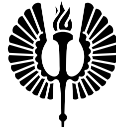
IMPLEMENTATION OF MODERN METHODS TO IDENTIFY DNA METHYLATION MARKS FOR EPISODIC MEMORY IMPAIRMENT AND ALZHEIMER'S DISEASE

Mikko Konki

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA - SER. D OSA - TOM. 1506 | MEDICA - ODONTOLOGICA | TURKU 2020



IMPLEMENTATION OF MODERN METHODS TO IDENTIFY DNA METHYLATION MARKS FOR EPISODIC MEMORY IMPAIRMENT AND ALZHEIMER'S DISEASE

Mikko Konki

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA - SER. D OSA – TOM. 1506 | MEDICA – ODONTOLOGICA | TURKU 2020

University of Turku

Faculty of Medicine
Medical Biochemistry and Genetics
Turku Doctoral Programme of Molecular Medicine
Turku Bioscience Centre

Supervised by

Docent Riikka Lund, PhD
Turku Bioscience Centre
University of Turku
Turku, Finland

Professor Juha Rinne, MD, PhD
Turku PET Centre
University of Turku
Turku University Hospital
Turku, Finland

Reviewed by

Professor Qihua Tan, MD, PhD
University of Southern Denmark
Odense, Denmark

Docent Nina Kaminen-Ahola, PhD
University of Helsinki
Helsinki, Finland

Opponent

Docent Juulia Jylhävä, PhD
Karolinska Institutet
Stockholm, Sweden

The originality of this publication has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

ISBN 978-951-29-8171-7 (PRINT)
ISBN 978-951-29-8172-4 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)
Painosalama Oy, Turku, Finland 2020

UNIVERSITY OF TURKU
Faculty of Medicine
Medical Biochemistry and Genetics
MIKKO KONKI: Implementation of modern methods to identify DNA methylation marks for episodic memory impairment and Alzheimer's disease
Doctoral Dissertation, 112 pp.
Turku Doctoral Programme of Molecular Medicine
Turku Bioscience Centre
September 2020

ABSTRACT

Alzheimer's disease (AD) is a progressive neurodegenerative disease and a common cause of dementia worldwide. The pathogenic mechanisms of AD are not completely understood. Both genetic and external factors, like lifestyle and other diseases, can alter AD risk. The impact of external factors may be mediated by epigenetic mechanisms, like DNA methylation. The first clinical symptoms of AD usually manifest as episodic memory impairment (EMI), however, AD can be diagnosed only at a later stage when neurodegeneration is far-progressed. New methods are required to enable early diagnosis and characterization of the disease mechanisms.

The main goals of this study were to implement new methods for characterization of disease marks and mechanisms and to identify new DNA methylation marks for EMI and AD with these methods. Reduced Representation Bisulfite Sequencing (RRBS) was first implemented by utilising human embryonic stem cells and by identifying DNA methylation changes in these cells during transformation to abnormal karyotype. RRBS was then utilised to identify blood DNA methylation marks for AD in Finnish disease discordant twin pairs. Such marks were detected in eleven genomic regions and the one in adenosine deaminase RNA specific B2 gene (*ADARB2*) was validated in Swedish twin cohorts. A new twin sample cohort was collected and genome-wide bisulphite sequencing method was implemented to identify plasma cell-free DNA methylation marks for EMI. No markers were detected. Patient-specific induced pluripotent stem cell lines and a brain organoid model were generated to study the pathogenic mechanisms of AD.

In conclusion, the results show that DNA methylation marks associated with AD can be detected not only in the brain but also in blood, however, more research is required to evaluate whether these marks can be utilised in diagnostics. Several methods enabling identification of disease marks were successfully implemented, samples from a new cohort were collected and pluripotent stem cell-based model was established to study AD. These resources will be valuable for future research aiming to identify mechanisms and markers for AD.

KEYWORDS: Alzheimer's disease, blood biomarkers, epigenomics, DNA methylation, human pluripotent stem cells, disease model.

TURUN YLIOPISTO
Lääketieteellinen tiedekunta
Lääketieteellinen biokemia ja genetiikka
MIKKO KONKI: Uusien menetelmien käyttöönotto Alzheimerin tautiin ja episodisen muistin heikkenemiseen liittyvien DNA-metylaatiomarkkereiden tunnistamiseksi
Väitöskirja, 112 s.
Molekyylilääketieteen tohtoriohjelma
Turku Bioscience Centre
Syyskuu 2020

TIIVISTELMÄ

Alzheimerin tauti (AT) on keskushermoston rappeumaan johtava sairaus, joka aiheuttaa yli puolet maailman dementiatapauksista. Taudin aiheuttajia ei vielä täysin tunneta. Sekä perintötekijät että ulkoiset tekijät, kuten elintavat ja muut sairaudet vaikuttavat AT-riskiin. Ulkoisten tekijöiden vaikutus saattaa johtua muutoksista epigeneettisissä mekanismeissa, kuten DNA:n metylaatiossa. Episodisen muistin heikkeneminen (EMH) aiheuttaa yleensä AT:n ensimmäiset kliiniset oireet, mutta sairaus pystytään diagnosoimaan vasta huomattavasti myöhemmin, jolloin hermosolutuho on edennyt jo pitkälle. Tarvitaan uusia menetelmiä AT:n mekanismien tutkimiseen sekä sairauden varhaiseen diagnosointiin.

Tämän tutkimuksen päätavoitteina oli ottaa käyttöön uusia menetelmiä tautimarkkerien ja -mekanismien tutkimiseen sekä etsiä niiden avulla DNA-metylaatiomarkkereita AT:lle sekä EMH:lle. RRBS (Reduced Representation Bisulfite Sequencing) otettiin ensin käyttöön tutkimalla DNA-metylaatiomuutoksia ihmisen alkion kantasoluissa niiden karyotyyppien muuttuessa epänormaaleiksi. Seuraavaksi RRBS:n avulla tunnistettiin AT:hen liittyviä metylaatiomarkkereita veressä suomalaisilla kaksospareilla, joista toisella on AT ja toisella ei. Markkereita havaittiin 11 genomin alueella. Näistä yksi, joka sijaitsi adenosine deaminase RNA specific B2 -geenissä (*ADARB2*), validoitiin ruotsalaisessa kaksosaineistossa. Keräsimme näytteitä uuteen kaksosaineistoon, jossa tutkimme EMH:hon liittyviä metylaatiomarkkereita plasman soluvapaassa DNA:ssa. Merkittäviä markkereita ei havaittu. Lopuksi perustimme indusoidut pluripotentit kantasolulinjat sekä aivokudosmallin AT:hen liittyvien mekanismien tutkimista varten.

Tämän tutkimuksen aikana otettiin käyttöön useita uusimpia genomin ja epigenomin tutkimusmenetelmiä. Keräsimme arvokkaan näyteaineiston, jota voidaan hyödyntää muistin heikentymisen tutkimuksessa. Tutkimuksen tulosten mukaan AT:hen liittyviä DNA-metylaatiomuutoksia on havaittavissa aivokudoksen lisäksi myös veressä. Tarvitaan kuitenkin lisätutkimuksia selvittämään, voidaanko metylaatiomarkkereita hyödyntää AT:n diagnostiikassa.

AVAINSANAT: Alzheimerin tauti, tautimarkkeri, DNA-metylaatio, epigenomiikka, ihmisen pluripotentit kantasolut, tautimallit

Table of Contents

Abbreviations **8**

List of Original Publications **9**

1 Introduction **10**

2 Review of the Literature **12**
2.1 Alzheimer's disease and dementia 12
2.1.1 Dementia – a worldwide burden 12
2.1.2 Alzheimer's disease aetiology 13
2.1.3 Alzheimer's disease diagnosis and treatment 14
2.2 Epigenetics and Alzheimer's disease 16
2.2.1 Epigenome 16
2.2.2 Epigenetic mechanisms 18
2.2.2.1 DNA methylation 18
2.2.2.2 Histone modifications 20
2.2.2.3 Non-coding RNAs 21
2.2.3 Epigenome and environment 21
2.2.4 Epigenetic alterations in Alzheimer's disease 23
2.2.5 Twin studies – a unique perspective into epigenetic variation 26
2.3 Unravelling the epigenome – DNA methylation analysis via bisulphite sequencing 26
2.3.1 Genome-wide DNA methylation analysis 26
2.3.2 Bisulphite sequencing 27
2.3.3 Bisulphite sequencing data analysis 30
2.4 Human pluripotent stem cells in disease modelling 31
2.4.1 Human pluripotent stem cells 31
2.4.2 Human pluripotent stem cell cultures and genomic stability 31
2.4.3 Human pluripotent stem cell models for Alzheimer's disease 33
2.5 Summary 33

3 Aims of the Study **35**

4 Materials and Methods **36**
4.1 Materials 36
4.1.1 Cell cultures (I) 36
4.1.2 Twin blood samples (I, II, III) 36

4.1.2.1 Finnish twin samples 36
4.1.2.2 Swedish twin samples 37
4.1.3 Human brain tissue samples (II) 37
4.1.4 Human iPSC lines reprogrammed from peripheral blood mononuclear cells 38
4.2 Methods 40
4.2.1 Human tissue sample processing 40
4.2.1.1 DNA isolation from EDTA blood samples (I, II) 40
4.2.1.2 Isolation of plasma and peripheral blood mononuclear cells from blood samples (III) 40
4.2.1.3 DNA isolation from plasma (III) 40
4.2.1.4 DNA isolation from frozen brain tissue samples (II) 40
4.2.2 Targeted sequencing of single nucleotide polymorphisms associated with Alzheimer's disease (II) 41
4.2.3 DNA methylation analysis 41
4.2.3.1 Reduced Representation Bisulfite Sequencing (I, II) 41
4.2.3.2 Targeted bisulphite pyrosequencing (II) 42
4.2.3.3 Plasma cell-free DNA bisulphite sequencing (III) 42
4.2.4 APOE genotyping (II) 43
4.2.5 Single-cell RNA sequencing (II) 43
4.2.6 Human pluripotent stem cell culturing and cerebral organoid differentiation 44
4.2.7 Cerebral organoid staining and imaging 44

5 Results 45
5.1 Implementation of RRBS for genome-wide DNA methylation analysis of human stem cell and blood samples (I) 45
5.2 Catalase gene is silenced via epigenetic mechanisms in karyotypically abnormal hESC lines 45
5.3 Genetic risk for Alzheimer's disease in Finnish disease-discordant twin pairs (II) 46
5.4 Peripheral blood DNA methylation differences in twin pairs discordant for Alzheimer's disease (II) 47
5.5 Alzheimer's disease-associated DNA methylation marks in the brain and correlation with blood DNA methylation markers (II) 48
5.6 Expression of the differentially methylated genes in peripheral blood (II) 49
5.7 Targeted validation of Alzheimer's disease-associated differential methylation in *ADARB2* gene (II) 49
5.8 Plasma cell-free DNA methylation markers for episodic memory impairment (III) 50
5.9 Human pluripotent stem cell models for Alzheimer's disease 51

6 Discussion 53
6.1 RRBS in genome-wide DNA methylation analysis 53
6.2 Peripheral blood DNA methylation markers for Alzheimer's disease 54
6.3 ADARB2 – a biomarker candidate for Alzheimer's disease 55
6.4 The differentially methylated genes in the blood are associated with neuronal functions and disorders 56
6.5 DNA methylation markers associated with Alzheimer's disease in the hippocampus 58
6.6 Plasma cell-free DNA methylation marks for episodic memory impairment 59
6.7 DNA methylation marks in hESCs and patient-derived iPSCs and brain organoid models 61
6.8 Future perspectives 61

7 Conclusions 63

Acknowledgements 65

References 67

Original Publications 81

Abbreviations

AD	Alzheimer's disease
AD-iPSC	AD patient-derived induced pluripotent stem cell
cfDNA	Cell-free DNA
aMCI	Amnestic mild cognitive impairment
CSF	Cerebrospinal fluid
DMR	Differentially methylated (genomic) region
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
EM	Episodic memory
EOAD	Early onset Alzheimer's disease
HDAC	Histone deacetylase
hESC	Human embryonic stem cells
iPSC	Induced pluripotent stem cells
LOAD	Late onset Alzheimer's disease
MCI	Mild cognitive impairment
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
ncRNA	Non-coding RNA
NGS	Next generation sequencing
PET	Positron emission tomography
PSC	Pluripotent stem cell
RNA	Ribonucleic acid
RRBS	Reduced Representation Bisulfite Sequencing
TET	Ten-eleven-translocation proteins
tRNA	Transfer RNA
WGBS	Whole-genome bisulfite sequencing

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Konki, Mikko; Pasumarthy, Kalyan; Malonzo, Maia; Sainio, Annele; Valensisi, Cristina; Söderström, Mirva; Emani, Maheswara Reddy; Stubb, Aki; Närvä, Elisa; Ghimire, Bishwa; Laiho, Asta; Järveläinen, Hannu; Lahesmaa, Riitta; Lähdesmäki, Harri; Hawkins, R. David; Lund, Riikka J. (2016) 'Epigenetic Silencing of the Key Antioxidant Enzyme Catalase in Karyotypically Abnormal Human Pluripotent Stem Cells', *Scientific Reports*. Nature Publishing Group, 6(1), p. 22190. doi: 10.1038/srep22190.

II Konki, Mikko; Malonzo, Maia; Karlsson, Ida K.; Lindgren, Noora; Ghimire, Bishwa; Smolander, Johannes; Scheinin, Noora M.; Ollikainen, Miina; Laiho, Asta; Elo, Laura L.; Lönnberg, Tapio; Röyttä, Matias; Pedersen, Nancy L.; Kaprio, Jaakko; Lähdesmäki, Harri; Rinne, Juha O.: Lund, Riikka J. (2019) 'Peripheral blood DNA methylation differences in twin pairs discordant for Alzheimer's disease', *Clinical Epigenetics*. BioMed Central, 11(1), p. 130. doi: 10.1186/s13148-019-0729-7.

III Konki, Mikko; Lindgren, Noora; Kyläniemi, Minna; Venho, Reija; Laajala, Essi; Ghimire, Bishwa; Lahesmaa, Riitta; Kaprio, Jaakko; Rinne, Juha O.; Lund, Riikka J. (2020) "Plasma cell-free DNA methylation marks for episodic memory impairment: a pilot twin study," Scientific Reports, 10(1), p. 14192. doi: 10.1038/s41598-020-71239-9.

The original publications have been reproduced with the permission of the copyright holders

1 Introduction

Alzheimer's disease (AD) is a neurodegenerative disease that causes more than half of all dementia cases worldwide. For the most part, AD is a disease of the elderly, even though there is a more uncommon familial form of the disease that has a lower age of onset. Thus, the number of AD patients is increasing as people live longer and populations are ageing. The decline in episodic memory (EM) performance causes the first clinical symptoms in AD. At the moment, AD can be clinically diagnosed only at a very late stage, when the neurodegeneration is far-progressed and patients' cognitive impairment is irreversible. In addition, most of the available therapies for AD only aim at alleviating the symptoms instead of stopping or even delaying the disease progression. There is an urgent need for new diagnostic markers for AD that would enable earlier detection and intervention. However, mechanisms that are involved in AD onset and progression are not yet completely understood, which hinders the development of new diagnostic and therapeutic strategies. Patient-derived induced pluripotent stem cells (iPSC) models, which can be further differentiated to neural cells or 3D organoids, provide new opportunities for in vitro disease modelling and drug development.

Even in the sporadic late-onset form of AD, genetic variants contribute to a major part of the disease risk. Genetics alone does not explain all AD cases, however, other external factors also have an impact on the disease risk. For example, environmental and lifestyle factors, as well as medical history, can either increase or decrease the risk and affect AD progression. According to several recent studies, this impact is at least partly mediated via epigenetic mechanisms, like DNA methylation. Epigenetic mechanisms control cellular development and function by regulating the packing of the genome and expression of genes without changing the nucleic acid sequence of the DNA. In comparison to the DNA nucleic acid sequence, epigenetic mechanisms are more dynamic and prone to transformation due to changes in the environment. The swift progression and development of new genome research tools, like next-generation sequencing (NGS) methods, has enabled the genome-wide characterisation of epigenetic markers in several diseases and other biological conditions. Epigenetic markers not only give new insights to the molecular

mechanism of diseases but can also reveal new diagnostic marker candidates, at least if they can be detected in peripheral tissues that are accessible for sampling.

The main goal of this study was to identify epigenetic markers associated with AD and EM impairment in peripheral blood. To reach this goal I implemented several methods for both genome-wide and targeted epigenetic analyses and characterised differences in Finnish and Swedish twin pairs discordant for the AD and EM impairment. Human embryonic stem cells (hESC) were utilised in the implementation of the genome-wide DNA methylation assay. Furthermore, another goal was to implement human pluripotent stem cell-derived (hPSC) AD models and characterise how the epigenetic markers contribute to disease progression.

2 Review of the Literature

2.1 Alzheimer's disease and dementia

2.1.1 Dementia – a worldwide burden

According to the World Health Organisation's (WHO) estimation, approximately 50 million people are suffering from dementia worldwide at the moment and close to 10 million dementia cases are diagnosed every year. The annual costs of dementia to healthcare and societies in 2015 was 818 billion US dollars, which equals to 1.1% of the global domestic product. The greatest risk factor for dementia is old age even though dementia is not a normal part of ageing. As people live longer, the prevalence of dementia is predicted to increase, reaching 82 million by the year 2030 and 152 million by the year 2050. Alzheimer's disease and other dementias are the 5th most common cause of death. About 60-70 % of dementia cases are caused by AD. Other common diseases causing dementia are vascular dementia, dementia with Lewy bodies and frontotemporal dementias. (WHO, 2019)

Even though there are different diseases causing dementia, the basic symptoms of the syndrome are relatively similar. The different disorders are characterised according to the pathophysiological findings that differ between diseases and are important for the disease progression. A common thing for dementia causing diseases is neurodegeneration, which leads to a typical decline in cognitive functions. The early symptoms start with troubles in remembering recent things and events and awareness of time. When the disease progresses, the patient will have problems surviving from everyday tasks due to periodical unawareness of time and place, not recognising friends and family, changes in personality and finally difficulties in performing basic motor functions, like walking. The patients' need for external help in their daily lives increases as the disease progresses and symptoms become more severe. Dementia does not affect only the patients but also the patients' families, caretakers and whole societies. (Alzheimer's Association, 2019; WHO, 2019)

All the common diseases that cause dementia are progressive, meaning neurodegeneration will only worsen by time. There are specific treatment strategies for different diseases, however, the disease progression in all of the cases is

unstoppable and any alleviation in cognitive symptoms is usually temporary. While there are differences in the pathophysiological processes and findings between these diseases, they are often very difficult to differentiate from each other in the clinics and mixed forms with typical findings from two or more of the diseases exist. The exact causes or triggers for all of the diseases remain unclear. (Alzheimer's Association, 2019)

2.1.2 Alzheimer's disease aetiology

The typical pathophysiological findings in the brain in AD include the accumulation of β-amyloid plaques inside neurons, the accumulation of tau tangles in the intercellular space, decreased glucose metabolism, inflammation caused by hyperactivated microglia, and brain tissue atrophy. Both β-amyloid and tau proteins have an impact on nervous system inflammation and neurodegeneration, however, it is not yet known what initially triggers the accumulation of these proteins. The changes start in brain regions that are involved in cognitive functions and spread to other regions along with disease progression. The pathological brain changes in AD may start even 20-30 years before the clinical symptoms appear (Villemagne *et al.*, 2013). The brain can compensate for the neuronal cell loss and dysfunction at first, and major cognitive decline manifests when this capacity is exceeded.

The AD progression is divided into three stages according to the patient's symptoms and pathophysiological findings. In preclinical AD, the pathological changes have started in the brain, however, the clinical symptoms are still unnoticeable. In the second stage that is mild cognitive impairment (MCI) due to AD, the progressive cognitive decline starts to cause mild clinical symptoms, like memory and thinking problems, that are often noticeable at least to friends and family members but do not hinder everyday activities remarkably. The third stage is dementia due to AD. At this point, the cognitive impairment is noticeable and the patients start to have significant problems in surviving daily life without external help. (Albert *et al.*, 2011; McKhann *et al.*, 2011; Sperling *et al.*, 2011)

Genetic variants have a significant impact on AD risk and progression. A familial autosomal dominant form of AD exists, and it is caused by mutations in amyloid precursor protein (*APP*), presenilin 1 (*PSEN1*) and/or presenilin 2 (*PSEN2*) genes that are directly involved in increased accumulation of β-amyloid in the neuronal cells. However, the familial form accounts for only about 0.1% of all the AD cases. The dominant hereditary AD has a typical early age of onset usually before 65 years, thus, it is often called early-onset AD (EOAD). (Blennow, Leon and Zetterberg, 2006) Genetic variance plays a significant role also in the sporadic late-onset AD (LOAD). According to twin and family studies, genetics may contribute up to 80 %

of the attributable risk also in the sporadic forms (Wilson *et al.*, 2011). Certain apolipoprotein E (*APOE*) genotypes, especially *APOE ε4*, are a well-known risk factor for AD, which explains up to 20 % of the risk (Genin *et al.*, 2011). Genome-wide association studies have identified close to 30 genomic loci contributing to the genetic AD risk, with varying effects, however, a large part of the genetic component in LOAD remains unclear (Lambert *et al.*, 2013; Ridge *et al.*, 2016; Jansen *et al.*, 2019).

The greatest risk factor for AD is old age even though AD is not a direct consequence of ageing itself (Nelson *et al.*, 2011). Alzheimer's disease prevalence increases from 3 % in people aged 65-74 to 17 % in people aged 75-84 and 32 % in people aged 85 or more (Hebert *et al.*, 2013). Family history is another significant risk factor for AD if close relatives have the disease, which may be due to both genetics and shared environmental conditions. External factors, some of which can be modulated by lifestyle choices, also affect AD risk. Cardiovascular disease and AD share many common risk factors, like diabetes, obesity in midlife, high cholesterol, hypertension and smoking (Solomon *et al.*, 2009; Rönnemaa *et al.*, 2011; Rusanen *et al.*, 2011; Gudala *et al.*, 2013; Vagelatos and Eslick, 2013; Beydoun *et al.*, 2014; Gottesman *et al.*, 2017; Abell *et al.*, 2018). Higher education, physical and social activity, cognitional engagement and a diet that promotes cardiovascular health have been associated with decreased risk for AD (Saczynski *et al.*, 2006; Sando *et al.*, 2008; Morris *et al.*, 2015; Yates *et al.*, 2016; Stephen *et al.*, 2017).

2.1.3 Alzheimer's disease diagnosis and treatment

At the moment, clinical AD diagnosis, as well as diagnosis of other forms of dementia, is based on neurophysiological and cognitive testing and information about the individual's family and medical history. Dementia itself is usually rather easy to diagnose, however, identifying the exact disease that causes the dementia is impossible with only these methods. Besides, dementia is usually detected at a very late stage, when neurodegeneration and cognitive impairment are already far progressed, which also decreases the efficacy of possible treatments (Villemagne *et al.*, 2013). Promising biomarkers for AD diagnosis have been identified, however, they are not yet in standard use.

A biomarker is a biological finding that indicates a specific condition or disease and can be measured reliably. In AD, the most studied biomarkers are β-amyloid and tau accumulation, brain tissue atrophy and decreased glucose metabolism in the brain. Increased amounts of tau and β-amyloid proteins can be detected in cerebrospinal fluid (CSF) in AD already at the early MCI stage. Also increased peripheral blood plasma tau level is a strong marker candidate for AD. (Olsson *et*

al., 2016) Another approach for AD biomarkers is to measure the pathophysiological changes directly in brain tissue via imaging. Magnetic resonance imaging (MRI) can be used to detect AD-associated brain atrophy, in other words, changes in brain tissue morphology and volume. With positron emission tomography (PET) it is possible to measure β-amyloid burden and changes in glucose metabolism in the brain. (Rathore *et al.*, 2017) Both imaging and CSF/blood biomarkers for AD are already in wide use in medical and clinical research, however, there are still a few obstacles to overcome before they can be utilised in clinical diagnostics worldwide. There are no standardised methods for measuring CSF and blood biomarkers for AD, which means that the analysis results may vary between research centres. Imaging, like PET and MRI, on the other hand, requires very specific and expensive equipment and expertise and may also be too time-consuming for clinical patient screening. In addition, most of these biomarkers can be detected only at a quite late stage in AD pathogenesis, whereas early diagnosis would be preferable (El Kadmiri *et al.*, 2018).

