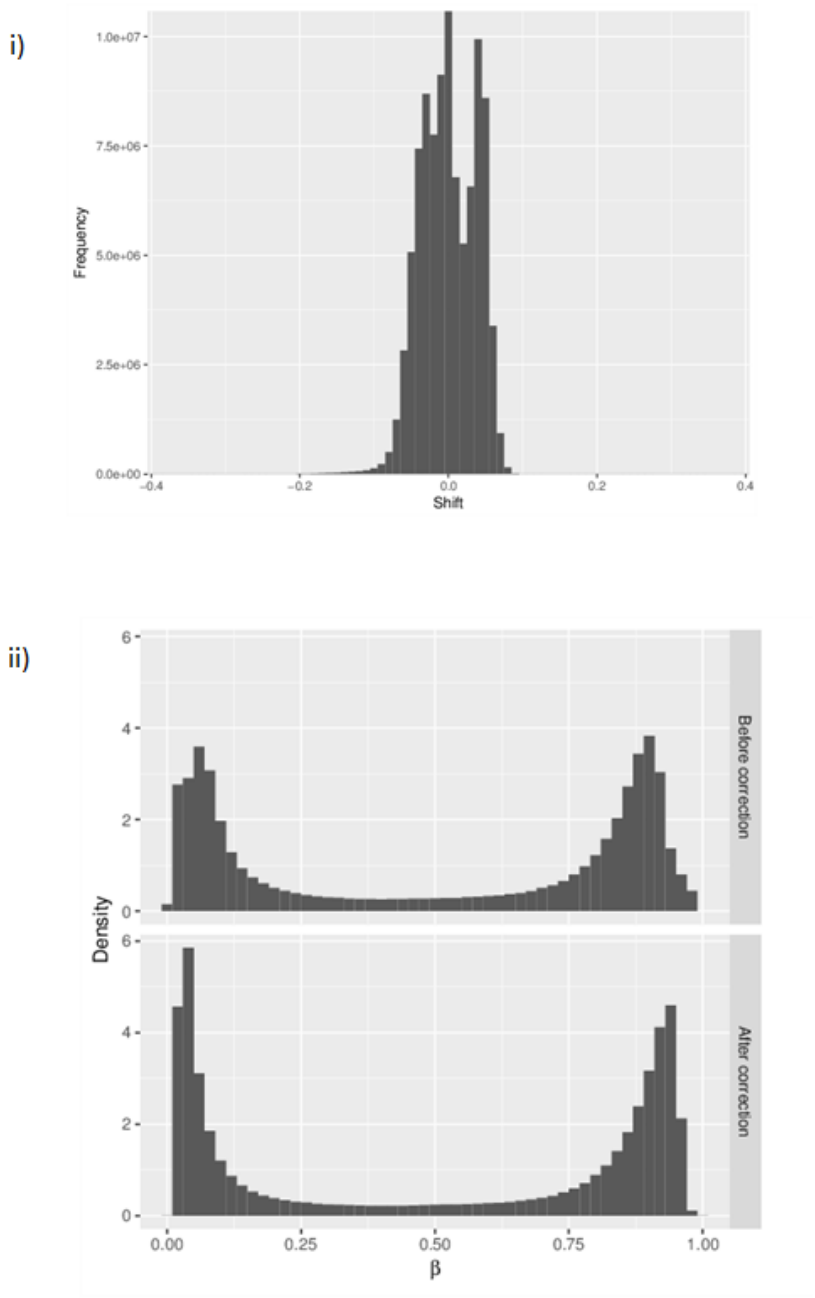


Figure 14: Histograms of beta shifts as a result of SWAN normalisation. i) Array-wide shifts in beta, and ii) beta distribution before and after normalisation.

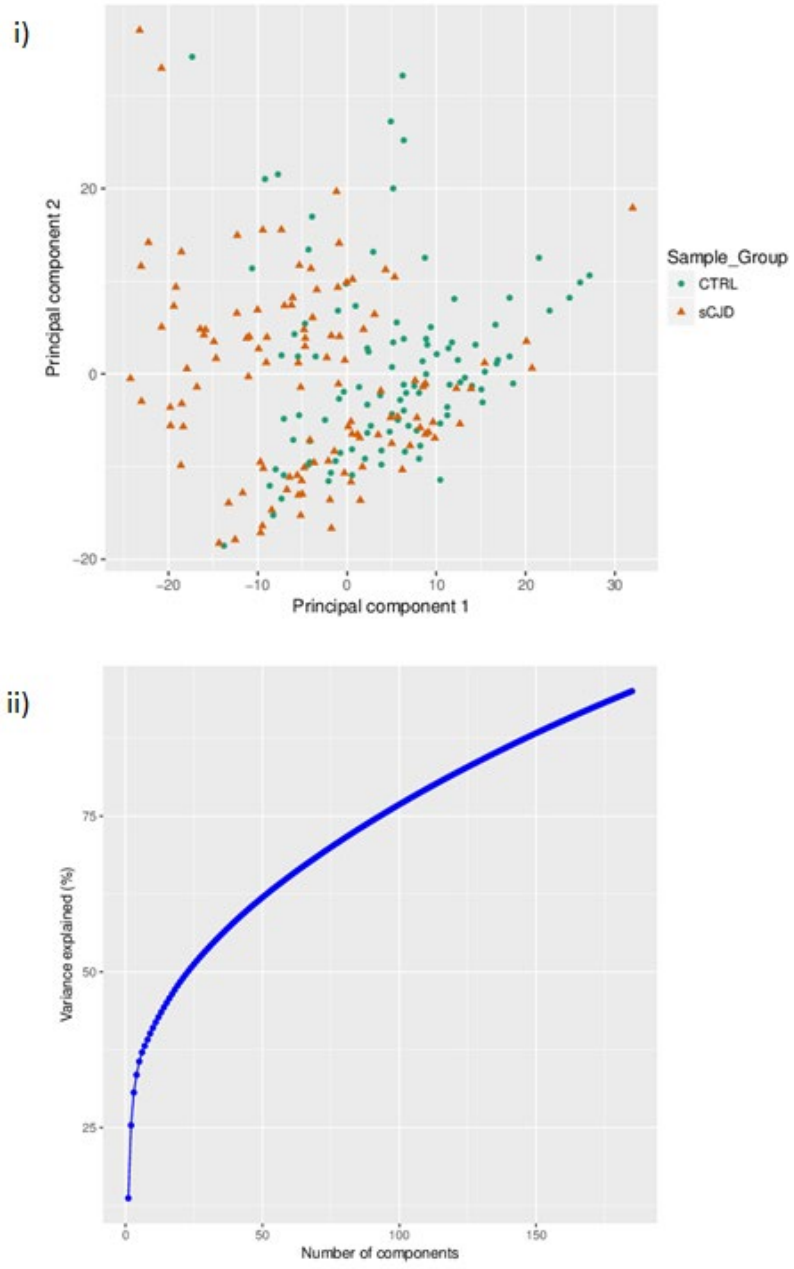


Figure 15: Principle component analysis of normalised RnBeads data. i) 2-dimensional scatterplot of sample distribution across principle components 1 & 2, and ii) cumulative explanatory variance of principle components 1 to 217.

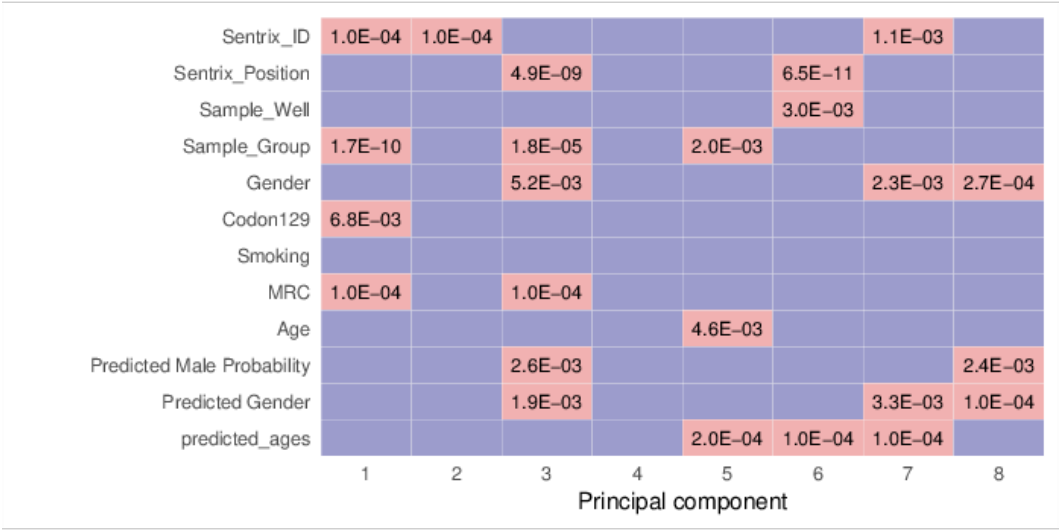


Figure 16: Chessboard plot of associations between the top 8 principle components and technical, biological and predicted metadata. Purple cells represent statistically insignificant associations, while the number in each magenta cell is the p value of a significant association.

Finally, normalised data was corrected for differences in cellular heterogeneity between samples. It is important to note that blood is a heterogeneous tissue and imbalances in cell proportions between individuals and sample groups are thought to be major sources of bias in studies of methylation data, for which correction is "critical" (Jaffe and Irizarry, 2014). While reference-based methods for correcting for differences in cell populations exist, RnBeads only permitted reference-free methods to be used alongside analyses featuring covariates such as age and sex. Therefore RefFreeEwas was used.

3.3.2 Analysis results

After processing and DMP calling with RefFreeEwas correction for cellular heterogeneity, RnBeads computed 3,842 differentially methylated positions which pass a genome-wide significance of $p < 0.05$ after Bonferroni correction for multiple testing. This is more than might be expected given the relatively modest sample size, particularly when compared to similar published studies. The genome-wide inflation factor value (λ) was calculated and found to be 3.37, indicative of appreciable *p* value inflation which can often be attributed to bias or the influence of unaccounted factors in a comparison. This inflation is also visually evident in the Manhattan plot of genome-wide significance shown in Figure 17. Without RefFreeEwas correction λ was almost doubled (5.97), as displayed in Figure 18.

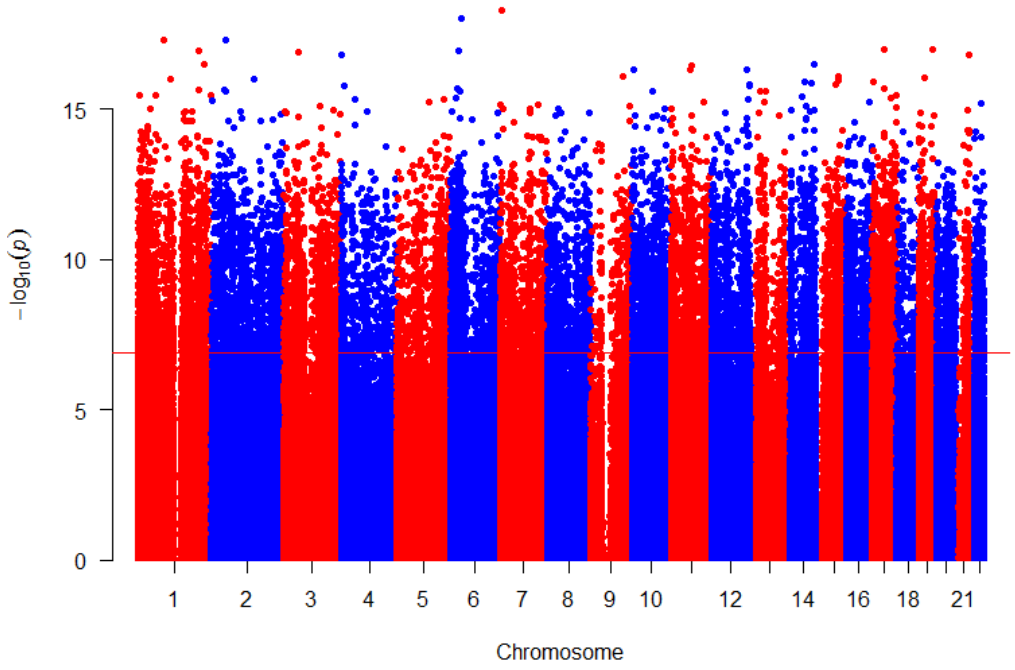


Figure 17: Manhattan plot of DMPs as computed by RnBeads.

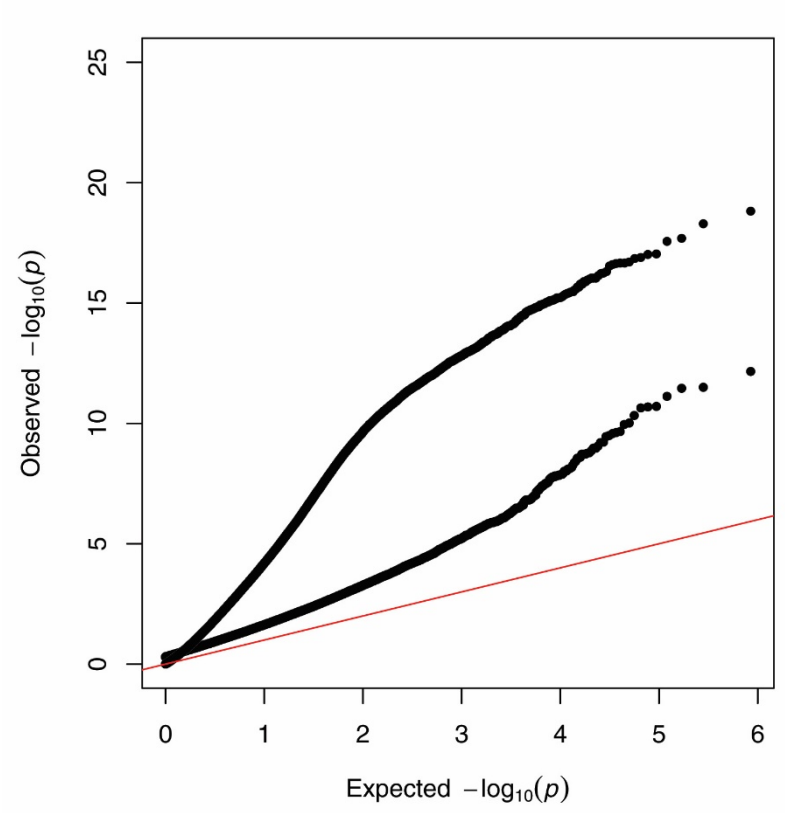


Figure 18: Quantile-quantile plot of DMP significance before (upper trace) and after (lower trace) correction for cellular heterogeneity using RefFreeEWAS.

To better interpret the biological relevance of these results, I used a package called Goldmine which allows association of genomic location with genetic and epigenetic features in user-supplied ChIP-Seq data. I integrated publicly available Roadmap Consortium data (Roadmap Epigenomics Consortium *et al.*, 2015) from Monocytes, CD4+ T Cells, CD8+ T Cells, B Cells, Natural Killer Cells and Granulocytes into Goldmine and found that 47.6% of DMPs were associated with CpG features (Island, Shore, Shelf) while 52.6% were situated in identified Transcription Factor Binding Sites and 61.2% at DNaseI hypersensitive sites. 27.6% of DMPs were positioned at enhancers. The majority of DMPs overlapped genomic features such as the promoter (34.1%), exons (3.5%), introns (33.7%) or the 3' region (12.2%), while the remaining 16.6% of sites were intergenic. Feature and Genomic distributions are displayed in Figure 19.

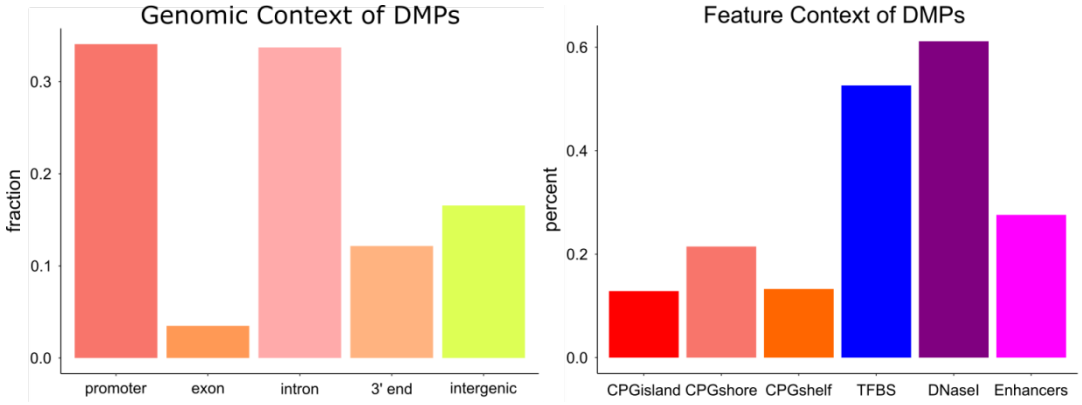


Figure 19: Feature and Genomic contexts of DMPs, computed using RnBeads results and ChIP-Seq data from the Roadmap Epigenomics Consortium. CpG Islands are regions of >200bp containing >50% GC composition and an observed to expected CpG ratio of >60%. Shores are regions 2kb from a CpG Island and Shelves are regions 2kb from a CpG Shore.

Ultimately the inflexibility of RnBeads and the long run times proved an obstacle to performing subanalyses, as if the pipeline encountered an error the entire process had to be reinitiated. Due to its one-step design, data could not be removed at different steps of analysis and compared or modified using other packages. In addition, the high level of genome-wide inflation ($\lambda = 3.37$) was a concern. As a result an alternative, lighter and more modular pipeline (ChAMP) was chosen for subsequent analyses.

3.4 Analysing 450K data using ChAMP

3.4.1 Pre-processing and Quality Control

As an alternative to RnBeads, I also analysed array data using ChAMP, a highly modular and comparatively lightweight analytical pipeline. Like RnBeads, ChAMP's champ.load() function imports .iDAT files and a .csv file containing sample metadata as described earlier, and excludes unreliable probes and samples before analysis. First probes with a detection $p < 0.01$ (based on iDAT file data) are discarded. Similarly samples where 10% or more probes have a detection $p < 0.01$ are discarded. Then poorly hybridising probes are removed (i.e. probes with less than three bead hybridisations in at least 5% of samples). Probes overlapping SNPs with a mean allele frequency over 5% (as described by Zhou et al. 2016) are also removed from the dataset (Zhou, Laird and Shen, 2016). Next, probes which are crossreactive with multiple genomic loci as reported by Nordlund *et al*. are discarded (Nordlund *et al.*, 2013). Finally probes on the X and Y chromosomes are discarded. The nature of these steps and numbers of probes discarded during loading sCJD and control data are shown in Figure 20.

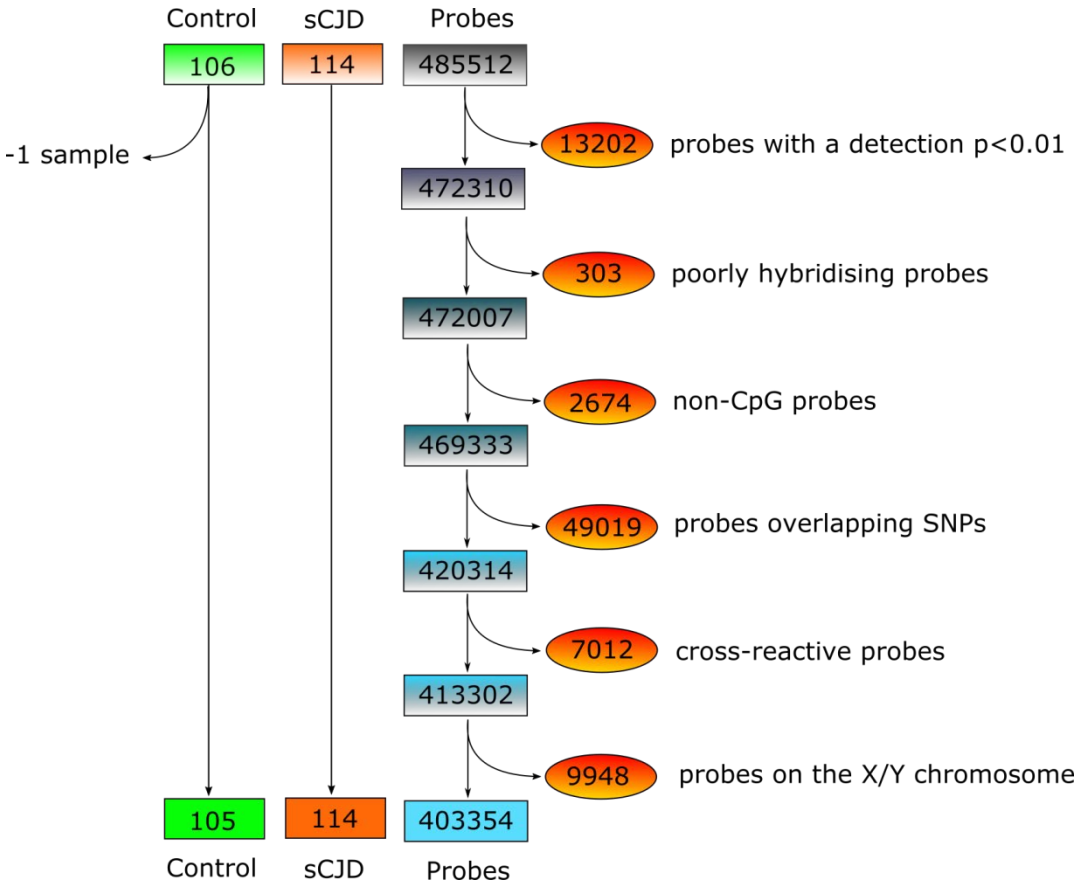


Figure 20: Exclusion of 82158 probes and one control sample as a result of the champ.load() function.

Once problematic probes and poorly-hybridised samples have been excluded red and green channel intensity are used to compute beta values per probe per sample. The champ.QC() function then generates several images for data visualisation including a display of array-wide beta density (Figure 21).

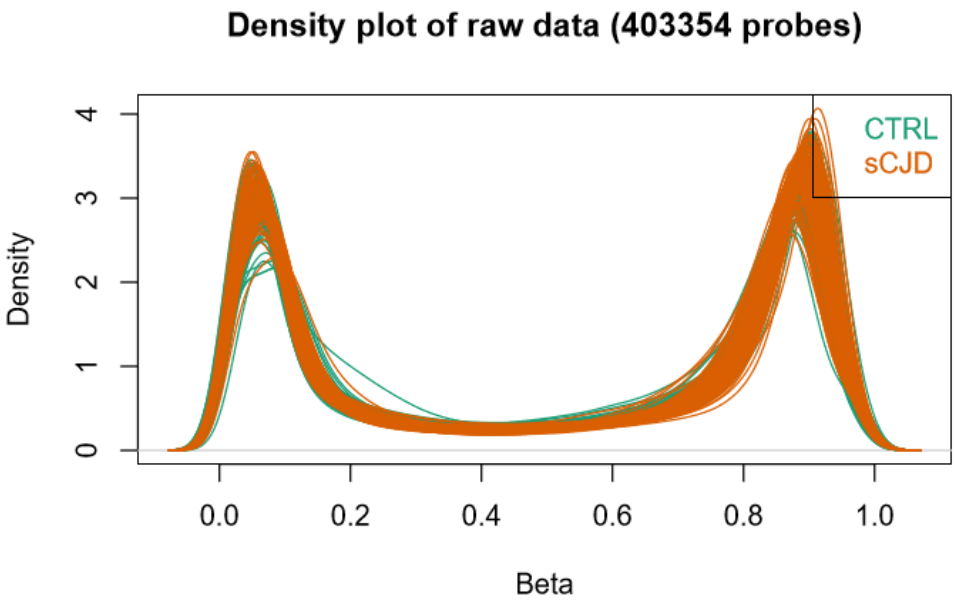


Figure 21: champ.QC()-generated plot of beta density across the loaded dataset. The majority of data trend towards a β of 0 (unmethylated) or 1 (methylated).

Once data were loaded type I and II probe imbalance was corrected through Beta-MIx Quantile dilation (BMIQ) normalisation (Teschendorff *et al.*, 2013). Compared to RnBeads' SWAN BMIQ performs similarly and is ChAMP's default normalisation method (Liu and Siegmund, 2016). Once normalised, I used Singular Value Decomposition (SVD) to associate biological and technical metadata with principle components. This serves two purposes: to identify which biological data are associated with variability in the dataset, and to identify any confounding biological or technical factors which should be corrected.

Despite the randomisation of samples within a 96 well plate as described in Section 2.1.5, the 12-slide format of the 450K beadchip array resulted in sCJD and control samples being inequally distributed between different slides on the array, and across different arrays. In addition, recently published work by Jiao *et al*. identifies strong positional effects on 450K array data, which further explain positional batch effects in our data as revealed by SVD using ChAMP and PCA in the RnBeads analysis (Jiao *et al.*, 2018). Principle components 1 and 2 were found to associate significantly with both slide coordinate and array barcode. I used an established statistical tool, ComBAT, to regress for positional effects. Reassuringly, remaining metadata associations with principle component 1 were (in decreasing order of significance) phenotype (Sample_Group), Codon129 Genotype, MRC Scale Score

and Gender, as shown in Figure 22. This shows that disease status and severity do indeed affect the methylome and suggests that sex as a covariate should be included in a final regression model.

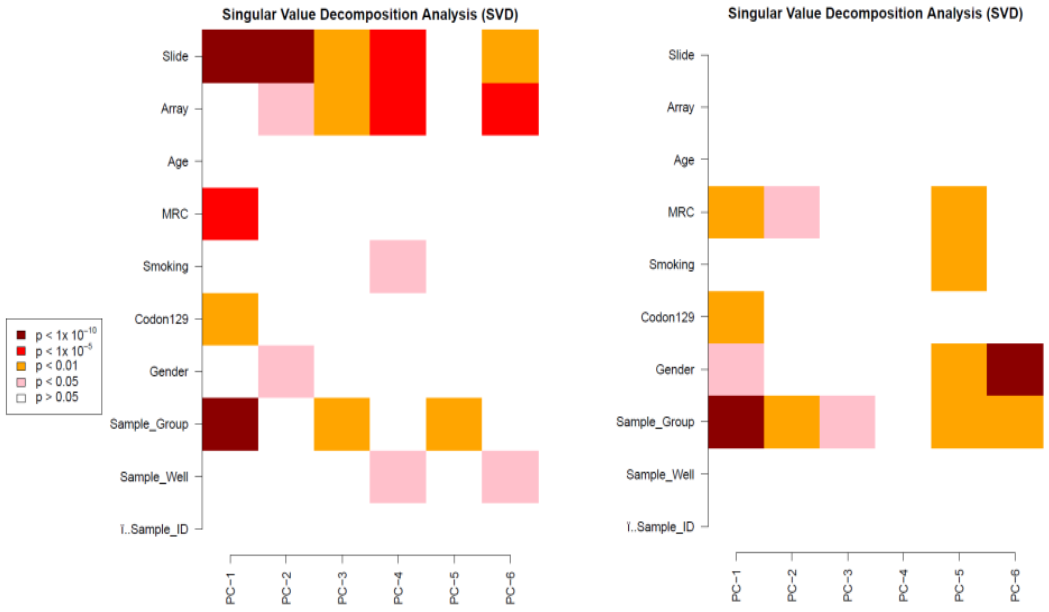


Figure 22: Chessboard plots of Singular Value Decomposition Analysis. These show association of principle components with sample metadata before and after ComBAT correction for positional batch effects at "Slide" and "Array" variables

As mentioned earlier, blood is a heterogeneous tissue and imbalances in cell proportions between individuals and sample groups are thought to be major sources of bias in studies of methylation data (Jaffe and Irizarry, 2014). I used two algorithms, the established Houseman algorithm and Teschendorff's more recent EpiDISH algorithm, to estimate cell proportions in batch-corrected data, after which differences in cell proportions were normalised through regression (Teschendorff *et al.*, 2017). These algorithms make use of 450K data acquired from DNA derived from blood cell populations as separated using FACS. This reference-based approach differs from RnBeads' RefFreeEWAS, which was also developed by Houseman who suggests that reference-based correction is more robust where a reference dataset exists, while RefFreeEWAS is suitable for DNA derived from tissue without a reference (Houseman, Molitor and Marsit, 2014; Houseman *et al.*, 2016). Both Houseman's RefBase and Teschendorff's EpiDISH algorithms estimated differences in cell proportions between case and control as well as

producing different estimations of cell fractions between each method when applied to case and control groups, as shown in Figure 23.

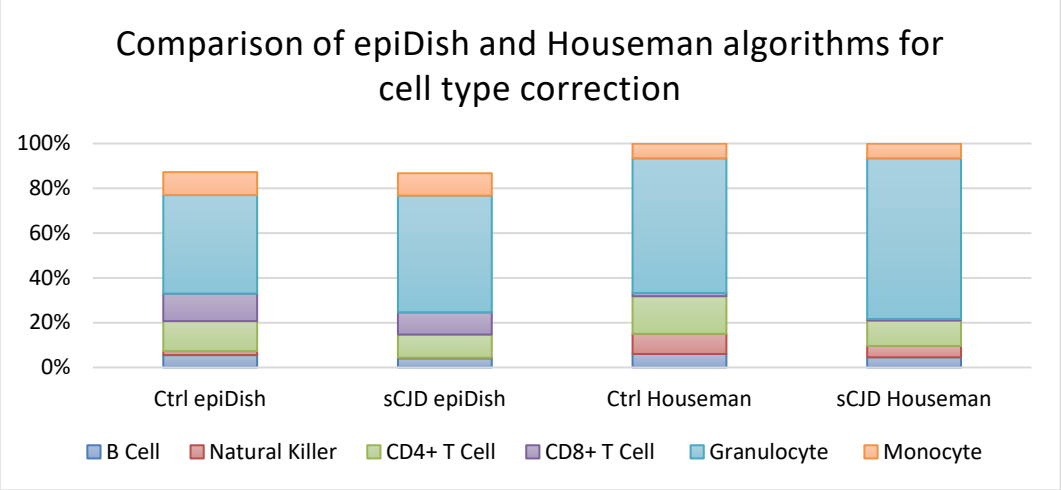


Figure 23: Average estimated cell proportions in the sCJD and Control groups based on epiDISH and Houseman algorithms.

Once data had been preprocessed, normalised and corrected for batch effects and cellular heterogeneity, I used limma to assess the effects of the pipeline by constructing linear regression models of case-control differences in beta after each processing step. Quantile-quantile analysis of *p* value inflation showed high levels of inflation after BMIQ normalisation (λ=4.07), which decreased after batch correction (λ=2.92) and reached minimum after "Refbase": reference-based correction for cell heterogeneity using the Houseman algorithm (λ=1.40). These plots are compared in Figure 24. EpiDish correction resulted in a comparatively inflated λ of 1.75, and so neither it nor RnBeads data (λ=3.37) were taken forward for use in downstream analysis. In making this decision, I made an assumption that the dataset exhibiting the lowest genome-wide inflation was more likely to reveal true biological effects.

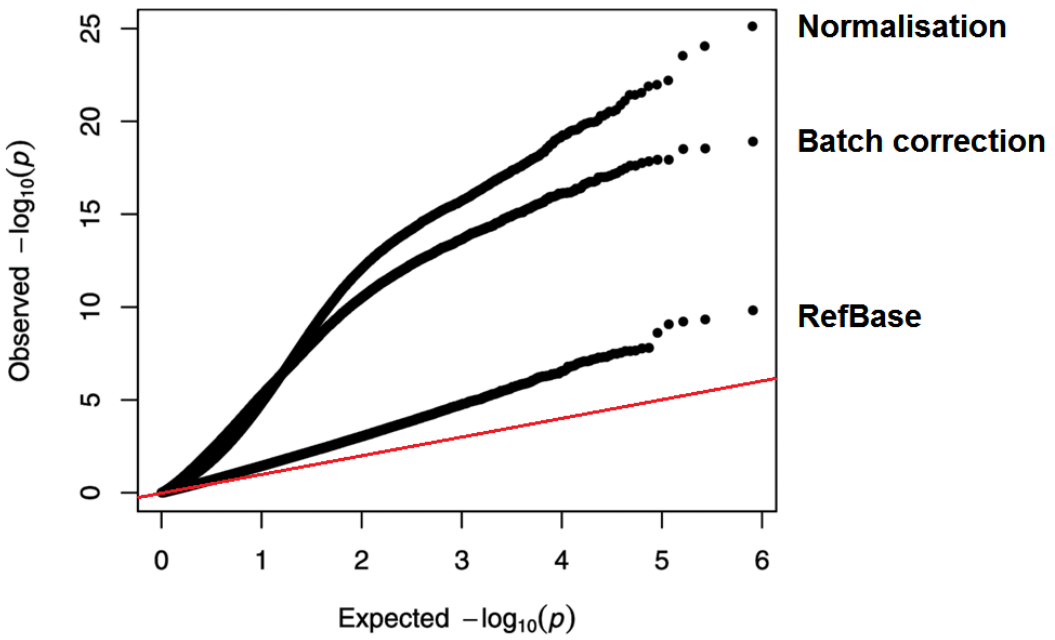


Figure 24: Quantile-quantile plots of limma linear model results. Traces correspond to results using data after BMIQ normalisation, then ComBAT batch correction and finally Houseman reference-based (RefBase) correction for cell heterogeneity.

3.4.2 Results: Differentially Methylated Positions in sporadic CJD

I investigated specific sites which exhibited differential methylation by constructing a linear regression model of β ~ Sample_Group + Age + Sex using limma, an established software package for analysis of microarray data (Ritchie *et al.*, 2015). This identified 38 DMPs (Table 7, Figure 25) which passed Bonferroni corrected significance threshold. Goldmine analysis using Roadmap Consortium data revealed that the majority of DMPs are within the promoter region of a gene and also overlap transcription factor binding sites or enhancers, as well as sites hypersensitive to DNAse1 (Figure 26). Two DMPs were situated upstream of the TSS of the same gene (*AIM2*), and sites in *AIM2*, *UHRF1* and *CANT1* were also found to be overlapped by binding sites for multiple transcription factors. The distributions of these features are displayed in Figure 27.

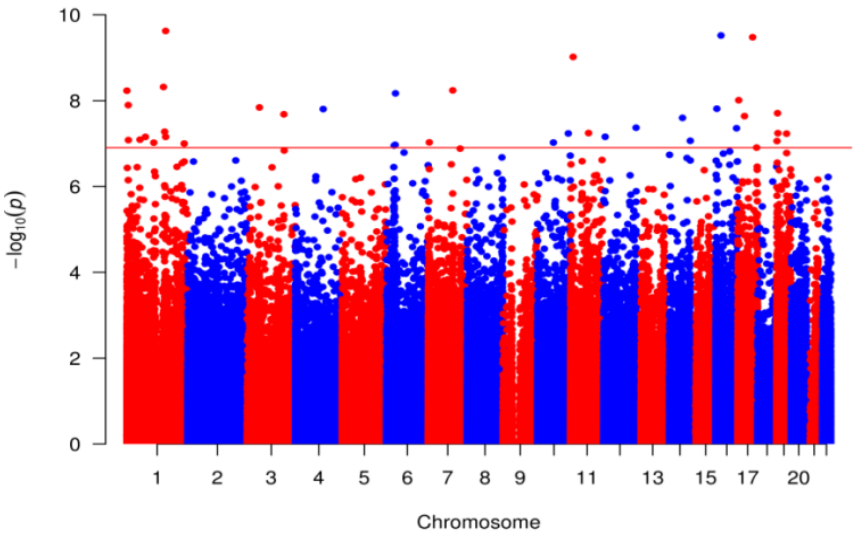


Figure 25: Manhattan plot of differential methylation between sCJD and control calculated using ChAMP. Genome-wide statistically significant DMPs lie over the red line (Bonferroni significance threshold of $p < 1.24x10^{-7}$).

Site	Δβ (%)	*p*	Gene	Region	Position	Feature
cg10636246	-4.05	2.38E-10	*AIM2*	TSS1500	1:159046973	Open Sea
cg02481950	1.98	3.03E-10	*METTL9; IGSF6*	Body; TSS1500	16:21665002	Open Sea
cg14427590	2.27	3.33E-10	Intergenic		17:60695089	Open Sea
cg05740793	4.39	9.57E-10	*MTRNR2L8*	TSS1500	11:10531091	Open Sea
cg13965201	2.91	4.80E-09	Intergenic		1:150210352	S.Shelf
cg21540367	0.99	5.76E-09	*LRCH4*	Body	7:100175462	S.Shelf
cg05001044	5.33	5.86E-09	*MIR1977*	TSS1500	1:567312	Open Sea
cg09048334	3.48	6.76E-09	Intergenic		6:37012640	Island
cg22519265	1.08	9.77E-09	*ATP2A3*	3'UTR	17:3828052	Island
cg02448796	3.28	1.27E-08	*KCNAB2*	Body	1:6101339	Open Sea
cg17641710	2.39	1.44E-08	*GNAI2*	Body	3:50279038	S.Shelf
cg03819286	2.68	1.54E-08	*MGRN1*	TSS1500	16:4673974	N.Shore
cg10855342	0.64	1.57E-08	*ALPK1*	5'UTR	4:113276902	Open Sea
cg15197458	1.4	1.96E-08	Intergenic		19:3464765	Island
cg00832928	2.88	2.08E-08	*SELT*	Body	3:150329458	Island
cg22688566	2.83	2.29E-08	*MYO18A*	Body	17:27459835	Open Sea
cg25966751	2.25	2.52E-08	Intergenic		14:74098320	N.Shelf
cg20056593	1.65	4.27E-08	Intergenic		12:132993150	S.Shelf
cg27229664	2.21	4.39E-08	*KIAA0513*	5'UTR	16:85096666	Open Sea
cg22505006	2.9	5.29E-08	*ZBTB7B*	5'UTR	1:154981829	Open Sea
cg05343106	2.03	5.69E-08	*DNAJB13*	TSS200	11:73661229	Open Sea
cg17714703	3.31	5.73E-08	*UHRF1*	Body	19:4912221	S.Shore
cg07081759	2.5	5.81E-08	*FAM53B*	Body	10:126330905	Open Sea

cg13444131	0.92	5.88E-08	*DYRK1B*	5'UTR	19:40322546	N.Shore
cg22434506	0.96	6.94E-08	*IFFO1*	Body	12:6657818	Island
cg17515347	-4.74	6.97E-08	*AIM2*	TSS1500	1:159047163	Open Sea
cg20003976	1.44	6.99E-08	*ACADM*	TSS1500	1:76188832	N.Shore
cg09007354	2.38	8.11E-08	*GLIS1*	5'UTR	1:54100163	Open Sea
cg20285559	1.01	8.30E-08	*THAP3*	Body	1:6688542	S.Shelf
cg19769147	1.95	8.61E-08	*PACS2*	Body	14:105860954	N.Shelf
cg24843003	2.94	8.72E-08	*DAZAP1*	Body	19:1409547	S.Shore
cg03393322	0.85	9.38E-08	*SDK1*	Body	7:4260883	Open Sea
cg04757081	1.4	9.47E-08	Intergenic		10:65424432	Open Sea
cg01084918	2.04	9.53E-08	*FAM40A*	TSS1500	1:110576366	N.Shore
cg01101459	2.89	1.00E-07	Intergenic		1:234871477	Open Sea
cg03546163	-5.35	1.07E-07	*FKBP5*	5'UTR	6:35654363	N.Shore
cg21393135	0.33	1.11E-07	*VARS*	Body	6:31747255	Open Sea
cg21155515	-0.75	1.24E-07	*CANT1; CANT1*	1stExon; 5'UTR	17:77005819	Island

Table 7: DMPs between sCJD and control. p values were adjusted using the Bonferroni method with a threshold of $p < 1.24x10^{-7}$. TSS200 and TSS1500 represent regions 200 nt and 1500 nt upstream of a gene's Transcription Start Site, respectively. Body and 1stExon relate to sites in the gene body or the first exon of the gene. 5'UTR and 3'UTR are the 5' and 3' untranslated regions of the mRNA transcript. Intergenic regions are regions that do not fall into the previous categories. CpG Islands are regions greater than 200 bp in width which contain greater than 50% GC composition and an observed:expected CpG ratio exceeding 60%. Shores are regions 2kb from a CpG Island, shelves are regions 2kb from a CpG Shore, and Open Sea regions are >2kb from a CpG Shore (Bird, 1986).

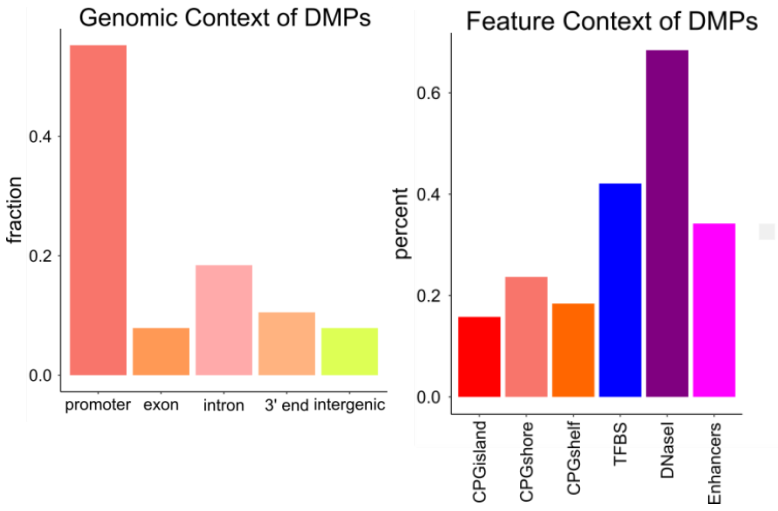


Figure 26: Feature and Genomic contexts of DMPs computed using ChAMP results and ChIP-Seq data from the Roadmap Epigenomics Consortium. CpG Islands are regions of >200bp containing >50% GC composition and an observed to expected CpG ratio of >60%. Shores are regions 2kb from a CpG Island and Shelves are regions 2kb from a CpG Shore.

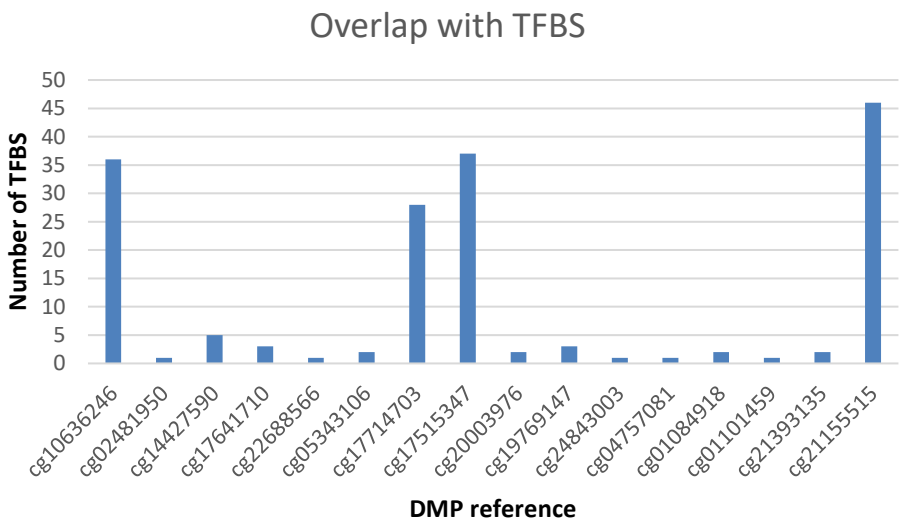


Figure 27: Number of overlapping Transcription Factor Binding Sites (TFBS) per DMP probe.

Having identified DMPs, I then associated beta values at these sites within the sCJD dataset with patient metadata from the Cohort Study, namely MRC Scale Score and MRC Scale Slope, and age of onset using bivariate Pearson's correlation. Codon 129 genotype was associated with beta values using a one-way ANOVA with a Dunnett's two-tailed post hoc test.

3.4.3 Association between MRC Scale Score and Slope with Methylation in sCJD patients

Association of DNA methylation with measures of disease severity and velocity could show promise as peripheral disease biomarkers and might potentially be useful in patient management and monitoring of any therapeutic interventions in clinical trials. Described briefly in Section 2.1.1, the MRC Scale was developed using a Rasch model to measure disease severity based on questionnaire data relating to cognitive, general, neurological and psychiatric function (Thompson *et al.*, 2013). By assessing patient status in relation to expected disease milestones (as displayed below in Figure 28), patients can be assigned a score between 20 (able to function independently) and 0 (moribund). The clinical milestones in a generalised order are displayed below in Figure 28.

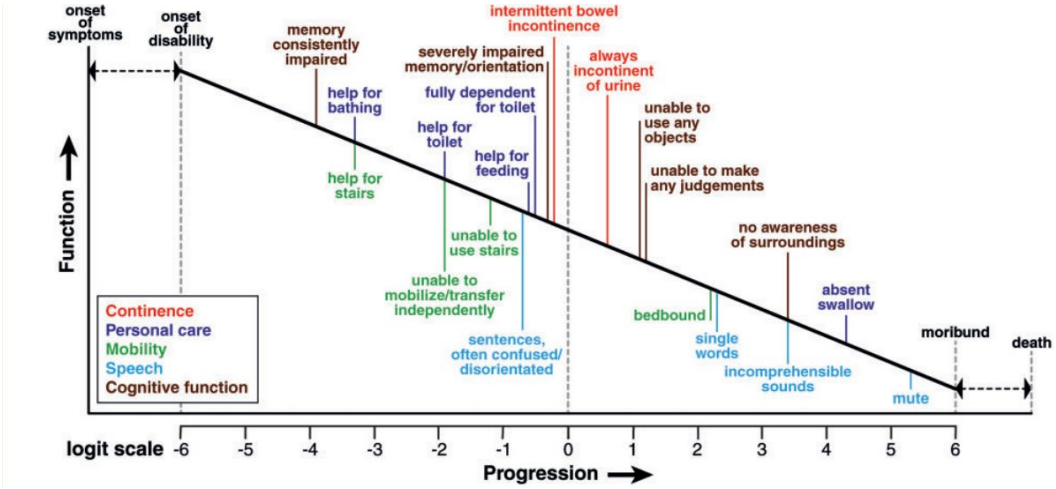


Figure 28: Disease milestones and functional criteria used to assign an MRC Scale score, from a maximum of 20 (highly functioning) to a minimum of 0 (zero independent function). Figure adapted from Thompson et al., 2013.

Moreover, by measuring the slope of decline in MRC Scale score, the rapidity of disease progression can be compared between patients. I therefore compared the MRC Scale scores and slopes of sCJD patients with methylation values at the 38 identified DMPs.

As shown in Figure 29 and Table 8, I found that demethylation at two probes within the *AIM2* promoter (cg01636246 & cg17515347) were found to correlate significantly with decreased MRC Scale Score, however no significant associations between DMPs and MRC Scale Slope were observed.

