Identifying DNA Methylation Signatures in Alzheimer's Disease Blood

Submitted by Janou Anna Yvonne Roubroeks

to the University of Exeter

as a thesis for the degree of

Doctor of Philosophy in Medical Studies

In June 2021

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: ……..

ABSTRACT

Alzheimer's disease is a complex, multifaceted disorder, which is estimated to affect over thirty six million people worldwide and is characterised by progressive neurodegeneration and cognitive decline. As the numerous genomic susceptibility loci that have been identified for the most common, sporadic form of Alzheimer's disease do not fully account for the disease risk, epigenetic and environmental factors have been suggested to be involved in the aetiology and development of Alzheimer's disease. Epigenetics, of which DNA methylation is perhaps the most studied mechanism, refers to transient, heritable changes in gene expression without the underlying genotype being altered. A growing number of epigenome-wide association studies have demonstrated robust differential DNA methylation in the brain of Alzheimer's disease patients, though limited studies have been undertaken in blood. The aim of this thesis was to characterise blood DNA methylation profiles in Alzheimer's disease, as well as individuals with mild cognitive impairment, who often progress to Alzheimer's disease. Disease-associated profiles were characterised on autosomal chromosomes as well as sex chromosomes, and the effects and interactions of the Alzheimer's disease risk factors sex and age were studied. The results from this thesis have provided novel insights into DNA methylation changes in blood related to Alzheimer's disease, mild cognitive impairment, and future progression to Alzheimer's disease. A region in the *HOXB6* gene was found to be differentially methylated in Alzheimer's disease, which presents an interesting target for future diagnostic biomarker studies. The results concerning the risk factors sex and age, and DNA methylation of the sex chromosomes, emphasise the importance of not only controlling for, but taking into account these factors.

Aan Joep,

Doe wóls minimaal eine paragraaf aan dich gewiedj in mien dankwaord,

van mich kriegs se het ganse book.

ACKNOWLEDGEMENTS

I count myself lucky to be surrounded by so many people who have advised me, supported me, or cheered me on throughout this journey. To all those people I would like to say: thank you.

First and foremost, I want to thank my supervisors, Professor Katie Lunnon, Professor Daniël van den Hove, and Dr. Liz Coulthard. Particularly Katie and Daniël, you have been absolutely wonderful supervisors, and your wisdom, support and advice have been invaluable. Even from the early days of my master's degree, you have both given me the opportunities to get me started on this journey, and I know that without you none of this would have been possible. You are both inspirational scientists, and I am very grateful to have ended up as your PhD student. Also, Daniël, thank you for all the mottos. Science *is* fun, and I will forever live by '*Alles kump good, mer neet vandaag'*!

I would like to acknowledge my funders, the MRC GW4 BioMed DTP, for my studentship and the ARUK for additional funding support during my PhD.

I would like to thank everyone within the Complex Disease Epigenetics group in Exeter, as well as the Neuroepigenetics group in Maastricht. You have all made the day to day work fun, and while I was essentially living separate lives in two separate countries, you have made me feel at home in both. Bex, there is so much to thank you for. First of all, I want to thank you for all the bioinformatics support and advice you've given me! Most of all, thank you for sharing your home with me these past few years. I have loved living with you, our game nights, and

the times where we just went to grab a drink and ended up standing in the hallway for ages just discussing everything from cheese to games to politics. Jenny, my (business) partner-in-crime and desk neighbour, never have I been so sad over a simple desk move, even when it was just a few metres separating us back then! I have loved talking about our shared interests from murder mysteries to funny socks, spending karaoke nights at the Ship and teaming up to bring some 'joy' to the desks of Adam and Sam (though they might see it differently). Sam, you have warned (threatened) me to keep my acknowledgements short, so thank you. Thank you for all the fun we had at work and outside of work, for always listening to my rants, and sharing your rants with me (I always like it when your sentences start with 'So' or 'Right'). You and Jenny have been the most supportive and amazing friends, and I am so glad that I met you both! I have so many wonderful memories of my time in Exeter with you, and I expect to make many more with the both of you in the future! To Adam, my go-to lab person, and my first friend in Exeter: thank you for all your help, your friendship, and for never, ever making me frown (I am sure the tally is at 0). I have loved all of our random discussions about absolutely nothing, singing stupid songs in the lab, and all the *a-ma-zing* science we have done together! Ehsan, thank you for all the help you've given me throughout this PhD, and most of all, thank you for your friendship. It's funny to think that we didn't really meet until I went to Exeter, when we had been working in the same group in Maastricht, but I'm so glad we did! I also want to thank you for the STATA analysis you ran for my master's thesis (better late than never, right?). Joe and Nyree, from the moment I met you, I instantly knew I had made two great friends. Thank you for sharing the beauty of Devon with me through all our walks on Dartmoor, for all the fun evenings spent at your home (making me cry with laughter from playing CAH), or just relaxing at the Pig with a

pot of tea and some piggy fours! I hope you can visit us in the Netherlands soon! Finally, before I switch to Limburgs, I want to thank everyone I started this journey with, Stefania, Philippos, Roy, Artemis, and Matt, and especially those I met later: Greg, Lachlan, Michael, Rob, Josh, and Leo. I have loved our nights out, but also the chaos in the office and all the completely random and inappropriate conversations. I doubt any other office will ever live up to the standards you have set!

Ich wil ouch mien leef femielje bedanke, mien sjoeanajers, en mien vrunj, de geitjes en de bukskes, die mich ummer höbbe aangemoedigdj en veural veur de nuuedige ontspanning höbbe gezurg! Oma, geer höb mich altied laote zeen waat doorzitte en volhaaje is, en det d'r veur èlk probleem waal ein oplossing te vinje is (desnoeads mit 10-seconde liem en duct tape). Det doorzittingsvermogen en probleemoplossendj vermoge höb ich bie deze PhD zeker nuuedig gadj! Aan Sharon, Daan, en natuurlik Joep: geer zeentj de bèste vrunj en bonus-breurs en -zus die ich mich oeajt haw kinne winse. Veer kinne ós al zoea lang det wäörd meistal neet mieër nuuedig zeen, mer toch mot ’t soms gezag waere: danke, veur alles.

Pap en mam, det ich dit höb bereik is gans aan uch te danke. Van jóngs aaf aan wól ich altied precies weite wie alles inein zoot en werktj, en geer höb mien nuujsjierigheid ummer aangemoedigd en gestimuleerd. Op uch steun en leefdje kin ich altied raekene, en geer höb mich ouch de kanse gegaeve óm dit allemaol te bereike: zonger uch waas dit noeajt meugelik gewaes.

Aan Kevin, d'r zeen gein wäörd genóg óm dig te bedanke of óm te zègke wie belangriek se veur mich bös. Dich höbs altied het volste vertroewe in mich gadj, zelfs op momente det ich det zelf neet haw. Doe höbs mich altied gesteund es ich door mos werke, en ouch gezurg det ich de ontspanning haw die ich nuuedig haw. Doe bös miene beste vrundj, mien (baetere) wederhelft, en ich haaj ontzettendj vääl van dich.

TABLE OF CONTENTS

ABSTRACT **2**

ACKNOWLEDGEMENTS **4**

TABLE OF FIGURES **15**

TABLE OF TABLES **21**

PUBLICATIONS ARISING FROM THIS THESIS **25**

DECLARATIONS **28**

ABBREVIATIONS **29**

CHAPTER 1. INTRODUCTION **33**

1.1. Alzheimer's disease 34

1.1.1. Types of AD 34

1.1.2. Epidemiology and burden of disease 35

1.1.2.1. Social and economic burden 36

1.1.3. Hallmarks of AD 37

1.1.4. Aetiology of AD 38

1.1.4.1. Causal mechanisms of FAD 38

1.1.4.2. The amyloid cascade hypothesis 41

1.1.4.3. The tau hypothesis 42

1.1.4.4. The inflammation hypothesis 45

1.1.5. Risk factors of AD 45

1.1.6. Diagnosis of AD 47

1.1.6.1. Mild Cognitive Impairment as a precursor of AD 48

1.1.6.2. Biomarkers of preclinical AD 50

1.2. Epigenetic Mechanisms 52

1.2.1. DNA modifications 52

1.2.2. Other epigenetic mechanisms 56

1.3. Epigenetics in Alzheimer's disease 57
1.3.1. Global measures of DNA methylation and hydroxymethylation 57
1.3.2. Candidate gene studies of DNA methylation 61
1.3.3. Epigenome-wide association studies of AD 64
1.4. Aims 72
CHAPTER 2. MATERIALS AND METHODS 73
2.1. The AddNeuroMed cohort 74
2.1.1. AddNeuroMed cohort demographics 74
2.2. Genome-wide DNA methylation profiling in AddNeuroMed 77
2.2.1. DNA preparation 77
2.2.2. Illumina HumanMethylation 450K arrays 77
2.2.3. Data pre-processing 81
2.2.3.1. Bisulfite conversion efficiency 84
2.2.3.2. Genetic distinctness of samples 86
2.2.3.3. Removal of cross-hybridising and SNP probes 88
2.2.3.4. Multidimensional scaling of sex chromosomes 88
2.2.3.5. P-filter 90
2.2.3.6. Outlier removal 90
2.2.3.7. Normalisation and final QC 92
2.2.4. Identification of covariates in the AddNeuroMed cohort 95
2.3. The ADNI cohort 102
2.3.1. DNA methylation profiling and sample selection in ADNI 103
2.3.2. ADNI data pre-processing 107
2.3.3. Identification of covariates in the ADNI cohort 116
2.4. General data analysis methods 122
2.4.1. Differential methylation analysis 122

2.4.2. Weighted gene correlation network analysis (WGCNA) 123
2.4.2.1. Generation of correlation networks in AddNeuroMed 123
2.5. Pyrosequencing validation 128
2.5.1. DNA bisulfite treatment 128
2.5.1.1. Sample preparation 129
2.5.1.2. Reagent preparation 129
2.5.1.3. Sodium bisulfite conversion 131
2.5.2. Pyrosequencing assay design 132
2.5.2.1. Polymerase chain reaction (PCR) 135
2.5.2.2. Gel electrophoresis 137
2.5.2.3. Assay optimisation and quality control 138
2.5.3. Pyrosequencing 141
2.5.3.1. Immobilising the PCR product to beads 141
2.5.3.2. Preparation of the vacuum workstation: 142
2.5.3.3. Sequencing primer preparation: 142
2.5.3.4. Combining the PCR products and sequencing primer 142
2.5.3.5. Setting up the pyrosequencer 143
2.5.4. Pyrosequencing data analysis 145
CHAPTER 3. AN EPIGENOME-WIDE ASSOCIATION STUDY OF ALZHEIMER'S DISEASE BLOOD HIGHLIGHTS ROBUST DNA HYPERMETHYLATION IN THE *HOXB6* GENE 146
3.1. Introduction 147
3.2. Aims 150
3.3. Materials and methods 151
3.3.1. Subjects 151

3.3.2. DNA methylation analysis 151
3.3.3. Generation of weighted gene correlation networks 154
3.3.4. Association of modules to traits of interest 154
3.3.5. Module membership and probe significance 155
3.3.6. Analysis of gene expression data and association with methylation data 156
3.3.7. Validation of the *HOXB6* DMR using pyrosequencing 157
3.4. Results 159
3.4.1. Identification of differentially methylated loci in MCI and AD blood 159
3.4.2. A number of significant DMRs can be identified in MCI and AD blood 165
3.4.3. Validation of the AD-associated DMR in *HOXB6* by pyrosequencing 182
3.4.4. Transcriptional differences in genes containing DMRs 187
3.4.5. Clusters of methylated loci associated with MCI and AD 195
3.4.6. Functional role of modules associated with MCI and AD 200
3.4.7. Investigating clusters of co-methylated loci associated with progression to AD 205
3.5. Discussion 207
CHAPTER 4. DNA METHYLOMIC PATTERNS IN BLOOD ASSOCIATED WITH AGE, SEX, AND ALZHEIMER'S DISEASE 212
4.1. Introduction 213
4.2. Aims 217
4.3. Materials and methods 218
4.3.1. Sex, age, and disease interaction model for identifying DMPs and DMRs 218
4.3.2. Replication of DMPs and DMRs in ADNI 219
4.3.3. Identification of altered biological pathways 219

4.4. Results 221
4.4.1. The effects of sex and age on DNA methylation 221
4.4.2. Interactions between sex, age, and diagnosis 229
4.4.3. Sex, age, and interaction DMRs 238
4.4.4. Replication of DMRs in ADNI 247
4.4.5. Pathways altered in association with sex, age and their interactions with disease 248
4.5. Discussion 254
CHAPTER 5. DNA METHYLATION ON THE X AND Y CHROMOSOME IN ALZHEIMER'S DISEASE AND MILD COGNITIVE IMPAIRMENT 259
5.1. Introduction 260
5.2. Aims 263
5.3. Materials and methods 264
5.3.1. Stratified analysis of the X and Y chromosomes 264
5.3.2. Analysis of the Xi DNA methylation profile 266
5.3.3. Correction of Xi inflation 268
5.3.4. Replication analysis 269
5.4. Results 272
5.4.1. Differentially methylated positions on the X and Y chromosomes in AddNeuroMed 272
5.4.1.1. Diagnostic group-specific DMPs in AddNeuroMed 277
5.4.2. Replication of X and Y linked DMPs in the ADNI cohort 285
5.4.3. DMPs on Xi in AddNeuroMed 292
5.4.4. Replication of Xi DMPs in ADNI 299
5.4.4.1. Diagnostic group-specific Xi DMPs replicated in ADNI 302
5.4.5. Regions on the Xi chromosome associated with diagnosis in AddNeuroMed 305
5.4.6. Replication of Xi DMRs in ADNI 312

5.5. Discussion 321

CHAPTER 6. DISCUSSION 327

6.1. Introduction 328

6.2. Key Findings from this Thesis 329

6.2.1. Chapter 3: An Epigenome-Wide Association Study of Alzheimer's Disease Blood Highlights Robust DNA Hypermethylation in the *HOXB6* Gene 329

6.2.2. Chapter 4: DNA Methylomic Patterns in Blood Associated with Age, Sex and Alzheimer's Disease 330

6.2.3. Chapter 5: DNA Methylation on the X And Y Chromosome in Alzheimer's Disease and Mild Cognitive Impairment 331

6.3. Strengths, limitations and future perspectives 334

6.3.1. Strengths 334

6.3.2. Limitations 335

6.3.2.1. Sample size 335

6.3.2.2. Cohort limitations 335

6.3.2.3. DNA methylation array limitations 337

6.3.2.4. Limitations related to methodology 338

6.3.2.5. Tissue- and cell-heterogeneity 339

6.3.2.6. Blood as a surrogate for brain tissue 340

6.3.2.7. Outlook on future perspectives 340

6.4. Conclusions 343

APPENDIX A: EPIGENETICS AND DNA METHYLOMIC PROFILING IN ALZHEIMER'S DISEASE AND OTHER NEURODEGENERATIVE DISEASES 344

APPENDIX B: AN EPIGENOME-WIDE ASSOCIATION STUDY OF ALZHEIMER'S DISEASE BLOOD HIGHLIGHTS ROBUST DNA HYPERMETHYLATION IN THE *HOXB6* GENE 358

APPENDIX C: ALZHEIMER'S DISEASE-ASSOCIATED (HYDROXY)METHYLOMIC CHANGES IN THE BRAIN AND BLOOD..... 379

APPENDIX D: XI DMPS ASSOCIATED WITH CTL, MCI, AND AD IN ADDNEUROMED..... 395

APPENDIX E: XI DMPS ASSOCIATED WITH AD IN ADDNEUROMED..... 401

APPENDIX F: XI DMPS ASSOCIATED WITH MCI IN ADDNEUROMED..... 404

APPENDIX G: XI DMPS ASSOCIATED WITH AD RELATIVE TO MCI IN ADDNEUROMED..... 408

APPENDIX H: DMRS ON THE XI CHROMOSOME ASSOCIATED WITH DIAGNOSTIC STATUS OF CTL, MCI, OR AD..... 411

REFERENCES..... 417

TABLE OF FIGURES

Figure 1.1 The amyloidogenic and non-amyloidogenic pathways of APP cleavage. 40

Figure 1.2 The formation of neurofibrillary tangles (NFTs) in AD. 44

Figure 1.3 Hypothetical model of biomarker presence prior to development of dementia..... 49

Figure 1.4 Cycle of cytosine modifications. 55

Figure 2.1 Illumina HumanMethylation 450K probe types. 80

Figure 2.2 Raw median methylated and unmethylated signal intensities in AddNeuroMed samples. 82

Figure 2.3 Density plots of raw β-values in the AddNeuroMed cohort. 83

Figure 2.4 Bisulfite conversion efficiency in AddNeuroMed samples. 85

Figure 2.5 Genetic distinctness of samples in the AddNeuromed cohort. 87

Figure 2.6 Multidimensional scaling of sex chromosomes in AddNeuroMed. ... 89

Figure 2.7 Outlyx plot for the identification of outliers in the AddNeuroMed cohort. 91

Figure 2.8 β-distribution following dasen normalisation in the AddNeuroMed cohort. 94

Figure 2.9 Median signal intensities in AddNeuroMed plotted by batch. 97

Figure 2.10 Median signal intensities in AddNeuroMed by 450K array chip. 98

Figure 2.11 Distribution of smoking scores across all 284 baseline samples in AddNeuroMed. 99

Figure 2.12 Correlation plot of principal components to potential confounders and variables of interest in AddNeuroMed. 101

Figure 2.13 Density plots of raw β-values in the ADNI cohort. 109

Figure 2.14 Raw median methylated and unmethylated signal intensities in ADNI samples. 110
Figure 2.15 Bisulfite conversion efficiency in ADNI samples. 111
Figure 2.16 Genetic distinctness of samples in the ADNI cohort. 112
Figure 2.17 Prediction of sex based on the X and Y chromosomes in ADNI. . 113
Figure 2.18 Outlyx plot for the identification of outliers in the ADNI cohort. 114
Figure 2.19 β-distribution following dasen normalisation in the ADNI cohort. . 115
Figure 2.20 Distribution of smoking scores across all 443 ADNI samples. 117
Figure 2.21 Median methylated signal intensities in ADNI by EPIC array chip. 118
Figure 2.22 Median unmethylated signal intensities in ADNI by EPIC array chip. 119
Figure 2.23 Correlation plot of principal components to potential confounders and variables of interest in ADNI. 121
Figure 2.24 Selection of the soft threshold for WGCNA module generation ... 125
Figure 2.25 Clusters (or 'modules') of highly co-methylated loci identified in the full dataset of 284 samples. 126
Figure 2.26 Modules identified in the subset of MCI-MCI and MCI-AD subjects. 127
Figure 2.27 Optimisation of the HOXB6 assay. 139
Figure 2.28 Gel electrophoresis quality control of PCR product of the HOXB6 amplicon. 140
Figure 3.1 Q-Q plots of p-values from the baseline diagnosis analysis and the analysis of conversion to AD. 153
Figure 3.2 The MOV10L1 DMR associated with baseline diagnostic status. . 167
Figure 3.3 The CBFA2T3 DMR associated with baseline diagnostic status. .. 168

Figure 3.4 The TPTEP2-CSNK1E DMR associated with baseline diagnostic status. 169
Figure 3.5 The HOXB6 DMR associated with AD relative to CTL. 170
Figure 3.6 The HOXB6 genomic region. 171
Figure 3.7 The CPT1B-CHKB DMR associated with future conversion to AD. 173
Figure 3.8 The TMEM184A DMR associated with future conversion to AD. .. 174
Figure 3.9 The KCNAB3 DMR associated with future conversion to AD. 175
Figure 3.10 The GABBR1 DMR associated with future conversion to AD. 176
Figure 3.11 The PRDM1 DMR associated with future conversion to AD. 177
Figure 3.12 The FLJ37453 DMR associated with future conversion to AD. ... 178
Figure 3.13 The OR56A3 and TRIM5 DMR associated with future conversion to AD. 179
Figure 3.14 The SMC1B and RIBC2 DMR associated with future conversion to AD. 180
Figure 3.15 The FIGN DMR associated with future conversion to AD. 181
Figure 3.16 Validation of the HOXB6 differentially methylated region (DMR). 184
Figure 3.17 Comparison of DNA methylation patterns quantified by the 450K array and pyrosequencing. 185
Figure 3.18 Correlations of DNA methylation patterns quantified by the 450K array and pyrosequencing. 186
Figure 3.19 Associations between gene expression and DNA methylation in CSNK1E. 190
Figure 3.20 Associations between gene expression and DNA methylation in HOXB6. 192
Figure 3.21 Correlation of DNA methylation and gene expression in CPT1B. 194

Figure 3.22 Correlations between module eigengenes and traits of interest.. 197
Figure 3.23 Modules showing positive correlations between MM and PS. 199
Figure 3.24 Pathways related to the brown MCI- and education years-associated module. 202
Figure 3.25 Pathways related to the yellow MCI- and MET-associated module. 203
Figure 3.26 Pathways related to the cyan APOE carrier status-associated module. 204
Figure 3.27 Significant KEGG terms related to the MCI to AD conversion-associated orange module. 206
Figure 4.1 The top DMP associated with sex in the AddNeuroMed cohort. ... 223
Figure 4.2 The cg02843237 DMP associated with age in the AddNeuroMed cohort. 225
Figure 4.3 The cg00664416 DMP associated with age in the AddNeuroMed cohort. 226
Figure 4.4 Q-Q plots of p-values of the main effects of sex (A) and age (B). . 228
Figure 4.5 The SHOX2 DMP showing a sex and age interaction in the AddNeuroMed cohort. 231
Figure 4.6 The top DMP showing an interaction between sex and diagnosis in the AddNeuroMed cohort. 233
Figure 4.7 The top DMP showing an interaction between age and diagnosis in the AddNeuroMed cohort. 235
Figure 4.8 Q-Q plots of sex, age, and diagnosis interaction effects. 237
Figure 4.9 Heatmap of module-trait relationships. 250
Figure 4.10 Module membership (MM) to probe significance (PS) correlations for modules associated with age in AddNeuroMed. 251

Figure 4.11 Module membership (MM) to probe significance (PS) correlations for the salmon module associated with a sex and diagnosis interaction in AddNeuroMed. 252
Figure 4.12 GO and KEGG terms enriched in the green module associated with age. 253
Figure 5.1 Q-Q plots of ANOVA p-values for the X and Y chromosomes in AddNeuroMed. 265
Figure 5.2 Q-Q plots of Xi ANOVA p-values in AddNeuroMed. 267
Figure 5.3 Q-Q plots of ANOVA p-values in ADNI. 270
Figure 5.4 Q-Q plots of Xi ANOVA p-values in ADNI. 271
Figure 5.5 Manhattan plots of ANOVA p-values for diagnosis in AddNeuroMed. 274
Figure 5.6 Manhattan plots of ANOVA p-values for diagnosis in the ADNI replication cohort. 287
Figure 5.7 Manhattan plot of the Xi ANOVA p-values in AddNeuroMed. 294
Figure 5.8 Manhattan plot of the Xi ANOVA p-values in the ADNI cohort. 301
Figure 5.9 The SRPK3 DMR on the Xi chromosome showed hypomethylation in AD in AddNeuroMed. 307
Figure 5.10 The MAGEE2 DMR on the Xi chromosome showed hypomethylation in AD and MCI in AddNeuroMed. 315
Figure 5.11 The MAGEE2 DMR on the Xi chromosome showed hypermethylation in AD and MCI in ADNI. 316
Figure 5.12 The ERAS DMR on the Xi chromosome showed hypomethylation in MCI relative to CTL in AddNeuroMed. 317
Figure 5.13 The ERAS DMR on the Xi chromosome showed hypomethylation in MCI and CTL relative to AD in ADNI. 318

Figure 5.14 The TAF1 DMR on the Xi chromosome showed hypermethylation in AD relative to MCI in AddNeuroMed. 319
Figure 5.15 The TAF1 DMR on the Xi chromosome showed hypomethylation in AD relative to CTL and MCI in ADNI. 320

TABLE OF TABLES

Table 1.1 Overview of studies examining global DNA modifications in AD. 60

Table 1.2 Overview of candidate gene studies of AD. 63

Table 1.3 Overview of epigenome-wide association studies of AD. 71

Table 2.1 AddNeuroMed cohort demographics. 76

Table 2.2 Principal components identified by PCA in the AddNeuroMed cohort. 100

Table 2.3 ADNI cohort demographics. 106

Table 2.4 Principal components identified by PCA in the ADNI cohort. 120

Table 2.5 CT conversion reagent. 130

Table 2.6 M-Wash Buffer reagents. 130

Table 2.7 HOXB6 primers for PCR and pyrosequencing. 134

Table 2.8 HOXB6 PCR reagents. 136

Table 2.9 Optimised PCR cycling conditions. 136

Table 2.10 PCR cycling conditions for HOXB6 assay optimisation. 139

Table 2.11 Volumes of PyroMark Gold Q24 reagents for the PyroMark Q24 Cartridge 144

Table 3.1 The top 10 most significant DMPs for diagnosis. 160

Table 3.2 The 10 most significant DMPs associated with MCI relative to CTL. 161

Table 3.3 The 10 most significant DMPs associated with AD relative to MCI. 162

Table 3.4 The 10 most significant DMPs associated with AD relative to CTL. 163

Table 3.5 The 10 most significant DMPs associated with future progression to AD. 164

Table 3.6 DMRs associated with baseline diagnosis in blood. 166

Table 3.7 DMRs associated with future conversion to AD in blood. 172
Table 3.8 Replication of AD-associated hypermethylation in HOXB6 by pyrosequencing. 183
Table 3.9 Expression of genes containing DMRs associated with baseline diagnosis. 189
Table 3.10 Correlation of methylation and expression for HOXB6 and TPTEP2-CSNK1E. 191
Table 3.11 Expression of genes containing DMRs associated with conversion from MCI to AD. 193
Table 3.12 Correlation of methylation and expression of DMRs associated with progression of MCI to AD. 193
Table 3.13 Correlations of MM and PS. 198
Table 4.1 Top 10 DMPs associated with sex in the AddNeuroMed cohort. 224
Table 4.2 Top 10 DMPs associated with age in the AddNeuroMed cohort. ... 227
Table 4.3 Top 10 DMPs associated with the sex and age interaction in the AddNeuroMed cohort. 232
Table 4.4 Top 10 DMPs associated with the sex and diagnosis interaction in the AddNeuroMed cohort. 234
Table 4.5 Top 10 DMPs associated with an interaction between age and diagnosis in the AddNeuroMed cohort. 236
Table 4.6 DMRs associated with sex in AddNeuroMed. 240
Table 4.7 DMRs associated with age in AddNeuroMed. 242
Table 4.8 DMRs associated with a sex by age interaction in AddNeuroMed. 243
Table 4.9 DMRs associated with a sex by diagnosis interaction in AddNeuroMed. 245