All the currently approved drug treatments for AD target towards alleviating the symptoms of neurodegeneration. They regulate the amounts of neurotransmitters in synapses and enhance signal transmission between neuronal cells in the brain. Memantine is the only agent that can also slow down neurodegeneration by decreasing harmful extrasynaptic glutamatergic activity in some cases of moderate to severe AD (Wang and Reddy, 2017). None of the current treatment methods can stop the progressive neurodegeneration in AD. Thus, the benefits of these therapies are only temporary. In recent years, a vast number of clinical trials on new AD drug therapies have failed due to low treatment efficiency. It is difficult to find effective treatments for AD since the mechanisms causing the disease are still unclear and current diagnostic methods are inadequate for accurate AD detection. (Alzheimer's Association, 2019) Recruiting AD patients in clinical trials is complex as the disease starts to progress tens of years before it can be diagnosed, and the most promising results may be achieved if treatments are started already at a very early stage of AD progression. Thus, predictive biomarkers for AD would be invaluable for both clinical AD research and clinical diagnostics. (Winblad *et al.*, 2016)

2.2 Epigenetics and Alzheimer's disease

2.2.1 Epigenome

The epigenome, by definition, refers to something that is "on the genome", which is fitting as the epigenetic mechanisms are relatively dynamic modifications that bind on the DNA. Inside the cells, DNA is packed into small basic units, called nucleosomes, around histone proteins. The core histones include four different proteins; H2A, H2B, H3 and H4 that form a complex of eight histones, two of each subunit and bind the DNA around them. The nucleosomes, DNA and histone proteins, together with other non-histone chromosomal proteins form tightly condensed chromatin that is packed into a certain number of chromosomes. The number of chromosomes varies between species. Chromatin may be in more loosely packed euchromatin or more condensed heterochromatin forms. Heterochromatin usually contains only a few genes and it is especially concentrated in certain functional parts of the chromosomes, like centromeres and telomeres. However, the packing of chromatin is dynamic and can be changed from silenced tightly condensed to more active and loosely packed forms, and vice versa. (Alberts et al., 2008)

Epigenetic mechanisms regulate the activity of the genome and gene transcription in two ways: by locally regulating the packing of the DNA, making it more or less open for gene reading, or directly altering the DNA's affinity to certain regulatory enzymes and other proteins that initiate or prevent transcription of the genes. The epigenetic mechanisms include methylation of certain nucleotides in the DNA, post-translational histone modifications and a variety of non-coding RNAs (ncRNA), as illustrated in **Figure 1**. All the epigenetic mechanisms are required for the normal functioning of cells as well as cell differentiation, development and division. Thus, defects in these systems often cause problems in cell homeostasis and may even lead to severe diseases. Epigenetic changes or distinct epigenetic patterns and profiles have been associated with several human diseases. In cancers, for example, the epigenetic changes are often genome-wide and promote cancer cell survival and proliferation. During the past 20 years, it has become evident that epigenetic mechanisms play important roles in almost all known diseases, from autoimmune diseases to cardiovascular problems and neurological disorders, like AD. (Urdinguio, Sanchez-Mut and Esteller, 2009; Brooks et al., 2010; Dorn et al., 2019)

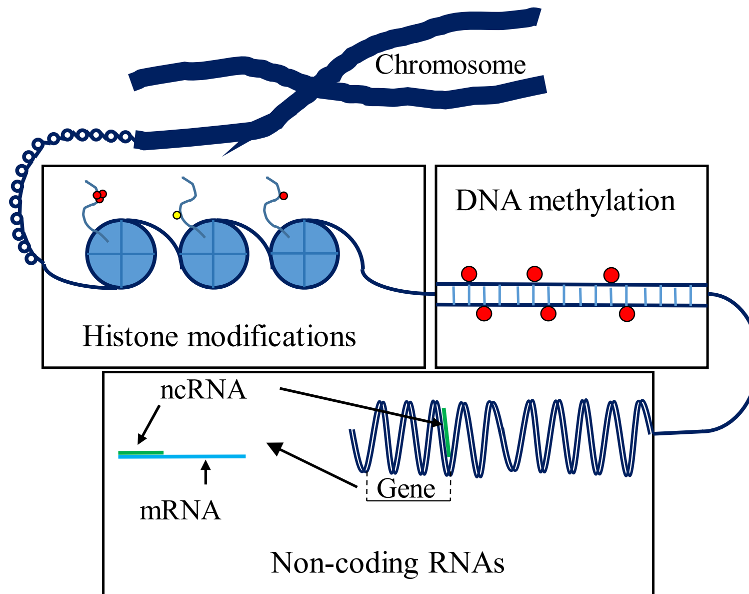


Figure 1. The epigenetic mechanisms. The basic epigenetic mechanisms include post-translational histone modifications, DNA methylation and a variety of non-coding RNAs. Histone modifications (methylation, acetylation etc.) that attach to the tails of the core histone proteins alter the packing and activity of the genome. DNA methylation also has an important role in regulating the packing of the genome but it can also locally modify DNAs affinity towards protein complexes that are involved in gene transcription. Non-coding RNAs regulate gene transcription by attaching to DNA or bind to mRNA products and control downstream processing, among various other mechanisms.

As the epigenome is tightly intertwined with the genome, they function as a single unit controlling each other's structure. Thus, genetic variation is a major cause of epigenetic variation between individuals and populations. Such variation can be seen, for example, when comparing the epigenomes of monozygotic and dizygotic twin pairs. Monozygotic twins share identical DNA and also their epigenomes are initially practically indistinguishable, whereas dizygotic twins have more differences in both their genomes and epigenomes. However, the epigenomes of DZ twins are more different also because they originate from two separate zygotes, whereas MZ twins originate from a single zygote. Each zygote has a unique epigenetic profile. (Kaminsky *et al.*, 2009) Genetic variation may influence the epigenetic mechanisms in a variety of different ways. For example, genes encoding enzymes that regulate the epigenetic mechanisms may have different variants with varying efficiencies or affinities towards certain genomic regions (Potter *et al.*, 2013). Epigenetic profiles may be affected by genetic variation at the sites where a modification binds to the DNA. However, genetic variants can also have an impact on more distant epigenetic features within the chromosome or even in other chromosomes. (Ng *et al.*, 2017)

The epigenetic mechanisms are both stable and dynamic. Epigenetic profiles are partly inheritable from cell to cell in mitotic cell division as well as from a parent to a child. However, epigenetic mechanisms also respond to internal and external stimuli, which causes changes and epigenetic variation. For example, nutrition has a major impact on the epigenome and it can even alter the risk for certain diseases, like cancers, via changes in the epigenetic mechanisms (Sapienza and Issa, 2016). Even though epigenetic profiles are partly heritable, a major epigenetic reprogramming occurs in the germline and fertilized egg cells. In this phenomenon, the epigenome resets to allow the cells to start to develop and differentiate into functional cells as parts of organs and organisms. (Heard and Martienssen, 2014) After all, cells in different organs and tissues have tissue-specific epigenetic profiles and gene expression profiles. Different cell types within an organism have very distinctive functions from one another, thus, they also require different genes to be active to fulfil their tasks. Since epigenetic mechanisms are susceptible to environmental changes and can also be heritable, some epigenetic changes caused by environmental factors can be carried through generations. However, this kind of transgenerational epigenetic inheritance has been mostly studied in animals whereas findings from human studies require further validation (Horsthemke, 2018; Perez and Lehner, 2019).

2.2.2 Epigenetic mechanisms

2.2.2.1 DNA methylation

DNA methylation means the addition of methyl groups to the base molecules of the nucleotides in DNA. Two of the four bases in DNA, cytosine and adenine, can be methylated (Wu *et al.*, 2016). Cytosine methylation is far more common and more extensively studied. Methyl group binds to the fifth atom, which is carbon, in the six-atom ring of a cytosine. DNA methylation does not affect cytosines binding affinity to other bases; methylated cytosines pair with guanines the same as unmethylated ones in the double-stranded DNA molecule. However, DNA methylation regulates the packing of the chromatin and changes DNA’s affinity to certain DNA-binding enzymes and proteins, like transcription factors. (Alberts *et al.*, 2008)

The most common form of DNA methylation is the methylation of CpG dinucleotides, where a methyl group is added to a cytosine that is followed by a guanine in the DNA sequence. However, also other cytosines can be methylated, which occurs especially in PSCs and brain tissue (Ziller *et al.*, 2011; Guo *et al.*, 2014). The CpG sites are often concentrated on certain functional regions, like gene promoters and enhancers, which regulate gene activity and transcription. CpG rich

regions are called CpG islands, shores or shelves, depending on their location and distance to the nearest gene. (Alberts *et al.*, 2008)

At a local level, methylation of a CpG rich gene promoter usually silences the gene, in other words, prevents gene transcription. About 60 % of human genes have CpG rich promoters and most of them remain unmethylated even if the gene is otherwise silenced. (Weber *et al.*, 2007; Hill *et al.*, 2018) When a gene is silenced CpG rich promoter is not necessary methylated, however, repression may occur due to Polycomb system-mediated addition of H3K27me3 histone marks for example (Lynch *et al.*, 2012). The Polycomb group family of proteins consists of Polycomp repressive complexes 1 and 2 that are important for regulating gene expression, especially during development (Moritz and Trievel, 2018). Gene bodies can also be relatively highly methylated, which occurs especially in highly transcribed genes in mammals, indicating that gene-body DNA methylation does not prevent transcription (Zemach *et al.*, 2010). DNA methylation is tissue-specific; however, only about 20 % of the detected CpG sites in the human genome are differentially methylated between normal tissues. These dynamic methylation differences that occur during development are associated with genomic regions that have tissue-specific functions, for example, genes that are expressed only in specific tissues. (Ziller *et al.*, 2013; Schultz *et al.*, 2015)

On a more widespread level, DNA methylation plays an important role in genomic imprinting and permanent repression of transposable elements. In genomic imprinting, DNA methylation, in cooperation with histone modifications and non-coding RNAs, can condense and silence large genomic regions or even whole chromosomes (Delaval *et al.*, 2007; Ideraabdullah, Vigneau and Bartolomei, 2008). The inflicted genes are thus expressed only from one of the two alleles found in diploid genomes. The human genome, for example, is diploid meaning that there are two copies, i.e. alleles, of each gene. One copy is inherited from the mother and the other from the father. Transposable elements are DNA sequences of varying lengths that can move in the genome. While activated, they decrease the genomic stability and can cause different kinds of mutations. They offer an important platform for evolutionary development; however, mutations can often be harmful to the organism. Most of the genomic regions containing transposable elements are heavily methylated, which keeps them inactive. (Bourque *et al.*, 2018)

Three active DNA methyltransferase enzymes, DNMT1, DNMT3A and DNMT3B, exist in human cells and catalyse the addition of methyl groups into cytosines in DNA. DNA methylation patterns can be inherited in both mitotic and meiotic cell division, where DNMT1 is responsible for copying the pattern to the replicated DNA strand. The other two enzymes, DNMT3A and DNMT3B, can add methyl groups later and alter the cells' responses to different external and internal stimuli. The methyl groups attached to cytosine nucleotides can also be converted to

hydroxymethyl groups. Ten Eleven Translocation (TET) family proteins, TET1, TET2 and TET3, can oxidate the methyl group. This reaction is crucial for active demethylation of the cytosines and it is required in many basic biological processes, like cell differentiation. (Rasmussen and Helin, 2016)

2.2.2.2 Histone modifications

Post-translational histone modifications are small molecule groups that are covalently added to the histone proteins. Histones have N-terminal amino acid side chains pointing out from the nucleosome. Modifications are usually attached into specific amino acids in these side chains, however, there can also be modifications closer to the nucleosomes core. Common post-translational modifications of histone side chains are mono-, di- and trimethylation of lysines and arginines, acetylation of lysines, phosphorylation of serines, threonines and tyrosines, and ubiquitylation and sumoylation of lysines. (Bannister and Kouzarides, 2011)

The histone modifications are usually dynamic and different enzymes mediate the addition and cleavage of the small molecules to and from the histone tails. For example, histone acetyltransferases, also known as HATs, add acetyl groups to lysines, whereas histone deacetylases, i.e. HDACs, cleave them off. (Grunstein, 1997) The modifications occur usually in only certain amino acids in the histones' side chains and have specific roles in chromatin functions and activation or inactivation of genomic regions. Acetylation of lysines in the histones side chain, for example, histone three lysine nine (H3K9ac), is usually a mark for active chromatin (Koch *et al.*, 2007), whereas methylation can be associated with either active or repressed chromatin, depending on the location and level of methylation. For example, monomethylation of histone three lysine nine (H3K9me) is found on active chromatin, and trimethylation of the same amino acid (H3K9me3) is associated with repressed chromatin. (Barski *et al.*, 2007)

Combining histone modification data from Chromatin Immunoprecipitation and Sequencing (ChIP-seq) with information about transcription factor binding sites and chromatin accessibility has been crucial in identifying histone modification combinations that mark different activity states of the chromatin. For example, H3K4me1 together with H3K27ac mark active gene enhancers, the combination of H3K4me1, H3K4me2 and a high ratio of H3K4me3 marks active promoters and H3K36me3 together with high RNA polymerase II activity indicate actively transcribed genome regions (Jiang and Mortazavi, 2018). With this information, it is possible to detect new regulatory regions in the genome.

There are a relatively low number of reports on specific histone modification alterations in other human diseases than cancers. This is partly because DNA methylation has been easier to study, which is why many groups focus on that rather

than histone modifications. Histone modifications are more dynamic than DNA methylation, thus, there may not be many stable histone markers in other human diseases, whereas in cancers the entire genetic and epigenetic regulation is harnessed to ensure survival and proliferation of the cancer cells. HDAC inhibitors have been approved for clinical treatment of certain lymphomas. (Jones, Issa and Baylin, 2016; T. Muka *et al.*, 2016; Taulant Muka *et al.*, 2016; Wen *et al.*, 2016) However, as histone modifications are important in regulating gene expression, characterising disease-specific histone profiles can reveal new insights to disease mechanisms and even new therapeutic targets in the future.

2.2.2.3 Non-coding RNAs

Non-coding RNAs are transcribed from the genome, however, instead of being translated into proteins they remain as RNAs and regulate the activity of the genome and gene expression via several different mechanisms. ncRNAs can, for example, impact the packing of the chromatin by regulating local DNA methylation and histone modifications, affect gene transcription by regulating enhancer-promoter interactions or control the levels of messenger RNAs transcribed from different genes (Rothschild and Basu, 2017; Ransohoff, Wei and Khavari, 2018). The length of the ncRNAs varies from 22 nucleotide-long microRNAs to more than 200 nucleotide-long long non-coding RNA (lncRNAs) and 5070 nucleotide-long 28S ribosomal RNA and everything in between. Non-coding RNAs are categorized into different groups according to their length and function in the cells. (Mattick and Makunin, 2006)

There are at least ten different RNA species categorised. Some of them, like transfer RNA (tRNA) and ribosomal RNA (rRNA) have already been detected tens of years ago, however, the number of known ncRNAs has been growing quickly during the recent years. The total number of ncRNAs is still unknown and new molecules are detected at a fast rate due to development in genome sequencing and data mining tools. (Sharp *et al.*, 1985; Lafontaine and Tollervey, 2001; Uszczynska-Ratajczak *et al.*, 2018) Non-coding RNAs are important in regulating normal development and function of the cells and organisms. Defects in ncRNA networks and functions have been detected in several human diseases, especially in cancers but also in AD (Millan, 2017; Anastasiadou, Jacob and Slack, 2018).

2.2.3 Epigenome and environment

The environment starts to modify the epigenome already before birth. The mother's diet and lifestyle have a significant impact on the developing offspring's epigenome. The epigenome of the offspring during pregnancy is especially vulnerable due to the

epigenetic reprogramming and high cell division rate. How the prenatal environment and stress can impact epigenome has been studied in animal models for a large part, however, there is some evidence from human studies as well. A classic example of such stress-induced impact on the epigenome in humans is the Dutch famine study that inspected the effect of inadequate maternal nutrition, during winter 1944-45, to foetal growth and the progeny's increased risk for several diseases in later life (Lumey, 1992; Roseboom, de Rooij and Painter, 2006). High-fat diet during pregnancy alters the expression of *leptin* and *adiponectin* genes in the offspring via epigenetic changes. These changes induce metabolic syndrome-like symptoms that persist in the mice for several generations even if the diet of the offspring is normalised. (Milagro *et al.*, 2009; Masuyama and Hiramatsu, 2012; Masuyama *et al.*, 2015). Inadequate nutrition of pregnant rats and mice also alters epigenetic profiles in glucocorticoid receptor genes of young and adult offspring (Lillycrop *et al.*, 2005; Burdge *et al.*, 2007).

Prenatal maternal stress and exposure to synthetic glucocorticoids, which are often used to model stress reactions in animals, have also been linked to distinctive epigenetic patterns in the offspring and even elevated risk for certain diseases and medical conditions during later life (Mueller and Bale, 2008; Oberlander *et al.*, 2008; Crudo *et al.*, 2013). Prenatal environment and maternal stress have a significant impact on nervous system development as well as immune and cardiovascular system functions and metabolism of the offspring (Cao-Lei *et al.*, 2016). In addition to maternal stress and diet, also certain behaviour, like the quality of maternal care over an infant or childhood abuse, and toxins can cause changes in the offspring's epigenome in both humans and animals (McGowan *et al.*, 2009, 2011; Faulk and Dolinoy, 2011).

Even though the epigenome in later life is not as susceptible to external stimuli as during prenatal development, the environment and lifestyle factors continue to modify the epigenome throughout adult life. Furthermore, in later life, the epigenetic changes usually occur in a subset of already differentiated cells, whereas during prenatal development all cells in the embryo/foetus may be affected. Genome-wide changes in histone H3 methylation have been linked to obesity and type 2 diabetes in adults (Jufvas *et al.*, 2013). Caloric restriction can hinder age-related DNA methylation changes in the adult mouse hippocampus (Chouliaras *et al.*, 2012). High caloric diet impacts DNA methylation also in human muscle cells (Jacobsen *et al.*, 2012). Global methylation levels in the repetitive DNA regions decrease during ageing of the individual, which is probably due to reduced DNMT1 activity (Casillas *et al.*, 2003; Bollati *et al.*, 2009). A healthy diet and physical activity have been linked to decreased risk for several cardiovascular, metabolic and neurophysiological diseases, which is most likely mediated by epigenetic mechanisms that alter the gene expression in all tissues (Kaliman *et al.*, 2011)

Exposure to toxic compounds can cause epigenetic changes. For example, tobacco use has a long-lasting genome-wide impact on DNA methylation profile, which can be detected even in peripheral blood (Breitling *et al.*, 2011). Alcohol can also induce increased histone acetylation in the brain of pregnant mice and the gestating foetuses (Mews *et al.*, 2019).

2.2.4 Epigenetic alterations in Alzheimer's disease

Genetic factors explain a large part of the LOAD risk, however, the mere existence of monozygotic twin pairs discordant for AD proves that also other factors contribute to the disease risk. These external factors, for example, diet, physical cognitive and social activity, medical history and lifestyle choices may cause specific changes in the epigenetic mechanisms that are involved in AD pathogenesis. The epigenetic mechanisms are important for cognitive functions, like memory formation and learning, and specific epigenetic changes occur in the brain during ageing (Zovkic, Guzman-Karlsson and Sweatt, 2013; Sen *et al.*, 2016)

Many recent studies have already discovered specific epigenetic changes associated with AD in brain tissue and there are also a few reports about AD-associated DNA methylation marks in peripheral blood. However, the findings on peripheral blood DNA methylation markers for dementia and AD have been inconsistent due to limitations in the study designs, like low power or too challenging case-control comparisons. (Fransquet *et al.*, 2018) DNA methylation marks associated with AD have been analysed in different brain regions. **Table 1** puts together differentially methylated genes that have been associated with AD to date. All the previous genome-wide DNA methylation analyses have been conducted with microarrays that detect methylation only at a predetermined set of CpG sites. In the targeted DNA methylation analyses, bisulphite pyrosequencing has been the most common method. While many of the differentially methylated genes have unknown functions and contribution to AD pathogenesis, some genes have known roles in nervous system functions. The wide range of findings from these studies shows that AD is a complex disease also from the epigenetics' point of view. However, AD-associated differential methylation in some of the genes, for example, ankyrin 1 (*ANK1*), WD repeat domain 81 (*WDR81*) and homeobox A3 (*HOXA3*), has been confirmed by several independent research groups, indicating that these genes are most likely involved in the disease process (De Jager *et al.*, 2014; Lunnon *et al.*, 2014; Semick *et al.*, 2019; Smith, Smith, Burrage, *et al.*, 2019; Smith, Smith, Pishva, *et al.*, 2019).

At first, the studies on DNA methylation perturbations in AD focused on the key genes amyloid-beta precursor protein (*APP*), *PSEN1* and microtubulin associated protein tau (*MAPT*) that are known to be important in the pathogenesis according to previous studies. Even though small-scale studies reported interesting findings of

distinctive DNA methylation patterns in for example *APP* gene, these findings could not be confirmed in a larger cohort (West, Lee and Maroun, 1995; Barrachina and Ferrer, 2009). Some groups have also detected differences in global genome-wide methylation levels in different brain regions in AD. However, findings from different groups have been inconsistent: some report decreased and others increased global methylation or hydroxymethylation in the brain (Mastroeni *et al.*, 2010; Chouliaras, Mastroeni, *et al.*, 2013; Coppieters *et al.*, 2014). This indicates that the possible brain DNA methylation changes in AD are relatively small and most likely are restricted into specific short genomic regions.

Alzheimer's disease-associated changes in histone modifications have been remarkably less studied than DNA methylation. DNA methylation is easier to quantify than histone markers, at least for now, and histones are not as stable as DNA methylation during processing and storage of the post-mortem brain tissue samples that the AD research material often consists of. Some groups have reported that HDAC inhibitors may have a protective role against AD and cognitive impairment and the basic HDAC levels increase by age in human and mice brain (Fischer *et al.*, 2007; Gräff *et al.*, 2012; Ricobaraza *et al.*, 2012; Chouliaras, van den Hove, *et al.*, 2013; Rumbaugh *et al.*, 2015). Gjoneska and her group (2015) studied post-translational histone marks in an AD mouse model and reported an increase in active chromatin marks H3K27ac and H3K4me3 associated with genes involved in immune responses. In recent years, many groups have reported differences in ncRNAs in AD. Changes in ncRNA functions and expression levels have been associated with the pathophysiological molecular and cellular changes in AD, even though their actual contribution to the disease outcome is still unclear. (Millan, 2017)

There is growing evidence that epigenetic mechanisms play an important role in AD pathogenesis. The major limitation of the studies on the epigenetic changes in AD so far is that the analysis of human post-mortem brain tissue samples cannot reveal whether epigenetic alterations are a cause or a consequence of the disease and at which point of the disease progression they appear. Animal models are valuable for studying longitudinal changes associated with diseases, however, the mechanisms between human diseases and animal disease models are not necessarily identical. Besides, the most commonly used AD animal models rely on inducing AD-like pathological changes in animals' nervous system via mutations in genes that have been associated with the disease, like *APP*, *MATP1* or *PSEN1*. Thus, they are better at modelling familiar AD instead of the sporadic disease that contributes to more than 90 % of the cases in humans. (Götz, Bodea and Goedert, 2018; Zhang *et al.*, 2020) Another possibility is to study epigenetic alterations in hPSC models, which enable the analysis of longitudinal epigenetic changes in human tissues during neuronal cell development and even different stages of neurodegeneration.

Table 1. Genes located closest to differentially methylated genomic regions in Alzheimer's disease.