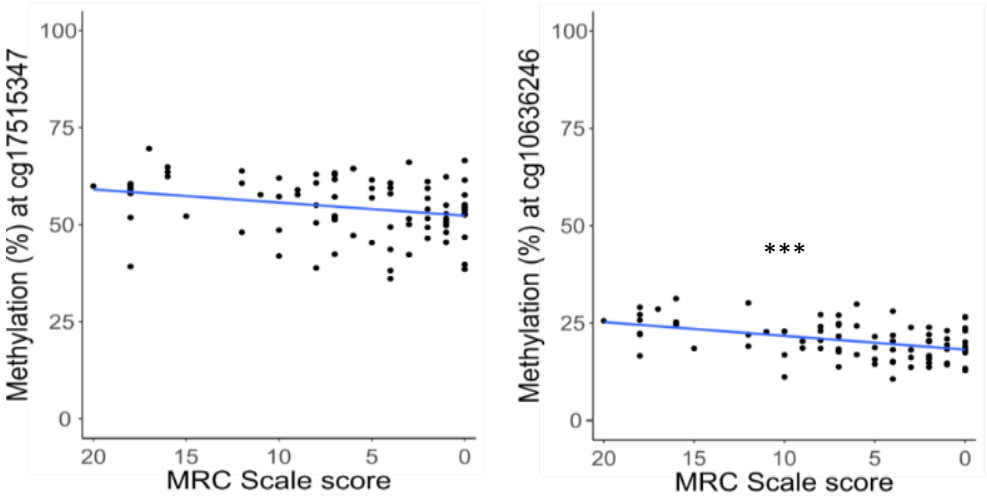


Figure 29: Coefficients of decreased MRC Scale Score with hypomethylation at AIM2. Shown are two probes (cg17515347 and cg10636246) with Pearson coefficients of 0.242 (p = 0.031) and 0.423 (p = 9.40x10-6) respectively. An MRC Scale score of 20 represents a patient who may have symptoms but is functionally independent for activities of daily life. An MRC Scale score below 3 indicates a comatose state near to death with a patient only able to take perhaps sips of fluid for nutrition and make incomprehensible sounds. Asterisks denote Bonferroni-adjusted significance, where an adjusted threshold of 0.05 has been exceeded (= p< 0.05, ** = p< 0.01, *** = p<0.001).*

Site	Gene	Scale Score coefficient	Scale *p*	Scale Slope coefficient	Slope *p*
cg10636246	*AIM2*	0.423	9.40E-06	-7.693	0.209
cg02481950	*METTL9;IGSF6*	0.009	0.939	3.209	0.318
cg14427590	Intergenic	-0.095	0.404	1.201	0.804
cg05740793	*MTRNR2L8*	0.142	0.21	0.524	0.927
cg13965201	Intergenic	0.029	0.798	11.382	0.481
cg21540367	*LRCH4*	0.4	0.725	4.521	0.308
cg05001044	*MIR1977*	0.179	0.112	4.999	0.570
cg09048334	Intergenic	0.067	0.555	18.431	0.285
cg22519265	*ATP2A3*	-0.023	0.843	2.794	0.547
cg02448796	*KCNAB2*	-0.2	0.075	-2.938	0.638
cg17641710	*GNAI2*	-0.184	0.102	-9.012	0.075
cg03819286	*MGRN1*	0.049	0.663	-3.230	0.576
cg10855342	*ALPK1*	-0.007	0.951	1.230	0.558
cg15197458	Intergenic	0.127	0.263	2.960	0.628
cg00832928	*SELT*	-0.084	0.456	16.096	0.403
cg22688566	*MYO18A*	0.071	0.532	-7.928	0.153
cg25966751	Intergenic	0.049	0.663	12.298	0.397
cg20056593	Intergenic	-0.007	0.951	1.959	0.680
cg27229664	*KIAA0513*	0.127	0.263	-10.778	0.008
cg22505006	*ZBTB7B*	-0.084	0.456	-8.057	0.195
cg05343106	*DNAJB13*	0.071	0.532	5.563	0.573
cg17714703	*UHRF1*	-0.046	0.688	-13.876	0.063
cg07081759	*FAM53B*	-0.107	0.343	0.652	0.905
cg13444131	*DYRK1B*	-0.003	0.979	2.402	0.532
cg22434506	*IFFO1*	-0.225	0.045	2.676	0.280
cg17515347	*AIM2*	0.242	0.031	-1.889	0.852
cg20003976	*ACADM*	-0.058	0.612	2.774	0.611

cg09007354	*GLIS1*	-0.014	0.899	-11.347	0.022
cg20285559	*THAP3*	0.067	0.556	1.434	0.820
cg19769147	*PACS2*	-0.124	0.273	3.273	0.547
cg24843003	*DAZAP1*	-0.161	0.154	-3.762	0.468
cg03393322	*SDK1*	0.218	0.052	0.577	0.727
cg04757081	Intergenic	0.055	0.627	6.759	0.303
cg01084918	*FAM40A*	0.041	0.72	14.902	0.246
cg01101459	Intergenic	-0.274	0.014	6.498	0.313
cg03546163	*FKBP5*	0.161	0.155	18.034	0.096
cg21393135	*VARS*	-0.054	0.633	1.220	0.277
cg21155515	*CANT1;CANT1*	-0.057	0.616	-3.158	0.460

Table 8: Coefficients of MRC Scale Score and MRC Scale with beta values at identified DMPs. p values are unadjusted, Bonferroni threshold is 0.0013.

3.4.4 Association between PRNP codon 129 homozygosity with Methylation in sCJD patients

Increased methylation at a probe within the promoter of *DNAJB13* (cg05343106) showed a trend towards association with homozygosity (MM, VV) at *PRNP* codon 129 via a one-way ANOVA with Dunnett's post hoc test, as displayed in Table 9 and Table 10. However, this trend was modest and the association would not pass correction for multiple testing. This might be due to low numbers of samples in each group as a result of splitting the data, but the very low levels of change between MM/VV and MV genotypes do not support differential methylation at *DNAJB13* between codon 129 alleles being biologically relevant.

Location		Sum of Squares	df	Mean Square	F	Sig.
cg10636246	*AIM2*	0.005	2	0.003	1.378	0.258
cg02481950	*METTL9*	0.000	2	0.000	0.043	0.958
cg14427590	Intergenic	0.002	2	0.001	1.626	0.203
cg05740793	*MTRNR2L8*	0.001	2	0.001	0.239	0.788
cg13965201	Intergenic	0.005	2	0.003	2.502	0.088
cg21540367	*LRCH4*	0.000	2	0.000	1.783	0.174
cg05001044	*MIR1977*	0.002	2	0.001	0.309	0.735
cg09048334	Intergenic	0.005	2	0.002	1.249	0.292
cg22519265	*ATP2A3*	0.001	2	0.000	2.633	0.078
cg02448796	*KCNAB2*	0.001	2	0.000	0.167	0.846
cg17641710	*GNAI2*	0.000	2	0.000	0.024	0.977
cg03819286	*MGRN1*	0.003	2	0.001	1.111	0.334
cg10855342	*ALPK1*	0.000	2	0.000	1.909	0.155
cg15197458	Intergenic	0.001	2	0.000	1.009	0.369
cg00832928	*SELT*	0.005	2	0.002	1.848	0.164
cg22688566	*MYO18A*	0.002	2	0.001	0.670	0.514
cg25966751	Intergenic	0.003	2	0.001	1.780	0.175
cg20056593	Intergenic	0.000	2	0.000	0.030	0.971
cg27229664	*KIAA0513*	0.002	2	0.001	1.635	0.201
cg22505006	*ZBTB7B*	0.000	2	0.000	0.036	0.965
cg05343106	*DNAJB13*	0.007	2	0.003	5.145	0.008
cg17714703	*UHRF1*	0.010	2	0.005	2.128	0.125
cg07081759	*FAM53B*	0.001	2	0.000	0.262	0.770
cg13444131	*DYRK1B*	0.000	2	0.000	1.598	0.208
cg22434506	*IFFO1*	0.000	2	0.000	0.840	0.435
cg17515347	*AIM2*	0.017	2	0.009	1.967	0.146
cg20003976	*ACADM*	0.002	2	0.001	2.498	0.088
cg09007354	*GLIS1*	0.000	2	0.000	0.019	0.981
cg20285559	*THAP3*	0.000	2	0.000	1.214	0.302
cg19769147	*PACS2*	0.001	2	0.000	0.621	0.540
cg24843003	*DAZAP1*	0.003	2	0.001	1.053	0.353
cg03393322	*SDK1*	0.000	2	0.000	0.601	0.550
cg04757081	Intergenic	0.001	2	0.000	1.531	0.222
cg01084918	*FAM40A*	0.002	2	0.001	1.727	0.184
cg01101459	Intergenic	0.006	2	0.003	1.800	0.172
cg03546163	*FKBP5*	0.009	2	0.004	0.793	0.456
cg21393135	*VARS*	0.000	2	0.000	1.845	0.164
cg21155515	*CANT1*	0.000	2	0.000	0.079	0.924

Table 9: ANOVA of beta values at identified DMPs between codon 129 genotypes. p values are unadjusted, Bonferroni threshold is 0.0013.

Multiple Comparisons

Dunnett t (2-sided)

Dependent Variable				Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
cg05343106	*DNAJB13*	MM	MV	-0.0228	0.0071	0.0036	-0.0387	-0.0069
		VV	MV	-0.0168	0.0076	0.0530	-0.0337	0.0002

Table 10: Dunnett's post-hoc test results for association between codon 129 genotype and increased methylation at DNAJB13 (cg05343106). In Dunnett's tests one group is used as a control and the other two genotypes compared against it.

3.4.5 Association between age of onset with Methylation in sCJD patients

Age of onset in the array dataset varies between 49 and 85, with most patients surviving between 6-12 months after clinical presentation. Age of onset was determined by the National Prion Clinic in the National Prion Monitoring Cohort study as the age at which the first symptom started that subsequently developed into the disease syndrome. A linear regression of Beta ~ Age + Sex was constructed in limma and one site achieved genome-wide statistical significance, as shown in Table 11.

Site	Coefficient (β)	*p* value	Adjusted *p* value	Gene
cg16867657	0.419699	7.13E-08	0.03	ELOVL2

Table 11: A significant association between DNA methylation at a CpG site overlapped by ELOVL2 and age of onset in sCJD patients. Bonferroni threshold was $p < 1.22x10^{-7}$.

ELOVL2 encodes Fatty Acid Elongase 2, and is well documented as becoming increasingly hypermethylated with age (Florath *et al.*, 2014; Bacalini, Boattini, *et al.*, 2015; Marttila *et al.*, 2015). Given the prior associations with hypermethylation at this gene with age, it is unlikely that methylation of this site predisposes to an earlier age of onset rather than simply reflects the age of the patients.

3.4.6 Differentially variable methylation in sCJD and Controls

iEVORA is an algorithm which detects significantly different variability at CpG sites, even where the median of the distribution at those sites may not differ between sCJD and control (Teschendorff *et al.*, 2016). Initially used to identify field defects – stochastic and heterogeneous alterations in DNA methylation amongst cells which are prone to become cancerous – Differentially Variably Methylated Positions (DVMCs) have also been observed in type 1 diabetes patients at genetic elements

which regulate cell metabolism (Teschendorff *et al.*, 2012; Paul *et al.*, 2016). I applied the iEVORA algorithm to the array dataset and identified 252 DVMPs. As a striking example, all probes overlapped by *MCCC1* showed increased variance in sCJD patients, as shown in Figure 30. DVMPs are displayed below in Table 12.

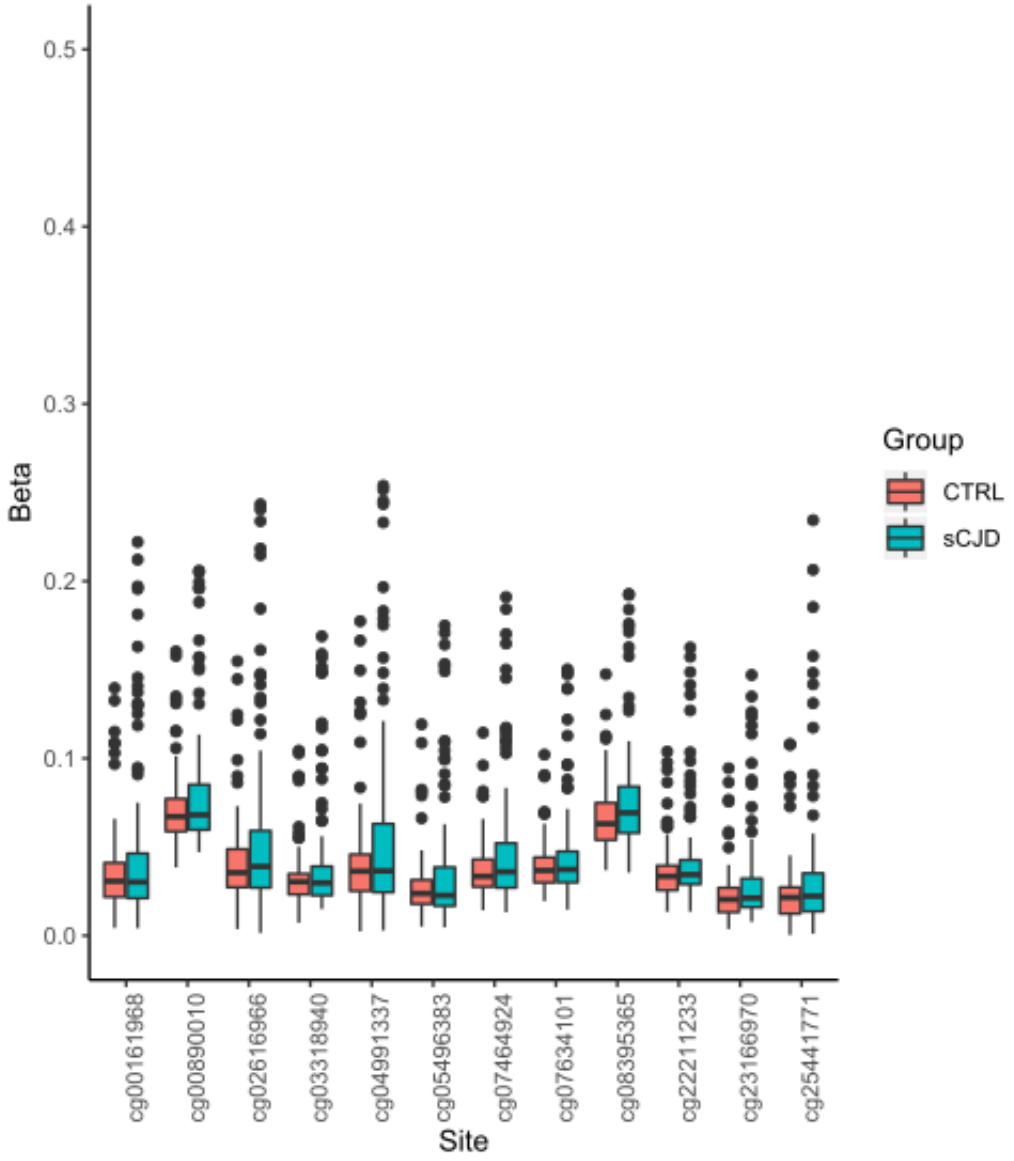


Figure 30: Increased variability in MCCC1 methylation in sCJD compared to control. Differences in mean methylation were not significant between groups.

	sCJD Variance (V0)	Control Variance (V1)	log[V1/V0]	*P* (Bartlett Test)	*q* (Bartlett Test)	Gene
cg19782061	0.120	0.095	1.781	8.80E-10	1.85E-06	Intergenic
cg07989851	0.314	0.297	1.350	2.45E-06	8.49E-04	Intergenic
cg16598420	0.888	0.872	-1.367	1.27E-06	5.18E-04	Intergenic
cg16740640	0.963	0.959	-1.316	3.03E-06	9.87E-04	Intergenic
cg25525687	0.015	0.017	-1.504	1.08E-07	7.74E-05	Intergenic
cg02294420	0.955	0.947	-1.528	6.82E-08	5.56E-05	Intergenic
cg06171329	0.942	0.935	-1.460	2.42E-07	1.52E-04	Intergenic
cg13447295	0.897	0.869	-1.749	8.07E-10	1.76E-06	Intergenic
cg21445918	0.973	0.968	-1.425	4.53E-07	2.40E-04	Intergenic
cg00961415	0.931	0.937	1.451	4.41E-07	2.37E-04	Intergenic
cg22681255	0.040	0.050	-1.764	5.90E-10	1.35E-06	Intergenic
cg15155360	0.933	0.923	-1.341	2.00E-06	7.24E-04	Intergenic
cg04825429	0.058	0.066	-1.719	1.52E-09	2.90E-06	Intergenic
cg11946719	0.174	0.144	1.486	2.37E-07	1.50E-04	Intergenic
cg06037631	0.033	0.029	1.433	5.97E-07	2.90E-04	Intergenic
cg03760138	0.932	0.922	-2.040	1.18E-12	7.10E-09	Intergenic
cg11467506	0.076	0.085	-1.580	2.48E-08	2.54E-05	Intergenic
cg26338473	0.841	0.816	-1.388	8.76E-07	3.91E-04	Intergenic
cg13910174	0.943	0.951	1.682	6.18E-09	8.98E-06	Intergenic
cg16651900	0.964	0.968	1.386	1.33E-06	5.39E-04	Intergenic
cg25448687	0.800	0.831	1.460	3.79E-07	2.15E-04	Intergenic
cg07136920	0.865	0.852	-1.503	1.09E-07	7.80E-05	Intergenic
cg05203206	0.964	0.958	-1.390	8.49E-07	3.83E-04	Intergenic
cg03632120	0.914	0.900	-1.432	4.05E-07	2.24E-04	Intergenic
cg19698348	0.025	0.028	-1.678	3.53E-09	5.69E-06	Intergenic
cg23439966	0.073	0.054	1.438	5.52E-07	2.76E-04	Intergenic
cg23721533	0.967	0.962	-1.338	2.08E-06	7.46E-04	Intergenic
cg02531516	0.078	0.086	-1.320	2.83E-06	9.36E-04	Intergenic
cg17117718	0.057	0.078	-1.727	1.29E-09	2.53E-06	Intergenic
cg17401808	0.924	0.931	1.435	5.83E-07	2.86E-04	Intergenic
cg24052284	0.972	0.969	-1.671	4.06E-09	6.38E-06	Intergenic
cg27273227	0.918	0.878	-2.043	1.10E-12	6.78E-09	Intergenic
cg03094193	0.020	0.025	-1.729	1.23E-09	2.44E-06	Intergenic
cg04115185	0.969	0.965	-1.762	6.16E-10	1.39E-06	Intergenic
cg12169233	0.051	0.056	-1.438	3.57E-07	2.05E-04	Intergenic
cg19632836	0.950	0.944	-2.603	5.66E-19	1.37E-14	Intergenic
cg23450377	0.918	0.892	-1.466	2.17E-07	1.40E-04	Intergenic
cg26879339	0.973	0.969	-2.159	6.61E-14	5.46E-10	Intergenic
cg24529615	0.912	0.905	-1.400	7.13E-07	3.36E-04	Intergenic
cg23402824	0.937	0.930	-1.798	2.81E-10	7.21E-07	Intergenic
cg26311454	0.936	0.940	1.661	9.31E-09	1.20E-05	Intergenic
cg26179679	0.025	0.029	-1.358	1.49E-06	5.88E-04	Intergenic

cg17738169	0.056	0.063	-1.747	8.49E-10	1.82E-06	Intergenic
cg09519326	0.957	0.942	-2.534	3.77E-18	7.08E-14	Intergenic
cg10130155	0.964	0.957	-1.541	5.28E-08	4.58E-05	Intergenic
cg27581660	0.907	0.927	1.375	1.61E-06	6.20E-04	Intergenic
cg24711482	0.940	0.925	-1.535	5.92E-08	5.03E-05	Intergenic
cg21173402	0.954	0.963	2.067	1.95E-12	1.10E-08	Intergenic
cg13390004	0.052	0.057	-1.647	6.57E-09	9.34E-06	Intergenic
cg08893692	0.054	0.057	-1.364	1.35E-06	5.43E-04	Intergenic
cg15593426	0.960	0.964	1.836	2.85E-10	7.26E-07	Intergenic
cg08063340	0.060	0.064	-1.321	2.77E-06	9.27E-04	Intergenic
cg08962087	0.043	0.046	-1.585	2.24E-08	2.36E-05	Intergenic
cg24684816	0.973	0.971	-1.441	3.42E-07	1.97E-04	Intergenic
cg21005683	0.902	0.862	-1.432	4.03E-07	2.23E-04	Intergenic
cg26576978	0.907	0.928	1.503	1.76E-07	1.18E-04	Intergenic
cg27422407	0.935	0.940	1.897	7.87E-11	2.46E-07	Intergenic
cg03268463	0.960	0.957	-1.467	2.14E-07	1.39E-04	Intergenic
cg06224751	0.043	0.046	-2.074	5.22E-13	3.46E-09	Intergenic
cg22230604	0.051	0.056	-1.321	2.80E-06	9.30E-04	Intergenic
cg10760240	0.821	0.875	1.604	2.75E-08	2.77E-05	AATF
cg06952291	0.025	0.028	-1.618	1.18E-08	1.44E-05	ABHD8
cg03053358	0.917	0.889	-1.493	1.30E-07	9.02E-05	ABR
cg04334422	0.963	0.960	-1.319	2.89E-06	9.50E-04	ACOX3
cg07595943	0.977	0.973	-1.625	1.03E-08	1.30E-05	ADAD2
cg02842496	0.133	0.143	-1.331	2.35E-06	8.22E-04	ADCY8
cg15691252	0.055	0.038	1.536	9.69E-08	7.23E-05	ADK;AP3M1
cg19245381	0.079	0.085	-1.648	6.53E-09	9.31E-06	ADORA2B
cg10791966	0.903	0.895	-1.645	6.94E-09	9.68E-06	ALDH3A1
cg07249488	0.944	0.938	-1.780	4.19E-10	1.00E-06	ANKLE2
cg11035303	0.016	0.041	-2.155	7.35E-14	5.92E-10	ANO10
cg19136183	0.022	0.024	-1.585	2.26E-08	2.37E-05	APAF1;IKBIP
cg13289321	0.950	0.943	-1.327	2.53E-06	8.66E-04	ASB10
cg00734483	0.020	0.023	-1.967	6.51E-12	2.97E-08	BACE2
cg15639842	0.925	0.917	-1.639	7.82E-09	1.06E-05	BAHCC1
cg26386740	0.016	0.017	-1.318	2.92E-06	9.59E-04	BCL11A
cg02781618	0.043	0.048	-1.692	2.63E-09	4.54E-06	BEND4
cg02699635	0.926	0.920	-1.524	7.34E-08	5.83E-05	C11orf94
cg02922913	0.951	0.942	-1.721	1.45E-09	2.81E-06	C19orf6
cg06651299	0.953	0.957	1.545	8.24E-08	6.39E-05	C19orf6
cg24304712	0.029	0.032	-1.547	4.69E-08	4.21E-05	C1orf101
cg19620452	0.027	0.029	-1.352	1.66E-06	6.32E-04	C1orf74
cg02192528	0.948	0.942	-1.398	7.35E-07	3.44E-04	C4orf23
cg12975295	0.035	0.040	-1.716	1.60E-09	3.02E-06	C9orf64
cg25160190	0.931	0.922	-1.554	4.11E-08	3.82E-05	CACNG4
cg07605269	0.896	0.877	-1.317	3.00E-06	9.77E-04	CAMTA1
cg17143900	0.931	0.921	-1.369	1.22E-06	5.04E-04	CAMTA1

cg17096979	0.917	0.894	-1.533	6.14E-08	5.11E-05	CAPN9
cg09444327	0.097	0.088	2.235	4.00E-14	3.47E-10	CCDC48
cg11616380	0.040	0.043	-1.614	1.28E-08	1.50E-05	CCDC48
cg17256192	0.952	0.947	-1.461	2.39E-07	1.51E-04	CDH6
cg09159068	0.037	0.031	1.374	1.64E-06	6.26E-04	CFD
cg27519392	0.806	0.795	-1.434	3.87E-07	2.18E-04	CHD7
cg04071104	0.957	0.963	2.246	3.08E-14	2.82E-10	CHDH
cg10418044	0.046	0.059	-1.521	7.78E-08	6.10E-05	CHRM2
cg14997413	0.169	0.183	-1.365	1.32E-06	5.38E-04	COBL
cg10364041	0.084	0.089	-1.487	1.47E-07	1.01E-04	COL12A1
cg09111258	0.040	0.044	-1.378	1.05E-06	4.51E-04	COPZ2;MIR152
cg26623575	0.853	0.831	-1.616	1.24E-08	1.48E-05	CST2
cg26923863	0.237	0.223	-2.261	5.18E-15	5.48E-11	CTBP1
cg23184518	0.934	0.924	-1.341	1.98E-06	7.19E-04	CTNNA1;LRRTM2
cg13132577	0.048	0.041	1.531	1.06E-07	7.65E-05	CYP2J2
cg07430760	0.077	0.063	1.591	3.54E-08	3.35E-05	CYP2J2
cg20690488	0.074	0.063	1.601	2.92E-08	2.90E-05	CYP2J2
cg18292664	0.066	0.073	-1.539	5.47E-08	4.72E-05	DBX1
cg05327967	0.938	0.933	-1.589	2.10E-08	2.25E-05	DLGAP4
cg17343385	0.927	0.894	-1.506	1.03E-07	7.50E-05	DNAJC5
cg19256292	0.055	0.060	-1.464	2.24E-07	1.44E-04	DNMT3A
cg19477942	0.967	0.962	-1.373	1.14E-06	4.82E-04	DSCAM
cg22355889	0.083	0.124	-2.164	5.90E-14	4.99E-10	ELMOD1;LOC643923
cg11058154	0.021	0.025	-1.585	2.25E-08	2.36E-05	FAM20A
cg05761971	0.947	0.940	-1.433	3.98E-07	2.22E-04	FAM82A1
cg19250315	0.954	0.958	1.395	1.15E-06	4.85E-04	FASN
cg22753962	0.939	0.924	-1.515	8.75E-08	6.68E-05	FOXK1
cg15277906	0.104	0.114	-1.454	2.71E-07	1.66E-04	GDF6
cg12949975	0.045	0.049	-1.614	1.28E-08	1.50E-05	GDF7
cg25308242	0.933	0.941	1.730	2.42E-09	4.24E-06	GLB1L
cg02612335	0.967	0.959	-1.683	3.17E-09	5.26E-06	GLT1D1
cg26234787	0.045	0.048	-1.448	3.02E-07	1.81E-04	GPR39
cg04657588	0.048	0.053	-1.457	2.53E-07	1.58E-04	GPRC5C
cg08230483	0.023	0.026	-1.361	1.41E-06	5.60E-04	GRID1
cg19640303	0.953	0.948	-1.510	9.50E-08	7.13E-05	GRIK3
cg19160878	0.979	0.976	-1.329	2.45E-06	8.49E-04	GTF3C3
cg20503652	0.941	0.914	-2.440	4.79E-17	7.72E-13	H1FNT
cg22101174	0.943	0.936	-1.451	2.83E-07	1.72E-04	HDAC4
cg17755964	0.025	0.028	-1.516	8.58E-08	6.58E-05	HES5
cg05428770	0.058	0.063	-1.370	1.20E-06	4.99E-04	HLA-L
cg17512353	0.035	0.039	-1.795	3.04E-10	7.50E-07	HLA-L
cg09463047	0.020	0.021	-1.475	1.83E-07	1.22E-04	HNF1B
cg21337717	0.896	0.884	-1.329	2.45E-06	8.49E-04	HRK
cg16595484	0.037	0.032	1.664	8.77E-09	1.16E-05	HSPBAP1
cg23448729	0.173	0.161	1.446	4.84E-07	2.52E-04	HTR1A

cg23639308	0.909	0.898	-1.520	7.89E-08	6.18E-05	IGSF9
cg06176987	0.937	0.951	1.781	8.66E-10	1.83E-06	INPP4B
cg04570208	0.961	0.955	-1.842	1.08E-10	3.20E-07	INPP5A
cg21528729	0.946	0.956	1.664	8.86E-09	1.17E-05	INSIG1
cg14908255	0.025	0.029	-1.620	1.15E-08	1.41E-05	IQSEC3
cg12984729	0.061	0.069	-1.418	5.21E-07	2.65E-04	ISL2
cg25395559	0.044	0.046	-1.414	5.59E-07	2.77E-04	ITGB8
cg23984080	0.921	0.915	-1.321	2.79E-06	9.30E-04	KIAA0495
cg14922732	0.890	0.879	-1.607	1.47E-08	1.65E-05	KIAA0562
cg13239126	0.810	0.864	1.434	5.90E-07	2.88E-04	KIAA1026
cg17001430	0.898	0.891	-1.462	2.33E-07	1.48E-04	KIF25
cg08287334	0.962	0.957	-1.354	1.59E-06	6.14E-04	KLF16
cg26832639	0.930	0.942	1.931	3.86E-11	1.36E-07	KLK5
cg25929399	0.146	0.079	1.608	2.54E-08	2.60E-05	KRT38
cg24827600	0.370	0.357	-1.846	9.99E-11	3.04E-07	LASS6
cg14492293	0.034	0.037	-1.663	4.81E-09	7.31E-06	LHX3
cg00894289	0.033	0.037	-1.545	4.94E-08	4.35E-05	LMO2
cg23497383	0.070	0.097	-1.801	2.63E-10	6.85E-07	LOC151174;LOC643387
cg13530938	0.046	0.066	-1.935	1.35E-11	5.42E-08	LOC151174;LOC643387
cg07323141	0.046	0.058	-1.426	4.47E-07	2.39E-04	LOC151174;LOC643387
cg11904906	0.058	0.079	-1.669	4.24E-09	6.60E-06	LOC151174;LOC643387
cg02249713	0.053	0.057	-1.705	2.01E-09	3.64E-06	LOC285780
cg15911153	0.982	0.980	-1.343	1.92E-06	6.99E-04	LOC389333
cg14379854	0.103	0.114	-1.419	5.07E-07	2.61E-04	LOC91149;RAPGEF4
cg27155168	0.975	0.972	-1.609	1.41E-08	1.61E-05	LRP5
cg15079934	0.853	0.840	-1.527	6.95E-08	5.63E-05	LSP1
cg26011692	0.937	0.932	-1.477	1.78E-07	1.19E-04	LTBP2
cg21944491	0.033	0.036	-1.392	8.24E-07	3.76E-04	LTBP4
cg11098525	0.958	0.955	-1.596	1.82E-08	2.00E-05	MAEA
cg02228913	0.842	0.781	-1.646	6.80E-09	9.58E-06	MAPT
cg08395365	0.078	0.066	1.645	1.28E-08	1.50E-05	MCCC1
cg23166970	0.031	0.023	1.639	1.42E-08	1.61E-05	MCCC1
cg02616966	0.056	0.042	1.926	4.27E-11	1.49E-07	MCCC1
cg25441771	0.034	0.023	1.915	5.44E-11	1.79E-07	MCCC1
cg22211233	0.044	0.036	1.639	1.44E-08	1.63E-05	MCCC1
cg04991337	0.056	0.042	1.714	3.31E-09	5.43E-06	MCCC1
cg05496383	0.037	0.028	2.006	7.45E-12	3.27E-08	MCCC1
cg03318940	0.041	0.033	1.974	1.50E-11	5.97E-08	MCCC1
cg00890010	0.081	0.072	1.456	4.03E-07	2.23E-04	MCCC1
cg07634101	0.046	0.039	2.004	7.80E-12	3.34E-08	MCCC1
cg07464924	0.048	0.037	2.241	3.46E-14	3.08E-10	MCCC1
cg00161968	0.047	0.035	1.773	1.02E-09	2.09E-06	MCCC1
cg00693240	0.963	0.959	-1.631	9.15E-09	1.19E-05	MGRN1
cg16208863	0.910	0.902	-1.571	2.98E-08	2.94E-05	MLLT1
cg03146452	0.034	0.040	-1.369	1.22E-06	5.03E-04	MSL3L2

cg21013756	0.097	0.112	-1.426	4.48E-07	2.39E-04	MSL3L2
cg01565659	0.948	0.955	1.403	1.01E-06	4.38E-04	MSRA
cg15020726	0.038	0.045	-1.700	2.25E-09	4.01E-06	MTMR15
cg17035899	0.074	0.093	-1.463	2.29E-07	1.45E-04	MYL5
cg09010067	0.966	0.963	-1.534	6.08E-08	5.10E-05	MYO1D
cg16904092	0.809	0.757	-1.717	1.58E-09	3.01E-06	MYPOP
cg06804705	0.749	0.772	1.404	9.85E-07	4.29E-04	NCRNA00114
cg16262215	0.934	0.941	2.156	2.53E-13	1.90E-09	NDUFA7;KANK3
cg20153196	0.069	0.073	-1.461	2.39E-07	1.51E-04	NEUROD2
cg04239375	0.068	0.074	-1.620	1.13E-08	1.40E-05	NIPSNAP3B
cg16174680	0.015	0.017	-1.715	1.66E-09	3.08E-06	NKX6-3
cg05421673	0.054	0.058	-1.391	8.33E-07	3.78E-04	NOG
cg14654471	0.918	0.878	-1.389	8.64E-07	3.88E-04	NPHP4
cg02739500	0.041	0.044	-1.625	1.03E-08	1.30E-05	NPW
cg22367989	0.028	0.030	-1.429	4.23E-07	2.31E-04	NRP2
cg01164094	0.934	0.946	1.367	1.84E-06	6.76E-04	NT5DC3
cg20405174	0.022	0.025	-1.719	1.51E-09	2.90E-06	PACSIN1
cg13670316	0.055	0.059	-1.556	4.00E-08	3.74E-05	PCSK5
cg27543403	0.030	0.033	-1.457	2.56E-07	1.59E-04	PDE4D
cg11153768	0.958	0.951	-1.846	9.86E-11	3.03E-07	PDE6B
cg05266497	0.937	0.932	-1.568	3.18E-08	3.08E-05	PGBD5
cg26268742	0.902	0.911	1.420	7.51E-07	3.49E-04	PLA2G4C
cg04202736	0.957	0.949	-1.473	1.90E-07	1.26E-04	PLA2G4D
cg01078276	0.036	0.039	-1.332	2.33E-06	8.18E-04	PLAU;C10orf55
cg16393928	0.952	0.947	-1.880	4.72E-11	1.61E-07	PRDM16
cg13583586	0.919	0.931	1.491	2.17E-07	1.40E-04	PTDSS2
cg08274544	0.967	0.959	-1.578	2.60E-08	2.65E-05	PTPN21
cg01878724	0.078	0.062	1.579	4.43E-08	4.02E-05	RAET1L
cg23231670	0.048	0.037	1.831	3.13E-10	7.68E-07	RAET1L
cg08119607	0.150	0.138	1.678	6.73E-09	9.53E-06	RAET1L
cg03118417	0.915	0.899	-2.291	2.38E-15	2.69E-11	RANBP3
cg03711182	0.017	0.020	-1.632	8.98E-09	1.18E-05	RASGRF1
cg05944207	0.051	0.059	-1.687	2.96E-09	4.95E-06	RASL10A
cg27395839	0.951	0.919	-1.349	1.73E-06	6.52E-04	RORA
cg24132791	0.062	0.052	1.459	3.81E-07	2.15E-04	RPSAP58
cg04490178	0.034	0.037	-1.519	8.02E-08	6.25E-05	S1PR4
cg25619287	0.742	0.763	1.771	1.08E-09	2.19E-06	SDF4
cg08129848	0.021	0.024	-1.695	2.51E-09	4.38E-06	SEPHS2
cg08113562	0.038	0.047	-1.564	3.40E-08	3.24E-05	SH3D20
cg03610867	0.088	0.097	-1.721	1.45E-09	2.81E-06	SHH
cg09327378	0.935	0.927	-1.410	5.94E-07	2.89E-04	SLC5A1
cg00007644	0.097	0.105	-1.450	2.90E-07	1.75E-04	SLC6A5
cg13823936	0.907	0.888	-1.423	4.76E-07	2.49E-04	SPOPL
cg25533220	0.041	0.044	-1.456	2.61E-07	1.61E-04	SPPL2B;LSM7
cg06092244	0.874	0.889	1.400	1.05E-06	4.51E-04	STON1-GTF2A1L

cg05806180	0.054	0.065	-1.495	1.27E-07	8.86E-05	SULF1
cg05845592	0.037	0.030	1.388	1.29E-06	5.26E-04	SULT1A1
cg22436429	0.049	0.052	-1.364	1.34E-06	5.40E-04	TFAP2E
cg14557510	0.037	0.040	-1.658	5.35E-09	7.93E-06	TFCP2L1
cg02567119	0.029	0.032	-1.691	2.70E-09	4.63E-06	TLX1;TLX1NB
cg03548490	0.955	0.949	-1.545	4.88E-08	4.31E-05	TMC7
cg08466517	0.937	0.945	2.073	1.68E-12	9.63E-09	TNIP3
cg07965986	0.949	0.955	2.328	4.29E-15	4.68E-11	TRAK1
cg06900571	0.968	0.963	-1.775	4.66E-10	1.10E-06	TRAPPC9
cg22309983	0.956	0.952	-1.776	4.58E-10	1.08E-06	TRPV1
cg22490255	0.054	0.057	-1.413	5.70E-07	2.81E-04	TSPAN31
cg18106434	0.069	0.057	1.453	4.29E-07	2.33E-04	TUBGCP5
cg22670128	0.866	0.848	-1.509	9.80E-08	7.26E-05	UBE2J2
cg19908577	0.042	0.046	-1.585	2.28E-08	2.38E-05	UGT8
cg18646207	0.055	0.059	-1.365	1.32E-06	5.37E-04	VAX1
cg08584627	0.060	0.064	-1.349	1.73E-06	6.51E-04	VWA3B
cg22867714	0.903	0.916	1.371	1.73E-06	6.51E-04	VWF
cg00405190	0.355	0.345	1.703	4.15E-09	6.49E-06	WIPF1
cg18579862	0.043	0.048	-1.544	5.05E-08	4.41E-05	ZIK1
cg19033875	0.027	0.029	-1.364	1.33E-06	5.39E-04	ZNF395
cg03405173	0.033	0.037	-1.344	1.88E-06	6.90E-04	ZNF441
cg25591573	0.061	0.069	-1.950	9.71E-12	4.00E-08	ZNF442
cg08847636	0.057	0.065	-1.873	5.44E-11	1.79E-07	ZNF442
cg00850039	0.038	0.041	-1.496	1.25E-07	8.70E-05	ZNF442
cg21727178	0.031	0.036	-1.825	1.59E-10	4.53E-07	ZNF709
cg02878907	0.028	0.033	-1.901	2.96E-11	1.11E-07	ZNF709
cg02934221	0.026	0.030	-1.511	9.35E-08	7.08E-05	ZNF709
cg17317439	0.043	0.047	-1.371	1.18E-06	4.94E-04	ZNF763
cg18483269	0.026	0.028	-1.689	2.80E-09	4.76E-06	ZNF823
cg04681963	0.057	0.064	-1.411	5.85E-07	2.87E-04	ZNF879

Table 12: Differentially Variably Methylated Positions (DVMPs), as computed by iEVORA. Variance is assigned as ±1.96 standard deviations (95% of a normal distribution), difference in variance is displayed as log[V1/V0], and p and FDR-corrected q values are displayed with a significance threshold of 0.001.

Goldmine analysis revealed that 53.6% of DVMPs were situated in gene promoters, 48.0% in CpG Islands, and 66.7% in transcription factor binding sites. Only 12.3% were situated in enhancers. Genomic and Feature associations are displayed in Figure 31.

I used Enrichr to perform enrichment analysis of genes overlapping DVMPs. Based on the ChEA dataset I observed enrichment of factors which target the polycomb protein SUZ12 (p = 1.6x10-7, Q=1.5x10-5, Z-score=-1.54, Combined Score=24.12), and EZH2 (p = 1.3x10-6, Q=6. 1x10-5, Z-score=-1.75, Combined Score=23.76) which alongside SUZ12 forms part of Polycomb Repressive Complex 2.

I also analysed genes overlapping DVMPs in terms of ontology using DAVID, though no ontologies were significantly associated after Benjamini-Hochberg correction for multiple testing.

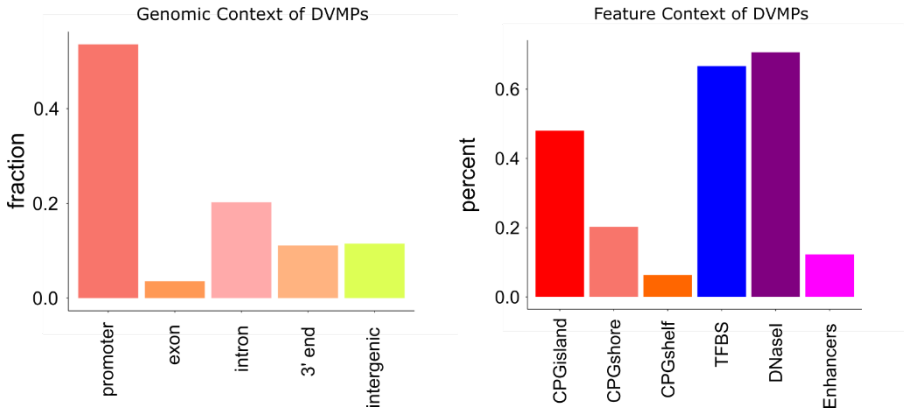


Figure 31: Genomic and Feature contexts of DVMPs computed using ChAMP results using iEVORA and ChIP-Seq data from the Roadmap Epigenomics Consortium. CpG Islands are regions of >200bp containing >50% GC composition and an observed to expected CpG ratio of >60%. Shores are regions 2kb from a CpG Island and Shelves are regions 2kb from a CpG Shore.

3.4.7 Inferring cellular origin of signal through overlap between DMPs and underlying cell-specific chromatin modifications

eFORGE (Breeze *et al.*, 2016) is a tool which associates CpG sites with chromatin modifications known to be present in a wide range of tissue types. I input the differentially methylated probes to eFORGE and found significant (Adjusted p > 0.001) associations between DMPs and H3K36me3 marks in Primary Monocytes and Primary B Cells, as well as marginal association (0.05 > p > 0.001) with the same mark in Primary Natural Killer Cells. This suggests a population-specific origin of differential methylation in whole blood which informed work detailed in later chapters.

3.4.8 Results: Differentially Methylated Regions in sporadic CJD

Regional differences in DNA methylation are thought to be particularly biologically relevant as they may reflect areas of epigenetic dysregulation which may overlap multiple DNA-protein interaction sites, or be indicative of a region of underlying chromatin modifications (Jones, 2012). I first used Bumphunter (Jaffe *et al.*, 2012) to tile adjacent probes and test for regional differences in methylation with Bonferroni threshold reduced by virtue of a reduction in data dimensionality through tiling. I identified 41 differentially methylated regions (DMRs), detailed in Table 13. The top-ranked DMR consisted of 48 probes (2379nt) overlapping the *HOXA5* promoter region (FWER = 0.012, $p = 3.3 \times 10^{-5}$), while subsequent DMRs had higher Family-Wise Error Rates and showed lower effect sizes between case and control. Genes overlapped by DMRs were input to Enrichr (E. Y. Chen *et al.*, 2013; Kuleshov *et al.*, 2016) for gene set enrichment analysis, which revealed a significant association with Leukotriene receptor partners ($p = 8.5 \times 10^{-5}$, $q = 0.007$, *Z*-score = -1.55, Combined Score = 14.57).

To assess the biological context of these DMRs, sequences were input to Goldmine using the previously mentioned Roadmap Consortium data. I found that all DMRs overlapped with transcription factor binding sites, and around 40% of DMRs overlapped enhancers in Monocytes, CD4+T Cells, CD8+ T Cells, B Cells, Neutrophils or Natural Killer Cells (Figure 32). Then I used MEME-Suite (Bailey *et al.*, 2009) to identify two motifs which are enriched in the DMRs (Figure 33), which I then associated with Gene Ontology Terms using MEME-Suite's GOMo function (Buske *et al.*, 2010). These are detailed in Table 14. To examine whether these motifs were represented in lists of transcription factor binding sites, I performed TOMTOM analysis which yielded two lists of transcription factors (Table 15) which I also input to Enrichr. The gene set of transcription factors with a *q*-value<0.05 (all of which corresponded to the first motif) was found to be significantly enriched for factors which target the polycomb protein SUZ12 ($p = 8.2 \times 10^{-9}$, $q = 6.0 \times 10^{-7}$, *Z*-score = -1.54, Combined Score = 28),based on the ChEA dataset (Lachmann *et al.*, 2010). These findings are intriguing as SUZ12 is involved in *HOXA5* silencing and suggest dysregulation of epigenetic control of *HOXA5* in sCJD.

DMR Number	Coordinates	Width	Strand	Value	Area	Cluster ID	Number of Probes	*p* Value	FWER	Genes
1	chr7:27183133-27185512	2379	*	0.195	9.337	183676	48	3.30E-05	0.012	HOXA-AS3; HOXA5; HOXA6
2	chr6:32063394-32064923	1529	*	0.122	5.742	171574	47	0.0001538	0.056	TNXB
3	chr6:32847377-32847845	468	*	0.180	3.068	171850	17	0.0001868	0.064	PPP1R2P1
4	chr6:31691198-31692365	1167	*	0.129	3.341	171384	26	0.0005054	0.172	C6orf25
5	chr6:41068646-41068752	106	*	-0.412	2.469	173080	6	0.0006702	0.22	NBR2
6	chr17:41277707-41278712	1005	*	0.131	2.744	89566	21	0.0006812	0.228	TSPAN32; C11orf21
7	chr11:2322500-2323459	959	*	0.122	2.321	32867	19	0.0013953	0.388	LY6G5C
8	chr6:31650735-31651362	627	*	0.113	2.267	171357	20	0.0014173	0.392	OR2L13
9	chr1:248100183-248100614	431	*	0.312	2.498	21131	8	0.001637	0.464	RP11-438H8.8
10	chr13:36871646-36872346	700	*	0.214	2.788	56535	13	0.001659	0.5	CCDC169-SOHLH2; CCDC169; SOHLH2
11	chr5:135415693-135416613	920	*	-0.206	2.884	163340	14	0.001681	0.516	VTRNA2-1
12	chr6:33084608-33085063	455	*	0.142	2.123	171927	15	0.0019227	0.52	HLA-DPB2
13	chr6:32632788-32633157	369	*	0.207	2.688	171780	13	0.0017909	0.528	HLA-DQB1
14	chr7:27169674-27170880	1206	*	0.138	2.070	183666	15	0.0020875	0.548	HOXA4; RP1-170O19.22; HOXA3; HOXA-AS3; HOXA-AS2
15	chr19:55972646-55973338	692	*	-0.258	2.319	107862	9	0.002538	0.628	ISOC2
16	chr11:70672365-70673256	891	*	0.216	2.372	39462	11	0.0026808	0.684	SHANK2
17	chr22:51016501-51017723	1222	*	0.156	2.187	134794	14	0.0029884	0.74	CPT1B; CHKB-CPT1B; CHKB
18	chr3:182817190-182817626	436	*	0.178	1.961	145497	11	0.0045376	0.824	MCCC1
19	chr18:77905119-77905947	828	*	0.192	1.919	98162	10	0.0047683	0.84	ADNP2; AC139100.2
20	chr17:6899085-6899888	803	*	0.134	1.878	85516	14	0.004153	0.852	ALOX12; AC027763.2; RP11-589P10.7

Table 13: DMRs calculated using Refbase-corrected 450K data and bumphunter.