Table 4.10 DMRs associated with an age by diagnosis interaction in AddNeuroMed. 246
Table 5.1 The top 10 most significant DMPs for diagnosis on the female X chromosome in AddNeuroMed. 275
Table 5.2 The top 10 most significant DMPs for diagnosis on the male X and Y chromosomes in AddNeuroMed. 276
Table 5.3 The top 10 most significant DMPs in MCI compared to CTL in females in AddNeuroMed. 279
Table 5.4 The top 10 most significant DMPs in MCI compared to CTL in males in AddNeuroMed. 280
Table 5.5 The top 10 most significant DMPs in AD compared to CTL in females in AddNeuroMed. 281
Table 5.6 The top 10 most significant DMPs in AD compared to CTL in males in AddNeuroMed. 282
Table 5.7 The top 10 most significant DMPs in AD compared to MCI in females in AddNeuroMed. 283
Table 5.8 The top 10 most significant DMPs in AD compared to MCI in males in AddNeuroMed. 284
Table 5.9. AddNeuroMed DMPs replicated in the ADNI cohort for the female X chromosome in the MCI to CTL comparison. 288
Table 5.10 AddNeuroMed DMPs replicated in the ADNI cohort for the female X chromosome in the AD to CTL comparison. 289
Table 5.11 AddNeuroMed DMPs replicated in the ADNI cohort for the female X chromosome in the AD to MCI comparison. 290
Table 5.12 AddNeuroMed DMPs replicated in the ADNI cohort for the male X chromosome..... 291

Table 5.13 Top 10 Xi DMPs associated with CTL, MCI, and AD in AddNeuroMed. 295
Table 5.14 Top 10 Xi DMPs associated with AD in AddNeuroMed. 296
Table 5.15 Top 10 Xi DMPs associated with MCI in AddNeuroMed. 297
Table 5.16 Top 10 Xi DMPs associated with differences in AD relative to MCI in AddNeuroMed. 298
Table 5.17 Overlap of Xi DMPs in the AddNeuroMed and ADNI cohorts. 304
Table 5.18 Top 20 DMRs on the Xi chromosome associated with diagnostic status in AddNeuroMed. 306
Table 5.19 DMRs on the Xi chromosome associated with diagnostic status of AD in AddNeuroMed. 308
Table 5.20 DMRs on the Xi chromosome associated with diagnostic status of AD compared to MCI in AddNeuroMed. 309
Table 5.21 DMRs on the Xi chromosome associated with diagnostic status of MCI in AddNeuroMed. 311
Table 5.22 Overlapping genes containing Xi DMRs in the ADNI and AddNeuroMed overall group comparisons. 314

PUBLICATIONS ARISING FROM THIS THESIS

Chapter 1 (published manuscript presented in Appendix A):

Roubroeks, J. A. Y., Smith, R. G., van den Hove, D. L. A., & Lunnon, K. (2017). Epigenetics and DNA methylomic profiling in Alzheimer's disease and other neurodegenerative diseases. *Journal of Neurochemistry*, *143*(2), 158–170. https://doi.org/10.1111/jnc.14148

Chapter 3 (published manuscript presented in Appendix B):

Roubroeks, J. A. Y., Smith, A. R., Smith, R. G., Pishva, E., Ibrahim, Z., Sattlecker, M., Hannon, E. J., Kłoszewska, I., Mecocci, P., Soininen, H., Tsolaki, M., Vellas, B., Wahlund, L.-O., Aarsland, D., Proitsi, P., Hodges, A., Lovestone, S., Newhouse, S. J., Dobson, R. J. B., … Lunnon, K. (2020). An epigenome-wide association study of Alzheimer's disease blood highlights robust DNA hypermethylation in the HOXB6 gene. *Neurobiology of Aging*, *95*, 26–45. https://doi.org/https://doi.org/10.1016/j.neurobiolaging.2020.06.023

First author articles published during the PhD (published manuscript presented in Appendix C):

Lardenoije, R., **Roubroeks, J. A. Y**., Pishva, E., Leber, M., Wagner, H., Iatrou, A., Smith, A. R., Smith, R. G., Eijssen, L. M. T., Kleineidam, L., Kawalia, A., Hoffmann, P., Luck, T., Riedel-Heller, S., Jessen, F., Maier, W., Wagner, M., Hurlemann, R., Kenis, G., … Van Den Hove, D. L. A. (2019). Alzheimer's disease-associated (hydroxy)methylomic changes in the brain and blood. *Clinical Epigenetics*, *11*(1), 164. https://doi.org/10.1186/s13148-019-0755-5

Co-authored articles arising from the data generated in this thesis:

Nabais, M. F., Laws, S. M., Lin, T., Vallerga, C. L., Armstrong, N. J., Blair, I. P., Kwok, J. B., Mather, K. A., Mellick, G. D., Sachdev, P. S., Wallace, L., Henders, A. K., Zwamborn, R. A. J., Hop, P. J., Lunnon, K., Pishva, E., **Roubroeks, J. A. Y.**, Soininen, H., Tsolaki, M., … McRae, A. F. (2021). Meta-analysis of genome-wide DNA methylation identifies shared associations across neurodegenerative disorders. *Genome Biology*, *22*(1), 90. https://doi.org/10.1186/s13059-021-02275-5

Co-authored articles published during the PhD:

Smith, A. R., Smith, R. G., Pishva, E., Hannon, E., **Roubroeks, J. A. Y.**, Burrage, J., Troakes, C., Al-Sarraj, S., Sloan, C., Mill, J., Van Den Hove, D. L., & Lunnon, K. (2019). Parallel profiling of DNA methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in Alzheimer's disease. *Clinical Epigenetics*, *11*(1), 52. https://doi.org/10.1186/s13148-019-0636-y

Paes, D., Lardenoije, R., Carollo, R. M., **Roubroeks, J. A. Y.**, Schepers, M., Coleman, P., Mastroeni, D., Delvaux, E., Pishva, E., Lunnon, K., Vanmierlo, T., van den Hove, D., & Prickaerts, J. (2021). Increased isoform-specific phosphodiesterase 4D expression is associated with pathology and cognitive impairment in Alzheimer's disease. *Neurobiology of Aging*, *97*. https://doi.org/10.1016/j.neurobiolaging.2020.10.004

Smith, R., Pishva, E., Shireby, G., Smith, A., **Roubroeks, J.**, Hannon, E., Wheildon, G., Mastroeni, D., Gasparoni, G., Riemenschneider, M., Giese, A., Sharp, A., Schalkwyk, L., Haroutunian, V., Viechtbauer, W., van den Hove, D., Weedon, M., Brokaw, D., Francis, P., … Lunnon, K. (2020). A

meta-analysis of epigenome-wide association studies in Alzheimer's disease highlights novel differentially methylated loci across cortex. *BioRxiv*, 2020.02.28.957894. https://doi.org/10.1101/2020.02.28.957894

DECLARATIONS

The blood samples used to generate the DNA methylation data used throughout this thesis were obtained from the AddNeuroMed cohort. Professor Katie Lunnon performed DNA methylation profiling using Illumina Infinium HumanMethylation450K arrays in this cohort. For the pyrosequencing validation carried out on the AddNeuroMed samples in Chapter 3, Dr. Adam Smith performed the DNA bisulfite treatment. The gel electrophoresis and part of the pyrosequencing were carried out by myself under the guidance of Dr. Adam Smith, who also performed part of the pyrosequencing. The DNA methylation data used for replication analyses in Chapters 4 and 5 were obtained from the ADNI study.

All bioinformatic and statistical analyses performed throughout this thesis were carried out by myself, under the guidance of Professor Katie Lunnon, Professor Daniël van den Hove, Dr. Rebecca Smith, and Dr. Ehsan Pishva.

ABBREVIATIONS

Abbreviation	Term
27K array	Illumina Infinium HumanMethylation27K array
450K array	Illumina Infinium HumanMethylation450K array
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
AC	Auditory cortex
AD	Alzheimer's disease
ADM	Adrenomedullin
ADNI	Alzheimer’s Disease Neuroimaging Initiative
AH	Anterior hippocampus
AICD	APP intracellular domain
ALTLC	Anterior lateral temporal lobe cortex
aMCI	Amnestic mild cognitive impairment
ANOVA	Analysis of variance
APOE	Apolipoprotein E
APP	Amyloid precursor protein
ATP	adenosine triphosphate
Aβ	Amyloid-beta
BA	Brodmann area
C	Cytosine
CA1	Cornu ammonis area 1
CA3	Cornu ammonis area 3
CDR	Clinical Dementia Rating
CER	Cerebellum
CERAD	Consortium to Establish a Registry for Alzheimer’s Disease
CGIs	CpG Islands
CI	Confidence interval
CJD	Creutzfeldt-Jakob disease
CpG	Cytosine-phosphate-guanine
CSF	Cerebrospinal fluid
CTF	C-terminal fragment
CTL	Control
DG	Dentate gyrus
DHR	Differentially hyrdoxymethylated region
DLB	Dementia with lewy bodies
DMP	Differentially methylated position
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	Deoxyribonucleotide triphosphate
DSM	Diagnostic and Statistical Manual of Mental Disorders
EC	Entorhinal cortex

EOAD	Early onset Alzheimer's disease
EPIC array	Illumina Infinium MethylationEPIC array
EWAS	Epigenome-wide association studies
FAD	Familial Alzheimer's disease
FC	Frontal cortex
FDG	Fluorodeoxyglucose
FDR	False discovery rate
FTD	Frontotemporal dementia
GC/MS	Gas chromatography/mass spectrometry
GDS	Geriatric Depression Scale
GO	Gene ontology
GWAS	Genome-wide association studies
HD	Huntington’s disease
HIP	Hippocampus
HT-12 arrays	Illumina Human HT-12 v3 Expression BeadChip arrays
iDMRs	Imprinting differentially methylated regions
IL-1	Interleukin-1
IPL	Inferior parietal lobe
ITL	Inferior temporal lobe
KEGG	Kyoto Encyclopedia of Genes and Genomes
LEV	Left entorhinal cortex
LHV	Left hippocampal volume
lnRNAs	Long non-coding RNA
LOAD	Late onset Alzheimer's disease
MAP	Microtubule-associated protein
MCI	Mild cognitive impairment
MCI-AD	Mild cognitive impairment to Alzheimer's disease converters
MCI-MCI	Mild cognitive impairment non-converters
MDS	Multidimensional scaling
MEs	Module eigengenes
MET	Median entorhinal thickness
MFG	Middle frontal gyrus
miRNAs	Micro RNA
MM	Module membership
MMSE	Mini-Mental State Examination
MRI	Magnetic resonance imaging
MR-proADM	Mid-regional pro-adrenomedullin
MTG	Middle temporal gyrus
ncRNA	Non-coding RNA
NFT	Neurofibrillary tangles
NIA	National Institute on Aging
NIH	National Institutes of Health
NINCDS-ADRDA	National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association

NTC	No template control
OC	Occipital cortex
PAMP	Proadrenomedullin N-terminal 20
PC	Parietal cortex
PCA	Principal component analysis
PCD	Premature centromere division
PCR	Polymerase chain reaction
PD	Parkinson's disease
PET	Positron emission tomography
PFC	Prefrontal cortex
PHFs	Paired helical fragments
PHG	Parahippocampal gyrus
PIB	Pittsburgh compound-B
piRNA	Piwi-interacting RNA
PPi	Pyrophosphate
PS	Probe significance
PSEN1	Presenilin 1
PSEN2	Presenilin 2
QC	Quality control
Q-Q	Quantile-quantile
RAVLT	Rey Auditory Verbal Learning Test
REV	Right entorhinal cortex
RHV	Right hippocampal volume
RNA	Ribonucleic acid
SAM	S-adenosyl methionine
SD	Standard deviation
SE	Standard error
siRNAs	small interfering RNA
SNP	Single nucleotide polymorphism
SPL	Superior parietal lobe
STG	Superior temporal gyrus
STG	Superior temporal gyrus
TBE	Tris-borate EDTA
TC	Temporal cortex
TET	Ten-eleven translocation
TEV	Total entorhinal cortex
THV	Total hippocampal volume
TL	Temporal lobe
TN	Temporal neocortex
TSS	Transcription start site
Tukey's HSD	Tukey's Honest Significant Difference
VV	Ventricular volume
WBV	Whole brain volume
WGCNA	Weighted gene correlation network analysis
WMS	Wechsler memory scale
Xa	Active X chromosome

XCI	X chromosome inactivation
Xi	Inactivated X chromosome

CHAPTER 1. INTRODUCTION

This chapter introduces the field of epigenetics, namely DNA methylation, in Alzheimer's disease (AD). It describes the research performed in this field to date, which has led to the research projects described in Chapters 3, 4, and 5. This chapter is based on a published review (Roubroeks et al., 2017), which can be found in Appendix A.

1.1. Alzheimer's disease

Dementia, of which AD is the most common subtype, is a term for complex cognitive decline that arises from progressive cortical and subcortical dysfunction. A number of neurodegenerative diseases are characterised by dementia; in addition to AD, dementia is observed in frontotemporal dementia (FTD), dementia with Lewy bodies (DLB), Creutzfeldt-Jakob disease (CJD), Huntington's disease (HD), and Parkinson's disease (PD). Each of these diseases characterised by dementia are distinct in their pathological hallmarks. For example, AD is characterised by the aggregation of extracellular amyloid-beta (Aβ) plaques and intracellular neurofibrillary tangles (NFTs) of hyperphosphorylated tau protein (Yates & McLoughlin, 2008).

1.1.1. Types of AD

AD is often classified into two subtypes, termed early onset (EOAD) and late onset AD (LOAD). Alternatively, AD is often classed as familial (FAD) or sporadic AD. Although EOAD and FAD are often used interchangeably, they are not synonymous. It is most often the case that individuals with the familial form of AD have a much earlier onset of the disease than those with the sporadic type of AD. However, not all cases of EOAD can be classified as FAD (Joshi et al., 2012;

Mendez, 2012). As the name suggests, the division of EOAD and LOAD is largely based on the age at onset of the first symptoms of the disorder. However, the age cut-off is often an arbitrary number, and most studies refer to those older than 65 as LOAD, and younger than 65 as EOAD. Age is not the only distinguishing mark of LOAD and EOAD, as a number of studies have shown that these subtypes also differ in their topographical patterns of grey matter atrophy (Frisoni et al., 2007) and presentation of clinical symptoms (Koedam et al., 2010).

The sporadic form of AD is the most common subtype, accounting for ~90-95% of AD cases (Goedert & Spillantini, 2006; Reitz & Mayeux, 2014), and its cause is multifaceted. This thesis focusses on individuals with LOAD, and its causes and symptoms will be further discussed in subsequent sections of this introduction. Unless specified otherwise, the term AD from hereon in will refer to this specific subtype.

1.1.2. Epidemiology and burden of disease

It is suggested that AD has the largest burden of disease amongst all neurodegenerative disorders, with an expected rise in dementia cases to over 131 million cases worldwide by 2050 (Prince et al., 2015). The prevalence of AD has been estimated to be lowest in Sub-Saharan Africa (3.1% of the population over 60 years old) and highest in Latin America (7.6% of the population over 60 years old). In Western Europe, the prevalence has been estimated to be 6.9% of the over 60 population, with women of every age group showing nearly twice the estimated prevalence of men (Prince et al., 2015). Similar results were found in a meta-analysis carried out over ten years from 2002-2012 by Takizawa *et al.*, which analysed data from six European countries and the USA (Takizawa et al.,

2014). The incidence of AD per 1,000 person years was found to vary greatly by country and by study, ranging from 0.04 in the UK for men and women between the ages of 45-65, to 16.8 in the USA in men and women aged 65 years and over. However, in each study the incidence in women was higher than the incidence in men, and on average the incidence in women was twice the incidence in men of a comparable age. The prevalence of AD across all studies examined varied from 3-7%. Both incidence and prevalence were shown to increase with age, and AD and mortality due to AD were both found to be more common in females compared to males, regardless of age.

1.1.2.1. Social and economic burden

Worldwide, the total economic costs of dementia were estimated to be $604 billion (£468 billion) in 2010, which amounts to ~$17,000 (~£13,155) per individual for both direct medical costs (e.g. hospital care, medication), and social care costs (e.g. home care, transport, nursing home care; Anders Wimo et al., 2013). In Europe alone, the total cost of illness (direct medical cost and social care cost) was estimated to be €177.2 billion (£100 billion) yearly. Of this total, the EU27 accounts for the majority of costs, of which 56% is spent on social care costs (A. Wimo et al., 2011). In addition to these care costs, it is estimated that the total cost of developing a new disease-modifying drug is $5.7 billion (~£4.3 billion) on average (Scott et al., 2014).

Even more important than the monetary costs of the disease are the personal and social costs, as the devastating effects of AD greatly impact the quality of life of both patients and their caregivers (Shin et al., 2005).

1.1.3. Hallmarks of AD

The pathological hallmarks of the disease include aggregations of extracellular Aβ protein, NFTs of hyperphosphorylated tau protein, and progressive neuronal cell death (Yates & McLoughlin, 2008). These changes in the brain are thought to occur years before a clinical diagnosis can be made (Amieva et al., 2008; Jack et al., 2010). In the early stages of disease, these hallmarks are most prevalent in the brainstem, hippocampal and entorhinal areas of the brain, and spread to the temporal, parietal and frontal cortex as the disease progresses (Dubois et al., 2010; Iatrou et al., 2017). In later stages, brain autopsies have shown that there is a dramatic shrinkage of most neocortical areas and a substantial expansion of the ventricles (Dallaire-Théroux et al., 2017). There is also considerable loss of subcortical structures, including the majority of the cells of the basal nucleus, the locus coeruleus and dorsal raphe. Spared regions of the brain include the substantia nigra. Though initially believed to originate in the entorhinal cortex (Hyman et al., 1984), it is now also hypothesised that the brainstem is the origin of AD pathology (Braak et al., 2011; Braak & Del Tredici, 2011; Grudzien et al., 2007; Simic et al., 2009).

The clinical symptoms that accompany these spatiotemporal pathological changes generally can be used to divide AD into three stages of progression, i.e. an early stage, a moderate stage, and a late stage of severe AD. Commonly, some of the first symptoms of the disease are short-term memory issues and occasional aphasia which interfere significantly with daily life (American Psychiatric Association, 2014). Many individuals will develop neuropsychiatric symptoms such as apathy, delusions, hallucinations, and aggression (Lyketsos et al., 2011), and in advanced stages, sleep and circadian rhythm issues

commonly occur (Vitiello & Borson, 2001; Volicer et al., 2001). In the final stages of the disease there is an increased risk of contracting other diseases, and the most common cause of death for individuals with AD is respiratory system disease, specifically bronchopneumonia (Brunnström & Englund, 2009).

1.1.4. Aetiology of AD

Though the cause of sporadic AD remains elusive, there are several hypotheses as to the possible causes. There is a clear genetic basis for the development of FAD, and therefore a number of studies have explored the role of genetics in the aetiology of sporadic AD. The most often studied theories of sporadic AD fall into three classes: the amyloid cascade hypothesis, the tau aggregation hypothesis, and the inflammation hypothesis. The cause for FAD and each of these three hypotheses for sporadic AD will be described in further detail below.

1.1.4.1. Causal mechanisms of FAD

FAD can be caused by a single mutation in one of three genes: the presenilin 1 (*PSEN1*) gene on chromosome 14, the presenilin 2 (*PSEN2*) gene on chromosome 1, and the amyloid precursor protein (*APP*) gene on chromosome 21. At the time of writing, over 400 mutations have been identified in these three genes, of which most pathogenic mutations have been found in the *PSEN1* gene (www.alzforum.org/mutations). Of note, individuals with Down's syndrome with trisomy 21 also often exhibit AD pathology by the age of 50 (Mann & Esiri, 1989). *APP* can be translated into three different isoforms, of which APP695 is highly expressed in neurons (Yoshikai et al., 1990). APP can be cleaved into fragments via the non-amyloidogenic (non-pathological) pathway or the amyloidogenic (pathogenic) pathway (Figure 1.1). In the non-amyloidogenic pathway, APP is

cleaved by α-secretase within the Aβ domain into sAPPα and an α-C-terminal fragment (CTF), which is further cleaved by γ-secretase, resulting in the fragments p3 and an *APP* intracellular domain (AICD; Müller et al., 2008). In the amyloidogenic pathway, *APP* is first cleaved by β-secretase, generating sAPPβ and β-CTF, followed by γ-secretase cleavage into an Aβ peptide and an AICD (G. F. Chen et al., 2017). *PSEN1* and *PSEN2* encode proteins that are important for the catalytic region of γ-secretase. Isoforms of various amino acid lengths of the Aβ peptide exist, though the most abundant isoforms are the 38-, 40-, and 42-amino acid forms of Aβ (Aβ-38, Aβ-40, and Aβ-42, respectively; Kummer & Heneka, 2014). In healthy controls, these have been found in the cerebrospinal fluid (CSF) at proportions of around 53% for Aβ-40, 17% for Aβ-38, and 11% for Aβ-42 (Bibl et al., 2012). Via the amyloidogenic pathway, mutations in the *APP* gene as well as the *PSEN1* and *PSEN2* genes generally lead to an overproduction of Aβ-42, which is more likely to aggregate and constitutes the primary component of Aβ plaques, thereby causing AD.

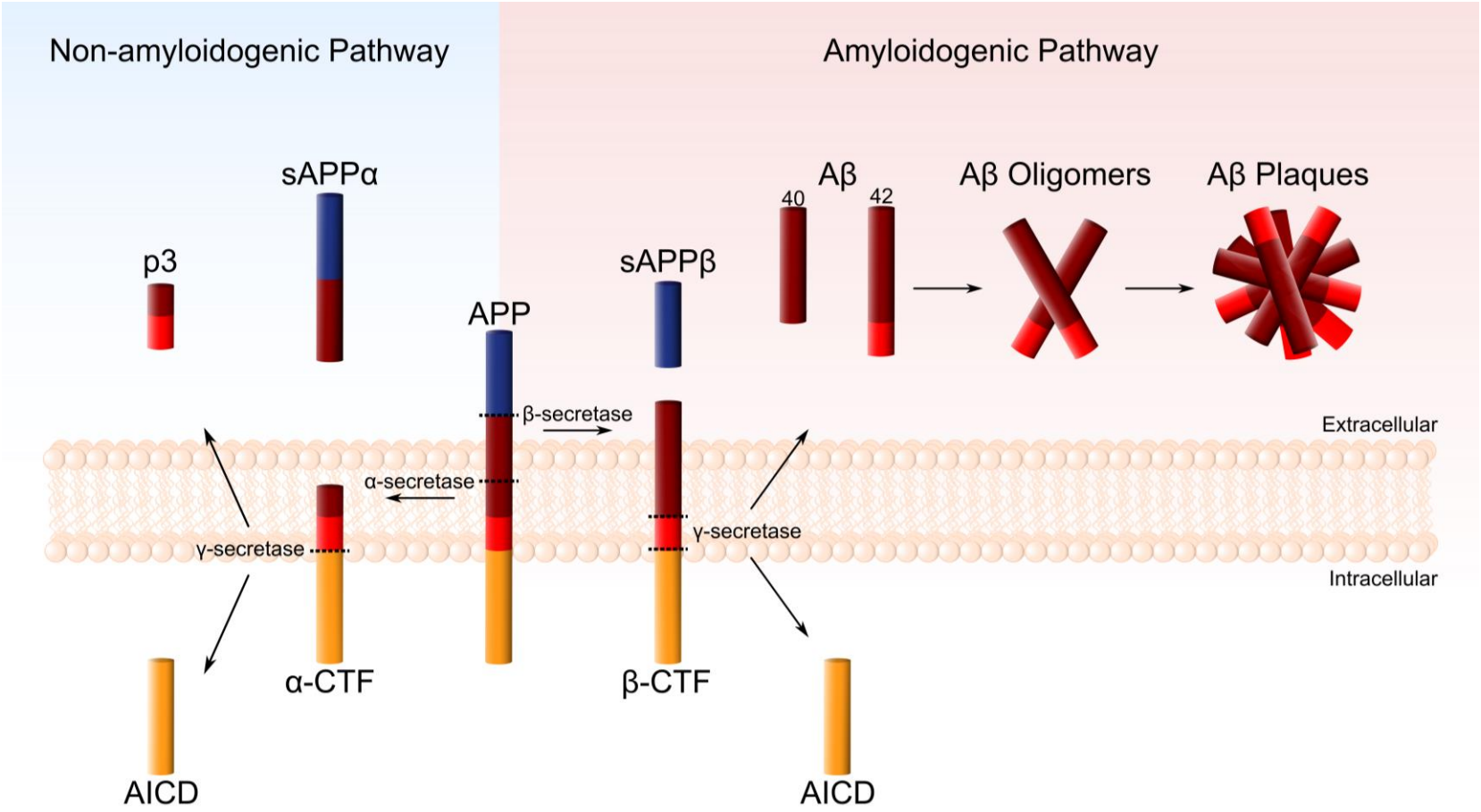

Figure 1.1 The amyloidogenic and non-amyloidogenic pathways of APP cleavage.

In the non-amyloidogenic pathway, α-secretase and γ-secretase cleave APP leading to the production of sAPPα, p3, and AICD. In the amyloidogenic pathway, β-secretase and γ-secretase cleave APP leading to the production of sAPPβ, Aβ and AICD. Mutations in the genes APP, PSEN1, and PSEN2 can lead to increased processing via the amyloidogenic pathway, resulting in increased production of Aβ oligomers and plaques.

1.1.4.2. The amyloid cascade hypothesis

As FAD has a similar phenotype to sporadic AD, with the exception of the age of onset, identification of the mutations causal to FAD gave support to the amyloid cascade hypothesis. This hypothesis focuses on Aβ, which, in a healthy state, has been associated with neurogenesis and memory enhancement (Morley et al., 2019; Morley & Farr, 2014; Rajasekhar et al., 2015), though the focus of Aβ research is predominantly placed on exploring the pathological role of Aβ rather than its physiological function. The amyloid cascade hypothesis states that the accumulation of Aβ peptides serves as a trigger for a cascade that leads to neuritic injury, neuronal dysfunction, the aggregation of tau protein, and eventual cell death (Beyreuther & Masters, 1991; J. A. Hardy & Higgins, 1992; J. Hardy & Allsop, 1991; Selkoe, 1991). While there is much support for the amyloid cascade hypothesis, several mouse models with Aβ depositions have been shown not to display cytotoxicity or develop tau tangles, nor display cognitive impairment (Bryan et al., 2009; Kim et al., 2013). Additionally, positron emission tomography (PET) studies have found amyloid deposits in healthy non-demented individuals (Chételat et al., 2013; Klunk et al., 2004). This has led to the idea that Aβ may not be the primary and/or sole toxic agent responsible for AD, but another product of the amyloidogenic pathway such as β-CTF or AICD. Indeed, it has been found that β-CTF can cause memory impairment and synaptic deficits (Tamayev et al., 2012), and a clinical trial of a γ-secretase inhibitor which cleaves CTF under normal circumstances in both *APP* processing pathways, found significantly worsened symptoms in AD patients (Doody et al., 2013). As for AICD, overexpression can lead to pathological hallmarks of AD such as tau aggregation, neurodegeneration and cognitive impairment in mouse models (Ghosal et al., 2009). Additionally, due to the spatial properties of the amyloidogenic and non-

amyloidogenic pathways, only AICD produced through the amyloidogenic pathway can be transported to the nucleus (Goodger et al., 2009). There, together with the proteins Fe65 and Tip60 it forms an AFT complex, which was found to influence expression of *APP* as well as β-secretase (von Rotz et al., 2004). In which manner and to which extent these substances may contribute to the aetiology of AD is still to be determined.