Genes	Tissue	Analysis method	Study/ reference
S100A2	Frontal cortex	MethyLight PCR	(Siegmund et al., 2007)
BDNF, COX2, CREB, NFKB, SYP	Frontal cortex	MSRE-PCR	(Rao et al., 2012)
TMEM59	Frontal cortex	Microarray	(Bakulski et al., 2012)
SORBS3	Entorhinal, temporal and prefrontal cortex Frontal cortex	Microarray, pyrosequencing MethyLight PCR	(Sanchez-Mut et al., 2013) (Siegmund et al., 2007)
SPTBN4 *TBXA2R*	Frontal cortex	Microarray, pyrosequencing	(Sanchez-Mut et al., 2013)
DUSP22 *IGFBP7*	Hippocampus	Microarray, pyrosequencing	(Sanchez-Mut et al., 2014)
ANK1	Entorhinal, temporal and prefrontal cortex	Microarray, pyrosequencing	(Lunnon et al., 2014)
ANK1, BIN1, CDH3, FOXK1, HMHA1/ABCA7, HOXA3, KDM2B, ITPRIPL2, PCNT/DIP2, RHBDF2, SLC2A1, SPG7/RPL13, WDR81/SERPINF1/SERPINF2	Entorhinal, temporal and prefrontal cortex	Microarray	(De Jager et al., 2014)
TREM2	Superior temporal gyrus	Microarray, pyrosequencing	(Smith et al., 2016)
NCAPH2/LMF2	Peripheral blood	Microarray, pyrosequencing	(Kobayashi et al., 2016)
NCAPH2/LMF2	Peripheral blood leukocytes	Pyrosequencing	(Shinagawa et al., 2016)
MOV10L1, B3GALT4, DUSP6, TBX15, HLA-J/ZNRD1-AS1, PRDM16, ELOVL1, RIBC2/SMC1B, KLK7, TRIM6, FBRSL1, AGPAT1/RNF5/RNF5P1, VAX2, PPT2/PPT2-EGFL8/PRRT1, C10orf105/CDH23, KIF25, NRG2, RNF39, CMYA5, TNXB, NAV2, TAP2, ZNF177/ZNF559-ZNF177, FLOT1/IER3	Superior temporal gyrus	Microarray	(Watson et al., 2016)
BMAL1	Frontal cortex	Microarray	(Cronin et al., 2017)
HOXA cluster	Prefrontal cortex and superior temporal gyrus	Microarray	(Smith et al., 2018)
MCF2L, ANK1, MAP2, LRRC8B, STK32C, S100B, HOXA3, APP, ADAM17	Frontal and temporal cortex, separated neurons and glial cells	Microarray	(Gasparoni et al., 2018)
ANK1, WNT5B, FBXL16, ALLC, ARID5B	Entorhinal cortex	Microarray	(Smith, Smith, Pishva, et al., 2019)
ANK1	Entorhinal cortex, superior temporal gyrus, cerebellum	Pyrosequencing	(Smith, Smith, Burrage, et al., 2019)
ANKRD30B, WDR81/SERPINF1/SERPINF2, ANK1, DUSP22, JRK, NAPRT, CSNK1G2	Prefrontal cortex, entorhinal cortex, hippocampus	Microarray	(Semick et al., 2019)

2.2.5 Twin studies – a unique perspective into epigenetic variation

Monozygotic (MZ) twins have identical genomes and yet there are phenotypic differences between the twins. The epigenetic profiles of MZ twins are practically indistinguishable at first and start to deviate from one another only by time. (Fraga *et al.*, 2005) Monozygotic twins are of same-sex and age and are usually exposed to the same environmental factors before birth and during early life. However, differences in environment and lifestyle in later life together with pure stochastic variation introduce deviation to their epigenetic profiles, phenotypes and risk for certain diseases as well. Dizygotic (DZ) twins share on average 50 % of the same genetic pool and have similar living environments during early life. As genetic and epigenetic features are intertwined, the epigenetic variation between individuals is for a large part determined by genetic differences. Thus, twin studies are beyond comparison when characterising epigenetic differences that occur by time and contribute to different phenotypic profiles.

Twin studies can be utilised in studying how somatic mutations induce epigenetic changes and analysing epigenetic differences between twin pairs discordant for certain trait or disease gives unique information about the epigenetic mechanisms' and external factors' contribution to the trait or disease of interest. (Kim *et al.*, 2016) Power estimation of an epigenome-wide association study using samples from discordant twin pairs has demonstrated the unbeatable sensitivity of a twin study design in comparison to basic case-control design. To reach certain power, a traditional case-control study may require up to ten times more samples than a disease-discordant twin study. However, the number of samples required is also dependent on the desired power level as well as the heredity of the disease of interest among other factors. (Li *et al.*, 2018) Furthermore, the scarce number of existing twin sample cohorts is probably the most significant limiting factor for the usage of this kind of study designs in epigenetic research.

2.3 Unravelling the epigenome – DNA methylation analysis via bisulphite sequencing

2.3.1 Genome-wide DNA methylation analysis

As DNA methylation is by far the most studied epigenetic mechanism, several different methods have been developed for DNA methylation analysis. All the methods have certain advances and limitations in comparison to each other, which makes all of them important for epigenetic research. Genome-wide analysis methods are very popular in discovering novel DNA methylation markers for diseases and

other conditions, whereas targeted methods are often utilised in validating these markers. The most popular methods quantify methylation at single-nucleotide resolution, i.e. at each methylation site, however, some methods that quantify methylation on larger regions are also utilised in some cases. (Bock *et al.*, 2010; Kurdyukov and Bullock, 2016)

The most commonly used genome-wide methods are bisulphite sequencing and methylation-sensitive bead microarrays. Both of these methods detect methylation at single-nucleotide resolution. The latest version of Illumina microarrays can quantify methylation at up to 850,000 of the 28 million CpG sites found in the human genome. These predetermined CpG sites are located in promoters and other functionally interesting regulatory regions of the genome. Microarrays provide highly reproducible data at low costs and the data analysis is standardised and straightforward. However, microarrays often cover only a few methylation sites in a specific regulatory region, which can give biased information on the whole region's methylation level. In addition, it is not possible to discover new methylation markers with microarrays as they target only specific methylation sites. (Pidsley *et al.*, 2016) Whole-genome bisulphite sequencing (WGBS) can detect methylation at each of the 28 million CpG sites found in the human genome, in principle. In practice, however, analysing all CpG sites with adequate certainty would increase the sequencing costs a great deal, making it usually more sensible to increase the number of samples instead of deeper sequencing. WGBS also produces vast amounts of data, which complicates data management and analysis. (Ziller *et al.*, 2015)

There are also other methods for genome-wide DNA methylation analysis that rely on chromatography and/or mass spectrometry or enzyme-linked immunosorbent assay (ELISA). These methods have lower resolution than bisulphite sequencing or microarrays as they quantify mean methylation in larger DNA regions rather than single nucleotides. Since there are commonly used protocols and even commercial kits for these methodologies, they can still be utilised to study changes in global whole-genome methylation levels when cultured cells are treated with agents that impact DNA methylation or demethylation on a large scale. (Kurdyukov and Bullock, 2016)

2.3.2 Bisulphite sequencing

DNA methylation analysis via bisulphite sequencing is based on two important steps: bisulphite conversion, where the unmethylated cytosines in DNA are chemically converted into uracils, and sequencing, where the base sequence of the genome or the region of interest is characterised. In bisulphite conversion, only unmethylated cytosines react with the conversion reagent, whereas the methyl group protects

methylated cytosines (Frommer *et al.*, 1992). Hydroxymethyl also protects the cytosine nucleotide from bisulphite conversion reagent similar to the methyl group. Thus, it is not possible to distinguish methylated and hydroxymethylated cytosines from each other with bisulphite sequencing. Converted uracils are replaced with thymines when the samples are amplified in polymerase chain reactions (PCR). When the DNA base sequence in the sample is sequenced, the methylation per cent of a CpG site can be calculated from the relative amount of cytosines and thymines detected in any specific location.

Bisulphite conversion is utilised in both genome-wide and targeted DNA methylation analyses. In targeted DNA methylation analysis a short genomic region of interest is amplified in PCR with target-specific primers after bisulphite conversion of the genomic DNA, and methylation levels are quantified via pyrosequencing, for example. (Bassil, Huang and Murphy, 2013) In genome-wide bisulphite sequencing, the whole genomic DNA is usually PCR amplified after conversion and the nucleic acid sequence is analysed with a high-throughput sequencing platform, like Illumina HiSeq or NovaSeq. Some protocols do not require PCR amplification since PCR can introduce bias in the methylation results (Olova *et al.*, 2018). In comparison to the DNA methylation-sensitive microarrays, the greatest strength of bisulphite sequencing is that it can detect new methylation sites, whereas microarrays target only a limited number of sites that are already known (Kurdyukov and Bullock, 2016).

WGBS produces vast amounts of data, especially if the number of samples is large. Only 2 % of the human genome encodes proteins and for a large part, the function of the non-coding genome remains unknown. Thus, analysing the whole genomes from large sample sets produces huge amounts of data that is difficult to store and analyse, while most of the information comes from the unknown non-coding regions or may not have any methylation sites. Sequencing of WGBS libraries is also still quite expensive. Reduced Representation Bisulfite Sequencing (RRBS) was developed to analyse the CpG rich regions of the genome, like gene promoters. The RRBS workflow, including sequencing library preparation and data analysis, is illustrated in **Figure 2**. It utilises a restriction digestion enzyme, MspI, which cuts the DNA between the two cytosine nucleotides in CCGG sequence sites. Thus, each DNA fragment in the sequencing library sample contains methylation information from at least one CpG site. The digested DNA fragments are size selected and amplified during the library preparation, which increases the amount of CpG rich DNA in the sample. More CpG rich regions in the library sample mean that after sequencing there will be more data from the coding regions and promoters that regulate gene expression. In comparison to WGBS, sequencing of RRBS libraries is less expensive, and the datasets are a lot easier to handle, which is why RRBS is frequently used even though the sequence

information from large parts of the genome is lost. (Meissner *et al.*, 2005; Boyle *et al.*, 2012)

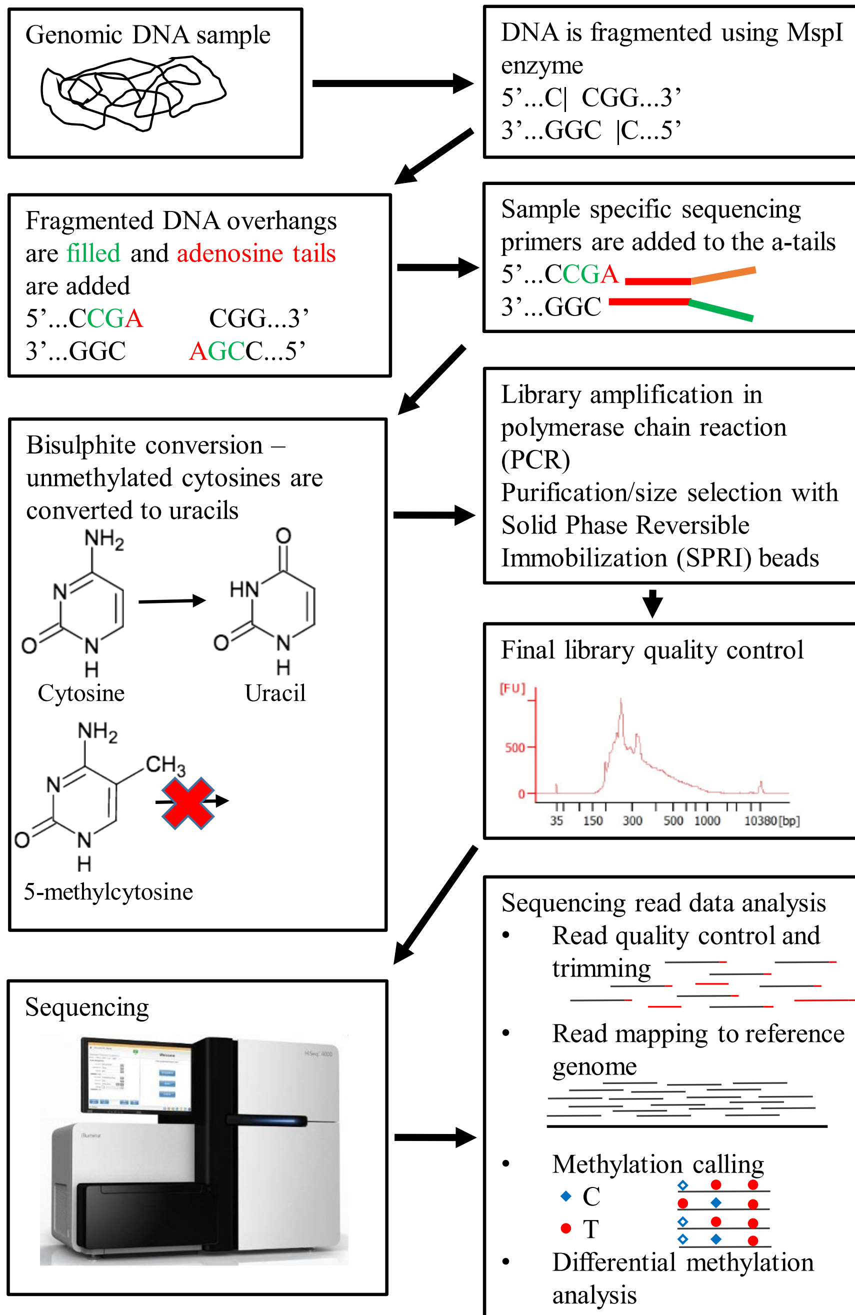


Figure 2. Reduced Representation Bisulfite Sequencing workflow.

2.3.3 Bisulphite sequencing data analysis

After sequencing the RRBS or WGBS libraries, the next step is data analysis. The same as with all high-throughput sequencing experiments, bisulphite sequencing data analysis starts with sequencing read quality checking and trimming and discarding low quality reads. Sequencing read is the nucleotide sequence of a single DNA fragment in the library sample. The pre-processed reads are then aligned and mapped to a reference genome. Bisulphite sequencing data requires a specific mapping software, for example, Bismark, since the DNA has been bisulphite treated and all unmethylated cytosines are converted to thymines and DNA sequence complexity is reduced. Bisulphite conversion efficiency is controlled by adding unmethylated lambda DNA to all library samples. The mapping software also goes through methylation calling, i.e. quantifies methylation in each methylation site, and outputs the methylation data, including at least the sites' location coordinates in the genome, methylation values, and sequencing coverages. Sequencing coverage is the number of reads covering a single site of interest, or in other words, how many times a specific site was sequenced from the library sample. (Krueger and Andrews, 2011)

The methylation data is then exported to a differential methylation analysis software. There is a wide variety of software available for genome-wide bisulphite sequencing data analysis, for example, methylKit, RnBeads, RadMeth and dmrseq (Akalin *et al.*, 2012; Dolzhenko and Smith, 2014; Korthauer *et al.*, 2019; Müller *et al.*, 2019). Bisulphite sequencing data analysis is not yet standardised and new methods are being developed all the time. The differential methylation analysis software packages differ from one another in how they handle the data. Some have been developed for small sample sets whereas others are optimised for tens or even hundreds of samples. Some programmes calculate methylation differences in individual sites and others focus on regional methylation differences. The most important difference between the programmes is, how they measure the significance of the differentially methylated sites and correct the results for multiple testing error. The statistical model each software uses also sets the limit to the number of confounding factors that can be included in the analysis. Thus, it is important to consider and test different analysis methods, which is the most suitable for specific data set and experimental design. (Akalin *et al.*, 2012; Song *et al.*, 2013; Korthauer *et al.*, 2019; Müller *et al.*, 2019)

2.4 Human pluripotent stem cells in disease modelling

2.4.1 Human pluripotent stem cells

Human pluripotent stem cells are an invaluable resource for regenerative medicine and disease modelling. Pluripotent stem cells have two defining characteristics: in theory, they can differentiate into all cell types and tissues found in the body, and they can proliferate indefinitely. There are two types of human pluripotent stem cells; embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs). Embryonic stem cells are collected from the inner cell mass of an embryo before implantation in the uterus (Thomson, 1998). Induced pluripotent stem cells, on the other hand, are reprogrammed from differentiated somatic cells in vitro and returned to the pluripotent state (Takahashi et al., 2007).

Pluripotent stem cells are already widely used to model the development of biological systems and diseases. Due to their ability to proliferate, PSCs can be cultured *in vitro,* and pluripotency enables the cells to be differentiated into any of the cell types found in the adult human body. Simpler model systems, like two-dimensional cultures of single-cell types differentiated from stem cells, have been around for over ten years. With more recently invented three dimensional (3D) stem cell-derived organoid cultures it is also possible to study the development of larger biological systems. For example, with cerebral organoids, it is possible to model the development and function of early human brain (Lancaster *et al.*, 2013; Lancaster and Knoblich, 2014).

A common feature for all stem cell applications is that the cells must be cultured *in vitro* in laboratory conditions. During culturing, the stem cell colonies are expanded to ensure enough cells for the application and passaged to provide the best environment for cell proliferation and to prevent spontaneous differentiation of the cells. However, prolonged *in vitro* culturing also provokes genetic and epigenetic changes in the PSCs. These changes may alter the behaviour of the cells by affecting their ability to differentiate and proliferate and compromise their suitability for research use. (Lund, Närvä and Lahesmaa, 2012)

2.4.2 Human pluripotent stem cell cultures and genomic stability

While PSCs can be maintained in *in vitro* cultures for tens of passages, *in vivo* they usually exist only transiently. Thus, the PSCs' ability to proliferate indefinitely is artificial and only seen in laboratory cultures. Prolonged culturing of PSCs has revealed a significant vulnerability in these cells as epigenetic reprogramming

exposes their genome to changes. (Heard and Martienssen, 2014; Weissbein, Benvenisty and Ben-David, 2014) Genomic aberrations occur on many levels in PSCs during *in vitro* culturing. Large karyotypic changes, e.g. the gain of chromosomes 12 and 17q, in ESC lines were reported as early as 2004 (Draper *et al.*, 2004). Additional large-scale amplifications have been detected in other chromosomes as well. The incidence rate that these changes occur seems to be similar both in ESC and iPSC lines at under 13 % of all lines. (Taapken *et al.*, 2011) An increased amount of smaller-scale copy number variations (CNVs) and single nucleotide polymorphisms (SNPs) have been detected in PSC lines as well. Interestingly, increased genetic variation has also been detected in genes that can have an impact on the cells' pluripotency and survivability, like BCL2 like 1 (*BCL2L1*), isopentenyl-diphosphate delta isomerase 1 (*IDI1*), Nanog homeobox pseudogene 1 *(NANOGP1*) and *DNMT3B.* (Chin *et al.*, 2009; Närvä *et al.*, 2010; Laurent *et al.*, 2011; Martins-Taylor *et al.*, 2011) The number of focal CNV is higher in iPSC lines than ESCs or somatic cells, which may be due to the reprogramming process or selection of somatic cells that were initially reprogrammed to iPSCs (Laurent *et al.*, 2011; Ji *et al.*, 2012; Young *et al.*, 2012).

A wide variety of epigenetic alterations have been detected in PSC lines *in vitro* as well. DNA methylation variation is seen on both protein-coding and non-coding regions. Local epigenetic aberrations may lead to changes in gene expression, whereas widespread deviations can cause loss of gene imprinting and even partial reactivation of an inactivated X chromosome (Enver *et al.*, 2005; Nazor *et al.*, 2012; Bar *et al.*, 2017; Weissbein *et al.*, 2017). The iPSCs' DNA methylation profile differs from ESCs' during early passages *in vitro*, however, the differences usually diminish at later passages. This variation may be caused by the reprogramming process or imperfect epigenetic reset from somatic cells to pluripotency. (Nishino *et al.*, 2011) Culturing conditions, like culturing media or oxygen level, may also cause changes to the epigenome (Allegrucci *et al.*, 2007; Nazor *et al.*, 2012). Genetic and epigenetic variance enables normal cells to react to external stimuli and alter their behaviour to survive and maintain the required functions. When the rate of these variations increases, the genetic and epigenetic profiles of PSCs start to resemble cancer cells (Harrison, Baker and Andrews, 2007; Yang *et al.*, 2008). When transplanted to mice, karyotypically abnormal human ESCs have elevated tumorigenic potential (Werbowetski-Ogilvie *et al.*, 2009). In a more recent study, Weissbein and his group (2019) identified intracellular signalling pathways, like Ras protein, Rho/Rho-associated coiled-coil containing protein kinase (Rho-ROCK), phosphatidylinositol-3-kinase/protein kinase B (PI3K-AKT) and Hippo, which influence the PCSs' survivability and tumorigenic potential in prolonged *in vitro* culture.

2.4.3 Human pluripotent stem cell models for Alzheimer's disease

Human pluripotent stem cell models offer new perspectives for studying chronic progressive neurodegenerative diseases like AD. Most of the pathophysiological changes in AD occur in the brain. Thus, it has been difficult to study the disease progression in human tissues, as only post-mortem samples have been available. With iPSC technologies, it is possible to produce neuronal cell lines and 3D tissues derived from AD patients and healthy individuals and compare cell and tissue development between the models. Stem cell-derived disease models have been widely used to study neurodegenerative disorders, like AD, Huntington's disease, amyotrophic lateral sclerosis and Parkinson's disease (Centeno, Cimarosti and Bithell, 2018). While the 3D organoid cultures are at the moment used to study organ development and patient-specific disease models, in the future they may enable building replacement tissues for therapeutic use (Madeline A. Lancaster and Knoblich, 2014).

The patient-derived disease models give new insights into the molecular and cellular disease mechanisms. Pluripotent stem cells have been utilised in a few AD studies already. In one of the earliest reports on the field, by Israel and his group (2012), iPSC derived neurons from familial and sporadic AD patients showed impaired β-amyloid and tau protein processing in comparison to controls. Alzheimer's disease-associated molecular disturbances have been detected in neuronal cells differentiated from AD patient-derived iPSCs (AD-iPSC) in other studies as well, even though most of them have focused on familial AD (Yagi et al., 2011; Muratore et al., 2014; Hossini et al., 2015; Yang et al., 2016). Also, 3D neuronal cell and tissue cultures differentiated from AD-iPSC have been detected to express phenotypes typical for AD, like the accumulation of β-amyloid and tau (Lee et al., 2016; Raja et al., 2016; Gonzalez et al., 2018).

2.5 Summary

Alzheimer's disease is one of the most important healthcare problems in the modern world. Despite extensive research, the factors leading to AD are not yet completely known and the tools for clinical AD diagnosis and treatment are inadequate. At the moment, AD can be clinically diagnosed at a very late stage, even 20-30 years after disease progression has started. At this point, current treatment methods can only alleviate the clinical symptoms temporarily before the patient's cognition starts to decline again. New methods that can detect AD at an earlier time point are needed urgently. If AD is diagnosed already at an early stage, the disease progression could be hindered or even stopped before major neurodegeneration has occurred and nervous tissue's regenerative capability has been depleted.

The swift development of the latest genome and epigenome analysis methods as well as modern stem cell technologies have enabled the detection of new markers for different biological conditions and diseases, like AD. Genetic variants contribute to a large part of the AD risk, however, the genetic risk factors alone do not determine an individual's AD risk profile. Epigenetic mechanisms, which regulate packing and activity of the genome, are also involved in the disease pathogenesis. As the pathological changes in AD appear mostly in the central nervous system during a time-span of several years, tracking the disease progression in the target tissues has been challenging. Epigenetic markers associated with AD have been identified in post-mortem brain tissue samples. These are invaluable findings and give new insights to the disease mechanisms, however, a biomarker for clinical AD diagnosis has to detectable in peripheral tissues, like blood, that can be sampled. Modern stem cell technologies, for example, patient-derived iPSC disease models, offer completely new possibilities for studying disease mechanisms in human cells and tissues. These models can also be utilised in biomarker research to validate how the markers are involved in the disease mechanisms within the target tissues.

3 Aims of the Study

The main goal of this study was to identify and evaluate peripheral blood DNA methylation markers that indicate or even predict AD progression. To achieve this goal several methods for genome and epigenome analysis were optimised and implemented. The specific goals for the subprojects were:

1. To implement the Reduced Representation Bisulfite Sequencing method for genome-wide DNA methylation analysis with human embryonic stem cell samples before proceeding to clinical samples. (I)
2. To characterise genetic risk for AD in the Finnish study population, identify blood DNA methylation markers for AD in twin pairs discordant for the disease with RRBS and implement and utilise bisulphite pyrosequencing in the validation of the differentially methylated regions (DMR). (II)
3. To identify AD-associated DNA methylation markers in hippocampus and compare blood and brain DNA methylomes in AD. (II)
4. To implement a method for plasma circulating cell-free DNA (cfDNA) methylation analysis and identify differences in twin pairs discordant for episodic memory performance. (III)
5. Implement patient-derived induced pluripotent stem cell lines and brain organoid models to study disease mechanisms in vitro.

In addition to these goals, during project III an additional objective was to collect a new Finnish twin cohort sample set including blood, plasma and peripheral blood mononuclear cell (PBMC) samples for studying epigenetic markers associated with EM impairment.