DMR Number	Coordinates	Width	Strand	Value	Area	Cluster ID	Number of Probes	*p* Value	FWER	Genes
1	chr7:27183133-27185512	2379	*	0.195	9.337	183676	48	3.30E-05	0.012	HOXA-AS3; HOXA5; HOXA6
2	chr6:32063394-32064923	1529	*	0.122	5.742	171574	47	0.0001538	0.056	TNXB
3	chr6:32847377-32847845	468	*	0.180	3.068	171850	17	0.0001868	0.064	PPP1R2P1

DMR Number	Coordinates	Width	Strand	Value	Area	Cluster ID	Number of Probes	*p* Value	FWER	Genes
21	chr11:117069780-117070046	266	*	0.222	1.780	42226	8	0.0057351	0.904	TAGLN
22	chr8:143859410-143859990	580	*	0.218	1.525	202288	7	0.0073722	0.94	LYNX1
23	chr15:81426347-81426669	322	*	0.181	1.626	72818	9	0.009163	0.96	CC15orf26
24	chr17:33759484-33760419	935	*	0.144	1.584	88187	11	0.0078886	0.96	SLFN12
25	chr6:32729105-32729659	554	*	0.109	1.421	171798	13	0.0077457	0.96	HLA-DQB2
26	chr10:530635-531569	934	*	0.142	1.559	21328	11	0.0082182	0.968	DIP2C
27	chr10:4868328-4869088	760	*	0.171	1.543	21999	9	0.010855	0.984	AKR1E2
28	chr5:178986131-178986906	775	*	0.156	1.402	167296	9	0.0144257	0.988	RUFY1
29	chr12:123215119-123215684	565	*	0.140	1.405	53316	10	0.011657	0.992	HCAR1
30	chr10:131567073-131568021	948	*	0.140	1.399	30761	10	0.0118218	0.992	RP11-109A6.3
31	chr4:5021084-5021328	244	*	0.171	1.369	148321	8	0.0161726	1	CYTL1
32	chr6:28602543-28602745	202	*	0.152	1.366	170405	9	0.0157332	1	
33	chr14:24780167-24780825	658	*	0.148	1.332	61366	9	0.0168538	1	CIDEB; LTB4R; LTB4R2
34	chr8:144659831-144660772	941	*	0.166	1.328	202569	8	0.0179195	1	NAPRT1
35	chr13:25670047-25670327	280	*	0.140	1.258	55700	9	0.0196774	1	PABPC3
36	chr2:84743445-84743743	298	*	0.135	1.214	114700	9	0.02171	1	DNAH6
37	chr1:1108820-1109676	856	*	0.151	1.209	236	8	0.0243578	1	TTLL10; TTLL10-AS1
38	chr1:247681242-247681931	689	*	0.169	1.180	21082	7	0.020908	1	GCSAML
39	chr4:99850801-99851211	410	*	0.168	1.177	152481	7	0.0210728	1	EIF4E; AC019131.1; RP11-571L19.7
40	chr2:241458886-241459847	961	*	0.162	1.133	124636	7	0.0240831	1	ANKMY1
41	chr2:239139911-239140369	458	*	-0.113	1.125	124149	10	0.0191281	1	AC016757.3; AC096574.4

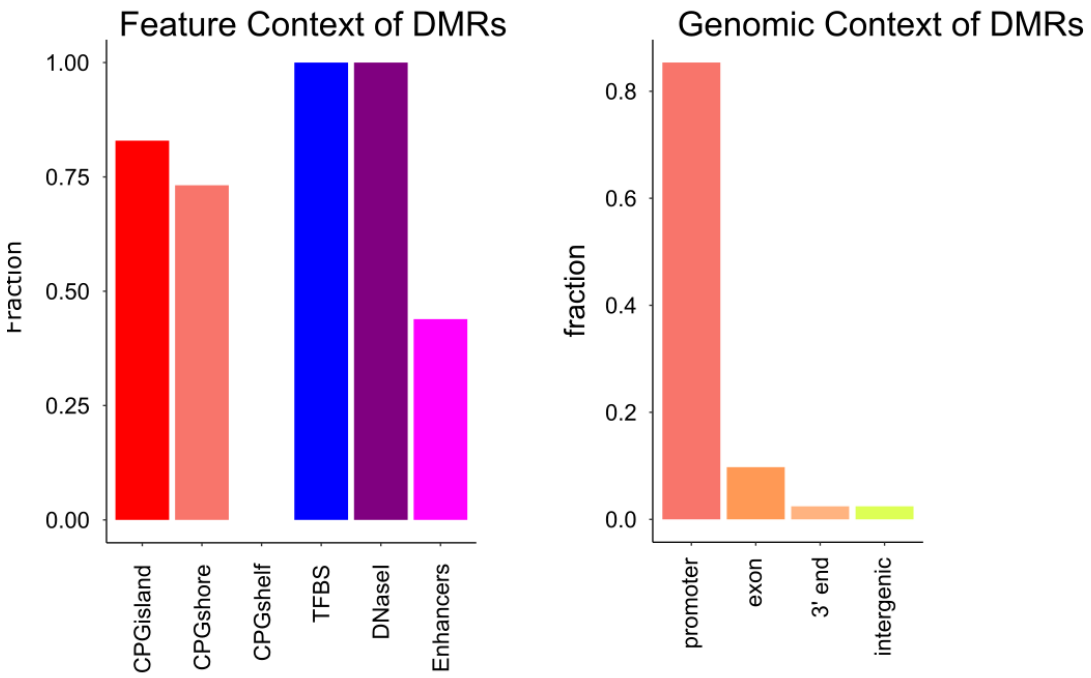
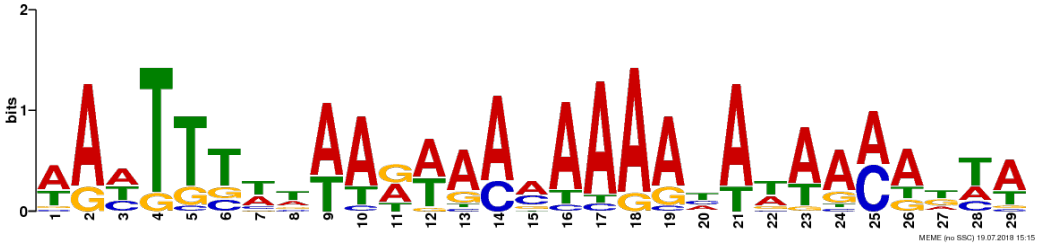


Figure 32: Feature and Genomic context of DMRs, computed using bumphunter results and ChIP-Seq data from the Roadmap Epigenome Consortium (Roadmap Epigenomics Consortium *et al.*, 2015). *CpG Islands are regions of >200bp containing >50% GC composition and an observed to expected CpG ratio of >60%. Shores are regions 2kb from a CpG Island and Shelves are regions 2kb from a CpG Shore.*

Motif 1



Motif 2

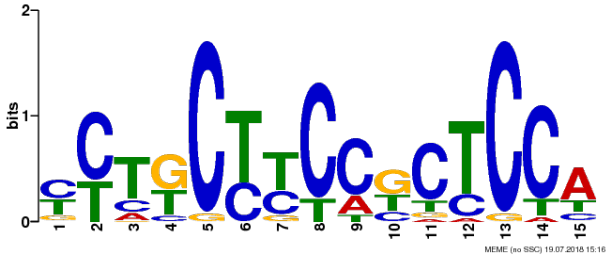


Figure 33: Motifs enriched in DMR sequences as called by MEME-Suite.

Motif	Associated GO Terms	5 most significant GO Terms	GO Type	*p* val	*q* Val	Specificity (%)
1	11	Olfactory receptor activity	MF	8.48e-08	1.24e-04	100
		Sensory perception of smell	BP	8.48e-08	1.24e-04	88
		Sensory perception of chemical stimulus	BP	8.48e-08	1.24e-04	25
		G-protein coupled receptor protein signaling pathway	BP	8.48e-08	1.24e-04	7
		Gene Expression	BP	2.19e-05	2.32e-02	1
2	673	Calcium ion binding	MF	8.48e-08	5.06e-06	100
		Protein heterodimerization activity	MF	8.48e-08	5.06e-06	100
		Positive regulation of transcription from RNA polymerase II promoter	BP	8.48e-08	5.06e-06	58
		Lung development	BP	8.48e-08	5.06e-06	16
		Signal transduction	BP	8.48e-08	5.06e-06	8

Table 14: Gene Ontology Results from MEME-SUITE's GOMo function. Term types are organised into Molecular Function (MF) or Biological Process (BP).

Motif	Gene	*p* value	*q* value	Overlap	Target Overlap	Strand
1	*CPEB1*	8.53E-08	2.15E-04	17	AAAAAAAAAAATAAAAA	-
1	*FUBP1*	5.45E-06	3.92E-03	12	AAAAAAACACAA	-
1	*ARI3A*	8.22E-06	3.92E-03	22	AATTAATCGAAATCAAATTAAA	+
1	*FOXL1*	9.96E-06	3.92E-03	19	AAAAGTAAACAAACAAACA	-
1	*ANDR*	1.09E-05	3.92E-03	18	AGCAAACAAAAAAGAACA	-
1	*FOXG1*	1.40E-05	4.42E-03	17	AAAAAAATATAAACAAT	-
1	*SRY*	2.69E-05	6.79E-03	9	AAAACAAAA	-
1	*LMX1A*	5.94E-05	0.01	19	TTAATTAAATAATTAATTA	-
1	*HXC10*	6.38E-05	0.01	19	GTAATAAAAAAATTAAAAA	-
1	*FOXJ3*	9.01E-05	0.02	13	AAAAAATAAACAA	-
1	*PRDM6*	1.36E-04	0.02	13	AAAAAGAAAAAAA	+
1	*AIRE*	1.68E-04	0.02	18	TTAACCAATATAACCAAT	-
1	*HMX1*	2.79E-04	0.04	17	ATTAAAAAGCAATTAAC	-
1	*FOXF1*	3.24E-04	0.04	11	AAAATAAACAT	-
1	*PO3F3*	3.91E-04	0.05	17	AAATTTGCATAATTTAT	-
1	*FOXJ3*	4.20E-04	0.05	13	TAAACAAAAACAA	-
1	*ZN713*	5.25E-04	0.06	19	ATAGAAAAAAGACATGAAA	+
1	*FOXJ2*	6.19E-04	0.06	10	TAAATAAACA	-
1	*FOXP1*	8.24E-04	0.08	9	ATAAACAAA	-
1	*ONEC3*	8.38E-04	0.08	12	AAAAAATCAATA	-
1	*Z354A*	9.60E-04	0.08	24	ACATTAAATGTAAATGGACTAAAT	-
1	*SOX5*	1.07E-03	0.08	8	TAACAATA	-
1	*ONEC2*	1.11E-03	0.08	20	AAAAAAAATCAATAACAAGAC	-
1	*LHX9*	1.14E-03	0.08	18	TAATTAATAGCTAATTAG	-
1	*STAT1*	1.22E-03	0.08	17	AAGAAAATGAAACTGAAAG	+
1	*BPTF*	1.31E-03	0.09	12	GAAAACAACAAA	-
1	*STAT2*	1.53E-03	0.10	17	AGGAAAATGAAACTGAAAG	+
1	*NFAC1*	1.75E-03	0.11	15	ATGGAAAAAAAGAAA	+
1	*SHOX*	2.01E-03	0.12	17	AAAAATTAACTAATTAG	-
1	*HXD11*	2.22E-03	0.13	11	AGTAATAAAAA	-
1	*DLX1*	2.47E-03	0.14	16	TAATTAGCATAATTTA	-
1	*HXC6*	2.55E-03	0.14	15	AAAGTAATAAATCAT	-
1	*HMX2*	2.78E-03	0.15	11	AACCAATTAAA	-
1	*FOXD1*	2.89E-03	0.15	15	CTTAAGTAAACAAAG	-
1	*FOXD2*	3.00E-03	0.15	19	TTAAATAAATATTTACATA	-
1	*MEF2D*	3.81E-03	0.18	12	CTAAAAATAGCA	-
1	*FOXD3*	3.92E-03	0.18	15	TGCTAAGTAAACAAA	-
1	*MEF2A*	4.28E-03	0.19	13	GCTAAAAATAGAA	-
1	*SHOX2*	4.52E-03	0.20	17	TAAAATTAACTAATTAG	-
1	*FOXQ1*	4.67E-03	0.20	12	AAATAAACAATT	-
1	*IRF1*	5.22E-03	0.21	19	GAAAATGAAAGTGAAAGTAA	+
1	*FOXP3*	5.59E-03	0.22	9	AAACAAATT	-
1	*GSX1*	5.60E-03	0.22	15	ATTAAAAACTAATTA	-
1	*HNF6*	5.73E-03	0.22	11	AAATCAATAAA	-
1	*MEF2C*	5.90E-03	0.22	13	GCTAAAAATAGCA	-

1	*MNX1*	6.33E-03	0.23	18	TTAAGGGTTATTAAAATAG	-
1	*PO3F4*	6.65E-03	0.24	17	TAATTTGCATAATTTAT	+
1	*ZFP28*	7.06E-03	0.25	20	TGACACAAGAAGAAATAGAA	-
1	*EVX2*	7.13E-03	0.25	13	AAGGCCATAAAAC	-
1	*MYNN*	7.40E-03	0.25	17	TTTCAAAATAAAAGTCC	-
1	*LEF1*	7.58E-03	0.25	14	AGCAAATCAAAGGA	-
2	*NKX25*	1.72E-04	0.08	8	CCTCTCCA	-
2	*ETS2*	1.96E-04	0.08	13	TCCTCTTCCTTCC	-
2	*E2F7*	2.66E-04	0.08	13	CCTTTCCCGCCCC	-
2	*ZN467*	2.86E-04	0.08	15	CCCCCCCCCCCCTCCCCTCCCC	-
2	*PTF1A*	3.76E-04	0.08	15	CCAGCTGCCCCCTTTCCC	+
2	*VEZF1*	3.83E-04	0.08	15	CCCCTCCCCCTCCCCCCTCCCC	-
2	*E2F1*	3.83E-04	0.08	14	CTTTCCCGCCCCCC	-
2	*SP4*	5.00E-04	0.08	15	CCCGGCCCCGCCCCCTTCCC	-
2	*ZBTB6*	5.07E-04	0.08	13	CGGCTCCAGCACC	-
2	*PATZ1*	5.08E-04	0.08	14	CCTCCCCCCCCCGCCCCCTCCCC	-
2	*WT1*	6.29E-04	0.08	15	CCCCCCCTCCTCCCCCGCCC	-
2	*TFDP1*	8.79E-04	0.11	14	TTTTCCCGCCCCCC	-
2	*E2F6*	1.18E-03	0.13	13	CCCTTCCCGCCCC	-
2	*KLF15*	1.38E-03	0.13	15	CCCCCCCCTGCTCCTCCCC	-
2	*ZFX*	1.45E-03	0.13	15	CCCCGGCCTCCGCCCCC	-
2	*MAZ*	1.50E-03	0.13	15	CCCCCCCCCCCCCCCCCTCCCCC	-
2	*ZN148*	1.86E-03	0.15	14	CCCCTCCCCCACCCC	-
2	*ZBT17*	1.94E-03	0.15	14	CTTCCCCTCCCCCACCCTC	-
2	*EGR2*	2.10E-03	0.15	14	CCCCCGCCCACGCCCC	-
2	*CREB3*	2.11E-03	0.15	13	TGCCACGTCACCA	-
2	*ERG*	2.41E-03	0.17	13	CCACTTCCTGCCC	-
2	*ZN263*	2.64E-03	0.18	15	CTCCTCCTCTCCCTCCTCCC	-
2	*ZN341*	3.22E-03	0.21	15	GCTCTTCCCTCCCCCCCCCCCC	-
2	*FLI1*	3.84E-03	0.23	13	CCACTTCCTGCCT	-
2	*EGR1*	3.85E-03	0.23	15	CCCCCGCCCACGCCCTC	-
2	*MAZ*	3.90E-03	0.23	11	CCCCCTCCCTC	-
2	*EGR2*	5.06E-03	0.28	15	CCCCTCCCACACCCCCCC	-
2	*NFAC1*	5.92E-03	0.31	14	TTTCTTTTTTTCCAT	-
2	*SP2*	6.10E-03	0.31	14	CCCCCGGCCCCGCCCCCCCCCC	-
2	*ETV5*	6.78E-03	0.31	14	CTCACTTCCTGCTC	-
2	*ETS1*	6.80E-03	0.31	13	CCACTTCCTGTCT	-
2	*NR0B1*	7.22E-03	0.31	10	TCTCCCACGC	-
2	*NFAT5*	7.30E-03	0.31	9	CCTTTTCCTCT	-
2	*ZN740*	7.32E-03	0.31	14	CCCACCCCCCCCCC	-
2	*VEZF1*	7.44E-03	0.31	11	CCCCCTCCCCCT	-

Table 15: Transcription Factors which recognise 2 motifs enriched in the DMR dataset as called by TOMTOM. p values are calculated through comparison of the query motif (1 or 2) with target overlap per transcription factor, while q value presents the significance after adjustment for False Discovery Rate.

3.5 Summary

During the early stage of the project I optimised a low-input high-yield protocol for extraction of genomic DNA from blood, as well as a bisulphite conversion protocol which produced bisulphite converted DNA of sufficient quality for 450K Beadchip array analysis. I went on to perform the first genome wide methylation association study conducted on human prion disease. I identified 38 Differentially Methylated Positions (DMPs), 41 Differentially Methylated Regions (DMRs) and 252 Differentially Variably Methylated Positions (DVMPs). As DVMPs are of ambiguous biological relevance and the DMRs I identified have high Family-Wise Error Rates, the most reliable results from this study are the DMPs, which will be taken forward for further analysis. Pathway, enrichment and motif analyses produced some positive associations between results and pathways, ontologies, functions and common sequences, but these associations were slight and not particularly convincing. Loci associated with sCJD were associated with disease severity and rate of decline, as loci unchanged between disease and control were thought unlikely to correlate with either metric under the study's statistical power. An association between demethylation at the *AIM2* promoter and increased disease severity shows promise as a peripheral prognostic biomarker. Effects seen in blood-derived DNA are relatively large compared to those seen in published studies of neurodegeneration. Yet genome-wide significance inflation was higher than optimal, and so to exclude false positives replicating these findings in a second study of DNA methylation in blood taken from a separate cohort of sCJD patients and controls was planned, as described in the next chapter.

4 Validating, replicating and exploring 450K Beadchip Array findings using CpG-sensitive pyrosequencing

Having successfully identified 38 loci differentially methylated between sporadic CJD patients and control volunteers, the logical next step was to replicate findings using a second cohort of samples and a different technology. This serves several purposes, the first being to minimise the false positive risk by controlling for variability unique to the first cohort or technically inherent to the 450K Beadchip array. Secondly, by using a sequencing approach rather than a probe-based approach, locally proximal CpG sites can also be interrogated. As CpG methylation can be highly locally correlated this provides the opportunity to strengthen the impact of the discovery findings by identifying additional adjacent loci with similarly differential levels of methylation (L. Zhang *et al.*, 2017; Song, Ren and Lei, 2017). Finally, as site-specific sequencing data cannot be adjusted for cellular heterogeneity or subjected to the preprocessing data from the 450K Beadchip array was, these results being an artifact of data manipulation can be ruled out.

In this chapter I aim to design pyrosequencing assays for biologically interesting loci identified during the previous chapter, validate pyrosequencing's complementarity to the 450K Beadchip array by assaying samples whose methylation levels at cg10636246 (AIM2) and cg03546163 (FKBP5) had been previously investigated, and replicate discovery study results in a second set of different sCJD patients and controls. Finally, I aim to test the specificity of any successfully replicated findings to sporadic CJD by performing the replication assays on blood-derived DNA from Alzheimer's Disease, iatrogenic CJD and Inherited Prion Disease patients.

4.1 Relevant Methods and Sample Demographics

Group	Number	Average Age (range)	Sex (% F)	Codon 129 (%) MM:MV:VV	Average MRC Scale Score (range)
sCJD	72	59.9 (26-86)	58.3	54:23:23	4.5 (0-20)
Control	114	78.2 (61-93)	64.9	Unknown	20
AD	59	71.8 (58-87)	47.5	Unknown	NA
iCJD	18	46.4 (41-53)	11.1	27:73:0	11.7 (1-18)
IPD	11	48.5 (38-68)	72.7	44:44:12	18.6 (13-20)

Table 16: Demographics of patients and controls whose samples were used in the replication/specificity phase.

4.2 Assay design

4.2.1 Probe Selection

Technologies for replication of observed effects at single loci are by definition less high-throughput than the 450K array. This places a limitation on the number of DMPs which could be brought forward to replication and necessitates selection of DMPs from the list of 38. I selected DMPs for replication based on statistical significance, effect size, genomic location and features, and biological function. Two probes in *AIM2* (cg10636246 & cg17515347) were selected due to their presence in the gene's promoter region and their association with MRC Scale score (Section 3.3.4). A CpG bordering the promoter of *FKBP5* (cg03546163) was selected as it exhibited the largest case-control difference ($\Delta\beta$ = - 0.055). Two additional probes overlapped by *FKBP5* (cg00052684 & cg25114611) were identified as differentially methylated using the EpiDISH algorithm (Box 2) and were also chosen for replication. A CpG in the body of *UHRF1* (cg04286737) and another in the body of *METTL9* (cg02481950) were selected due to their biological function: *UHRF1* facilitates DNMT1 recruitment to hemi-methylated DNA whilst *METTL9*, while uncharacterised, bears homology to methyltransferases. A probe in the body of *KCNAB2* (cg02448796) was chosen for replication as bioinformatic analysis of sCJD GWAS data (currently unpublished, personal communication from Dr Holger Hummerich, MRC Prion Unit at UCL) indicated an enrichment in factors associated with potassium channels, and a CpG upstream of *MTRNR2L8* – Humanin-like protein 8 – was also chosen due to previous association with humanin-like proteins with neuroprotection in prion disease and Alzheimer's disease (Hashimoto *et al.*, 2001; Sponne *et al.*, 2004; Bodzioch *et al.*, 2009). Because of their relevance to prion disease and epigenetics of neurodegeneration, and partly because no

statistically significant change was observed at these loci in the discovery stage, probes in *ANK1* (cg11823178) and *PRNP* (cg00168514) were also selected (Lloyd, Mead and Collinge, 2011; De Jager *et al.*, 2014; Lunnon *et al.*, 2014). Distribution of beta values between groups at these probes are visualised in Figure 34 and a summary of selection rationale is displayed in Table 17.

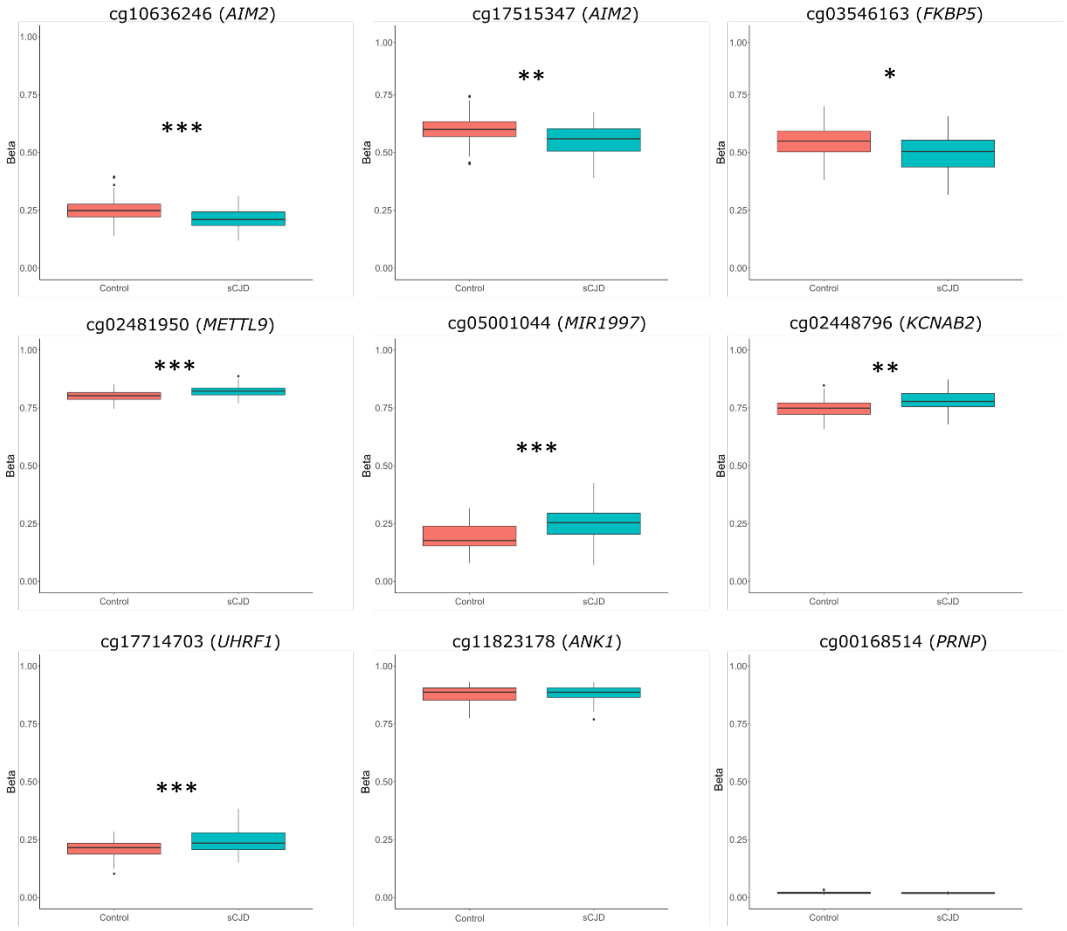


Figure 34: Array-derived methylation values at selected DMPs, as well as loci in ANK1 and PRNP which showed no significant change between control (red) and sCJD (turquoise) groups. Asterisks denote Bonferroni-adjusted significance, where an adjusted threshold of 0.05 has been surpassed (= p < 0.05, ** = p < 0.01, *** = p < 0.001)*

Probe	Gene	Bonferroni Significant?	CpG within TSS?	|Δβ|> 0.03?	Functional connection to:	
					Neurodegeneration?	Epigenetics?
cg10636246	*AIM2*	Yes	Yes	Yes	Yes	No
cg02481950	*METTL9*	Yes	No	No	No	Yes
cg05740793	*MTRNR2L8*	Yes	Yes	Yes	Yes	No
cg05001044	*MIR1977*	Yes	Yes	Yes	No	Yes
cg02448796	*KCNAB2*	Yes	No	Yes	No	No
cg17714703	*UHRF1*	Yes	No	Yes	No	Yes
cg17515347	*AIM2*	Yes	Yes	Yes	Yes	No
cg03546163	*FKBP5*	Yes	Yes	Yes	Yes	No
cg11823178	*ANK1*	No	No	No	Yes	No
cg00168514	*PRNP*	No	No	No	Yes	No

Table 17: Criteria for replication as fulfilled per CpG site chosen for replication.

4.2.2 Bisulphite-PCR primer design

Producing an amplicon less than 200 bp in length at probe sites surrounded by particularly dense CpG clusters is technically challenging as forward and reverse PCR primers cannot overlap CpG sites. This would bias the reaction towards amplification of methylated or unmethylated fragments, depending on whether the complementary nucleotide at the CpG's cytosine is A (unmethylated) or G (methylated) after bisulphite conversion. This was the case for many of the *PRNP* probes, including cg00168514 which had been chosen as a negative control. Instead an assay was designed for cg04286737, which is located upstream of *PRNP*'s TSS. Fortunately, an assay for the second negative control CpG in *ANK1* (cg11823178) was designed without difficulty.

MTRNR2L8 presented a different challenge, as it is one of 15 nuclear-encoded descendants of a mitochondrial peptide (Bodzioch *et al.*, 2009). As such the DNA sequence surrounding cg05740793 was highly conserved across the Humanin homologues and generating a specific amplicon shorter than 200 bp proved impossible. Moreover, the bordering sequences were enriched for cytosine and thymine, which after bisulphite conversion resulted in stretches of thymine homo-polymers which are refractory to primer design. This is due in part to increased sequence redundancy and thus reduced hybridisation specificity, but also because interactions between A/T rich sequences are weaker than those between G/C rich sequences. Nested bisulphite PCR of a larger specific amplicon and then production of a shorter amplicon for sequencing was attempted but in all cases the second

reaction failed. Regrettably upstream differential methylation of *MTRNRL2L8* was therefore not able to be replicated. The local genomic contexts of *PRNP* and *MTRNRL2L8* probes as displayed in the PyroMark assay design software are shown in Figure 35.

Despite the technical challenges of designing primers for bisulphite PCR followed by sequencing, assay design for the remaining selected probes was less problematic and in some instances allowed profiling of multiple adjacent CpG sites (up to a total of 5 CpGs in the *ANK1* assay, also shown in Figure 35).

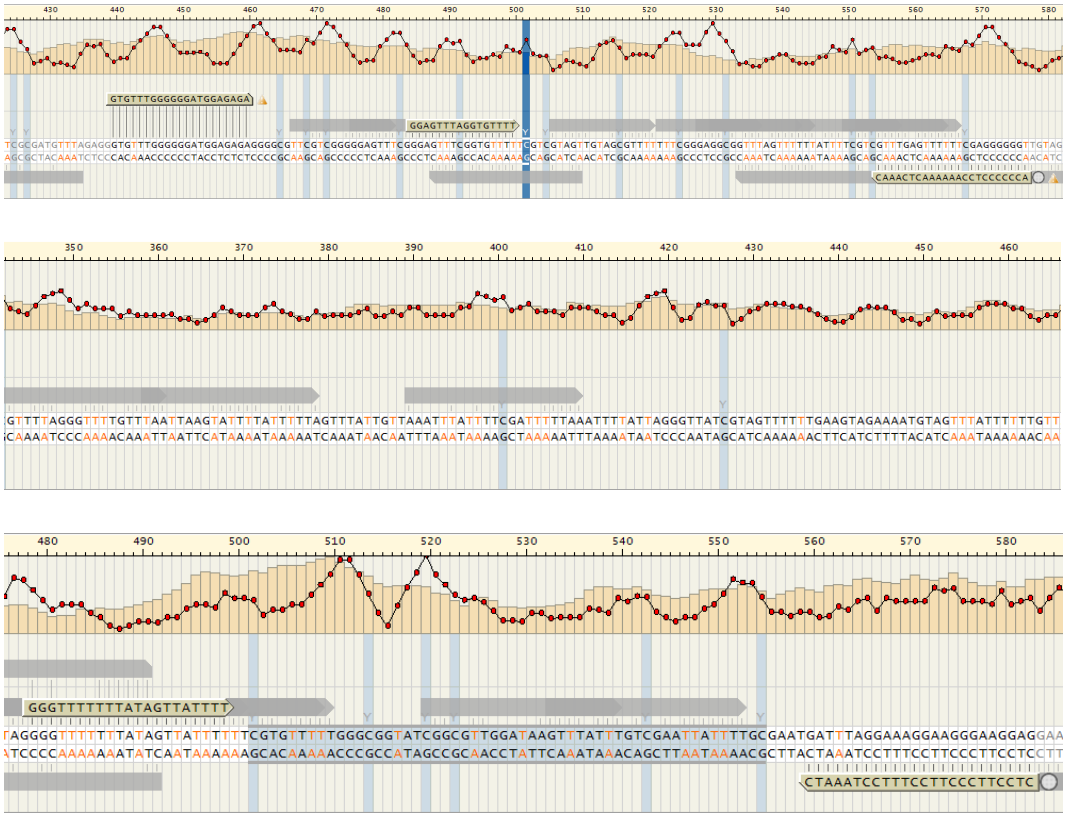


Figure 35: Screenshots from PyroMark Assay Design 2.0 software. Numbers correspond to nucleotide position around the CpG of interest (CpG ± 500 nt), orange bars and red circles correspond to melting temperature and Gibbs free energy per location. Blue bars correspond to CpG cytosines, shaded blue sequence represents target sequences. Grey shadows and beige blocks correspond to potential primer positions and generated primer sequences. (i) PRNP probe cg00168514 (position 501) is situated in a CpG island and so surrounded by CpG sites (light blue shading), over which many of the generated primer sequences (grey shadows, beige blocks) overlap. (ii) MTRNRL2L8 probe cg05740793 (position 501), as well as being surrounded by sequences conserved elsewhere in the genome, is flanked by homo-thymine polymers after bisulphite conversion which are refractory to specific primer hybridisation. The downstream flank is shown; orange colour in the sequence indicates paired thymine/adenine produced by bisulphite conversion of paired cytosine/guanine. (iii) Multiple sites adjacent to ANK1 probe cg11823178 (position 501) are eligible for profiling during sequencing extension. As maximum pyrosequencing read length is approximately 40 nucleotides, of the targeted five adjacent CpGs only the first four generated sufficient quality data for analysis.

4.3 Validation of *AIM2* and *FKBP* array-derived methylation values using pyrosequencing

To lend confidence in the results of a genome-wide discovery study, it is important to firstly validate and then replicate the results in a second cohort of samples, ideally using a different technology. Not only does this reduce the influence of the basal false positive rate, but in the latter instance discounts the effects of technical biases particular to the technology used from contributing to this error rate.

Pyrosequencing was originally used to validate raw 450K array data and found to have a Spearman's R of 0.86 and *p* below 0.0001 (Roessler *et al.*, 2012). However, it is possible that the corrective techniques which have emerged since the platform's release may have an effect on complementarity between the two technologies. As a pilot study, 10 sCJD samples and 6 control samples which had been profiled on the array were assayed at cg10636246 (*AIM2*) and cg03546163 (*FKBP5*) using pyrosequencing. Pyrosequencing provides DNA methylation values in percentage rather than β, and data is not processed as with array values. Perhaps as a consequences of this, absolute methylation values were approximately 10-20% lower as measured by pyrosequencing, but differences between means remained consistent between array and pyrosequencing results for *FKBP5* (Δβ = -0.069, Δ% = -6.1%) and *AIM2* (Δβ = -0.095, Δ% = -8.2%), as shown in Figure 36. Spearman's R between pyrosequenced and array values were 0.844 ($p = 2.83 \times 10^{-4}$) and 0.911 ($p = 1.87 \times 10^{-7}$) respectively. This demonstrates that despite the differences between the two platforms, pyrosequencing remains an appropriate method of validating and replicating 450K array data, at least at these two DMPs. As such a second, separate cohort of sCJD and control blood-derived DNA was assembled for replication of DMPs.

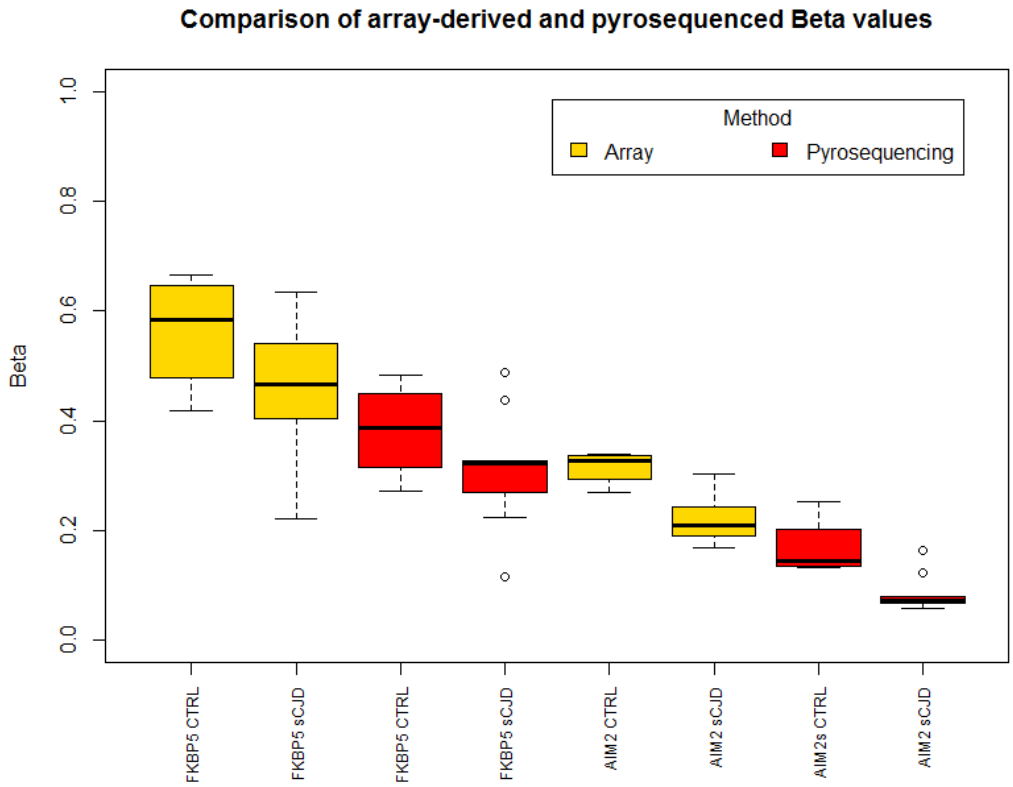


Figure 36: Validation of 450K array results using pyrosequencing. Shown are control and sCJD methylation values (Beta) at FKBP5 and AIM2 DMPs as measured by 450K array (gold) and Pyrosequencing (red).

4.4 Replication of DMPs using pyrosequencing

4.4.1 Power calculations and cohort selection

Now that estimates of effect sizes and beta distributions were known, it became possible to calculate sample numbers (N) needed to achieve 95% power per DMP assay. Using G*Power 3.1 (Faul *et al.*, 2009) N per assay was calculated, which ranged from 7 to 50 as shown in Table 18. In the event of altered distribution of array methylation values and pyrosequencing values per assay or methylation distributions in the new sample group being different, an excess of samples (72 sCJD, 114 controls) were selected and bisulphite DNA stocks generated as previously described. In addition, 59 Alzheimer's disease (AD) patient-derived samples (see methods for sample sources and diagnosis of patients) were converted in order to test the specificity of observed changes to prion disease, rather being a general consequence of neurodegeneration. To test the specificity of these DMPs to sporadic prion disease, 18 iatrogenic CJD (iCJD) and 9 inherited

prion disease (IPD) patient samples (carrying the following *PRNP* mutations: A117V, 5-octapeptide repeat insertion, D178N, P102L (x 6)) were also selected for comparison, with the caveat that not all assays would be powered in these comparisons and that iCJD caused by inoculation with contaminated cadaveric pituitary gland-derived growth hormone disproportionately affected men, thus only 2 of 18 patients were female. Iatrogenic CJD is a very rare disease which also limited sample size.

Probe	Gene	Control β	Control SD	sCJD β	sCJD SD	SD used	sCJD N	Control N	Power
cg10636246	*AIM2*	0.25	0.02	0.21	0.02	0.02	7	7	0.969
cg02481950	*METTL9*	0.82	0.03	0.8	0.02	0.03	50	50	0.95
cg05740793	*MTRNR2L8*	0.16	0.02	0.21	0.025	0.025	7	7	0.969
cg05001044	*MIR1977*	0.19	0.025	0.245	0.035	0.035	10	10	0.958
cg02448796	*KCNAB2*	0.749	0.034	0.781	0.4	0.4	35	35	0.95
cg17714703	*UHRF1*	0.21	0.016	0.242	0.02	0.02	10	10	0.96
cg17515347	*AIM2*	0.6	0.04	0.55	0.045	0.045	19	19	0.956
cg03546163	*FKBP5*	0.55	0.047	0.49	0.05	0.05	16	16	0.95

Table 18: Power calculations for replication assays. Shown are sample numbers (N) required to replicate findings per site at 95% power calculated from Δβ between sCJD and control array values. To ensure stringency the largest of the two standard deviations (SD used) was used in the power calculation.

4.4.2 Replication of DMPs in second control and sCJD cohort

Raw methylation values in percent were output by the PyroMark Q96. Failed calls were discarded manually whilst calls highlighted for manual checking were discarded based on unrealistic values (e.g. 0%) or common discrepancies in the assay. For example, if pyrogram peak height was consistently different from expected height at a position up/downstream of the assay site, this was not considered to be disqualifying. sCJD and control samples were converted in a single batch on two 96-well plates and no assays which contained a non-CpG cytosine as a control for bisulphite conversion efficiency (cg01636246, cg11823178, cg00052684, cg25114611, cg05001044, cg02448796, cg02481950) showed errors related to incomplete bisulphite conversion. As eight genes in total were being assayed, albeit in some instances at several proximal sites, nominal significance threshold was adjusted to $p < 0.00625$ (0.05 ÷ 8) and analysis was performed using a linear regression model with age and sex as covariates.

Of 9 DMPs, 6 were replicated, and the two negative controls remained unaffected by disease/control status.

In *FKBP5*, hypomethylation of cg03546163 was found to be increased from -5.55% in the first cohort to -11.00% in the second, with unadjusted statistical significance being of the order of $p \sim 10^{-8}$ in both groups. In addition, a CpG 4 nucleotides downstream was also found to be significantly hypomethylated (-7.08%, $p = 6.42 \times 10^{-10}$). Of the two additional loci identified using EpiDISH, cg00052684 showed similar effect size between cohorts (-4.95% and -5.35% for the first and second cohorts respectively), but significance, although nominal, was reduced (unadjusted $p = 4.86 \times 10^{-8}$ (first cohort) to $p = 0.002$ (second cohort)). The second EpiDISH CpG (cg25114611) did not pass the threshold for statistical significance of $p < 0.00625$ (-3.31%, $p = 2.45 \times 10^{-10}$ (first cohort), -3.29%, $p = 0.029$ (second cohort)).

Both DMPs in *AIM2* were replicated in the second cohort. Hypomethylation of cg17515347 was also found to be increased from -4.72% to -10.33% and significance was increased by two orders of magnitude ($p = 7.56 \times 10^{-8}$ (first cohort) to $p = 2.02 \times 10^{-10}$ (second cohort)). In addition, a CpG 14 nucleotides downstream was found to be similarly altered (-7.74%, $p = 7.69 \times 10^{-10}$). In contrast, cg10636246 also replicated at increased significance and with a greater effect size (-4.05%, $p = 2.38 \times 10^{-10}$ (first cohort), -5.99%, $p = 5.51 \times 10^{-13}$ (second cohort)), but a CpG 36 nucleotides upstream was not found to be significantly altered between sCJD and control.

The DMP at cg02481950 in *METTL9* was also replicated in the second cohort with similar effect size and significance (+1.93%, $p = 8.65 \times 10^{-10}$ (first cohort), +2.41%, $p = 8.00 \times 10^{-6}$ (second cohort)), but cg02448796 in *KCNAB2* (+3.15%, $p = 4.36 \times 10^{-8}$ (first cohort), +1.96, $p = 0.087$ (second cohort)) and cg05001044 in *MIR1977* (+5.23%, $p = 9.82 \times 10^{-9}$ (first cohort), +0.70, $p = 0.34$ (second cohort)) did not show differential methylation on replication. All three CpGs in the *UHRF1* assay showed hypermethylation (+2.06%, + 4.34%, +2.12%) which was nominally significant ($p = 0.002, 0.001, 0.003$), but reduced compared to the array results corresponding to the second CpG (cg17714703) in the pyrosequencing assay (+3.44%, $p = 2.15 \times 10^{-8}$).

Although both cg04286737 and cg00168514 are located in the body of *PRNP*, the former is 2,107 bp downstream of the latter and so methylation values at these sites cannot be expected to correlate. However, cg04286737 was not found to be differentially methylated between sCJD and control (-0.47%, $p = 0.037$). In *ANK1* lack of differential methylation at cg11823178 was confirmed (-0.46%, $p = 0.819$). Thus, of seven DMPs calculated using Houseman correction five were replicated

and two did not replicate, while of two additional EpiDISH-calculated DMPs one replicated and one did not. Replication results including array data for comparison and values at novel probe-adjacent CpG sites are summarised in Table 19.

CpG site	Coordinates (hg19)	Gene	Array Δβ (%)	Array *p* (Unadj)	Pyro ΔM (%)	Pyro *p* (Unadj)
	6:35,654,360	*FKBP5*			-7.08	$6.42x10^{-10}$
cg03546163	6:35,654,364	*FKBP5*	-5.55	$3.71x10^{-8}$	-11.00	$1.70x10^{-8}$
cg00052684	6:35,694,246	*FKBP5*	-4.95	$4.86x10^{-8}$	-5.35	0.002
cg25114611	6:35,696,871	*FKBP5*	-3.31	$2.45x10^{-10}$	-3.29	0.029
	6:35,696,887	*FKBP5*			-0.68	0.764
cg17515347	1:159,047,164	*AIM2*	-4.72	$7.56x10^{-8}$	-10.33	$2.02x10^{-10}$
	1:159,047,178	*AIM2*			-7.74	$7.69x10^{-10}$
	1:159,046,937	*AIM2*			+1.66	0.349
cg10636246	1:159,046,973	*AIM2*	-4.02	$3.43x10^{-10}$	-5.99	$5.51x10^{-13}$
cg04286737	20:4,665,646	*PRNP*			-0.47	0.037
	20:4,665,649	*PRNP*			-0.21	0.28
	19:4,912,212	*UHRF1*			+2.06	0.002
cg17714703	19:4,912,222	*UHRF1*	+3.44	$2.15x10^{-8}$	+4.34	0.001
	19:4,912,224	*UHRF1*			+2.12	0.003
cg05001044	1:567,312	*MIR1977*	+5.23	$9.82x10^{-9}$	+0.70	0.34
	1:567,347	*MIR1977*			+0.92	0.004
	1:567,357	*MIR1977*			+0.79	0.014
	1:567,375	*MIR1977*			+0.12	0.93
cg02481950	16:21,665,003	*METTL9*	+1.93	$8.65x10^{-10}$	+2.41	$8.0x10^{-6}$
	1:6,101,332	*KCNAB2*			+2.00	0.91
cg02448796	1:6,101,339	*KCNAB2*	+3.15	$4.36x10^{-8}$	+1.96	0.087
cg11823178	8:41,519,399	*ANK1*	-0.29	0.54	-0.46	0.819
	8:41,519,411	*ANK1*			+1.04	0.290
	8:41,519,417	*ANK1*			-4.05	0.023
	8:41,519,420	*ANK1*			-1.75	0.113
	8:41,519,440	*ANK1*			-0.80	0.286

Table 19: Results from DMP replication study. Probes selected for replication are in the first column, where assay permitted measurement of DNA methylation at adjacent CpG sites their genomic coordinates are displayed in column two. For ease of comparison methylation values are given in percent for both array (Δβ x 100) and pyrosequencing data, and p values are unadjusted with the significance threshold being $p < 1.24x10^{-7}$ for array results and $p < 0.00625$ for pyrosequencing results.