1.1.4.3. The tau hypothesis

Tau, a microtubule-associated protein (MAP), is expressed by the *MAPT* gene located on chromosome 17. As the name suggests, the tau hypothesis states that tau, and specifically hyperphosphorylated tau, is the principal cause of AD. In a healthy state, tau is important for the maintenance of neuronal structure through the binding of microtubules. Phosphorylation of tau is important for microtubule dynamics, and is known to be elevated during development (Brion et al., 1994). In disease, tau becomes hyperphosphorylated, causing decreased binding to microtubules and their destabilisation (Drewes et al., 1995; Ksiezak-Reding et al., 2003; Sengupta et al., 1998). Detached tau self-aggregates to form paired helical fragments (PHFs) leading to the formation of NFTs (Figure 1.2; Iqbal, Liu, Gong, & Grundke-Iqbal, 2010). Hyperphosphorylated tau can further cause disruption to microtubules and inhibit their assembly (B. Li et al., 2007).

The tau hypothesis is supported by (i) the fact that the accumulation of NFTs in the brain strongly corresponds to cognitive impairment seen in AD, unlike Aβ plaques (Arriagada et al., 1992; Ghoshal et al., 2002), (ii) hyperphosphorylated tau in CSF has been found to correlate with the extent of cognitive impairment (Maccioni et al., 2006), and (iii) tau aggregation has been reported to occur before

Aβ aggregation (Braak & Del Tredici, 2014). Additionally, a study by Götz *et al.* (2001) found that the injection of Aβ-42 in a mouse model expressing NFTs accelerated the formation of NFTs. While there is much support for the role tau may play in the development of AD, the initial triggering mechanism of the pathogenic tau aggregation pathway remains unclear.

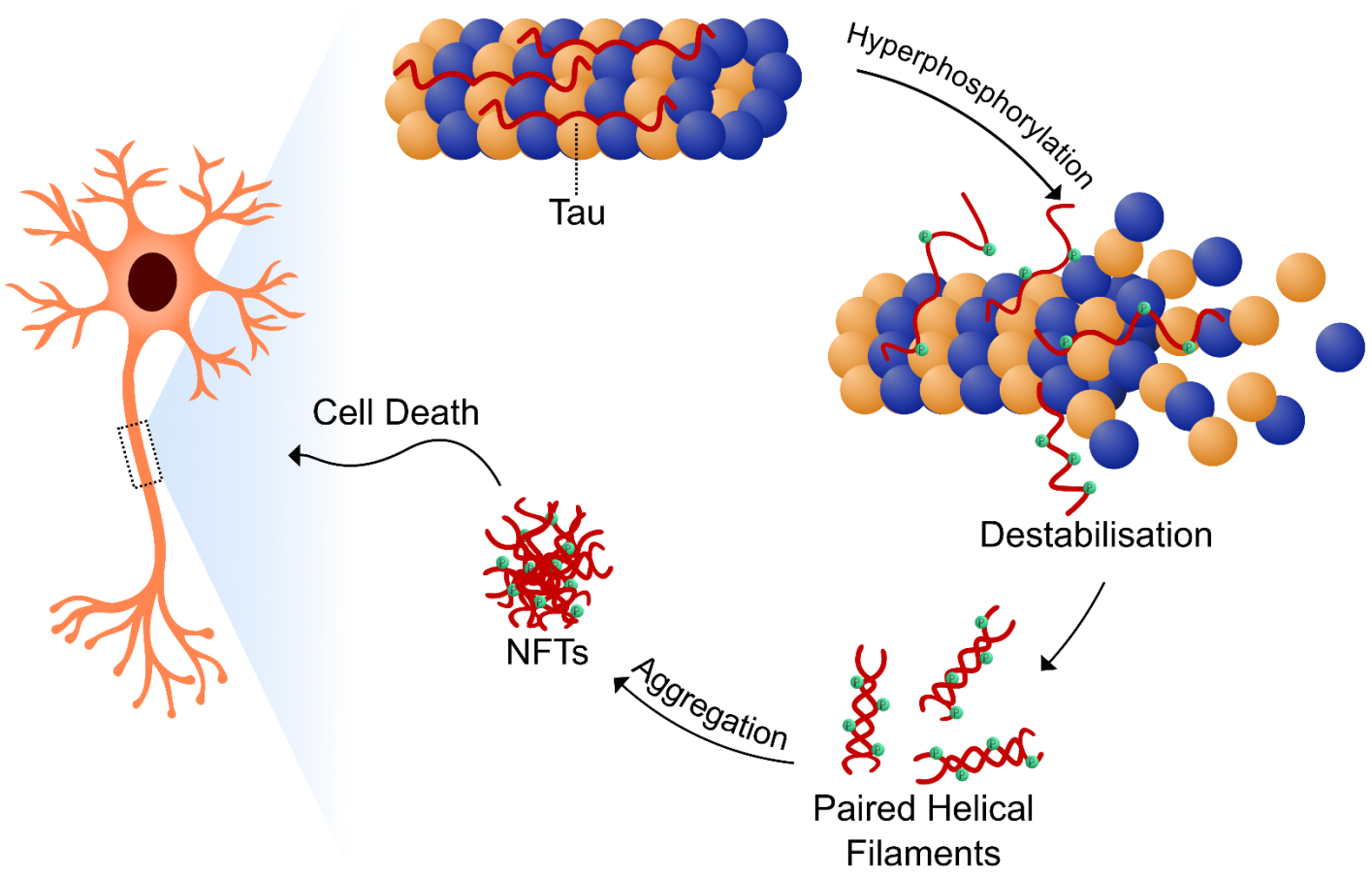

Figure 1.2 The formation of neurofibrillary tangles (NFTs) in AD.

Hyperphosphorylation of tau protein leads to destabilisation of microtubules and the formation of paired helical filaments (PHFs), which subsequently aggregate into the neurotoxic neurofibrillary tangles (NFTs) which are a hallmark of AD.

1.1.4.4. The inflammation hypothesis

Multiple studies have observed evidence of a sustained inflammatory response in AD, and activation of microglia, the brain resident macrophages, is thought to play a major role starting in the early stages of the disease (Arends et al., 2000; Cagnin et al., 2001). It is hypothesised that the existence of Aβ plaques is the main catalyst for microglial activation. In the early stages of AD pathogenesis, microglia are able to phagocytose and clear Aβ (Lee & Landreth, 2010). However, in more advanced stages the microglia may be unable to process Aβ, leading to a loop of increased Aβ accumulation and continuous activation of microglia and pro-inflammatory signalling, which in turn causes neuronal damage (Hickman et al., 2008; Krabbe et al., 2013; Meda et al., 1995; Michelucci et al., 2009). While microglia are increasingly unable to clear Aβ, their ability to produce pro-inflammatory cytokines remains constant, exacerbating the neurodegenerative effects. This is supported by studies of ageing, which have found that microglia in aged brains reflect a primed state in which they are hypersensitive to inflammatory signals (Norden & Godbout, 2013). Microglia in this primed state can display a magnified inflammatory response in the presence of NFTs and Aβ plaques (Perry & Holmes, 2014). Furthermore, there is some evidence that increased quantities of the cytokines interleukin-1 (IL-1) and IL-1β may lead to increased Aβ production and tau phosphorylation, respectively. As such, the inflammation hypothesis provides a potential link between the accumulation of Aβ plaques and NFTs.

1.1.5. Risk factors of AD

Though there appears to be no single genetic cause of sporadic AD, quantitative models examining twins have estimated the heritability of AD to lie between 58%

and 79%. In recent years, this has led to many genome-wide association studies (GWAS), through which a number of genetic risk factors have been identified. Of these factors, the largest risk is associated with the apolipoprotein E (*APOE*) ε4 allele, where the presence of two ε4 alleles is associated with an increased risk of developing AD (Corder et al., 1993; Harold et al., 2009), for both EOAD and LOAD (Chartier-Hariln et al., 1994). Conversely, the presence of an *APOE* ε2 allele is associated with decreased risk, while the most common ε3 allele does not increase nor decrease risk (Chartier-Hariln et al., 1994; Corder et al., 1994). While two copies of the *APOE* ε4 allele are neither sufficient nor necessary to cause AD, their presence may reduce the age of onset of AD (Blacker et al., 1997; Corder et al., 1993; Meyer et al., 1998). In addition to *APOE*, several other genetic risk variants have been identified and validated in meta-analyses of GWAS (Harold et al., 2009; Hollingworth et al., 2011; Jansen et al., 2019; Jun et al., 2010; Kunkle et al., 2019; Lambert et al., 2013; Naj et al., 2011). The most recent meta-analysis of AD risk genes confirmed the risk of several previously identified genes including *BIN1, INPP5D, TREM2, PTK2B, EPHA1, SORL1, PICALM, ABCA7, and CASS4*, and identified new risk loci near *NCK2, SPRED2, TSPAN14,* and *CCDC6* (Schwartzentruber et al., 2021). Furthermore, genome and exome analysis of a large cohort of AD patients and controls has identified a significantly increased risk of AD associated with rare variants within the *TREM2* gene (Guerreiro et al., 2013). In addition to genetic risk factors, many studies have focused on non-genetic factors, the largest influencers being age and sex (J. H. Chen et al., 2009; Mazure & Swendsen, 2016), which will be discussed in detail in Chapters 4 and 5. Other risk factors have also been examined, such as pathogenic microbes (Hill et al., 2014), obesity (Alford et al., 2018), vascular

factors (Larsson & Markus, 2018), stress (Caruso et al., 2018), and other psychosocial factors (Burke et al., 2018).

1.1.6. Diagnosis of AD

Currently, a definitive diagnosis of AD can only be made through post-mortem pathological assessment and confirmation of the presence of Aβ plaques according to the Consortium to Establish a Registry for Alzheimer's disease (CERAD) score, and the presence of NFTs according to Braak staging (Hyman et al., 2012). Braak staging is a classification method for the spatiotemporal progression of tau accumulation in the brain, and identifies six stages: the preclinical stages I-II where tau accumulation is present in brainstem and transentorhinal regions, stages III-IV indicate accumulation in limbic regions, and stages V-VI are the late-disease stages of AD with neocortical tau accumulation (Braak & Braak, 1995). In a clinical setting, a diagnosis can be made using different guidelines, such as the diagnostic and statistical manual of mental disorders (DSM-5; American Psychiatric Association, 2013), as well as the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria (McKhann et al., 2011). The latter set of criteria describes three stages of AD; a preclinical stage where pathological changes have emerged, but clinical symptoms are not evident; a middle stage referring to mild cognitive impairment (MCI; see section 1.1.6.1.); and a final stage of dementia. These criteria have also taken into consideration the use of biomarkers for the detection of a preclinical phase of AD, though only recommends these to be used in research and not in a clinical setting. This is recommended as biomarkers need to be further validated and standardised in their usage, and not all diagnostic centres

have access to all biomarkers. However, it is acknowledged that some fluid and imaging biomarkers may supplement clinical diagnosis in certain cases in research settings. An overview of the most commonly used biomarkers is described in section 1.1.6.2.

1.1.6.1. Mild Cognitive Impairment as a precursor of AD

MCI is defined as a condition in which an extent of cognitive impairment is present that surpasses normal age-associated cognitive decline (Albert et al., 2011). Individuals with MCI generally do not experience extensive impairments in their daily life and show no signs of dementia, unlike those who suffer from AD. However, like AD, MCI currently cannot be diagnosed using a definitive test, and is assessed clinically. This is often done using the Petersen criteria for amnestic MCI (aMCI; Petersen, 2004; Petersen et al., 1999), or the NINCDS-ADRDA criteria for MCI due to AD (Albert et al., 2011), which are nearly identical. A clinical diagnosis of MCI due to AD can be assisted through biomarker measurement, particularly by Aβ deposition and neuronal injury measures (section 1.1.6.2.). Individuals with MCI can further be distinguished based on the type of cognitive decline. Generally, when memory is affected, this is classified as aMCI, and when other cognitive domains are affected this can be classified as non-amnestic MCI.

While some individuals with MCI may remain stable over time, it is known that some can develop a neurodegenerative disorder other than AD, or progress to AD. This is particularly true for individuals with aMCI (Jicha et al., 2006). In these cases AD has not fully manifested yet, but the clinical symptoms reflect underlying pathological changes that occur in early AD (Jack et al., 2010; see Figure 1.3).

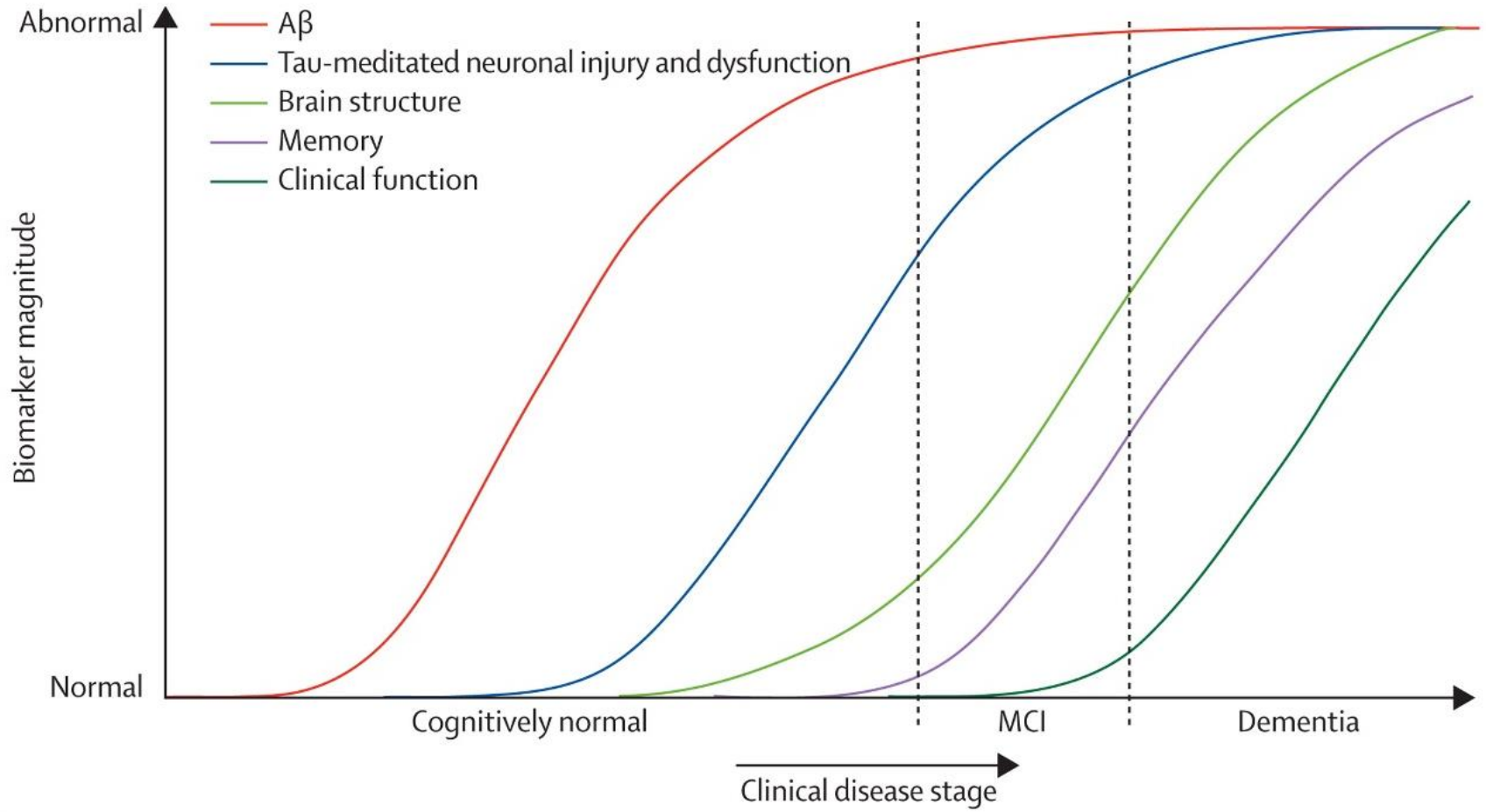

Figure 1.3 Hypothetical model of biomarker presence prior to development of dementia.

Reprinted from Jack et al. (2010). The model shows a representation of the onset of pathological hallmarks of AD which may be assessed with neuroimaging or CSF measurements, prior to the emergence of clinical symptoms.

1.1.6.2. Biomarkers of preclinical AD

The most commonly studied biomarkers of AD can be classified into three groups, based on the type of pathophysiology measured: *A,* for biomarkers targeting Aβ; *T,* for biomarkers targeting tau; *N,* for biomarkers targeting neurodegeneration. This model, proposed by Jack *et al.* (2016), suggests a diagnosis can be assisted by creating a score in which an individual receives a positive or negative indication for A, T, and N biomarker presence. It should be noted that not all (future) biomarkers may fit into these categories. For example, a DNA methylation biomarker which is suggested in Chapter 3, would not classify into A, T, or N. However, this classification system still provides a useful system for biomarker-assisted diagnosis of AD using the most commonly used biomarkers, as described in Jack *et al.* (2018).

Class A consists of CSF measures of Aβ, which detect the presence of Aβ-42 or calculate a ratio of Aβ-42 to Aβ-40 (Fagan et al., 2007; Mattsson et al., 2009). Additionally, a ratio of Aβ-42 to total tau may also be used (Visser et al., 2009). Levels of amyloid detected through PET imaging, using the Pittsburgh Compound-B (PIB) for example (Klunk et al., 2004; Villain et al., 2012), are also included in this category. Class T consists of measures of tau through PET imaging (Maass et al., 2017), or phosphorylated tau detection in CSF (Hampel et al., 2010). Though perhaps counterintuitive, total tau levels detected in CSF are included in class N, not class T. While total tau may also be classified in the T class (Dubois et al., 2014), it is included in the N category here due to its similarity to other biomarkers in class N in relation to cognitive symptoms of AD. Class N includes general measures of neurodegeneration or neuronal injury, such as PET

imaging of fluorodeoxyglucose (FDG), and atrophy detected by magnetic resonance imaging (MRI).

Though these biomarkers have been studied extensively, variable ranges of specificity and sensitivity have been reported particularly when it comes to detection of early AD (Bloudek et al., 2011). Furthermore, accuracy has been reported to vary between diagnostic centres for CSF measurements (Mattsson et al., 2009) in addition to MRI and PET techniques not being widely available to all diagnostic centres. An ideal target for new biomarker development would therefore be blood or saliva, which are both easily accessible, and may be obtained more easily in an elderly population of MCI and AD individuals. A study by A. K. Smith *et al*. (2015) found that DNA methylation profiles obtained from saliva samples were more similar to brain tissue DNA methylation than DNA methylation in blood. This appeared to be dependent on the proportion of epithelial cells collected within a sample, which can be highly variable. However, as DNA methylation profiles obtained from blood and saliva samples correlate well, and blood samples are more readily available in large cohorts, the primary focus of this thesis is the assessment of blood DNA methylation. This thesis aims to explore the epigenetic mark of DNA methylation in the blood of individuals with AD and MCI, which could have implication for novel biomarkers in the future.

1.2. Epigenetic Mechanisms

The term epigenetics was first coined by Waddington in 1942 (Waddington, 2012), and it is now used to describe the study of heritable changes in the phenotype without any changes in the genotype. In other words, epigenetics refers to mitotically and meiotically heritable changes in gene expression without alterations in the underlying DNA sequence. This allows for alterations in gene expression in response to environmental variation, for example stress, diet or exposure to environmental chemicals. The essence of epigenetics as a regulator of gene expression may also provide a mechanism for the missing heritability found in some diseases such as AD. Additionally, the transient nature of epigenetics gives rise to the idea that pathological changes might be reversible, making epigenetic processes interesting targets for drug discovery studies. Of the various known epigenetic mechanisms, DNA modifications of the cytosine base are perhaps the most investigated and are the focus of this thesis.

1.2.1. DNA modifications

Currently, the cytosine base is known to exist in five states: as unmodified cytosine (C), 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC). The most common state of cytosine in the brain, after the unmodified state, is 5mC, which is mainly located in cytosine-phosphate-guanine (CpG) dinucleotides. Genome-wide distribution studies have found 5mC in a large majority of CpG dinucleotides, with the exception of high density CpG areas, known as CpG islands (CGIs), which largely remain unmethylated (Y. Li et al., 2010; Lister et al., 2009; Meissner et al., 2008). Unmodified DNA can be methylated through DNA methyltransferases (DNMTs)

that transfer a methyl group from S-adenosyl methionine (SAM) to the 5 position of cytosine (Figure 1.4). DNMT3A and DNMT3B are *de novo* DNTMs, acting on unmodified cytosine, while DNMT1 is thought to be largely responsible for the maintenance of methylation through cell divisions (Z. X. Chen & Riggs, 2011). DNA methylation plays a crucial role in several key processes, such as genomic imprinting and X-chromosome inactivation (Bonasio et al., 2010). Although it was originally believed to be solely associated with transcriptional repression, recent evidence suggests a differential effect on gene expression depending on location in or around the gene (Ziller et al., 2013), with intragenic DNA methylation shown to modulate alternative splicing (Maunakea et al., 2013) and gene body methylation associated with increased expression (Varley et al., 2013). Demethylation, the transition from methylation to unmodified cytosine, can occur either passively, or actively (Delatte & Fuks, 2013). Passive demethylation takes place over several DNA replication cycles, during which DNMT1 is inhibited or absent from the process, leaving the newly synthesised DNA strand unmethylated. Active demethylation refers to the enzymatic conversion of 5mC to cytosine, which is initiated by oxidation of 5mC into 5hmC by a family of ten-eleven translocation (TET) proteins (Delatte & Fuks, 2013; J. U. Guo et al., 2011; Tahiliani et al., 2009). Initially believed to be a transient step in the demethylation process, in recent years, 5hmC has been shown to be a potentially independent and functional epigenetic marker (Sun et al., 2014; van den Hove et al., 2014). It has been found to be present in most tissues and cell types, but is particularly abundant in the brain (Wen & Tang, 2014), where it shows a genomic region-specific distribution (Lunnon et al., 2016). Interestingly, the presence of 5hmC appears to be negligible in blood (Nestor et al., 2012). 5hmC can be further oxidised into 5fC by TET enzymes, which in turn can be oxidised into 5caC in the

demethylation process (He et al., 2011; Ito et al., 2011). Finally, the demethylation process can be completed through decarboxylation of 5caC into unmodified cytosine (Ito et al., 2011; S. C. Wu & Zhang, 2010). The role of 5fC and 5caC in human development and their potential role in diseases such as AD remains unclear (for a review see Breiling & Lyko, 2015), and these are beyond the scope of this thesis.

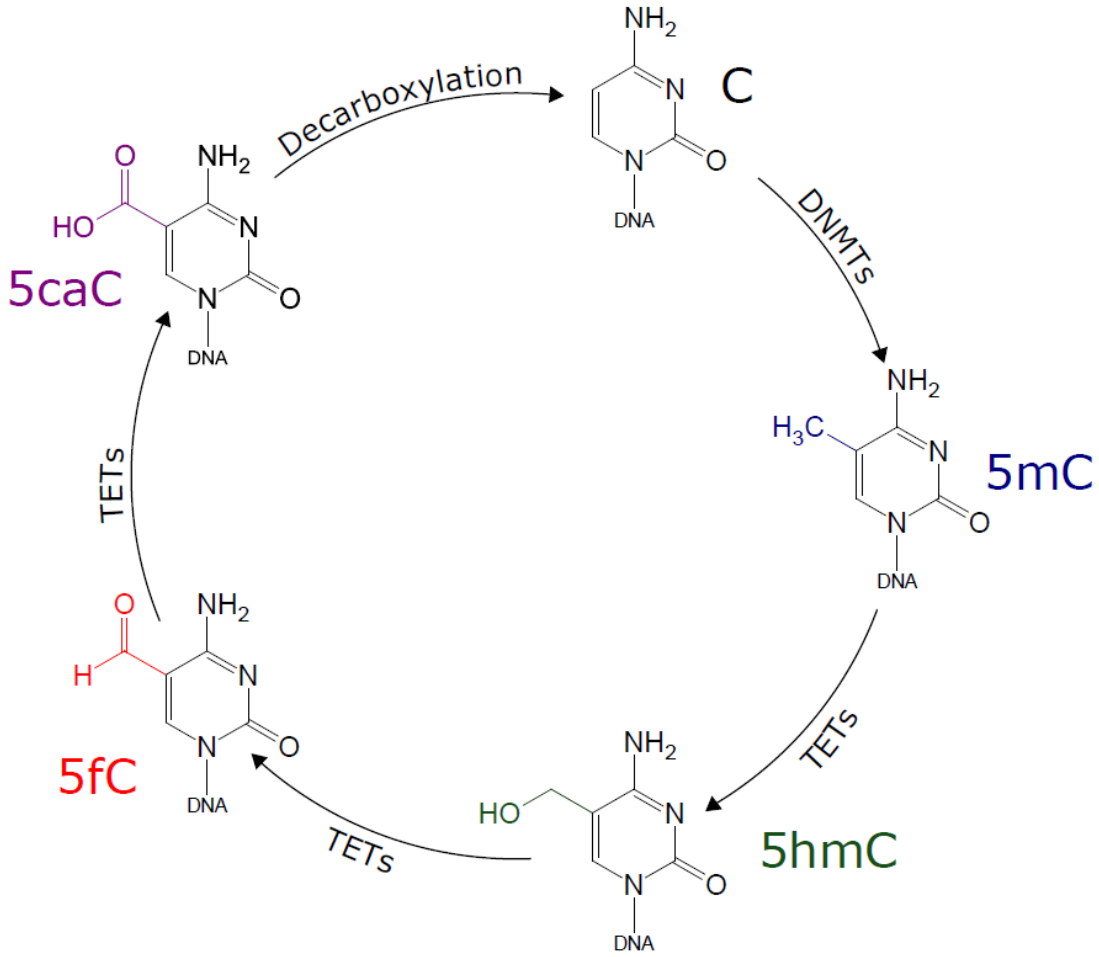

Figure 1.4 Cycle of cytosine modifications.

Unmethylated cytosine can be converted into 5mC through DNMTs. Active demethylation of 5hmC, 5fC, and 5caC occurs via TET enzymes and decarboxylation.