4 Materials and Methods

4.1 Materials

4.1.1 Cell cultures (I)

The HS360 and H9 hESC lines were cultured in Turku Centre for Biotechnology on feeder cultures with human foreskin fibroblasts that had been mitotically inactivated. Before experiments, the cells were cultured at least two passages in feeder-free conditions in mTeSR1 medium on Matrigel-coated plates. The H7, H14 and Shef5 lines had been cultured in Sheffield UK as described previously (Lund *et al.*, 2013). In addition, NT2D1 human myeloma, CCRF-CEM human acute lymphocytic leukaemia and 2120Ep human embryonal carcinoma cell lines were included in the differential methylation analysis in study I. The cell samples were utilised in implementing the RRBS protocol and characterising DNA methylation differences between normal and karyotypically abnormal hPSCs. H9 and HS360 hESC lines were cultured and utilised in cerebral organoid model implementation.

4.1.2 Twin blood samples (I, II, III)

4.1.2.1 Finnish twin samples

The ethylenediaminetetraacetic acid (EDTA) stored blood samples from the Finnish twin pairs discordant for AD, who were included in the peripheral blood DNA methylation analysis (II), were obtained from the Older Finnish twin cohort (born 1922-1937), excluding four additional twin pairs born 1915-1950. During the cohort study, a total number of 2,483 individuals had been screened via phone interviews (Järvenpää *et al.*, 2002). Discordant twin pairs were invited to neurophysiological examinations and PET and MRI characterisation and blood sampling between the years 2000 and 2008. Alzheimer's disease diagnosis was made according to the overall examination results for 29 pairs. (Kaprio and Koskenvuo, 2002; Järvenpää *et al.*, 2003, 2004; Virta *et al.*, 2008; Scheinin *et al.*, 2011; Kaprio, 2013)

During study III, additional blood samples from Finnish twins were collected (**Table 2**). A total number of 1,817 individuals from the Older Finnish twin cohort were screened via phone interviews between the years 2013 and 2017 (Kaprio and Koskenvuo, 2002; Lindgren *et al.*, 2019). Seventeen twin pairs and eight non-twin controls participated PET, MRI and neurophysiological testing and blood sampling at Turku PET Centre. In addition, two single twins took part in the examination since the co-twins could not attend imaging. From the 17 twin pairs, 8 were monozygotic and 9 dizygotic. The twins were defined with amnestic mild cognitive impairment (aMCI) if the z-scores in the two EM tests were less than or equal to minus one. The two tests included the delayed word list recall from the Consortium to Establish a Registry for Alzheimer's disease Neuropsychological Battery (CERAD-NB) and the Logical Memory delayed recall from the Wechsler Memory Scale-Revised (WMS-R) (Jak *et al.*, 2009; Sotaniemi *et al.*, 2012). Seven twin pairs, two monozygotic and five dizygotic, who participated in the examinations at PET centre before July 2016, were selected for characterisation of the plasma cfDNA methylation profiles.

4.1.2.2 Swedish twin samples

The peripheral blood DNA samples from Swedish twin pairs discordant for AD (born 1907-1953) were obtained from the Swedish Twin Registry (Magnusson *et al.*, 2013). The samples had been collected before (120 pairs) or after (29 pairs) AD diagnosis was made and combined from three Swedish twin studies: the Swedish Adoption/Twin Study of Aging (SATSA), the Study of Dementia in Swedish Twins (HARMONY) and TwinGene (Gatz *et al.*, 1997, 2005; Finkel *et al.*, 2005; Magnusson *et al.*, 2013).

4.1.3 Human brain tissue samples (II)

The frozen post-mortem human brain tissue samples analysed in study II were obtained from NIH NeuroBioBank. The samples consisted of tissue from hippocampus, including dentate gyrus. The samples had been collected from six AD cases and six age-matched controls (57-90 years). Both case and control groups included two females and four males. One of the AD case samples was excluded from the analyses because the sample was in fact from the amygdala, according to the neuropathological examination.

4.1.4 Human iPSC lines reprogrammed from peripheral blood mononuclear cells

Induced pluripotent stem cells were reprogrammed from PBMCs collected from three individuals: a dizygotic twin pair (DZ-1/DZ-2) and a single monozygotic twin (MZ-1). **Figure 3** shows a microscope image of the iPSCs reprogrammed from DZ-1 peripheral blood mononuclear cells. The iPSC lines were reprogrammed at Glykos Finland Oy with the CytoTune™ -iPS Sendai Reprogramming Kit and stored in liquid nitrogen for later use.

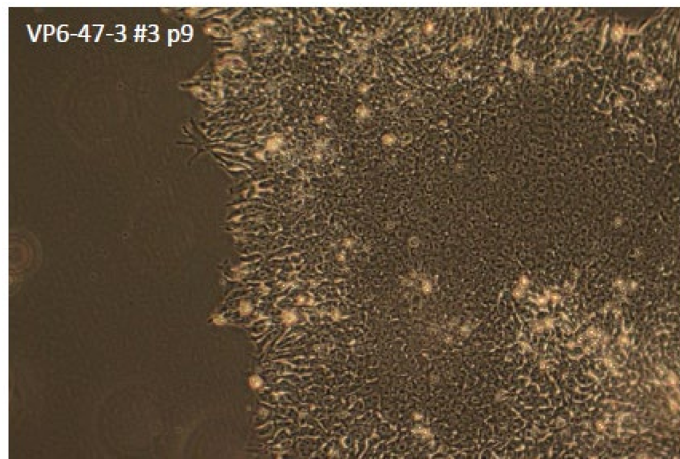


Figure 3. Induced pluripotent stem cells reprogrammed from DZ-1 peripheral blood mononuclear cells.

Table 2. Characterisation of the twin cohort samples collected in project III.

ID	Twin pair ID	Age	Gender	APOE*	CERAD-NB z-score	WMS-R z-score	EM status	Plasma cfDNA
MZ-1	*1*	*74*	*male*	*33*	*-2.5*	*-1.0*	*aMCI*	*X*
MZ-2	*1*	*74*	*male*	*33*	*-2.5*	*-2.8*	*aMCI*	*X*
MZ-3	*2*	*74*	*male*	*33*	*0.3*	*-1.0*	*normal*	
MZ-4	*2*	*74*	*male*	*33*	*-2.5*	*-2.3*	*aMCI*	
MZ-5	*3*	*75*	*male*	*33*	*0.8*	*1.6*	*normal*	*X*
MZ-6	*3*	*75*	*male*	*33*	*-0.3*	*0.1*	*normal*	*X*
MZ-7	*4*	*74*	*female*	*33*	*-1.4*	*-1.3*	*aMCI*	
MZ-8	*4*	*74*	*female*	*33*	*0.3*	*-2.3*	*normal*	
MZ-9	*5*	*71*	*female*	*34*	*-0.8*	*-1.3*	*normal*	
MZ-10	*5*	*71*	*female*	*34*	*-1.4*	*-1.0*	*aMCI*	
MZ-11	*6*	*73*	*male*	*44*	*-1.4*	*-1.0*	*aMCI*	
MZ-12	*6*	*73*	*male*	*44*	*-0.8*	*-0.5*	*normal*	
MZ-13	*7*	*71*	*male*	*34*	*-0.3*	*2.7*	*normal*	
MZ-14	*7*	*71*	*male*	*34*	*-2.5*	*1.6*	*normal*	
DZ-1	*8*	*75*	*female*	*33*	*1.4*	*3.5*	*normal*	
DZ-2	*8*	*75*	*female*	*33*	*-1.4*	*-1.8*	*aMCI*	
DZ-3	*9*	*75*	*female*	*33*	*-1.4*	*-1.0*	*aMCI*	*X*
DZ-4	*9*	*75*	*female*	*33*	*0.3*	*0.8*	*normal*	*X*
DZ-5	*10*	*74*	*male*	*33*	*-0.8*	*-0.2*	*normal*	*X*
DZ-6	*10*	*74*	*male*	*33*	*-0.8*	*-0.2*	*normal*	*X*
DZ-7	*11*	*74*	*female*	*33*	*-0.3*	*-0.2*	*normal*	
DZ-8	*11*	*74*	*female*	*33*	*-2.5*	*-0.2*	*normal*	
DZ-9	*12*	*74*	*male*	*34*	*0.3*	*0.3*	*normal*	*X*
DZ-10	*12*	*74*	*male*	*34*	*-3.1*	*-3.1*	*aMCI*	*X*
DZ-11	*13*	*72*	*female*	*33*	*1.4*	*-0.2*	*normal*	*X*
DZ-12	*13*	*72*	*female*	*34*	*-2.5*	*-3.1*	*aMCI*	*X*
DZ-13	*14*	*72*	*female*	*44*	*-1.9*	*-2.3*	*aMCI*	*X*
DZ-14	*14*	*72*	*female*	*34*	*1.4*	*0.8*	*normal*	*X*
DZ-15	*15*	*72*	*female*	*34*	*-1.4*	*-1.3*	*aMCI*	
DZ-16	*15*	*72*	*female*	*44*	*-0.3*	*-1.5*	*normal*	
DZ-17	*16*	*75*	*male*	*34*	*-0.8*	*-1.3*	*normal*	
DZ-18	*16*	*75*	*male*	*34*	*0.3*	*-0.5*	*normal*	
HC-1		*70*	*male*	*33*	*0.8*	*2.2*	*normal*	
HC-2		*73*	*female*	*23*	*0.8*	*0.3*	*normal*	
HC-3		*76*	*female*	*33*	*0.8*	*0.6*	*normal*	
HC-4		*74*	*male*	*33*	*-0.3*	*0.6*	*normal*	
HC-5		*71*	*female*	*33*	*0.8*	*4.0*	*normal*	
HC-6		*73*	*female*	*23*	*0.8*	*3.7*	*normal*	
HC-7		*72*	*male*	*13*	*1.9*	*1.9*	*normal*	
HC-8		*76*	*male*	*33*	*-0.3*	*0.1*	*normal*	

MZ: monozygotic twin, DZ: dizygotic twin, HC: healthy control, **APOE* genotype

4.2 Methods

4.2.1 Human tissue sample processing

4.2.1.1 DNA isolation from EDTA blood samples (I, II)

DNA was isolated from EDTA blood samples with Qiagen QIAamp Blood Mini and Maxi kits according to the manufacturer's protocol. Isolated DNA concentrations we measured with Thermo Scientific NanoDrop 2000 and genomic DNA quality checked with gel electrophoresis.

4.2.1.2 Isolation of plasma and peripheral blood mononuclear cells from blood samples (III)

Blood samples were collected at Turku PET Centre into BD Vacutainer® CPT™ tubes and processed at Turku BioScience Centre within one hour of sampling. To separate plasma and PBMCs, the tubes were first centrifuged 20 minutes at 1,500 RCF (relative centrifugal force) in room temperature. Plasma was transferred into one-millilitre aliquots in low-binding Eppendorf tubes, centrifuged again 10 min at 1,300 RCF in +4°C to remove residual cells, transferred into clean Eppendorf tubes, snap-frozen on dry ice and stored in -80°C. The number of PBMCs was counted in a cell counting chamber, PBMCs were transferred into 0.5 ml aliquots in cryotubes, including 10 % DMSO (dimethyl sulfoxide), and frozen in cell freezing container in -80°C. After two days the PBMC cryotubes were transferred into -150°C for long term storage.

4.2.1.3 DNA isolation from plasma (III)

Circulating cfDNA was isolated from plasma samples with Qiagen QIAamp Circulating Nucleic Acid Kit using Qiagen QIAvac 24 Plus vacuum system according to the manufacturer's protocol. Isolated DNA amounts were measured with Invitrogen Qubit 2.0 fluorometer dsDNA HS assay and quality checked with Agilent Technologies Bioanalyzer 2100 DNA High Sensitivity assay.

4.2.1.4 DNA isolation from frozen brain tissue samples (II)

The fresh-frozen brain tissue samples were cut on microscope slides in the microtome. The frozen tissue slides were stained with anti-hyaluronan tags and taken into neuropathological examination by a pathologist, to ensure that the tissue samples were from the correct brain region. DNA was isolated from frozen brain

tissue slides with Qiagen QIAamp DNA Micro Kit according to the manufacturer's protocol with few modifications. Lysis buffer (ATL) was pipetted on the tissue slide, tissue was detached from the glass slide with a cell scraper and the mix containing the tissue sample and lysis buffer was transferred into 1.5 ml Eppendorf tube for the following steps.

4.2.2 Targeted sequencing of single nucleotide polymorphisms associated with Alzheimer's disease (II)

The 21 SNPs (single nucleotide polymorphism) previously associated with AD and *APOE* genotypes were characterised with Illumina TruSeq Custom Amplicon sequencing assay from the Finnish twin samples, included in the blood DNA methylation RRBS analysis. The sequencing panel was designed in Illumina DesignStudio software. The samples were sequenced with Illumina MiSeq sequencing platform and 500-cycle chemistry. Variants were called in Illumina BaseSpace cloud computing service. Genetic risk scores (GRS) were calculated and GRS association with AD was tested with a generalised linear model in R software version 3.4.3 using packages lmtest, multiwayvcov and fmsb (R Core Development Team, 2013; Zeileis and Hothorn, 2015; Nakazawa, 2016; Graham, Arai and Hagströmer, 2018).

4.2.3 DNA methylation analysis

4.2.3.1 Reduced Representation Bisulfite Sequencing (I, II)

RRBS was utilised in the characterisation of genome-wide DNA methylation profiles from different cell lines (I), human peripheral blood samples (II) and human post-mortem brain tissue samples (II). The sequencing libraries were prepared as described previously (Boyle *et al.*, 2012). However, the samples were pooled together only after confirming the quality of each final library instead of pooling directly after the ligation of indexed sequencing adapters. Library qualities were checked with Agilent Technologies Bioanalyzer 2100 High Sensitivity DNA Assay and concentrations measured with Invitrogen Qubit 2.0 Fluorometer dsDNA HS Assay. The RRBS libraries were sequenced with 1x50 bp chemistry in Illumina HiSeq 2000/2500/3000 machines. Sequencing reads were trimmed with Trim Galore software and mapped to hg19 (cells, blood) or hg38 (brain) reference genomes using Bismark (Krueger and Andrews, 2011; Krueger, 2012). Differential methylation between samples was analysed using methylKit (cells and brain) or RADMeth (blood) software (Akalin *et al.*, 2012; Dolzhenko and Smith, 2014). While analysing

the blood DNA methylation data, BACON software was utilised in calculating and correcting for bias caused by unobserved factors, like cell type heterogeneity (van Iterson, van Zwet and Heijmans, 2017).

4.2.3.2 Targeted bisulphite pyrosequencing (II)

Targeted bisulphite pyrosequencing was used to validate one of the DMRs that were detected with RRBS in peripheral blood of AD-discordant twin pairs. The genomic region chr10:1405336-1405409 (hg19) was analysed with pyrosequencing. PCR primers (5′-gtaatttagtggtgttgttgaat-3′ and 5′-biotin-cctaacccccaaccaacttcttactac-3′) and sequencing primer (5′-gggttgagttaagtgtgtttggtaga-3′) were designed in Qiagen PyroMark AssayDesign SW 2.0. DNA samples were bisulphite converted with Qiagen Epitect Fast DNA Bisulfite Kit and amplified with Qiagen PyroMark PCR kit. Amplicons were sequenced with Qiagen PyroMark Q24 Advanced platform and methylation values called with Qiagen PyroMark Q24 Advanced 3.0.0 software. Statistical analysis of differential methylation association with AD was carried out in R version 3.4.3 and packages lme4 v1.1-15, car v2.1-6, survival 2.42-4, coxme 2.2-10 (Fox and Weisberg, 2011; R Core Development Team, 2013; Bates *et al.*, 2015; Therneau, 2015, 2018).

4.2.3.3 Plasma cell-free DNA bisulphite sequencing (III)

Plasma circulating cfDNA methylation profiles were analysed with a custom protocol that was implemented during the project III. The sequencing libraries were prepared from seven twin pairs, including two monozygotic twin pairs concordant for aMCI, one dizygotic twin pair concordant for aMCI and four dizygotic twin pairs discordant for aMCI. The isolated DNA samples were first end-repaired and A-tailed with Klenow fragment. Unmethylated lambda DNA was added to the samples to control for bisulphite conversion efficiency. Illumina TruSeq sequencing adapters were ligated to the a-tailed DNA fragments and library samples purified with AMPure XP Beads. Libraries were bisulphite converted with Invitrogen MethylCode Bisulfite Conversion Kit and amplified in an 18 cycle PCR. Finally, the libraries were purified twice with AMPure XP beads, with 1.2 and 1.5 bead concentrations, and final library quality was characterised with Agilent Technologies Bioanalyzer 2100 DNA High Sensitivity Assay and concentrations measured with Qubit 2.0 ds DNA HS assay. The libraries were sequenced in Illumina HiSeq 2500 and 3000, sequencing reads trimmed with Trim Galore version 0.4.1 and trimmed reads mapped into reference genome (hg38) with Bismark 0.14.5 (Krueger and Andrews, 2011; Krueger, 2012). Different R software packages were tested for differential methylation analysis, including RnBeads version 1.6.1, dmrseq version 1.0.12 and

methylKit version 1.7.9, however, the final analyses were carried out with PQLseq version 1.1 (Akalin *et al.*, 2012; Korthauer *et al.*, 2019; Müller *et al.*, 2019; Sun *et al.*, 2019). Methylation differences associated with the episodic memory z-score were tested within the whole seven-twin pair sample set using a generalised linear mixed-effects model.

4.2.4 APOE genotyping (II)

Two methods were utilised in the characterisation of the *APOE* genotypes (rs7412 and rs429358) from the human DNA samples. 404 DNA samples collected from Finnish and Swedish twins were sent to LGC Genomics to be genotyped with a custom qPCR based KASP assay. However, KASP genotyping failed from a few samples, thus *APOE* genotypes from four individuals were characterised with pyrosequencing. Pyrosequencing primers had been designed in PyroMark AssayDesign 2.0 software. The PCR primers for rs7412 were 5'-CTCCGCGATGCCGATGAC-3' and 5'-biotin-CCCCGGCCTGGTACACTG-3' and sequencing primer 5'-CGATGACCTGCAGAA-3'. For rs429358 the PCR primers were 5'-biotin-CGCGGACATGGAGGACGT-3' and 5'-CCTCGCCGCGGTACTGCA-3' and the sequencing primer 5'-ACTGCACCAGGCGGC-3'. The DNA samples were amplified with Qiagen PyroMark PCR kit according to the manufacturer's protocol and sequenced in Qiagen PyroMark Q24 Advanced. Variants were called with Qiagen PyroMark Advanced 3.0.0 software.

4.2.5 Single-cell RNA sequencing (II)

10X Genomics Chromium single-cell RNA sequencing platform was utilised in analysing the single-cell RNA expression profiles of PBMCs from two Finnish monozygotic twin pairs discordant for cognitive function. Chromium™ controller and Single Cell 3' Reagents kit (10x Genomics) were used in the preparation of the single-cell RNA sequencing libraries. Libraries were sequenced with Illumina HiSeq2500 and the sequencing data pre-processed with Cell Ranger software (v. 1.2.0, 10x Genomics) and mapped into hg38 reference genome. Single-cell gene expression profiles for each sample were normalised and analysed with Seurat R package v. 1.4.0.9 (Butler *et al.*, 2018).

4.2.6 Human pluripotent stem cell culturing and cerebral organoid differentiation

H9 and HS360 hESC lines were cultured in feeder-free conditions on Matrigel (Corning) coated plates with mTeSR1 culturing medium (STEMCELL Technologies). Culturing medium was changed daily and the cells were passaged using dispase. The hESCs were differentiated into cerebral organoids according to the protocol designed by Lancaster and her group (2014). Briefly, PSCs are differentiated into embryoid bodies (EB) on a 96-well plate. The EBs are cultured on 96-well plate until germ layers start to differentiate. Next, the EBs are transferred into larger wells, where neuroepithelial differentiation is inducted. The primitive neuroepithelial tissues are transferred into Matrigel droplets, expanded, and after further differentiation, they are moved into stationary cultures in 125 ml spinning bioreactors.

4.2.7 Cerebral organoid staining and imaging

Cerebral organoids were collected from the bioreactor, washed with PBS and fixed with paraformaldehyde. After fixing, the organoids were cryoprotected by incubating in 30 % sucrose and PBS in +4º C overnight. After cryoprotection, the fixed organoids were mounted into freezing blocks in mounting medium for cryotomy. Blocks were frozen in isopentane on dry ice and stored in -80º C. The blocks containing the organoids were cut into 40 µm thick slides in a cryostat and stained with anti-DAPI, -SOX2, -TUJ1 and -DCX antibodies. The organoids were imaged using a Zeiss LSM 780 confocal microscope.

5 Results

5.1 Implementation of RRBS for genome-wide DNA methylation analysis of human stem cell and blood samples (I)

Before working with the clinical blood DNA samples isolated from the twin pairs discordant for AD, the RRBS protocol was tested and implemented with DNA from human cell lines and peripheral whole blood. The cell lines included hESC and cancer cell lines. **Figure 4** shows the Bioanalyzer electropherograms of RRBS libraries prepared from human stem cell and peripheral blood DNA. The overall size distribution is very similar between all samples even though the concentrations differ. Sequencing of the RRBS libraries prepared from human cell samples produced 9.8 to 24.4 million raw reads per sample. After quality processing and trimming, between 5.5 to 14.04 million reads per sample were uniquely mapped into the reference genome (hg19) with 56.6%-73.9% mapping efficiency. Between 945K and 1.5 million CpG sites were detected with 5x sequencing coverage in each sample.

5.2 Catalase gene is silenced via epigenetic mechanisms in karyotypically abnormal hESC lines

With the DNA methylation data from the cell samples, we compared differences between karyotypically normal and abnormal human hESC lines as well as hESC lines before the transformation, however, with a tendency to accumulate karyotypic abnormalities. By comparing DNA methylation and gene transcription profiles, we discovered that a key antioxidant enzyme catalase is silenced in abnormal hESCs via DNA methylation and the changes can be detected already before karyotypic abnormalities become evident. Methylation of the CpG site in *catalase* (*CAT*) gene promoter was increased in abnormal cells by 68 % ($p=2.33*10^{-5}$) in comparison to normal cell lines. *CAT* expression was also decreased in abnormal cells according to the transcriptome data (fold change: -8.47, $p=9.08*10^{-23}$).

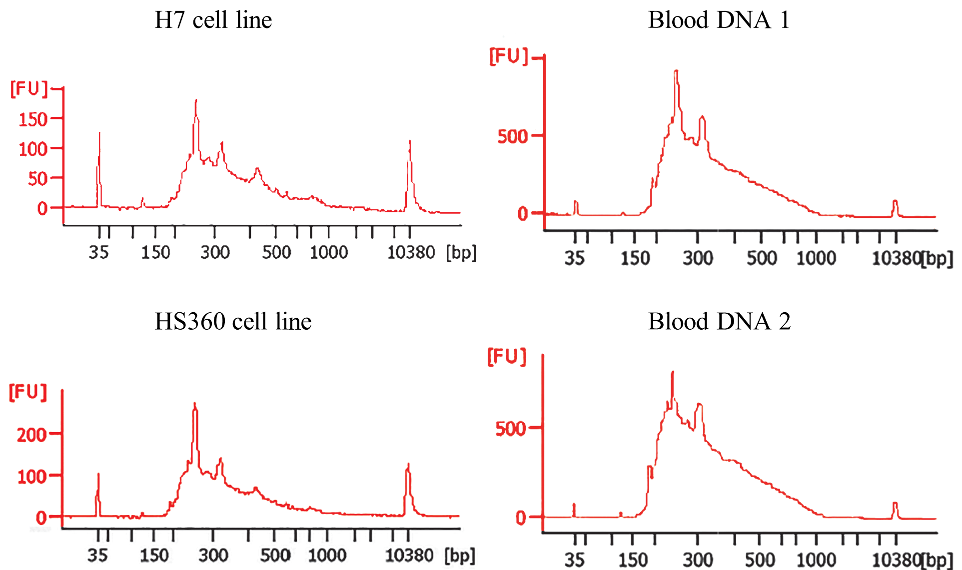


Figure 4. Bioanalyzer electropherograms of RRBS libraries prepared from human stem cell and blood DNA.