I then tested the association between MRC Scale score and hypomethylation at the *AIM2* probe cg10636246 in data from sCJD-derived samples in the replication cohort. Although effect size and significance were diminished compared to the discovery cohort, this effect was replicated in the second cohort (Pearson = 0.281, p = 0.013, compared to Pearson = 0.423, $p = 9.40 \times 10^{-6}$ (first cohort)) as shown in Figure 37.

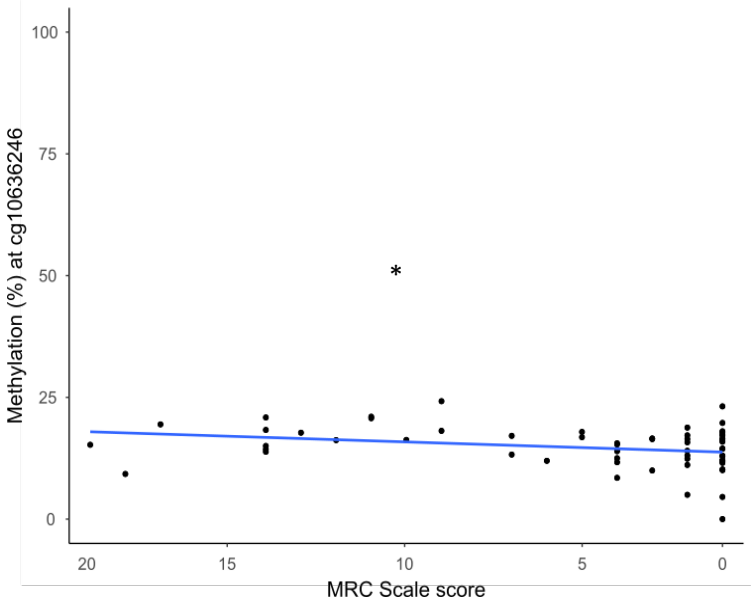


Figure 37: Correlation of hypomethylation at cg10636246 (AIM2) with decrease in MRC Scale score is replicated in a second cohort of sCJD patient samples (r = 0.281, p = 0.013). An MRC Scale score of 20 represents a patient who may have symptoms but is functionally independent for activities of daily life. An MRC Scale score below 3 indicates a comatose state near to death with a patient only able to take perhaps sips of fluid for nutrition and make incomprehensible sounds. Asterisks denote significance (= p < 0.05)*

4.5 DNA methylation at replicated DMPs in iatrogenic CJD, inherited prion disease and Alzheimer's disease

Five replicated assays at *AIM2* (cg17515347, cg10636246), *FKBP5* (cg03546163), *METTL9* (cg02481950) and *UHRF1* (cg17714703) loci were applied to DNA derived from iCJD, IPD and AD patient blood. In summary, changes observed in sCJD were not observed in the other diseases profiled, although the *METTL9* assay was not powered in the case of iCJD and IPD. In this instance, analysis of iCJD data excluded sex as a covariate from the regression model as there was a strong imbalance of sex between iCJD and control groups (as explained in Section 4.3.1). As no models using data from any of the assays showed any strong coefficients

with sex, it is unlikely that this exclusion masked any biological effect. As five assays were used, Bonferroni threshold for disease group specificity assays was set at 0.01. Comparisons of raw pyrosequencing data between groups per assay are displayed in Figures 38-41, while summary statistics are tabulated in Table 20: Statistics for differences between AD, iCJD, IPD compared to control group per replicated DMP. p values are unadjusted, Bonferroni-adjusted significance threshold is $p < 0.01$

.

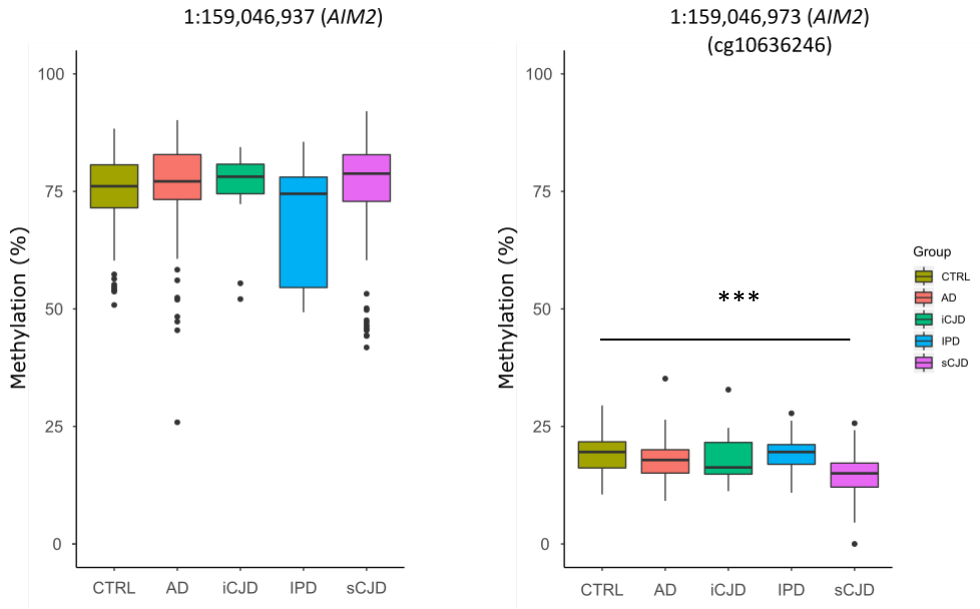


Figure 38: Pyrosequencing results at cg10636246 (AIM2). The assay shows no significant differential methylation at the first interrogated CpG, while at the probe-targeted CpG differential methylation is only observed between control and sCJD groups (ΔM = -5.99%, p = 5.5x10^{-13}). Bonferroni threshold is 0.00625 for sCJD-control comparisons and 0.01 for comparisons between control and non-sCJD diseases. Asterisks denote adjusted significance, where a threshold of 0.05 has been surpassed (= p< 0.05, ** = p< 0.01, *** = p<0.001).*

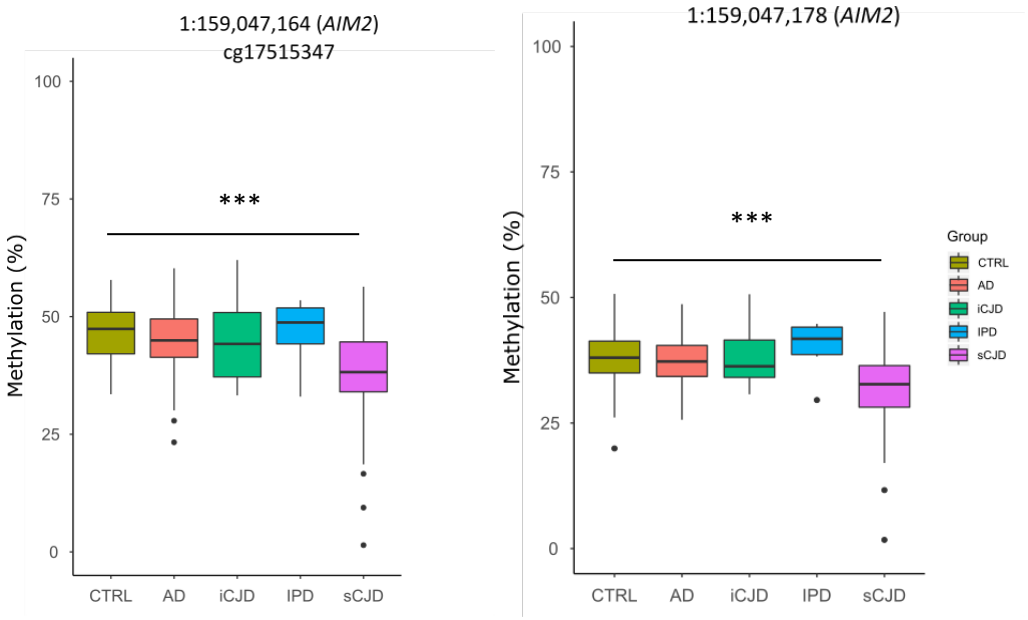


Figure 39: Pyrosequencing results at cg17515347 (AIM2). The assay shows significant hypomethylation at both targeted CpG sites in the sCJD group only (ΔM = -10.33%, p = 2.02x10^{-10}; ΔM = -7.74%, p = 7.69x10^{-10}). Bonferroni threshold is 0.00625 for sCJD-control comparisons and 0.01 for comparisons between control and non-sCJD diseases. Asterisks denote Bonferroni-adjusted significance, where an adjusted threshold of 0.05 has been surpassed (= p< 0.05, ** = p< 0.01, *** = p<0.001).*

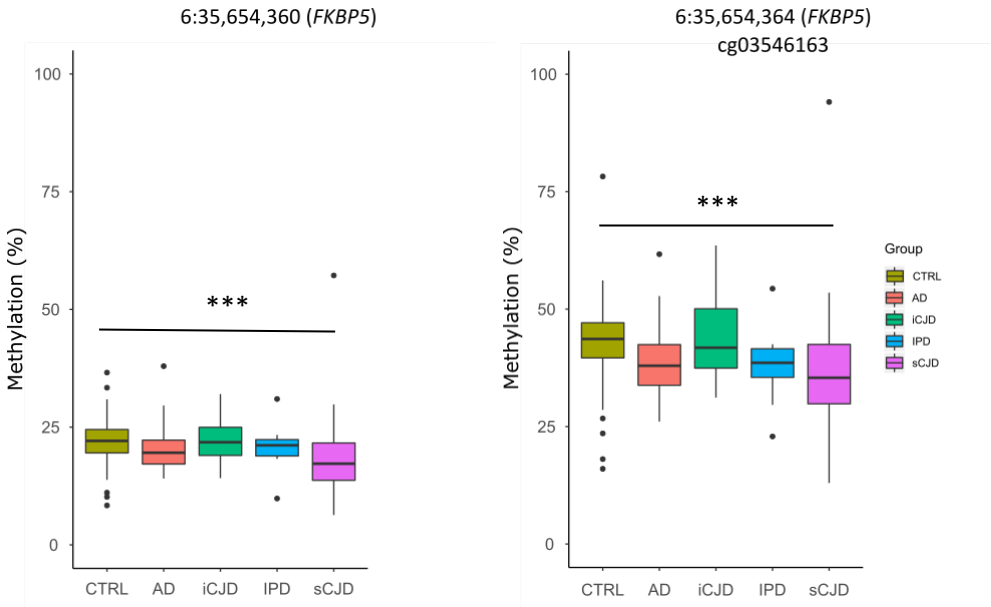


Figure 40: Pyrosequencing results at cg03546163 (FKBP5). The assay shows significant hypomethylation at both targeted CpG sites in the sCJD group only (ΔM = -7.08%, p = 6.42x10^{-10}; -11.00%, p = 1.70x10^{-8}). Bonferroni threshold is 0.00625 for sCJD-control comparisons and 0.01 for comparisons between control and non-sCJD diseases. Asterisks denote Bonferroni-adjusted significance, where an adjusted threshold of 0.05 has been surpassed (= p< 0.05, ** = p< 0.01, *** = p<0.001).*

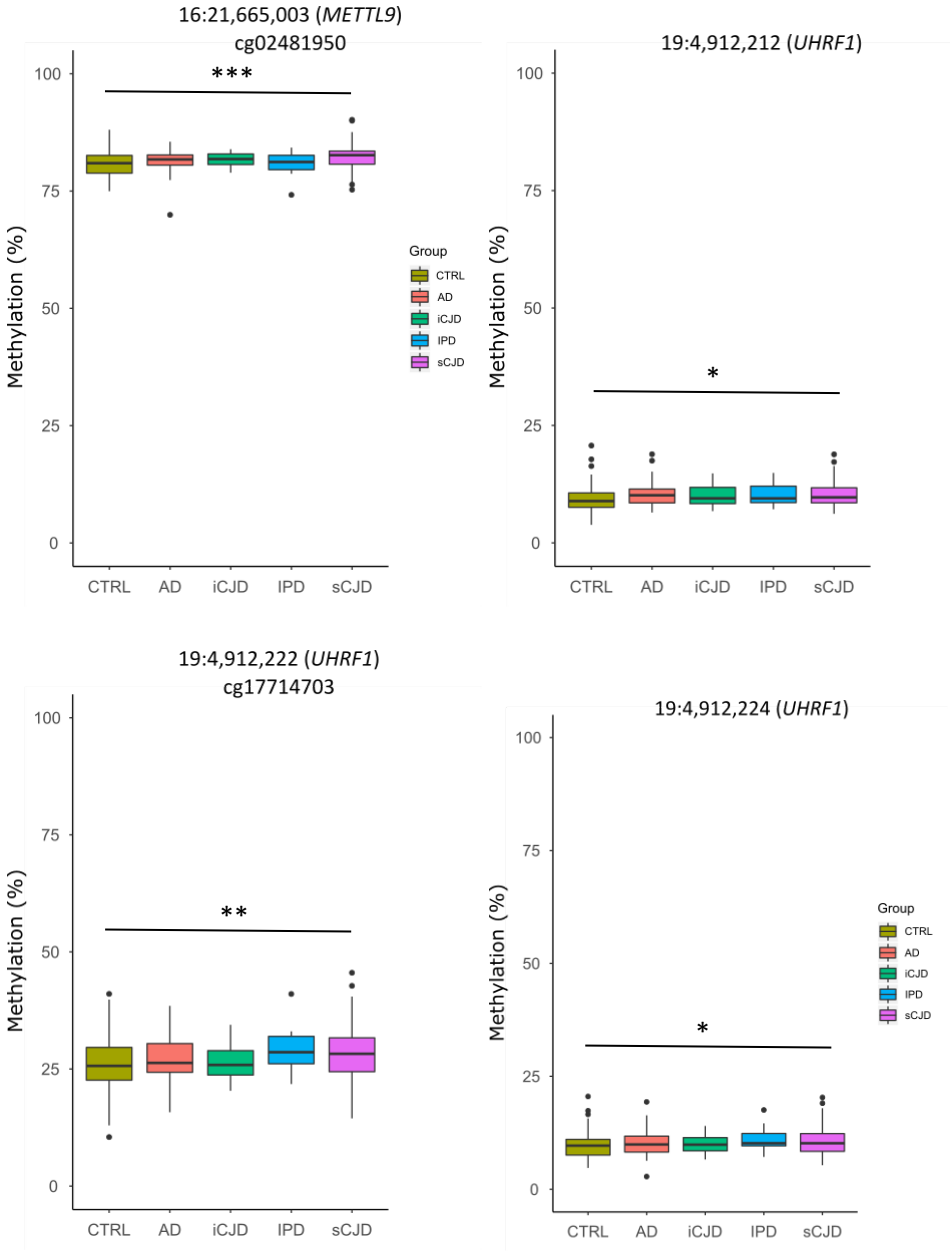


Figure 41: Pyrosequencing results at cg02481950 (METTL9) and cg17714703 (UHRF1). The assay for cg02481950 shows significant hypermethylation in the sCJD group only (ΔM = +2.41%, p = 8.0x10^{-6}), while the assay targeting cg17714703 shows nominally significant but minor hypermethylation in the sCJD group only at all three loci (ΔM = +2.06%, p = 0.002; ΔM = +4.34%, p = 0.001; ΔM = +2.12%, p = 0.003). Bonferroni threshold is 0.00625 for sCJD-control comparisons and 0.01 for comparisons between control and non-sCJD diseases. Asterisks denote Bonferroni-adjusted significance, where an adjusted threshold of 0.05 has been surpassed
(= p< 0.05, ** = p< 0.01, *** = p<0.001).*

			AD		iCJD		IPD	
Gene & Assay	Array Probe	Coordinates	Pyro Δβ (%)	Pyro *p*	Pyro Δβ (%)	Pyro *p*	Pyro Δβ (%)	Pyro *p*
FKBP5(1)		6:35,654,360	-2.4	0.18	0.8	0.76	-1.3	0.52
FKBP5(1)	cg03546163	6:35,654,364	-5.9	0.06	0.8	0.76	1.1	0.83
AIM2 (1)		1:159,046,937	-0.4	0.8	-4.2	0.39	-5.2	0.24
AIM2 (1)	cg10636246	1:159,046,973	-2.4	0.14	-3.1	0.21	0	0.14
AIM2 (2)	cg17515347	1:159,047,164	-2.1	0.05	-1.8	0.28	-7	0.05
AIM2 (2)		1:159,047,178	-0.9	0.28	-0.1	0.94	-3.1	0.32
UHRF1		19:4,912,212	0.9	0.14	-0.5	0.82	-0.2	0.92
UHRF1	cg17714703	19:4,912,222	1.3	0.2	-0.9	0.81	1.5	0.6
UHRF1		19:4,912,224	1.4	0.03	-2.1	0.34	0.1	0.97
METTL9	cg02481950	16:21,665,003	0.6	0.16	0.5	0.75	-0.2	0.9

Table 20: Statistics for differences between AD, iCJD, IPD compared to control group per replicated DMP. p values are unadjusted, Bonferroni-adjusted significance threshold is $p < 0.01$

4.6 Summary

In summary, two thirds of DMPs taken forward for replication successfully replicated, and five novel CpG sites captured in the pyrosequencing assays were also found to be differentially methylated between sCJD patients and control volunteers. Excitingly, association of hypomethylation at the promoter of *AIM2* (cg10636246) with decreased MRC Scale score was also observed in the second cohort of sCJD patients, validating this site's potential as a biomarker of disease severity. The differential methylation observed in both sCJD cohorts was not found to be a feature of Alzheimer's disease, iatrogenic CJD or Inherited Prion Disease. This is significant not just because it suggests the signature observed in sCJD is specific to sporadic prion disease rather than prion disease or neurodegeneration in general, but also because drugs prescribed to treat symptoms of sCJD, IPD and iCJD are broadly similar (anti-epileptics, benzodiazepines), meaning the observed signature is unlikely to be a result of differences in therapeutic intervention between disease groups.

Having confirmed several of the array findings, the logical next step was to consider the functional relevance of implicated genes. Based on *FKBP5* promoter demethylation I decided to investigate serum cortisol levels, while also measuring methylation at replicated sites in brain-derived DNA and further dissecting the confirmed blood-derived methylation signature.

5 Further investigation: Blood-brain concordance, cortisol concentrations and cell-specific effects

As explained in Section 1.8, blood was selected as the tissue of study during this project because of its comparative accessibility in a clinical setting, making it an ideal tissue for biomarker discovery. Moreover, all sCJD patients have blood taken for diagnostics and in many cases consent is given for research samples to be taken too, while postmortem brain tissue is available from fewer patients and is a much more difficult tissue to work with, on account of infectivity risk. Using blood-derived DNA allowed sufficient sample numbers to power a genome-wide study of DNA methylation and now that candidate loci have been identified and replicated it is much easier to study brain-derived DNA in search of similar effects. The logical next step is to profile DNA samples taken from brain tissue (specifically frontal cortex, as this region of brain is known to contain high prion titre and exhibit spongiform degeneration (Hill *et al.*, 2003)) using the pyrosequencing assays designed during the 450K replication stage.

As mentioned in Section 3.3.8, eFORGE revealed that the array DMPs overlapped with functional elements known to be active in monocytes. As monocytes circulate in the periphery but are capable of infiltrating tissues, it would be interesting if the observed changes were amplified in this leukocyte population compared to other circulating cells: not only would this explain the relatively modest effect size seen in whole blood but could also provide more insight into why changes in DNA methylation in the periphery are observed. Frozen blood cannot be sorted into cell fractions as cells are lysed on freezing, so this experiment would require fresh blood samples to be prospectively collected from patients and relatives/spouses during clinic visits. While this presents an issue regarding collecting sufficient samples from patients of a rare disease to power pyrosequencing assays, if the effect size is magnified in a particular cell lineage differences may be detectable at lower sample numbers than calculated previously (Table 18).

But firstly, the discovery of differential methylation of *FKBP5* lead me to consider profiling serum cortisol in sCJD patients compared to controls. Hypomethylation of *FKBP5*'s promoter suggests upregulation, and a known stimulus for this is increased levels of circulating cortisol.

5.1 Relevant Methods and Sample Demographics

Group	Number	Average Age (range)	Sex (% F)	Codon 129 (%) MM:MV:VV	Average MRC Scale Score (range)
sCJD	39	68.9 (49-79)	64.1		
Control	52	55.3 (24-98)	53.8	Unknown	20
FFI	6				

Table 21: Demographics of patients and controls whose samples were used in cortisol screening.

Group	Number	Average Age (range)	Sex (% F)	Codon 129 (%) MM:MV:VV	Average MRC Scale Score (range)
sCJD	58	68.8 (40-87)	42.9	65:25:10	0
Control	33	74.0 (41-89)	54.2	Unknown	0

Table 22: Demographics of patients and controls whose samples were used to investigate brain-derived DNA methylation.

Group	Number	Average Age (range)	Sex (% F)	Codon 129 (%) MM:MV:VV	Average MRC Scale Score (range)
sCJD	7		42.9		
Control	11		36.4	Unknown	

Table 23: Demographics of patients and controls whose samples were used to investigate leukocyte-enriched DNA methylation.

5.2 Comparing circulating cortisol concentrations between sCJD patients and healthy controls

5'-UTR hypomethylation of *FKBP5* is in sCJD is suggestive of gene upregulation, and indeed *FKPB5* transcript levels are found to be increased in sCJD patient blood (currently unpublished, personal communication from Dr Emmanuelle Viré, MRC Prion Unit at UCL). *FKBP5* has many functions, one of which is inhibition of the cellular response to cortisol through binding to the Glucocorticoid Receptor (Wochnik *et al.*, 2005). Hypercortisolaemia is a reported feature of sheep scrapie (Gayrard *et al.*, 2000), prompting us to investigate whether cortisol concentration in sCJD blood is elevated, as a potential causal mechanism for hypomethylation at *FKBP5* as these two observations might suggest. One of the hypotheses underpinning my work concerns the relevance of blood methylation to a brain disease. The opportunity to identify a mechanism of *FKBP5* involvement through the effect of CJD on the hypothalamic-pituitary axis, the endocrine system, and thus peripheral gene expression and methylation was important to explore. Increased *FKBP5* expression and demethylation of its promoter may be a result of a negative feedback response to increased circulating cortisol levels.

As well as 39 sCJD and 52 control sera, serum from 6 patients with Fatal Familial Insomnia (FFI) were profiled as the thalamus is dramatically affected in FFI, thus disruption of the hypothalamic-pituitary-adrenal axis might be expected. 20 of the sCJD patients were part of the cohort which had been profiled on the 450K array as reported in Chapter 3, allowing us to compare *FKBP5* methylation and cortisol concentration in these individuals. Controls were healthy and without cognitive impairment. No sCJD patients had received steroid injections. Blood samples were taken between approximately 10am-4pm. Cortisol was measured by The Doctors' Laboratory (Sonic Healthcare Ltd.) using an electrochemiluminescence immunoassay (Elecsys Cortisol II assay, Roche) and serum concentrations were delivered by email. Cortisol was found to be elevated in sCJD sera within and above the normal circadian range of 133-537 nM. In control sera the mean concentration was 239.8 nM, while in sCJD sera the mean concentration was 387.6 nM (p = 6.6×10^{-5}). FFI mean sera concentration was 271.3 nM (p = 0.434). Interestingly, methylation at cg03546163 (*FKBP5*) did not correlate with [Cortisol] (Pearson's Correlation = -0.047, p = 0.845). Distributions of values per group and [Cortisol] compared with methylation at cg03546163 are shown below in Figure 42 and Figure 43.

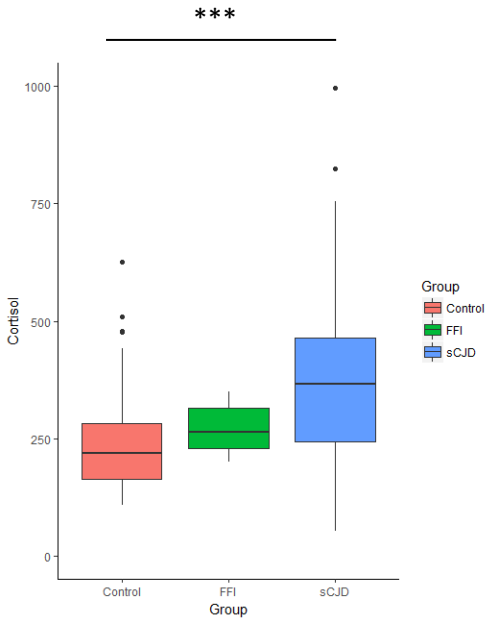


Figure 42: [Cortisol] (nM) in sera extracted from Controls, FFI and sCJD patients. Cortisol was significantly elevated in sCJD patients compared to controls (+147.8 nM, p = $6.6x10^{-5}$). Significance threshold was set at p < 0.05. Asterisks denote significance (= p< 0.05, ** = p< 0.01, *** = p<0.001).*

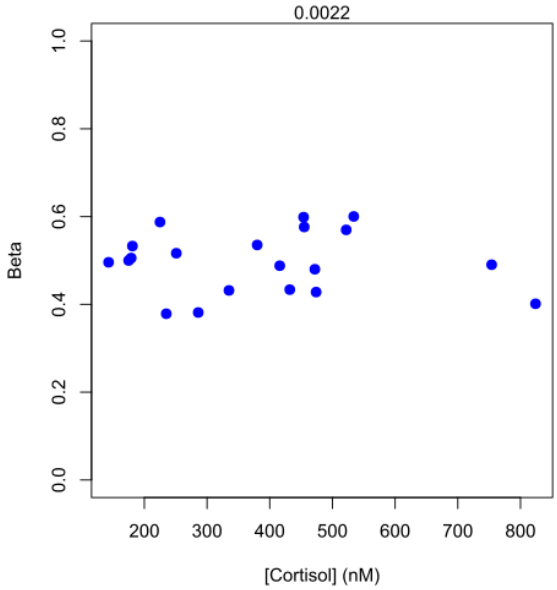


Figure 43: No correlation of methylation at cg03546163 (FKBP5) with [Cortisol] (nM) in 20 sCJD patients profiled on the 450K array, R^2=0.0022.

5.3 Measuring DNA methylation at DMPs in post-mortem brain-derived DNA from sCJD and control donors

In some instances DNA is extracted from post-mortem brain, typically if the patient is too ill to have blood taken or if neuropathology differs from what is typically seen in sporadic CJD. Having described epigenetic changes in blood, I decided to test for the same changes in brain-derived DNA. A concordant signature would suggest that peripheral DNA methylation mirrors cortical DNA methylation, with obvious implications for the suitability of sites other than *AIM2* (at which DNA hypomethylation correlates with increased disease severity) as peripheral biomarkers of sCJD. It would also open the possibility that altered DNA methylation may emerge from risk factors present in the germline that not detected in the sCJD GWAS, rather than be a product of the disease. Conversely, if DMPs are not observed in the brain then perhaps the signature observed is unique to the periphery, which suggests a second, non-central aspect to the disease and would prompt further investigation into the interplay between brain and blood in sCJD.

This experiment comes with a number of caveats: firstly although DNA was taken from the same brain region (frontal cortex, a region susceptible to prion aggregation and deposition (Hill *et al.*, 2003)), and same subregion by the same pathology team (Professor S. Brandner and Dr Z. Jaunmuktane), the tissue remains heterogeneous in terms of neuronal layers, subtypes and gray/white matter. As these samples are to be pyrosequenced at single loci these effects cannot be regressed out as per the original array study of blood as it was not possible to also pyrosequence the Houseman reference probes. Similarly, post mortem examination in CJD can be delayed related to patient locations around the UK. If delay in hours is known it can be adjusted for by including it as a covariate in a linear regression model (Wockner *et al.*, 2014), but research using mouse, pig and human brain tissue has shown that DNA methylation levels remain stable more than 72 hours post mortem (Jarmasz *et al.*, 2019). Conversely, final illnesses that result in death (eg. respiratory failure, sepsis etc.) could have profound influences on methylation profiles, unrelated to prion disease. Nevertheless, as an exploratory study of DNA methylation in 32 control and 54 sCJD brain samples were profiled at replicated DMP sites using pyrosequencing. As seven genes were being assayed I continued to use a strict Bonferroni threshold of $p < 0.007$, albeit with the consideration that this experiment is limited by sample availability and potentially confounded by cellular heterogeneity, and thus trends would also be of interest.

In *FKBP5*, significant hypomethylation of cg03546163 was not observed (-3.06%, *p* = 0.457), nor was hypomethylation of the downstream adjacent CpG (-1.84%, *p* = 0.641). cg10636246 was not found to be differentially methylated in *AIM2* (-3.72%, *p* = 0.439), similarly cg17515347 and a second downstream CpG showed no case-control difference (-1.42%, *p* = 0.711; +0.11%, *p* = 0.976).

None of the three sites at *UHRF1* which showed nominally significant difference in sCJD blood were found to be similarly affected in the brain (-0.08%, *p* = 0.903; +0.54%, *p* = 0.747; 0.28, *p* = 0.699), nor was cg02481950 in *METTL9* (+0.68%, *p* = 0.426). Interestingly the second site in the *PRNP* assay for cg04286737 showed a nominally significant decrease in methylation (-9.38%, *p* = 0.054; -12.01%, *p* = 0.021), although not after correction for multiple testing. These results are displayed below in Figure 44.

ANK1 was also profiled as hypermethylation in the entorhinal and prefrontal cortex is associated with increase in Braak Staging in Alzheimer's Disease (De Jager *et al.*, 2014; Lunnon *et al.*, 2014). The specific CpG, cg11823178, was not found to be differentially methylated (-0.14%, *p* = 0.929), but interestingly of the 4 adjacent CpGs one approached and one passed nominal significance but not after consideration of multiple testing (+1.96%, *p* = 0.056; +0.01%, *p* = 0.994; +3.42%, *p* = 0.04; +0.69%, *p* = 0.652). The effect sizes of these two sites are in the same direction of and of approximately similar magnitude to those observed in the cohorts studied by Lunnon *et al*. (London: +2.05%, Mt Sinai: +2.97%). Distribution of methylation values at these sites is shown in Figure 45.

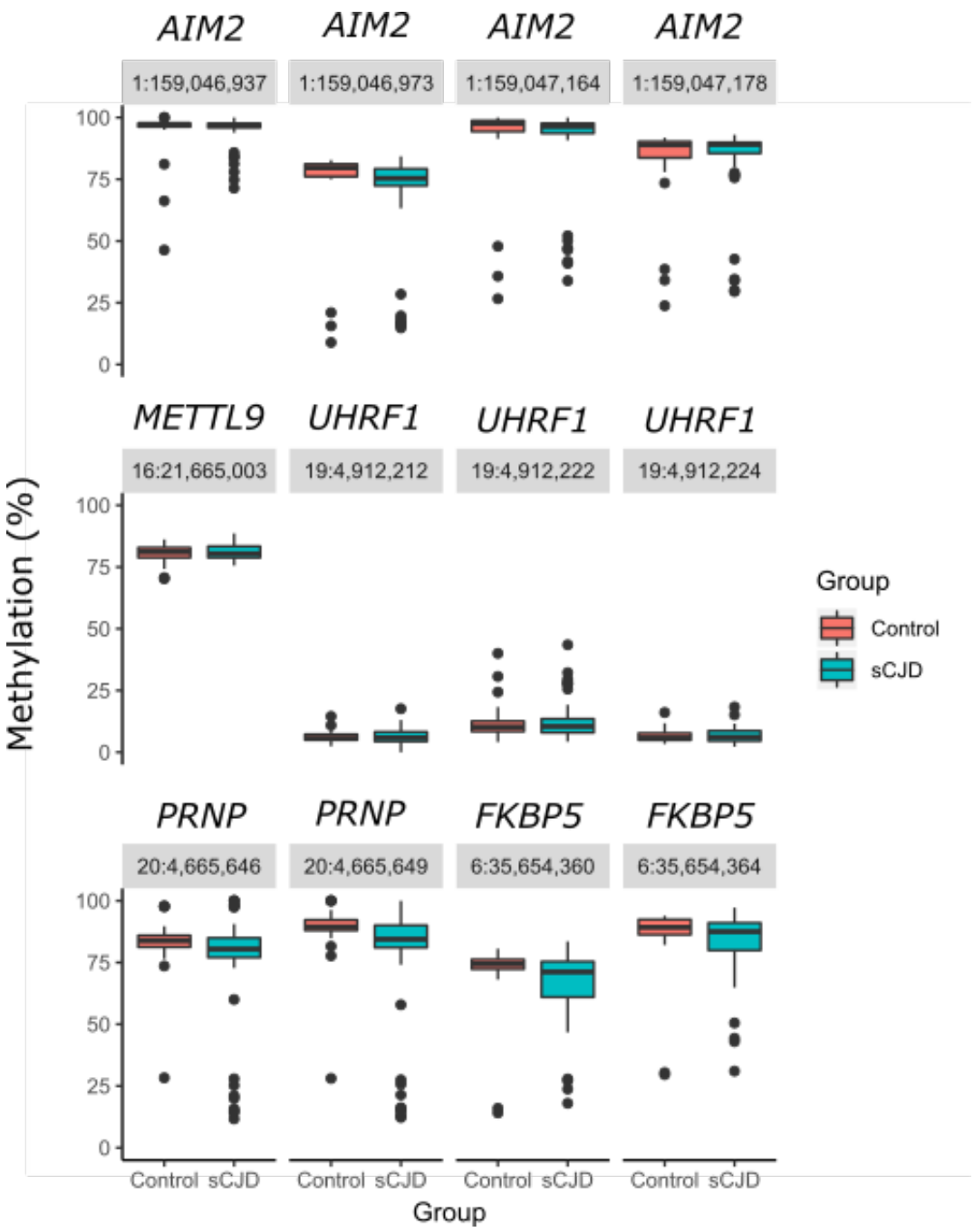


Figure 44: Methylation of control and sCJD brain-derived DNA as profiled in assays for replicated DMPs and PRNP. Bonferroni significance threshold was set at $p < 0.007$ and no sites reached significance, although two sites at PRNP trended towards significance ($\Delta M = -9.38\%$, $p = 0.054$; $\Delta M = -12.01\%$, $p = 0.021$).

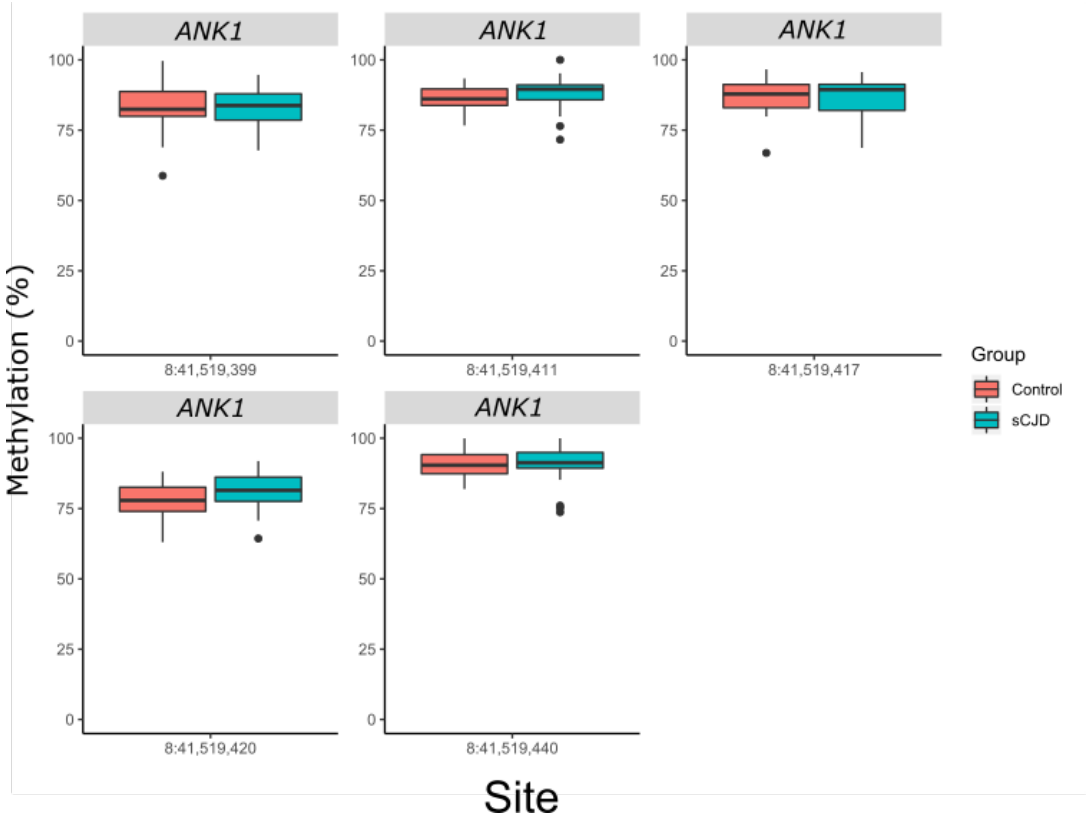


Figure 45: Methylation of control and sCJD brain-derived DNA at cg11823178 (ANK1) and 4 adjacent sites. Bonferroni significance threshold was set at p < 0.007 and no sites reached significance, although two sites at 8:41,519,411 and 8:41,519,420 trended towards significance (ΔM = +1.96%, p = 0.056; ΔM = +3.42%, p = 0.04).

5.4 Determining cell-specific origin of differential methylation in whole blood from sCJD patients

In most contexts blood cells are sorted using Fluorescence Assisted Cell Sorting (FACS), but due to health and safety issues connected to using blood derived from prion disease patients, FACS was not a viable option. Instead I opted to sort cells using Magnet Assisted Cell Sorting (MACS), which uses disposable columns and permits relatively rapid fractionation of leukocytes from fresh whole blood.

5.4.1 MACS protocol optimisation

Blood donated from two individuals was used to optimise MACS separation of B Cells, T Cells, Monocytes and Granulocytes. The first individual's blood was collected into an EDTA vacutainer tubes before being separated after venepuncture and then again after being stored for 1 week at 4ºC, while the second blood sample was collected into two vacutainers, separated after venepuncture, after storage at 4ºC for 3 days, and finally after 1 week of storage. As well as the effects of storage

time on DNA yield, the effects of including the optional filtration and pellet/wash steps were tested for. During the first separation trial (Individual 1, Day 0) due to a procedural error the magnetic columns were not separated from the QuadroMACS separator before fraction elution of the no filter/no wash and filter/no wash samples, meaning cells were retained in the column matrix and were discarded along with the columns. This is reflected in Table 24 and Figure 46, which detail yields of DNA per ml of input blood.

		DNA/1ml input (ng/µl)			
Individual	Day	NFNW	FNW	NFW	FW
1	0	0	0	650	610
1	7	450	650	675	590
1	0	325	242.5	635	840
2	3	292.5	282.5	585	600
2	7	248	482	570	510
AVG		328.875	414.25	623	630

Table 24: Comparing the performance of MACS protocols. Shown are yields of DNA (ng/µl) from 1 ml of whole blood from two individuals across four separation methods (NFNW = No Filter & No Wash; FNW = Filter & No Wash; NFW = No Filter & Wash; FW = Filter & Wash).

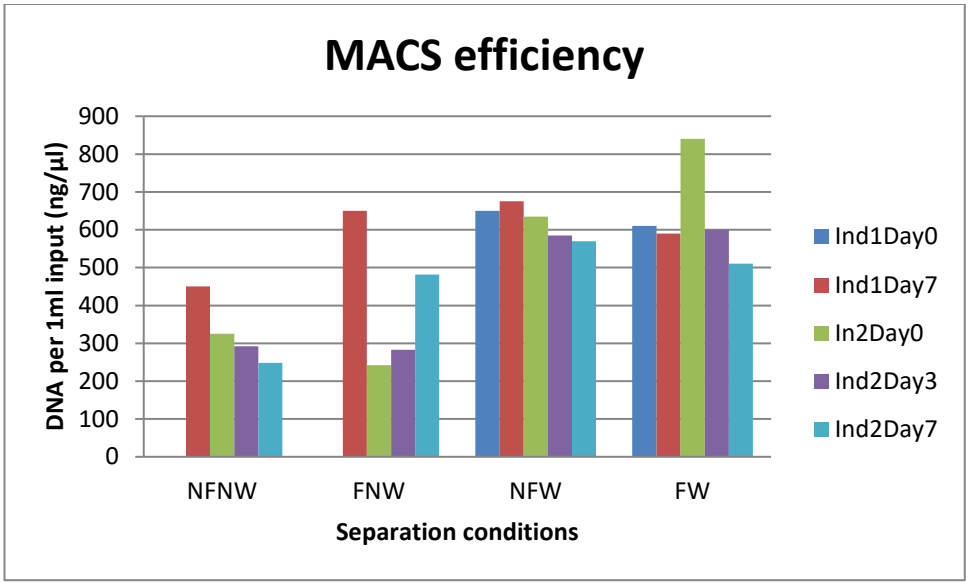


Figure 46: Comparison of MACS method efficiency per method and days between venepuncture and separation. NFNW = No Filter & No Wash; FNW = Filter & No Wash; NFW = No Filter & Wash; FW = Filter & Wash.

The wash steps appeared to have the greatest effect in terms of increased DNA yield, perhaps as resuspension of cells in separation buffer may improve retention of labelled cells in the column matrix. The effect of filtration may be negligible in the samples studied but was included in subsequent separations as filtration is recommended by the manufacturer as it reduces the presence of cell aggregates. As such all subsequent samples were pre-filtered and pelleted and resuspended in separation buffer after microbead annealing.

5.4.2 Pyrosequencing of *AIM2* and *FKBP5* in sorted cell fractions

The average of control sample methylation values per fraction were subtracted from each of six sCJD patient blood samples to produce % change in methylation compared to the control group. These were compared via one-way ANOVA with a post-hoc Dunnett t-test, where whole blood values were designated as controls and expected effect direction was negative, as both targets are hypomethylated in sCJD blood compared to control. The threshold for significance was $p < 0.05$.

The assay targeting cg17515347 (*AIM2*) was applied to separated blood fractions and ANOVA significance was 0.017. The post hoc Dunnett test showed no significant effect in T cells (+5.76%, $p = 0.977$), monocytes (-1.73%, $p = 0.68$) or B cells (+0.97%, $p = 0.845$), but a nominally significant effect was seen in neutrophils (-14.6%, $p = 0.029$). The second downstream CpG captured by the assay also showed nominal ANOVA significance ($p = 0.013$) and, as before, no post-hoc effect in T cells (+6.99%, $p = 0.987$), monocytes (-1.86%, $p = 0.673$) or B cells (+2.84%, $p = 0.916$), while the effect compared to that in whole blood was nominally significant for neutrophils (-14.13%, $p = 0.037$). The differences between means per fraction are displayed in Figure 47.

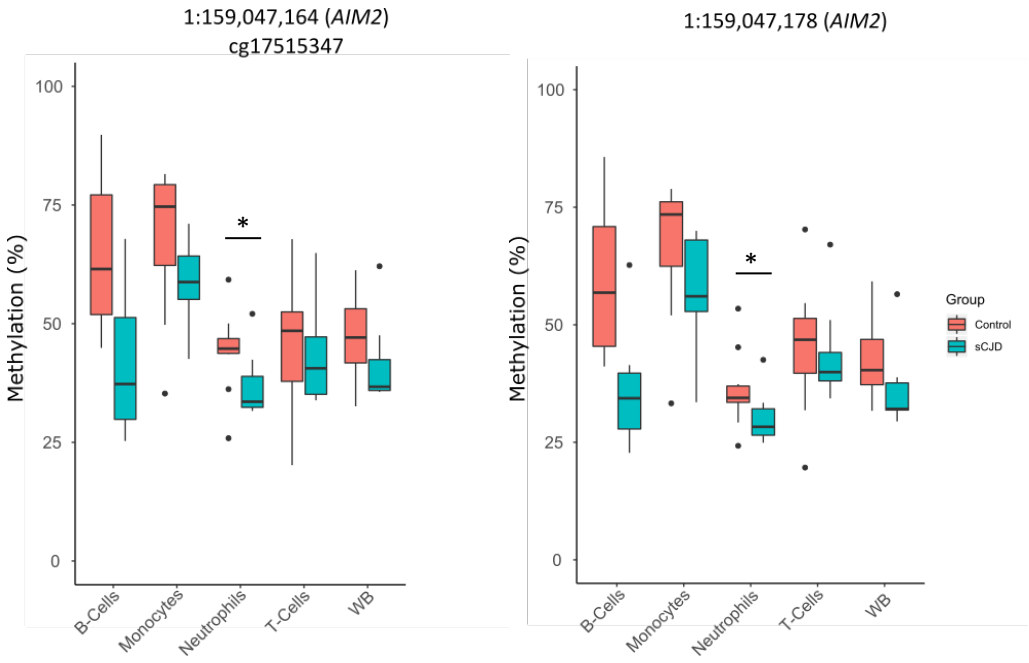


Figure 47: DNA methylation in separated sCJD and control blood fractions at the AIM2 probe cg17515347 and a CpG 14 nucleotides downstream. Changes in Neutrophils compared to Whole Blood (WB) were statistically significant at both sites (ΔM = -14.6%, p = 0.029; ΔM = -14.13%, p = 0.037). Asterisks denote significance (= p< 0.05, ** = p< 0.01, *** = p<0.001).*

Conversely, ANOVA for both positions captured in the assay for *FKBP5*, cg03546163 and a CpG 4 nucleotides upstream, did not pass nominal significance (p = 0.400, p = 0.182). Nevertheless, post-hoc tests were conducted and as expected, effects significantly different to those observed in whole blood were not found at either position in T cells (+2.43%, p = 0.937; +4.72%, p = 0.956), monocytes (-0.94%, p = 0.721; -6.61%, p = 0.376), B cells (-5.52%, p = 0.260; -10.36%, p = 0.173), or neutrophils (+1.08%, p = 0.871; -3.22%, p = 0.616). Differences between means are shown in Figure 48.