1.2.2. Other epigenetic mechanisms

In addition to cytosine modifications, there are several other epigenetic mechanisms that play a role in the regulation of gene expression, such as histone modifications and non-coding RNA regulation. As the focus of this PhD thesis is on DNA methylation, other regulating mechanisms of gene expression are outside of the scope of this review and are only briefly highlighted here. However, it is acknowledged that these factors may also contribute to the development and progression of AD.

Histones, which form octamers around which DNA is wound, can be modified post-translationally. The octamer structure with the DNA wound around it, the nucleosome, is necessary to fit entire DNA strands into the small space of a nucleus, and its conformation affects the overall chromatin structure as well as the transcriptional capability of the DNA. The conformation of the histone can be changed through several methods, the most well studied of these methods being histone acetylation. Bannister and Kouzarides (2011) and Sadakierska-Chudy and Filip (2014) provide excellent reviews on chromatin regulation and histone modifications.

Non-coding RNAs (ncRNAs) exist in a multitude of classes, including long non-coding RNAs (lnRNAs), micro RNAs (miRNAs) and small interfering RNAs (siRNAs), for example (Mattick, 2011). Just as there are a multitude of classes of ncRNAs, so there are numerous ways in which these can regulate gene expression, such as transcriptional- and post-transcriptional regulation, and chromatin remodelling (for reviews of these mechanisms see Kaikkonen et al., 2011 and Mattick, 2011).

1.3. Epigenetics in Alzheimer's disease

1.3.1. Global measures of DNA methylation and hydroxymethylation

Initially, most human epigenetic studies on AD focused on identifying changes in global DNA methylation in post-mortem brain tissue, employing immunohistochemical techniques with antibodies specific for 5mC or 5hmC (Table 1.1). These studies highlight the potential tissue-specificity of epigenetic modifications, as some studies of the hippocampus (5mC + 5hmC; Chouliaras et al., 2013), temporal neocortex (5mC; Mastroeni, McKee, Grover, Rogers, & Coleman, 2009) and entorhinal cortex (5mC; Mastroeni et al., 2010, 5hmC; Condliffe et al., 2014) have found a global decrease in DNA modifications in AD patients compared to controls. However, a study of the middle temporal gyrus and middle frontal gyrus demonstrated elevated levels of 5mC and 5hmC in neurons of AD patients (Coppieters et al., 2014). Other studies have published conflicting evidence; Bradley-Whitman and Lovell showed increased 5mC and 5hmC in the hippocampus in AD (Bradley-Whitman & Lovell, 2013), whilst Lashley and colleagues showed no global changes in either modification in the entorhinal cortex (Lashley et al., 2015). A recently published study using the more specific technique of gas chromatography/mass spectrometry (GC/MS) examined global levels of methylation and hydroxymethylation in the superior temporal gyrus and middle temporal gyrus, hippocampus/parahippocampal gyrus, cerebellum, and the inferior parietal lobe across various stages of AD (Ellison et al., 2017). It was found that brain regions affected by AD pathology can show global alterations in 5mC and 5hmC in a temporal pattern. Particularly for 5hmC, hyper-hydroxymethylation was found in the early stages of AD, yet in later stages of the disease these global levels reversed to levels observed in

controls. This disease progression-related change in global DNA methylation and hydroxymethylation levels could contribute to the contradictory results found in the previous immunohistochemical studies, although methodological differences in tissue processing may also play an important role.

Global measurements of DNA modifications							
Article	**Sample type**	**Number of Samples**		**Type of modification**	**Approach**	**Method**	**Main findings**
		AD	**Control**				
Mastroeni *et al.* (2009)	TN, CER	1	1	5mC	Global assessment in discordant monozygotic twin pair	Immunohistochemistry	Decreased global 5mC in neurons in AD
Mastroeni *et al.* (2010)	EC	20	20	5mC	Global assessment	Immunohistochemistry	Decreased global 5mC in neurons in AD
Chouliaras *et al.* (2013)	HIP (CA1, CA3, DG)	10	11	5mC, 5hmC	Global, cell-specific analysis in AD/control samples and a discordant monozygotic twin pair	Immunohistochemistry	Decreased global 5mC and 5hmC in glia and neurons in AD
Bradley-Whitman & Lovell (2013)	HIP/PHG, CER	12 (5 Preclinical + 7 late-stage AD)	5	5mC, 5hmC, 5fC, 5caC	Global assessment	Immunohistochemistry	Increased global 5mC and 5hmC, and decreased 5fC and 5caC in all AD subjects

Global measurements of DNA modifications (Continued)							
Article	**Sample type**	**Number of Samples**		**Type of modification**	**Approach**	**Method**	**Main findings**
		AD	**Control**				
Coppieters *et al.* (2014)	MTG, MFG	29	29	5mC, 5hmC	Global assessment	Immunohistochemistry	Increased global 5mC and 5hmC in AD neurons
Condliffe *et al.* (2014)	EC, CER	13	8	5mC, 5hmC, 5fC, 5caC	Global assessment	Immunohistochemistry	Decreased global 5hmC in EC and CER in AD
Lashley *et al.* (2015)	EC, CER	12	14	5mC, 5hmC	Global assessment	Immunohistochemistry, ELISA	No global neuronal changes in 5mC and 5hmC
Ellison *et al.* (2017)	STG, MTG, HIP/PHG, IPL, CER	29 (8 Preclinical AD + 8 MCI + 11 Late stage AD)	10	C, 5mC, 5hmC	Global assessment	Gas chromatography/mass spectrometry	Global levels of 5mC and 5hmC altered in early disease stages, return to basal levels at later stages

Table 1.1 Overview of studies examining global DNA modifications in AD.

Abbreviations: CA1: cornu ammonis area 1, CA3: cornu ammonis area 3, CER: cerebellum, DG: dentate gyrus, EC: entorhinal cortex, HIP: hippocampus, IPL: inferior parietal lobe, MFG: middle frontal gyrus, MTG: middle temporal gyrus, PHG: parahippocampal gyrus, STG: superior temporal gyrus, TN: temporal neocortex.

1.3.2. Candidate gene studies of DNA methylation

Although global DNA modification studies in AD have highlighted a potential role for epigenetic mechanisms in AD, it is, however, important that changes in individual genes are studied, to allow us to better understand mechanisms and pathways. Candidate gene studies in AD have mainly focused on methylation profiling of a vast array of genes, the majority of which were studied based on previous association with either EOAD or LOAD (see Table 1.2). Initially, studies focussed on *APP* gene methylation in AD, although these have been relatively inconsistent; an early study of *APP* methylation found AD-associated hypomethylation of APP in the temporal lobe (West et al., 1995), whilst more recent studies have found no AD-related changes in DNA methylation (Barrachina & Ferrer, 2009; Brohede et al., 2010), or even increased *APP* DNA methylation (Iwata et al., 2014). However, the use of different techniques in each of these studies, the limited sample numbers and the use of heterogeneous tissue could be responsible for conflicting results, in addition to the previously discussed temporally sensitive DNA methylation profile. Since the initial study of *APP*, a wide range of other candidate genes have been studied. DNA methylation alterations have been found in a number of genes, including *PP2AC* (Sontag et al., 2004), *S100A2* and *SORBS3* (Siegmund et al., 2007), *BDNF*, *SYP*, NF-*κβ*, and *COX*-2 (Rao et al., 2012) and *TREM2* (Celarain et al., 2016; A. R. Smith et al., 2016). Studies of alterations in *MAPT* DNA methylation in AD are currently inconclusive, with reports of both AD-associated hypomethylation (Iwata et al., 2014) and no AD-related changes (Barrachina & Ferrer, 2009).

Candidate gene studies of AD							
		Number of Samples					
Article	Sample type	AD	Control	Type of modification	Approach	Method	Main findings
West *et al.* (1995)	TL (BA38)	1	1 control, 1 non-AD dementia	5mC	*APP* targeted	Southern blot	AD-associated decreased *APP* methylation
Sontag *et al.* (2004)	MFG, STG, CER	48	24	5mC	*PP2AC* targeted methylation analysis	Western blot	AD-associated decreased *PP2AC* methylation
Siegmund *et al.* (2007)	ALTLC	18	39	5mC*	50 target genes	RT-PCR (MethyLight)	AD-associated decreased *S100A2* methylation, and increased *SORBS3* methylation
Wang *et al.* (2008)	PFC, lymphocytes	Blood: 6 Brain: 24	Blood: 6 Brain: 10	5mC*	12 target genes previously associated with AD	MALDI-TOF mass spectrometry	Identified epigenetic drift from the norm in late-onset AD
Barrachina & Ferrer (2009)	FC, HIP	Stage I-II: 17 Stage III-IV: 15 Stage V-VI: 12	26	5mC*	*MAPT*, *APP*, *PSEN1*, *RAGE*, *ADORA2A* and *UCHL1* targeted analysis of 5mC in AD and other tauopathies	MALDI-TOF mass spectrometry	No AD-associated methylation differences found in any of the target genes
Brohede *et al.* (2010)	FC, PC, TC, CER	6	-	5mC*	5mC analysis targeted at CpG island at 5' end of *APP* gene	Capillary electrophoresis	No methylation at *APP* CpG island in any of the investigated brain regions
Rao *et al.* (2012)	FC (BA9)	10	10	5mC	Global 5mC, and 8 target genes known for differential expression in AD	RT-PCR (MethyLight)	Increased methylation of *BDNF* and synaptophysin, decreased methylation of *NF-κβ* and *COX-2*
Furuya, da Silva, Payão, Bertolucci *et al.* (2012)	EC, AC, HIP, blood	Blood: 34 Brain: 10	Blood: 22 (young), 23 (elderly) Brain: 10	5mC*	*SNAP25* promoter targeted	MALDI-TOF mass spectrometry	No AD-associated methylation differences in *SNAP25* promoter in brain or blood

Furuya, da Silva, Payão, Rasmussen *et al.* (2012)	EC, AC, HIP, blood	Blood: 36 Brain: 12	Blood: 25 (young), 23 (elderly) Brain: 10	5mC*	*SORL1* and *SIRT1* promoter targeted	MALDI-TOF mass spectrometry	No AD-associated methylation differences in *SORL1* and *SIRT1* promoters in brain or blood
Da Silva *et al.* (2013)	EC, AC, HIP	12	10	5mC*	*CNP* and *DPYSL2* targeted	MALDI-TOF mass spectrometry	No AD-associated methylation differences in *CNP* and *DPYSL2* promoters
Da Silva *et al.* (2014)	EC, AC, HIP, blood	Blood: 34 Brain: 12	Blood: 23 Brain: 10	5mC*	Targeted at promoter regions of *HSPA8* and *HSPA9*	MALDI-TOF mass spectrometry	No AD-associated methylation differences in *HSPA8* and *HSPA9* promoters in brain or blood
Iwata *et al.* (2014)	grey matter from ITL, CER, SPL	CER: 45 SPL: 59 ITL: 56	CER: 71 SPL: 76 ITL: 74	5mC*	Genes related to sporadic or familial AD	FACS, pyrosequencing	Increased *APP* methylation and decreased *MAPT* methylation in both neuronal and non-neuronal cells
Celarain *et al.* (2016)	HIP	10	6	5mC*, 5hmC	5mC and 5hmC analysis of *TREM2*	Cloning-based Sanger sequencing, 5hMeDIP	AD-associated increase of 5mC+5hmC in *TREM2* TSS-associated region
A.R. Smith *et al.* (2016)	STG	Cohort 1: 66 Cohort 2: 44 Cohort 3: 117	Cohort 1: 29 Cohort 2: 59 Cohort 3: 75	5mC*	Meta-analysis of 3 cohorts targeting *TREM2*	Illumina 450K array, pyrosequencing	AD-associated hypermethylation of *TREM2*

Table 1.2 Overview of candidate gene studies of AD.

*Studies examining DNA methylation differences between individuals with AD and control groups targeted at specific genes or regions of interest. Abbreviations: AC: auditory cortex, ALTLC: anterior lateral temporal lobe cortex BA: Brodmann area, CER: cerebellum, EC: entorhinal cortex, FC: frontal cortex, HIP: hippocampus, ITL: inferior temporal lobe, MFG: middle frontal gyrus, PC: parietal cortex, SPL: superior parietal lobe, STG: superior temporal gyrus, TC: temporal cortex, TL: temporal lobe. *The methods used in these studies could not distinguish 5mC and 5hmC.*

1.3.3. Epigenome-wide association studies of AD

In the past decade, major technological advances have allowed the first epigenome-wide association studies (EWAS) in AD (Table 1.3; R. G. Smith & Lunnon, 2017). The first AD EWAS utilised the Illumina Infinium HumanMethylation 27K array (27K array) to study >27,000 CpG sites in 14,475 genes (Bakulski et al., 2012). The study identified 948 CpG sites spanning 918 unique genes associated with late-onset AD, in a comparison of frontal cortex tissue from 12 AD cases and 12 age- and sex-matched controls. The most significant AD-associated CpG site showed hypomethylation in the *TMEM59* gene, with a 7.3% difference on the 27K array between AD patients and controls, and a 2.7% difference shown via pyrosequencing validation. Sanchez-Mut *et al.* (2014) also used the 27K array, and examined hippocampal tissue of five Braak stage I-II cases, five Braak stage III-IV cases, five Braak stage V-VI cases and five controls. Braak-associated DNA methylation alterations were found in four loci, of which two resided in *DUSP22*, and one locus each in *CLDN15* and *QSCN6*. The group also reported that hypermethylation of *DUSP22* correlated with its decreased RNA expression.

The 27K array was superseded by the Illumina Infinium HumanMethylation 450K array (450K array), which has been the most widely used method for EWAS in AD to date, although the Illumina Infinium MethylationEPIC array (EPIC array) has been released with additional coverage. The 450K array interrogates more than 485,000 CpG sites covering 98.9% of known UCSC RefGenes and 96% of CpG islands (Bibikova et al., 2011). Lunnon *et al*. used this technique to study a cohort of ~120 donors in a cross-tissue approach, using tissue from the superior temporal gyrus, entorhinal cortex, prefrontal cortex, cerebellum and pre-mortem

blood (Lunnon et al., 2014). These regions are known to be affected differentially by AD, with the entorhinal cortex showing pathology early in the disease process, and the cerebellum remaining relatively unaffected (Wenk, 2003). Initially focusing on the entorhinal cortex, the study found two neighbouring differentially methylated positions (DMPs) located in the *ANK1* gene among the most significant Braak stage-associated sites. These results were replicated in the prefrontal cortex and superior temporal gyrus in the same individuals, yet no disease-associated changes in these loci were found in the cerebellum or pre-mortem blood. The study also validated AD-associated *ANK1* hypermethylation in a further three validation cohorts, including one that used another technology (bisulfite pyrosequencing). Genetic variation in *ANK1* has been associated with diabetes (Harder et al., 2013; Imamura et al., 2012) and, interestingly, DNA methylation changes in this gene were implicated in AD in the study by De Jager and colleagues (De Jager et al., 2014). Their EWAS study determined dorsolateral prefrontal cortex DNA methylomic profiles associated with Aβ plaque burden in 708 individuals, and identified 71 plaque burden-associated CpG sites. Twelve of these nominated loci were validated in data taken from the Lunnon *et al.* study, reaching Bonferroni significance. Of the genes identified, *ANK1*, *CHD23*, *DIP2A*, *RHBDF2*, *RPL13*, *SERPINF1* and *SERPINF2* were found to show significant AD-associated gene expression changes, indicating a potential functional role for the DNA modifications identified in AD. Furthermore, Watson *et al.* (2016) examined superior temporal gyrus tissue from 34 patients with LOAD, and an equal number of age-, race-, and sex-matched non-demented controls. They analysed neighbouring AD-associated CpG sites and identified 479 differentially methylated regions (DMRs) with an average size of 927 bp, the majority of which were found to be hypermethylated in AD. Of the 25 most

significant DMRs, eight genes (*LOC100507547*, *PRDM16*, *PPT2*, *PPT2-EGFL8*, *PRRT1*, *C10orf105*, *CDH23*, and *RNF39*) had been previously reported in the first AD EWAS studies (De Jager et al., 2014; Lunnon et al., 2014).

One caveat of the EWAS studies described above is that they have all used bisulfite-treated DNA, which means that 5mC and 5hmC are indistinguishable, and results actually represent the sum of the two modifications. A recent adaptation to the protocol, whereby DNA is first oxidised prior to bisulfite treatment, allows measurement of 5mC alone. Furthermore, by performing an oxidative-bisulfite treatment in parallel with a bisulfite treatment, one can subtract one value from the other to generate a measurement of 5hmC in isolation (Fukuzawa et al., 2016). At the time of writing, three EWAS of 5hmC in AD have been published. The first study used high-throughput sequencing to examine the distribution of 5hmC in dorsolateral prefrontal cortex tissue of 20 AD patients, four MCI patients, and six non-demented controls (J. Zhao et al., 2017). A total of 517 plaque-associated differentially hydroxymethylated regions (DHRs) were identified, along with 60 NFT-associated DHRs (at $q<0.05$). However, due to the low sample size these results will require replication in further studies. It should also be noted that due to low sequencing resolution, this study was not able to sensitively differentiate 5mC and 5hmC. Two further EWAS studies have been carried out since, both examining 5mC and 5hmC in parallel. A study by Smith *et al.* (2019) examined the entorhinal cortex, finding a differentially methylated locus in *WNT5B* associated with Braak stage, and DMRs in *ARID5B* and *ANK1,* where the latter overlapped with the findings of De Jager *et al.* (2014) and Lunnon *et al.* (2014) and was validated using pyrosequencing. Furthermore, a DHR was found in the *FBXL16* gene in association with Braak stage. The third EWAS assessed

5mC and 5hmC in the middle temporal gyrus (MTG) of controls and AD patients, in addition to examining blood DNA methylation in parallel in a separate longitudinal cohort of preclinical AD patients and controls (Lardenoije et al., 2019). As a joint first author of this paper, a copy has been added in Appendix C of this thesis. A case-control analysis in the brain identified several AD-associated genomic regions, including one DMR in *OXT* and one DHR in *CHRNB1*. In blood, controls were compared to individuals who were non-demented at baseline, but who had developed AD by a follow-up point at least 4.5 years later. In the baseline analysis, three genome-wide significant DMPs and 15 DMRs were detected, with a further 266 DMPs and 21 DMRs at follow-up. Interestingly, the DMR found in *OXT* was detected in the MTG as well as the blood baseline measurement, showing hypomethylation in the MTG in AD, and hypermethylation in the blood of future converters.

In addition to the study by Lardenoije (Lardenoije et al., 2019), several other EWAS on blood DNA methylation in AD have been carried out. The study by Kobayashi *et al.* (2016) generated 450K array data in blood from 12 individuals, classified into an AD group, an aMCI group, and a control group. At a nominally significant level ($p < 0.05$), 1021 loci were found to show differential methylation patterns between the three groups. Four of these loci were located in the *NCAPH2/LMF2* gene and were validated using pyrosequencing, with lower methylation levels found in the MCI and AD groups compared to controls. It should be noted that the analyses carried out do not account for cell type proportions. Second, Madrid et al (2018) examined DNA methylation in whole blood samples from 45 LOAD patients and 39 controls using the Illumina EPIC array, investigating over 850,000 CpG sites. A total of 466 CpG sites showed

differential methylation levels between groups, with several sites located in genes previously associated with LOAD, for example *B3GALT4*, *FLOT1*, *DLG2,* and *OXT.* Of the total differentially methylated sites, 17 sites also showed associations with various CSF measures and the Rey Auditory Verbal Learning Test (RAVLT). Through pyrosequencing, seven of these 17 sites located in the *B3GALT4* gene were validated. Third, Konki *et al.* (2019) studied DNA methylation profiles in blood of 23 AD-discordant twin pairs using reduced representation bisulfite sequencing. Based on *APOE* genotypes and 21 risk genes for AD, discordant twins were found to have similar risk scores for AD. An analysis of both monozygotic and dizygotic twin pairs discordant for AD yielded 11 DMPs, located in or near the genes *DEFA1, TSNARE1, DEAF1, ARAP2, CNPY1, ADARB2, ARHGAP8, GTF3C2, ACTA1, SEMA5A,* and *CLIP2.* Investigation of DNA methylation in a separate cohort of six AD and six control samples from the anterior hippocampus identified a gene altered in both brain and blood, with a DMP found in *ADARB2* near the DMP identified in the analysis of blood samples. A fourth study, conducted by Vasanthakumar and colleagues (2020) performed an EWAS on blood DNA methylation data from the Alzheimer's disease Neuroimaging Initiative (ADNI) cohort, which has been used in Chapters 4 and 5 of this thesis. The study by Vasanthakumar *et al.* used data from 653 individuals, and compared DNA methylation profiles from controls, individuals with MCI, and individuals with AD. A total of 42 DMPs were associated with AD relative to control, 25 DMPs were associated with MCI relative to control, and 13 DMPs were related to differences between MCI and AD. This study was conducted independently, but at the same time as the AD, MCI, and control EWAS presented in Chapter 3 of this thesis.

Epigenome-wide association studies of AD

Article	Sample type	Number of Samples		Type of modification	Approach	Method	Main findings
		AD	Control				
Bakulski *et al.* (2012)	FC	12	12	5mC*	AD-control EWAS	Illumina 27K array	948 AD-associated CpG sites representing 918 unique genes, *TMEM59*
Sanchez-Mut *et al.* (2014)	HIP	Braak stage I-II: 5 Braak stage III-IV: 5 Braak stage V-VI: 5	5	5mC*	Braak stage EWAS	Illumina 27K array	4 Braak-associated CpG sites in *DUSP22, CLDN15, QSCN6 genes*
De Jager *et al.* (2014)	DLPFC	708		5mC*	Plaque burden EWAS	Illumina 450K array	71 Plaque burden-associated CpG sites, *BIN1*, *RHBDF2*, *ANK1*
Lunnon *et al.* (2014)	EC, STG, PFC, CER, blood	122		5mC*	Braak stage EWAS	Illumina 450K array	*ANK1, MIR486, RHBDF2*
Kobayashi *et al.* (2016)	Blood	AD: 4 MCI: 4	4	5mC*	AD-control EWAS	Illumina 450K array, pyrosequencing	Hypomethylation in *NCAPH2*/*LMF2* region in MCI and AD
Watson *et al.* (2016)	STG	34	34	5mC*	AD-control EWAS	Illumina 450K array	479 AD-associated DMRs
Yu *et al.* (2016)	CD4+ T-cells, DLPFC	41 (longitudinal)		5mC*	Plaque burden EWAS	Illumina 450K array	DLPFC results were not replicated in CD4+ lymphocytes.
Zhao *et al.* (2017)	DLPFC	AD: 20 MCI: 4	6	5hmC*	Plaque & NFT EWAS	High-throughput sequencing	517 plaque-associated DHRs, 60 NFT-associated DHRs

Epigenome-wide association studies of AD (continued)

Article	Sample type	Number of Samples		Type of modification	Approach	Method	Main findings
		AD	Control				
Gasparoni *et al.* (2018)	FC, TC, cell-sorted OC	FC: 63 TC: 65 OC: 31		5mC*	Cell-sorted ageing and Braak stage EWAS	Illumina 450K array	Replication of differential methylation in *HOXA3, APP,* and *ANK1* in AD. Novel AD-associated DMPs in *LRRC8B* and *MCF2L,* and age-related DMPs in *CLU.*
Madrid *et al.* (2018)	Blood	45	39	5mC*	AD-control EWAS	Illumina EPIC array, pyrosequencing	477 AD-associated DMPs (i.a. in *B3GALT4*, *FLOT1*, *DLG2*, *OXT*). *B3GALT4* sites associated with CSF & cognitive measures.
R.G. Smith *et al.* (2018)	PFC, STG	147		5mC*	Braak stage EWAS	Illumina 450K array	Large neuropathology-associated DMR in the *HOXA* gene cluster spanning 48 kb.
Altuna *et al.* (2019)	HIP	26	12	5mC*	AD-control EWAS	Illumina 450K array	118 AD-related DMPs in 159 genes, many of which were related to neural development and neurogenesis.
A.R. Smith *et al.* (2019)	EC	Discovery cohort: 96 Validation cohorts: 104, 96		5mC,5hmC	Braak stage EWAS	Illumina 450K array	DMP in *WNT5B*, DMRs in *ANK1, ARID5B*, DHR in *FBXL16*, DUR in *ALLC, JAG2, ARID5B. ANK1* hypermethylation and hypohydroxymethylation in AD.
Konki *et al.* (2019)	Blood, AH	Blood: 23 discordant twin pairs AH: 6 AD, 6 control		5mC*	AD-control EWAS	Illumina HiSeq2500/3000, pyrosequencing	11 AD-associated DMPs detected in blood, though no association was found with gene expression. Differential methylation in *ADARB2*, in both brain and blood.

Epigenome-wide association studies of AD (continued)

Article	Sample type	Number of Samples		Type of modification	Approach	Method	Main findings
		AD	Control				
Lardenoije *et al.* (2019)	MTG, blood	MTG: 45 Blood: 42	MTG: 36 Blood: 54	5mC, 5hmC	AD-control EWAS, EWAS of future conversion to AD	Illumina 450K array	MTG: one DMR in *OXT*, one DHR in *CHRNB1*, 11 DURs. Blood: 3 DMPs and 15 DMRs at baseline prior to conversion, 266 DMPs and 21 DMRs after conversion to AD. OXT DMR found in both brain and blood.
Vasanthakumar *et al.* (2020)	Blood	AD: 94 MCI: 333	220	5mC*	AD-MCI-control EWAS	Illumina EPIC array	42 DMPs associated with AD, 25 with MCI, and 13 with MCI vs. AD differences. DMPs enriched near brain tissue-specific genes.
Roubroeks *et al.* (2020)	Blood	AD: 86 MCI: 109	89	5mC*	AD-MCI-control EWAS, EWAS of future conversion to AD	Illumina 450K array	Hypermethylated DMR in *HOXB6* in AD, DMRs associated with conversion to AD and altered biological pathways.
Zhang *et al.* (2020)	PFC, blood	PFC: 1030 Blood: 69		5mC*	Braak stage meta-analysis of data from Lunnon *et al*. (2014), De Jager *et al.* (2014), R.G. Smith *et al.* (2018), and Gasparoni *et al.* (2018).	Illumina 450K array	3751 DMPs and 119 DMRs associated with Braak stage in AD brain, including the genes *MAMSTR, AGAP2,* and *AZU1.* Replication of DMRs in *HOXA3, ANK1, RHBDF2, BIN1.* Low overlap between brain and blood DNA methylation.

Table 1.3 Overview of epigenome-wide association studies of AD.