5.3 Genetic risk for Alzheimer's disease in Finnish disease-discordant twin pairs (II)

Since genetics have such a significant impact on AD risk, we characterised genetic AD risk in the Finnish twin study participants before analysing their blood DNA methylation profiles. The 21 SNP positions that had previously been associated with altered AD risk were characterised from 9 MZ and 12 DZ twin pairs discordant for AD, 9 un-related controls and 18 un-related AD cases (**Figure 5**). Monozygotic twins had identical genotypes in comparison to the co-twins, as expected. Dizygotic twin pairs had differences in 2-10 variant positions compared to the co-twins. Genetic risk scores were calculated for the different study groups; however, the scores were not associated with AD status according to generalised linear regression model (Wald test z value < 1, Pr(>[z]) > 0.5). (II: table 2)

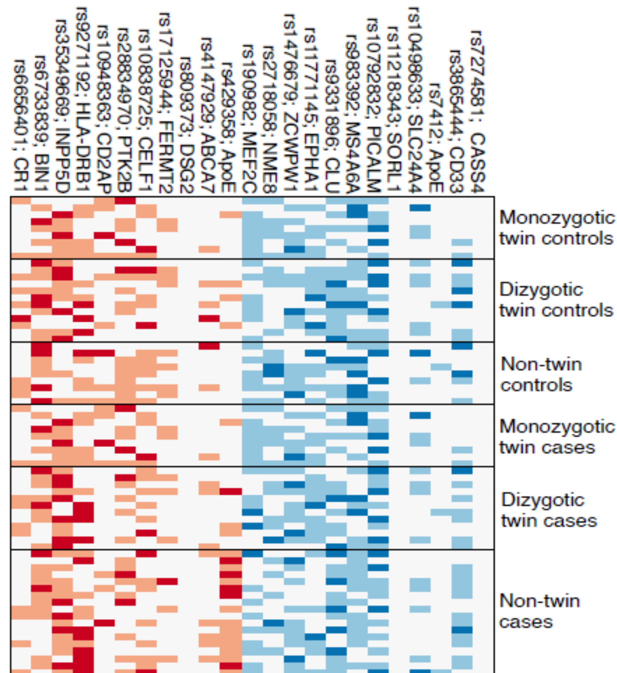


Figure 5. Genetic Alzheimer's disease risk profiles of the Finnish study participants. SNPs associated with increased AD risk are marked red and protective SNPs are blue. Dark colours indicate biallelic risk or protective variant and light colours indicate monoallelic variant.

5.4 Peripheral blood DNA methylation differences in twin pairs discordant for Alzheimer's disease (II)

To identify possible DNA methylation markers for AD in peripheral blood, we characterised genome-wide DNA methylation profiles from blood samples and compared differences within 11 MZ and 12 DZ twin pairs discordant for AD. The differences between MZ and DZ twins were analysed separately at first. Only CpG sites detected in at least four twin pairs, both twins of the pair, per study group with 10x or higher sequencing coverage were included in the differential methylation analyses. With this threshold, we detected 838,967 sites in MZ twin and 817,103 in DZ twin groups. The differential methylation cut-off criteria were median methylation difference of at least 15 % between discordant twins and Benjamin Hochberg false discovery rate adjusted $p \leq 0.05$ Within MZ twins, we detected 912 differentially methylated sites, 434 of which were more and 478 were less methylated in AD cases versus controls. A total number of 2112 CpG sites were differentially methylated between AD discordant DZ twins. 1344 of these sites were more and 768 were less methylated in cases when compared to controls. Comparison

of the two datasets from MZ and DZ twins revealed 11 CpG sites that were differentially methylated in both groups. Seven CpG sites were more and four were less methylated in AD cases than in controls. The closest or overlapping genes to these 11 CpG sites were defensin alpha 1 (*DEFA1*), t-SNARE domain containing 1 (*TSNARE1*), DEAF1 transcription factor (*DEAF1*), ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 2 (*ARAP2*), canopy FGF signalling regulator 1 (*CNPY1*), adenosine deaminase RNA specific B2 (*ADARB2*), Rho GTPase activating protein 8 (*ARHGAP8*), general transcription factor IIIC subunit 2 (*GTF3C2*), actin alpha 1, skeletal muscle (*ACTA1*), semaphorin 5A (*SEMA5A*) and CAP-Gly domain-containing linker protein 2 (*CLIP2*). (**Table 3**, II: Figure 1)

Table 3. Differentially methylated CpG sites in peripheral blood of Finnish twin pairs discordant for Alzheimer's disease.

CpG site annotation		MZ twins		DZ twins		All twins	
Location (hg19)	***Closest gene***	***Meth. Diff. %***	***p-value***	***Meth. Diff. %***	***p-value***	***Meth. Diff. %***	***p-value***
chr8:143407666	*TSNARE1*	*24.92*	*8.89E-03*	*20.40*	*0.00*	*22.66*	*2.57E-04*
chr7:155327302	*CNPY1*	*18.18*	*2.55E-02*	*17.46*	*0.01*	*17.82*	*3.19E-03*
chr22:45193257	*ARHGAP8*	*17.11*	*5.98E-04*	*18.40*	*0.01*	*17.75*	*8.64E-05*
chr10:1405392	*ADARB2*	*17.03*	*3.05E-03*	*23.67*	*0.05*	*20.35*	*1.43E-03*
chr4:33809696	*ARAP2*	*16.18*	*1.29E-04*	*16.60*	*0.01*	*16.39*	*1.17E-05*
chr11:646372	*DEAF1*	*15.99*	*4.05E-03*	*15.82*	*0.00*	*15.91*	*2.35E-06*
chr8:6875660	*DEFA1/-1B/-3*	*15.49*	*2.55E-02*	*21.25*	*0.02*	*18.37*	*3.90E-03*
chr7:73825342	*CLIP2*	*-15.95*	*1.98E-02*	*-20.54*	*0.04*	*-18.25*	*6.13E-03*
chr1:229540297	*ACTA1*	*-18.18*	*9.10E-04*	*-19.02*	*0.02*	*-18.60*	*1.71E-04*
chr2:27580515	*GTF3C2*	*-22.45*	*6.61E-04*	*-29.82*	*0.02*	*-26.14*	*1.27E-04*
chr5:8851435	*SEMA5A*	*-27.75*	*6.55E-03*	*-31.67*	*0.01*	*-29.71*	*8.55E-04*

5.5 Alzheimer's disease-associated DNA methylation marks in the brain and correlation with blood DNA methylation markers (II)

Next, the goal was to find out whether the blood DNA methylation marks associated with AD can also be detected in brain tissue. Genome-wide DNA methylation profiles were analysed in the hippocampus, including dentate gyrus tissue samples from four AD cases and six un-related controls. Differential methylation cut-off was defined as mean methylation difference of at least 15 % between cases and control, consistent methylation difference between all cases and controls and multiple test corrected $p<0.05$. With these criteria, we detected 114 CpG sites that were less methylated and 87 that were more methylated in AD cases in comparison to controls. These 201 sites located within or close to 176 different genes. Comparison of the

gene lists from differentially methylated sites in peripheral blood of Finnish twin pairs discordant for AD and brain tissue from un-related AD cases and controls revealed one common gene *ADARB2*. In both brain and blood, the differentially methylated sites were located within the same region chr10:1404752-1405717 in exon 3 of the *ADARB2* gene. (II: Figure 2)

5.6 Expression of the differentially methylated genes in peripheral blood (II)

The next research question was whether the genes located closest to DMRs in the blood are also differentially expressed in blood cells in AD. At first, we examined the expression of the eleven differentially methylated genes (**Table 3**) in a published peripheral blood expression data from 121 AD cases and 122 controls (Sood *et al.*, 2015). Nine of the eleven genes were expressed in the blood; however, there were no differences in expression of these genes between AD cases and controls. Expression of these genes was also characterised with single-cell RNA sequencing in PBMC samples from two Finnish twin pairs discordant for EM performance. The single-cell RNA sequencing analysis revealed no major differences in the proportions of PBMC populations or expression of the differentially methylated genes in PBMCs between EM impaired and control twins. Expression of six of the eleven differentially methylated genes, including *ARAP2*, *CLIP2*, *DEAF1*, *GTF3C2*, *TSNARE1* and *ADARB2*, was detected in the PBMC samples.

5.7 Targeted validation of Alzheimer's disease-associated differential methylation in *ADARB2* gene (II)

Since the genomic region chr10:1,405,405-1,405,366 (hg19) in exon three of *ADARB2* gene was differentially methylated both in the blood of Finnish twin pairs discordant for AD and brain tissue samples from AD cases versus controls, this region was selected for validation assay. The next question was whether the peripheral blood methylation difference in the region can be validated in an extended twin cohort, including Swedish twins, and is the methylation status predictive for AD. DNA methylation in the region chr10:1,405,405-1,405,366 (hg19) was analysed with targeted bisulphite pyrosequencing in peripheral blood samples of 62 twin pairs discordant for AD. This included 33 Finnish and 29 Swedish twin pairs. Statistical testing was carried out with a linear mixed-effects model (lme) (Bates *et al.*, 2015).

The methylation status of the region was confirmed to be associated with AD (Wald test t value 4.68, II: Fig. 3). In the final statistical model, also gender, zygosity and age influenced the methylation status. DNA methylation difference was greater between

dizygotic than monozygotic twin pairs (estimate 7.67%, standard error (SE) 1.44, Wald t value 5.34). Analysing males and females separately revealed that methylation was increased in male cases (Wald t value 6.10), however, in females the methylation was increased only in cases who had *APOE* ε34/44 genotype (Wald t value 3.44) but not in cases who had *APOE* ε33 genotype (Wald t value -1.67). Additional analysis of 60 twin pairs from whom smoking data were available revealed that smoking also interacts with the disease (estimate 8.10%, SE 1.43, and Wald t value 5.68), however, the disease-methylation status association is not attenuated even if smoking is included in the statistical model (estimate 64.51%, SE 9.23, and Wald t value 6.99). DNA methylation in the region was also characterised in blood samples collected from 120 Swedish discordant twin pairs 0.5-18.5 years before AD diagnosis. However, the results show that methylation the status of the region before AD diagnosis does not predict disease outcome (hazard ratio = 1.00, 95% confidence interval = 1.00–1.00, p = 0.63).

5.8 Plasma cell-free DNA methylation markers for episodic memory impairment (III)

Episodic memory impairment usually causes the first clinical symptoms in dementia due to AD. However, AD can currently be clinically diagnosed only at later stages when neurodegeneration is more progressed. To identify early markers for AD, we implemented a genome-wide bisulphite sequencing method for plasma cfDNA methylation analysis and compared differences between Finnish twin pairs discordant for EM impairment. The WGBS protocol was successfully implemented for the methylation analysis of plasma circulating cfDNA. We characterised methylation profiles of 14 twins, including two monozygotic and five dizygotic twin pairs, from three nanograms of plasma DNA per individual. The number of CpG sites detected with 10x sequencing coverage per sample varied between 4,023 and 13,414. A vast majority of the detected CpG sites are located in certain hotspots, especially centromere and telomere regions of most chromosomes, instead of distributing evenly on the genome (III: Fig. 1).

CpG sites located in low complexity DNA regions, like repeating elements, from the data. Differential methylation associated with episodic memory z-score was tested with PQLseq R package that implements a binomial mixed-effects model. Episodic memory z-score was set as the test variable in the analysis. Gender, age, zygosity and *APOE* genotype included as fixed effects and twin pair ID as a random effect in the model. Differential methylation was analysed over 2.089 CpG sites that were detected in all fourteen twin samples with 10x sequencing coverage. 99.52 % of these sites located in intergenic regions of the genome. We detected no statistically significant differentially methylated CpG sites (multiple test corrected $p<0.05$) associated with EM performance in the twins.

5.9 Human pluripotent stem cell models for Alzheimer’s disease

Peripheral blood mononuclear samples were collected from seven monozygotic and nine dizygotic twin pairs and eight non-twin individuals, a total number of 40 individuals. Induced pluripotent stem cells were successfully reprogrammed from three individuals, a dizygotic twin pair (DZ-1/DZ-2) and a single monozygotic twin (MZ-1), at Glykos Finland Oy (**Table 2**). The iPSC cell lines expressed high levels of PSC markers TRA-1-60, SSEA-4 and H type 1 (**Figure 6**). Cerebral organoids were successfully differentiated from hESCs and maintained in the bioreactors for three months. Staining the organoids revealed a structured organisation of neurons and neuronal progenitor cells in the developing organoid tissue (**Figure 7**).

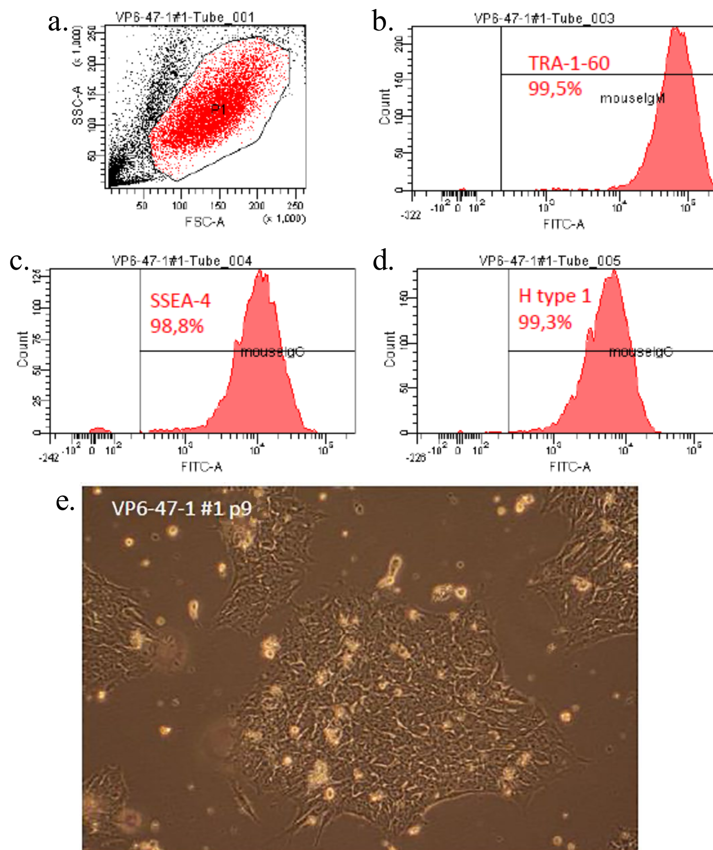


Figure 6. Induced pluripotent stem cell line reprogrammed from MZ-1 peripheral blood mononuclear cells. The iPSCs population was identified and selected according to side scatter (SSC) and forward scatter (FSC) values via flow cytometry (**a**.). The iPSCs expressed high levels of pluripotent stem cell markers TRA-1-60 (**b.**), SSEA-4 (**c.**) and H type 1 (**d.**) and formed typical colonies in *in vitro* culture (**e.**).

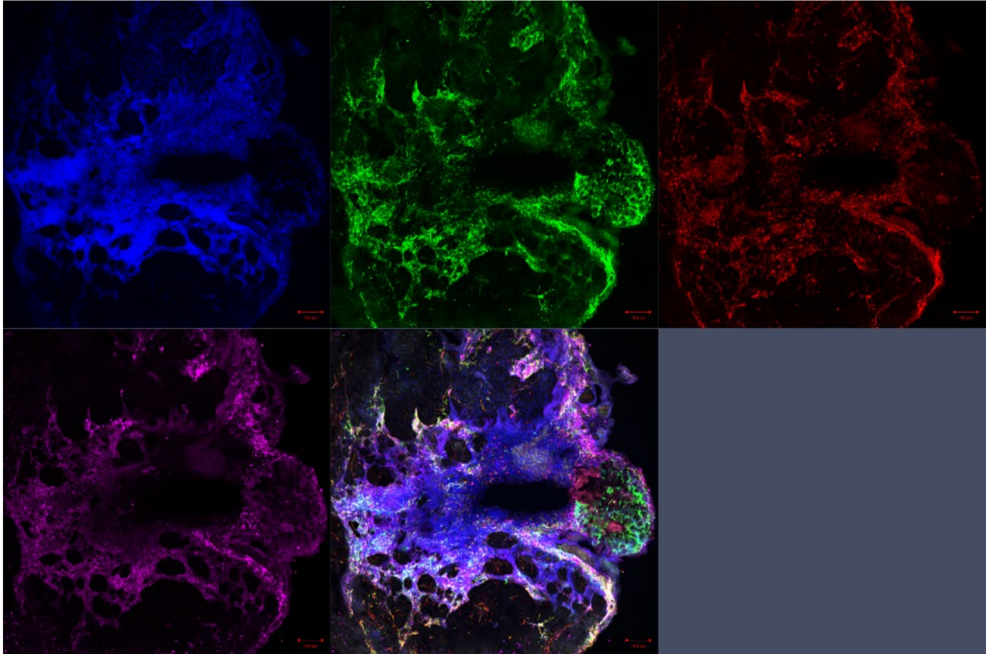


Figure 7. Cerebral organoid staining for DNA (DAPI: blue), neurons (TUJ1: green), neuronal progenitors (DCX: red/SOX2: purple)

6 Discussion

6.1 RRBS in genome-wide DNA methylation analysis

RRBS method was successfully implemented for genome-wide DNA methylation analysis of hESC, blood and post-mortem brain tissue DNA samples, during this project. In the peripheral blood DNA samples, we detected 838,967 sites in MZ twins and 817,103CpG in DZ twins with required 10x sequencing coverage in at least four twin pairs per twin group. This number is very close to the number of sites detected with the highest coverage microarrays, which is currently 850,000 sites, showing that the genomic coverage we reached with RRBS is similar to microarrays. The preparation of first of the blood DNA RRBS libraries was started already before the current highest coverage Illumina Infinium MethylationEpic microarray method was released. While the analysis of the methylation data would most likely have been a lot easier and more straightforward with microarrays, we would have detected only about half the number of CpG sites with the Illumina 450K array that was available at the time.

In comparison to WGBS, the genomic coverage we reached in the different sample sets is quite low, as expected. However, WGBS libraries are still quite expensive to sequence as reaching an adequate sequencing depth requires deeper sequencing. Deeper sequencing produces more sequencing data, which requires more storage and computing capacity to process the data. WGBS data also contains vast amounts of sequence reads that do not contain any methylation sites and are discarded in methylation calling step in data analysis. Overall, RRBS is a well-suited method for what it has been designed for: it quantifies DNA methylation at CpG rich genomic regions, which are, at least according to current knowledge, the most interesting in terms of functional relevance. Another possible option for this project would have been to analyse the samples with array-based methods. The per sample analysis costs for RRBS is in the same range as the newest version of microarrays, whereas WGBS is a lot more expensive. RRBS also provides a possibility to scale the costs of the experiments, since the most expensive step in the protocol is sequencing. Deeper sequencing will be more expensive but also will detect more

sites, whereas shallow sequencing is not so expensive but may also give interesting results at least if the number of samples per study group is adequate.

Even though the coverage in terms of the number of the methylation sites detected is similar between RRBS and microarrays, the actual coverage, where the detected sites are located in the genome, can still be quite different. Methylation arrays are for the most part targeted towards regions where methylation most likely has a functional impact, like gene promoters. In RRBS analysis these same regions that usually have high CpG density are also enriched, however, the selection of the target regions is also randomised as no regions are specifically targeted. (Stirzaker *et al.*, 2014) This allows RRBS to detect new methylation markers, not included in the methylation arrays, but may also cause unreliability issues in the analysis since specific CpG sites are detected only in a subset of the samples instead of all of them. Thus, in our peripheral blood DNA methylation analysis, the CpG sites had to be detected in at least four discordant twin pairs in a twin group to be included in the analysis.

To conclude, RRBS was a well-suited method for discovery of methylation markers candidates for AD. While methylation arrays could also have been a good option, none of the DMRs we detected with RRBS is included in the arrays that are currently available. The array-based analysis could also reveal as interesting differences as we detected with RRBS, however, the results would most likely be different.

6.2 Peripheral blood DNA methylation markers for Alzheimer's disease

We discovered DNA methylation differences in 11 genomic regions in peripheral blood of twin pairs discordant for AD, indicating that epigenetic markers associated with the disease can be detected in peripheral tissues. Whether the methylation levels in these sites affect gene expression or not, requires further inspection. However, our results show no differences in expression of the genes closest to the DMRs in PBMCs of twins who have EM impairment in comparison to controls. Expression of 9 out of the 11 differentially methylated genes was detected in peripheral blood in a study by Sood and her group (2015), yet they discovered no differences between AD patients and controls. Overall, the preliminary results indicate that AD-associated differential methylation in these regions does not affect gene expression in peripheral blood.

One of the 11 DMRs in blood was also differentially methylated in hippocampus tissue samples from AD patients in comparison to controls. Alzheimer's disease-associated differential methylation in this region, located in the third exon of the *ADARB2* gene, was also validated in a larger cohort including Swedish twin pairs

discordant for the disease. Even though the result was confirmed in the extended cohort, there is a lot of variation in the methylation status of this region between individuals. Our results also showed that the methylation status does not predict AD before the disease can be diagnosed. Thus, utilising peripheral blood methylation markers in AD diagnosis would most likely require a combinatory analysis of several distinct regions.

The reason why DNA methylation markers associated with AD can be detected in peripheral blood remains unclear. It may be that external factors, like environment, lifestyle, other diseases etc. give rise to these epigenetic changes in not only neuronal tissues but also other peripheral tissues like blood; however, the responses to these changes vary from tissues to tissues. In other words, the problems arise only in those tissues, where the affected genes are functionally important. A study by Braun and her group (2019) concluded that blood DNA methylation is indeed well correlated with DNA methylation in living brain tissue. How the genes that are differentially methylated in blood in AD contribute to the disease pathogenesis, requires still more research, even though many of the differentially methylated genes have known roles in neuronal tissue development and functions.

6.3 ADARB2 – a biomarker candidate for Alzheimer's disease

Our most promising finding was the AD-associated differential methylation in the third exon of the *ADARB2* gene in the blood of twin pairs discordant for the disease. However, it still requires more research to evaluate this marker's potential as a diagnostic indicator for AD. By definition, AD biomarkers have to have certain important characteristics: a biomarker has to have a proven function in the pathogenesis of the disease and the test has to be sensitive enough to detect AD at the expected rate and specific to distinguish AD from other dementias. Besides, a biomarker test should be relatively simple to perform, non-invasive for the patients, reliable, reproducible and not too expensive. (Davies *et al.*, 1998)

There is growing evidence that *ADARB2* may have important functions in regulating memory and learning and even cognitive impairment in AD. *ADARB2* is a member of RNA editing enzymes. *ADARB2* is expressed specifically in the brain where it may regulate RNA editing. In contrast to the other two members of the ADAR enzyme family ADARB2, also known as ADAR3, does not have the same catalytic domain, thus its functions remain unknown even though there is evidence of ADARB2 RNA editing activity in human tissues. (Chen *et al.*, 2000; Savva, Rieder and Reenan, 2012; Oakes *et al.*, 2017) Interestingly, a study by Khermesh and his group (2016) found out that adenosine to inosine RNA editing, which is mediated by ADAR enzymes, is distorted in the hippocampus in AD. A mouse model

has also shown that ADARB2 is involved in the regulation of memory and learning in mice. (Mladenova *et al.*, 2018) A variant of the *ADARB2* gene has also been associated with an elevated rate of MCI conversion to AD (Lee *et al.*, 2017).

Even though previous studies indicate that *ADARB2* may be functionally important in AD pathogenesis, the actual mechanisms have not been studied, and even the gene product's function in normal conditions is not completely understood. Due to inter-individual variability in the methylation level of this region, it is clear that this marker alone is not enough to detect AD with sufficient sensitivity. However, possible mechanistic studies could in the future give new information on whether and how *ADARB2* affects AD pathogenesis in neuronal tissue. It remains to be studied whether DNA methylation analysis of the *ADARB2* region in combination with other promising regions could be utilised in AD diagnostics.

6.4 The differentially methylated genes in the blood are associated with neuronal functions and disorders

In addition to the DMR in *ADARB2*, several of the other DMRs are close to genes that are highly expressed in the brain and are involved in neuronal functions. This supports our findings and the hypothesis that peripheral blood DNA methylation markers can indicate perturbations in the nervous system and neurodegenerative diseases, like AD. The DMR in *DEAF1* gene had the lowest multiple corrected test corrected significance value in the whole blood RRBS analysis. *DEAF1* encodes a zinc finger transcription factor that is important in embryonal development and is highly expressed in brain tissue (Fagerberg *et al.*, 2014). Different variants of *DEAF1* have been associated with several neuronal and neurodevelopmental disorders, like DEAF1 associated neurodevelopmental disorder, autism, epilepsy, basal ganglia dysfunction and intellectual disability (Rajab *et al.*, 2015; Gund *et al.*, 2016; Chen *et al.*, 2017).