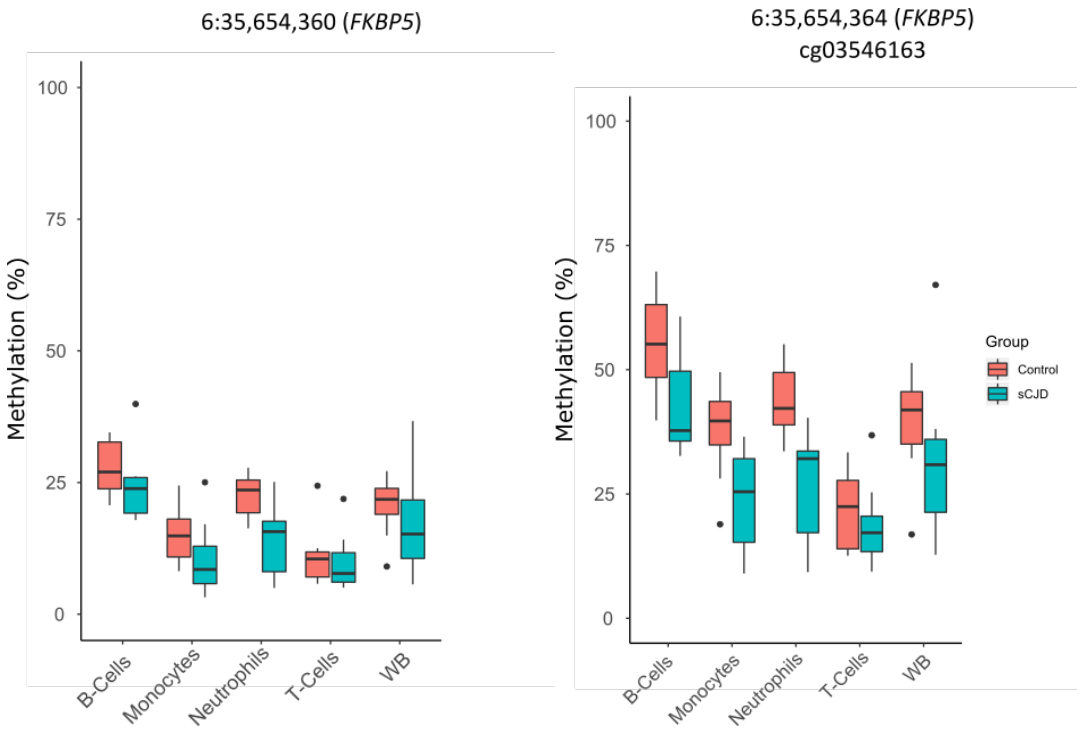


Figure 48: DNA methylation in separated sCJD and control blood fractions at a CpG 14 nucleotides upstream of cg03546163 and at the probe target itself. There was no statistically significant change in any enriched leukocyte fraction compared to Whole Blood (WB)

While not nominally significant, B Cell fraction mean difference is markedly greater than the effect observed in whole blood in the *FKBP5* assay. However, also of note is the potentially decreased separation between sCJD and control values observed in the T Cell fraction in both *FKBP5* and *AIM2* assays. Ultimately, this experiment is extremely limited in terms of N, and further recruitment of patient and control samples may be necessary to power the assays for changes specific to different cell fractions.

5.5 Summary

The observation of elevated serum cortisol (albeit partially within a normal 10am-4pm circadian range) in sCJD lends further weight to the connection between *FKBP5* dysregulation and sCJD. Cortisol is an anti-inflammatory agent and a chemical messenger connecting the brain with the periphery: dysregulation of cortisol metabolism has exciting implications in sCJD which will be discussed later in the thesis.

While the MACS experiment outcome is extremely preliminary, it is surprising that the cell lines inferred to carry more sizable differential methylation by eFORGE were not observed to be markedly affected compared to other cell lines. Sadly limitations

in this experiment make it difficult to draw firm conclusions, but the roles of B cells and granulocytes may be worth considering when trying to frame other data in a physiological context.

Finally, failing to detect differential methylation in sites affected in sCJD blood in brain-derived DNA raises several questions and potentially the exciting prospect of a new dimension to sCJD pathology. Taken in tandem with the suggestion of cortisol dysmetabolism and the connection between *AIM2* hypomethylation and disease severity, connections can be drawn which will be discussed later in the thesis.

Having further investigated the physiology of sCJD epigenetics based on the replication stage of the project, I decided to return once more to the 450K data to see if, as well as be useful in prognosis of sCJD in the case of cg10636246 (*AIM2*), array probes might be useful in diagnosis.

6 Classifying sporadic CJD status using 450K array data and machine learning

6.1 Machine Learning classification

While the linear regression models used in previous chapters are also examples of machine learning, classification is a commonly encountered form of supervised learning. Whether it be social media algorithms trained to recognise faces in photographs, targeted advertising or domestic appliances such as Amazon's Alexa, much effort has been dedicated to developing ways of training models to make decisions based on complex and multifactorial input data. Clinically, this approach has been enormously successful in oncology, where machine learning has been demonstrated to estimate susceptibility, prognosis and recurrence likelihood with high accuracy using a variety of data types across a range of cancer types (Cruz and Wishart, 2006; Exarchos *et al.*, 2014). Using 450K array data it is possible to train a model to classify disease/control status; although aspects of model design must be carefully considered in relation to data type and complexity between groups.

6.1.1 Trees or neurons?

Two popular classifier models are the Random Forest and Neural Network models. Random Forest models operate by building a "forest" of randomly generated decision trees which make classification decisions based on different combinations of input data (Ho, 1995; Breiman, 2001). Unlike a standard decision tree, by using the mode of the forest's classification decisions to make a final classification Random Forests can make decisions in situations where some input data are missing or novel, unseen data is included.

Gradient Boosted Machine learning (GBM) is similar to Random Forest (RF) classification, but where RF classification builds a model from a "forest" of independent decision trees which are considered in parallel, GBM generates its decision trees sequentially, with each tree's design informed by the usefulness of predictor variables in previous trees. These training differences between models are displayed in Figure 49.

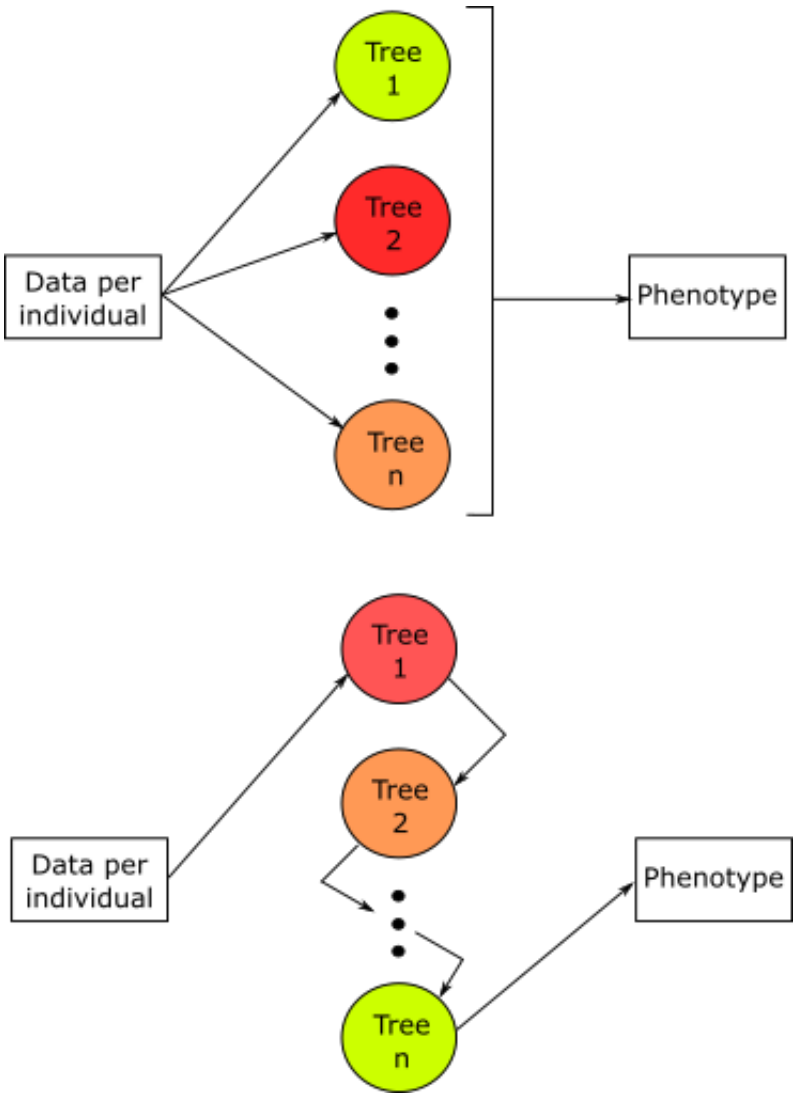


Figure 49: Random Forest (upper) and Gradient Boosted Machine (lower) models. Random Forest models pass data to randomly generated decision trees, their output in parallel is then used to classify phenotype based on a function such as modal averaging. Gradient Boosted Machine models generate decision trees sequentially based on the performance of previous trees, the best performing of which is used to classify phenotype.

Neural Networks are loosely analogous to biological neuronal networks as they are composed of layers of “neurons”. Input data can trigger activation of different combinations of input neurons which connect to output neurons via a user-defined number of hidden layers, between all of which are connections weighted during model training. In this way activation thresholds of each neuron are adjusted so that complex, low variance datasets can be classified. A simplified schematic of a neural network is displayed below in Figure 50.

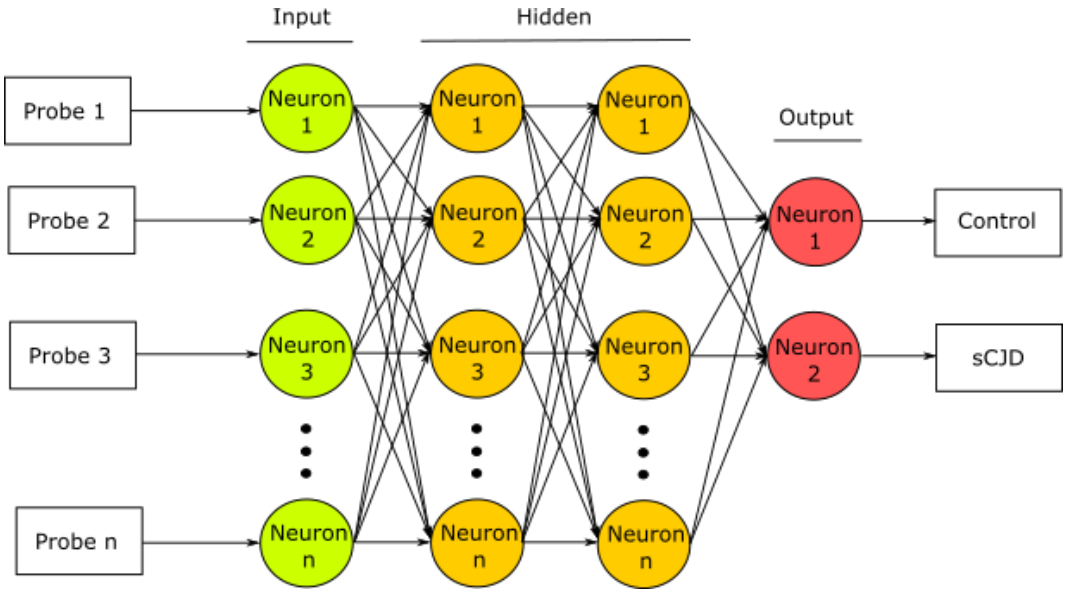


Figure 50: Simplified neural network. Input data (probes) can trigger activation of input neurons, which will then transmit a signal to hidden layers of neurons based on weighted connection thresholds, which are adjusted during model training. Signal from the hidden layers is passed to the output layer and phenotype is classified.

Neural Networks typically require greater computational power and training and prediction runtimes than Random Forest models but are capable of superior classification accuracy when data variance between classes is subtle. However, Neural Networks are more sensitive to bias and prone to overfitting errors, for example when training datasets are limited. In this context, Random Forests are bias-resistant and tend towards superior accuracy (Han *et al.*, 2018).

Differences between sCJD and control in 450K array data are indeed relatively subtle as the greatest significant mean change with phenotype is -5.35%, suggesting a Neural Network approach may be suitable. However, the dataset is limited in terms of sample number and only 38 probes passed genome-wide significance, which in turn suggests it may be appropriate to train a Random Forest model. I therefore decided to test a combination of models and datasets to explore the effects changes in model design would have on accuracy of classification.

6.1.2 Feature Selection

423,742 probes are available in our 450K array dataset, but to train the model using all these data would be both computationally expensive and result in the inclusion of many redundant, insignificant and false positive variables. Random Forests require model features to be defined before model training, which can be achieved in several ways. The most obvious is to only include probes which exhibit genome-

wide significant differences between phenotypes. However, there may be probes which are different between sCJD and control groups but fail to pass significance thresholds due to, for example, the heterogeneity of the disease or the limited number of samples studied. A second approach, therefore, is to include genome-wide insignificant probes based on unadjusted *p* or magnitude of change. As *p* is a proxy indicator of interquartile range separation between groups, selection of probes for inclusion based on unadjusted *p* may be more likely to include probes with diagnostically useful yet sub-genome wide significant variance between sCJD and control.

Good practice, particularly when working with extremely large datasets, is to remove data which correlate. For example, cg10636246 and cg17515347 are both located in *AIM2* and methylation values at both probes are correlated (Pearson's coefficient = 0.737, $p = 8.78 \times 10^{-39}$). Due to this high degree of correlation these probes may be redundant: methylation values at cg10636246 provide as much useful variance between phenotypes as both probes combined. While one additional probe might not cause much of a problem, many redundant probes would reduce the efficiency and increase the runtime of the model, so highly correlating probes should be excluded from the dataset.

When classifying groups based on data with known associations (e.g. classifying sex using age and height data) manual feature selection may be justified. In the absence of prior associations or hypotheses, exactly which probes to include in the model to provide optimum accuracy can be determined by Recursive Feature Elimination (RFE). This process begins by training the model with all features – in this case probes – present, before iteratively removing a user-defined number of features at random and then retraining and testing the model's performance. This continues until an optimum minimum number of probes is reached. To enable a training/testing cycle, the training data is typically split into a user-defined number of "folds". For example, in 10-fold RFE the training data is split into deciles, and each decile is sequentially used to test models built from the remaining 9 folds. As RFE's iterative feature removal process is random, each cycle is repeated several times (the number of which is again defined by the user), which serves to reduce bias. In summary, the number of probes to remove per iteration, the number of permutations of training/testing data to use, and the number of times RFE should be repeated per permutation can all be defined by the user, allowing for stringent yet computationally expensive feature selection or less granular and computationally faster feature

selection. The fold and repeat aspects of feature selection are displayed in Figure 51.

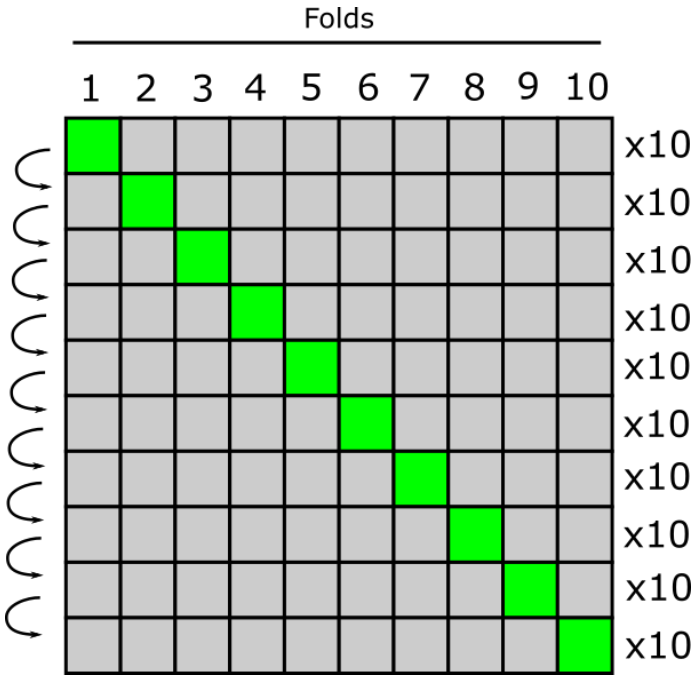


Figure 51: Model of 10-fold 10-repeat crossvalidation. Crossvalidation of model training and tuning involves the division of individuals in the training set into 10 groups, or "folds". 9 folds (grey) are used to train a model which is tested on the remaining fold (green). This random training process is repeated 10 times, before the second fold is used for testing models built on the remaining 9 folds, which is again repeated 10 times. This continues until all folds have been used for model testing.

To test the effect of different feature selection approaches, features were selected from the following combinations of input data:

1. 38 genome-wide significant DMPs
2. Features selected from 500 probes ranked by P value after removal of correlated probes and 10-fold 10-repeat RFE removing 1 probe per iteration
3. Features selected from 5000 probes ranked by P value without further feature selection

An important note is that these processes rely on random number generators (RNGs) which, while themselves are random, can be "seeded" to provide the same series of random numbers in future experiments for the sake of repeatability and reproducibility. The RNG seed for all models was set as 4444.

6.2 Relevant Methods and Sample Demographics

2.1.1 Sample Selection (page 54)

2.2.2 Case-Control Study Using ChAMP (page 61)

2.2.8 Machine Learning Classification of 450K Dataset (page 64)

2.7 Statistics and Graphics (page 72)

Group	Number	Average Age (range)	Sex (% F)	Codon 129 (%) MM:MV:VV	Average MRC Scale Score (range)
sCJD	114	68.0 (49-85)	50.9	46:23:32	6 (0-20)
Control	105	69.0 (41-83)	55.7	Unknown	20

Table 25: Demographics of patients and controls whose 450K Beadchip Array data were used to train disease classification models.

6.3 Gradient Boosted Machine classification

6.3.1 Gradient Boosted Machine classification using 38 genome-wide significant DMPs

Beta values from the genome-wide significant DMPs per individual, age and sex were used as predictor variables. Individuals were partitioned into a training set (75%) of 165 and a testing set (25%) of 54, and the first set was used to train a GBM model with 10-fold 10-repeat cross-validation. 50,100 and 150 trees (boosting iterations) were constructed with between 1 and 3 decision nodes (tree depth). Model accuracy across the training process did not vary appreciably across training parameters, as graphically represented in Figure 52.

A confusion matrix, principle component analysis (PCA) and model report were generated: model accuracy as measured by Receiver Operating Characteristic (ROC) was found to be 81.48% versus a No Information Rate of 51.85% (p = 5.869×10^{-6}). Sensitivity was 76.92% and specificity was 85.71%. PCA found that 28 variables could explain 95% of total variance. The confusion matrix, which indicates true (reference) versus model (prediction) classification, is displayed below in Table 26 where true positive and negative predictions are shown in lower right and upper left cells, while false positive and false negative predictions are shown in lower left and upper right cells, respectively. Positive Predictive Value (PPV) and Negative Predictive Value (NPV) was 85.7% and 83.3% respectively, calculated by dividing

true positive and true negative predictions by total positive and negative predictions respectively.

	Reference	
Prediction	Control	sCJD
Control	20	4
sCJD	6	24

Table 26: Confusion matrix of untuned gbm model classification of sCJD status using 38 DMPs, age and sex as predictor variables.

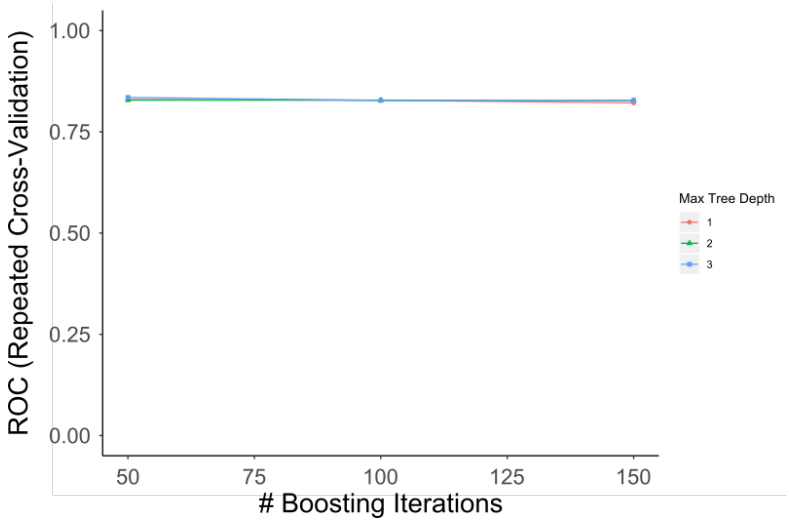


Figure 52: Accuracy of gbm model trained using DMPs as measured by ROC. Shown are values across a number of boosting iterations (number of trees sequentially generated) and tree depth, where tree depth is the number of nodes between the root node and decision nodes.

Simple models such as this can be improved through tuning. This is achieved by adjusting model parameters to improve the fit and can be technically challenging and lead to model overfitting. In this instance the two major parameters as shown above in Figure 52 are number of boosting iterations and tree depth. A more complex model might perform better or worse, depending on the complexity of the data, so I decided to increase iterations to 250 in incrementations of 50 iterations, and increase tree depth to 5. Once again, the model was trained using 10-repeat

10-fold cross-validation and a depth of 1 with 50 boosting iterations were selected. Interestingly the tuned model had decreased accuracy of 74.07% (p = 7.006x10^{-4}), and although sensitivity was improved (80.77%), specificity had decreased (67.86%). The confusion matrix and model accuracy across tuning are displayed below in Table 27 and Figure 53 respectively. Positive Predictive Value (PPV) and Negative Predictive Value (NPV) was 79.6% and 80.7% respectively, calculated by dividing true positive and true negative predictions by total positive and negative predictions respectively.

	Reference	
Prediction	Control	sCJD
Control	21	9
sCJD	5	19

Table 27: Confusion matrix of tuned GBM model classification of sCJD status using 38 DMPs, age and sex as predictor variables.

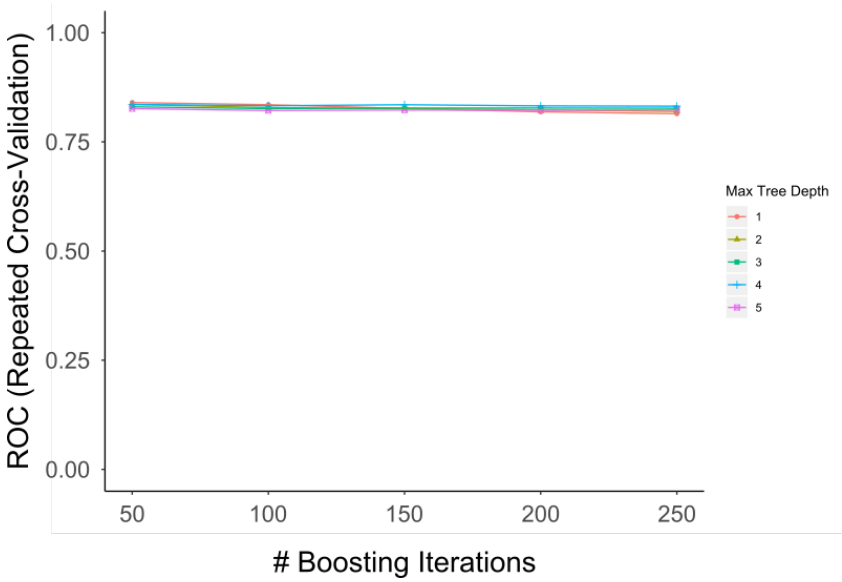


Figure 53: Accuracy of gbm model trained using DMPs after tuning as measured by ROC. Shown are values across a number of boosting iterations (number of trees sequentially generated) and tree depth, where tree depth is the number of nodes between the root node and decision nodes.

This model was trained using manual feature selection (38 DMPs). Next, I decided to select features from the top 500 most significantly ranked probes using RFE.

6.3.2 Gradient Boosted Machine classification using 33 probes identified by Recursive Feature Elimination

In the top 500 probes, 28 probes were found to have Pearson's correlations exceeding 0.75 and were removed. The remaining 472 probes were processed using RFE with 10-repeat 10-fold cross-validation. As elimination iterations removed one probe per turn this was a computationally demanding process which took two days to complete, after which 33 probes were selected for training (displayed below in Table 28).

Using the same training/testing partitions as before, a gbm model was trained using 10-repeat 10-fold cross-validation. Again, model accuracy across the training process and final variable weights were extracted and are graphically represented in Figure 54.

A confusion matrix, principle component analysis (PCA) and model report were generated: on classification of the testing set model accuracy was found to be 79.63% versus a No Information Rate of 51.85% ($p = 2.263 \times 10^{-5}$). Sensitivity was 84.62% and specificity was 75.00%. PCA found that 29 variables could explain 95% of total variance. The confusion matrix, which indicates true/false positive/negative classification, is displayed below in Table 29. Positive Predictive Value (PPV) and Negative Predictive Value (NPV) was 75% and 84.6% respectively, calculated by dividing true positive and true negative predictions by total positive and negative predictions respectively.

Probe	Gene	Location	Feature
cg05948372	*ABHD3*	Body	N_Shore
cg20003976	*ACADM*	TSS1500	N_Shore
cg15691252	*ADK;AP3M1*	TSS1500;5'UTR	N_Shore
cg10636246	*AIM2*	TSS1500	
cg17515347	*AIM2*	TSS1500	
cg02243276	*AZIN1*	5'UTR	N_Shelf
cg04057956	*CD9*	Body	
cg13599258	*CUX2*	Body	
cg04137490	*DOCK5*	Body	
cg00052684	*FKBP5*	5'UTR	N_Shore
cg03546163	*FKBP5*	5'UTR	N_Shore
cg25114611	*FKBP5;LOC285847*	TSS1500;Body	S_Shore
cg18023065	*FUT4*	1stExon	Island
cg19382697	*GNG7*	5'UTR	S_Shore
cg14195992	*KIAA0146*	Body	
cg05336268	*KIAA0556*	Body	
cg13696490	*LOC201651*	TSS1500	
cg05218245	*LY6G6D*	Body	
cg02481950	*METTL9;IGSF6*	Body;TSS1500	
cg20290360	*NFASC*	Body	
cg03891318	*NMNAT3*	5'UTR	
cg17163138	*PIN1*	Body	S_Shelf
cg03071209	*THADA*	TSS200	Island
cg17714703	*UHRF1*	Body	S_Shore
cg21643513	*ZNF547;TRAPPC2P1*	TSS1500;TSS1500	N_Shore
cg05740793	*MTRNR2L8*	TSS1500	
cg04590451			S_Shore
cg11227822			Island
cg12992827			
cg09964921			
cg10972973			N_Shore
cg25771026			
cg14427590			

Table 28: Probes identified as useful predictors of sCJD status through Recursive Feature Elimination.

	Reference	
Prediction	Control	sCJD
Control	22	7
sCJD	4	21

Table 29: Confusion matrix of untuned gbm model classification of sCJD status using 33 probes selected from the 500 most significantly altered between sCJD and control, age and sex as predictor variables.

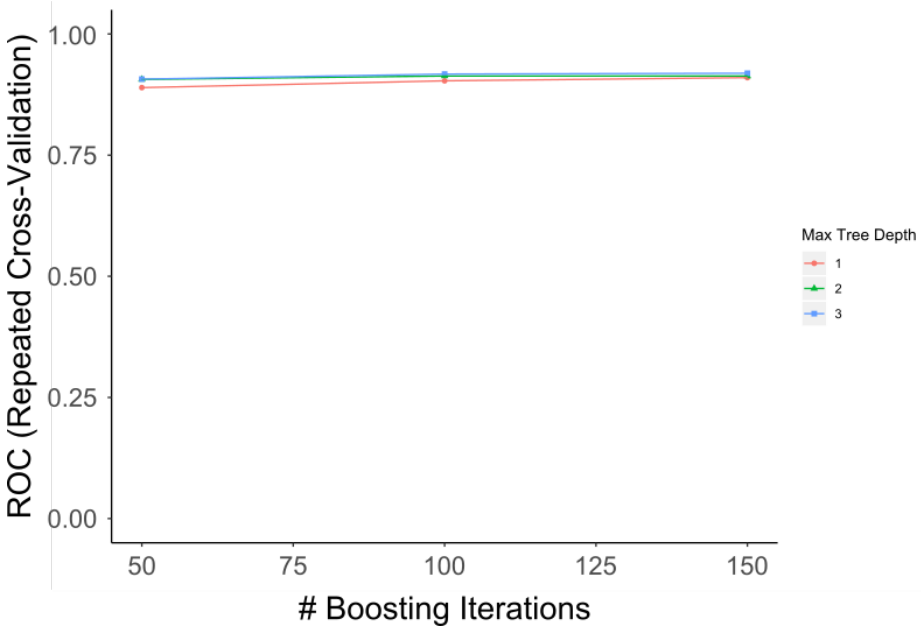


Figure 54: Accuracy of gbm model trained using 33 RFE-identified probes as measured by ROC. Shown are values across a number of boosting iterations (number of trees sequentially generated) and tree depth, where tree depth is the number of nodes between the root node and decision nodes.

Interestingly this model performs slightly worse overall than one trained using the DMPs. To test whether the performance could be improved by training a more complex model, I performed the same tuning process as before of increasing boosting iterations to 250 in incrementations of 50 iterations and increasing tree depth to 5. Once again, the model was trained using 10-repeat 10-fold cross-validation and a depth of 5 with 150 boosting iterations were selected. After tuning model accuracy was found to be 87.04% versus a No Information Rate of 51.85% ($p = 4.922 \times 10^{-8}$), while sensitivity was 88.46% and specificity was 85.71%. The confusion matrix is displayed below in Table 30, and model accuracy over the tuning process are displayed in Figure 55. Positive Predictive Value (PPV) and

Negative Predictive Value (NPV) was 85.7% and 88.5% respectively, calculated by dividing true positive and true negative predictions by total positive and negative predictions respectively.

	Reference	
Prediction	Control	sCJD
Control	23	4
sCJD	3	24

Table 30: Confusion matrix of tuned gbm model classification of sCJD status using 33 probes selected from the 500 most significantly altered between sCJD and control, age and sex as predictor variables.

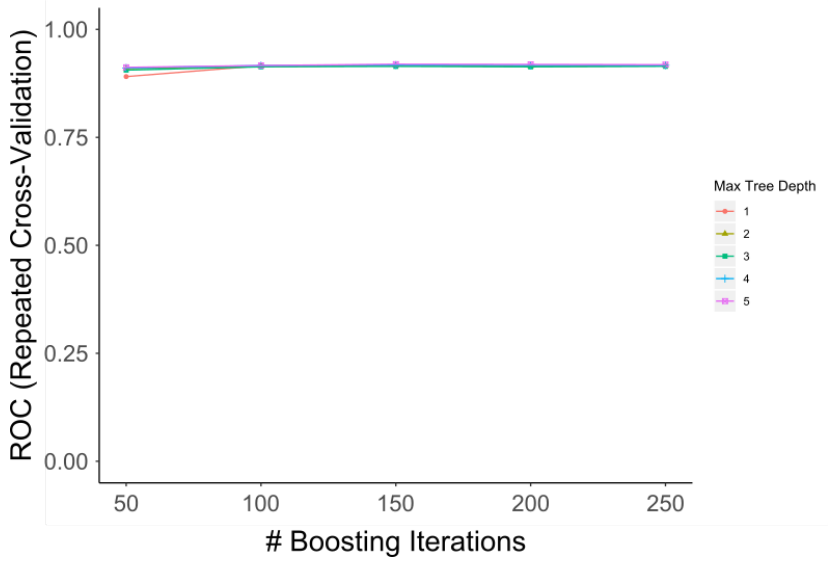


Figure 55: Accuracy of tuned gbm model trained using 33 RFE-identified probes as measured by ROC. Shown are values across a number of boosting iterations (number of trees sequentially generated) and tree depth, where tree depth is the number of nodes between the root node and decision nodes.

Encouragingly accuracy after tuning was superior to any of the previous three models. However, 33 and 38 probes are very small fractions of the total number of array probes, and it is possible that large numbers of small effects at less significantly ranked probes may still prove useful predictors. Thus, the final dataset - the top 5,000 probes ranked by significance - was run through the same pipeline.

6.3.3 Gradient Boosted Machine classification using 5,000 probes as ranked by genome-wide significance

The top 5000 most differentially methylated probes, ranked by significance, were used to train a third model. This model performed poorly with an accuracy of 65.38% (p = 0.018), a sensitivity of 80.77% and a specificity of 50.00%. PCA required only 127 variables to explain 95% of the variance. The confusion matrix of classification efficacy is presented below in Table 31 and model accuracies over the training process are displayed in Figure 56. Positive Predictive Value (PPV) and Negative Predictive Value (NPV) was 50% and 80.8% respectively, calculated by dividing true positive and true negative predictions by total positive and negative predictions respectively.

	Reference	
Prediction	Control	sCJD
Control	21	13
sCJD	5	13

Table 31: Confusion matrix of GBM model classification of sCJD status using 5000 probes most significantly altered between sCJD and control, age and sex as predictor variables.

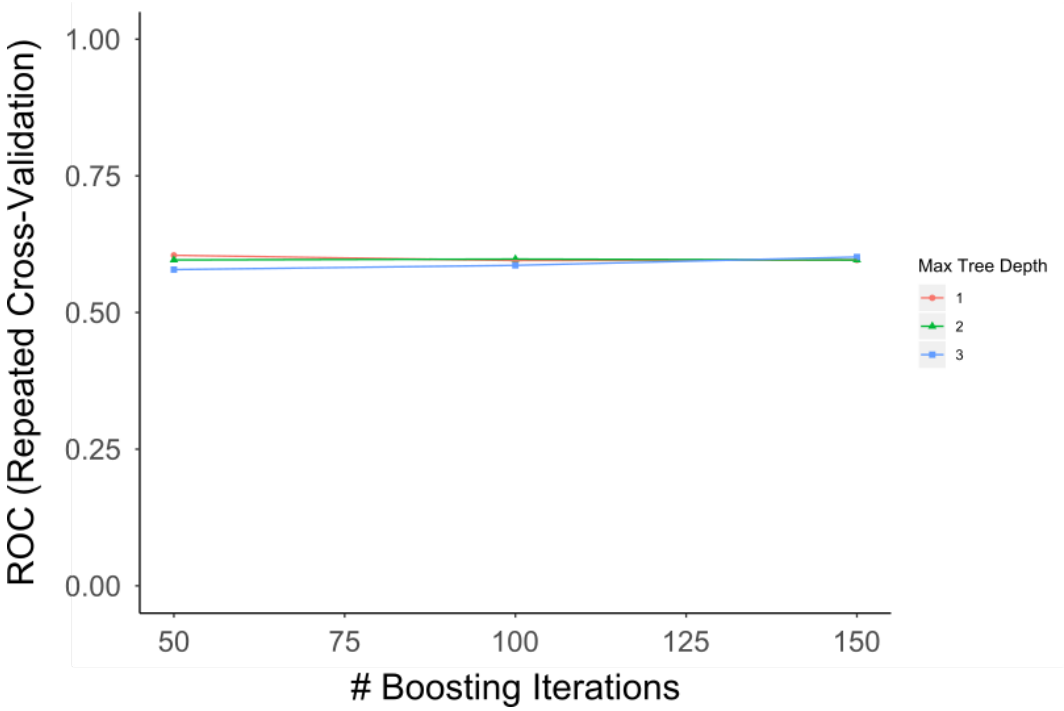


Figure 56: Accuracy of model trained using the 5000 most differentially methylated probes (as ranked by significance) as measured by ROC. Values are shown across a number of boosting iterations (number of trees sequentially generated) and tree depth, where tree depth is the number of nodes between the root node and decision nodes.

I attempted to tune the model in the same manner as the previous two, but after 48 hours the process was still running so, given the model's poor performance at the outset, tuning was abandoned.

6.4 Averaged Neural Network classification

6.4.1 Averaged Neural Network classification using 38 genome-wide significant DMPs

An averaged neural network was trained using this dataset with tuning parameters set as the number of hidden neuronal layers and weight decay value. Weight decay is a number approaching 1 serving as a function by which weights are adjusted after each training iteration to prevent them from growing too large. Tuning results set optimum parameters as 8 layers and a weight decay value of 0; accuracy across tuning and training is displayed in Figure 57.

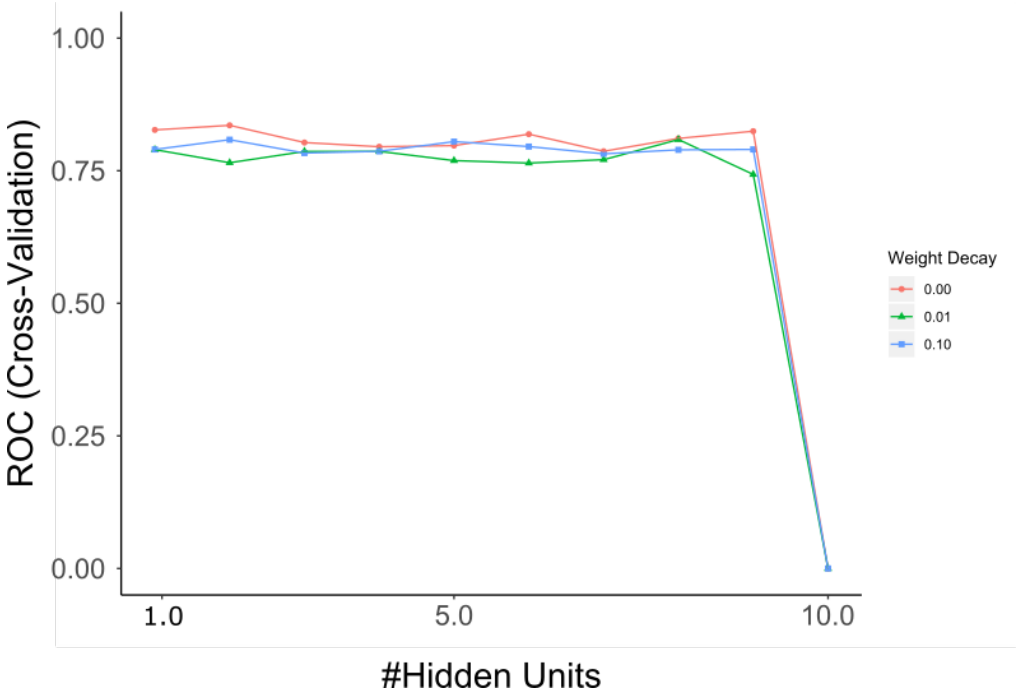


Figure 57: Accuracy of tuned avNNet model trained using 38 DMPs as measured by ROC. Values are shown across a number of hidden neuronal layers and weight decay. At 10 hidden units the model fails, as this number of hidden neuronal layers cannot be supported by only 38 variables.

The model was then used to classify the testing set and a confusion matrix was generated, displayed below in Table 32. The model's accuracy was 77.78% versus a No Information Rate of 51.85% (p = 7.839x10^{-5}), with sensitivity of 76.92% and specificity of 78.57%. Positive Predictive Value (PPV) and Negative Predictive Value (NPV) was 78.5% and 76.9% respectively, calculated by dividing true positive and true negative predictions by total positive and negative predictions respectively.

	Reference	
Prediction	Control	sCJD
Control	20	6
sCJD	6	22

Table 32: Confusion matrix of tuned averaged Neural Network model classification of sCJD status using 38 DMPs

6.4.2 Averaged Neural Network classification using 33 probes identified by Recursive Feature Elimination

Using the dataset of 33 probes selected by RFE, a model was tuned in an identical way. Here parameters were again optimised to 8 layers and a weight decay value of 0, but the model's accuracy was increased to 81.48% (p = 5.869x10^{-6}), sensitivity increased to 84.62% while specificity remained the same as before at 75.87%. The confusion matrix and model accuracies over the training process are displayed below in Figure 58 and Table 33 respectively. Positive Predictive Value (PPV) and Negative Predictive Value (NPV) were 78.6 and 84.6% respectively, calculated by dividing true positive and true negative predictions by total positive and negative predictions respectively.

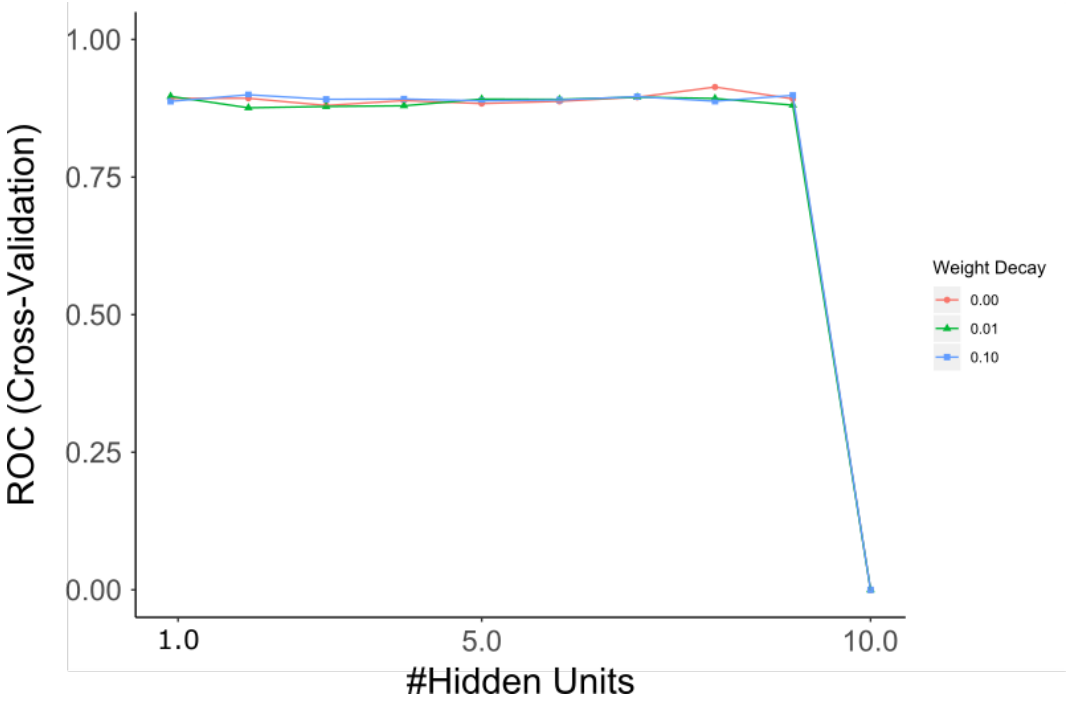


Figure 58: Accuracy of tuned avNNet model trained using 33 RFE-identified probes as measured by ROC. Values are shown across a number of hidden neuronal layers and weight decay. At 10 hidden units the model fails, as this number of hidden neuronal layers cannot be supported by only 33 variables.

	Reference	
Prediction	Control	sCJD
Control	22	6
sCJD	4	22

Table 33: Confusion matrix of tuned avNNet model classification of sCJD status trained using 33 probes identified using RFE from the top 500 probes ranked by significance.

6.4.3 Averaged Neural Network classification using top 5000 probes as ranked by significance

I attempted to tune a neural network model using the third dataset, but as with the GBM model tuning runtime exceeded 48 hours, and so this attempt was abandoned.

6.5 Putting it all together: Stacking models using generalised linear regression

6.5.1 General Linear Model using 33 probes identified by Recursive Feature Elimination

As the previous sections illustrate, model performance (and thus utility) depend not only on the data used to train them, but also the type of model used, and the parameters used to tune it. In this final section the performance of five different models will be compared and then unified using a generalised linear model (GLM) to provide a meta-model, which in principle should perform at least as well as any of the models produced so far.

The RFE-identified dataset of 33 probes, having produced the best model so far, was used to train models using five models: a random forest model (ranger), a gradient boosted model (gbm), a multivariate adaptive regression splines model (earth), a support vector machine model (svmRadial), and a neural network model (avNNet). Once again, model training was executed using 10-fold 10-repeat cross-validation, and performance metrics were exported. In terms of mean accuracy

ranger performed the best (94.75%). Gbm was the most sensitive (83.61%) while the most specific model was ranger (89.64%). Accuracy, sensitivity and specificity of these models are displayed in Figure 59.

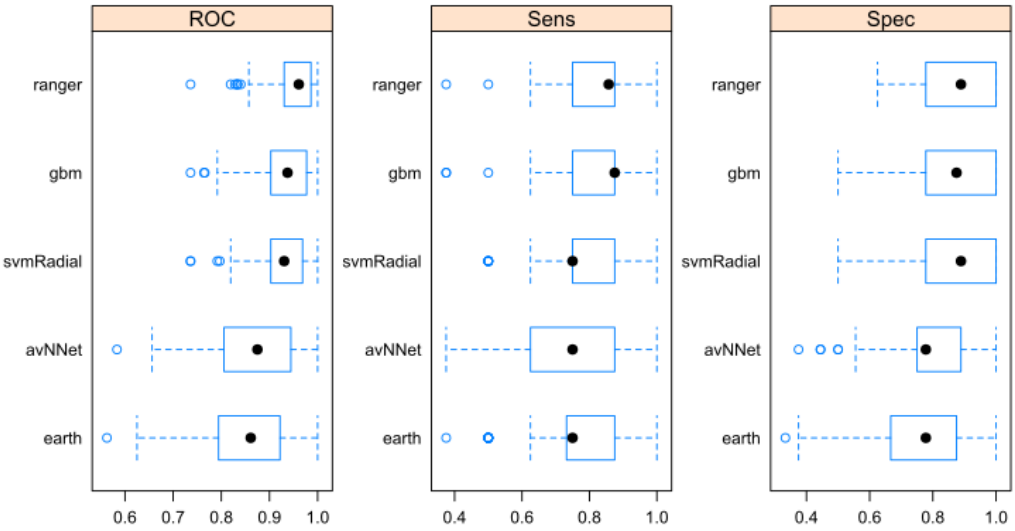


Figure 59: Accuracy (ROC), sensitivity and specificity of models trained using 33 RFE-selected probes.

A generalised linear model was then constructed using these models and used to classify the test dataset, again using 10-fold 10-repeat cross-validation. Accuracy was 87.04% (p = 7.839x10^{-5}), sensitivity was 82.14% and specificity was 92.31%. The confusion matrix is presented below as Table 34. Positive Predictive Value (PPV) and Negative Predictive Value (NPV) was 82.1% and 92.3% respectively, calculated by dividing true positive and true negative predictions by total positive and negative predictions respectively.

	Reference	
Prediction	Control	sCJD
Control	24	5
sCJD	2	23

Table 34: Confusion matrix of generalised linear model classification of sCJD status using five models trained with 33 probes selected using RFE from the top 500 probes ranked by significance

6.5.2 General Linear Model using 38 DMPs

For the sake of comparison, I trained a second series of models in the same way using the 38 DMPs dataset. Once again, ranger provided best mean accuracy (86.00%), gbm was the most sensitive (74.30%) and ranger was most specific (81.00%). Model accuracies during the training process are displayed in Figure 60.

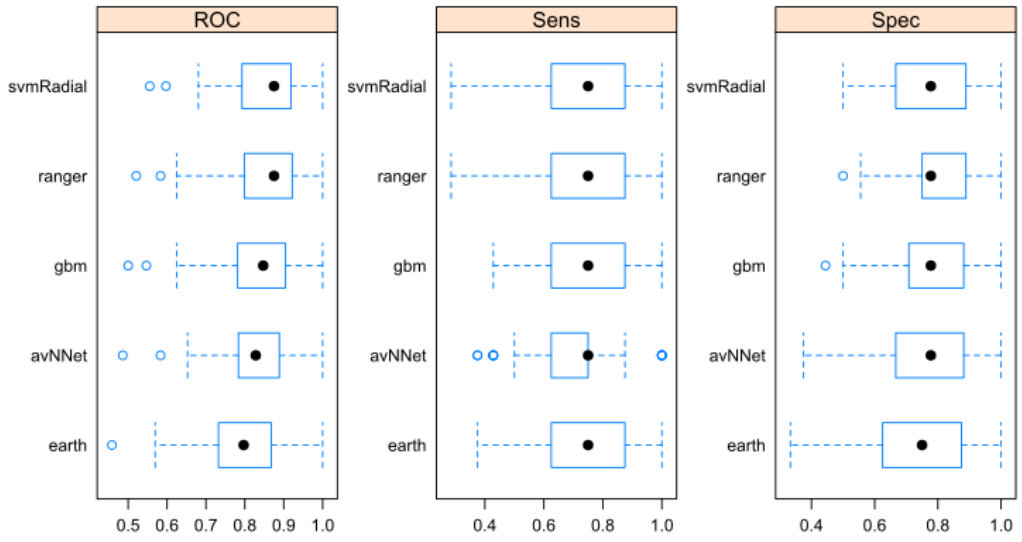


Figure 60: Accuracy (ROC), sensitivity and specificity of models trained using 38 DMPs.