*Studies examining epigenome-wide DNA methylation profiles in AD. Abbreviations: AH: anterior hippocampus, CER: cerebellum, DLPFC: dorsolateral prefrontal cortex, EC: entorhinal cortex, FC: frontal cortex, HIP: hippocampus, MTG: middle temporal gyrus, OC: occipital cortex, PFC: prefrontal cortex, STG: superior temporal gyrus, TC: temporal cortex. *The methods used in these studies could not distinguish 5mC and 5hmC.*

1.4. Aims

This thesis aims to investigate DNA methylation differences in the blood of AD, MCI and control individuals, and to determine the effects of the risk factors of age and sex on blood DNA methylation. The specific aims of this project are as follows:

1. To perform the first EWAS of its size on blood DNA methylation data generated in AD, MCI and control individuals, identifying differences at the levels of single CpG sites and regions within the DNA. In addition, weighted gene correlation network analysis (WGCNA) will be performed to identify clusters of co-methylated loci, with biological pathway analyses run subsequently (Chapter 3).
2. To examine the effects of age and sex on blood DNA methylation of the autosomes, and their interaction with disease (Chapter 4).
3. To identify sex-specific and disease-associated differences in blood DNA methylation on the sex chromosomes and on the inactivated X chromosome (Xi; Chapter 5)

CHAPTER 2. MATERIALS AND METHODS

This chapter provides an overview of the samples, laboratory work, and data analysis methods used in multiple chapters of this thesis. Data analyses specific to a single chapter will be described in detail in the relevant chapter.

2.1. The AddNeuroMed cohort

The data and samples used to conduct the research described in this thesis include DNA methylation data which was generated by Professor Katie Lunnon from whole blood samples from the AddNeuroMed cohort (2.1, Chapters 3, 4, and 5) as well as previously generated DNA methylation data from the ADNI study which was used for replication purposes (described in section 2.3; Chapters 4 and 5).

2.1.1. AddNeuroMed cohort demographics

The AddNeuroMed cohort was formed by the European Innovative Medicines (InnoMed) initiative, with the goal of diagnostic and prognostic biomarker discovery for AD (Lovestone et al., 2007, 2009). This initiative was funded by the European Union as well as by members of the European Federation for Pharmaceutical Industries and Associations. The study represents a collaboration between six centres across Europe where participants were recruited and assessed according to standardised procedures: the University of Eastern Finland (Kuopio, Finland), the University of Łódź (Łódź, Poland), King's College London (London, United Kingdom), Università degli Studi di Perugia (Perugia, Italy), Aristotle University of Thessaloniki (Thessaloniki, Greece), and Centre Hospitalier Universitaire de Toulouse (Toulouse, France). Informed consent was obtained from all participants according to the Declaration of

Helsinki (1991), and all procedures were approved by the relevant institutional boards at each site. Participants included individuals with AD, MCI, and elderly CTL. AD was defined by the NINCDS-ADRDA criteria, MCI as Petersen amnestic and non-amnestic types, and CTL was defined as having a Mini-Mental State Examination (MMSE) score greater or equal to 28 points, and showing no symptoms of dementia. Subjects were excluded from the study if depression or another neurological syndrome was present at baseline which would call the diagnosis into question.

Of the full AddNeuroMed cohort, a subset of 301 individuals were selected for genome-wide assessment of DNA methylation, on the basis of additional availability of transcriptomic (Lunnon et al., 2012, 2013), genomic (Furney, Simmons, et al., 2011), and imaging data (Furney, Kronenberg, et al., 2011; Furney, Simmons, et al., 2011; Westman et al., 2011). A further 16 samples were removed post-profiling as they met the exclusion criterion of being younger than 65 years of age at baseline. After one sample was removed during data pre-processing (as described in 2.2.3.5), 284 samples remained which were used in all baseline analyses described in Chapters 3, 4, and 5. A subset of individuals with MCI had received an AD diagnosis at follow-up assessments (MCI-AD, n = 42). For four of these individuals, the time of conversion was unknown and therefore these samples were excluded from conversion analyses, whilst the remaining 38 individuals progressed to AD within one year after baseline assessment and were thus included in any conversion analysis. For an overview of samples used, see Table 2.1.

AddNeuroMed				
	CTL	MCI		AD
		MCI-MCI	MCI-AD	
N	89	67	42†	86
Sex (M/F)	34/55	34/33	16/26	30/56
Age (Mean ± SD)	73.8 ± 5.3	75.1 ± 5.6	76.3 ± 5.3	76.8 ± 5.6
MMSE (Mean ± SD)	29 ± 1.2	27.3 ± 1.7	26.3 ± 2.2	20.8 ± 4.5

Table 2.1 AddNeuroMed cohort demographics.

Subject characteristics of 284 whole blood DNA methylation samples that passed data pre-processing. Of the original 301 samples, 16 did not meet inclusion criteria, and one sample did not pass quality control (QC). Shown are the number of samples (N), sex (Males/Females), mean age ± standard deviation (SD), and MMSE score ± SD. Of the 109 MCI subjects, 67 remained stable, and 42 progressed to AD. For conversion analyses comparing MCI-MCI to MCI-AD, four MCI-AD subjects were removed due to unknown conversion time, the remaining 38 individuals converted within one year after baseline measurement.

2.2. Genome-wide DNA methylation profiling in AddNeuroMed

Of the full AddNeuroMed cohort, a subset of 301 samples were selected for DNA methylation profiling, which was performed by Professor Katie Lunnon (sections 2.2.1 and 2.2.2).

2.2.1. DNA preparation

DNA had previously been extracted from venous blood baseline samples as described by Furney *et al.* (2011), and was tested for degradation and purity. For each sample, 500ng DNA was sodium bisulfite-treated using the Zymo EZ-96 DNA methylation kit (Zymo Research, CA, USA) according to the manufacturer's standard protocol. First described by Frommer *et al.* (1992), bisulfite treatment can be used to selectively identify the presence of the DNA cytosine modifications 5mC and 5hmC, although levels of 5hmC are negligible in blood (Nestor et al., 2012). With this method, unmodified cytosine bases are converted to uracil, while modified bases remain unchanged. During polymerase chain reaction (PCR) amplification, only modified cytosine sites will be amplified as cytosine, as uracil will be amplified as thymine.

2.2.2. Illumina HumanMethylation 450K arrays

Genome-wide DNA methylation profiles were obtained using the Illumina Infinium HumanMethylation450K BeadChip array (450K array; Illumina, CA, USA). This method is able to assess DNA methylation at more than 485,000 targeted CpG sites within the genome, which cover 98.9% of known UCSC RefGenes and 96% of CpG islands at the time of development (Bibikova et al., 2011). The 450K array makes use of two types of probes to assess methylation, i.e. type I and type II

probes (Figure 2.1). Type I probes originate from the 450K array's predecessor, the Illumina Infinium HumanMethylation27 BeadChip (27K array). For this probe type, every CpG locus is covered by two beads, one to assess a methylated locus (M bead probe), and another to assess an unmethylated locus (U bead probe). The U bead probe contains an adenine base at its 3' terminus, which will bind a query DNA strand containing an unmethylated locus (in which the unmethylated cytosine has been converted to thymine by bisulfite treatment and PCR). This enables single base elongation with labelled dideoxynucleotide triphosphates (ddNTPs), resulting in a fluorescent signal. If the query CpG locus is methylated, this will cause a 3' mismatch pairing, preventing elongation. Conversely, the M bead probe contains a guanine base at the 3' terminus. If the query CpG locus is methylated (and thus the methylated locus was converted to cytosine following bisulfite treatment and PCR), the DNA strand will bind to the probe and result in single base elongation and a fluorescent signal. In contrast, type II probes require only one type of bead per query CpG locus. The probes will bind to targeted regions, and single base extension will take place at the site of the query CpG locus. If unmethylated, the binding of adenosine to thymine will result in a red fluorescent signal, and if methylated, the binding of guanine to cytosine will result in a green signal. Using the two types of probes, the 450K array chips can bind bisulfite-treated DNA, and, through measurement of the resulting fluorescent signals with an Illumina HiScan (Illumina, CA, USA), DNA methylation levels at specific loci are assessed.

For the generation of DNA methylation data in the AddNeuroMed cohort samples, bisulfite treatment of the DNA was performed in four batches, and all samples

were randomised with respect to sex, centre of sample collection, and diagnostic status on the 450K array BeadChips, which assess 12 samples per chip.

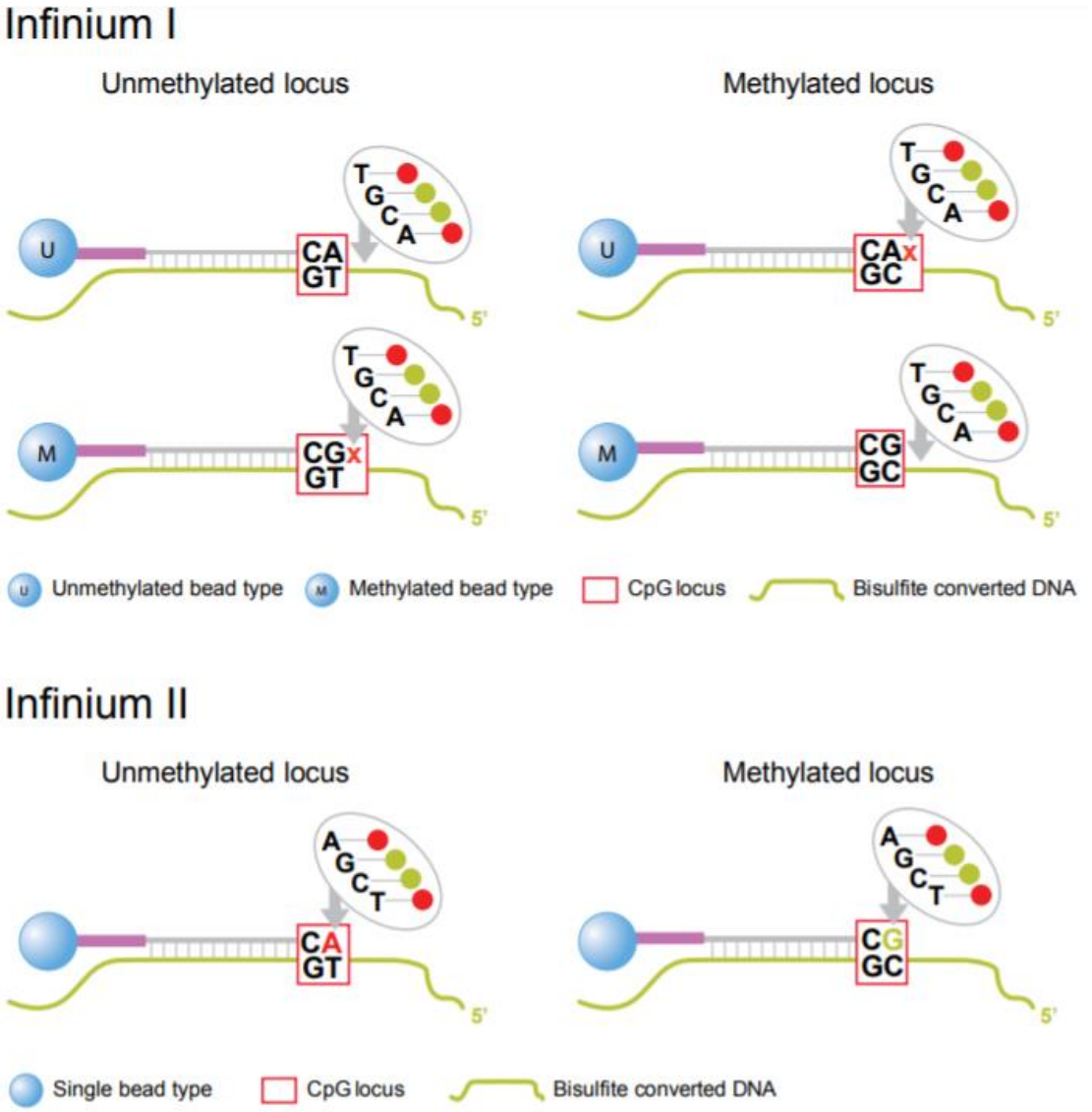

Figure 2.1 Illumina HumanMethylation 450K probe types.

The 450K array employs two probe types, type I (Infinium I), and type II (Infinium II). The U bead in type I probes assesses unmethylated query CpG sites, while the M bead assesses methylated query sites. For type II probes, one bead type assesses both methylated and unmethylated query CpG sites. Taken from Infinium® HumanMethylation450 BeadChip Data Sheet (Illumina).

2.2.3. Data pre-processing

Unless mentioned otherwise, all data processing was performed in the R statistical environment (version ≥ 3.5.2; R Development Core Team 3.0.1., 2013). R code for the computational methods described in this chapter, and Chapters 3, 4, and 5 can be found on https://github.com/JanouR/PhD_Thesis.

Following HiScan measurement, raw signal intensity data files were imported into R as a methylumi object using the *methylumi* package (Davis, S., Du, P., Bilke, S., Triche, T., Bootwalla, M, 2015). DNA methylation values were generated by calculating β-values as a ratio of the methylated and unmethylated signal intensities (M and U, respectively), according to the following equation:

$$\beta = \frac{M}{M + U + 100}$$

β-Values range from zero to one, indicating no methylation present at a given loci at the zero value, and a given loci is fully methylated at the one value. Median methylated and unmethylated signal intensities were visually inspected to ensure no samples displayed extremely low intensities which clearly deviated from the main cluster of samples (Figure 2.2), and a density plot of raw β-values can be seen in Figure 2.3.

Quality control (QC) and pre-processing was performed using the *methylumi* and *wateRmelon* packages (Schalkwyk et al., 2013), following the steps described below in sections 2.2.3.1 - 2.2.3.7.

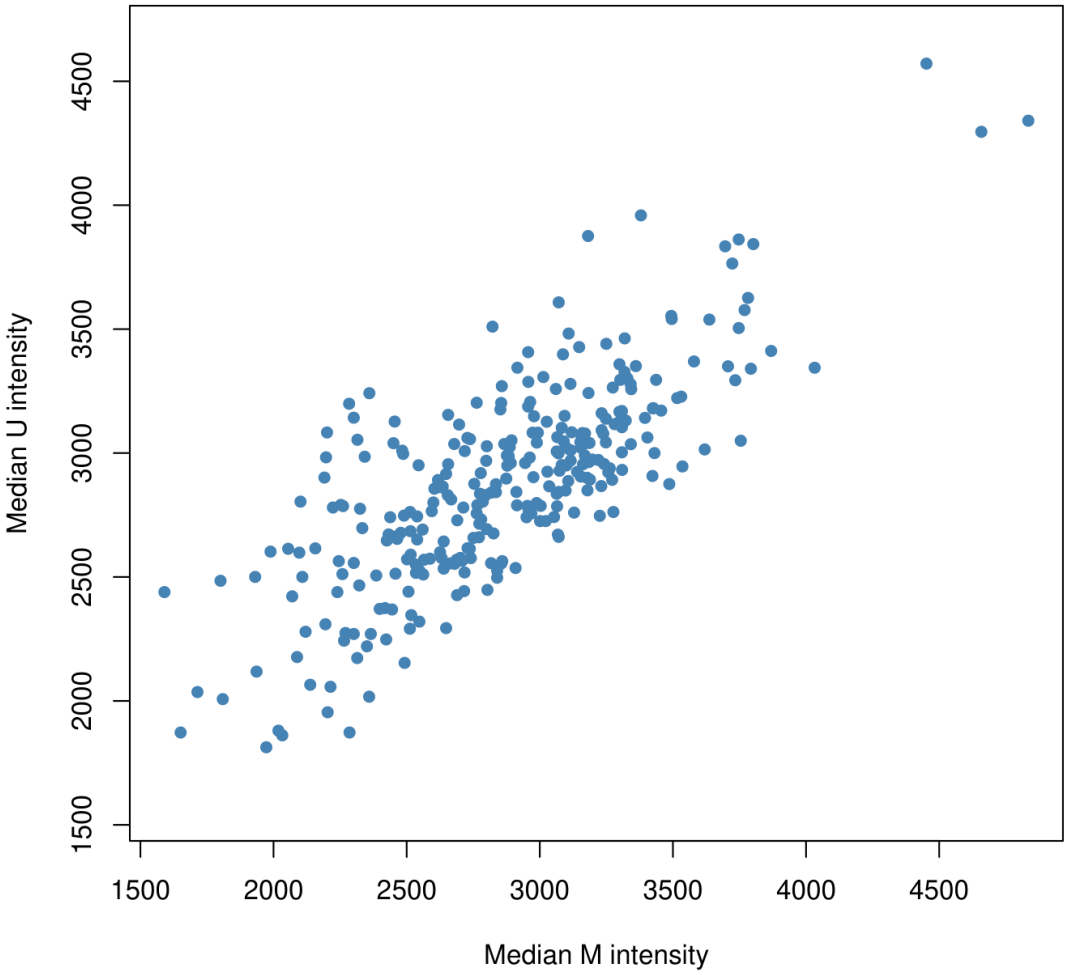

Figure 2.2 Raw median methylated and unmethylated signal intensities in AddNeuroMed samples.

None of the samples showed extremely low intensity levels that deviate from the main cluster of samples. Therefore, all samples could be taken forward in the QC pipeline.

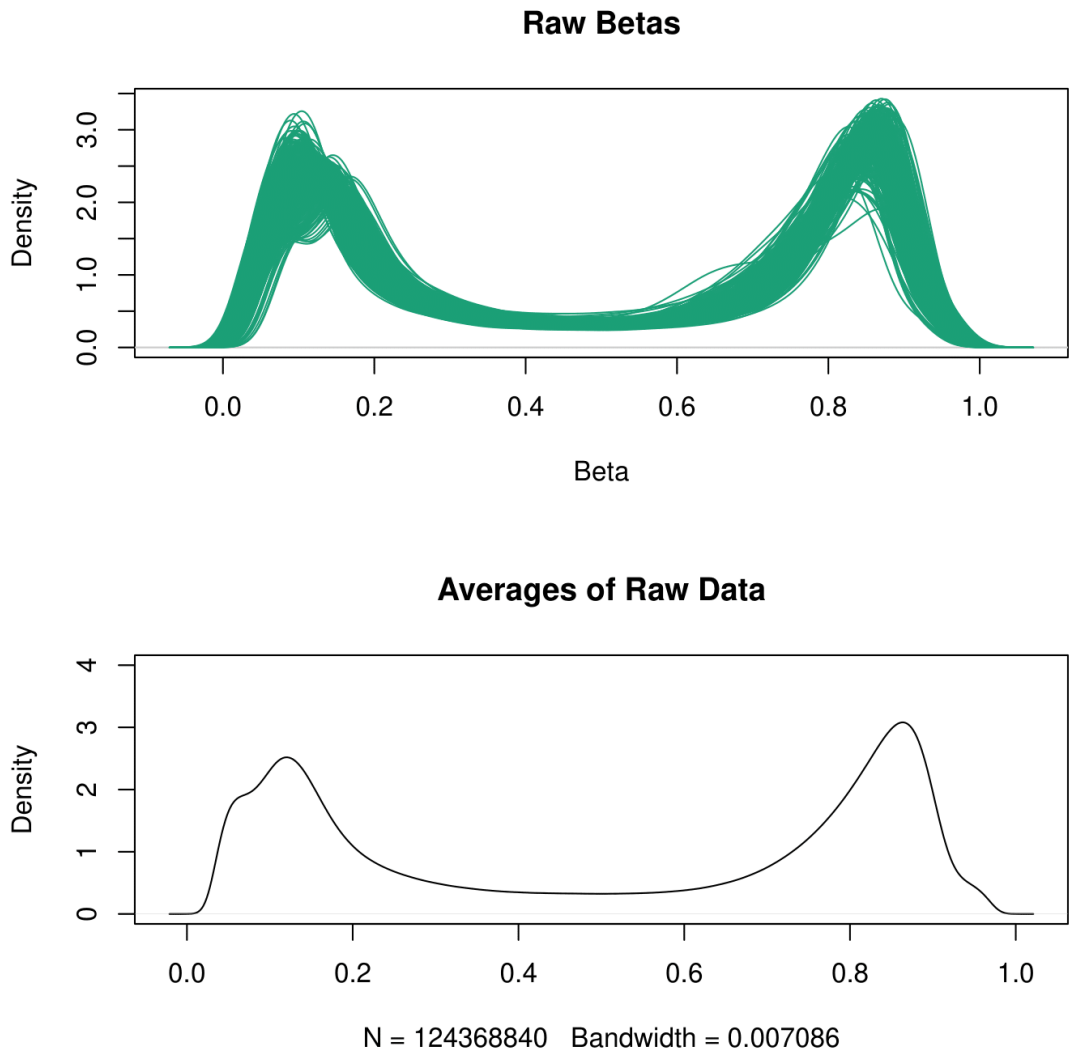

Figure 2.3 Density plots of raw β-values in the AddNeuroMed cohort.

Density plots displaying the frequency of DNA methylation values across all individual AddNeuroMed blood samples (top), and for the average of all samples (bottom).

2.2.3.1. Bisulfite conversion efficiency

The efficiency of the bisulfite treatment of the DNA was assessed using the *bscon* function in the *wateRmelon* package. This function makes use of a set of fully methylated control probes that are included on the array, calculating a β-value for these probes, indicative of conversion efficiency. A threshold of a median value of ≥ 80% for each sample is commonly used, which all samples passed (Figure 2.4).

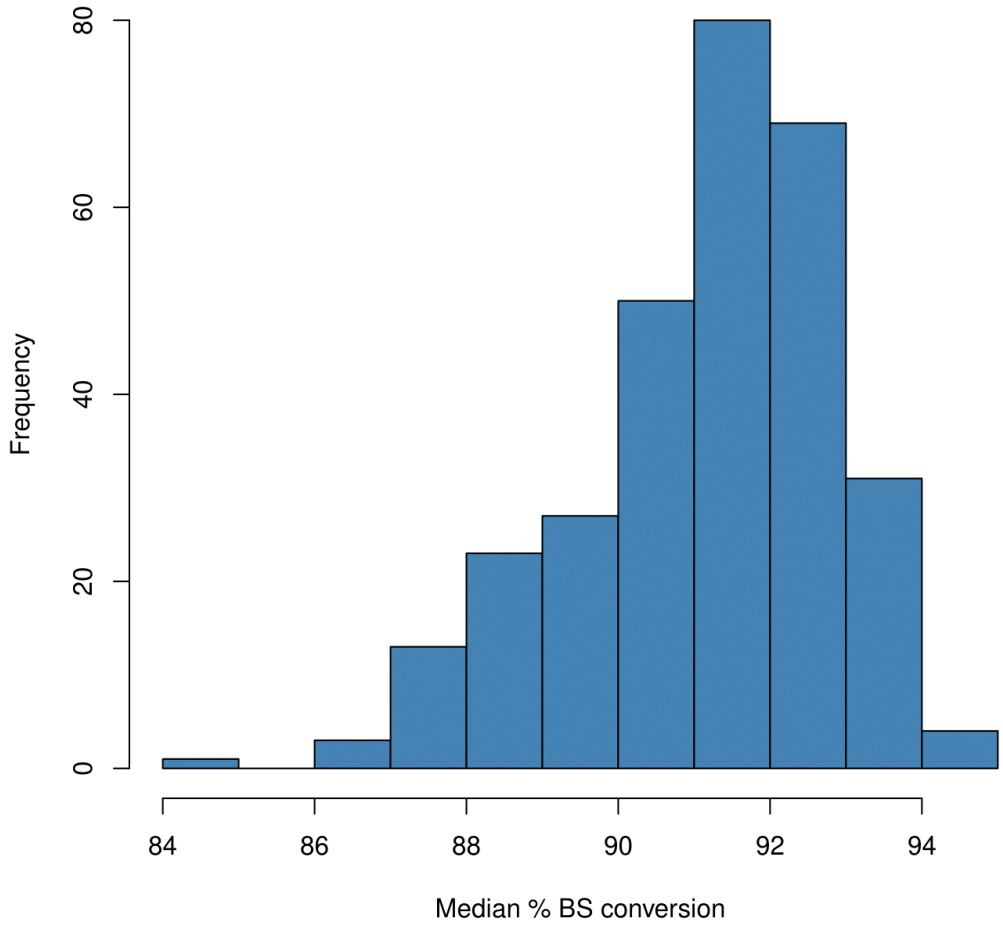

Figure 2.4 Bisulfite conversion efficiency in AddNeuroMed samples.

Distribution of median methylation percentage across samples in fully methylated control probes on the 450K array. All samples passed the minimum median methylation percentage of 80%.

2.2.3.2. Genetic distinctness of samples

The 450K array contains 65 probes that interrogate single nucleotide polymorphism (SNP) sites by design, which can be used to assess whether any samples are genetically identical. The β-values for these samples were extracted and correlated, and the maximum correlation for each sample with all other samples was extracted. Typically, a value >0.95 indicates that samples may be genetically identical, which was, as expected, not found for the samples in this study (Figure 2.5).

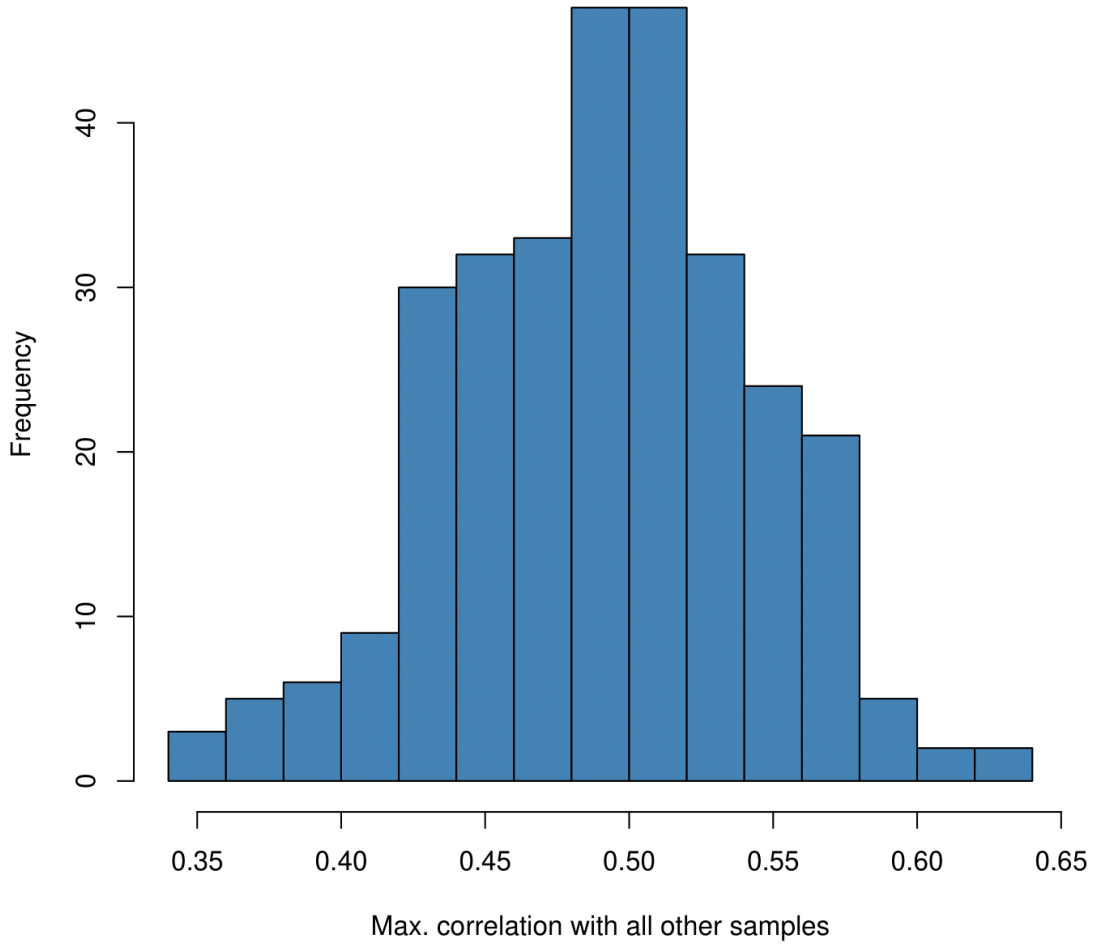

Figure 2.5 Genetic distinctness of samples in the AddNeuromed cohort.