TSNARE1, on the other hand, is expressed in several tissues around the body, and specific variants of the gene have been associated with schizophrenia and bipolar disorder (Sleiman *et al.*, 2013; Gu *et al.*, 2015). Increased levels of alpha-defensins, encoded by *DEFA1* gene, have been detected in CSF of AD patients (Szekeres *et al.*, 2016). *SEMA5A* gene, which encodes the semaphorin 5A cell membrane protein, is expressed widely in different tissues; however, it is very important for neuronal development (Goldberg *et al.*, 2004; Fagerberg *et al.*, 2014). Specific variants of the genes have been linked to increased Parkinson's disease risk and *SEMA5A* expression is decreased in autism (Melin *et al.*, 2006; Yu, Wang and Zhang, 2014). Semaphorin family of proteins has an important function in the nervous system and gene variants and defects in the functional networks of these proteins have been

associated with different neurodegenerative diseases, like AD, Parkinson's disease and amyotrophic lateral sclerosis (Quintremil *et al.*, 2019).

The genome is stored together as a three-dimensional structure where genomic regions that are located far away from each other may interact with each other. Thus, some of the DMRs we have detected in the blood may also regulate other more distant genomic regions and genes, instead of the closest genes that we have mainly focused on. The importance of long-range genomic interactions in the three-dimensional genome has become more and more apparent during recent years due to development in genome research methods (Schmitt, Hu and Ren, 2016; Schoenfelder and Fraser, 2019; Zheng and Xie, 2019). More research is still required to characterise and understand the interactions between distal genomic regions and possible disturbances in these mechanisms in diseases like AD.

Since several of the DMRs are near genes that are expressed in the brain and have been linked to neuronal functions and diseases, they may be also functionally relevant in AD pathogenesis. This also supports our hypothesis that DNA methylation markers for AD can be detected in the blood. Due to limited time and resources, we only managed to validate AD-associated DNA methylation difference in the *ADARB2* gene in the extended cohort, including the Swedish twins. The other DMRs, discovered with RRBS, should also be validated to confirm the differential methylation and to evaluate their sensitivity in AD diagnosis.

While targeted bisulphite pyrosequencing was a reliable, reproducible and cost-efficient method for targeted validation of the *ADARB2* regions, the method has its weak points as well. The platform that we used, Qiagen PyroMark Q24 Advanced, can only analyse up to 24 samples in a sequencing run, which is not very high-throughput. A higher throughput Q96 version of the machine also exists; however, Q24 was the only one available for us. The most difficult and time-consuming part of the pyrosequencing protocol is designing and setting-up the sequencing assays for each target region. Depending on the genomic sequence, some regions can be even impossible to analyse with bisulphite pyrosequencing. For example, the DMR close to *TSNARE1* gene (chr8:143407666, hg19) consists of highly repetitive DNA sequence, making it very difficult to design a working assay for PCR and sequencing of the bisulphite converted DNA. However, there are not many competing methods available for targeted DNA methylation analysis that the user can target and design the assay for, in principle, any target region in the genome. After all, a working pyrosequencing assay can produce accurate and quantitative methylation data at single-nucleotide resolution.

More research is also required to understand better, how DNA methylation differences in blood correlate to DNA methylation in the brain in AD, whether the expression of these genes is regulated via DNA methylation and what are the exact functional mechanisms that link these genes to the disease pathogenesis. Studying

the functional mechanisms of AD pathogenesis is difficult because even despite the extensive amount of research on the field, the disease-causing mechanisms remain unclear.

6.5 DNA methylation markers associated with Alzheimer's disease in the hippocampus

In the hippocampus samples, we detected 201 differentially methylated CpG sites in AD patients versus the controls: 114 of the sites were hypomethylated and 87 hypermethylated in AD. The main goal of this comparison was to find out whether DMRs in blood in AD overlap with DNA methylation differences in a brain region that is important for memory and learning. Comparison of the AD-associated DNA methylation markers in blood and brain revealed only one common region, which located in 3rd exon of the *ADARB2* gene. Since the gene is differentially methylated in both blood and brain in AD, *ADARB2* could have an important role in the disease pathogenesis.

However, the number of hippocampus samples we could acquire and analyse was very limited. In the final differential methylation comparison, we included only four AD patients and six controls. These samples were acquired from NIH Neurobiobank and, while they were matched against age and gender, we do not have information on the individuals' ethnic backgrounds or genetic AD risk profiles and could not adjust the analysis for these variables. Thus, there can be relatively high genetic and epigenetic variance between the individuals that we do not know of. Due to the limited information on the individuals' backgrounds and the small number of samples, we cannot make profound conclusions about the mechanisms of DNA methylation perturbations in the hippocampus in AD according to the analysis. However, the results are reliable enough to support the findings from the blood methylation analysis.

We also correlated our results from the blood DNA methylation analysis with previously published findings on AD-associated DNA methylation markers in the brain. However, only two of the eleven regions, the ones closest to *DEAF1* and *DEFA1* genes, we identified with RRBS in the blood are targeted in Illumina 450K methylation arrays, which have been used in the previous studies. On the other hand, previously identified AD DNA methylation marker regions, for example, *ANK1* and *RHBDF2*, were not detected in our RRBS data. (De Jager *et al.*, 2014; Lunnon *et al.*, 2014; Smith, Smith, Burrage, *et al.*, 2019)

6.6 Plasma cell-free DNA methylation marks for episodic memory impairment

A new method was implemented for the analysis and DNA methylation profiles were characterised from as little as three nanograms of plasma DNA. All in all, the amount of DNA that we were able to isolate from the plasma samples and the number of CpG sites we detected from the bisulphite sequencing libraries, was quite low. This indicates that there is no major leak of free DNA from the degenerating neuronal cells into blood circulation at this early stage of memory impairment. Majority of the methylation sites we detected in the plasma DNA samples were located in the telomere and centromere regions of most of the chromosomes, showing that these regions are for some reason heavily represented in the plasma. It may be that centromere and telomere regions are more stable and resistant to enzymatic degradation due to condensed packing of the chromatin, in comparison to other genomic regions.

We found no statistically significant differences in the plasma circulating cfDNA methylomes associated with EM impairment. This indicates that the degenerating neuronal cells do not release free DNA into the peripheral blood circulation at this stage of cognitive impairment, or the cfDNA degrades so quickly in plasma that it cannot be measured. However, the number of discordant twin pairs included in the differential methylation analysis was very limited: only four pairs, and, to our knowledge, this has not been studied elsewhere either. Thus, a more comprehensive comparison of a larger study population is needed to form more reliable conclusions. However, as the number of detected CpG sites varied remarkably between individuals and only 2.089 sites were detected in all individuals, our results indicate that the inter-individual variety in plasma cfDNA populations makes it difficult to compare cfDNA methylation profiles of larger groups of individuals. We initially selected the seven twin pairs for this comparison according to primary knowledge on their cognitive status and discordance. The twin pairs were recruited for imaging studies and neurophysiological testing in Turku PET centre using a telephone questionnaire (Järvenpää *et al.*, 2002). However, at the time when we had already processed the samples for plasma cfDNA methylation analysis, the neurophysiological test results revealed that only four of the seven twin pairs were discordant according to aMCI criteria.

The custom whole-genome bisulphite sequencing method was successfully implemented in this project. However, there are a few uncertainties to take into consideration. The overall number of CpG sites detected with sufficient sequencing coverage was quite low. Of course, this may also be due to a biological characteristic of the samples and there are no more DNA fragments or CpG sites to be detected in plasma cfDNA. When implementing the library preparation method, we also looked into the effect of starting DNA amount on the final library quality. According to the

results, the starting DNA amount has little to no impact on the number of detected CpG sites in the final library. This indicates that only a very small portion of the genome is present as cfDNA in the plasma. However, more research is also needed in this field as there are no comprehensive reports about the genomic coverage when sequencing plasma cfDNA.

The processing of the plasma DNA methylation sequencing data proved to be more challenging than we expected. Mapping of short read bisulphite sequencing data is difficult because bisulphite conversion reaction decreases the nucleotide sequence complexity of the sequencing reads. Thus, a high proportion of the sequencing reads can be identical and align perfectly with several different genomic locations, when mapped to a reference genome, which would cause bias in the sequencing results. By default, sequencing reads that align with several different genomic locations, in other words non-uniquely mapped reads, are discarded from the data after data pre-processing and mapping. Yet even from our uniquely mapped reads, which align only with one unique genomic location, a large proportion aligned with genomic regions like centromeres and telomeres that often consist of simple DNA sequences and repeating elements. These regions share high sequence similarity even between different chromosomes while the short repeating regions may differ between individuals, making bisulphite sequencing read mapping unreliable.

Due to the possible reliability issue in sequencing read mapping, we decided to exclude the reads aligning with repeating DNA elements from the data, which decreased the number of detected CpG sites by half of what was initially in the data. While this was probably the best option, a lot of data was lost and it is also possible that these regions would contain interesting information. Mapping this kind of bisulphite sequencing data would require more inspection. This problem has only arisen after the release of the newest human reference genome hg38 since many of the low sequence-complexity genomic regions were not included in the previous versions. Also, we did not find any published studies that would describe the mapping of plasma cfDNA bisulphite sequencing data, making it impossible to compare our data to others', even though similar methods have been utilised before in studies on other diseases and biological conditions. Plasma cfDNA methylation marks have been studied as potential biomarkers for prenatal diagnostics, detection of different cancers and assessment of organ transplant rejection, for example (Lun *et al.*, 2013; Sun *et al.*, 2015; Guo *et al.*, 2017).

In conclusion, we found no differences in plasma cfDNA methylation of twin pairs discordant for EM function. However, a more comprehensive comparison of a larger number of samples is required to validate or disprove these results. While the method that we implemented and utilised in the plasma DNA methylation analysis seemed to work well, a more comprehensive comparison of sequencing coverage

and read mapping between our method and other similar methods would be advised before utilising it in a larger sample cohort.

6.7 DNA methylation marks in hESCs and patient-derived iPSCs and brain organoid models

RRBS analysis of hESCs before and after transformation to abnormal karyotype revealed epigenetic silencing of the *CAT* gene, which encodes an important antioxidant enzyme. Although further studies are required to elucidate whether this alteration regulates genomic integrity, the finding provides a valuable marker of epigenetic integrity of hPSC lines for further studies and helps the identification of cell lines prone to accumulate genomic abnormalities.

Patient-derived iPSC lines were generated from PBMCs and the method to generate 3D cerebral organoids was successfully implemented. Cerebral organoids were generated from hESC lines. The initial goal was to utilise this approach in modelling AD *in vitro* and examine how epigenetic changes contribute to the disease mechanisms. During the progression of this study, however, findings by other research groups revealed that cerebral organoid models recapitulate the phenotype of developing foetal brain and require a lot more development before they can be utilised in modelling adult neurodegenerative diseases. Therefore, the project was put on hold until methods enabling in vitro ageing of the models have emerged (Luo *et al.*, 2016).

6.8 Future perspectives

We detected peripheral blood DNA methylation differences at eleven genomic regions in Finnish twin pairs discordant for AD. One of these DMRs, which located in the *ADARB2* gene, was validated in an extended cohort, including Swedish twins. Targeted validation would be required for the other DMR regions as well to more precisely evaluate their value for AD diagnostics. Since the methylation marker in the *ADARB2* gene is on its own not sensitive enough for AD diagnosis, targeted validation of the other regions would also enable characterising whether a combined analysis of several of the DMRs increases the diagnosis sensitivity to a sufficient level or not. It would also be interesting to see if these peripheral blood methylation markers are specific for AD or are they present in other dementias and cognitive disorders as well. After all, an AD biomarker should be specific for the disease and enable differential diagnosis from other dementias and neurodegenerative disorders.

More research is also needed to find out how the differentially methylated genes are involved in AD pathogenesis. Studying the pathogenic mechanisms of AD has been very difficult. The recent discoveries in stem cell research have produced new

tools, like patient-derived iPSC lines and cerebral organoids, which are valuable for studying neuronal development and functions. However, even these have limitations in modelling neurodegenerative diseases, like AD, that develop during a time-span of tens of years. While iPSC lines have been derived from AD patients and these cells have disturbances in molecular mechanisms associated with AD, like amyloid-β and tau processing, the *in vitro* stem cell and organoid cultures are simplified models of very complicated neuronal tissues that are degenerating in AD. Besides, the *in vitro* neuronal cell and organoid cultures can only be maintained for a limited amount of time. Thus, they model neuronal cell and tissue development at relatively early stages making it currently impossible to study the effects of ageing and external factors over time that are very important in AD pathogenesis. (Lancaster *et al.*, 2013; Lee *et al.*, 2016; Raja *et al.*, 2016; Gonzalez *et al.*, 2018) The initial goal of the stem cell project was to establish neuronal cell and cerebral organoid cultures from patient-derived iPSC lines and study how epigenetic mechanisms and the marker genes discovered in peripheral blood DNA methylation analysis of discordant twin pairs impact AD pathogenesis. However, the organoid cultures were successfully implemented using hESC lines and this method can be invaluable in other research projects studying, for example, neurodevelopmental disorders and defects.

There is an urgent need for reliable AD models, was it human stem cell or animal-derived, to better characterise the molecular pathogenic changes that trigger the neurodegeneration in the disease. Even though there are similarities between different AD cases, like the disturbances in the processing of amyloid-β and tau, the disease is indeed multifactorial and the factors that trigger the disease may vary between individuals. Thus, patient-derived iPSC lines and disease models will be invaluable in molecular AD research to characterise the similarities and differences between individual AD cases in the future, even though the current utility of these models in AD research is limited.

7 Conclusions

The main goal of this study was to identify peripheral blood DNA methylation markers for AD and EM impairment. To achieve this goal, several methods were implemented to analyse genome-wide DNA methylation profiles from the human cell and tissue samples and to validate the findings in a targeted manner. The RRBS method was successfully implemented for genome-wide DNA methylation analysis of hESC, human brain tissue and blood samples. The RRBS results from hESC lines were validated by comparing DNA methylation profiles and mRNA and protein levels, as we discovered that expression of an important antioxidant enzyme, catalase, is silenced in abnormal cells via increased DNA methylation. RRBS was further utilised in characterising DNA methylation markers associated with AD in peripheral blood and brain tissue.

We detected peripheral blood DNA methylation differences in Finnish twin pairs discordant for AD at eleven genomic regions. One of the DMRs was also validated in the blood of discordant Swedish twin pairs and brain tissue from non-twin American AD cases versus controls. Thus, the results show that DNA methylation markers for the disease can be detected in peripheral blood, and correlate with AD-associated DNA methylation marks in the brain. According to our results, the methylation markers do not associate with gene expression differences in blood between AD patients and controls. More research is required to characterise how these methylation differences and the affected genes contribute to AD, which can give new insights to the disease pathogenesis in the future. Further studies are also needed to find out whether a peripheral blood DNA methylation marker or a combination of several markers could be sensitive and specific enough to be utilised in clinical AD diagnosis.

We found no methylation differences in the circulating plasma cfDNA fraction in twin pairs discordant for EM performance. However, in the course of this study, we collected a valuable twin sample cohort, including blood, plasma and PBMC samples from 40 individuals and PBMC-derived iPSC lines from three individuals. This sample cohort can be utilised in studying new peripheral blood biomarkers for EM impairment and functional changes in iPSC derived neuronal cell and tissue development in the future. A human PSC derived cerebral organoid tissue model was

implemented for studying epigenetic changes and molecular mechanisms in AD, however; due to current limitations in the methodology the models are more suitable for studying early nervous tissue development rather than progressive neurodegeneration. Furthermore, all the genome and epigenome research methods that were implemented during this PhD project are available for the customers of Finnish Functional Genomics Centre.

Acknowledgements

This work was carried out at Turku Bioscience Centre, University of Turku and Åbo Akademi, during years 2013-2020. I sincerely thank all the twin pairs and other individuals who have participated in our study and made this research possible.

I thank my supervisors for giving me the opportunity to work with this fascinating research. Adjunct professor Riikka Lund, PhD, your expertise on genome research made this study possible and kept it in focus. I thank you for allowing me to learn about a wide variety of research methods required in this field and guiding me through this extensive project. I hope that at least a bit of your knowledge has been transferred into me. Professor Juha Rinne, MD, PhD, I thank you for guiding me and sharing your expertise especially in the clinical aspects of Alzheimer's disease research and collecting such an invaluable and unique study material that made this study one of a kind.

I acknowledge all the co-authors of the original study articles that were included in this thesis. Special thanks to Professor Jaakko Kaprio whose invaluable work with the Finnish twin cohort studies has been crucial for this study. I thank Noora Lindgren and other personnel at Turku PET Centre for your extensive work on interviewing the twin study participants and clinical sample collection. I acknowledge Research Director, PhD Eleanor Coffey and PhD Francesca Marchisella for their help with organoid preparation and imaging.

I acknowledge Professor Qihua Tan, MD, PhD, and Docent Nina Kaminen-Ahola, PhD for reviewing this thesis. Thank you for offering your time and expertise for this work. Your feedback improved the quality of this thesis remarkably.

I thank the steering committee of my thesis, Professor Jaakko Kaprio, MD, PhD and Professor Heli Skottman, PhD, for guiding and supporting the progression of this work.

I thank all the past and present members of the Biomedical Epigenomics research group and personnel at the Finnish Functional Genomics Centre with whom I have had the pleasure to work with: Eveliina Virtanen, Sanna Vuorikoski, Miina Nurmi, Minna Niemelä, Minna Kyläinpää, Leena Kytömäki, Oso Rissanen, Päivi Junni, Riina Kaukonen, Tapio Lönnberg, Leni Kauko, Reija Venho, Samuli Laaksonen, Raili Kronström, Mikko Antinluoma and Lauri Kakko. Without your help and

expertise, this work would not have been possible. I also thank all the other personnel at Turku Bioscience Centre for the research facilities and atmosphere that promotes good science and discoveries.

I thank University of Turku personnel and doctoral students at Turku Doctoral Programme of Molecular Medicine for providing a motivating and supportive atmosphere with scientific as well as leisure- time activities, study guidance and funding possibilities. Special thanks to my friends and colleagues, Joni and Olli, for the motivational discussions during lunch and coffee breaks.

I thank my parents for providing a loving home as I grew up and always supporting me and my sisters in our life endeavours. Your love and support have made me the person that I am. My dearest sisters, Henriikka and Marjaana, I am very grateful for the bond and understanding that we share. I express my deepest love and gratitude to my fiancée, Marianna. Thank you for sharing everyday life with me and supporting and understanding me in everything that I do.

I thank the funding agencies; Turku Doctoral Programme of Molecular Medicine, Jenny and Antti Wihuri Foundation, Finnish Cultural Foundation, Varsinais-Suomi Regional Fund, State Research Funding, Hospital District of Southwest Finland, Yrjö Jahnsson Foundation and Tirkkonen family for supporting this research work.

September 2020
Mikko Konki

References

Abell, J. G. *et al.* (2018) 'Association between systolic blood pressure and dementia in the Whitehall II cohort study: role of age, duration, and threshold used to define hypertension', *European Heart Journal*. Narnia, 39(33), pp. 3119–3125. doi: 10.1093/eurheartj/ehy288.

Akalin, A. *et al.* (2012) 'methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles', *Genome Biology*. BioMed Central, 13(10), p. R87. doi: 10.1186/gb-2012-13-10-r87.

Albert, M. S. *et al.* (2011) 'The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease', *Alzheimer's & Dementia*. Elsevier, 7(3), pp. 270–279. doi: 10.1016/J.JALZ.2011.03.008.

Alberts, B. *et al.* (2008) *Molecular biology of the cell*, *Molekularbiologie der Zelle*. Garland Science.

Allegrucci, C. *et al.* (2007) 'Restriction landmark genome scanning identifies culture-induced DNA methylation instability in the human embryonic stem cell epigenome', *Human Molecular Genetics*. Narnia, 16(10), pp. 1253–1268. doi: 10.1093/hmg/ddm074.

Alzheimer's Association (2019) '2019 Alzheimer's disease facts and figures', *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 15, pp. 321–387. doi: 10.1016/j.jalz.2019.01.010.

Anastasiadou, E., Jacob, L. S. and Slack, F. J. (2018) 'Non-coding RNA networks in cancer', *Nature Reviews Cancer*. Nature Publishing Group, 18(1), pp. 5–18. doi: 10.1038/nrc.2017.99.

Bakulski, K. M. *et al.* (2012) 'Genome-wide DNA methylation differences between late-onset Alzheimer's disease and cognitively normal controls in human frontal cortex.', *Journal of Alzheimer's disease : JAD*, 29, pp. 571–88. doi: 10.3233/JAD-2012-111223.

Bannister, A. J. and Kouzarides, T. (2011) 'Regulation of chromatin by histone modifications', *Cell Research*, pp. 381–395. doi: 10.1038/cr.2011.22.

Bar, S. *et al.* (2017) 'Large-Scale Analysis of Loss of Imprinting in Human Pluripotent Stem Cells', *Cell Reports*. Cell Press, 19(5), pp. 957–968. doi: 10.1016/J.CELREP.2017.04.020.

Barrachina, M. and Ferrer, I. (2009) 'DNA Methylation of Alzheimer Disease and Tauopathy-Related Genes in Postmortem Brain', *Journal of Neuropathology & Experimental Neurology*. Narnia, 68(8), pp. 880–891. doi: 10.1097/NEN.0b013e3181af2e46.

Barski, A. *et al.* (2007) 'High-Resolution Profiling of Histone Methylations in the Human Genome', *Cell*. Cell Press, 129(4), pp. 823–837. doi: 10.1016/J.CELL.2007.05.009.

Bassil, C. F., Huang, Z. and Murphy, S. K. (2013) 'Bisulfite pyrosequencing', *Methods in Molecular Biology*, 1049, pp. 95–107. doi: 10.1007/978-1-62703-547-7_9.

Bates, D. *et al.* (2015) 'Fitting Linear Mixed-Effects Models Using lme4', *Journal of Statistical Software*, 67(1), pp. 1–48. doi: 10.18637/jss.v067.i01.

Beydoun, M. A. *et al.* (2014) 'Epidemiologic studies of modifiable factors associated with cognition and dementia: systematic review and meta-analysis', *BMC Public Health*. BioMed Central, 14(1), p. 643. doi: 10.1186/1471-2458-14-643.

Blennow, K., Leon, M. J. De and Zetterberg, H. (2006) 'Seminar Alzheimer ' s disease', 368, pp. 387–403.

Bock, C. *et al.* (2010) ‘Quantitative comparison of genome-wide DNA methylation mapping technologies’, *Nature Biotechnology*. Nature Publishing Group, 28(10), pp. 1106–1114. doi: 10.1038/nbt.1681.

Bollati, V. *et al.* (2009) ‘Decline in genomic DNA methylation through aging in a cohort of elderly subjects’, *Mechanisms of Ageing and Development*. NIH Public Access, 130(4), pp. 234–239. doi: 10.1016/j.mad.2008.12.003.

Bourque, G. *et al.* (2018) ‘Ten things you should know about transposable elements’, *Genome biology*. BioMed Central, p. 199. doi: 10.1186/s13059-018-1577-z.

Boyle, P. *et al.* (2012) ‘Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling’, *Genome Biology*. BioMed Central Ltd, 13(10), p. R92. doi: 10.1186/gb-2012-13-10-r92.

Braun, P. R. *et al.* (2019) ‘Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals’, *Translational Psychiatry*. Nature Publishing Group, 9(1), pp. 1–10. doi: 10.1038/s41398-019-0376-y.

Breitling, L. P. *et al.* (2011) ‘Tobacco-Smoking-Related Differential DNA Methylation: 27K Discovery and Replication’, *The American Journal of Human Genetics*. Cell Press, 88(4), pp. 450–457. doi: 10.1016/J.AJHG.2011.03.003.

Burdge, G. C. *et al.* (2007) ‘Dietary protein restriction of pregnant rats in the F0 generation induces altered methylation of hepatic gene promoters in the adult male offspring in the F1 and F2 generations’, *British Journal of Nutrition*. Cambridge University Press, 97(3), pp. 435–439. doi: 10.1017/S0007114507352392.

Butler, A. *et al.* (2018) ‘Integrating single-cell transcriptomic data across different conditions, technologies, and species’, *Nature Biotechnology*. Nature Publishing Group, 36(5), pp. 411–420. doi: 10.1038/nbt.4096.