Noting that these models overall seemed to perform less well than those trained with the RFE-selected probes, I once again constructed a GLM from the models which could classify with an accuracy of 77.78%, a sensitivity of 73.08%, and a specificity of 82.14%. The confusion matrix is displayed below in Table 35. Positive Predictive Value (PPV) and Negative Predictive Value (NPV) was 82.1% and 73.1%

respectively, calculated by dividing true positive and true negative predictions by total positive and negative predictions respectively.

	Reference	
Prediction	Control	sCJD
Control	19	5
sCJD	7	23

Table 35: Confusion matrix of generalised linear model classification of sCJD status using five models trained with 38 DMPs.

6.6 Summarising suitability of different datasets and different modelling approaches to machine learning classification of sCJD status

Overall, I found that the top 5000 probes were least suitable for classification, possibly due to insufficiently significant differences between sCJD and control groups. The list of 38 DMPs performed well in models that did not require tuning, but the 33 probes selected by RFE outperformed any dataset when model tuning was part of the pipeline. Had RFE been been performed earlier in the course of the project, it would have been interesting to see whether using unsupervised variable selection to choose candidates for replication yields a higher successful replication rate than manual selection based on *p*-values, effect sizes and potential biological relevance. Interestingly the generalised linear model combination of five different models produced the joint greatest accuracy, providing increased specificity but decreased sensitivity compared to a tuned GBM model alone. The summary statistics per model and dataset are compared in Table 36 and ROC area under the curve values per predictor for the tuned 33-RFE identified variable-trained GBM model are displayed below in Figure 61. It should be noted that while ROC AUC values seem extremely high (in one instance 100%) in this plot, these values relate to the classification efficiency during cross-validation of the training set rather than the model's classification efficacy of the unseen test data.

		Trained			Tuned		
Model	Dataset	Accuracy (%)	Specificity (%)	Sensitivity (%)	Accuracy (%)	Specificity (%)	Sensitivity (%)
gbm	38 DMPs	81.48	76.92	85.71	74.07	67.86	80.77
gbm	33 probes	79.63	84.62	75.00	87.04	88.46	85.71
gbm	5000 probes	65.38	80.77	50.00			
avNNet	38 DMPs				77.78	76.92	78.57
avNNet	33 probes				81.48	84.62	78.57
avNNet	5000 probes						
GLM	33 probes	87.04	92.31	82.14			
GLM	38 DMPs	77.78	73.08	82.14			

Table 36: Comparison of summary statistics of models trained using different datasets. Shown are performance metrics of gradient boosted machine (gbm), averaged neural network (avNNet) and generalised linear models (GLMs) after training and tuning. The 5000 probe dataset and GLMs could not be tuned, and avNNet model parameters were tuned as part of training.

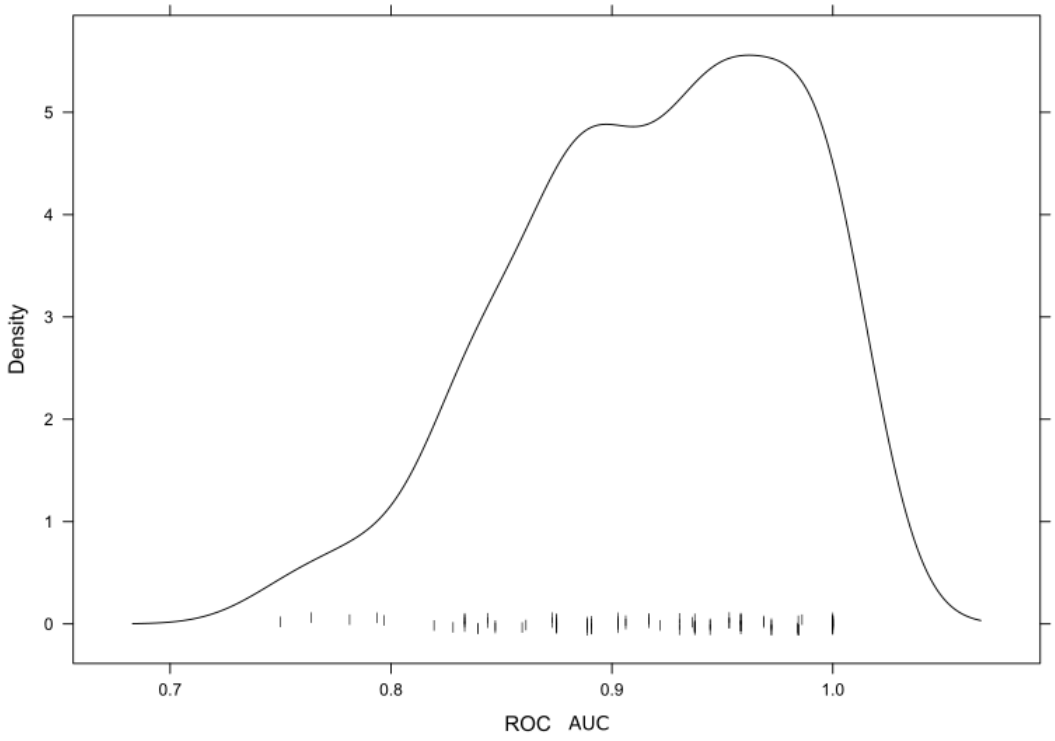


Figure 61: Density plot of area under the curve of receiver-operator characteristic per variable identified using recursive feature elimination used to train the most sensitive and accurate machine learning classifer of sCJD status.

7 Discussion

7.1 Summary

In this project I initially profiled genome-wide DNA methylation in 114 sCJD patients and 105 non-cognitively impaired controls. A case-control comparison using a linear regression model, with age and sex included as covariates, identified 38 loci that were significantly differentially methylated between the two groups. Interestingly, loss of DNA methylation at one of these loci (*AIM2*, cg10636246) correlated with disease severity in sCJD patients only, as measured by the MRC Scale score. Using a tiling approach, I also identified 42 regions across which DNA methylation was significantly altered.

Of these 38 differentially methylated positions (DMPs), I chose 8 to take forward for replication based on the size of mean difference between sCJD and control groups, and the biological role of underlying genes. Although assays could only be designed for 7 loci, 5 were found to replicate in a second cohort of 72 sCJD patients and 114 controls. In addition, using pyrosequencing for the replication stage allowed interrogation of adjacent CpG sites, revealing 5 additional sites which were significantly altered in sCJD. The association of hypomethylation of cg10636246 (in *AIM2*) with disease severity was also observed in this second cohort of sCJD patients

Having confirmed 5 out of seven array-profiled differentially methylated CpG loci and discovered 5 additional loci, I then tested the specificity of these changes to sCJD by repeating the DMP assays using samples derived from Alzheimer's disease, inherited prion disease and iatrogenic prion disease. Although some assays were underpowered (cg02481950 (*METTL9*) for iCJD and IPD, cg17515347 (*AIM2*) for iCJD and IPD, cg03546163 (*FKBP5*) for IPD), none of the changes observed in sCJD were found in other disease groups suggesting that the observed changes are either specific to sCJD rather than to prion disease or neurodegeneration in general.

One of the most interesting findings was a decrease in DNA methylation upstream of the transcription start site of *FKBP5*. This gene plays a role in supressing cellular response to cortisol, leading me to wonder whether circulating cortisol levels are affected in sCJD patients. I curated a group of serum samples from sCJD patients and controls and sent them to The Doctor's Laboratory for cortisol profiling. Cortisol was indeed elevated in sCJD patients within and above the 10am-4pm range (which

is when most samples would have been taken) compared to controls, although cortisol concentration did not correlate with *FKBP5* methylation levels in patients profiled on the array where serum was also available. Because cortisol levels vary according to a circadian pattern, it is perhaps not surprising that blood-derived DNA methylation at *FKBP5* does not correlate with cortisol concentration in sCJD patients: if blood for DNA extraction and blood for serum fractionation were taken at the same time for each patient a correlation might be revealed. Similarly, taking blood for serum fractionation from patients and controls at a fixed time of day would reduce the effects of the circadian pattern and may result in a larger separation of mean concentrations between the groups.

I then addressed the question of where the difference in DNA methylation was coming from. First, to test whether the signature observed in the blood was reflective of a signal in the central nervous system, I repeated the pyrosequencing assays on DNA samples derived from frontal cortex of sCJD patients and non-demented individuals. Strikingly, while none of the effects observed in the blood were found in the brain, sites both in *ANK1* and *PRNP* were found to be significantly differentially methylated in sCJD. While in blood *ANK1* methylation does not differ between sCJD and control, *ANK1* hypermethylation has been repeatedly observed in Alzheimer's disease cortex (De Jager *et al.*, 2014; Lunnon *et al.*, 2014) and observing it in sCJD suggests it is involved in neurodegeneration, in line with recent findings in Huntington's disease and Parkinson's disease (Smith *et al.*, 2019). *PRNP* has an obvious role in sCJD and so observed hypomethylation upstream of *PRNP* in sCJD cortex is an intriguing discovery. However, *PRNP* mRNA has been reported to be decreased in sCJD cortex (Llorens *et al.*, 2013), which, if both findings are correct, runs counter to the convention of demethylation of the 5'-UTR leading to upregulation. Wider profiling of *PRNP* methylation in brain-derived DNA may provide an explanation.

Having found no evidence of the peripheral signal being reflective of changes in the brain, I then decided to investigate the cellular origin of the signal by assaying blood fractions enriched for certain lineages of leukocytes. Sample numbers were extremely limited, so only assays targeting *FKBP5* and *AIM2* were performed. While all fractions showed hypomethylation in sCJD compared to control, these changes were significantly magnified at *AIM2* in granulocytes.

Finally, I trained several machine learning models to classify sCJD/control status using 450K array data, testing the accuracy of each model using different configurations of input data and then tuning parameters to increase accuracy. Using recursive feature elimination, I identified a set of 33 array probes which could be used to predict phenotype with greater accuracy than the 38 significant findings. Using an optimised random forest model, I could classify patient/control status with an accuracy of 87.04%.

7.2 Genome-wide investigation revealed site-specific differences in methylation of blood-derived DNA from sCJD patients

38 DMPs were identified using a linear regression model with age and sex as covariates, and with Bonferroni correction for multiple testing. While statistically stringent, to have confidence that the genome-wide loci identified as differentially methylated were not false positive results it was important to replicate the experiment using a second set of patient/control samples and a second technology. Regrettably there was only scope to reassess a fraction of the 38 DMPs, which were selected based on whether the CpG was in the promoter of a gene, the biological role of that gene's product, the effect size between sCJD and control at that CpG, and association with clinical metrics of disease severity. The expanded rationale for selecting each of the CpGs is detailed below.

7.3 Possible relevance of differentially-methylated genes to sCJD

7.3.1 *AIM2*: cg10636246 and cg17515347

AIM2 is an acronym for Absent in Melanoma 2, despite this gene being highly expressed in primary melanoma (De Koning *et al.*, 2014). AIM2 is one of several pattern recognition receptors (PRRs) which identify molecular signatures of infection or cell damage and form cytokine-mobilising inflammasomes. In particular, AIM2 nucleates and encourages inflammasome assembly on binding to cytosolic double-stranded DNA (Hornung *et al.*, 2009), or spontaneously when upregulated, by interferons for example (Veeranki *et al.*, 2011). PRRs are complexes composed of a PRR, procaspase-1 and ASC, although some PRRs such as NLRC4 and NLRP1 can form inflammasomes without ASC (Latz, Ts and Stutz, 2013). ASC is a prion-like protein (Cai *et al.*, 2014) which forms complexes of dimers referred to as ASC specks, through which it brings procaspase-1 monomers together allowing trans-hydrolysis and release of caspase-1 (Dick *et al.*, 2016). Caspase-1 is responsible for interleukin-1β production and its sustained activation can lead to apoptosis

(Pablo *et al.*, 2009). A model of AIM2 inflammasome activation is shown below in Figure 62.

Aim2 is the most highly expressed PRR in murine neuronal culture (Wu *et al.*, 2016), and its ablation has been found to reduce brain injury in response to ischemia (Denes *et al.*, 2015). In culture Aim2 promotes axonal but inhibits dendritic growth, and deletion in mice promotes anxious behaviour such as hugging corners of boxes and increased passing of urine and stools (Wu *et al.*, 2016). This gene therefore appears to have a neurotrophic role with regards to its capacity for inflammation, but also functions as a mediator of psychological stress. Indeed, in humans a study of post-traumatic stress disorder (PTSD) found that serum C-reactive protein, a biomarker of peripheral inflammation, was elevated in US veterans suffering from PTSD and that this increase and PTSD diagnosis both significantly associated with demethylation of blood-derived DNA at cg10636246 in *AIM2* (Miller *et al.*, 2018).

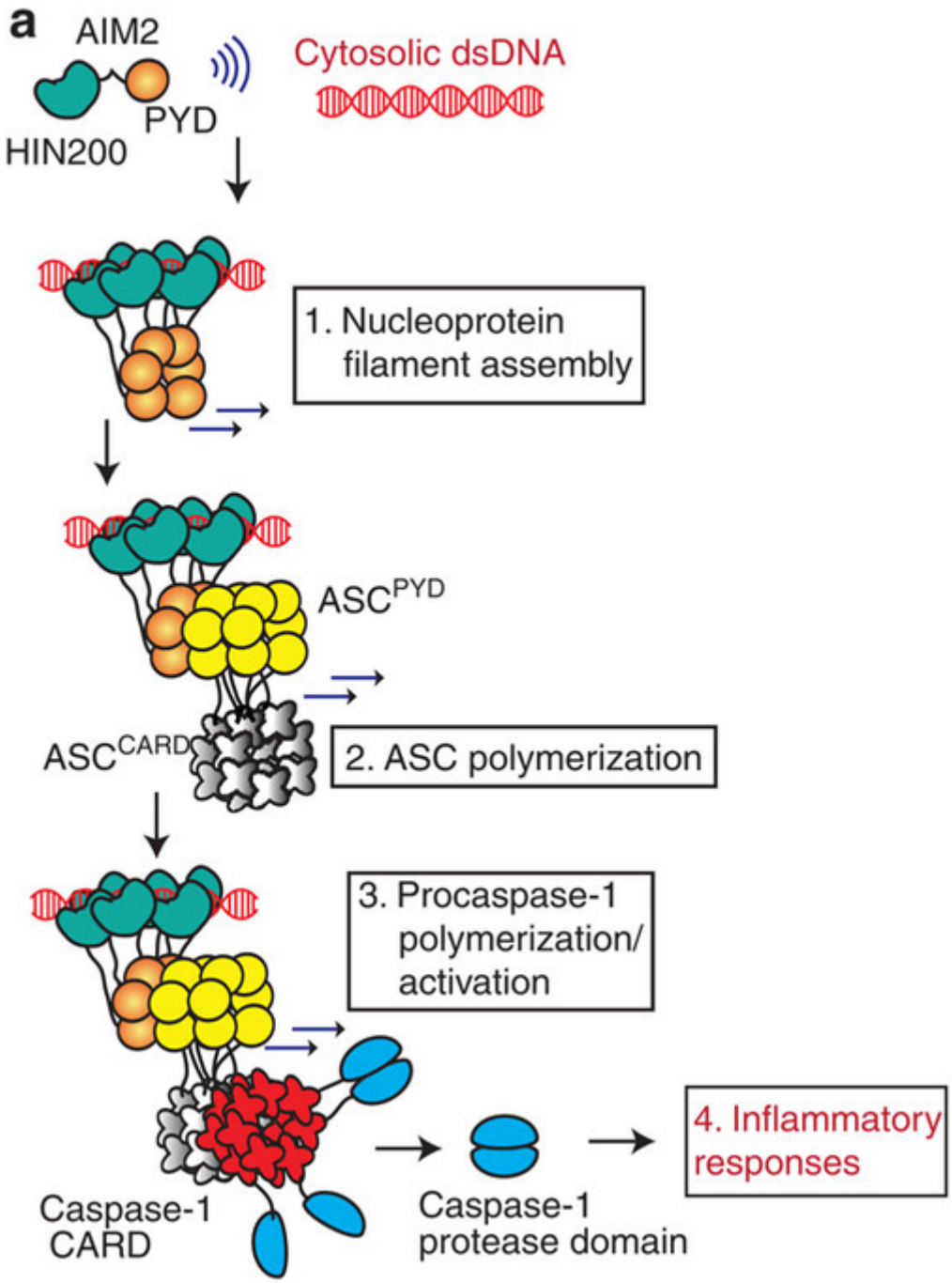


Figure 62: A model of AIM2 inflammasome formation. Shown is AIM2 oligomerisation leading to ASC filament nucleation, recruitment of procaspase-1 dimers and release of caspase-1. Although in this figure AIM2 hexamers form around cytosolic double-stranded DNA, with increased expression an increased expression an increased basal oligomerisation rate becomes sufficient for inflammasome activation. This figure is adapted from Figure 1a from Morrone et al., 2015.

Prion disease has a strong inflammatory component associated with microgliosis and apoptosis of neurons (Aguzzi, Barres and Bennett, 2013). Implication of *AIM2* in sCJD is therefore interesting, particularly as two probes in the transcription start site of the gene are affected, one of which (cg1036246) has been associated with peripheral inflammation in the literature. Association of hypomethylation at these probes with decrease in MRC Scale score, a measure of disease severity, supports a genuine connection between demethylation at this gene's promoter region and sCJD phenotype. It is, however, surprising to not observe differential methylation of *AIM2* in the sCJD brain, which suggests peripheral inflammation in sCJD patients is distinct from neuroinflammation, or occurs in a cell type more abundant in the periphery than the brain.

The role of cytokines in prion disease is unclear, partially due to studies reporting cytokine profiles from different transgenic mice of different backgrounds, infected with different prion strains at different stages of the disease. An early study found that levels of IL-1α, IL-1β, and TNF-α mRNA in the brain were elevated at clinical onset in swiss (SWR/j) mice infected with the mouse-adapted Chandler strain of scrapie (Campbell *et al.*, 1994). PrP^{Sc} deposition was later linked spatially and temporally with increased production of IL-1β and TNF-α by perivascular macrophages in a VM/DK scrapie mouse model of infected with the 301V strain of scrapie (Williams *et al.*, 1997). A study of IM mice describes elevated levels of NF-κβ and IL-6 brains 300 days after infection with the mouse-adapted 87V strain of scrapie (Kim *et al.*, 1999). Conversely, no increase in IL-1β, IL-6 or IFN-γ was observed in a study of black-6 (C57BL/6J) mice sacrificed between 56 and 168 days after inoculation with mouse-adapted ME7 prions (Walsh, Betmouni and Perry, 2001). A study which followed CV/DK mice from inoculation with mouse-adapted ME7 prions all the way until terminal disease found steady elevation of IL-1β mRNA levels from 180 days post inoculation (preclinical stage) to a maximum at end-stage, with some mice showing slight elevations in TNF-α and IL-6 mRNA at end-stage (Brown *et al.*, 2003). The authors comment that only IL-1β is markedly elevated and not to such an extent what could explain extensive microgliosis. Functional insight is provided by observation that knocking IL-10 out of the 129/Sv agouti mouse halved the disease duration upon inoculation with RML 5.0 prions compared to 129/Sv controls (Klein *et al.*, 2004). Taking into account the periods of disease studied, it would seem that cytokine production is likely a feature of the toxic stage of prion pathogenesis, rather than the infectious stage, as per the model proposed by Sandberg *et al*. (Sandberg *et al.*, 2014). This fits neatly with the association of *AIM2*

demethylation with disease severity in humans, suggesting activation of this inflammatory pathway may be a feature of clinically presenting prion disease. Of course, other features of sCJD (incontinence, reduced coughing, predisposition to bedsores) may also cause susceptibility to infections and therefore peripheral inflammation. This is an important point and will be addressed in Section 7.6.

7.3.2 *FKBP5*: cg03546163 (and cg00052684 and cg25114611)

Demethylation of 5.35% in the 5' region of *FKBP5* is the largest significant change seen in this dataset. *FKBP5* was originally identified as a component of progesterone receptor complexes and named immunophilin p54 (Smith, Faber and Toft, 1990). Now mostly referred to as *FKBP5* or *FKBP51* (FK506-Binding Protein 5), this gene has been extensively studied, particularly in its original context as a receptor chaperone. *FKBP5* was found to decrease the affinity of the Glucocorticoid Receptor (GR) to ligand agonists and delay translocation of activated GR to the nucleus (Wochnik *et al.*, 2005), in dynamic opposition to the similarly named homologue *FKBP52* (Smith *et al.*, 1993), now more frequently named *FKBP4*. FKBP5 also serves as a chaperone for AKT and PHLPP, a protein phosphatase which inactivates AKT and thus negatively regulates multiple biological survival and growth pathways (Wang, 2011). *FKBP5* thus serves as a major regulator of the glucocorticoid response pathway: a simplified model of this is presented below in Figure 63.

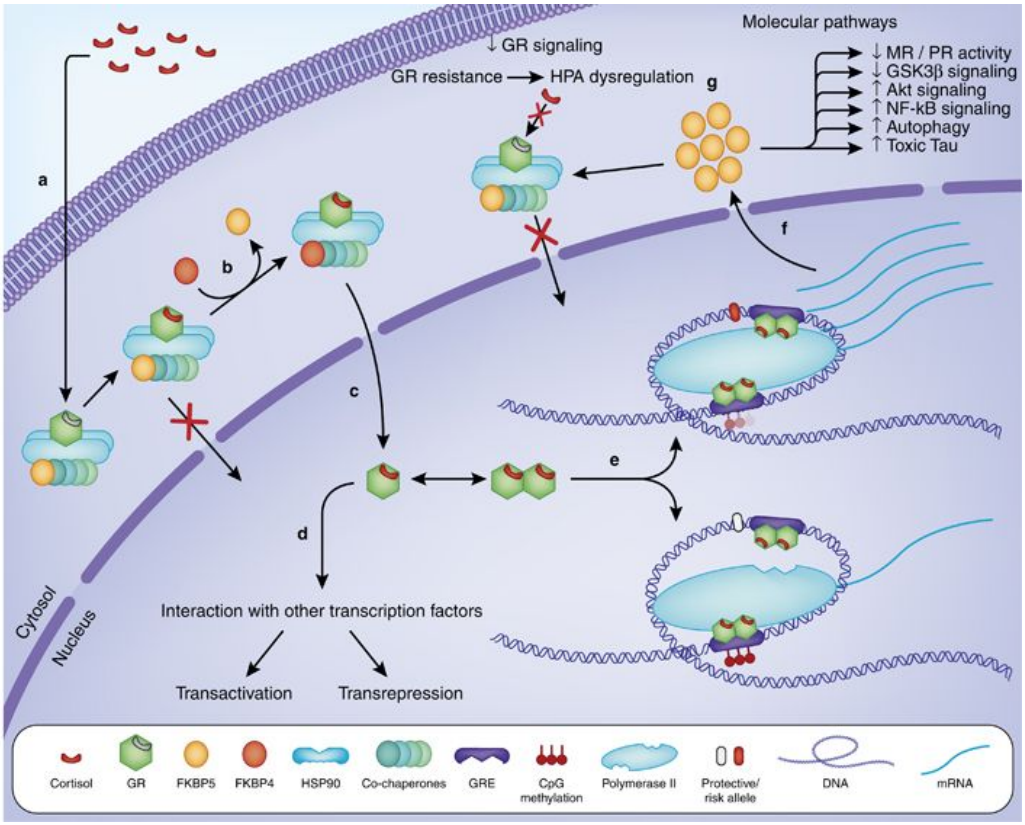


Figure 63: Model of FKBP5's role in inhibition of the glucocorticoid response. Glucocorticoids enter the cytoplasm (a) and activate the glucocorticoid receptor (GR) complex. FKBP5 can associate with the complex and reduce binding affinity of glucocorticoids to the GR, thus delaying translocation of the GR to the nucleus. Exchange of FKBP5 for FKBP4 (b) permits GR translocation to the nucleus (c). The GR can either interact as a monomer with other transcription factors (d) or form a homodimer that binds to DNA at glucocorticoid response elements, allowing transactivation or transrepression of a large number of genes. The FKBP5 *gene is highly responsive to GR, but responsiveness depends on* FKBP5 *polymorphisms and methylation status (e).* FKBP5 *mRNA translocates to the cytoplasm (f) where it is translated into FKBP5 protein. FKBP5 then inhibits GR activity by not only forming an ultra-short, intracellular negative feedback loop of GR signaling but also modulating several other biological pathways (g). This figure and legend are adapted from an excellent review of FKBP5's function and regulation by Zannas, Wiechmann, Gassen and Binder 2016.*

As well as functioning as a scaffold for protein assembly, FKBP5 has a cis/trans-peptidyl-prolyl-isomerase (PPIase) function, common to the FKBP family which are named according to the inhibition of the PPIase domain by FK506. PPIase function also contributes to *FKBP5*'s role as a chaperone, for example in its inactivation of cyclin-dependent kinase 4 by isomerisation of the T172-P173 peptide bond, by which phosphorylation of T172 is inhibited, thus promoting myogenesis (Ruiz-Estevez *et al.*, 2018).

There is a connection between *FKBP* family members and prion disease in the literature. FK506 (also known by its pharmaceutical name, tacrolimus) was identified in a drug screen in murine N2a cell culture as preventing prion propagation (Karapetyan *et al.*, 2013). Further work revealed that FK506 decreases levels of PrP^{C} through blocking import into the endoplasmic reticulum, leading to increased degradation in the cytoplasm, identifying endoplasmic reticulum-associated Fkbp10 as the target of FK506 (Stocki *et al.*, 2016). The same work revealed that FK506 could also decrease levels of PrP^{C} and PrP^{Sc} in chronically infected ScN2a cells. Knockdown of *Fkbp9* in a prion-propagating derivative of N2a cells (PK1) was later found to significantly reduce prion propagation while not altering base levels of PrP^{C} (Brown *et al.*, 2014). Worth noting is that mutations at P102 or P105 of PrP can cause Gerstmann-Straussler-Scheinker disease, an inherited form of prion disease (Kraus *et al.*, 2015). Of the canonical amino acids, proline exerts the greatest constraint on conformational potential at the level of secondary structure, suggesting that these disease-causing polymorphisms may act through increasing the conformational freedom of PrP. If this is the case, the effects of *FKBP* family PPIase activity on PrP conformation may be worth considering.

FKBP5 has also been associated with other neurodegenerative and psychiatric diseases. Upregulation of *Fkbp5* in mice expressing humanised P301L tau resulted in accumulation of toxic tau oligomers and progressive neurodegeneration (Blair *et al.*, 2013). Conversely, *Fkbp5*$^{-/-}$ mice have lower levels of total brain tau and phospho-tau and show resilience to experimental stress conditions compared to wildtype mice in terms of reduced depressive behaviour and anxiety, improved quality of sleep and lower levels of circulating corticosterone, the rodent equivalent of cortisol (O'Leary *et al.*, 2011; Touma *et al.*, 2011; Albu *et al.*, 2014). Overexpression of *Fkbp5* was observed in the hippocampus and prefrontal cortex of rats subjected to chronic stress (Guidotti *et al.*, 2013). Strikingly, significant and correlated decreases in DNA methylation of *Fkbp5* were observed in the blood and hippocampi of mice supplied with corticosterone-laced water (Ewald *et al.*, 2014). In humans dysregulation of *FKBP5* is also observed in the context of stress, notably in survivors of the holocaust and their offspring, where a CpG overlapped by one of *FKBP5*'s glucocorticoid response elements (GRE) is hypermethylated and hypomethylated in parents and children respectively (Yehuda *et al.*, 2015). Dawn cortisol levels in children are also elevated compared to age-matched non-exposed Jewish controls, in inverse correlation with demethylation of the GRE. In Cushing's

Syndrome, a psychiatric disorder caused by hypercortisolaemia, *FKBP5* is also found to be hypomethylated. (Resmini *et al.*, 2016).

Because of its diverse functions, its existing connection with prion research, and the magnitude of hypomethylation observed in the array results, *FKBP5* is an exciting candidate gene for further study. It is also interesting to note that cortisol is an anti-inflammatory agent, and hypercortisolaemia has been observed in scrapie-infected sheep (Gayrard *et al.*, 2000). There may therefore be an inflammatory component connecting dysregulation of *AIM2* and *FKBP5*.

One CpG (cg03546163) was identified in the sCJD-control comparison using Houseman-corrected data. However, EpiDISH correction for cellular heterogeneity identified this probe and two others (cg00052684 and cg25114611). While it was encouraging to observe additionally affected loci in the same gene, this also presented an opportunity to assess the ability of EpiDISH to identify significantly affected loci which the Houseman algorithm could not, therefore all three probes were taken forward as candidates for replication.

7.3.3 *MTRNR2L8:* cg05740793

MTRNR2L8, or humanin-like Protein 8, is a paralogue of 15 nuclear-encoded descendants of a putative ancestral mitochondrial gene (Bodzioch *et al.*, 2009). A mere 24 amino acids in length, the original humanin was identified through expression screening of a cDNA library generated from the occipital cortex of an Alzheimer's disease patient, and allegedly named after its discoverer's hope that this peptide was the key to restoring the humanity of AD patients (Hashimoto *et al.*, 2001; Lee, Yen and Cohen, 2013). Hashimoto *et al*. found that treatment with purified humanin was sufficient to protect cells transfected with amyloidogenic mutant amyloid-precursor protein or treated with amyloid-beta oligomers. Later work found that similar treatments of humanin could protect cortical neurons from apoptosis induced by a codon 118-135 fragment of PrP (Sponne *et al.*, 2004). *In vivo*, interperitoneal injection of a synthetic humanin mimic peptide was found to reduce memory deficits, amyloid plaque counts and neuroinflammation in an AD mouse model (Zhang *et al.*, 2012).

Its neuroprotective functions make Humanin a fascinating candidate gene, but investigation was stymied by the high degree of conservation of the sequence around the CpG of interest. I was unable to design a pyrosequencing assay specific for humanin-like protein 8. This is particularly unfortunate as, despite differential

expression of 10 of the humanin proteins across different tissues (five are thought to be pseudogenes), which humanin paralogue encodes the “original” humanin peptide isolated from the brain remains a matter of contention.

7.3.4 *UHRF1* and *METTL9*: cg17714703 and cg02481950

UHRF1 was introduced at the introduction of this thesis: it recruits DNMT1 to hemimethylated DNA to establish symmetrical DNA methylation at CpG sites. Methyltransferase-like 9 is so called because it contains an SAM-dependent methyltransferase domain, however which biological substrate it methylates has not been confirmed. *METTL9* is also named DORA Reverse Strand Protein (*DREV1*) as it is encoded by the opposite strand of a region overlapping *IGSF6*, or *DORA*, an immunoglobulin associated with inflammatory bowel disease which is specifically expressed in haemoatopoetic cells (Bates *et al.*, 2000). While *UHRF1* has been subject to much research, *METTL9* remains uncharacterised. However, its potential role in DNA methylation, alongside *UHRF1*’s established role, made them interesting candidates for replication. Dysregulation of *UHRF1* in particular might explain the increased genome-wide inflation seen in the array results: if spontaneous DNA methylation is being differentially converted to symmetrical DNA methylation at CpG sites by dysregulated UHRF1, I would expect to see greater background differences in DNA methylation between sCJD and control groups.

7.3.5 *KCNAB2* and *MIR1977*: cg02448796 and cg05001044

KCNAB2 encodes Kvβ2, a subunit of potassium voltage-gated channels, and was selected primarily as bioinformatic analyses of other data gathered from sCJD samples indicated enrichment of potassium voltage-gated channel family members. While potassium channels manage a wealth of biological functions, *Kcnab2* deletion was found to impair memory and learning in mice (Perkowski and Murphy, 2011), while both polycomb-targeted repression and knockdown of *Kcnab2* was found to improve resistance to ischaemia in murine culture (Stapels *et al.*, 2010).

Conversely, *MIR1977* is a mitochondrially-encoded microRNA which is thought to target and silence the mitochondrial transfer RNA gene *TRNN* (Bandiera *et al.*, 2011), although this has not been confirmed. While this miRNA’s function remains unknown, the Δβ between sCJD and control is the second largest in the dataset and so this site is a suitable candidate for further investigation.

7.4 Genome-wide significance inflation in 450K datasets is common and is unlikely to detract from this study

As shown in Section 3.3.1, pre-processing suppressed genome-wide significance inflation (λ) to a final value of 1.4, which is higher than reported in most Genome-Wide Association Studies (GWAS). While this may reflect the clinical heterogeneity of the disease or "noisy" epigenetic dysregulation across the genome, *p* value inflation in Epigenome-Wide Association Studies (EWAS) is a recognised phenomenon (Iterson, Zwet and Heijmans, 2017). While GWAS are studies of genotypes (ordinal data), EWAS of DNA methylation study continuous numerical data which is more vulnerable to the biases described in Chapter 2. Alternatively, sCJD could have extremely heterogeneous effects on genome-wide DNA methylation. In terms of whether this inflation is greater than expected, λ of 1.3-1.4 have been reported in other studies using the 450K array and blood-derived DNA (Ahsan *et al.*, 2016; Joehanes *et al.*, 2016). It is therefore unclear whether further correction of λ would reflect a legitimate decrease in technical bias or over-fitting of the data to conform to standards established in GWAS, although such correction could be achieved using statistical techniques such as surrogate variable analysis.

The greatest decrease in λ is observed before and after correction for leukocyte population heterogeneity. Briefly, this uses reference methylome datasets from FACS purified leukocyte-derived DNA to weight 450K array data in terms of estimated leukocyte proportions, and then normalises the data based on those estimates. In doing so, differential methylation based on differences in cell populations, rather than epigenetic dysregulation, is discounted (Jaffe and Irizarry, 2014). In our dataset both Houseman and EpiDISH algorithms estimate significant differences between leukocyte proportions in sCJD and control groups, albeit within normal physiological ranges. This raises two questions: are there legitimate differences in cell populations in sCJD compared to control, and what is driving the discrepancy between the Houseman and EpiDISH algorithms? Regrettably blood counts of sCJD patients were not accessible at the time of the study, and prospectively counting blood cell populations to statistically power a difference between sCJD and controls, assuming both are within normal physiological ranges, would take months to years. To answer the second question, a large battery of blood samples from more than one phenotypic group should be subjected to leukocyte population counting and 450K or 850K profiling, so as to rate the comparative accuracy of the two algorithms. This has been done on a small scale

by the authors of the EpiDISH algorithm, but bears repeating (Teschendorff *et al.*, 2017).

7.5 Genome-wide investigation revealed regional differences in methylation of blood-derived DNA from sCJD patients

41 differentially methylated regions (DMRs) were identified using the Bumphunter package. The reduction of data dimensionality through tiling regions of adjacent CpG probes allowed sites that did not exhibit genome-wide significance after Bonferroni correction for multiple testing to pass a false-discovery rate corrected significance threshold of 0.05. *p* values are generated by creating a 250 series of "null" regions and calculating the percentage of null regions that are greater in width and have a higher value than the observed DMR. The package's authors advise that such *p* values "should be interpreted with care as the theoretical properties are not well understood" (Aryee *et al.*, 2019). A second metric of error, the Family Wise Error Rate (FWER) is the proportion of series of null regions that contain at least one region as wide and with an equal or greater value than the observed DMR. FWER is a more stringent measure of error and the one by which I decided to interpret DMR results.

The most significant DMR by both metrics ($p = 3.3 \times 10^{-5}$, FWER = 0.012) overlaps three genes: *HOXA-AS3*, *HOXA5* and *HOXA6*. Less significant DMRs tend to have sequentially higher *p* values, while FWER increases linearly. The second most significant DMR as ranked by Bumphunter has an FWER of 0.056, meaning that 5.6% of null series contain an equivalently "significant" DMR in terms of difference between mean methylation and total width. The FWER reaches 90.4% at DMR number 21 and so, whilst nominally significant in terms of *p* value, it seems probable that many of the identified DMRs are unreliable.

Nevertheless, I went on to study the genomic distributions of these regions. Perhaps unsurprisingly most DMRs overlapped CpG-dense islands rather than CpG-depleted open sea regions, and as a result also overlapped gene promoters. Using publicly available ChIP-Seq data from purified blood cell types in the Goldmine package (Bhasin and Ting, 2016), I also found that all DMRs contained transcription factor binding sites (TFBSs). I decided to see whether these DMRs shared any common motifs or binding partners and identified two motifs as being enriched across the sequence, namely TATTTTTTTATTAACAAAATATAACATTA and TTTTCTTCCTCTCCA (Motifs 1 and 2 from Figure 33 respectively). 86 transcription factors were significantly associated with the motifs, although after adjustment for

false discovery rate only 14 remained significantly associated with the first motif only.

7.5.1 *HOXA5*

The DMR least likely to be a false positive (1.2% of null regions show equivalent or greater effects) overlaps *HOXA5*. *HOX* genes are developmental transcription factors (Krumlauf and Hill, 1994) which are expressed in different combinations across distinct anatomical segments, thus directing local development of anatomical structure. *HOXA5* in particular contributes to the development of the respiratory and digestive systems, thyroid and mammary glands, and regulates the balance between myelopoiesis and erythropoiesis in the bone marrow (Jeannotte, Gotti and Landry-Truchon, 2016). Besides their normal physiological role in development, dysregulation of *HOX* genes, including *HOXA5*, have been implicated in neurodegeneration. In blood-derived DNA from patients with Down Syndrome, which confers a high risk of developing Alzheimer's Disease with age, hypomethylation (as opposed to hypermethylation as observed in sCJD) has been described across *HOXA2*, *HOXA4*, *HOXA5* and *HOXA6* (Bacalini, Gentilini, *et al.*, 2015). In Huntington's disease *HOX* genes have been found to be upregulated in prefrontal cortex, alongside upregulation of five *HOX*-targeting miRNAs, some of which were not expressed at all in non-Huntington's control cortex (Hoss *et al.*, 2014). Perhaps most striking is the recent publication of a 48 kilobase DMR spanning the *HOXA* cluster in prefrontal cortex and superior temporal gyrus-derived DNA across Braak stage in Alzheimer's Disease patients (Smith *et al.*, 2018). The seventh DMR (as ranked by significance) identified in this report overlaps *HOXA5* and *HOXA-AS3* (Chr7:27,183,133-27,184,853), a region of DNA overlapped by the DMR identified in this study (Chr7:27,183,133-27,185,512). Both DMRs are regions of hypermethylation.

7.5.2 Further work

It is clear from the studies cited above that *HOXA5* plays a neurotrophic role. It may be that hypermethylation at binding sites within DMRs is reflective of hypermethylation at other binding sites which could connect to differential expression of genes not identified in this study. However, this would only be the case if the substrate binding of these transcription factors is methyl-sensitive. This assumption is driven solely by *in silico* observations but could be tested through a number of follow-up experiments:

1. Replication of the *HOXA5* DMR using targeted bisulphite sequencing: Because bisulphite treatment fragments DNA, a 2,379 bp region may require shotgun sequencing or primer walking, where small sections of this region are sequenced and fragments combined to create a profile of the whole region.
2. qPCR of *HOXA5*: If *HOXA5* is indeed dysregulated, this should be confirmed by measuring its expression levels in sCJD and control blood.
3. Chromatin Immunoprecipitation of DNA sequences bound to identified transcription factors, e.g. *CPEB1*: if hypermethylation of motifs within DMRs is altering binding of transcription factors to these sequences, precipitation of CPEB1-bound chromatin and subsequent sequencing (or qPCR targeted around these motifs) should give an indication of magnitude and direction of effect.

Even given these hypotheses, the disease relevance of blood-based changes in methylation, expression and activity of these genes is not easy to connect to prion pathology in the brain. It may be possible the observed changes are present in the brain too, or that they are symptomatic of PCR2 dysfunction related to altered haematopoiesis in response to neuronal or peripheral inflammation, which will be discussed in the next section.

7.6 Five out of seven candidate DMPs were found to be replicable in a second cohort of patients and controls

As mentioned in Section 4.1.2, assays for all loci were designed and optimised with the exception of the assay for *MTRNR2L8,* due to CpG density in the surrounding region. A *MTRNR2L8*-specific 300bp amplicon was generated but failed to sequence, while nested PCR of a smaller amplicon from a larger, specific amplicon failed to provide a product.

Five out of seven assays based on the Houseman-corrected data showed significant differences between sCJD and control, while two (*MIR1977*, *KCNAB2*) failed to replicate. One of the two EpiDISH identified assays (cg00052684) replicated but at dramatically reduced significance (p = 0.002). The second assay (cg25114611) did not reach significance after correction for multiple testing (p = 0.029). It is possible that, based on a lower replication rate and a higher λ in the EpiDISH dataset, that EpiDISH is subject to a higher false positive rate, although more EpiDISH DMPs would need to be replicated to confirm this.

Because of the extensive correction measures employed during array data analysis, overfitting of the data was a concern. Apart from inclusion of age and sex as components in the linear model, pyrosequencing data is not corrected. Therefore, the replicated DMPs are not artefacts of data as they are present in both corrected and uncorrected data.

It is possible that the replicated DMPs are consequences of prion disease, dementia or illness in general. To exclude these possibilities the replicated assays were performed on groups of iatrogenic CJD and inherited prion disease patients, as well as Alzheimer's disease patients. There are two caveats to this experiment: firstly only the AD group was powered in the assay for *METTL9*, this requiring 50 samples for both disease and control groups. Secondly, because the majority of iCJD patients are male, in these analyses sex was not included as a covariate in the linear model to prevent bias. As sex did not correlate significantly with differences in methylation between disease/control replication groups, this is unlikely to have adversely affected results.

Strikingly, no DMPs identified in sCJD patients were observed in other disease groups. Not only does this support the specificity of the DMPs to sCJD, but also rules out the effects being induced by disease treatment as drugs prescribed to treat symptoms of sCJD, IPD and iCJD are broadly similar (anti-epileptics, benzodiazepines). Consultation with clinical fellows who treat prion disease patients suggests no form of the disease carries a disproportionate risk of infection as a comorbidity, which could also be responsible for altered DNA methylation (particularly of *AIM2* and *FKBP5*). These results and observations support the specificity of these DMPs to sporadic human prion disease.

7.7 Differential DNA methylation in sCJD patient blood is not observed in frontal cortex-derived DNA

Differential DNA methylation in blood is a useful biomarker and can give insight into downstream consequences of the disease or germline susceptibility factors, while altered DNA methylation in the disease tissue may identify mechanistically relevant candidate genes and pathways. Having identified DMPs in peripheral blood of sCJD patients, investigating these loci in brain tissue was a natural next step.

Fortunately frontal-cortex derived sCJD and non-sCJD control DNA was available, and in sufficient sample numbers to power assays for *FKBP5, AIM2,* and *UHRF1*. Assays for *ANK1* and *PRNP* were included, but this time not as negative controls.

cg11823178 in ANK1 is hypermethylated in Braak stage IV-VI Alzheimer's disease frontal cortex compared to Braak stage I-III (De Jager *et al.*, 2014; Lunnon *et al.*, 2014), and DNA methylation of *PRNP* has yet to be investigated in human brain.

In an ideal, prospective study of brain-derived DNA, tissue would be sampled from the same region of the frontal cortex, specifically the grey matter which contains the neuronal soma rather than the white matter which contains neuronal axons. In this case the DNA is derived from bulk frontal cortex which would not have been systematically sampled in this way. Despite this limitation, it is interesting to observe no changes in DNA methylation levels at *AIM2*, *UHRF1* or *FKBP5* loci between sCJD and non-sCJD control brain-derived DNA. While the *METTL9* assay was not powered, the distribution of the data when plotted does not suggest that increasing sample numbers would reveal an effect.

Interestingly, I found that a CpG located 21 bp downstream of cg11823178 (*ANK1*) was found to be hypermethylated in sCJD brain. This CpG (8:41,519,420) has recently been identified as hypomethylated in entorhinal cortex in Huntington's disease (HD) and Parkinson's disease (PD) (Smith *et al.*, 2019), though notably the three upstream CpGs in this assay, while hypermethylated in HD, PD and AD, are not in sCJD. Laser capture of cells from AD hippocampi revealed a four-fold upregulation of *ANK1* mRNA compared to controls in microglia, but not neurons or astrocytes (Mastroeni *et al.*, 2017). As extensive microgliosis is a feature of sCJD, it is possible that *ANK1* may also be worthy of study in different regions of the brain and at different omic levels in prion disease.

PRNP showed significant hypomethylation at position 20:4,665,649, 3 base pairs downstream of cg04286737. While methylation at a CpG 576 bp upstream of *Prnp*'s transcription start sight has been found to affect expression in mice (Dalai *et al.*, 2017), this assay probes a site even further upstream of *PRNP*. Yet studies of cortical expression of *PRNP* show decreased mRNA levels in the cortex, which in turn correlates with decreased PrP^C in cerebrospinal fluid, which in turn correlates with sCJD severity (Llorens *et al.*, 2013). This discrepancy between observations in mice and humans warrants further study of *PRNP* methylation and expression *in vivo* and *in vitro*, as Llorens' observations may reflect normal or increased transcription of PRNP but subsequent mRNA dysmetabolism, as observed in other neurodegenerative diseases (Liu, Cali and Lee, 2017).

7.8 Differences in DNA methylation may be associated with specific leukocyte classes

eFORGE analysis suggests that the DMPs are enriched in loci overlapped by the histone moiety H3K36me3 in monocytes, B cells and natural killer cells. Methylation of lysine 36 in Histone protein 3 is associated with active transcription, as SET2, a H3K36 methyltransferase, binds to phosphorylated RNA Polymerase II during transcript elongation (Krogan *et al.*, 2003; Li *et al.*, 2003). Resultantly H3K36me3 abundance correlates with expression levels and, mirroring RNAP II association, this mark is more frequently found in exons rather than introns (Schwartz, Meshorer and Ast, 2009). H3K36me3 is also recognised by MRG15 which recruits PTP, a negative regulator of alternative splicing (Luco *et al.*, 2010). Altered DNA methylation at these sites may indicate changes in chromatin organisation from condensed to unpacked, which in turn may allow methylation or demethylation of H3K34me3.

The presence of functional chromatin modifications at these DMPs in particular lineages of leukocytes allows inference of a cell-specific origin for the observed changes in DNA methylation. This would also explain the comparatively modest effect sizes observed: monocytes comprise ~5% of total leukocytes in adults, so a change in DNA methylation of 10% in *FKBP5* specifically in monocytes would actually be a change of 50% between sCJD and control, diluted out by no change in DNA from other cell fractions. I set out to test this by prospectively enriching leukocyte populations using Magnet Assisted Cell Sorting (MACS).