Maximum correlations found between samples for 65 probes interrogating SNP loci. As none of the samples showed correlations higher than r > 0.95, this indicates that the samples are not genetically identical, as expected.

2.2.3.3. Removal of cross-hybridising and SNP probes

The next step in the QC process included the removal of probes that have been previously reported to be technically questionable in 450K array data studies. These probes either contain a SNP within the target sequence, or may co-hybridise, as they have a highly similar genetic sequence to other genomic sites. These probes have been described in Price *et al*. (2013) and Chen *et al.* (2013), and were removed from the original 485,577 probes on the array, leaving 413,510 probes for further QC and analysis.

2.2.3.4. Multidimensional scaling of sex chromosomes

The next step in the QC process involves checking whether the sex predicted by the 450K array data (based on an individual's DNA methylation profile on the X and Y chromosomes) matched the clinical record for that individual, which can help to detect mislabelling of samples. This was done by the creation of a distance matrix (*dist* function) and multidimensional scaling (MDS) of sex chromosome methylation using the *cmdscale* function. A visual inspection of the MDS coordinates plots, coloured by clinically defined sex, confirmed correct labelling of all samples (Figure 2.6).

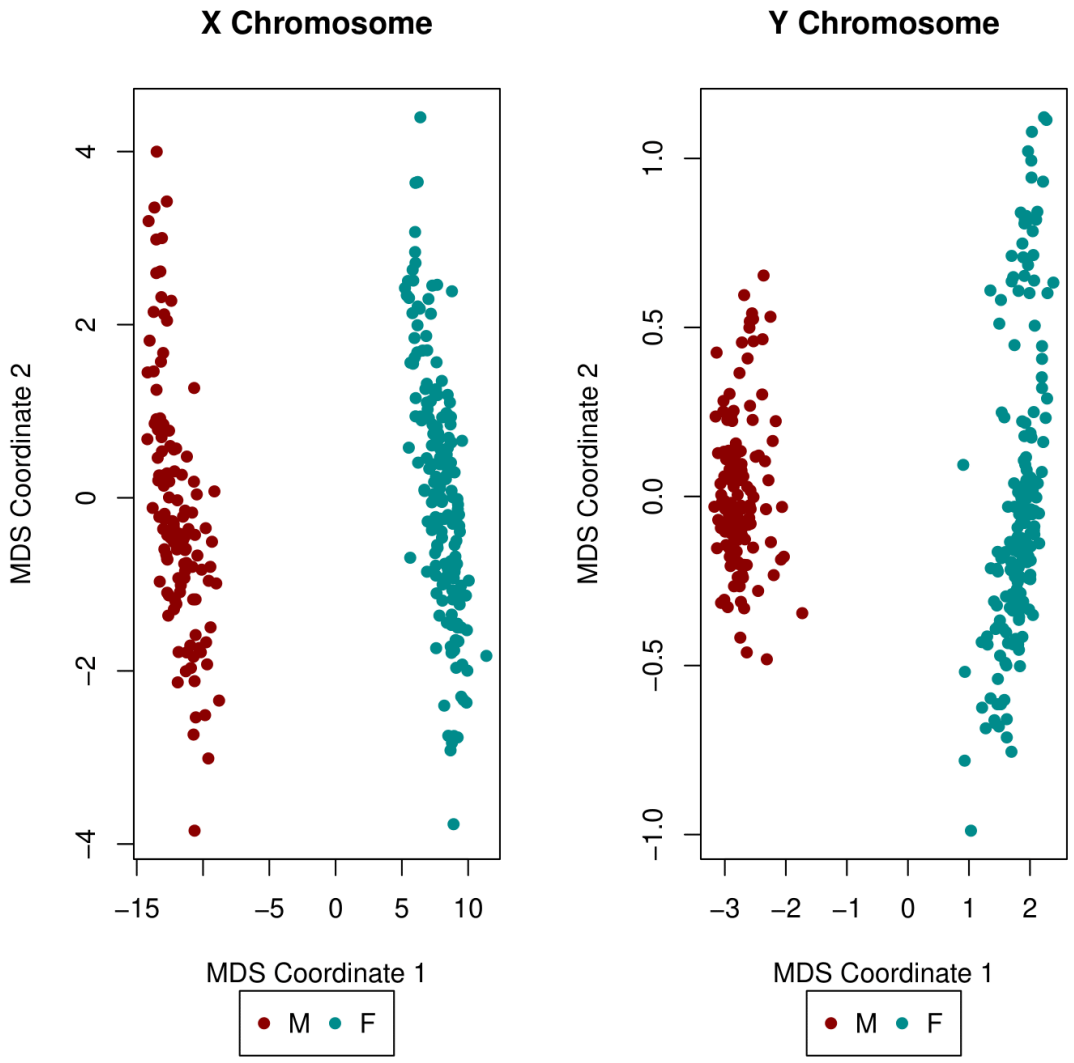

Figure 2.6 Multidimensional scaling of sex chromosomes in AddNeuroMed.

MDS coordinates calculated from the DNA methylation profile the sex chromosomes using the cmdscale function are plotted. Profiles that are more similar cluster together, and as expected two clusters are seen for each chromosome, which correspond to clinically reported sex (i.e. colour of the samples).

2.2.3.5. P-filter

The *pfilter* function in *wateRmelon* is a further step of the QC process, which removes both probes and samples according to the following criteria:

- Samples having 1% of probes with a detection *p*-value greater than 0.05;
- Probes having a beadcount <3 in 5% of samples;
- Probes having 1% of samples with a detection *p*-value greater than 0.05.

Following the application of this function, one sample (out of the original 301 profiled) was removed as a consequence of not fulfilling the first criterion.

2.2.3.6. Outlier removal

As a final control prior to normalisation of the data, the *outlyx* function in *wateRmelon* was applied to identify any outliers in the data. This function uses the first principal component of the data and determines outliers by interquartile ranges. No outliers were present in the final dataset of 284 individuals.

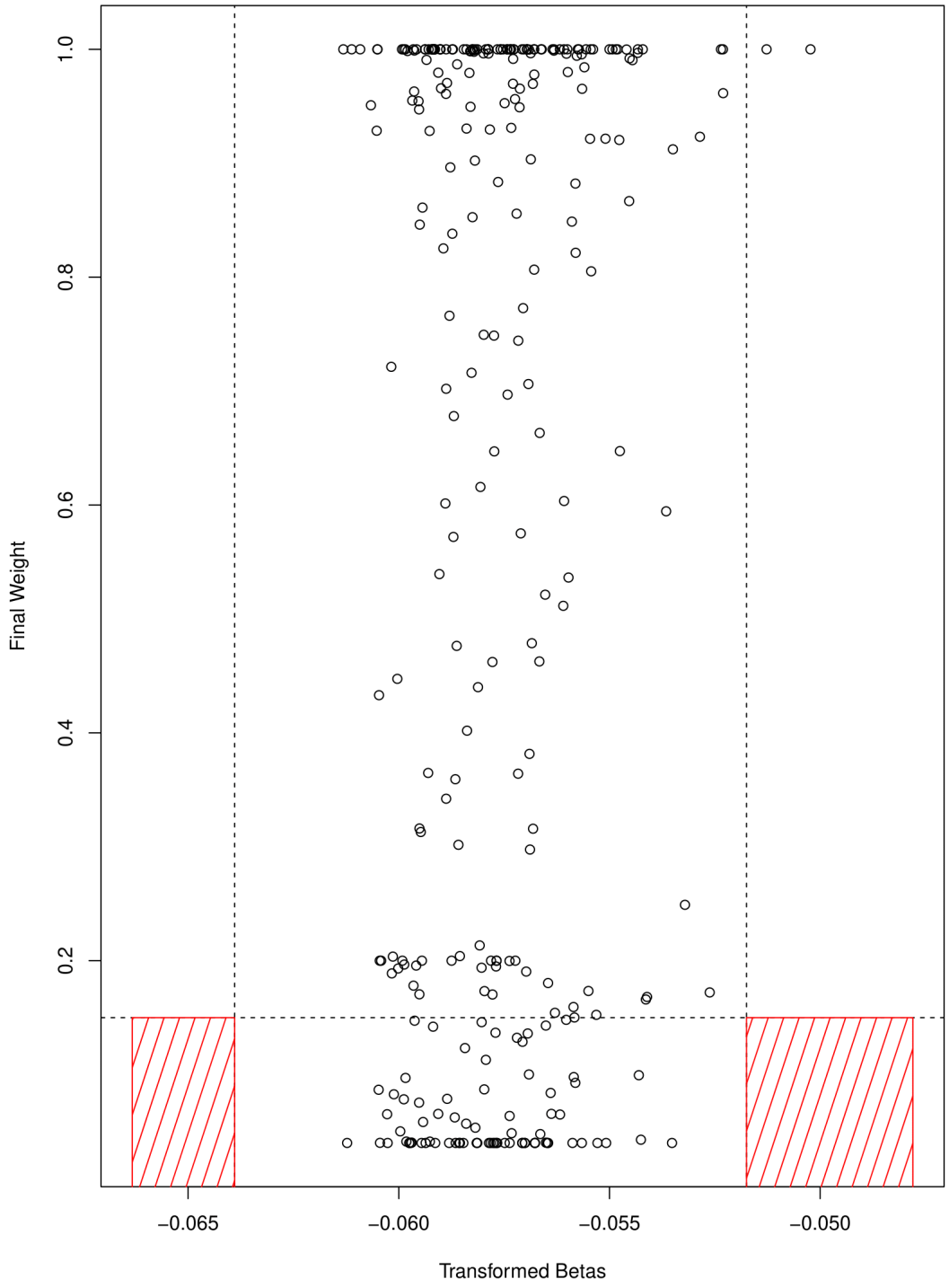

Figure 2.7 Outlyx plot for the identification of outliers in the AddNeuroMed cohort.

The outlyx function was applied to the raw (p-filtered) DNA methylation data to detect any outlying samples. Outlier zones based on interquartile ranges are indicated in red, none of the samples were classified as outliers.

2.2.3.7. Normalisation and final QC

Quantile normalisation of the data was performed using the *dasen* function of the *wateRmelon* package (Figure 2.8; see Pidsley et al. (2013) for a detailed description). Briefly, this method performs a background signal adjustment for both M and U intensity signals, for type I and type II probes, as they perform slightly differently. The background difference between the two types of probes is calculated and added to the type I intensity. The function then performs quantile normalisation of M and U intensity signals separately, as well as separately by probe type I and II. Although the *wateRmelon* package offers a range of functions for normalisation, the *dasen* function was found to perform best in the Pidsley *et al.* (2013) study as well as in our own studies within our research group.

Finally, two further performance metrics were applied as data QC steps. The *dmrse* function in *wateRmelon* assesses β-values at 227 probes that are known to be located in imprinting differentially methylated regions (iDMRs), which are expected to be hemi-methylated (β = 0.5). As a measure of data quality, the function calculates the standard error (SE) of the β-values at these sites, with a low SE found in the current dataset indicating a positive performance (SE = 0.0074, SE type I probes = 0.0073, SE type II probes = 0.0073). In addition, the *genki* function makes use of the 65 SNP probes described in section 2.2.3.2. This function was applied to a separate data object, in which the SNP probes had not been removed, but which had been normalised according to the methods described above. As the samples will represent two homozygous and one heterozygous group for each SNP, the signals for these SNP probes are expected to cluster into three groups. The *genki* function performs one-dimensional K-means clustering on each SNP probe, and calculates an SE-like

statistic for each of the three genotype groups, with low values indicating low technical variation. For the current dataset, the overall *genki* SE values were: 3.47×10^{-5}, 4.21×10^{-5}, and 2.06×10^{-5} (type I probes: 5.10×10^{-5}, 4.53×10^{-5}, 2.96×10^{-5}, type II probes: 2.49×10^{-5}, 4.00×10^{-5}, 1.49×10^{-5}). Considering the low SE values found here, it is concluded that this QC step does not indicate there is any undue technical variation.

After finalising data QC and pre-processing, two sets of data were created for subsequent analyses: one dataset for investigating effects in relation to baseline diagnostic status (*n* = 284 samples: CTL = 89, MCI = 109, AD = 86, see Table 2.1), and one dataset including only the MCI individuals, for analyses examining future progression to AD (*n* = 105 samples). In the subset created for analyses related to future progression to AD, four MCI samples were removed from the dataset as their exact conversion date was unknown, leaving a total of 105 samples (MCI-MCI = 67, MCI-AD = 38). For the analysis conducted in Chapters 3 and 4, 9,676 probes located on the sex chromosomes were removed from the data, leaving 401,266 probes on the autosomes, whereas only the probes located on the sex chromosomes were included in the analyses for Chapter 5.

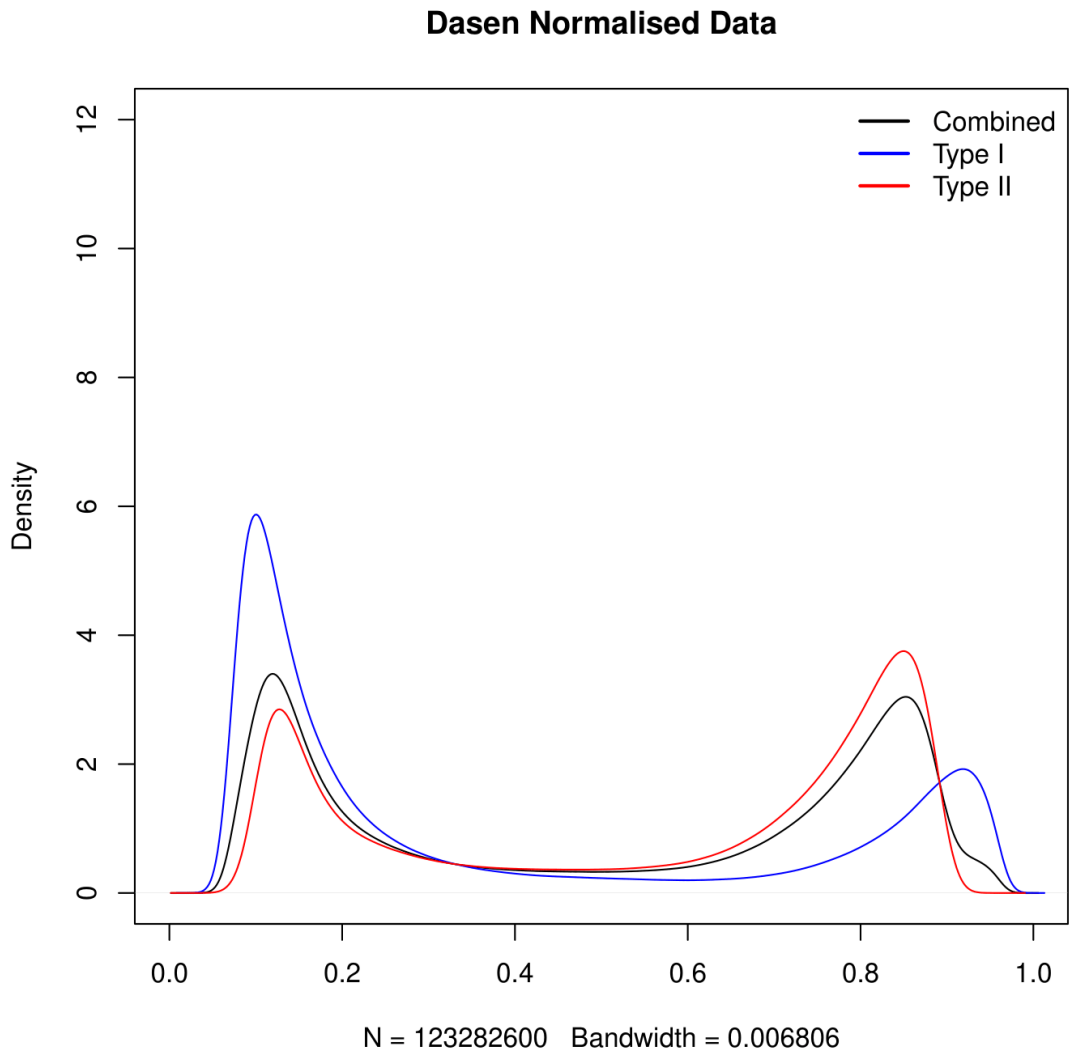

Figure 2.8 β-distribution following dasen normalisation in the AddNeuroMed cohort.

Density plots displaying the frequency of dasen-normalised DNA methylation values. The overall distribution is shown in black, and the data separated by type I or type II probes in blue and red, respectively.

2.2.4. Identification of covariates in the AddNeuroMed cohort

For all analyses of differential methylation, a number of covariates were included based on the identification of variables influencing DNA methylation using principal component analysis (PCA). As it is known that variables such as smoking (Hannon et al., 2018; Zeilinger et al., 2013) and cell type proportion within samples (Houseman et al., 2012) may cause variation, a smoking score was calculated for each individual according to (Elliott et al., 2014; Figure 2.11), and cell type proportions were estimated according to Houseman *et al.* (2012). Additionally, as both bisulfite treatment batch (plate) and array chip may cause technical variation, median methylated and unmethylated intensity signals calculated per plate and BeadChip were visually inspected (Figure 2.9 and Figure 2.10, respectively).

To determine which of these variables, in addition to the well-known confounding variables of sex and age, would be taken into account as covariates in subsequent analyses, an unrotated PCA was performed. Using the *prcomp* function, principal components were generated, each of which account for a certain proportion of the variance in the dataset (Table 2.2). I decided to focus on the first two principal components, as these accounted for more than 5% of the total variance each (11.2% and 5.9%, respectively). For visualisation purposes, the first five principal components were then correlated to potential confounders as well as diagnosis and other variables of interest (Figure 2.12), and only the variables correlated to the first two principal components were identified as covariates to include in subsequent analyses. Based on this information, the variables of age, sex, cell type proportion and bisulfite treatment batch (plate) were taken into account in all analyses. Of note, a small but significant correlation

was observed between samples originating from Toulouse and the first principal component. However, the inclusion of source as a covariate would add five dummy variables to any models, thereby decreasing power substantially. Furthermore, in trial runs of the differential methylation analyses it was found that adding source as a covariate only increased *p*-value inflation in models where inflation was already present (data not shown). Taking into account that the first principal component also strongly correlated with cell type, it was decided to omit the source dummy variables from further analyses.

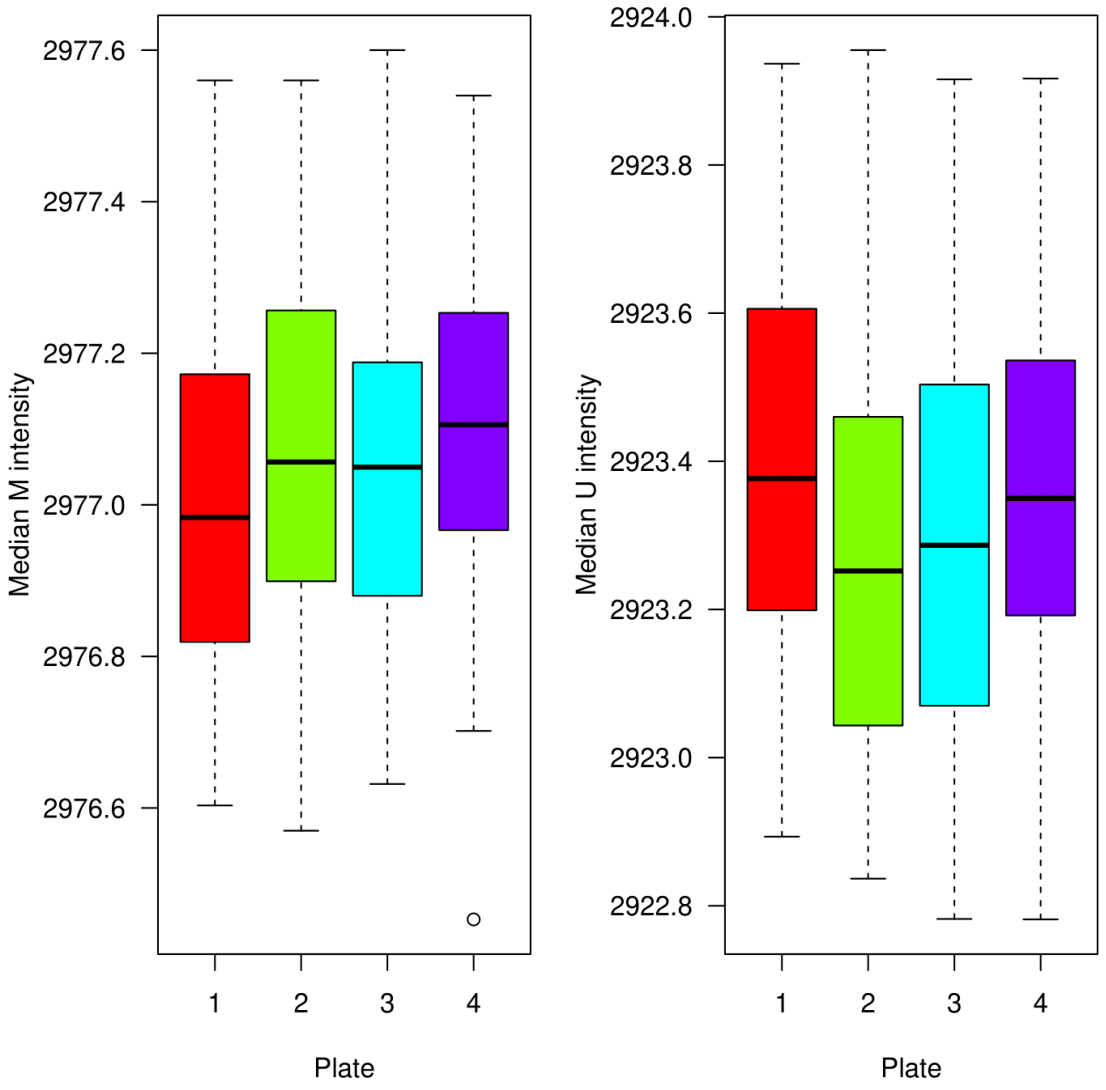

Figure 2.9 Median signal intensities in AddNeuroMed plotted by batch.

Median methylated (M) and unmethylated (U) signal intensities are plotted and coloured by bisulfite treatment plate (batch).

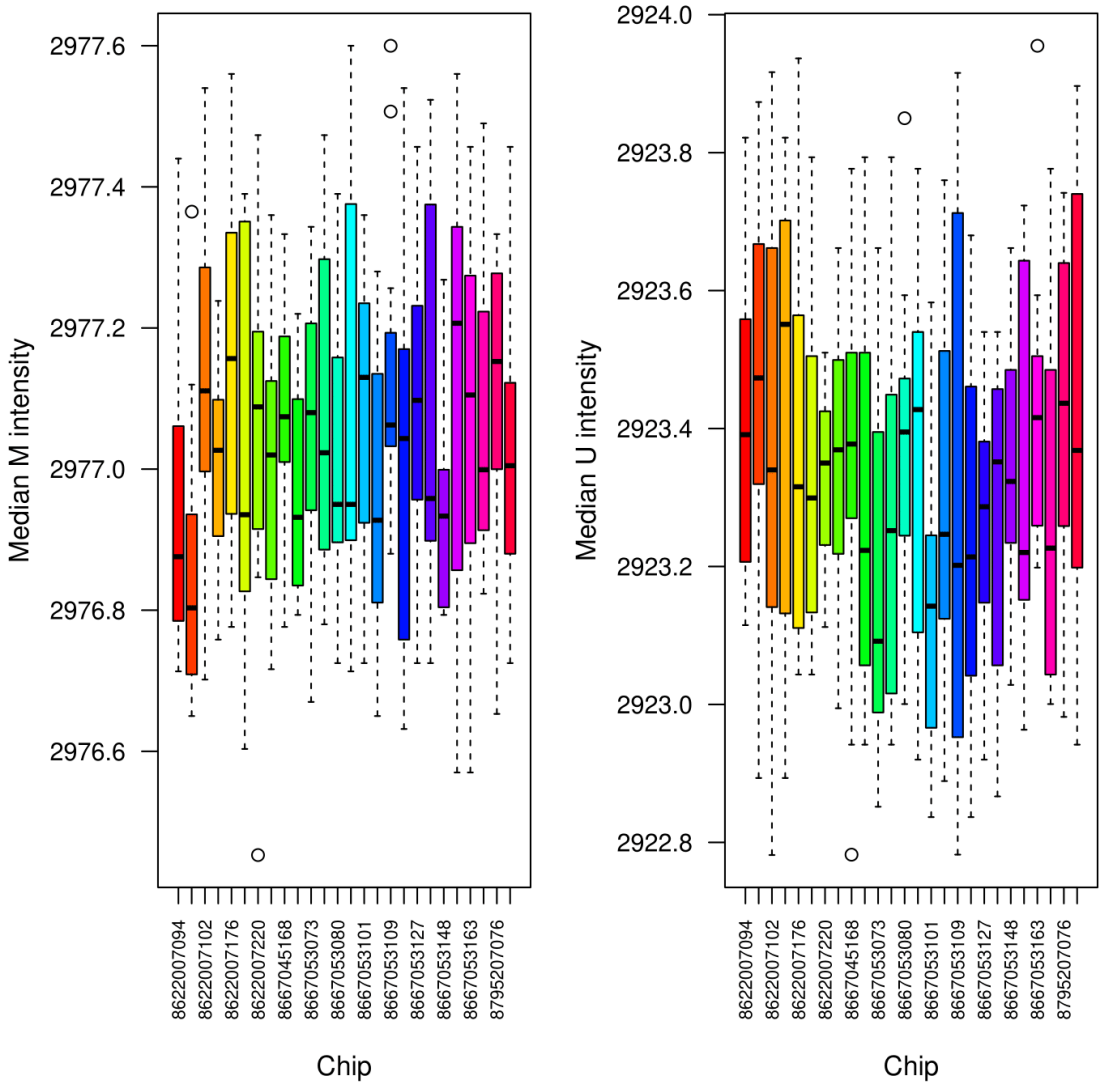

Figure 2.10 Median signal intensities in AddNeuroMed by 450K array chip.