Cao-Lei, L. *et al.* (2016) ‘Prenatal stress and epigenetics’, *Neuroscience and Biobehavioral Reviews*, 18 May. doi: 10.1016/j.neubiorev.2017.05.016.

Casillas, M. A. *et al.* (2003) ‘Transcriptional control of the DNA methyltransferases is altered in aging and neoplastically-transformed human fibroblasts’, *Molecular and Cellular Biochemistry*. Kluwer Academic Publishers, 252(1–2), pp. 33–43. doi: 10.1023/A:1025548623524.

Chen, C. X. *et al.* (2000) ‘A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains’, *RNA*, 6(5), pp. 755–767. doi: 10.1017/S1355838200000170.

Chen, L. *et al.* (2017) ‘Functional analysis of novel DEAF1 variants identified through clinical exome sequencing expands DEAF1-associated neurodevelopmental disorder (DAND) phenotype’, *Human Mutation*. John Wiley and Sons Inc., 38(12), pp. 1774–1785. doi: 10.1002/humu.23339.

Chin, M. H. *et al.* (2009) ‘Induced Pluripotent Stem Cells and Embryonic Stem Cells Are Distinguished by Gene Expression Signatures’, *Cell Stem Cell*. Cell Press, 5(1), pp. 111–123. doi: 10.1016/J.STEM.2009.06.008.

Chouliaras, L. *et al.* (2012) ‘Age-related increase in levels of 5-hydroxymethylcytosine in mouse hippocampus is prevented by caloric restriction.’, *Current Alzheimer research*. NIH Public Access, 9(5), pp. 536–44. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22272625 (Accessed: 25 July 2019).

Chouliaras, L., Mastroeni, D., *et al.* (2013) ‘Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer’s disease patients’, *Neurobiology of Aging*. Elsevier Ltd, 34(9), pp. 2091–2099. doi: 10.1016/j.neurobiolaging.2013.02.021.

Chouliaras, L., van den Hove, D., *et al.* (2013) ‘Histone Deacetylase 2 in the Mouse Hippocampus: Attenuation of Age- Related Increase by Caloric Restriction’, *Current Alzheimer Research*. Bentham Science Publishers, 10(8), pp. 868–876. doi: 10.2174/1567205011310080009.

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

Cronin, P. *et al.* (2017) 'Circadian alterations during early stages of Alzheimer's disease are associated with aberrant cycles of DNA methylation in BMAL1', *Alzheimer's & Dementia*. Elsevier, 13(6), pp. 689–700. doi: 10.1016/J.JALZ.2016.10.003.

Crudo, A. *et al.* (2013) 'Glucocorticoid programming of the fetal male hippocampal epigenome', *Endocrinology*. Narnia, 154(3), pp. 1168–1180. doi: 10.1210/en.2012-1980.

Davies, P. *et al.* (1998) 'Consensus report of the working group on: "Molecular and biochemical markers of Alzheimer's disease"', in *Neurobiology of Aging*, pp. 109–116. doi: 10.1016/S0197-4580(98)00022-0.

Delaval, K. *et al.* (2007) 'Differential histone modifications mark mouse imprinting control regions during spermatogenesis', *EMBO Journal*. John Wiley & Sons, Ltd, 26(3), pp. 720–729. doi: 10.1038/sj.emboj.7601513.

Dolzhenko, E. and Smith, A. D. (2014) 'Using beta-binomial regression for high-precision differential methylation analysis in multifactor whole-genome bisulfite sequencing experiments', *BMC Bioinformatics*. BioMed Central, 15(1), p. 215. doi: 10.1186/1471-2105-15-215.

Draper, J. S. *et al.* (2004) 'Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells', *Nature Biotechnology*. Nature Publishing Group, 22(1), pp. 53–54. doi: 10.1038/nbt922.

Enver, T. *et al.* (2005) 'Cellular differentiation hierarchies in normal and culture-adapted human embryonic stem cells', *Human Molecular Genetics*. Narnia, 14(21), pp. 3129–3140. doi: 10.1093/hmg/ddi345.

Fagerberg, L. *et al.* (2014) 'Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics', *Molecular and Cellular Proteomics*, 13(2), pp. 397–406. doi: 10.1074/mcp.M113.035600.

Faulk, C. and Dolinoy, D. C. (2011) 'Timing is everything: The when and how of environmentally induced changes in the epigenome of animals', *Epigenetics*, pp. 791–797. doi: 10.4161/epi.6.7.16209.

Finkel, D. *et al.* (2005) 'The longitudinal relationship between processing speed and cognitive ability: Genetic and environmental influences', *Behavior Genetics*. Kluwer Academic Publishers-Plenum Publishers, 35(5), pp. 535–549. doi: 10.1007/s10519-005-3281-5.

Fischer, A. *et al.* (2007) 'Recovery of learning and memory is associated with chromatin remodelling', *Nature*. Nature Publishing Group, 447(7141), pp. 178–182. doi: 10.1038/nature05772.

Fox, J. and Weisberg, S. (2011) *An R Companion to Applied Regression, 2nd edition*. Sage Publications.

Fraga, M. F. *et al.* (2005) 'Epigenetic differences arise during the lifetime of monozygotic twins.', *Proceedings of the National Academy of Sciences of the United States of America*, 102(30), pp. 10604–10609. doi: 10.1073/pnas.0500398102.

Fransquet, P. D. *et al.* (2018) 'Blood DNA methylation as a potential biomarker of dementia: A systematic review', *Alzheimer's and Dementia*. Elsevier, pp. 81–103. doi: 10.1016/j.jalz.2017.10.002.

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 of the United States of America*. National Academy of Sciences, 89(5), pp. 1827–31. doi: 10.1073/pnas.89.5.1827.

Gasparoni, G. *et al.* (2018) 'DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex', *Epigenetics & Chromatin*. BioMed Central, 11(1), p. 41. doi: 10.1186/s13072-018-0211-3.

Gatz, M. *et al.* (1997) 'Heritability for Alzheimer's disease: The study of dementia in Swedish twins', *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*. Narnia, 52(2), pp. M117–M125. doi: 10.1093/gerona/52A.2.M117.

Gatz, M. *et al.* (2005) 'Complete ascertainment of dementia in the Swedish Twin Registry: The HARMONY study', *Neurobiology of Aging*. Elsevier, 26(4), pp. 439–447. doi: 10.1016/j.neurobiolaging.2004.04.004.

Genin, E. *et al.* (2011) ‘APOE and Alzheimer disease: a major gene with semi-dominant inheritance’, *Molecular Psychiatry*. Nature Publishing Group, 16(9), pp. 903–907. doi: 10.1038/mp.2011.52.

Gjoneska, E. *et al.* (2015) ‘Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer’s disease’, *Nature*. Nature Publishing Group, 518(7539), pp. 365–369. doi: 10.1038/nature14252.

Goldberg, J. L. *et al.* (2004) ‘An oligodendrocyte lineage-specific semaphorin, sema5A, inhibits axon growth by retinal ganglion cells’, *Journal of Neuroscience*, 24(21), pp. 4989–4999. doi: 10.1523/JNEUROSCI.4390-03.2004.

Gonzalez, C. *et al.* (2018) ‘Modeling amyloid beta and tau pathology in human cerebral organoids’, *Molecular Psychiatry*. Nature Publishing Group, 23(12), pp. 2363–2374. doi: 10.1038/s41380-018-0229-8.

Gottesman, R. F. *et al.* (2017) ‘Association Between Midlife Vascular Risk Factors and Estimated Brain Amyloid Deposition’, *JAMA*. American Medical Association, 317(14), p. 1443. doi: 10.1001/jama.2017.3090.

Götz, J., Bodea, L. G. and Goedert, M. (2018) ‘Rodent models for Alzheimer disease’, *Nature Reviews Neuroscience*. Nature Publishing Group, pp. 583–598. doi: 10.1038/s41583-018-0054-8.

Gräff, J. *et al.* (2012) ‘An epigenetic blockade of cognitive functions in the neurodegenerating brain’, *Nature*, 483, pp. 222–226. doi: 10.1038/nature10849.

Graham, N., Arai, M. and Hagströmer, B. (2018) *multiwayvcov: Multi-Way Standard Error Clustering version 1.2.3*. Available at: https://sites.google.com/site/npgraham1/research/code (Accessed: 4 September 2019).

Grunstein, M. (1997) ‘Histone acetylation in chromatin structure and transcription’, *Nature*. Nature Publishing Group, 389(6649), pp. 349–352. doi: 10.1038/38664.

Gu, L. Z. *et al.* (2015) ‘TSNARE1 polymorphisms are associated with schizophrenia susceptibility in Han Chinese’, *Journal of Neural Transmission*. Springer-Verlag Wien, 122(6), pp. 929–932. doi: 10.1007/s00702-014-1348-2.

Gudala, K. *et al.* (2013) ‘Diabetes mellitus and risk of dementia: A meta-analysis of prospective observational studies’, *Journal of Diabetes Investigation*. John Wiley & Sons, Ltd (10.1111), 4(6), pp. 640–650. doi: 10.1111/jdi.12087.

Gund, C. *et al.* (2016) ‘Identification of a syndrome comprising microcephaly and intellectual disability but not white matter disease associated with a homozygous c.676C>T p.R226W DEAF1 mutation’, *American Journal of Medical Genetics, Part A*. Wiley-Liss Inc., 170(5), pp. 1330–1332. doi: 10.1002/ajmg.a.37580.

Guo, J. U. *et al.* (2014) ‘Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain’, *Nature Neuroscience*. Nature Publishing Group, 17(2), pp. 215–222. doi: 10.1038/nn.3607.

Guo, S. *et al.* (2017) ‘Identification of methylation haplotype blocks AIDS in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA’, *Nature Genetics*. Nature Publishing Group, 49(4), pp. 635–642. doi: 10.1038/ng.3805.

Harrison, N. J., Baker, D. and Andrews, P. W. (2007) ‘Culture adaptation of embryonic stem cells echoes germ cell malignancy’, *International Journal of Andrology*. John Wiley & Sons, Ltd (10.1111), 30(4), pp. 275–281. doi: 10.1111/j.1365-2605.2007.00762.x.

Heard, E. and Martienssen, R. A. (2014) ‘Transgenerational epigenetic inheritance: Myths and mechanisms’, *Cell*. Cell Press, pp. 95–109. doi: 10.1016/j.cell.2014.02.045.

Hebert, L. E. *et al.* (2013) ‘Alzheimer disease in the United States (2010-2050) estimated using the 2010 census.’, *Neurology*. Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology, 80(19), pp. 1778–83. doi: 10.1212/WNL.0b013e31828726f5.

Hill, P. W. S. *et al.* (2018) ‘Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte’, *Nature*. Nature Publishing Group, 555(7696), pp. 392–396. doi: 10.1038/nature25964.

Horsthemke, B. (2018) ‘A critical view on transgenerational epigenetic inheritance in humans’, *Nature Communications*. Nature Publishing Group, pp. 1–4. doi: 10.1038/s41467-018-05445-5.

Ideraabdullah, F. Y., Vigneau, S. and Bartolomei, M. S. (2008) ‘Genomic imprinting mechanisms in mammals’, *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*. Elsevier, pp. 77–85. doi: 10.1016/j.mrfmmm.2008.08.008.

van Iterson, M., van Zwet, E. W. and Heijmans, B. T. (2017) ‘Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution’, *Genome Biology*. BioMed Central, 18(1), p. 19. doi: 10.1186/s13059-016-1131-9.

Jacobsen, S. C. *et al.* (2012) ‘Effects of short-term high-fat overfeeding on genome-wide DNA methylation in the skeletal muscle of healthy young men’, *Diabetologia*. Springer-Verlag, 55(12), pp. 3341–3349. doi: 10.1007/s00125-012-2717-8.

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), p. advance online publication. doi: 10.1038/nn.3786.

Jak, A. J. *et al.* (2009) ‘Quantification of Five Neuropsychological Approaches to Defining Mild Cognitive Impairment’, *The American Journal of Geriatric Psychiatry*. Elsevier, 17(5), pp. 368–375. doi: 10.1097/JGP.0B013E31819431D5.

Jansen, I. E. *et al.* (2019) ‘Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk’, *Nature Genetics*. Nature Publishing Group, 51(3), pp. 404–413. doi: 10.1038/s41588-018-0311-9.

Järvenpää, T. *et al.* (2002) ‘Characteristics of Two Telephone Screens for Cognitive Impairment’, *Dementia and Geriatric Cognitive Disorders*. Karger Publishers, 13(3), pp. 149–155. doi: 10.1159/000048646.

Järvenpää, T. *et al.* (2003) ‘Regional cerebral glucose metabolism in monozygotic twins discordant for Alzheimer’s disease’, *Dementia and Geriatric Cognitive Disorders*, 16(4), pp. 245–252. doi: 10.1159/000072809.

Järvenpää, T. *et al.* (2004) ‘Hippocampal MRI volumetry in cognitively discordant monozygotic twin pairs.’, *Journal of neurology, neurosurgery, and psychiatry*. BMJ Publishing Group Ltd, 75(1), pp. 116–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14707319 (Accessed: 30 August 2019).

Ji, J. *et al.* (2012) ‘Elevated Coding Mutation Rate During the Reprogramming of Human Somatic Cells into Induced Pluripotent Stem Cells’, *STEM CELLS*. John Wiley & Sons, Ltd, 30(3), pp. 435–440. doi: 10.1002/stem.1011.

Jiang, S. and Mortazavi, A. (2018) ‘Integrating ChIP-seq with other functional genomics data’, *Briefings in Functional Genomics*. Narnia, 17(2), pp. 104–115. doi: 10.1093/bfgp/ely002.

Jones, P. A., Issa, J. P. J. and Baylin, S. (2016) ‘Targeting the cancer epigenome for therapy’, *Nature Reviews Genetics*. Nature Publishing Group, pp. 630–641. doi: 10.1038/nrg.2016.93.

Jufvas, Å. *et al.* (2013) ‘Global differences in specific histone H3 methylation are associated with overweight and type 2 diabetes’, *Clinical Epigenetics*. BioMed Central, 5(1), p. 15. doi: 10.1186/1868-7083-5-15.

El Kadmiri, N. *et al.* (2018) ‘Biomarkers for Alzheimer Disease: Classical and Novel Candidates’ Review’, *Neuroscience*. Elsevier Ltd, pp. 181–190. doi: 10.1016/j.neuroscience.2017.07.017.

Kaliman, P. *et al.* (2011) ‘Neurophysiological and epigenetic effects of physical exercise on the aging process’, *Ageing Research Reviews*. Elsevier, 10(4), pp. 475–486. doi: 10.1016/J.ARR.2011.05.002.

Kaminsky, Z. a *et al.* (2009) ‘DNA methylation profiles in monozygotic and dizygotic twins.’, *Nature genetics*, 41(2), pp. 240–245. doi: 10.1038/ng.286.

Kaprio, J. (2013) ‘The Finnish Twin Cohort Study: An Update’, *Twin Research and Human Genetics*. Cambridge University Press, 16(1), pp. 157–162. doi: 10.1017/thg.2012.142.

Kaprio, J. and Koskenvuo, M. (2002) 'Genetic and Environmental Factors in Complex Diseases: The Older Finnish Twin Cohort', *Twin Research*. Cambridge University Press, 5(05), pp. 358–365. doi: 10.1375/twin.5.5.358.

Khermesh, K. *et al.* (2016) 'Reduced levels of protein recoding by A-to-I RNA editing in Alzheimer's disease', *RNA*. Cold Spring Harbor Laboratory Press, 22(2), pp. 290–302. doi: 10.1261/rna.054627.115.

Kim, K. *et al.* (2016) 'Intersection of genetics and epigenetics in monozygotic twin genomes', *Methods*. Academic Press, pp. 50–56. doi: 10.1016/j.ymeth.2015.10.020.

Kobayashi, N. *et al.* (2016) 'Development of Biomarkers Based on DNA Methylation in the NCAPH2/LMF2 Promoter Region for Diagnosis of Alzheimer's Disease and Amnesic Mild Cognitive Impairment', *PLOS ONE*. Edited by T. Arendt. Public Library of Science, 11(1), p. e0146449. doi: 10.1371/journal.pone.0146449.

Koch, C. M. *et al.* (2007) 'The landscape of histone modifications across 1% of the human genome in five human cell lines', *Genome Research*, 17(6), pp. 691–707. doi: 10.1101/gr.5704207.

Korthauer, K. *et al.* (2019) 'Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing', *Biostatistics*. Narnia, 20(3), pp. 367–383. doi: 10.1093/biostatistics/kxy007.

Krueger, F. (2012) *Babraham Bioinformatics – Trim Galore!* Available at: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (Accessed: 4 September 2019).

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

Kurdyukov, S. and Bullock, M. (2016) 'DNA methylation analysis: Choosing the right method', *Biology*. MDPI AG. doi: 10.3390/biology5010003.

Lafontaine, D. L. J. and Tollervey, D. (2001) 'The function and synthesis of ribosomes', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 2(7), pp. 514–520. doi: 10.1038/35080045.

Lambert, J. C. *et al.* (2013) 'Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease.', *Nature genetics*, 45, pp. 1452–8. doi: 10.1038/ng.2802.

Lancaster, M. A. *et al.* (2013) 'Cerebral organoids model human brain development and microcephaly', *Nature*. Nature Publishing Group, 501(7467), pp. 373–379. doi: 10.1038/nature12517.

Lancaster, M. A. and Knoblich, J. A. (2014) 'Generation of cerebral organoids from human pluripotent stem cells', *Nature Protocols*. Nature Publishing Group, 9(10), pp. 2329–2340. doi: 10.1038/nprot.2014.158.

Laurent, L. C. *et al.* (2011) 'Dynamic Changes in the Copy Number of Pluripotency and Cell Proliferation Genes in Human ESCs and iPSCs during Reprogramming and Time in Culture', *Cell Stem Cell*. Cell Press, 8(1), pp. 106–118. doi: 10.1016/J.STEM.2010.12.003.

Lee, E. *et al.* (2017) 'Single-nucleotide polymorphisms are associated with cognitive decline at Alzheimer's disease conversion within mild cognitive impairment patients', *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*. Elsevier Inc, 8, pp. 86–95. doi: 10.1016/j.dadm.2017.04.004.

Lee, H.-K. *et al.* (2016) 'Three Dimensional Human Neuro-Spheroid Model of Alzheimer's Disease Based on Differentiated Induced Pluripotent Stem Cells', *PLOS ONE*. Edited by R. Yan. Public Library of Science, 11(9), p. e0163072. doi: 10.1371/journal.pone.0163072.

Li, W. *et al.* (2018) 'On the power of epigenome-wide association studies using a disease-discordant twin design', *Bioinformatics*, 34(23), pp. 4073–4078. doi: 10.1093/bioinformatics/bty532.

Lillycrop, K. A. *et al.* (2005) 'Dietary Protein Restriction of Pregnant Rats Induces and Folic Acid Supplementation Prevents Epigenetic Modification of Hepatic Gene Expression in the Offspring', *The Journal of Nutrition*. Narnia, 135(6), pp. 1382–1386. doi: 10.1093/jn/135.6.1382.

Lindgren, N. *et al.* (2019) 'Immediate verbal recall and familial dementia risk: population-based study of over 4000 twins.', *Journal of neurology, neurosurgery, and psychiatry*. BMJ Publishing Group Ltd, 90(1), pp. 90–97. doi: 10.1136/jnnp-2018-319122.

Lumey, L. H. (1992) 'Decreased birthweights in infants after maternal in utero exposure to the Dutch famine of 1944–1945', *Paediatric and Perinatal Epidemiology*. John Wiley & Sons, Ltd (10.1111), 6(2), pp. 240–253. doi: 10.1111/j.1365-3016.1992.tb00764.x.

Lun, F. M. *et al.* (2013) 'Noninvasive Prenatal Methylomic Analysis by Genomewide Bisulfite Sequencing of Maternal Plasma DNA', *Clinical Chemistry*, 59(11), pp. 1583–1594. doi: 10.1373/clinchem.2013.212274.

Lund, R. J. *et al.* (2013) 'Karyotypically abnormal human ESCs are sensitive to HDAC inhibitors and show altered regulation of genes linked to cancers and neurological diseases', *Stem Cell Research*. Elsevier, 11(3), pp. 1022–1036. doi: 10.1016/J.SCR.2013.07.002.

Lund, R. J., Närvä, E. and Lahesmaa, R. (2012) 'Genetic and epigenetic stability of human pluripotent stem cells', *Nature Reviews Genetics*. Nature Publishing Group, 13(10), pp. 732–744. doi: 10.1038/nrg3271.

Lunnon, K. *et al.* (2014) 'Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease', *Nature Neuroscience*. Nature Publishing Group, 17(9), pp. 1164–1170. doi: 10.1038/nn.3782.

Luo, C. *et al.* (2016) 'Cerebral Organoids Recapitulate Epigenomic Signatures of the Human Fetal Brain', *Cell Reports*. Elsevier B.V., 17(12), pp. 3369–3384. doi: 10.1016/j.celrep.2016.12.001.

Lynch, M. D. *et al.* (2012) 'An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment', *The EMBO Journal*. John Wiley & Sons, Ltd, 31(2), pp. 317–329. doi: 10.1038/emboj.2011.399.

Magnusson, P. K. E. *et al.* (2013) 'The Swedish Twin Registry: Establishment of a Biobank and Other Recent Developments', *Twin Research and Human Genetics*. Cambridge University Press, 16(1), pp. 317–329. doi: 10.1017/thg.2012.104.

Martins-Taylor, K. *et al.* (2011) 'Recurrent copy number variations in human induced pluripotent stem cells', *Nature Biotechnology*, 29(6), pp. 488–491. doi: 10.1038/nbt.1890.

Mastroeni, D. *et al.* (2010) 'Epigenetic changes in Alzheimer's disease: Decrements in DNA methylation', *Neurobiology of Aging*. Elsevier Inc., 31(12), pp. 2025–2037. doi: 10.1016/j.neurobiolaging.2008.12.005.

Masuyama, H. *et al.* (2015) 'The Effects of High-Fat Diet Exposure In Utero on the Obesogenic and Diabetogenic Traits Through Epigenetic Changes in Adiponectin and Leptin Gene Expression for Multiple Generations in Female Mice', *Endocrinology*. Endocrine Society, 156(7), pp. 2482–2491. doi: 10.1210/en.2014-2020.

Masuyama, H. and Hiramatsu, Y. (2012) 'Effects of a High-Fat Diet Exposure in Utero on the Metabolic Syndrome-Like Phenomenon in Mouse Offspring through Epigenetic Changes in Adipocytokine Gene Expression', *Endocrinology*. Oxford Academic, 153(6), pp. 2823–2830. doi: 10.1210/en.2011-2161.

Mattick, J. S. and Makunin, I. V. (2006) 'Non-coding RNA.', *Human molecular genetics*. Narnia, pp. R17–R29. doi: 10.1093/hmg/ddl046.

McGowan, P. O. *et al.* (2009) 'Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse', *Nature Neuroscience*. Nature Publishing Group, 12(3), pp. 342–348. doi: 10.1038/nn.2270.

McGowan, P. O. *et al.* (2011) 'Broad Epigenetic Signature of Maternal Care in the Brain of Adult Rats', *PLoS ONE*. Edited by A. Sirigu. Public Library of Science, 6(2), p. e14739. doi: 10.1371/journal.pone.0014739.

McKhann, G. M. *et al.* (2011) 'The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease', *Alzheimer's and Dementia*. Elsevier Ltd, 7(3), pp. 263–269. doi: 10.1016/j.jalz.2011.03.005.

Meissner, A. *et al.* (2005) 'Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis', *Nucleic Acids Research*, 33(18), pp. 5868–5877. doi: 10.1093/nar/gki901.

Melin, M. *et al.* (2006) 'Constitutional downregulation of SEMA5A expression in autism', *Neuropsychobiology*, 54(1), pp. 64–69. doi: 10.1159/000096040.

Mews, P. *et al.* (2019) 'Alcohol metabolism contributes to brain histone acetylation', *Nature*. Nature Publishing Group, 574(7780), pp. 717–721. doi: 10.1038/s41586-019-1700-7.

Milagro, F. I. *et al.* (2009) 'High fat diet-induced obesity modifies the methylation pattern of leptin promoter in ratsLa obesidad inducida por dieta alta en grasas modifica el patrón de metilación del promotor de la leptina en la rata', *Journal of Physiology and Biochemistry*. Springer Netherlands, 65(1), pp. 1–9. doi: 10.1007/bf03165964.