7.8.1 Strengths and limitations of MACS methodology and results

The greatest strength of MACS is that it allows fractionation of leukocytes rapidly, with minimal preparation and under less mechanically stressful conditions than FACS. However, validation of MACS efficiency is rarely performed in the literature and the manufacturer's validation uses FACS. One of the practical reasons for choosing MACS to fractionate cells is that FACS was not a viable method due to health and safety issues connected with handling of prion patient samples, ruling this out as a validation technique. Instead, qPCR of cell surface markers was decided upon as a means of measuring enrichment of the cell type of interest in each fraction. This required the eluted fraction to be effectively halved in volume to allow for both DNA and RNA extraction. While DNA yields were sufficient for bisulphite conversion and pyrosequencing, RNA yields were low. Eventually, time

did not permit for the qPCR assays to be performed, meaning the MACS results have not been validated and should be considered with caution.

Another limitation is that blood samples had to be recruited prospectively as leukocytes lyse on freezing, meaning frozen samples could not be used. While clinic visits to patients are relatively frequent given the rarity of the disease (1-2 per week), on several occasions patients were too unwell to have blood taken. Because of this only 11 controls and 9 sCJD patient bloods were collected. Moreover, in some instances the volume of blood taken was insufficient to purify all chosen fractions, in which case monocyte and B cell fractions were prioritised. A final limitation was delay before sample processing, as some samples were processed within hours of venepuncture whereas others were processed a few days after venepuncture.

Surprisingly all cell fractions tested showed a trend towards demethylation in the *AIM2* loci, but a statistically significant difference was only observed in granulocytes, a cell lineage not identified by eFORGE. Conversely, while all cell lineages showed a trend towards demethylation at *FKBP5*, the ANOVA was not significant. However, the post hoc test, and the data when visualised, point towards an enrichment of demethylation in B cells. This assay is underpowered and recruiting more samples may lead to a significant difference in this lineage being revealed.

The granulocyte lineage is made up of neutrophils, eosinophils and basophils, all characterised by their ability to secrete granules containing cytotoxic or signalling molecules (Geering *et al.*, 2013). In blood neutrophils are the most abundant granulocyte and persist for up to 5 days in circulation. However, spontaneous apoptosis of neutrophils is inhibited by both glucocorticoids and Interleukin-1β (Cox *et al.*, 1995; William *et al.*, 1998), signalling molecules whose effects are regulated by *FKBP5* and *AIM2* respectively. Both EpiDISH and Houseman algorithms estimated greater proportions of granulocytes in sCJD (+8% and +11% respectively), suggesting that increased granulocyte (and specifically neutrophil) survival and thus prevalence may be a peripheral feature of sCJD.

7.8.2 The value of further work as a function of cost

This experiment is underpowered and has yet to be validated, but preliminary results are encouraging. As granulocytes seem to be affected and (based on *in silico* estimates) elevated in sCJD, it may be worth prospectively comparing proportions of leukocyte populations between sCJD patients and controls. This will

both give a clearer picture as to what peripheral immunological processes are occurring in sCJD and may also prove useful in supporting diagnosis.

However, such work is time-consuming, requires prospective blood sampling from patients of a rare disease, and although the leukocyte lineage is estimated to be increased in sCJD, populations in both sCJD and control bloods are within physiological ranges. It is also surprising that eFORGE indicated that observed changes would be most amplified in monocytes rather than granulocytes. These preliminary results are perhaps insufficient to justify expanding this work without validation of the efficiency of MACS enrichment, which could be done using control blood and validated by FACS rather than qPCR of cell surface markers.

7.9 *FKBP5* and *AIM2*: Peripheral inflammation in neurodegenerative disease

Hypomethylation of *FKBP5* and *AIM2*, and an enrichment of the latter in granulocytes, point towards inflammation in the periphery. Cortisol, a hormone with anti-inflammatory properties, was also found to be elevated in sCJD sera compared to control sera. The relationship between inflammation and neurodegeneration is well established, with chronic neuroinflammation observed in most neurodegenerative diseases (Frank-Cannon *et al.*, 2009). Recently, peripheral inflammation in neurodegenerative disease has become a topic of interest. In Alzheimer's Disease and Lewy Body Dementia (LBD) inflammatory cytokines such as IL-1β were found to be elevated in plasma from prodromal patients but not in those who were cognitively impaired, compared to controls (King *et al.*, 2018). However, a meta-analysis of 175 AD studies found that cytokines were in fact elevated in clinically presenting patients, with IL-6 levels correlating inversely with Mini Mental State Examination scores (Lai *et al.*, 2017). In Huntington's Disease patients chemokine levels in plasma were found to increase linearly across disease progression and correlate with clinical scores (Wild *et al.*, 2011). A meta-analysis of 25 studies found elevated levels of cytokines were also elevated in ALS (Hu *et al.*, 2017), and treatment of peripheral blood mononuclear cells with α-synuclein significantly increased IL-6 and IL-1β production.

It is therefore perhaps unsurprising that demethylation of *AIM2* correlates with MRC Scale score, given the prior reports of similar associations between clinical scores and inflammatory markers in AD, HD and LBD. But how is peripheral inflammation connected to neuroinflammation? During this project I considered several hypotheses.

7.9.1 Potential causes and roles of peripheral inflammation in sCJD

The first is that immune cells infiltrate the brain, are affected by the disease microenvironment, and return to circulation. This hypothesis was strengthened by the eFORGE prediction of an enriched differential methylation signature in monocytes, a leukocyte class which are known to infiltrate tissues and in particular to be recruited into the brain early on in prion pathogenesis (Williams, Ryder and Blakemore, 1995). This hypothesis explains the relatively small effect sizes observed, as only a small fraction of cells would infiltrate the brain and a smaller fraction of them would return to the periphery. However, preliminary MACS results do not support this hypothesis. The lack of concordant blood/brain changes in methylation are also problematic, although it could be that leukocytes respond differently to the disease microenvironment than native neuronal, astrocyte or glial cells.

The second is that neuroinflammatory factors leak into the periphery, perhaps due to impairment of the blood brain barrier (BBB), triggering epigenetic changes. As IL-1β regulates BBB permeability, neuroinflammation should result in cytokines passing into the periphery, triggering systemic inflammation. Again, these molecules would be diluted across the total volume of the circulatory system which may also explain the relatively modest effect sizes observed. IL-1β is known to increase survival of neutrophils, so this hypothesis does explain both the estimated increase in granulocyte proportions in sCJD and the preliminary MACS results.

Thirdly, it is possible that prions are present in the circulatory system. Prions have been detected in the blood of vCJD patients (Concha-Marambio *et al.*, 2016; Nicot *et al.*, 2016), but whether they are present in the blood of sCJD patients remains unclear. A bioassay where plasma from sCJD patients was injected intracerebrally into transgenic mice expressing human *PRNP* showed prion infectivity in 2 out of 4 patient samples (Douet *et al.*, 2014). Varying circulating prion titre between patients may also explain the relatively modest degree of separation between sCJD and control groups, and as cited above Aβ and α-synuclein (both prion-like proteins) are sufficient to trigger PRRs and thus activate inflammasomes. This hypothesis is also supported by the estimated increase in granulocyte proportions in sCJD and the preliminary MACS results. Whether or not prions are present in the circulatory system of sporadic prion disease patients is an important question both in terms of pathophysiology and research biosafety.

At this stage, hypotheses two and three seem to be the most plausible and rely on fewer assumptions than hypothesis one. Profiling cytokine concentrations and leukocyte proportions in sCJD blood across the disease course would clarify whether there is indeed IL-1β driven inflammation and increased neutrophil survival, and whether these correlate with *AIM2* demethylation. Comparison of IL-1β levels in human sCJD brain compared to control brains with no expected BBB impairment using immunohistochemistry would not only reveal whether cytokines and chemokines could leak from the brain into the periphery, but correlation of IL-1β presence in the brain with levels in blood could reveal a linear connection.

Continuing and validating the MACS experiment is important to clarify which cells may be functionally affected by differential methylation. Enriched fractions of leukocytes can also be used, alongside plasma, in RT-QuIC or PMCA reactions – sensitive *in vitro* assays for PrP^{Sc} seeding – to establish whether prions are present in the periphery, and whether particular types of leukocyte are vulnerable to infection.

7.10 Elevated cortisol in sCJD suggests dysregulation of the Hypothalamic-Pituitary-Adrenal Axis

Hypomethylation of the promoter of *FKBP5* and elevated levels of cortisol in sCJD patient sera both point towards potential dysregulation of the Hypothalamic-Pituitary-Adrenal (HPA) axis. Briefly, this system coordinates signals generated by environmental or molecular stress stimuli – such as IL-1β – from various neural circuits to the hypothalamus, which releases corticotrophin-releasing hormone and arginine vasopressin. These stimulate the release of adrenocorticotrophic hormone (ACTH) by the pituitary gland, which travels through the circulatory system to the adrenal glands, which respond by producing and releasing glucocorticoids, such as cortisol (Bellavance and Rivest, 2014). It is highly likely that damage to the brain, such as occurs during sCJD, would through inflammation and IL-1β affect the function of the HPA axis. Because the HPA axis operates according to circadian and ultradian rhythms, many models of its function exist: a simplified model is displayed below in Figure 64.

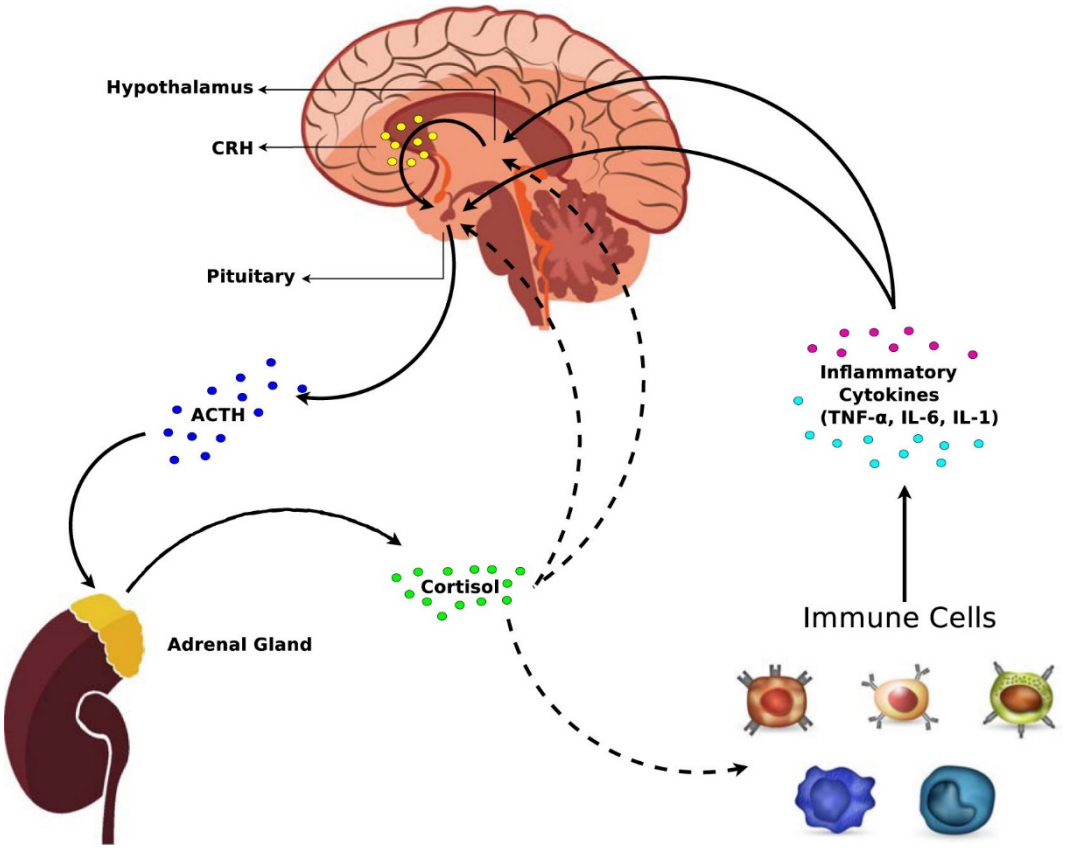


Figure 64: Simplified model of the Hypothalamic-Pituitary-Adrenal axis. Here the hypothalamus receives stressor input (in this case inflammatory cytokines) and produces corticotrophin-releasing hormone (CRH). This stimulates the pituitary gland to secrete adrenocorticotrophic hormone (ACTH), which travels through the blood to the adrenal glands and promotes the secretion of glucocorticoids, such as cortisol. This has inhibitory effects (dashed lines) on CRH production, ACTH production, and leukocyte cytokine production. Figure is adapted from Malek et al., 2015.

Glucocorticoids are potent anti-inflammatories, acting by silencing expression of chemokines and other factors required for infiltration of tissues by circulating leukocytes (Cronstein *et al.*, 1992; Jahnsen *et al.*, 1999), and by downregulating inflammatory and upregulating anti-inflammatory cytokines (Franchimont, 2004). This suppression of inflammation, provoked by inflammation, is in turn negatively regulated by *FKBP5*. It is therefore likely that sustained activation of the HPA axis leads to increased cortisol production in sCJD patients, which in turn results in demethylation of the *FKBP5* promoter, upregulation of *FKBP5* and a suppression of glucocorticoid receptor activity.

Investigating this pathway in the context of sCJD is complicated by the breadth of inputs to the HPA axis and the levels at which it, and notably *FKBP5* (Zannas *et al.*, 2016), is regulated. However, that the sizable change in methylation at *FKBP5* does not associate with disease severity implies that dysregulation occurs in tandem with

pathogenesis. As such study of the axis's effects on pathogenesis would best be done in mouse models, where the effects of corticosterone or adrenaline-enriched diets or *Fkbp5* ablation on prion propagation and toxicity could reveal whether HPA axis activation is a protective response to, a downstream effect of or a contributing factor to prion pathology. Further investigation of HPA activity in patients, for example profiling of circulating ACTH or adrenaline levels, may clarify which elements of the pathway are being affected: do sCJD patients have decreased or increased levels of adrenaline? Systematic sampling of sCJD and control blood would also improve resolution, as while it is likely that blood from controls and patients was taken during "visiting hours" (10am-4pm), HPA axis activity and subsequent release of hormones fluctuate. Moreover, the circadian pattern of HPA axis activity is sensitive to disruption, with stress-induced HPA axis activation prompting dramatically greater elevations (602%) of circulating glucocorticoid levels during secretory phases of the ultradian cycle compared to nonsecretory phases in rats (Windle *et al.*, 1998). The PPIase function of *FKBP5* further complicates the picture. Does upregulation of this gene as a response to neuroinflammation increase the rate of PrP misfolding? Intriguingly, dysregulation of the HPA axis has also been consistently observed in Alzheimer's Disease, suggesting that further research in the context of both diseases may provide new insight into pathogenesis and identify therapeutic targets, biomarkers or risk factors (De Leon *et al.*, 1988; Breitner *et al.*, 1994; Green *et al.*, 2006).

7.11 Machine learning classification of sCJD status has utility and sets precedence for future work

Being able to train a classification model with an accuracy of 87.04% reinforces the validity of results from the 450K array analysis. However, the recursive feature elimination (RFE) identified probes overlap imperfectly with the Bonferroni-adjusted DMPs. This raises the question: as more useful predictive variables in sCJD/control classification, is DNA methylation at these probes rather than the DMPs a better molecular signature of sCJD status? Use of RFE to select targets for replication rather than manually selecting them as I did in the replication stage of the project may be a less biased approach and potentially result in the identification of biologically relevant candidates despite no prior association with neurodegeneration. It is also striking that the replicated DMPs in the promoter of *AIM2* (cg10636246, cg17515347), the body of *UHRF1* (cg17714703) and *METTL9* (cg02481950), and the three sites identified between the Houseman and EpiDISH algorithm in the promoter of *FKBP5* (cg03546163, and cg00052684 and

cg25114611 respectively) were all included in the list of 33 RFE-identified probes, as was promising candidate *MTRNR2L8* (cg05740793). It would be interesting to test RFE's ability to identify replicable candidate probes in future work.

The utility of this model in sCJD diagnosis is debatable. In clinical practice, it is not difficult to place an individual into the categories of "healthy" or "sCJD". A better model would be one capable of classifying sCJD/non-prion dementia. That said, this model's performance is a solid proof of concept and sets precedent for future work with the 850K EPIC array, perhaps considering classification discrimination between sCJD patient subgroups, sCJD and AD patients, or between sporadic, iatrogenic, inherited and variant prion diseases. An interesting phenomenon is the prevalence of PrP^{Sc} immunoreactivity in UK appendixes removed during routine appendectomies after exposure to the BSE epidemic (16 positive of 32441 studied), which can be extrapolated to estimate that 1 in 10,000 UK residents carries PrP^{Sc} in their gut (Noel Gill *et al.*, 2013). Yet intriguingly, appendixes removed before exposure to the BSE epidemic (1970-1979) show immunoreactivity at roughly the same levels, i.e. 1 in 7000 (Advisory Committee on Dangerous Pathogens TSE Subgroup, 2016). Because of the rarity of these specimens and the rarity of vCJD samples, machine learning based on methylation of appendix-derived DNA between non-immunoreactive individuals, immunoreactive individuals and individuals who developed vCJD could potentially shed light on differences by which subclinical carriers of prions resist infection from the periphery compared to vCJD patients. Indeed, while sporadic CJD is near impossible to anticipate, identification of susceptibility factors using machine learning in combination with other risk factors such as exposure to prion epidemics, contaminated medical products, or inheritance of a disease-causing mutation could greatly aid in diagnosis, prognosis, and ultimately improved quality of life.

7.12 Conclusions

Over the course of this work I identified differential methylation at several genes in blood-derived DNA from sCJD patients. These differences are replicable and specific to the sporadic form of human prion disease. Three genes in particular (*HOXA5*, *FKBP5*, *AIM2*) have neurotrophic properties and/or play roles in mediating inflammation and the immune response, functions known to mediate neurodegeneration. They thus serve as biomarkers of sporadic CJD. Hypomethylation at cg10636246 in *AIM2* correlates with MRC Scale score and can be considered a biomarker of disease severity. A preliminary attempt to use 450K array data for sCJD/control classification shows diagnostic potential, and further profiling of other patients from prion subgroups or other dementias may result in clinically useful models. While these experiments suggest that DNA methylation in sCJD may have translational potential, the most actionable outputs from this work are the hypotheses that emerge from the results discussed above. At a molecular level, the roles of *AIM2* and *FKBP5* in sporadic CJD and in the latter case prion pathogenesis specifically merit further explanation. At a physiological level, the melting of central pathology into the periphery shows promise not just for the discovery of biomarkers (be they cellular or molecular) but also suggests that dysregulation of the HPA axis, long suspected to predispose towards Alzheimer's Disease, may be a component or consequence of prion disease pathogenesis. Prion disease bioassays, both *in vitro* and *in vivo*, are sufficiently established to allow measurement of the effects of manipulation of the HPA axis in terms of prion infectivity and toxicity. While this has been predominantly a study of blood, it is exciting to think that some of the changes observed may not simply be useful as biomarkers but may also point towards therapeutic targets.

8. Bibliography

Aapola, U. *et al.* (2000) 'Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family', *Genomics*, 65(3), pp. 293–298. doi: 10.1006/geno.2000.6168.

Advisory Committee on Dangerous Pathogens TSE Subgroup (2016) *Updated position statement on occurrence of vCJD and prevalence of infection in the UK*. Available at: https://app.box.com/s/hhhhg857fjpu2bnxhv6e/file/91796156506.

Agency, E. M. (2013) *Public summary of opinion on orphan designation Decitabine for the treatment of acute myeloid leukaemia*. Available at: https://www.ema.europa.eu/documents/orphan-designation/eu/3/06/370-public-summary-positive-opinion-orphan-designation-decitabine-treatment-acute-myeloid-leukaemia_en.pdf.

Aguzzi, A., Barres, B. A. and Bennett, M. L. (2013) 'Microglia: Scapegoat, Saboteur, or Something Else?', *Science*, 339(6116), pp. 156–161. doi: 10.1126/science.1227901.Microglia.

Ahsan, M. *et al.* (2016) 'Epigenome-wide association study reveals differential DNA methylation in individuals with a history of myocardial infarction', *Human Molecular Genetics*, 25(21), p. ddw302. doi: 10.1093/hmg/ddw302.

Albu, S. *et al.* (2014) 'Deficiency of FK506-binding protein (FKBP) 51 alters sleep architecture and recovery sleep responses to stress in mice', *Journal of Sleep Research*, 23(2), pp. 176–185. doi: 10.1111/jsr.12112.

Alibhai, J. *et al.* (2016) 'Distribution of Misfolded Prion Protein Seeding Activity Alone Does Not Predict Regions of Neurodegeneration', *PLoS Biology*, 14(11), pp. 1–25. doi: 10.1371/journal.pbio.1002579.

Almasri, J. *et al.* (2015) 'Comparative Analysis of Azacitidine and Decitabine in Myelodysplastic Syndromes: A Systematic Review and Network Meta-Analysis', *Blood*, 126(23), pp. 1692 LP – 1692. Available at: http://www.bloodjournal.org/content/126/23/1692.abstract.

Amir, R. E. *et al.* (1999) 'Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2', *Nature Genetics*, 23(2), pp. 185–188. doi: 10.1038/13810.

Arányi, T. *et al.* (2006) 'The BiSearch web server', *BMC Bioinformatics*, 7, pp. 1–7.

doi: 10.1186/1471-2105-7-431.

Arita, K. *et al.* (2008) 'Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism', *Nature*, 455(7214), pp. 818–821. doi: 10.1038/nature07249.

Aryee, M. *et al.* (2019) *Package ' bumphunter '*. Available at: http://bioconductor.org/packages/release/bioc/manuals/bumphunter/man/bumphunter.pdf (Accessed: 4 March 2019).

Assenov, Y. *et al.* (2014) 'Comprehensive analysis of DNA methylation data with RnBeads.', *Nature methods*, 11(11), pp. 1138–40. doi: 10.1038/nmeth.3115.

Avvakumov, G. V. *et al.* (2008) 'Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1', *Nature*, 455(7214), pp. 822–825. doi: 10.1038/nature07273.

Bacalini, M. G., Boattini, A., *et al.* (2015) 'A meta-analysis on age-associated changes in blood DNA methylation: Results from an original analysis pipeline for Infinium 450k data', *Aging*, 7(2), pp. 97–109. doi: 10.18632/aging.100718.

Bacalini, M. G., Gentilini, D., *et al.* (2015) 'Identification of a DNA methylation signature in blood cells from persons with down syndrome', *Aging*, 7(2), pp. 82–96. doi: 10.1016/j.jss.2009.04.022.

Bailey, T. L. *et al.* (2009) 'MEME Suite: Tools for motif discovery and searching', *Nucleic Acids Research*, 37(SUPPL. 2), pp. 202–208. doi: 10.1093/nar/gkp335.

Bandiera, S. *et al.* (2011) 'Nuclear outsourcing of RNA interference components to human mitochondria', *PLoS ONE*, 6(6). doi: 10.1371/journal.pone.0020746.

Barrachina, M. and Ferrer, I. (2009) 'DNA methylation of Alzheimer disease and tauopathy-related genes in postmortem brain', *Journal of Neuropathology and Experimental Neurology*, 68(8), pp. 880–891. doi: 10.1097/NEN.0b013e3181af2e46.

Bates, E. E. M. *et al.* (2000) 'The mouse and human IGSF6 (DORA) genes map to the inflammatory bowel disease 1 locus and are embedded in an intron of a gene of unknown function', *Immunogenetics*, 52(1–2), pp. 112–120. doi: 10.1007/s002510000259.

Beck, E. *et al.* (1966) 'Experimental "Kuru" In Chimpanzees: A Pathological Report',

The Lancet, (1959), pp. 1056–1059. doi: 10.1016/S0140-6736(66)92031-9.

Beers, D. R. *et al.* (2017) 'ALS patients' regulatory T lymphocytes are dysfunctional, and correlate with disease progression rate and severity', *Jci*, 2(5), p. e89530. doi: 10.1172/jci.insight.89530.

Bellavance, M. A. and Rivest, S. (2014) 'The HPA - immune axis and the immunomodulatory actions of glucocorticoids in the brain', *Frontiers in Immunology*, 5(MAR), pp. 1–13. doi: 10.3389/fimmu.2014.00136.

de Bem, C. M. B. E. *et al.* (2016) 'The synergistic risk effect of apolipoprotein ε4 and DNA (cytosine-5-)-methyltransferase 3 beta (DNMT3B) haplotype for Alzheimer's disease', *Molecular Biology Reports*, 43(7), pp. 653–658. doi: 10.1007/s11033-016-3999-6.

Bestor, T. *et al.* (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', *Journal of Molecular Biology*, 203(4), pp. 971–983. doi: 10.1016/0022-2836(88)90122-2.

Bestor, T. H. and Verdine, G. L. (1994) 'DNA methyltransferases', *Current Opinion in Cell Biology*, 6(3), pp. 380–389. doi: 10.1016/0955-0674(94)90030-2.

Bhasin, J. M. and Ting, A. H. (2016) 'Goldmine integrates information placing genomic ranges into meaningful biological contexts', *Nucleic Acids Research*, 44(12), pp. 5550–5556. doi: 10.1093/nar/gkw477.

Bibikova, M. *et al.* (2006) 'High-throughput DNA methylation profiling using universal bead arrays', pp. 1–11. doi: 10.1101/gr.4410706.

Bibikova, M., Barnes, B., *et al.* (2011) 'High density DNA methylation array with single CpG site resolution', *Genomics*. Elsevier Inc., 98(4), pp. 288–295. doi: 10.1016/j.ygeno.2011.07.007.

Bibikova, M., Le, J., *et al.* (2011) 'High density DNA methylation array with single CpG site resolution', *Genomics*, 98, pp. 177–200.

Bird, a P. (1986) 'CpG-rich islands and the function of DNA methylation', *Nature*, 321(6067), pp. 209–213. doi: 10.1038/321209a0.

Blair, L. J. *et al.* (2013) 'Accelerated neurodegeneration through chaperone-mediated oligomerization of tau', *Journal of Clinical Investigation*, 123(10), pp.

4158–4169. doi: 10.1172/JCI69003.

Bockman, JM. , Kingsbury, DT., McKinley, MP., Bendheim, PE., Prusiner, S. (1985) 'Creutzfeldt-Jakob Disease prion proteins in human brains', *New England Journal of Medicine*, 312(2), pp. 73–78.

Bodzioch, M. *et al.* (2009) 'Evidence for potential functionality of nuclearly-encoded humanin isoforms', *Genomics*. Elsevier Inc., 94(4), pp. 247–256. doi: 10.1016/j.ygeno.2009.05.006.

Bostick, M. *et al.* (2007) 'UHRF1 plays a role in maintaining DNA methylation in mammalian cells.', *Science (New York, N.Y.)*, 317(5845), pp. 1760–4. doi: 10.1126/science.1147939.

Bourc'his, D. *et al.* (2001) 'Dnmt3L and the establishment of maternal genomic imprints', *Science*, 294(5551), pp. 2536–2539. doi: 10.1126/science.1065848.

Breeze, C. E. *et al.* (2016) 'eFORGE: A Tool for Identifying Cell Type-Specific Signal in Epigenomic Data', *Cell Reports*, 17(8), pp. 2137–2150. doi: 10.1016/j.celrep.2016.10.059.

Breiman, L. (2001) 'Random Forests', *Machine Learning*, 45, pp. 5–32.

Breitner, J. C. S. *et al.* (1994) 'Inverse association of anti-inflammatory treatments and Alzheimer's disease', *Neurology*, 44(2), pp. 227 LP – 227. doi: 10.1212/WNL.44.2.227.

Bremer, J. *et al.* (2010) 'Axonal prion protein is required for peripheral myelin maintenance', *Nature Neuroscience*. Nature Publishing Group, 13(3), pp. 310–318. doi: 10.1038/nn.2483.

Brown, A. R. *et al.* (2003) 'Inducible cytokine gene expression in the brain in the ME7/CV mouse model of scrapie is highly restricted, is at a strikingly low level relative to the degree of gliosis and occurs only late in disease', *Journal of General Virology*, 84(9), pp. 2605–2611. doi: 10.1099/vir.0.19137-0.

Brown, C. A. *et al.* (2014) 'In vitro screen of prion disease susceptibility genes using the scrapie cell assay', *Human molecular genetics*, 23(19), pp. 5102–5108. doi: 10.1093/hmg/ddu233.

Brown, P. *et al.* (2006) 'Iatrogenic Creutzfeldt – Jakob disease The waning of an era', *Neurology*, 67, pp. 389–393. doi: 10.1212/01.wnl.0000231528.65069.3f.

Brown, R. and Strathdee, G. (2015) ‘Epigenomics and epigenetic therapy of cancer’, *Trends in Molecular Medicine*. Elsevier, 8(4), pp. S43–S48. doi: 10.1016/S1471-4914(02)02314-6.

Bruce, M. E. *et al.* (1997) ‘Transmissions to mice indicate that “new variant” CJD is caused by the BSE agent’, *Nature*, 389(6650), pp. 498–501. doi: 10.1038/39057.

Bruegl, A. S. *et al.* (2014) ‘Utility of MLH1 methylation analysis in the clinical evaluation of Lynch Syndrome in women with endometrial cancer’, *Current pharmaceutical design*, 20(11), pp. 1655–1663. Available at: https://www.ncbi.nlm.nih.gov/pubmed/23888949.

Buske, F. A. *et al.* (2010) ‘Assigning roles to DNA regulatory motifs using comparative genomics’, *Bioinformatics*, 26(7), pp. 860–866. doi: 10.1093/bioinformatics/btq049.

Cai, X. *et al.* (2014) ‘Prion-like polymerization underlies signal transduction in antiviral immune defense and inflammasome activation’, *Cell*. Elsevier Inc., 156(6), pp. 1207–1222. doi: 10.1016/j.cell.2014.01.063.

Campbell, I. L. *et al.* (1994) ‘Activation of cerebral cytokine gene expression and its correlation with onset of reactive astrocyte and acute-phase response gene expression in scrapie.’, *Journal of virology*, 68(4), pp. 2383–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=236715&tool=pmcentrez&rendertype=abstract.

Capper, D. *et al.* (2018) ‘DNA methylation-based classification of central nervous system tumours’, *Nature*. 2018/03/14, 555(7697), pp. 469–474. doi: 10.1038/nature26000.

Chahrour, M. *et al.* (2008) ‘MeCP2, a Key Contributor to Neurological Disease, Activates and Represses Transcription’, 320(5880), pp. 1224–1229. doi: 10.1126/science.1153252.

Chakravarthy, A. *et al.* (2018) ‘Pan-cancer deconvolution of tumour composition using DNA methylation’, *Nature Communications*, 9(1). doi: 10.1016/j.matchar.2018.05.010.

Chang, L. *et al.* (2014) ‘Elevation of Peripheral BDNF Promoter Methylation Links to the Risk of Alzheimer’s Disease’, *PLoS ONE*. Edited by L. Chiariotti. San Francisco, USA: Public Library of Science, 9(11), p. e110773. doi:

10.1371/journal.pone.0110773.

Chen, C. M. *et al.* (2003) 'Methylation target array for rapid analysis of CpG island hypermethylation in multiple tissue genomes', *American Journal of Pathology*. American Society for Investigative Pathology, 163(1), pp. 37–45. doi: 10.1016/S0002-9440(10)63628-0.

Chen, E. Y. *et al.* (2013) 'Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool', *BMC Bioinformatics*, 14. doi: 10.1186/1471-2105-14-128.

Chen, Y. A. *et al.* (2013) 'Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray', *Epigenetics*, 8(2), pp. 203–209. doi: 10.4161/epi.23470.

Clark, C. N. and Warren, J. D. (2013) 'A Hypnic Hypothesis of Alzheimer's Disease', *Neuro-Degenerative Diseases*. Allschwilerstrasse 10, P.O. Box · Postfach · Case postale, CH–4009, Basel, Switzerland · Schweiz · Suisse, Phone: +41 61 306 11 11, Fax: +41 61 306 12 34, karger@karger.ch: S. Karger AG, 12(4), pp. 165–176. doi: 10.1159/000350060.

Collinge, J. *et al.* (1996) 'Molecular analysis of prion strain variation and the aetiology of "new variant" CJD', *Nature*, pp. 685–690. doi: 10.1038/383685a0.

Collinge, J., Palmer, M. S. and Dryden, A. J. (1991) 'Genetic predisposition to iatrogenic Creutzfeldt-Jakob disease', *The Lancet*, 337(8755), pp. 1441–1442. doi: 10.1016/0140-6736(91)93128-V.

Comb, M. and Goodman, H. M. (1990) 'CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2', *Nucleic Acids Research*, 18(13), pp. 3975–3982. doi: 10.1093/nar/18.13.3975.

Concha-Marambio, L. *et al.* (2016) 'Detection of prions in blood from patients with variant Creutzfeldt-Jakob disease', *Science Translational Medicine*, 8(370), pp. 1–8. doi: 10.1126/scitranslmed.aaf6188.

Cooper, D. N. and Youssoufian, H. (1988) 'The CpG dinucleotide and human genetic disease.', *Human genetics*, 78(2), pp. 151–155. doi: 10.1007/BF00278187.

Corp., I. (2017) 'IBM SPSS Statistics for Windows'. Armonk, NY.

Cox, G. *et al.* (1995) 'Glucocorticoid treatment inhibits apoptosis in human neutrophils. Separation of survival and activation outcomes.', *Journal of immunology*

(Baltimore, Md. : 1950), 154(9), pp. 4719–25. doi: 10.4049/jimmunol.181.6.4089.

Creutzfeldt, H. G. (1920) 'Über eine eigenartige herdförmige erkrankung des zentralnervensystems (Vorläufige mitteilung)', *Zeitschrift für die gesamte Neurologie und Psychiatrie*, 57(1), pp. 1–18. doi: 10.1007/BF02866081.

Cronstein, B. N. *et al.* (1992) 'A mechanism for the antiinflammatory effects of corticosteroids: the glucocorticoid receptor regulates leukocyte adhesion to endothelial cells and expression of endothelial-leukocyte adhesion molecule 1 and intercellular adhesion molecule 1', *Proceedings of the National Academy of Sciences of the United States of America*, 89(21), pp. 9991–9995. Available at: https://www.ncbi.nlm.nih.gov/pubmed/1279685.

Cross, S. H. *et al.* (1997) 'A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins', 16(July), pp. 256–259.

Cruz, J. A. and Wishart, D. S. (2006) 'Applications of machine learning in cancer prediction and prognosis', *Cancer Informatics*, 2, pp. 59–77. doi: 10.1177/117693510600200030.

Cuillé, J. and Chelle, P. L. (1936) 'La malade dite tremblante du mouton: est-elle inoculable?', *Comptes Rendus, Académie des Sciences*, 203, pp. 1552–1554.

Dalai, W. *et al.* (2017) 'CpG site DNA methylation patterns reveal a novel regulatory element in the mouse prion protein gene', *Journal of Veterinary Medical Science*, 79(1), pp. 100–107. doi: 10.1292/jvms.16-0390.

Dedeurwaerder, S. *et al.* (2011) 'Evaluation of the Infinium Methylation 450K technology', *Epigenomics*, 3(6), pp. 771–784.

Dejesus-hernandez, M. *et al.* (2011) 'Expanded GGGGCC hexanucleotide repeat in non-coding region of C9ORF72 causes chromosome 9p-linked frontotemporal dementia and amyotrophic lateral sclerosis', *Neuron*, 72(2), pp. 245–256. doi: 10.1016/j.neuron.2011.09.011.Expanded.

Denes, A. *et al.* (2015) 'AIM2 and NLRC4 inflammasomes contribute with ASC to acute brain injury independently of NLRP3.', *Proceedings of the National Academy of Sciences of the United States of America*, 112(13), pp. 4050–5. doi: 10.1073/pnas.1419090112.

Desplats, P. *et al.* (2011) ‘α-synuclein sequesters Dnmt1 from the nucleus: A novel mechanism for epigenetic alterations in Lewy body diseases’, *Journal of Biological Chemistry*, 286(11), pp. 9031–9037. doi: 10.1074/jbc.C110.212589.

Devlin, B. and Roeder, K. (1999) ‘Genomic Control for Association’, *International Biometric Society*, 55(4), pp. 997–1004.

Dick, M. S. *et al.* (2016) ‘ASC filament formation serves as a signal amplification mechanism for inflammasomes’, *Nature Communications*, 7(May). doi: 10.1038/ncomms11929.

Dombroski, B. *et al.* (1991) ‘Isolation of an active human tranposable element’, *Science*, 254, pp. 1805–1808.

Douet, J. Y. *et al.* (2014) ‘Detection of infectivity in blood of persons with variant and sporadic Creutzfeldt-Jakob disease’, *Emerging Infectious Diseases*, 20(1), pp. 114–117. doi: 10.3201/eid2001.130353.

Du, P. *et al.* (2010) ‘Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis’, *BMC Bioinformatics*. BioMed Central, 11, p. 587. doi: 10.1186/1471-2105-11-587.

Eckhardt, F. *et al.* (2011) ‘DNA methylation profiling of human chromosomes 6, 20 and 22: author manuscript’, *Nature genetics*, 38(12), pp. 1378–1385. doi: 10.1038/ng1909.DNA.

Ehrlich, M. *et al.* (1982) ‘Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells’, *Nucleic Acids Res*, 10(8), pp. 11–14.

Estey, E. H. (2013) ‘Epigenetics in clinical practice: The examples of azacitidine and decitabine in myelodysplasia and acute myeloid leukemia’, *Leukemia*. Nature Publishing Group, 27(9), pp. 1803–1812. doi: 10.1038/leu.2013.173.

Evans-Galea, M. V *et al.* (2012) ‘FXN methylation predicts expression and clinical outcome in Friedreich ataxia’, *Annals of Neurology*. Wiley Subscription Services, Inc., A Wiley Company, 71(4), pp. 487–497. doi: 10.1002/ana.22671.

Ewald, E. R. *et al.* (2014) ‘Alterations in DNA methylation of Fkbp5 as a determinant of blood-brain correlation of glucocorticoid exposure’, *Psychoneuroendocrinology*, 44, pp. 112–122. doi: 10.1016/j.psyneuen.2014.03.003.

Exarchos, T. P. *et al.* (2014) ‘Machine learning applications in cancer prognosis and

prediction', *Computational and Structural Biotechnology Journal*. Elsevier B.V., 13, pp. 8–17. doi: 10.1016/j.csbj.2014.11.005.

Faul, F. *et al.* (2009) 'Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses', *Behavior Research Methods*, 41(4), pp. 1149–1160. doi: 10.3758/BRM.41.4.1149.

van der Flier, W. M. *et al.* (2011) 'Early-onset versus late-onset Alzheimer's disease: The case of the missing APOE ε4 allele', *The Lancet Neurology*. Elsevier Ltd, 10(3), pp. 280–288. doi: 10.1016/S1474-4422(10)70306-9.

Florath, I. *et al.* (2014) 'Cross-sectional and longitudinal changes in DNA methylation with age: An epigenome-wide analysis revealing over 60 novel age-associated CpG sites', *Human Molecular Genetics*, 23(5), pp. 1186–1201. doi: 10.1093/hmg/ddt531.

Franchimont, D. (2004) 'Overview of the actions of glucocorticoids on the immune response: A good model to characterize new pathways of immunosuppression for new treatment strategies', *Annals of the New York Academy of Sciences*, 1024, pp. 124–137. doi: 10.1196/annals.1321.009.

Frank-Cannon, T. C. *et al.* (2009) 'Does neuroinflammation fan the flame in neurodegenerative diseases?', *Molecular Neurodegeneration*, 4(1), pp. 1–13. doi: 10.1186/1750-1326-4-47.

Fridman, W. H. *et al.* (2017) 'The immune contexture in cancer prognosis and treatment', *Nature Reviews Clinical Oncology*, 14(12), pp. 717–734. doi: 10.1038/nrclinonc.2017.101.

Friedman, S. (1981) 'The inhibition of DNA (cytosine-5) methylases by 5-azacytidine', *Molecular pharmacology*, 19, pp. 314–320.

Frommer, M. *et al.* (1992) 'A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands.', *Proceedings of the National Academy of Sciences*, 89(5), pp. 1827–1831. doi: 10.1073/pnas.89.5.1827.

Fujita, H. *et al.* (2003) 'Antithetic Effects of MBD2a on Gene Regulation', *Molecular and Cellular Biology*, 23(8), pp. 2645–2657. doi: 10.1128/mcb.23.8.2645-2657.2003.

Fuso, A. *et al.* (2005) 'S-adenosylmethionine/homocysteine cycle alterations modify

DNA methylation status with consequent deregulation of PS1 and BACE and beta-amyloid production', *Molecular and Cellular Neuroscience*, 28(1), pp. 195–204. doi: 10.1016/j.mcn.2004.09.007.

Gajdusek, D. C. and Zigas, V. (1957) 'Degenerative disease of the central nervous system in new guinea', *NEJM*, 257(20), pp. 974–978.

Gambetti, P. *et al.* (2003) 'Sporadic and familial CJD: classification and characterisation', *British Medical Bulletin* , 66(1), pp. 213–239. doi: 10.1093/bmb/66.1.213.

Gardiner-Garden, M. and Frommer, M. (1987) 'CpG Islands in vertebrate genomes', *Journal of Molecular Biology*, 196(2), pp. 261–282. doi: 10.1016/0022-2836(87)90689-9.

Gayrard, V. *et al.* (2000) 'Major hypercorticism is an endocrine feature of ewes with naturally occurring scrapie', *Endocrinology*, 141(3), pp. 988–994. doi: 10.1210/en.141.3.988.

Geering, B. *et al.* (2013) 'Living and dying for inflammation: Neutrophils, eosinophils, basophils', *Trends in Immunology*, 34(8), pp. 398–409. doi: 10.1016/j.it.2013.04.002.

Gibbs Jr., C. J. *et al.* (1968) 'Creutzfeldt-Jakob Disease (Spongiform Encephalopathy): Transmission to the Chimpanzee', *Science*, 161(10).

Gillies, G. E. *et al.* (2014) 'Sex differences in Parkinson's disease', *Frontiers in Neuroendocrinology*. Elsevier Inc., 35(3), pp. 370–384. doi: 10.1016/j.yfrne.2014.02.002.

Green, K. N. *et al.* (2006) 'Glucocorticoids Increase Amyloid-beta and Tau Pathology in a Mouse Model of Alzheimer's Disease', *Journal of Neuroscience*, 26(35), pp. 9047–9056. doi: 10.1523/JNEUROSCI.2797-06.2006.

Griffith, J. S. (1967) 'Nature of the Scrapie Agent: Self-replication and Scrapie', *Nature*, 215(5105), pp. 1043–1044. doi: 10.1038/2151043a0.

Gründemann, J. *et al.* (2008) 'Elevated α-synuclein mRNA levels in individual UV-laser-microdissected dopaminergic substantia nigra neurons in idiopathic Parkinson's disease', *Nucleic Acids Research*, 36(7), pp. 1–16. doi: 10.1093/nar/gkn084.

Le Guezennec, X. *et al.* (2005) 'MBD2/NuRD and MBD3/NuRD, Two Distinct Complexes with Different', 26(3), pp. 843–851. doi: 10.1128/MCB.26.3.843.

Guidotti, G. *et al.* (2013) 'Glucocorticoid Receptor and FKBP5 Expression Is Altered Following Exposure to Chronic Stress: Modulation by Antidepressant Treatment', *Neuropsychopharmacology*. Nature Publishing Group, 38(4), pp. 616–627. doi: 10.1038/npp.2012.225.

Han, T. *et al.* (2018) 'Comparison of random forest , artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery', 40(8), pp. 2681–2693. doi: 10.1177/0142331217708242.

Harrington, B. et. al. (2005) 'Inkscape'. Available at: http://www.inkscape.org/.

Hashimoto, H. *et al.* (2008) 'The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix', *Nature*, 455(7214), pp. 826–829. doi: 10.1038/nature07280.

Hashimoto, H. *et al.* (2012) 'Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation', *Nucleic Acids Research*, 40(11), pp. 4841–4849. doi: 10.1093/nar/gks155.

Hashimoto, Y. *et al.* (2001) 'A rescue factor abolishing neuronal cell death by a wide spectrum of familial Alzheimer's disease genes and A ', *Proceedings of the National Academy of Sciences*, 98(11), pp. 6336–6341. doi: 10.1073/pnas.101133498.

Hata, K. *et al.* (2002) 'Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice.', *Development (Cambridge, England)*, 129, pp. 1983–1993.

He, Y.-F. *et al.* (2011) 'Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA', *Science*, 333(6047), pp. 1303 LP – 1307. Available at: http://science.sciencemag.org/content/333/6047/1303.abstract.

Hellman, A. and Chess, A. (2007) 'Gene Body – Specific Methylation on the Active X Chromosome', *Science*, 315(5815), pp. 1141–1143. doi: 10.1126/science.1136352.

Hendrich, B. *et al.* (1999) 'The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites', *Nature*, 401(6750), pp. 301–304. doi: 10.1038/45843.

Hendrich, B. *et al.* (2001) 'Closely related proteins MBD2 and MBD3 play distinctive

but interacting roles in mouse development', *Genes and Development*, 15(6), pp. 710–723. doi: 10.1101/gad.194101.

Hendrich, B. and Bird, A. (2015) 'Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins', *Molecular and Cellular Biology*, 18(11), pp. 6538–6547. doi: 10.1128/mcb.18.11.6538.

Hill, A. F. *et al.* (1997) 'The same prion strain causes vCJD and BSE', *Nature*. Macmillan Magazines Ltd., 389(6650), pp. 448–450. Available at: http://dx.doi.org/10.1038/38925.

Hill, A. F. *et al.* (2003) 'Molecular classification of sporadic Creutzfeldt-Jakob disease', *Brain*, 126(6), pp. 1333–1346. doi: 10.1093/brain/awg125.

Hinrichs, A. L. *et al.* (2012) 'NIH Public Access', *Changes*, 29(6), pp. 997–1003. doi: 10.1016/j.biotechadv.2011.08.021.Secreted.

Ho, T. K. (1995) 'Random Decision Forests', in *Proceedings of 3rd International Conference on Document Analysis and Recognition*.

Hollenbach, P. W. *et al.* (2010) 'A comparison of azacitidine and decitabine activities in acute myeloid leukemia cell lines', *PLoS ONE*, 5(2). doi: 10.1371/journal.pone.0009001.

Hornung, V. *et al.* (2009) 'AIM2 recognizes cytosolic dsDNA and forms a caspase-1 activating inflammasome with ASC', *Nature*, 458(7237), pp. 514–518. doi: 10.1038/nature07725.AIM2.

Hoss, A. G. *et al.* (2014) 'MicroRNAs Located in the Hox Gene Clusters Are Implicated in Huntington's Disease Pathogenesis', *PLoS Genetics*, 10(2). doi: 10.1371/journal.pgen.1004188.

Houseman, E. A. *et al.* (2012) 'DNA methylation arrays as surrogate measures of cell mixture distribution.', *BMC bioinformatics*, 13(1), p. 86. doi: 10.1186/1471-2105-13-86.