Median methylated (M) and unmethylated (U) signal intensities are plotted and coloured by 450K array Chip ID.

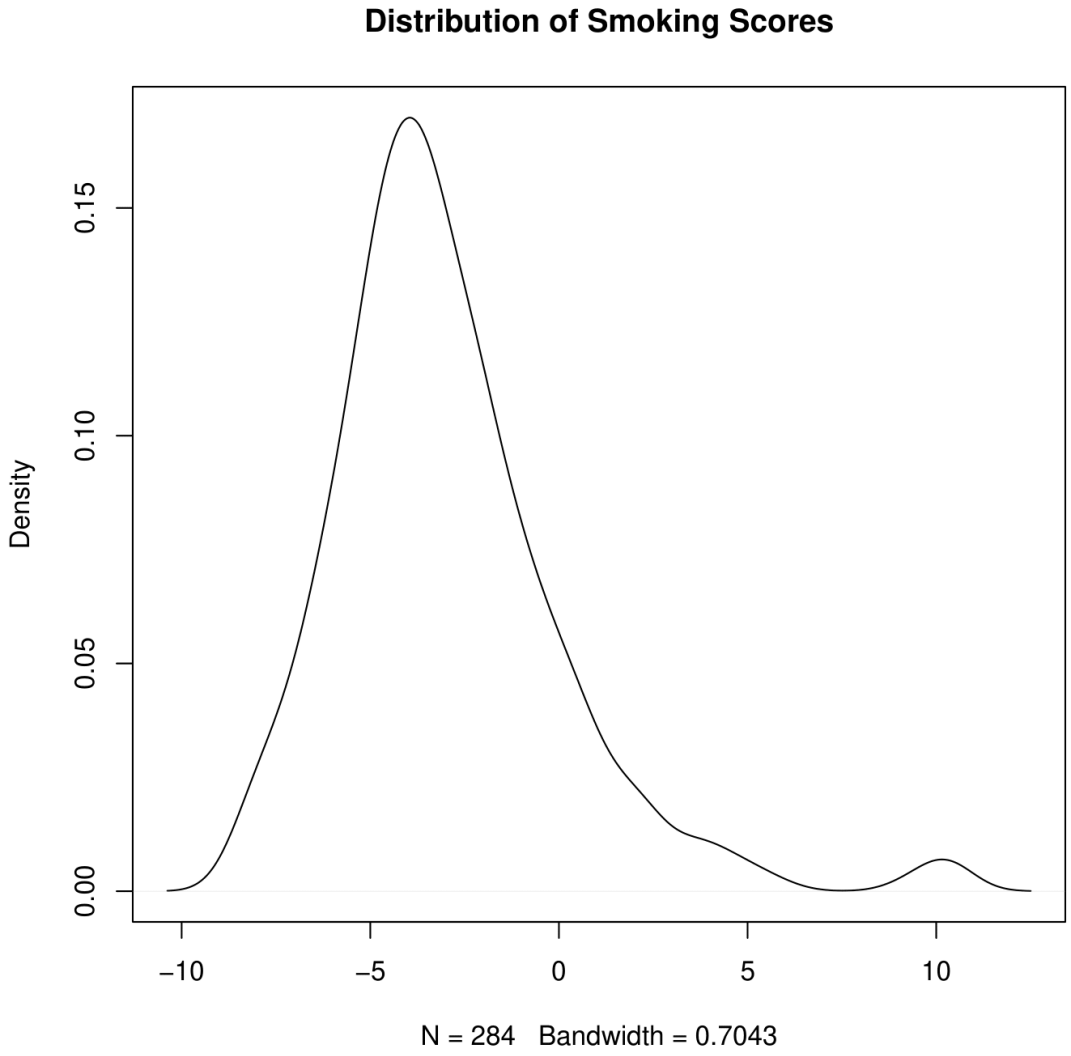

Figure 2.11 Distribution of smoking scores across all 284 baseline samples in AddNeuroMed.

Smoking scores were calculated based on methods described by Elliott et al. (2014). Values toward the left end of the graph correspond to non-smokers.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Standard deviation	6.719	4.875	3.743	2.833	2.591	2.039	1.987	1.827	1.771	1.735
Proportion of Variance	0.112	0.059	0.035	0.020	0.017	0.010	0.010	0.008	0.008	0.007
Cumulative Proportion	0.112	0.171	0.206	0.226	0.243	0.253	0.263	0.271	0.279	0.287

Table 2.2 Principal components identified by PCA in the AddNeuroMed cohort.

PCA was used to identify sources of variance in the data. Shown for the first ten principal components (PC1 - PC10) are the standard deviation, proportion of variance, and the cumulative proportion of variance. As the first two principal components accounted for the arbitrary threshold of more than 5% of variance in the data, these were used to select which covariates would be included in analyses of differential methylation.

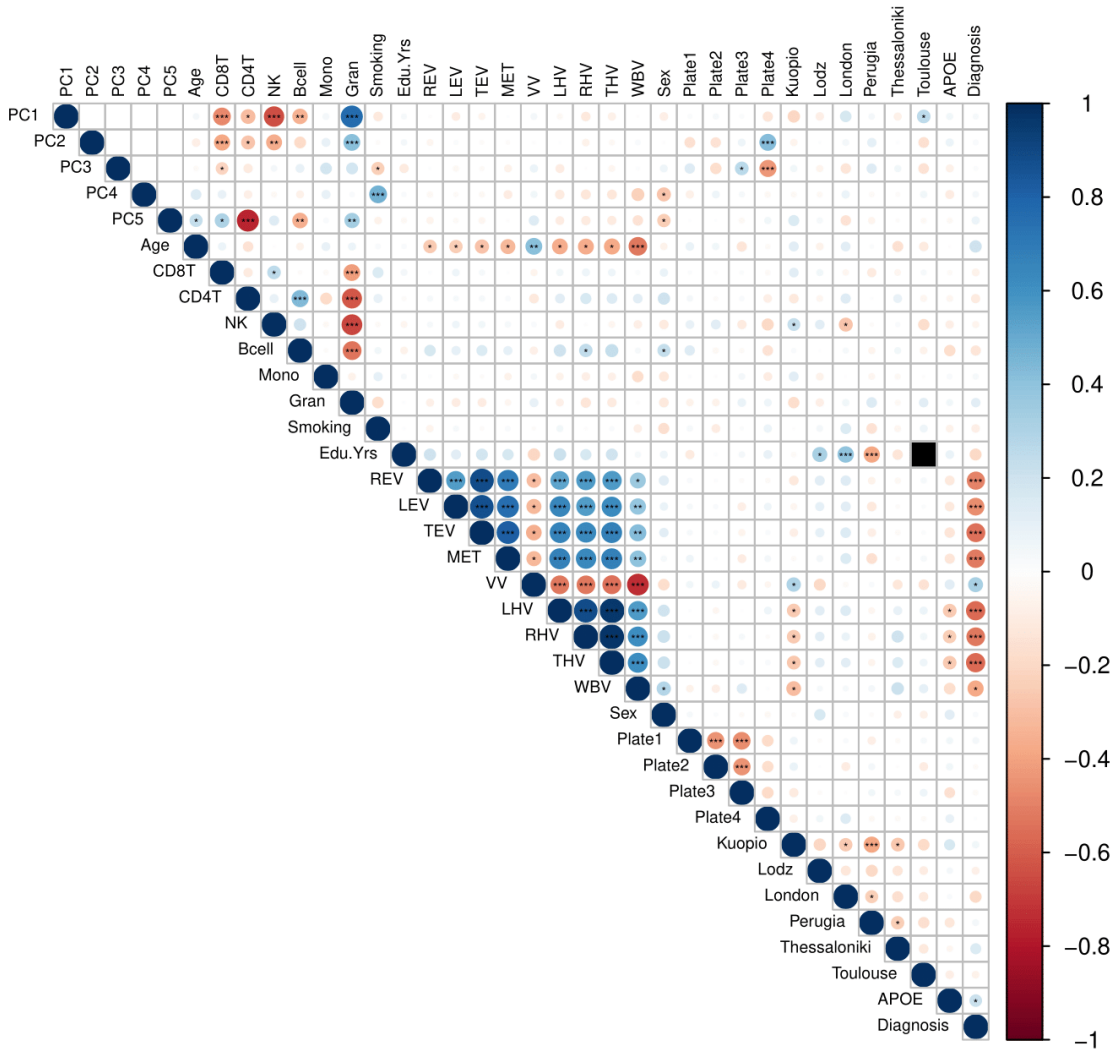

Figure 2.12 Correlation plot of principal components to potential confounders and variables of interest in AddNeuroMed.

*The first five principal components resulting from PCA were correlated to diagnosis as well as all potential confounders: age, cell types (CD8T, CD4T, natural killer (NK), B cells, monocytes (mono), granulocytes (gran)), smoking, number of education years (Edu.Yrs), MRI data (right/left/total entorhinal cortex volume (REV/LEV/TEV), median entorhinal thickness (MET), ventricular volume (VV), right/left/total hippocampal volume (RHV/LHV/THV), whole brain volume (WBV)) , sex, batch (Plate 1 – 4), sample source (Kuopio, Lodz, London, Perugia, Thessaloniki, Toulouse), and the number of APOE ε4 alleles. As cell types and batch correlated strongly with the first two principal components (accounting for > 5% of variance), these were selected as covariates in addition to age and sex. Significance: * = 3.23×10^{-4} (Bonferroni-corrected p-value), ** = 1.0×10^{-8}, *** = 1.0×10^{-10}.*

2.3. The ADNI cohort

The ADNI study was initiated in 2004, in a partnership of private companies and the National Institute on Aging (NIA) as well as the National Institutes of Health (NIH) in the United States of America and Canada. The study, led by Dr. Michael W. Weiner, has the aim of developing biomarkers for the early diagnosis of AD using cross-sectional and longitudinal data from different modalities (http://adni.loni.usc.edu/). The ADNI cohort includes CTL subjects, individuals with MCI, and individuals with AD. The study design, selection of participants and data collection procedures have been described in Petersen *et al.* (2010). Briefly, participants were only included if they were between 55 and 90 years old, showed no signs of depression on the Geriatric Depression Scale (GDS), and were of good general health with no diseases precluding enrolment. The inclusion criteria specific to the CTL group were: no uncommon memory complaints for an individual's age, normal memory functioning as determined by the revised Wechsler Memory Scale (WMS), and MMSE score between 24 and 30, a Clinical Dementia Rating (CDR) of 0, and no significant impairment in cognitive functions or activities of daily living. Participants with MCI were only included if they had memory complaints, showed abnormal memory function on the WMS, scored between 24 and 30 on the MMSE and had a CDR of 0.5, but general cognition and performance did not indicate a diagnosis of AD. AD was diagnosed through the NINCDS/ADRDA criteria for probable AD, with subjects having memory complaints, showing abnormal memory function on the WMS, scoring between 20 and 26 on the MMSE and had a CDR of 0.5 – 1.0.

2.3.1. DNA methylation profiling and sample selection in ADNI

DNA methylation profiling of blood samples was performed on 653 individuals in the ADNI cohort using Illumina EPIC arrays. DNA was collected for DNA methylation profiling at baseline, and at two follow-up points, each one year apart. Samples were randomised with a modified incomplete balanced block design, with all samples from the same subject being placed on the same BeadChip. Spaces on the chip were filled with age- and sex-matched samples. In order to avoid confounding, subjects from different diagnostic groups were placed on the same chip, and any remaining spaces on the chip were filled with replicated DNA samples for technical reproducibility assessment.

As the ADNI cohort was intended as a replication cohort for the analyses performed in Chapters 4 and 5, a subset of samples from the cohort were selected based on several criteria in order to match the cohort demographics of the AddNeuroMed cohort as closely as possible. The overall inclusion criteria were:

1. Only individuals classified as CTL, or MCI, or AD (i.e. no individuals with subjective memory complaints).
2. White/Caucasian ethnicity.
3. Age 65 or older.

As multiple time points were available for participants, the following process was used to select samples to obtain a cross-sectional cohort similar to AddNeuroMed (with CTL, MCI-MCI, MCI-AD, and AD groups). First, samples were divided into those with a consistent classification (i.e. the same diagnostic group at every clinical assessment), and those with an inconsistent classification. For every

subsequent step described below, the following applies: if multiple suitable methylation samples were available for an individual, the methylation sample with the age closest to the average age of the respective diagnostic group in AddNeuroMed was selected. Additionally, individuals were only included in one of the diagnostic groups.

For samples with a consistent classification:

- All samples were allocated to their respective diagnostic groups (CTL, MCI-MCI, and AD).
    - The CTL and MCI-MCI groups exclusively contained samples with a consistent classification.
    - The AD group contained some samples with a consistent classification, with the rest of the group composed of inconsistent classified samples as described below

For samples with an inconsistent classification:

- To generate the rest of the AD group, I selected methylation samples that had at least one earlier diagnosis of AD, and a stable diagnosis of AD after that time point (i.e. no reversion to MCI or CTL at a later time point).
- For the MCI-AD group:
    - Only individuals with methylation profiling performed less than one year prior to conversion to AD were selected.
    - Individuals with a chronological development from CTL to MCI to AD were included, though any individuals with a chronological development from MCI to CTL, or AD to CTL or MCI were excluded.
    - Where two samples were available less than one year before conversion to AD, the earliest sample was selected.

This selection process left a total of 443 samples for the replication analyses of Chapters 3 and 4, of which 140 were CTL, 216 were MCI, and 87 were AD. Within the MCI group the ratio of MCI-AD converters to MCI-MCI individuals were the same in ADNI (38.9%) as in AddNeuroMed (38.8%; Fisher's exact test: $p = 1$). The average age of the CTL group in ADNI ($M = 76.3$, $SD = 5.9$) was higher than the CTL group in AddNeuroMed ($M = 73.8$, $SD = 5.3$; $t\ (227) = 3.32$, $p = .001$). However, the average age between the MCI groups (AddNeuroMed: $M = 75.5$, $SD = 5.5$, ADNI: $M = 75.9$, $SD = 5.9$) did not differ significantly ($t\ (323) = 0.50$, $p = .616$), nor did the average age between the AD groups (AddNeuroMed: $M = 78.4$, $SD = 6.1$, ADNI: $M = 76.8$, $SD = 5.6$, $t\ (171) = 1.85$, $p = .066$). The MMSE scores were also compared between each diagnostic group. No differences between cohorts were found for the CTL groups (AddNeuroMed: $M = 29$, $SD = 1.2$, ADNI: $M = 29.1$, $SD = 1.3$, $t\ (227) = 1.85$, $p = .56$), the MCI groups (AddNeuroMed: $M = 26.9$, $SD = 2$, ADNI: $M = 27.4$, $SD = 2.1$, $t\ (323) = 1.85$, $p = .56$), or the AD groups (AddNeuroMed: $M = 20.8$, $SD = 4.5$, ADNI: $M = 20.1$, $SD = 5$, $t\ (170) = -1.07$, $p = .288$). Finally, the distribution of the number of *APOE* ε4 alleles was compared between cohorts for each diagnostic group. No differences were found in the CTL groups (Fisher's exact test: $p = .244$), the MCI groups (Fisher's exact test: $p = .470$), or the AD groups (Fisher's exact test: $p = .071$).

ADNI				
	CTL	MCI		AD
		MCI-MCI	MCI-AD	
N	140	132	84	87
Sex (M/F)	67/73	81/51	54/30	55/32
Age (Mean ± SD)	76.3 ± 5.9	75.4 ± 5.9	76.6 ± 6	78.4 ± 6.1
MMSE (Mean ± SD)	29.1 ± 1.3	28.1 ± 1.8	26.3 ± 2.1	20.1 ± 5

Table 2.3 ADNI cohort demographics.

Subject characteristics of 443 whole blood DNA methylation samples that all passed data pre-processing. Shown are the number of samples (N), sex (Males/Females), mean age ± standard deviation (SD), and mini-mental state examination score (MMSE) ± SD. Of the 216 MCI subjects, 132 remained stable (MCI-MCI), and 84 progressed to AD within one year after DNA profiling.

2.3.2. ADNI data pre-processing

To keep differences between the ADNI and AddNeuroMed cohorts to a minimum, the raw data from the ADNI cohort was processed in the same manner as described in section 2.2.3, but adapted for EPIC arrays where necessary. Briefly, raw signal intensity data was imported into R as a methylumi object, and β-values were calculated across all 866,895 probes on the EPIC array (Figure 2.13). The median methylated and unmethylated signal intensities were visually inspected, and all samples were deemed acceptable. Although signal intensities are slightly lower on the EPIC array compared to the 450K array, this matches what has been seen previously in other EPIC array data analyses in our group. Signal intensity may vary from experiment to experiment, and samples should only be removed if they clearly deviate from the main cluster of samples on the intensity plot, which was not the case in this dataset (Figure 2.14).

Next, bisulfite conversion efficiency was calculated (see section 2.2.3.1), and all samples passed the minimal conversion threshold of ≥ 80% median methylation in the fully methylated control probes (Figure 2.15). The complete set of samples also passed the threshold of SNP probe correlations <0.95 (section 2.2.3.2), indicating none of the samples were genetically identical (Figure 2.16). Subsequently, cross-hybridising probes and SNP probes were removed from the data based on the lists for the EPIC array provided by McCartney *et al.* (2016). This left 811,909 probes for further processing.

The next QC step tested sample mislabelling by examining whether clinically recorded sex matched sex predicted from the DNA methylation data. However, the multidimensional scaling as detailed in section 2.2.3.4 does not work as well

for EPIC array data as it does for the 450K array data. Therefore, the *getSex* and *plotSex* functions from the *minfi* package were used, which generate an estimation of sex based on median intensity values on the X and Y chromosomes. For all samples, the clinically reported sex matched the sex predicted by the *getSex* function (Figure 2.17). Following the sex check, the *pfilter* function was applied (see section 2.2.3.5) which did not remove any samples, but removed 162 probes with a beadcount < 3 in 5% of samples, and 1,605 probes that had 1% of samples with a detection *p*-value greater than 0.05. A total of 810,140 probes remained. Finally, the *outlyx* function was used to identify outliers (section 2.2.3.6), which were not found in the ADNI dataset (Figure 2.18). The dataset of 443 samples and 810,140 probes was then *dasen*-normalised (Figure 2.19), and the final QC performance checks were calculated as detailed in section 2.2.3.7. Low SEs were reported from the application of the *dmrse* function, indicating a positive performance (SE = 0.0056, SE type I probes = 0.0059, SE type II probes = 0.0049), and the results from the *genki* function indicated low technical variation. The *genki* SE values in the ADNI dataset were: 3.19×10^{-5}, 4.6×10^{-5}, and 1.73×10^{-5} (type I probes: 5.2×10^{-5}, 5.13×10^{-5}, 2.61×10^{-5}, type II probes: 2.18×10^{-5}, 4.32×10^{-5}, 1.17×10^{-5}).

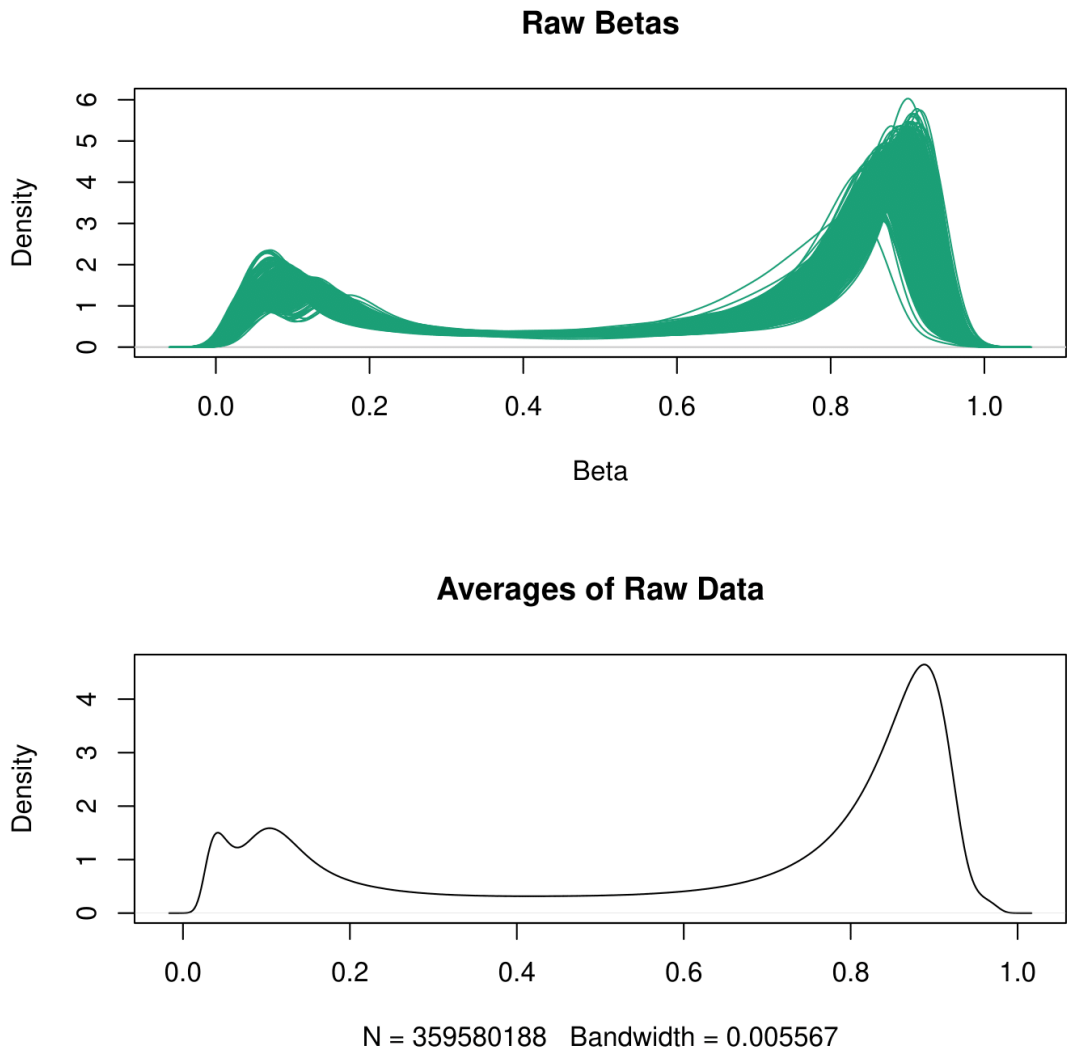

Figure 2.13 Density plots of raw β-values in the ADNI cohort.

Density plots displaying the frequency of DNA methylation values across all individual ADNI blood samples (top), and for the average of all samples (bottom).

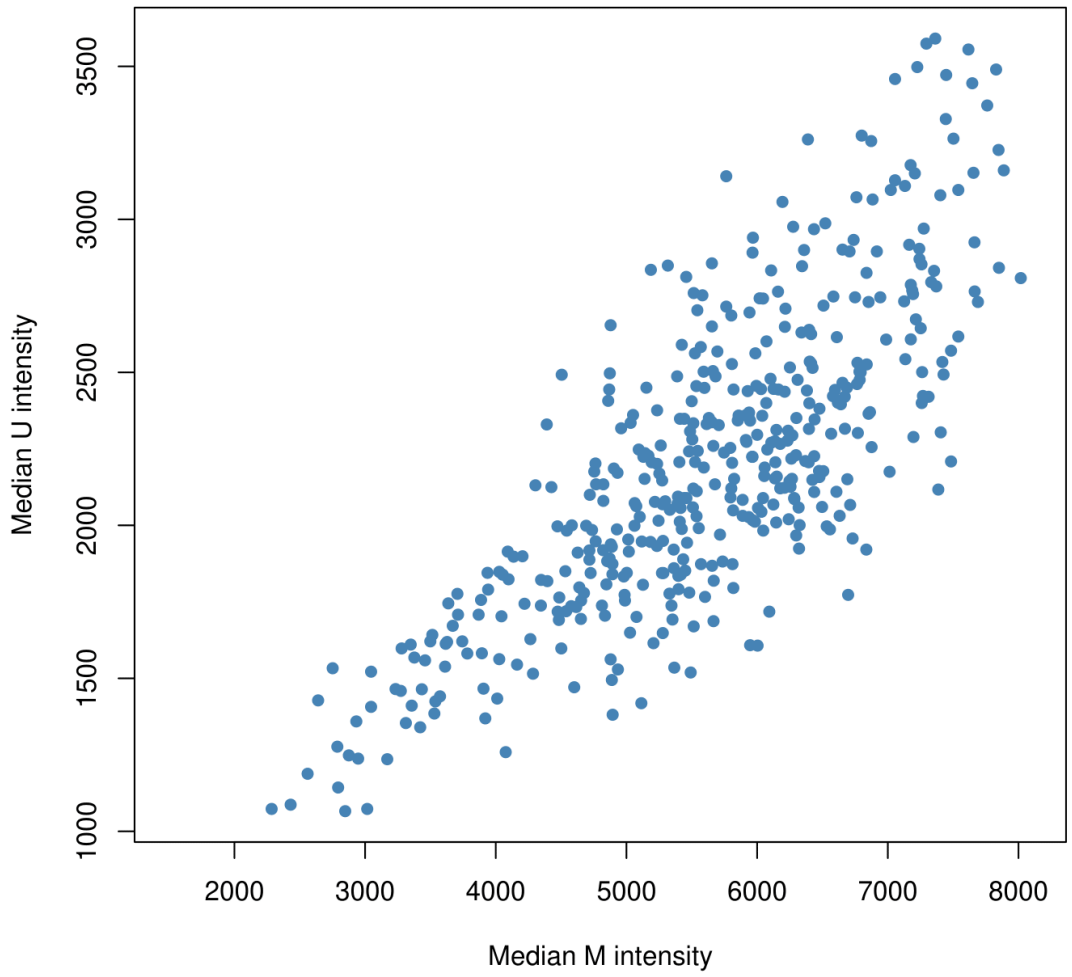

Figure 2.14 Raw median methylated and unmethylated signal intensities in ADNI samples.

None of the samples show extreme deviations from the main cluster of samples, therefore, all samples are taken forward in the QC pipeline.