Millan, M. J. (2017) 'Linking deregulation of non-coding RNA to the core pathophysiology of Alzheimer's disease: An integrative review', *Progress in Neurobiology*. Pergamon, pp. 1–68. doi: 10.1016/j.pneurobio.2017.03.004.

Mladenova, D. *et al.* (2018) 'Adar3 is involved in learning and memory in mice', *Frontiers in Neuroscience*. Frontiers Media S.A., 12(APR). doi: 10.3389/fnins.2018.00243.

Moritz, L. E. and Trievel, R. C. (2018) 'Structure, mechanism, and regulation of polycomb-repressive complex 2', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology Inc., pp. 13805–13814. doi: 10.1074/jbc.R117.800367.

Morris, M. C. *et al.* (2015) 'MIND diet associated with reduced incidence of Alzheimer's disease', *Alzheimer's & Dementia*. Elsevier, 11(9), pp. 1007–1014. doi: 10.1016/J.JALZ.2014.11.009.

Mueller, B. R. and Bale, T. L. (2008) 'Sex-Specific Programming of Offspring Emotionality after Stress Early in Pregnancy', *Journal of Neuroscience*, 28(36), pp. 9055–9065. doi: 10.1523/jneurosci.1424-08.2008.

Muka, Taulant *et al.* (2016) 'The role of epigenetic modifications in cardiovascular disease: A systematic review', *International Journal of Cardiology*. Elsevier, 212, pp. 174–183. doi: 10.1016/j.ijcard.2016.03.062.

Muka, T. *et al.* (2016) 'The role of global and regional DNA methylation and histone modifications in glycemic traits and type 2 diabetes: A systematic review', *Nutrition, Metabolism and Cardiovascular Diseases*. Elsevier, pp. 553–566. doi: 10.1016/j.numecd.2016.04.002.

Müller, F. *et al.* (2019) 'RnBeads 2.0: comprehensive analysis of DNA methylation data', *Genome Biology*. BioMed Central, 20(1), p. 55. doi: 10.1186/s13059-019-1664-9.

Nakazawa, M. (2016) *fmsb: Functions for Medical Statistics Book with some Demographic Data version 0.6.3 from CRAN*. Available at: https://rdrr.io/cran/fmsb/ (Accessed: 4 September 2019).

Närvä, E. *et al.* (2010) 'High-resolution DNA analysis of human embryonic stem cell lines reveals culture-induced copy number changes and loss of heterozygosity', *Nature Biotechnology*. Nature Publishing Group, 28(4), pp. 371–377. doi: 10.1038/nbt.1615.

Nazor, K. L. *et al.* (2012) 'Recurrent Variations in DNA Methylation in Human Pluripotent Stem Cells and Their Differentiated Derivatives', *Cell Stem Cell*. Cell Press, 10(5), pp. 620–634. doi: 10.1016/J.STEM.2012.02.013.

Nelson, P. T. *et al.* (2011) 'Alzheimer's disease is not "brain aging": neuropathological, genetic, and epidemiological human studies', *Acta Neuropathologica*. Springer-Verlag, 121(5), pp. 571–587. doi: 10.1007/s00401-011-0826-y.

Ng, B. *et al.* (2017) 'An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome', *Nature Neuroscience*. Nature Publishing Group, 20(10), pp. 1418–1426. doi: 10.1038/nn.4632.

Nishino, K. *et al.* (2011) 'DNA Methylation Dynamics in Human Induced Pluripotent Stem Cells over Time', *PLoS Genetics*. Edited by J. M. Greally. Public Library of Science, 7(5), p. e1002085. doi: 10.1371/journal.pgen.1002085.

Oakes, E. *et al.* (2017) 'Adenosine deaminase that acts on RNA 3 (adar3) binding to glutamate receptor subunit B Pre-mRNA Inhibits RNA editing in glioblastoma', *Journal of Biological Chemistry*.

American Society for Biochemistry and Molecular Biology Inc., 292(10), pp. 4326–4335. doi: 10.1074/jbc.M117.779868.

Oberlander, T. F. *et al.* (2008) 'Prenatal exposure to maternal depression, neonatal methylation of human glucocorticoid receptor gene (NR3C1) and infant cortisol stress responses', *Epigenetics*, 3(2), pp. 97–106. doi: 10.4161/epi.3.2.6034.

Olova, N. *et al.* (2018) 'Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data', *Genome Biology*. BioMed Central, 19(1), p. 33. doi: 10.1186/s13059-018-1408-2.

Olsson, B. *et al.* (2016) 'CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis', *The Lancet Neurology*. Elsevier, 15(7), pp. 673–684. doi: 10.1016/S1474-4422(16)00070-3.

Perez, M. F. and Lehner, B. (2019) 'Intergenerational and transgenerational epigenetic inheritance in animals', *Nature Cell Biology*. Nature Publishing Group, pp. 143–151. doi: 10.1038/s41556-018-0242-9.

Pidsley, R. *et al.* (2016) 'Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling', *Genome Biology*. BioMed Central Ltd., 17(1). doi: 10.1186/s13059-016-1066-1.

Potter, C. *et al.* (2013) 'Influence of DNMT Genotype on Global and Site Specific DNA Methylation Patterns in Neonates and Pregnant Women', *PLoS ONE*. Public Library of Science, 8(10). doi: 10.1371/journal.pone.0076506.

Quintremil, S. *et al.* (2019) 'Roles of Semaphorins in Neurodegenerative Diseases', in *Neurons - Dendrites and Axons*. IntechOpen. doi: 10.5772/intechopen.82046.

R Core Development Team (2013) 'A language and environment for statistical computing.' Available at: ftp://ftp.uvigo.es/CRAN/web/packages/dplR/vignettes/intro-dplR.pdf (Accessed: 4 September 2019).

Raja, W. K. *et al.* (2016) 'Self-Organizing 3D Human Neural Tissue Derived from Induced Pluripotent Stem Cells Recapitulate Alzheimer's Disease Phenotypes', *PLOS ONE*. Edited by J. Padmanabhan. Public Library of Science, 11(9), p. e0161969. doi: 10.1371/journal.pone.0161969.

Rajab, A. *et al.* (2015) 'Recessive DEAF1 mutation associates with autism, intellectual disability, basal ganglia dysfunction and epilepsy', *Journal of Medical Genetics*. BMJ Publishing Group, 52(9), pp. 607–611. doi: 10.1136/jmedgenet-2015-103083.

Ransohoff, J. D., Wei, Y. and Khavari, P. A. (2018) 'The functions and unique features of long intergenic non-coding RNA', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 19(3), pp. 143–157. doi: 10.1038/nrm.2017.104.

Rao, J. S. *et al.* (2012) 'Epigenetic modifications in frontal cortex from Alzheimer's disease and bipolar disorder patients', *Translational Psychiatry*. Nature Publishing Group, 2(7), pp. e132–e132. doi: 10.1038/tp.2012.55.

Rasmussen, K. D. and Helin, K. (2016) 'Role of TET enzymes in DNA methylation, development, and cancer', *Genes and Development*, pp. 733–750. doi: 10.1101/gad.276568.115.

Rathore, S. *et al.* (2017) 'A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages', *NeuroImage*. Academic Press, 155, pp. 530–548. doi: 10.1016/J.NEUROIMAGE.2017.03.057.

Ricobaraza, A. *et al.* (2012) 'Phenylbutyrate rescues dendritic spine loss associated with memory deficits in a mouse model of Alzheimer disease', *Hippocampus*. John Wiley & Sons, Ltd, 22(5), pp. 1040–1050. doi: 10.1002/hipo.20883.

Ridge, P. G. *et al.* (2016) 'Assessment of the genetic variance of late-onset Alzheimer's disease', *Neurobiology of Aging*. Elsevier, 41, pp. 200.e13-200.e20. doi: 10.1016/J.NEUROBIOLAGING.2016.02.024.

Rönnemaa, E. *et al.* (2011) 'Vascular Risk Factors and Dementia: 40-Year Follow-Up of a Population-Based Cohort', *Dementia and Geriatric Cognitive Disorders*. Karger Publishers, 31(6), pp. 460–466. doi: 10.1159/000330020.

Roseboom, T., de Rooij, S. and Painter, R. (2006) 'The Dutch famine and its long-term consequences for adult health', *Early Human Development*, pp. 485–491. doi: 10.1016/j.earlhumdev.2006.07.001.

Rothschild, G. and Basu, U. (2017) 'Lingering Questions about Enhancer RNA and Enhancer Transcription-Coupled Genomic Instability', *Trends in Genetics*. Elsevier Current Trends, 33(2), pp. 143–154. doi: 10.1016/J.TIG.2016.12.002.

Rumbaugh, G. *et al.* (2015) 'Pharmacological Selectivity Within Class I Histone Deacetylases Predicts Effects on Synaptic Function and Memory Rescue', *Neuropsychopharmacology*. Nature Publishing Group, 40(10), pp. 2307–2316. doi: 10.1038/npp.2015.93.

Rusanen, M. *et al.* (2011) 'Heavy Smoking in Midlife and Long-term Risk of Alzheimer Disease and Vascular Dementia', *Archives of Internal Medicine*. American Medical Association, 171(4), pp. 333–339. doi: 10.1001/archinternmed.2010.393.

Saczynski, J. S. *et al.* (2006) 'The Effect of Social Engagement on Incident Dementia', *American Journal of Epidemiology*. Narnia, 163(5), pp. 433–440. doi: 10.1093/aje/kwj061.

Sanchez-Mut, J. V. *et al.* (2013) 'DNA methylation map of mouse and human brain identifies target genes in Alzheimer's disease', *Brain*. Narnia, 136(10), pp. 3018–3027. doi: 10.1093/brain/awt237.

Sanchez-Mut, J. V. *et al.* (2014) 'Promoter hypermethylation of the phosphatase DUSP22 mediates PKA-dependent TAU phosphorylation and CREB activation in Alzheimer's disease', *Hippocampus*. John Wiley & Sons, Ltd, 24(4), pp. 363–368. doi: 10.1002/hipo.22245.

Sando, S. B. *et al.* (2008) 'Risk-reducing effect of education in Alzheimer's disease', *International Journal of Geriatric Psychiatry*. John Wiley & Sons, Ltd, 23(11), pp. 1156–1162. doi: 10.1002/gps.2043.

Sapienza, C. and Issa, J.-P. (2016) 'Diet, Nutrition, and Cancer Epigenetics', *Annual Review of Nutrition*. Annual Reviews, 36(1), pp. 665–681. doi: 10.1146/annurev-nutr-121415-112634.

Savva, Y. A., Rieder, L. E. and Reenan, R. A. (2012) 'The ADAR protein family', *Genome biology*, p. 252. doi: 10.1186/gb-2012-13-12-252.

Scheinin, N. M. *et al.* (2011) 'Early detection of Alzheimer disease: 11C-PiB PET in twins discordant for cognitive impairment', *Neurology*. Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology, 77(5), pp. 453–460. doi: 10.1212/WNL.0b013e318225118e.

Schmitt, A. D., Hu, M. and Ren, B. (2016) 'Genome-wide mapping and analysis of chromosome architecture', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, pp. 743–755. doi: 10.1038/nrm.2016.104.

Schoenfelder, S. and Fraser, P. (2019) 'Long-range enhancer–promoter contacts in gene expression control', *Nature Reviews Genetics*. Nature Publishing Group. doi: 10.1038/s41576-019-0128-0.

Schultz, M. D. *et al.* (2015) 'Human body epigenome maps reveal noncanonical DNA methylation variation', *Nature*. Nature Publishing Group, 523(7559), pp. 212–216. doi: 10.1038/nature14465.

Semick, S. A. *et al.* (2019) 'Integrated DNA methylation and gene expression profiling across multiple brain regions implicate novel genes in Alzheimer's disease', *Acta Neuropathologica*. Springer Berlin Heidelberg, 137(4), pp. 557–569. doi: 10.1007/s00401-019-01966-5.

Sen, P. *et al.* (2016) 'Epigenetic Mechanisms of Longevity and Aging', *Cell*. Cell Press, 166(4), pp. 822–839. doi: 10.1016/J.CELL.2016.07.050.

Sharp, S. J. *et al.* (1985) 'Structure and Transcription of Eukaryotic tRNA Gene', *Critical Reviews in Biochemistry*. Taylor & Francis, 19(2), pp. 107–144. doi: 10.3109/10409238509082541.

Shinagawa, S. *et al.* (2016) 'DNA methylation in the NCAPH2/LMF2 promoter region is associated with hippocampal atrophy in Alzheimer's disease and amnesic mild cognitive impairment patients', *Neuroscience Letters*. Elsevier, 629, pp. 33–37. doi: 10.1016/J.NEULET.2016.06.055.

Siegmund, K. D. *et al.* (2007) 'DNA Methylation in the Human Cerebral Cortex Is Dynamically Regulated throughout the Life Span and Involves Differentiated Neurons', *PLoS ONE*. Edited by T. Reh. Public Library of Science, 2(9), p. e895. doi: 10.1371/journal.pone.0000895.

Sleiman, P. *et al.* (2013) 'GWAS meta analysis identifies TSNARE1 as a novel Schizophrenia / Bipolar susceptibility locus', *Scientific Reports*. Nature Publishing Group, 3. doi: 10.1038/srep03075.

Smith, A. R. *et al.* (2016) 'Increased DNA methylation near TREM2 is consistently seen in the superior temporal gyrus in Alzheimer's disease brain', *Neurobiology of Aging*. Elsevier, 47, pp. 35–40. doi: 10.1016/J.NEUROBIOLAGING.2016.07.008.

Smith, A. R., Smith, R. G., Burrage, J., *et al.* (2019) 'A cross-brain regions study of ANK1 DNA methylation in different neurodegenerative diseases', *Neurobiology of Aging*. Elsevier, 74, pp. 70–76. doi: 10.1016/J.NEUROBIOLAGING.2018.09.024.

Smith, A. R., Smith, R. G., Pishva, E., *et al.* (2019) 'Parallel profiling of DNA methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in Alzheimer's disease', *Clinical Epigenetics*. BioMed Central, 11(1), p. 52. doi: 10.1186/s13148-019-0636-y.

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, 14(12), pp. 1580–1588. doi: 10.1016/J.JALZ.2018.01.017.

Solomon, A. *et al.* (2009) 'Midlife Serum Cholesterol and Increased Risk of Alzheimer’s and Vascular Dementia Three Decades Later', *Dementia and Geriatric Cognitive Disorders*. Karger Publishers, 28(1), pp. 75–80. doi: 10.1159/000231980.

Song, Q. *et al.* (2013) 'A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics.', *PloS one*. Public Library of Science, 8(12), p. e81148. doi: 10.1371/journal.pone.0081148.

Sood, S. *et al.* (2015) 'A novel multi-tissue RNA diagnostic of healthy ageing relates to cognitive health status', *Genome Biology*. BioMed Central Ltd., 16(1). doi: 10.1186/s13059-015-0750-x.

Sotaniemi, M. *et al.* (2012) 'CERAD-neuropsychological battery in screening mild Alzheimer's disease', *Acta Neurologica Scandinavica*. John Wiley & Sons, Ltd (10.1111), 125(1), pp. 16–23. doi: 10.1111/j.1600-0404.2010.01459.x.

Sperling, R. A. *et al.* (2011) 'Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease', *Alzheimer's & Dementia*. Elsevier, 7(3), pp. 280–292. doi: 10.1016/J.JALZ.2011.03.003.

Stephen, R. *et al.* (2017) 'Physical Activity and Alzheimer's Disease: A Systematic Review', *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*. Narnia, 72(6), p. glw251. doi: 10.1093/gerona/glw251.

Stirzaker, C. *et al.* (2014) 'Mining cancer methylomes: Prospects and challenges', *Trends in Genetics*, pp. 75–84. doi: 10.1016/j.tig.2013.11.004.

Sun, K. *et al.* (2015) 'Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 112(40), pp. E5503–E5512. doi: 10.1073/pnas.1508736112.

Sun, S. *et al.* (2019) 'Heritability estimation and differential analysis of count data with generalized linear mixed models in genomic sequencing studies', *Bioinformatics*, 35(3), pp. 487–496. doi: 10.1093/bioinformatics/bty644.

Szekeres, M. *et al.* (2016) 'Relevance of defensin β-2 and α defensins (HNP1-3) in Alzheimer's disease', *Psychiatry Research*. Elsevier Ireland Ltd, 239, pp. 342–345. doi: 10.1016/j.psychres.2016.03.045.

Taapken, S. M. *et al.* (2011) 'Karotypic abnormalities in human induced pluripotent stem cells and embryonic stem cells', *Nature Biotechnology*, pp. 313–314. doi: 10.1038/nbt.1835.

Therneau, T. M. (2015) 'A Package for Survival Analysis in S'. Available at: https://cran.r-project.org/package=survival.

Therneau, T. M. (2018) 'coxme: Mixed Effects Cox Models'. Available at: https://cran.r-project.org/package=coxme.

Uszczynska-Ratajczak, B. *et al.* (2018) 'Towards a complete map of the human long non-coding RNA transcriptome', *Nature Reviews Genetics*. Nature Publishing Group, 19(9), pp. 535–548. doi: 10.1038/s41576-018-0017-y.

Vagelatos, N. T. and Eslick, G. D. (2013) 'Type 2 Diabetes as a Risk Factor for Alzheimer's Disease: The Confounders, Interactions, and Neuropathology Associated With This Relationship', *Epidemiologic Reviews*. Narnia, 35(1), pp. 152–160. doi: 10.1093/epirev/mxs012.

Villemagne, V. L. *et al.* (2013) 'Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study', *The Lancet Neurology*. Elsevier, 12(4), pp. 357–367. doi: 10.1016/S1474-4422(13)70044-9.

Virta, J. J. *et al.* (2008) 'Cerebral glucose metabolism in dizygotic twin pairs discordant for Alzheimer's disease', *Dementia and Geriatric Cognitive Disorders*. Karger Publishers, 25(1), pp. 9–16. doi: 10.1159/000111114.

Wang, R. and Reddy, P. H. (2017) 'Role of Glutamate and NMDA Receptors in Alzheimer's Disease', *Journal of Alzheimer's Disease*. NIH Public Access, pp. 1041–1048. doi: 10.3233/JAD-160763.

Watson, C. T. *et al.* (2016) 'Genome-wide DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease', *Genome Medicine*. BioMed Central, 8(1), p. 5. doi: 10.1186/s13073-015-0258-8.

Weber, M. *et al.* (2007) 'Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome', *Nature Genetics*. Nature Publishing Group, 39(4), pp. 457–466. doi: 10.1038/ng1990.

Weissbein, U. *et al.* (2017) 'Culture-induced recurrent epigenetic aberrations in human pluripotent stem cells', *PLOS Genetics*. Edited by A. Clark. Public Library of Science, 13(8), p. e1006979. doi: 10.1371/journal.pgen.1006979.

Weissbein, U. *et al.* (2019) 'Genome-wide Screen for Culture Adaptation and Tumorigenicity-Related Genes in Human Pluripotent Stem Cells', *iScience*. Elsevier, 11, pp. 398–408. doi: 10.1016/J.ISCI.2018.12.031.

Weissbein, U., Benvenisty, N. and Ben-David, U. (2014) 'Genome maintenance in pluripotent stem cells', *Journal of Cell Biology*. Rockefeller University Press, pp. 153–163. doi: 10.1083/jcb.201310135.

Wen, K. X. *et al.* (2016) 'The role of DNA methylation and histone modifications in neurodegenerative diseases: A systematic review', *PLoS ONE*. Edited by A. Kumar. Public Library of Science, 11(12), p. e0167201. doi: 10.1371/journal.pone.0167201.

Werbowetski-Ogilvie, T. E. *et al.* (2009) 'Characterization of human embryonic stem cells with features of neoplastic progression', *Nature Biotechnology*. Nature Publishing Group, 27(1), pp. 91–97. doi: 10.1038/nbt.1516.

West, R. L., Lee, J. M. and Maroun, L. E. (1995) 'Hypomethylation of the amyloid precursor protein gene in the brain of an alzheimer's disease patient', *Journal of Molecular Neuroscience*. Humana Press, 6(2), pp. 141–146. doi: 10.1007/BF02736773.

WHO (2019) *Dementia fact sheet*. Available at: https://www.who.int/en/news-room/fact-sheets/detail/dementia (Accessed: 1 August 2019).

Wilson, R. S. *et al.* (2011) 'Heritability of different forms of memory in the Late Onset Alzheimer's Disease Family Study.', *Journal of Alzheimer's disease : JAD*. NIH Public Access, 23(2), pp. 249–55. doi: 10.3233/JAD-2010-101515.

Winblad, B. *et al.* (2016) 'Defeating Alzheimer's disease and other dementias: a priority for European science and society.', *The Lancet. Neurology*. Elsevier, 15(5), pp. 455–532. doi: 10.1016/S1474-4422(16)00062-4.

Wu, T. P. *et al.* (2016) 'DNA methylation on N(6)-adenine in mammalian embryonic stem cells.', *Nature*. NIH Public Access, 532(7599), pp. 329–33. doi: 10.1038/nature17640.

Yang, S. *et al.* (2008) 'Tumor progression of culture-adapted human embryonic stem cells during long-term culture', *Genes, Chromosomes and Cancer*. John Wiley & Sons, Ltd, 47(8), pp. 665–679. doi: 10.1002/gcc.20574.

Yates, L. A. *et al.* (2016) ‘Cognitive leisure activities and future risk of cognitive impairment and dementia: systematic review and meta-analysis’, *International Psychogeriatrics*. Cambridge University Press, 28(11), pp. 1791–1806. doi: 10.1017/S1041610216001137.

Young, M. A. *et al.* (2012) ‘Background Mutations in Parental Cells Account for Most of the Genetic Heterogeneity of Induced Pluripotent Stem Cells’, *Cell Stem Cell*. Cell Press, 10(5), pp. 570–582. doi: 10.1016/J.STEM.2012.03.002.

Yu, X., Wang, F. and Zhang, J.-P. (2014) ‘Meta analysis of the association of rs7702187 SNP in SEMA5A gene with risk of Parkinson’s disease.’, *European review for medical and pharmacological sciences*, 18(6), pp. 900–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24706317 (Accessed: 2 November 2019).

Zeileis, A. and Hothorn, T. (2015) ‘Diagnostic Checking in Regression Relationships’. Available at: https://www.semanticscholar.org/paper/Diagnostic-Checking-in-Regression-Relationships-Zeileis-Hothorn/4b9f2a6954c05fd2bc228a701d89d598e18b9be5 (Accessed: 4 September 2019).

Zemach, A. *et al.* (2010) ‘Genome-wide evolutionary analysis of eukaryotic DNA methylation’, *Science*. American Association for the Advancement of Science, 328(5980), pp. 916–919. doi: 10.1126/science.1186366.

Zhang, L. *et al.* (2020) ‘Advance of sporadic Alzheimer’s disease animal models’, *Medicinal Research Reviews*. John Wiley and Sons Inc., pp. 431–458. doi: 10.1002/med.21624.

Zheng, H. and Xie, W. (2019) ‘The role of 3D genome organization in development and cell differentiation’, *Nature Reviews Molecular Cell Biology*. Nature Publishing Group. doi: 10.1038/s41580-019-0132-4.

Ziller, M. J. *et al.* (2011) ‘Genomic distribution and Inter-Sample variation of Non-CpG methylation across human cell types’, *PLoS Genetics*. Edited by D. Schübeler. Public Library of Science, 7(12), p. e1002389. doi: 10.1371/journal.pgen.1002389.

Ziller, M. J. *et al.* (2013) ‘Charting a dynamic DNA methylation landscape of the human genome’, *Nature*. Nature Publishing Group, 500(7463), pp. 477–481. doi: 10.1038/nature12433.

Ziller, M. J. *et al.* (2015) ‘Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing’, *Nature Methods*. Nature Publishing Group, 12(3), pp. 230–232. doi: 10.1038/nmeth.3152.

Zovkic, I. B., Guzman-Karlsson, M. C. and Sweatt, J. D. (2013) ‘Epigenetic regulation of memory formation and maintenance.’, *Learning & memory (Cold Spring Harbor, N.Y.)*. Cold Spring Harbor Laboratory Press, 20(2), pp. 61–74. doi: 10.1101/lm.026575.112.

UNIVERSITY OF TURKU

ISBN 978-951-29-8171-7 (PRINT)
ISBN 978-951-29-8172-4 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)

Painosalama Oy, Turku, Finland 2020