Houseman, E. A. *et al.* (2016) 'Reference-free deconvolution of DNA methylation data and mediation by cell composition effects', *BMC Bioinformatics*. BMC Bioinformatics, 17(1), pp. 1–15. doi: 10.1186/s12859-016-1140-4.

Houseman, E. A., Molitor, J. and Marsit, C. J. (2014) 'Reference-free cell mixture adjustments in analysis of DNA methylation data', *Bioinformatics*, 30(10), pp. 1431–

1439. doi: 10.1093/bioinformatics/btu029.

Hu, Y. *et al.* (2017) 'Increased peripheral blood inflammatory cytokine levels in amyotrophic lateral sclerosis: A meta-analysis study', *Scientific Reports*, 7(1), pp. 12–15. doi: 10.1038/s41598-017-09097-1.

Huang, T. H. M. *et al.* (1997) 'Identification of DNA methylation markers for human breast carcinomas using the methylation-sensitive restriction fingerprinting technique', *Cancer Research*, 57(6), pp. 1030–1034.

Iguchi-Ariga, S. M. and Schaffner, W. (1989) 'CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation.', *Genes & development*, 3(5), pp. 612–619. doi: 10.1101/gad.3.5.612.

Illumina (2012) *Infinium Human Methylation 450 data sheet*. Available at: https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_humanmethylation450.pdf.

Illumina, I. (2015) 'Infnium® HD Assay Methylation Protocol Guide', *ILLUMINA PROPRIETARY Material # 20002138 Document # 15019519 v01 November 2015*, (November). Available at: https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/infinium_assays/infinium_hd_methylation/infinium-hd-methylation-guide-15019519-01.pdf.

Iterson, M. Van, Zwet, E. W. Van and Heijmans, B. T. (2017) 'Controlling bias and inflation in association studies using the empirical null distribution', *Genome Biology*. Genome Biology, pp. 1–13. doi: 10.1186/s13059-016-1131-9.

Jaffe, A. E. *et al.* (2012) 'Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies', *International Journal of Epidemiology*, 41(1), pp. 200–209. doi: 10.1093/ije/dyr238.

Jaffe, A. E. and Irizarry, R. A. (2014) 'Accounting for cellular heterogeneity is critical in epigenome-wide association studies', *Genome Biology*, 15(2), pp. 1–9. doi: 10.1186/gb-2014-15-2-r31.

De Jager, P. L. *et al.* (2014) 'Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci.', *Nature neuroscience*. Nature Publishing Group, 17(9), pp. 1156–63. doi: 10.1038/nn.3786.

Jahnsen, F. L. *et al.* (1999) 'Glucocorticosteroids inhibit mRNA expression for eotaxin, eotaxin-2, and monocyte-chemotactic protein-4 in human airway inflammation with eosinophilia.', *Journal of immunology (Baltimore, Md. : 1950)*, 163(3), pp. 1545–51. doi: ji_v163n3p1545 [pii].

Jarmasz, J. S. *et al.* (2019) 'DNA methylation and histone post-translational modification stability in post-mortem brain tissue', *Clinical Epigenetics*. Clinical Epigenetics, 11(1), pp. 1–23. doi: 10.1186/s13148-018-0596-7.

Jeannotte, L., Gotti, F. and Landry-Truchon, K. (2016) 'Hoxa5: A Key Player in Development and Disease', *Journal of Developmental Biology*, 4(2), p. 13. doi: 10.3390/jdb4020013.

Jiang, C.-L. *et al.* (2005) 'MBD3L2 Interacts with MBD3 and Components of the NuRD Complex and Can Oppose MBD2-MeCP1-mediated Methylation Silencing', *Journal of Biological Chemistry*, 280(13), pp. 12700–12709. doi: 10.1074/jbc.m413492200.

Jiang, C. L., Jin, S. G. and Pfeifer, G. P. (2004) 'MBD3L1 is a transcriptional repressor that interacts with methyl-CpG-binding protein 2 (MBD2) and components of the NuRD complex', *Journal of Biological Chemistry*, 279(50), pp. 52456–52464. doi: 10.1074/jbc.M409149200.

Jiao, C. *et al.* (2018) 'Positional effects revealed in Illumina methylation array and the impact on analysis', *Epigenomics*, 10(5), pp. 643–659. doi: 10.2217/epi-2017-0105.

Joehanes, R. *et al.* (2016) 'Epigenetic Signatures of Cigarette Smoking', *Circulation: Cardiovascular Genetics*, p. CIRCGENETICS.116.001506. doi: 10.1161/CIRCGENETICS.116.001506.

Johnson, W. E., Li, C. and Rabinovic, A. (2007) 'Adjusting batch effects in microarray expression data using empirical Bayes methods', *Biostatistics*, 8(1), pp. 118–127. doi: 10.1093/biostatistics/kxj037.

Jones, P. A. (2012) 'Functions of DNA methylation: islands, start sites, gene bodies and beyond.', *Nature reviews. Genetics*. Nature Publishing Group, 13(7), pp. 484–92. doi: 10.1038/nrg3230.

Jowaed, A. *et al.* (2010) 'Methylation Regulates Alpha-Synuclein Expression and Is Decreased in Parkinson's Disease Patients' Brains', *Journal of Neuroscience*,

30(18), pp. 6355–6359. doi: 10.1523/JNEUROSCI.6119-09.2010.

Kalbe, E. *et al.* (2004) 'DemTect: A new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia.', *International Journal of Geriatric Psychiatry*, 19(2), pp. 136–143. doi: 10.1002/gps.1042.

KAMINSKAS, E., FARRELL, A. T. and WANG, YONG-CHENG SRIDHARA, RAJESHWARI PAZDUR, R. (2005) 'FDA Drug Approval Summary: Azacitidine (5-azacytidine, Vidaza™) for Injectable Suspension', *The Oncologist*, 10, pp. 176–182. doi: 10.1634/theoncolo-gist.2007-0248.

Karapetyan, Y. E. *et al.* (2013) 'Unique drug screening approach for prion diseases identifies tacrolimus and astemizole as antiprion agents', *Proceedings of the National Academy of Sciences*, 110(17), pp. 7044–7049. doi: 10.1073/pnas.1303510110.

Karimi, M. *et al.* (2006) 'LUMA (LUminometric Methylation Assay)-A high throughput method to the analysis of genomic DNA methylation', *Experimental Cell Research*, 312(11), pp. 1989–1995. doi: 10.1016/j.yexcr.2006.03.006.

Katscher, F. (1998) 'It's Jakob's disease, not Creutzfeldt's', *Nature*, 393(May), p. 11.

Kim, Jae Il *et al.* (1999) 'Expression of cytokine genes and increased nuclear factor-kappa B activity in the brains of scrapie-infected mice', *Molecular Brain Research*, 73(1–2), pp. 17–27. doi: 10.1016/S0169-328X(99)00229-6.

Kimberlin, R. H. (1982) 'Scrapie agent: prions or virinos?', *Nature*, pp. 107–108. doi: 10.1038/297107a0.

King, E. *et al.* (2018) 'Peripheral inflammation in prodromal Alzheimer's and Lewy body dementias', *Journal of Neurology, Neurosurgery and Psychiatry*, 89(4), pp. 339–345. doi: 10.1136/jnnp-2017-317134.

Kishi, N. and Macklis, J. D. (2004) 'MECP2 is progressively expressed in post-migratory neurons and is involved in neuronal maturation rather than cell fate decisions', *Molecular and Cellular Neuroscience*, 27(3), pp. 306–321. doi: 10.1016/j.mcn.2004.07.006.

Klein, M. A. *et al.* (2004) 'Accelerated Prion Disease in the Absence of Interleukin-10', *Journal of Virology*, 78(24), pp. 13697–13707. doi: 10.1128/jvi.78.24.13697-

13707.2004.

de Koning, A. P. J. *et al.* (2011) ‘Repetitive elements may comprise over Two-Thirds of the human genome’, *PLoS Genetics*, 7(12). doi: 10.1371/journal.pgen.1002384.

De Koning, H. D. *et al.* (2014) ‘Absent in Melanoma 2 is predominantly present in primary melanoma and primary squamous cell carcinoma, but largely absent in metastases of both tumors’, *Journal of the American Academy of Dermatology*, pp. 1012–1015. doi: 10.1016/j.jaad.2014.06.012.

Kraus, A. *et al.* (2015) ‘Prion protein prolines 102 and 105 and the surrounding lysine cluster impede amyloid formation’, *Journal of Biological Chemistry*, 290(35), pp. 21510–21522. doi: 10.1074/jbc.M115.665844.

Kretzschmar, H. A., Neumann, M. and Stavrou, D. (1995) ‘Codon 178 mutation of the human prion protein gene in a German family (Backer family): Sequencing data from 72-year-old celloidin-embedded brain tissue’, *Acta Neuropathologica*, 89(1), pp. 96–98. doi: 10.1007/s004010050220.

Krogan, N. J. *et al.* (2003) ‘Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II’, *Molecular and cellular biology*. American Society for Microbiology, 23(12), pp. 4207–4218. doi: 10.1128/MCB.23.12.4207-4218.2003.

Krumlauf, R. and Hill, M. (1994) ‘Hox Genes in Vertebrate Development Review’, 78, pp. 191–201.

Kuhn, M. (2008) ‘Building Predictive Models in R Using the caret Package’, *Journal Of Statistical Software*, 28(5), pp. 1–26. doi: 10.1053/j.sodo.2009.03.002.

Kuleshov, M. V. *et al.* (2016) ‘Enrichr: a comprehensive gene set enrichment analysis web server 2016 update’, *Nucleic acids research*, 44(W1), pp. W90–W97. doi: 10.1093/nar/gkw377.

Kumar, S. *et al.* (1994) ‘The DNA (cytosine-5) methyltransferases’, 22(1), pp. 1–10.

Kuo, K. C. *et al.* (1980) ‘Quantitative reversed-phase high performance liquid chromatographic determination of major and modified deoxyribonucleosides in DNA.’, *Nucleic acids research*, 8(20), pp. 4763–76. doi: 10.1093/nar/8.20.4763.

Kurdyukov, S. and Bullock, M. (2016) ‘DNA Methylation Analysis: Choosing the

Right Method', *Biology*, 5(1), p. 3. doi: 10.3390/biology5010003.

Lachmann, A. *et al.* (2010) 'ChEA: Transcription factor regulation inferred from integrating genome-wide ChIP-X experiments', *Bioinformatics*, 26(19), pp. 2438–2444. doi: 10.1093/bioinformatics/btq466.

Lai, K. S. P. *et al.* (2017) 'Peripheral inflammatory markers in Alzheimer's disease: A systematic review and meta-analysis of 175 studies', *Journal of Neurology, Neurosurgery and Psychiatry*, 88(10), pp. 876–882. doi: 10.1136/jnnp-2017-316201.

Lam, L. *et al.* (2016) 'Epigenetic changes in T-cell and monocyte signatures and production of neurotoxic cytokines in ALS patients', *The FASEB Journal*. Bethesda, MD, USA: Federation of American Societies for Experimental Biology, 30(10), pp. 3461–3473. doi: 10.1096/fj.201600259RR.

Larsson, S. C. *et al.* (2017) 'Modifiable pathways in Alzheimer's disease: Mendelian randomisation analysis', *BMJ (Clinical research ed.)*, 359, p. j5375. doi: 10.1136/bmj.j5375.

Latz, E., Ts, X. and Stutz, A. (2013) 'Activation and regulation of the inflammasomes', *Nature Reviews Immunology*, 13(6), pp. 1–30. doi: 10.1038/nri3452.Activation.

Laurén, J. *et al.* (2009) 'Cellular prion protein mediates impairment of synaptic plasticity by amyloid-β oligomers', *Nature*, 457(7233), pp. 1128–1132. doi: 10.1038/nature07761.

Lee, A. and Gilbert, R. M. (2016) 'Epidemiology of Parkinson Disease', *Neurologic Clinics*. Elsevier Inc, 34(4), pp. 955–965. doi: 10.1016/j.ncl.2016.06.012.

Lee, C., Yen, K. and Cohen, P. (2013) 'Humanin: A harbinger of mitochondrial-derived peptides?', *Trends in Endocrinology and Metabolism*, 24(5), pp. 222–228. doi: 10.1016/j.tem.2013.01.005.

Lei, H. *et al.* (1996) 'De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells', 3205, pp. 3195–3205.

De Leon, M. J. *et al.* (1988) 'Abnormal Cortisol Response in Alzheimer'S Disease Linked To Hippocampal Atrophy', *The Lancet*, 332(8607), pp. 391–392. doi: 10.1016/S0140-6736(88)92855-3.

Lewis, J. D. *et al.* (1992) 'Purification, sequence, and cellular localization of a novel

chromosomal protein that binds to Methylated DNA', *Cell*, 69(6), pp. 905–914. doi: 10.1016/0092-8674(92)90610-O.

Li, B. *et al.* (2003) 'The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II', *Journal of Biological Chemistry*, 278(11), pp. 8897–8903. doi: 10.1074/jbc.M212134200.

Li, Y. *et al.* (2014) 'An Epigenetic Signature in Peripheral Blood Associated with the Haplotype on 17q21.31, a Risk Factor for Neurodegenerative Tauopathy', *PLoS Genetics*. Edited by G. P. Copenhaver. San Francisco, USA: Public Library of Science, 10(3), p. e1004211. doi: 10.1371/journal.pgen.1004211.

Lin, J. C. *et al.* (2007) 'Role of Nucleosomal Occupancy in the Epigenetic Silencing of the MLH1 CpG Island', *Cancer cell*, 12(5), pp. 432–444. doi: 10.1016/j.ccr.2007.10.014.

Liu, E. Y., Cali, C. P. and Lee, E. B. (2017) 'RNA metabolism in neurodegenerative disease', *Disease Models & Mechanisms*, 10(5), pp. 509–518. doi: 10.1242/dmm.028613.

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

Llorens, F. *et al.* (2013) 'PrP mRNa and protein expression in brain and PrPcin CSF in Creutzfeldt-Jakob disease MM1 and VV2', *Prion*, 7(5), pp. 383–393. doi: 10.4161/pri.26416.

Lloyd, S., Mead, S. and Collinge, J. (2011) 'Genetics of Prion Disease', in Tatzelt, J. (ed.) *Prion Proteins SE - 157*. Springer Berlin Heidelberg (Topics in Current Chemistry), pp. 1–22. doi: 10.1007/128_2011_157.

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

Luchsinger, J. A. and Gustafson, D. R. (2009) 'Adiposity and Alzheimer's disease', *Current Opinion in Clinical Nutrition and Metabolic Care*, 12(1), pp. 15–21. doi: 10.1097/MCO.0b013e32831c8c71.

Luco, R. F. *et al.* (2010) 'Regulation of Alternative Splicing by Histone Modifications', *Science*, 327, pp. 996–1001. doi: 10.1126/science.1184208.

Lue, L. F. *et al.* (1996) 'Inflammation, Aβ deposition, and neurofibrillary tangle formation as correlates of Alzheimer's disease neurodegeneration', *Journal of Neuropathology and Experimental Neurology*, 55(10), pp. 1083–1088. doi: 10.1097/00005072-199655100-00008.

Lunnon, K. *et al.* (2014) 'Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease', *Nat Neurosci*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 17(9), pp. 1164–1170. Available at: http://dx.doi.org/10.1038/nn.3782.

Luo, Y., Lu, X. and Xie, H. (2014) 'Dynamic Alu Methylation during Normal Development, Aging, and Tumorigenesis', *BioMed Research International*. Hindawi Publishing Corporation, 2014. doi: 10.1155/2014/784706.

Maksimovic, J., Gordon, L. and Oshlack, A. (2012) 'SWAN: Subset-quantile Within Array Normalization for Illumina Infinium HumanMethylation450 BeadChips', *Genome Biology*, 13(6), p. R44. doi: 10.1186/gb-2012-13-6-r44.

Malek, H. *et al.* (2015) 'Dynamics of the HPA axis and inflammatory cytokines: Insights from mathematical modeling', *Computers in Biology and Medicine*, 67, pp. 1–12. doi: 10.1016/j.compbiomed.2015.09.018.

Marina, R. J. *et al.* (2015) 'TET-catalyzed oxidation of intragenic 5-methylcytosine regulates CTCF-dependent alternative splicing', *The EMBO Journal*, 35(3), pp. 1–21. doi: 10.15252/embj.

Marttila, S. *et al.* (2015) 'Ageing-associated changes in the human DNA methylome: Genomic locations and effects on gene expression', *BMC Genomics*, 16(1), pp. 1–17. doi: 10.1186/s12864-015-1381-z.

Masliah, E. *et al.* (2013) 'Distinctive patterns of DNA methylation associated with Parkinson disease: identification of concordant epigenetic changes in brain and peripheral blood leukocytes.', *Epigenetics*. Taylor & Francis, 8(10), pp. 1030–8. doi: 10.4161/epi.25865.

Mastrianni, J. *et al.* (2003) *Genetic Prion Diseases*, *GeneReviews*®. Available at: https://www.ncbi.nlm.nih.gov/books/NBK1229/ (Accessed: 19 April 2018).

Mastroeni, D. *et al.* (2017) 'ANK1 is up-regulated in laser captured microglia in Alzheimer's brain; the importance of addressing cellular heterogeneity', *PLoS ONE*, 12(7), pp. 1–11. doi: 10.1371/journal.pone.0177814.

Matsumoto, L. *et al.* (2010) ‘CpG demethylation enhances alpha-synuclein expression and affects the pathogenesis of Parkinson’s disease’, *PLoS ONE*, 5(11), pp. 1–9. doi: 10.1371/journal.pone.0015522.

McDonald, A. J. and Millhauser, G. L. (2014) ‘PrP overdrive :Does inhibition of a-Cleavage contribute to PrPC toxicity and prion disease?’, *Prion*, 8(2), pp. 1–9. doi: 10.4161/pri.28796.

Mcgill, R., Tukey, J. W. and Larsen, W. A. (1978) ‘Variations of Box Plots’, *The American Statistician*. Taylor & Francis, 32(1), pp. 12–16. doi: 10.1080/00031305.1978.10479236.

McMillan, C. T. *et al.* (2015) ‘C9orf72 promoter hypermethylation is neuroprotective: Neuroimaging and neuropathologic evidence’, *Neurology*, 84, pp. 1622–1630. doi: 10.1212/WNL.0000000000001495.

Mead, S. (2006) ‘Prion disease genetics.’, *European journal of human genetics : EJHG*, 14(3), pp. 273–281. doi: 10.1038/sj.ejhg.5201544.

Mead, S. *et al.* (2016) ‘Clinical Trial Simulations Based on Genetic Stratification and the Natural History of a Functional Outcome Measure in Creutzfeldt-Jakob Disease.’, *JAMA Neurol*, 73(4), pp. 447–455. doi: 10.1001/jamaneurol.2015.4885.

Meehan, R. R. *et al.* (1989) ‘Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs’, *Cell*, 58(3), pp. 499–507. doi: 10.1016/0092-8674(89)90430-3.

Ter Meulen, V. and Hall, W. W. (1978) ‘Slow virus infections of the nervous system: Virological, immunological and pathogenetic considerations’, *Journal of General Virology*, 41(1), pp. 1–25. doi: 10.1099/0022-1317-41-1-1.

Miller, M. W. *et al.* (2018) ‘Brain , Behavior , and Immunity CRP polymorphisms and DNA methylation of the AIM2 gene influence associations between trauma exposure , PTSD , and C-reactive protein’, *Brain Behavior and Immunity*, 67, pp. 194–202. doi: 10.1016/j.bbi.2017.08.022.

Mok, T. H. *et al.* (2017) ‘Variant Creutzfeldt–Jakob Disease in a Patient with Heterozygosity at PRNP Codon 129’, *New England Journal of Medicine*, 376(3), pp. 290–292. doi: 10.1056/NEJMc1615251.

Morris, T. J. *et al.* (2014) ‘ChAMP: 450k Chip Analysis Methylation Pipeline’,

Bioinformatics, 30(3), pp. 428–430. doi: 10.1093/bioinformatics/btt684.

Morrone, S. R. *et al.* (2015) 'Assembly-driven activation of the AIM2 foreign-dsDNA sensor provides a polymerization template for downstream ASC', *Nature Communications*, 6(September). doi: 10.1038/ncomms8827.

Müller, T., Woitalla, D. and Kuhn, W. (2003) 'Benefit of folic acid supplementation in parkinsonian patients treated with levodopa', *Journal of Neurology Neurosurgery and Psychiatry*, 74(4), p. 549. doi: 10.1136/jnnp.74.4.549.

Nabel, C. S. *et al.* (2012) 'AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation', *Nature Chemical Biology*, 8(9), pp. 751–758. doi: 10.1038/nchembio.1042.

Ng, H.-H., Jeppesen, P. and Bird, A. (2002) 'Active Repression of Methylated Genes by the Chromosomal Protein MBD1', *Molecular and Cellular Biology*, 20(4), pp. 1394–1406. doi: 10.1128/mcb.20.4.1394-1406.2000.

Ng, H. H. *et al.* (1999) 'MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex', *Nature Genetics*, 23(1), pp. 58–61. doi: 10.1038/12659.

Nicot, S. *et al.* (2016) 'Detection of prions in the plasma of presymptomatic and symptomatic patients with variant Creutzfeldt-Jakob disease.', *Science translational medicine*, 8(370), p. 370ra182. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=28003547&retmode=ref&cmd=prlinks%0Apapers3://publication/doi/10.1126/scitranslmed.aag1257.

Noel Gill, O. *et al.* (2013) 'Prevalent abnormal prion protein in human appendixes after bovine spongiform encephalopathy epizootic: Large scale survey', *BMJ (Online)*, 347(7929), pp. 1–12. doi: 10.1136/bmj.f5675.

Nordlund, J. *et al.* (2013) 'Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia', *Genome Biology*, 14(9). doi: 10.1186/gb-2013-14-9-r105.

O'Leary, J. C. *et al.* (2011) 'A new anti-depressive strategy for the elderly: Ablation of FKBP5/FKBP51', *PLoS ONE*, 6(9). doi: 10.1371/journal.pone.0024840.

Obeid, R. *et al.* (2009) 'Methylation status and neurodegenerative markers in

Parkinson disease', *Clinical Chemistry*, 55(10), pp. 1852–1860. doi: 10.1373/clinchem.2009.125021.

Okano, M., Xie, S. and Li, E. (1998) 'Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases', *Nature Genetics*, 19(3), pp. 219–220. doi: 10.1038/890.

Ooi, S. K. T. *et al.* (2007) 'DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA', *Nature*, 448(7154), pp. 714–717. doi: 10.1038/nature05987.

Pablo, J. *et al.* (2009) 'Therapeutic Neutralisation of the NLRP1 Inflammasome Reduces the Innate Immune Response After Traumatic Brain Injury', *Journal of cerebral blood flow and metabolism*, 29(7), pp. 1251–1261. doi: 10.1038/jcbfm.2009.46.THERAPEUTIC.

Parry, H. B. (1962) 'SCRAPIE: A TRANSMISSIBLE AND HEREDITARY DISEASE', *Heredity*, 17, pp. 75–105.

Paul, D. S. *et al.* (2016) 'Increased DNA methylation variability in type 1 diabetes across three immune effector cell types', *Nature Communications*, 7, p. 13555. doi: 10.1038/ncomms13555.

Perkowski, J. J. and Murphy, G. G. (2011) 'Deletion of the Mouse Homolog of KCNAB2, a Gene Linked to Monosomy 1p36, Results in Associative Memory Impairments and Amygdala Hyperexcitability', *J Neurosci.*, 31(1), pp. 46–54. doi: 10.1016/j.dcn.2011.01.002.The.

Pezzi, J. C. *et al.* (2014) 'DNA methyltransferase haplotype is associated with Alzheimer's disease', *Neuroscience Letters*. Elsevier Ireland Ltd, 579, pp. 70–74. doi: 10.1016/j.neulet.2014.07.013.

Pidsley, R. *et al.* (2013) 'A data-driven approach to preprocessing Illumina 450K methylation array data', *BMC Genomics*, 14(1), p. 293. doi: 10.1186/1471-2164-14-293.

Plummer, P. J. G. (1946) 'Scrapie—A Disease of Sheep', *Canadian journal of comparative medicine and veterinary science*, (1).

Powell-Jackson, J., Kennedy, P. and Whitcombe, E. M. (1985) 'CREUTZFELDT-JAKOB DISEASE AFTER ADMINISTRATION OF HUMAN GROWTH HORMONE',

The Lancet, 326(8449), pp. 244–246.

Prusiner, S. (1991) 'Molecular biology of prion diseases', *Science*, 252(5012), pp. 1515–1522. doi: 10.1126/science.1675487.

Prusiner, S. B. *et al.* (1981) 'Scrapie agent contains a hydrophobic protein.', *Proceedings of the National Academy of Sciences of the United States of America*, 78(11), pp. 6675–9. doi: 10.1073/pnas.78.11.6675.

Prusiner, S. B. (1982) 'Novel proteinaceous infectious particles cause scrapie.', *Science (New York, N.Y.)*, 216(4542), pp. 136–44. doi: 10.1126/science.6801762.

Prusiner, S. B. and Scott, M. R. (1997) 'GENETICS OF PRIONS', *Annual Review of Genetics*, 31, pp. 139–175. doi: 10.1146/annurev.genet.31.1.139.

Prusiner, S. and Stahl, N. (1991) 'Prions and prion proteins', *FASEB J.*, 5(13), pp. 2799–2807.

Pushie, M. J. *et al.* (2011) 'Prion protein expression level alters regional copper, iron and zinc content in the mouse brainw', *Metallomics*, 3(3), pp. 206–214. doi: 10.1039/c0mt00037j.

Qazi, T. J. *et al.* (2018) 'Epigenetics in Alzheimer's Disease: Perspective of DNA Methylation', *Molecular Neurobiology*. Molecular Neurobiology, 55(2), pp. 1026–1044. doi: 10.1007/s12035-016-0357-6.

Racko, D. *et al.* (2018) 'Transcription-induced supercoiling as the driving force of chromatin loop extrusion during formation of TADs in interphase chromosomes', *Nucleic Acids Research*. Oxford University Press, 46(4), pp. 1648–1660. doi: 10.1093/nar/gkx1123.

Ramsahoye, B. H. *et al.* (2000) 'Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a.', *Proceedings of the National Academy of Sciences of the United States of America*, 97(10), pp. 5237–5242. doi: 10.1073/pnas.97.10.5237.

Rangam, G. *et al.* (2012) 'Aid enzymatic activity is inversely proportional to the size of cytosine c5 orbital cloud', *PLoS ONE*, 7(8), pp. 3–8. doi: 10.1371/journal.pone.0043279.

Reese, B. E. *et al.* (2003) 'The Methyl-CpG Binding Protein MBD1 Interacts with the p150 Subunit of Chromatin Assembly Factor 1', *Molecular and Cellular Biology*,

23(9), pp. 3226–3236. doi: 10.1128/mcb.23.9.3226-3236.2003.

Renton, A. E. *et al.* (2011) 'A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD', *Neuron*, 72(2), pp. 257–268. doi: 10.1016/j.neuron.2011.09.010.A.

Resmini, E. *et al.* (2016) 'Reduced DNA methylation of FKBP5 in Cushing's syndrome', *Endocrine*, 54(3), pp. 768–777. doi: 10.1007/s12020-016-1083-6.

Ritchie, M. E. *et al.* (2015) 'limma powers differential expression analyses for RNA-sequencing and microarray studies', *Nucleic acids research*, 43(7), p. e47. doi: 10.1093/nar/gkv007.

Roadmap Epigenomics Consortium *et al.* (2015) 'Integrative analysis of 111 reference human epigenomes', *Nature*, 518(7539), pp. 317–329. doi: 10.1038/nature14248.

Roessler, J. *et al.* (2012) 'Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc.', *BMC Research Notes*, 5(1), p. 210. doi: 10.1186/1756-0500-5-210.

Roman-Gomez, J. *et al.* (2008) 'Repetitive DNA hypomethylation in the advanced phase of chronic myeloid leukemia', *Leukemia Research*, 32(3), pp. 487–490. doi: 10.1016/j.leukres.2007.07.021.

Rudge, P. *et al.* (2015) 'Iatrogenic CJD due to pituitary-derived growth hormone with genetically determined incubation times of up to 40 years', *Brain*, 138(11), pp. 3386–3399. doi: 10.1093/brain/awv235.

Ruiz-Estevez, M. *et al.* (2018) 'Promotion of Myoblast Differentiation by Fkbp5 via Cdk4 Isomerization', *Cell Reports*. ElsevierCompany., 25(9), pp. 2537-2551.e8. doi: 10.1016/j.celrep.2018.11.006.

Safar, J. *et al.* (1998) 'Eight prion strains have PrP(Sc) molecules with different conformations.', *Nature medicine*, 4(10), pp. 1157–1165. doi: 10.1038/2654.

Sandberg, M. K. *et al.* (2014) 'Prion neuropathology follows the accumulation of alternate prion protein isoforms after infective titre has peaked', *Nature Communications*. Nature Publishing Group, 5(May), pp. 1–7. doi: 10.1038/ncomms5347.

Sasai, N. *et al.* (2010) 'The Human Proteins MBD5 and MBD6 Associate with

Heterochromatin but They Do Not Bind Methylated DNA', *PLoS ONE*, 5(8), p. e11982. doi: 10.1371/journal.pone.0011982.

Saunders, A. M. *et al.* (1993) 'Association of apolipoprotein E allele 4 with late-onset familial and sporadic Alzheimer's disease', *Neurology*, 43(8), pp. 1467–1467. doi: 10.1212/wnl.43.8.1467.

Saxonov, Serge, Berg, P. and Brutlag, D. L. (2006) 'A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 103(5), pp. 1412–1471. doi: 10.1073/pnas.87.12.4692.

Saxonov, S., Berg, P. and Brutlag, D. L. (2006) 'A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters', *Proceedings of the National Academy of Sciences*, 103(5), pp. 1412–1417. doi: 10.1073/pnas.0510310103.

Saya, H. *et al.* (2002) 'Mechanism of Transcriptional Regulation by Methyl-CpG Binding Protein MBD1', *Molecular and Cellular Biology*, 20(14), pp. 5107–5118. doi: 10.1128/mcb.20.14.5107-5118.2000.

Schwartz, S., Meshorer, E. and Ast, G. (2009) 'Chromatin organization marks exon-intron structure', *Nature Structural and Molecular Biology*. Nature Publishing Group, 16(9), pp. 990–995. doi: 10.1038/nsmb.1659.

Shen, S. Y. *et al.* (2018) 'Sensitive tumour detection and classification using plasma cell-free DNA methylomes', *Nature*. Springer US, 563(7732), pp. 579–583. doi: 10.1038/s41586-018-0703-0.

Shukla, S. *et al.* (2011) 'CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing', *Nature*. Nature Publishing Group, 479(7371), pp. 74–79. doi: 10.1038/nature10442.

Simpkins, S. B. *et al.* (2002) 'MLH1 Promoter Methylation and Gene Silencing is the Primary Cause of Microsatellite Instability in Sporadic Endometrial Cancers', *Human Molecular Genetics*, 8(4), pp. 661–666. doi: 10.1093/hmg/8.4.661.

Singh-Manoux, A. *et al.* (2017) 'Trajectories of depressive symptoms before diagnosis of dementia: A 28-year follow-up study', *JAMA Psychiatry*, 74(7), pp. 712–718. doi: 10.1001/jamapsychiatry.2017.0660.

Skene, P. J. *et al.* (2010) 'Neuronal MeCP2 Is Expressed at Near Histone-Octamer Levels and Globally Alters the Chromatin State', *Molecular Cell*. Elsevier Ltd, 37(4), pp. 457–468. doi: 10.1016/j.molcel.2010.01.030.

Smith, A. R. *et al.* (2019) 'A cross-brain regions study of ANK1 DNA methylation in different neurodegenerative diseases', *Neurobiology of Aging*. Elsevier Inc, 74, pp. 70–76. doi: 10.1016/j.neurobiolaging.2018.09.024.

Smith, D. F. *et al.* (1993) 'Two FKBP-related proteins are associated with progesterone receptor complexes', *Journal of Biological Chemistry*, 268(24), pp. 18365–18371.

Smith, D. F., Faber, L. E. and Toft, D. O. (1990) 'Purification of unactivated progesterone receptor and identification of novel receptor-associated proteins', *Journal of Biological Chemistry*, 265(7), pp. 3996–4003.

Smith, R. G. *et al.* (2018) 'Elevated DNA methylation across a 48-kb region spanning the HOXA gene cluster is associated with Alzheimer's disease neuropathology', *Alzheimer's & Dementia*. Elsevier Inc., (March), pp. 1–9. doi: 10.1016/j.jalz.2018.01.017.

Sofi, F. *et al.* (2011) 'Physical activity and risk of cognitive decline: A meta-analysis of prospective studies', *Journal of Internal Medicine*, 269(1), pp. 107–117. doi: 10.1111/j.1365-2796.2010.02281.x.

Song, L. *et al.* (2005) 'Specific Method for the Determination of Genomic DNA Methylation by Liquid Chromatography-Electrospray Ionization Tandem Mass Spectrometry', *Analytical Chemistry*. American Chemical Society, 77(2), pp. 504–510. doi: 10.1021/ac0489420.

Song, Y., Ren, H. and Lei, J. (2017) 'Collaborations between CpG sites in DNA methylation', *bioRxiv*. doi: https://doi.org/10.1101/149815.

Sontag, E. *et al.* (2007) 'Protein Phosphatase 2A Methyltransferase Links Homocysteine Metabolism with Tau and Amyloid Precursor Protein Regulation', *Journal of Neuroscience*, 27(11), pp. 2751–2759. doi: 10.1523/JNEUROSCI.3316-06.2007.

Sponne, I. *et al.* (2004) 'Humanin rescues cortical neurons from prion-peptide-induced apoptosis', *Molecular and Cellular Neuroscience*, 25(1), pp. 95–102. doi: 10.1016/j.mcn.2003.09.017.

Stadler, M. B. *et al.* (2011) 'DNA-binding factors shape the mouse methylome at distal regulatory regions', *Nature*, 480(7378), pp. 490–495. doi: 10.1038/nature10716.

Stapels, M. *et al.* (2010) 'Polycomb group proteins as epigenetic mediators of neuroprotection in ischemic tolerance', *Science signaling*, 3(111), pp. ra15–ra15. doi: 10.1126/scisignal.2000502.

Steele, A. D. *et al.* (2006) 'Prion protein (PrPc) positively regulates neural precursor proliferation during developmental and adult mammalian neurogenesis.', *Proceedings of the National Academy of Sciences of the United States of America*, 103(9), pp. 3416–21. doi: 10.1073/pnas.0511290103.

Stein, R. *et al.* (1982) 'Clonal inheritance of the pattern of DNA methylation in mouse cells.', *Proceedings of the National Academy of Sciences*, 79(1), pp. 61–65. doi: 10.1073/pnas.79.1.61.

Stocki, P. *et al.* (2016) 'Inhibition of the FKBP family of peptidyl prolyl isomerases induces abortive translocation and degradation of the cellular prion protein.', *Molecular biology of the cell*, 27(5), pp. 757–67. doi: 10.1091/mbc.E15-10-0729.

Sved, J. and Bird, A. (1990) 'The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model (DNA methylation/factor IX/a-globin gene/pseudogene/neutral hypothesis)', *Evolution*, 87(June 1990), pp. 4692–4696. Available at: http://www.pnas.org/content/87/12/4692.long.

Team, R. C. (2016) 'R: A language and environment for statistical computing'. Vienna, Austria: R Foundation for Statistical Computing.

Terry, C. *et al.* (2016) 'Ex vivo mammalian prions are formed of paired double helical prion protein fibrils', *Open Biology*, 6(5), p. 160035. doi: 10.1098/rsob.160035.

Teschendorff, A. E. *et al.* (2012) 'Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation', *Genome Medicine*, 4(3). doi: 10.1186/gm323.

Teschendorff, A. E. *et al.* (2013) 'A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data', *Bioinformatics*, 29(2), pp. 189–196. doi: 10.1093/bioinformatics/bts680.

Teschendorff, A. E. *et al.* (2016) 'DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer', *Nature Communications*. Nature Publishing Group, 7, p. 10478. doi: 10.1038/ncomms10478.

Teschendorff, A. E. *et al.* (2017) 'A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies', *BMC Bioinformatics*. BMC Bioinformatics, 18(1), p. 105. doi: 10.1186/s12859-017-1511-5.

Teschendorff, A. E. and Zheng, S. C. (2017) 'Cell-type deconvolution in epigenome-wide association studies: a review and recommendations', *Epigenomics*, 9, p. epi-2016-0153. doi: 10.2217/epi-2016-0153.

Thakur, A. K. *et al.* (2011) 'Copper alters aggregation behavior of prion protein and induces novel interactions between its N- and C-terminal regions', *Journal of Biological Chemistry*, 286(44), pp. 38533–38545. doi: 10.1074/jbc.M111.265645.

The National CJD Research and Surveillance Unit (2017) *26th Annual Report 2017: CREUTZFELDT-JAKOB DISEASE SURVEILLANCE IN THE UK*. Available at: https://www.cjd.ed.ac.uk/sites/default/files/report26.pdf.

The University of Edinburgh (2017) 'Variant CJD cases Worldwide.', *The National CJD Research & Surveillance Unit*, p. 1. Available at: http://www.cjd.ed.ac.uk/sites/default/files/worldfigs_2.pdf.

Thompson, A. G. B. *et al.* (2013) 'The medical research council prion disease rating scale: A new outcome measure for prion disease therapeutic trials developed and validated using systematic observational studies', *Brain*, 136(4), pp. 1116–1127. doi: 10.1093/brain/awt048.

Tobiasson, M. *et al.* (2017) 'Comprehensive mapping of the effects of azacitidine on DNA methylation, repressive/permissive histone marks and gene expression in primary cells from patients with MDS and MDS-related disease', *Oncotarget*, 8(17), pp. 28812–28825. doi: 10.18632/oncotarget.15807.

Touma, C. *et al.* (2011) 'FK506 binding protein 5 shapes stress responsiveness: Modulation of neuroendocrine reactivity and coping behavior', *Biological Psychiatry*. Elsevier Inc., 70(10), pp. 928–936. doi: 10.1016/j.biopsych.2011.07.023.

Triche, T. J. *et al.* (2013) 'Low-level processing of Illumina Infinium DNA Methylation BeadArrays', *Nucleic Acids Research*, 41(7), pp. 1–11. doi: 10.1093/nar/gkt090.

Turner, S. D. (2018) ‘qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots’, *Journal of Open Source Software*, 3(25), p. 731. doi: 10.21105/joss.00731.

Vázquez-Fernández, E. *et al.* (2016) ‘The Structural Architecture of an Infectious Mammalian Prion Using Electron Cryomicroscopy’, *PLoS Pathogens*, 12(9), pp. 1–21. doi: 10.1371/journal.ppat.1005835.

Veeranki, S. *et al.* (2011) ‘IFI16 protein mediates the anti-inflammatory actions of the type-I interferons through suppression of activation of caspase-1 by inflammasomes’, *PLoS ONE*, 6(10). doi: 10.1371/journal.pone.0027040.

Waddington, C. H. (1942) ‘The epigenotype’, *Endeavour*, pp. 18–20. doi: 10.1093/ije/dyr184.

Walsh, C. P., Chaillet, J. R. and Bestor, T. H. (1998) ‘Transcription of IAP endogenous retroviruses is constrained by cytosine methylation [4]’, *Nature Genetics*, 20(2), pp. 116–117. doi: 10.1038/2413.

Walsh, D. T., Betmouni, S. and Perry, V. H. (2001) ‘Absence of detectable IL-1β production in murine prion disease: A model of chronic neurodegeneration’, *Journal of Neuropathology and Experimental Neurology*, 60(2), pp. 173–182. doi: 10.1093/jnen/60.2.173.

Wang, H. *et al.* (2011) ‘Smoking and risk of amyotrophic lateral sclerosis: a pooled analysis of five prospective cohorts’, *Arch Neurol.*, 68(2), pp. 207–213. doi: 10.1001/archneurol.2010.367.Smoking.

Wang, L. (2011) ‘FKBP51 regulation of AKT/Protein Kinase B Phospharylation’, *Curr Opin Pharmacol*, 11(4), pp. 360–364. doi: 10.1016/j.pmrj.2014.02.014.Lumbar.

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wild, E. *et al.* (2011) ‘Abnormal peripheral chemokine profile in Huntington’s disease’, *PLoS currents*. Public Library of Science, 3, pp. RRN1231–RRN1231. doi: 10.1371/currents.RRN1231.

Will, R. . *et al.* (1996) ‘A new variant of Creutzfeldt-Jakob disease in the UK’, *The Lancet*, 347(9006), pp. 921–925. doi: 10.1016/S0140-6736(96)91412-9.

William, R. *et al.* (1998) ‘The IL-1 beta-converting enzyme (caspase-1) inhibits

apoptosis of inflammatory neutrophils through activation of IL-1 beta.', *Journal of immunology (Baltimore, Md. : 1950)*, 161(2), pp. 957–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9670975.

Williams, A. *et al.* (1997) 'Immunocytochemical appearance of cytokines, prostaglandin E2 and lipocortin-1 in the CNS during the incubation period of murine scrapie correlates with progressive PrP accumulations', *Brain Research*, 754(1–2), pp. 171–180. doi: 10.1016/S0006-8993(97)00067-X.

Williams, A. E., Ryder, S. and Blakemore, W. F. (1995) 'Monocyte recruitment into the scrapie-affected brain', *Acta Neuropathologica*, 90(2), pp. 164–169. doi: 10.1007/BF00294316.

Windle, R. J. *et al.* (1998) 'Ultradian rhythm of basal corticosterone release in the female rat: Dynamic interaction with the response to acute stress', *Endocrinology*, 139(2), pp. 443–450. doi: 10.1210/endo.139.2.5721.

Wochnik, G. M. *et al.* (2005) 'FK506-binding proteins 51 and 52 differentially regulate dynein interaction and nuclear translocation of the glucocorticoid receptor in mammalian cells', *Journal of Biological Chemistry*, 280(6), pp. 4609–4616. doi: 10.1074/jbc.M407498200.

Wockner, L. F. *et al.* (2014) 'Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients', *Translational Psychiatry*, 4(September 2013), pp. 1–8. doi: 10.1038/tp.2013.111.

World Health Organisation (2003) 'WHO manual for surveillance of human transmissible spongiform encephalopathies including variant Creutzfeldt-Jakob disease', *WHO manual for surveillance of human transmissible spongiform encephalopathies*, p. 105.

Wu, H. and Zhang, Y. (2011) 'Tet1 and 5-hydroxymethylation: A genome-wide view in mouse embryonic stem cells', *Cell Cycle*, 10(15), pp. 2428–2436. doi: 10.4161/cc.10.15.16930.

Wu, H. and Zhang, Y. (2014a) 'Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions', *Cell*, 156(0), pp. 45–68. doi: 10.1016/j.cell.2013.12.019.

Wu, H. and Zhang, Y. (2014b) 'Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions', *Cell*, 156(0), pp. 45–68. doi:

10.1016/j.cell.2013.12.019.

Wu, P.-J. *et al.* (2016) 'AIM 2 inflammasomes regulate neuronal morphology and influence anxiety and memory in mice.', *Scientific reports*. Nature Publishing Group, 6(August), p. 32405. doi: http://dx.doi.org/10.1038/srep32405.

Xi, Z. *et al.* (2013) 'Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion.', *American journal of human genetics*, 92(6), pp. 981–9. doi: 10.1016/j.ajhg.2013.04.017.

Yamada, M. *et al.* (2009) 'Dura mater graft-associated Creutzfeldt-Jakob disease in Japan: clinicopathological and molecular characterization of the two distinct subtypes.', *Neuropathology : official journal of the Japanese Society of Neuropathology*, 29(October 2008), pp. 609–618. doi: 10.1111/j.1440-1789.2008.00987.x.

Yang, A. S. *et al.* (2004) 'A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements', *Nucleic Acids Research*, 32(3), pp. 38e – 38. doi: 10.1093/nar/gnh032.

Yehuda, R. *et al.* (2015) 'Holocaust Exposure Induced Intergenerational Effects on FKBP5 Methylation', *Biological Psychiatry*. Elsevier, 80(5), pp. 372–380. doi: 10.1016/j.biopsych.2015.08.005.

Zannas, A. S. *et al.* (2016) 'Gene–Stress–Epigenetic Regulation of FKBP5: Clinical and Translational Implications', *Neuropsychopharmacology*, 41(1), pp. 261–274. doi: 10.1038/npp.2015.235.

Zanusso, G. *et al.* (2016) 'Advanced tests for early and accurate diagnosis of Creutzfeldt–Jakob disease', *Nature Reviews Neurology*. Nature Publishing Group, 12(May), pp. 325–333. doi: 10.1038/nrneurol.2016.65.

Zarei, S. *et al.* (2015) 'A comprehensive review of amyotrophic lateral sclerosis', *Surgical Neurology International*. India: Medknow Publications & Media Pvt Ltd, 6, p. 171. doi: 10.4103/2152-7806.169561.

Zhang, H. *et al.* (2010) 'TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine', *Cell Research*, 20(12), pp. 1390–1393. doi: 10.1038/cr.2010.156.

Zhang, L. *et al.* (2017) 'DNA Methylation Landscape Reflects the Spatial

Organization of Chromatin in Different Cells', *Biophysical Journal*. Biophysical Society, 113(7), pp. 1395–1404. doi: 10.1016/j.bpj.2017.08.019.

Zhang, P. *et al.* (2017) 'Methyl-CpG binding domain protein 1 regulates localization and activity of Tet1 in a CXXC3 domain-dependent manner', *Nucleic Acids Research*, 45(12), pp. 7118–7136. doi: 10.1093/nar/gkx281.

Zhang, Wenjun *et al.* (2012) 'S14G-Humanin improves cognitive deficits and reduces amyloid pathology in the middle-aged APPswe/PS1dE9 mice', *Pharmacology Biochemistry and Behavior*. Elsevier Inc., 100(3), pp. 361–369. doi: 10.1016/j.pbb.2011.09.012.

Zhang, Y. *et al.* (1999) 'Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation', *Genes & Development*, 13, pp. 1924–1935. doi: 10.1101/gad.13.15.1924.

Zhou, W., Laird, P. W. and Shen, H. (2016) 'Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes', *Nucleic Acids Research*, 45(4), p. gkw967. doi: 10.1093/nar/gkw967.

Zhu, H., Wang, G. and Qian, J. (2016) 'Transcription factors as readers and effectors of DNA methylation.', *Nature reviews. Genetics*. Nature Publishing Group, 17(9), pp. 551–65. doi: 10.1038/nrg.2016.83.

Zilberman, D. *et al.* (2008) 'Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks', *Nature*, 456(7218), pp. 125–129. doi: 10.1038/nature07324.

Zomosa-Signoret, V. *et al.* (2008) 'Physiological role of the cellular prion protein', *Veterinary Research*, 39(4), p. 09. doi: 10.1051/vetres:2007048.