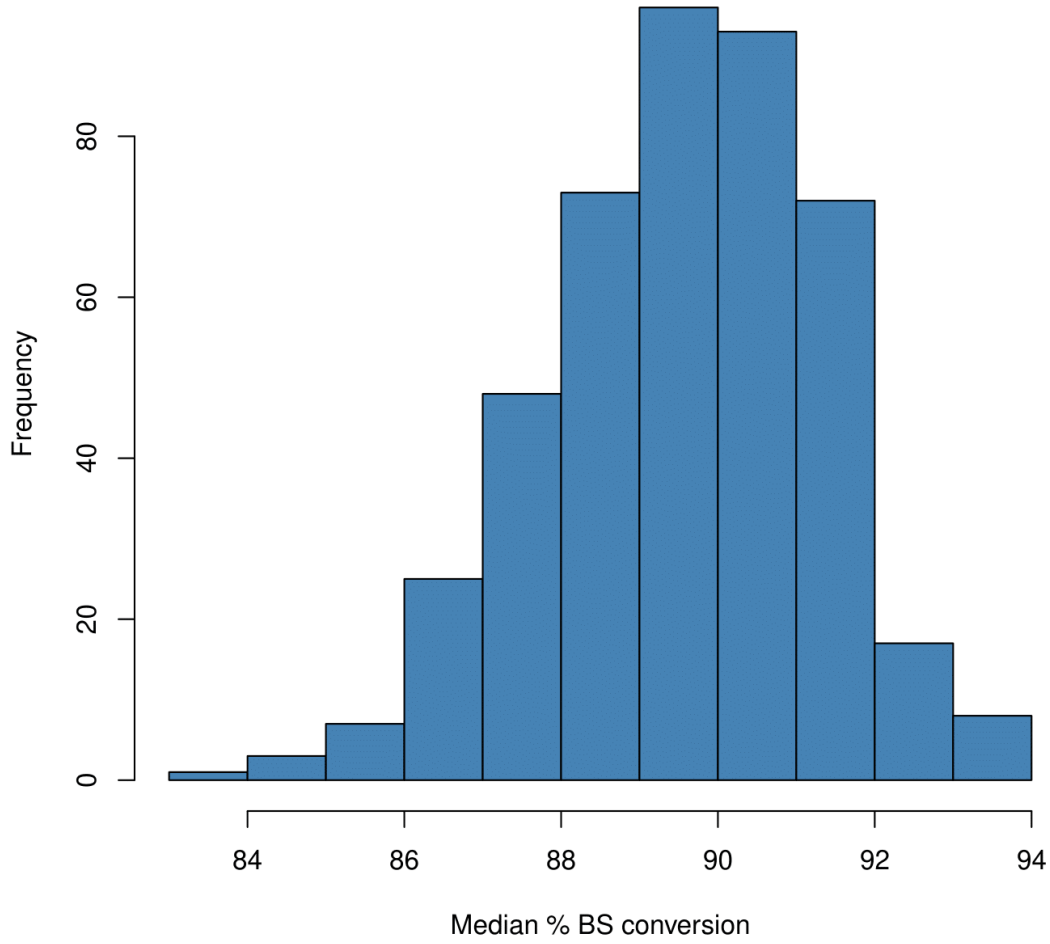

Figure 2.15 Bisulfite conversion efficiency in ADNI samples.

Distribution of median methylation percentage across samples in fully methylated control probes on the EPIC array. All samples passed the minimum median methylation percentage of 80%.

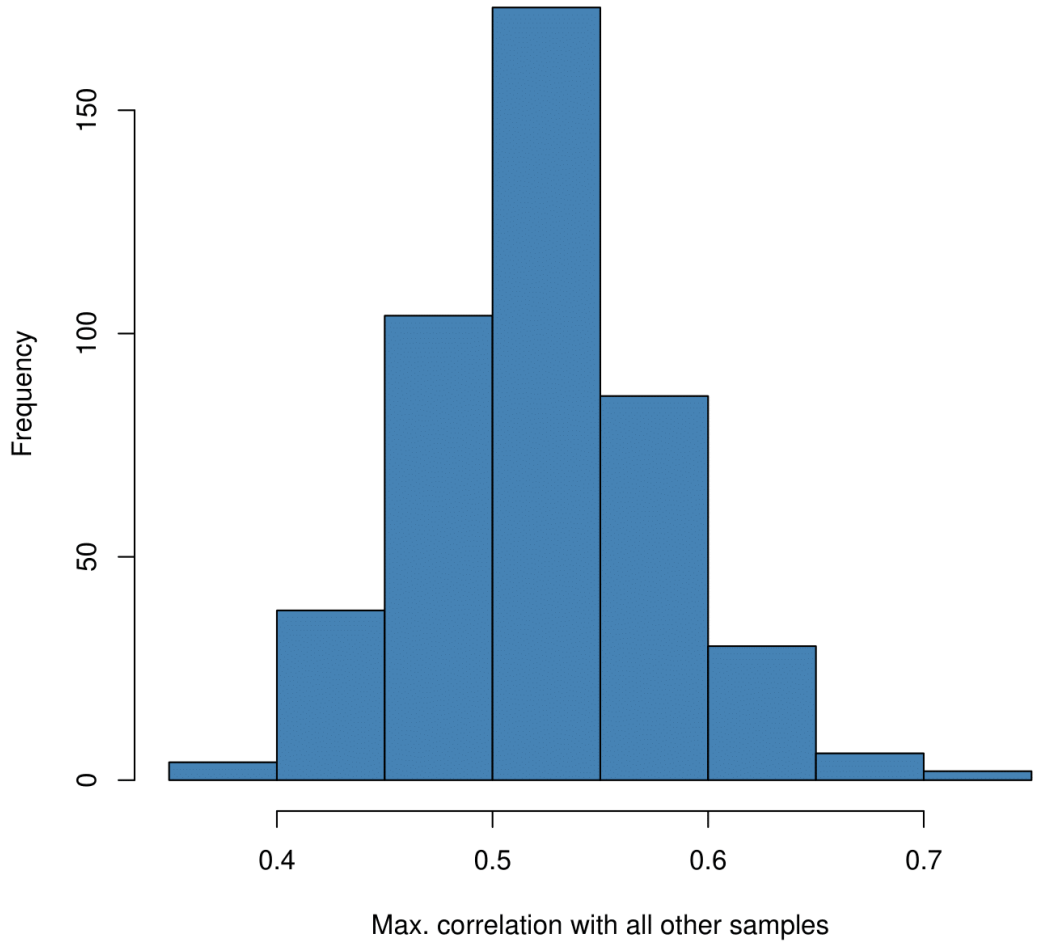

Figure 2.16 Genetic distinctness of samples in the ADNI cohort.

Maximum correlations found between samples for 59 probes interrogating SNP loci on the EPIC array. As none of the samples showed correlations higher than r > 0.95, this indicates that the samples are not genetically identical, as expected.

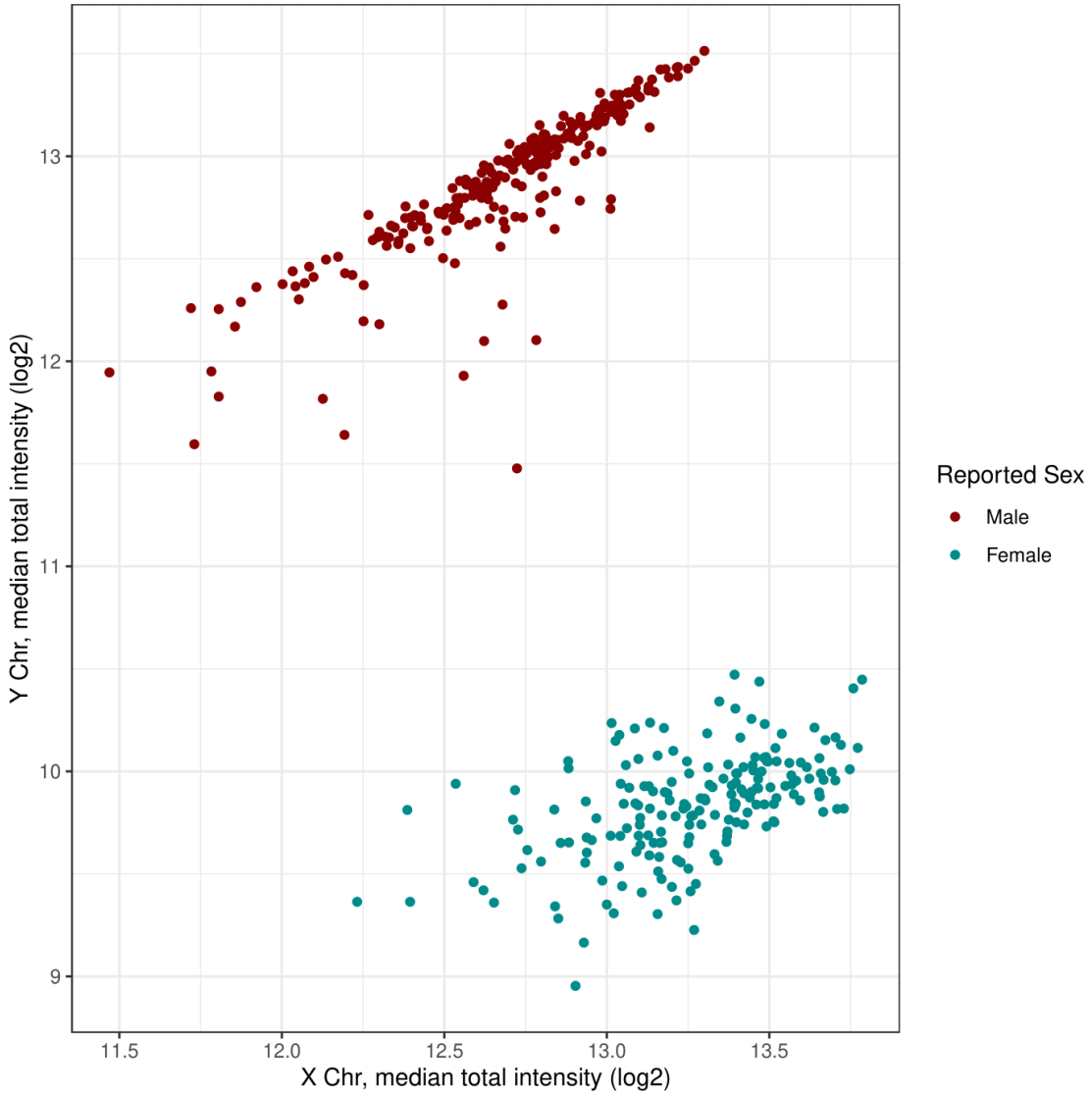

Figure 2.17 Prediction of sex based on the X and Y chromosomes in ADNI.

Median intensity values of probes on the X and Y chromosomes are plotted, and coloured by clinically reported sex for each sample. Profiles that are more similar cluster together, and as expected two clusters are seen which correspond to clinically defined sex.

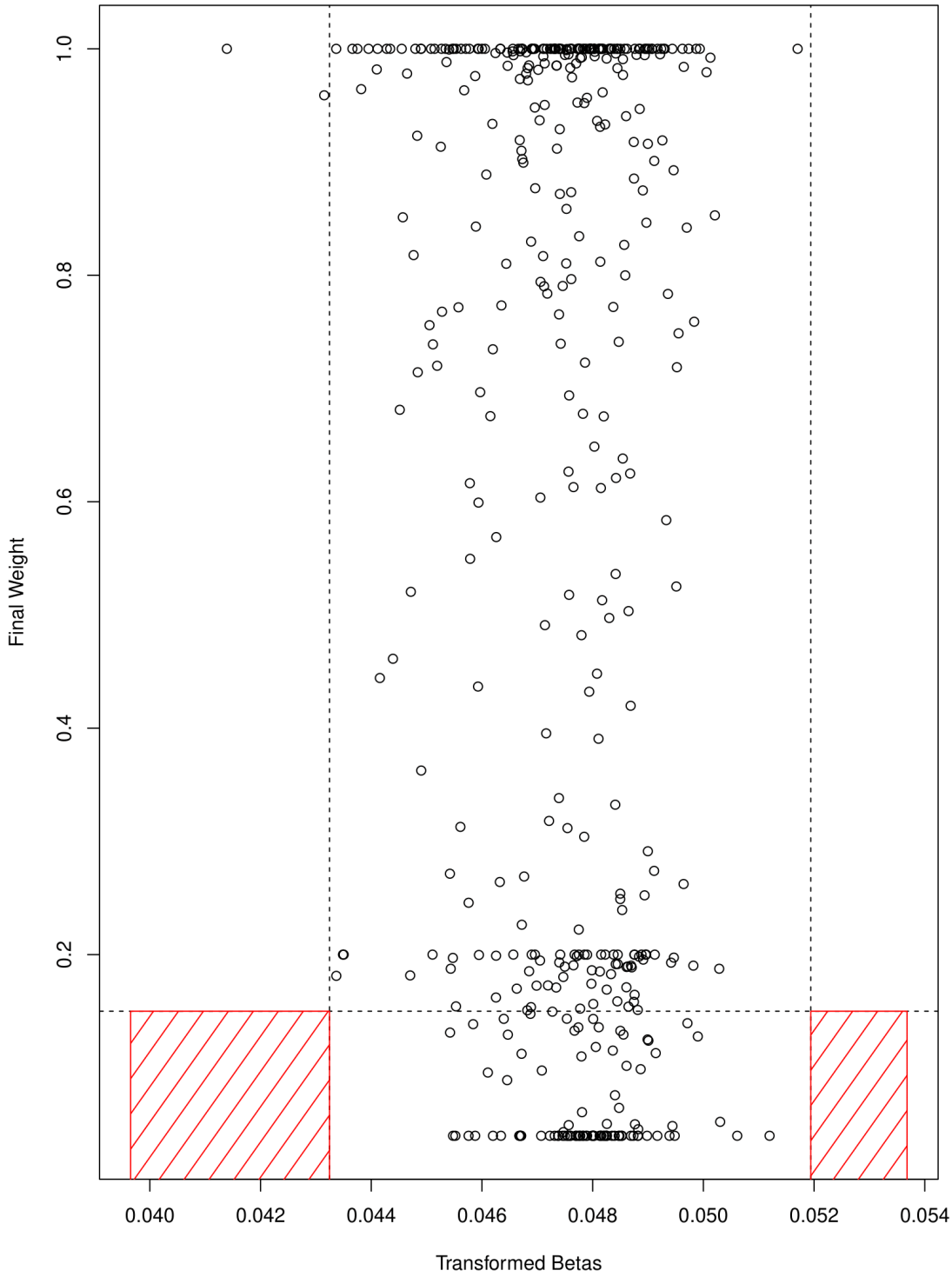

Figure 2.18 Outlyx plot for the identification of outliers in the ADNI cohort.

The outlyx function was applied to the raw (p-filtered) DNA methylation data to detect any outlying samples. Outlier zones based on interquartile ranges are indicated in red, none of the samples were classified as outliers.

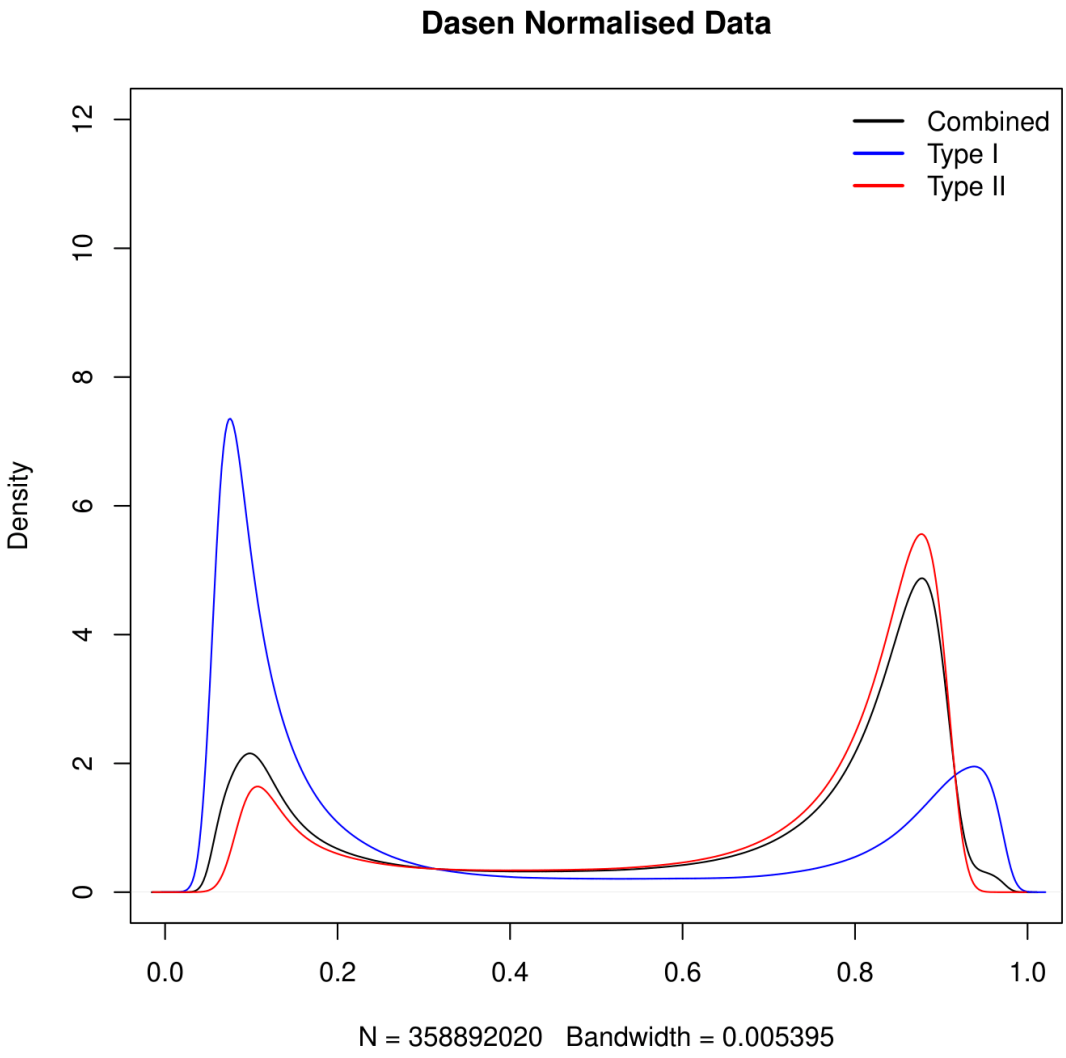

Figure 2.19 β-distribution following dasen normalisation in the ADNI cohort.

Density plots displaying the frequency of dasen-normalised DNA methylation values. The overall distribution is shown in black, and the data separated by type I or type II probes in blue and red, respectively.

2.3.3. Identification of covariates in the ADNI cohort

To identify covariates in the ADNI cohort, a smoking score (Figure 2.20) and cell type proportions were generated based on the DNA methylation data, as detailed in section 2.2.4. No data on bisulfite treatment plates (batch) was available in the ADNI cohort, but median methylated and unmethylated intensity signals calculated per BeadChip were visually inspected (Figure 2.21 and Figure 2.22, respectively).

As in the AddNeuroMed cohort, covariates were identified using PCA performed on the DNA methylation data from the ADNI cohort, as described in section 2.2.4. The first 10 principal components are shown in Table 2.4. As in AddNeuroMed, the threshold of 5% proportion of variance was chosen to select which principal components are of interest. For visualisation purposes, the five largest principal components were correlated to diagnosis as well as potential confounders, which included age, an individual's number of education years, Houseman cell type proportions, smoking score, time to conversion from MCI to AD, sex, source of samples (site), BeadChip ID, and number of *APOE* ε4 alleles (Figure 2.23). In addition to the variables age and sex, the cell type proportion variables were selected as covariates to be used in all differential methylation analysis models as they correlated with the first two principal components.

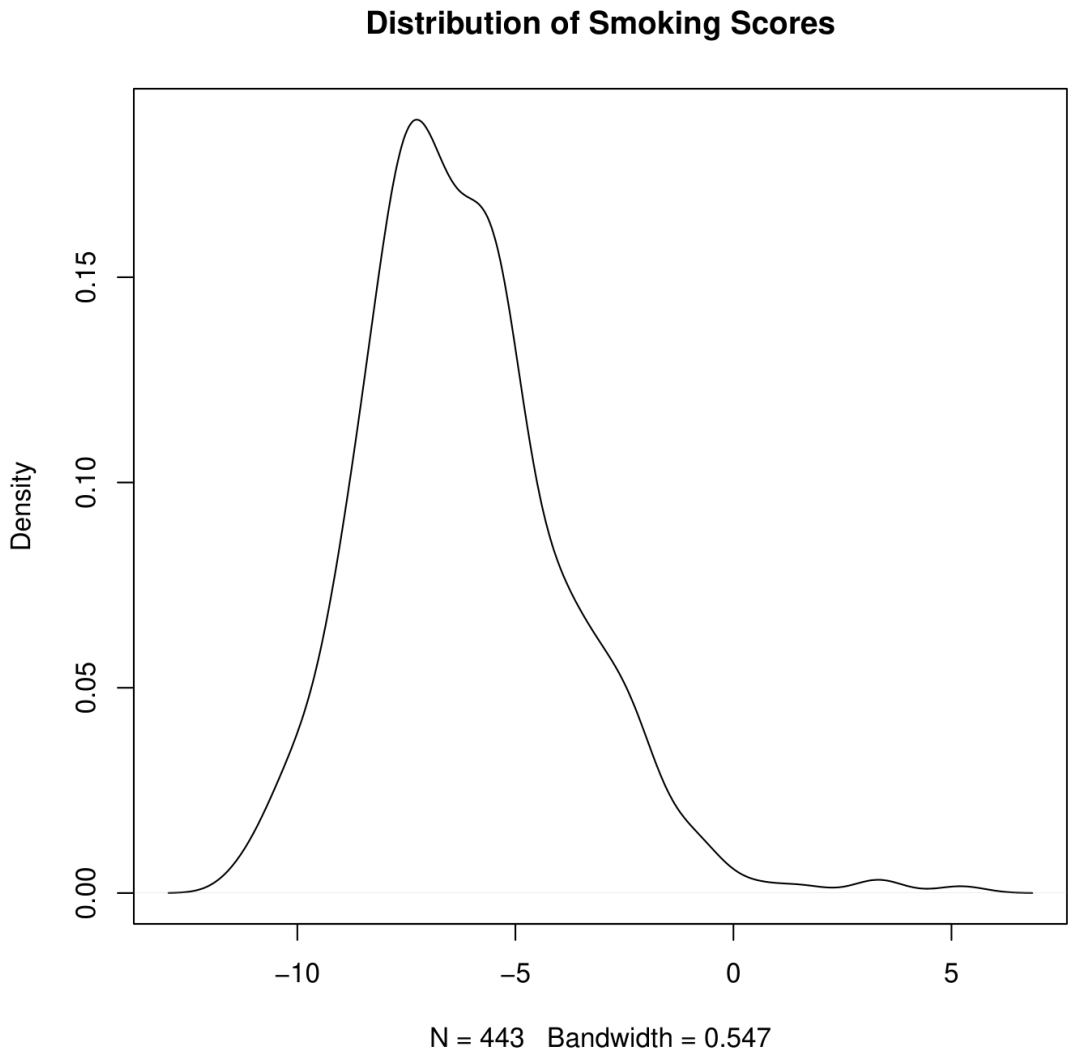

Figure 2.20 Distribution of smoking scores across all 443 ADNI samples.

Smoking scores are calculated based on methods described by Elliott et al. (2014). Values toward the left end of the graph correspond to non-smokers.

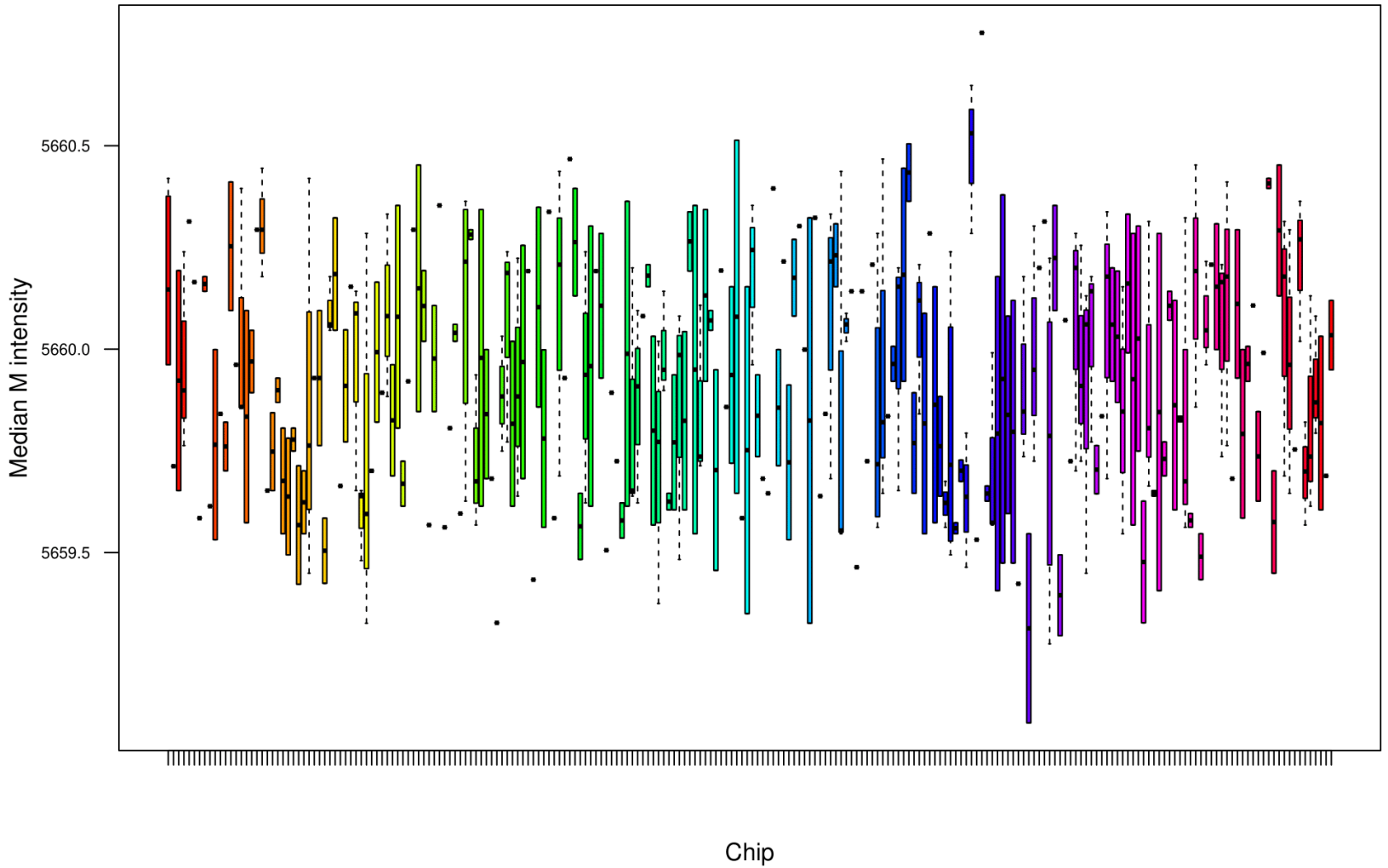

Figure 2.21 Median methylated signal intensities in ADNI by EPIC array chip.

Median methylated (M) signal intensities are plotted and coloured by EPIC array Chip ID.

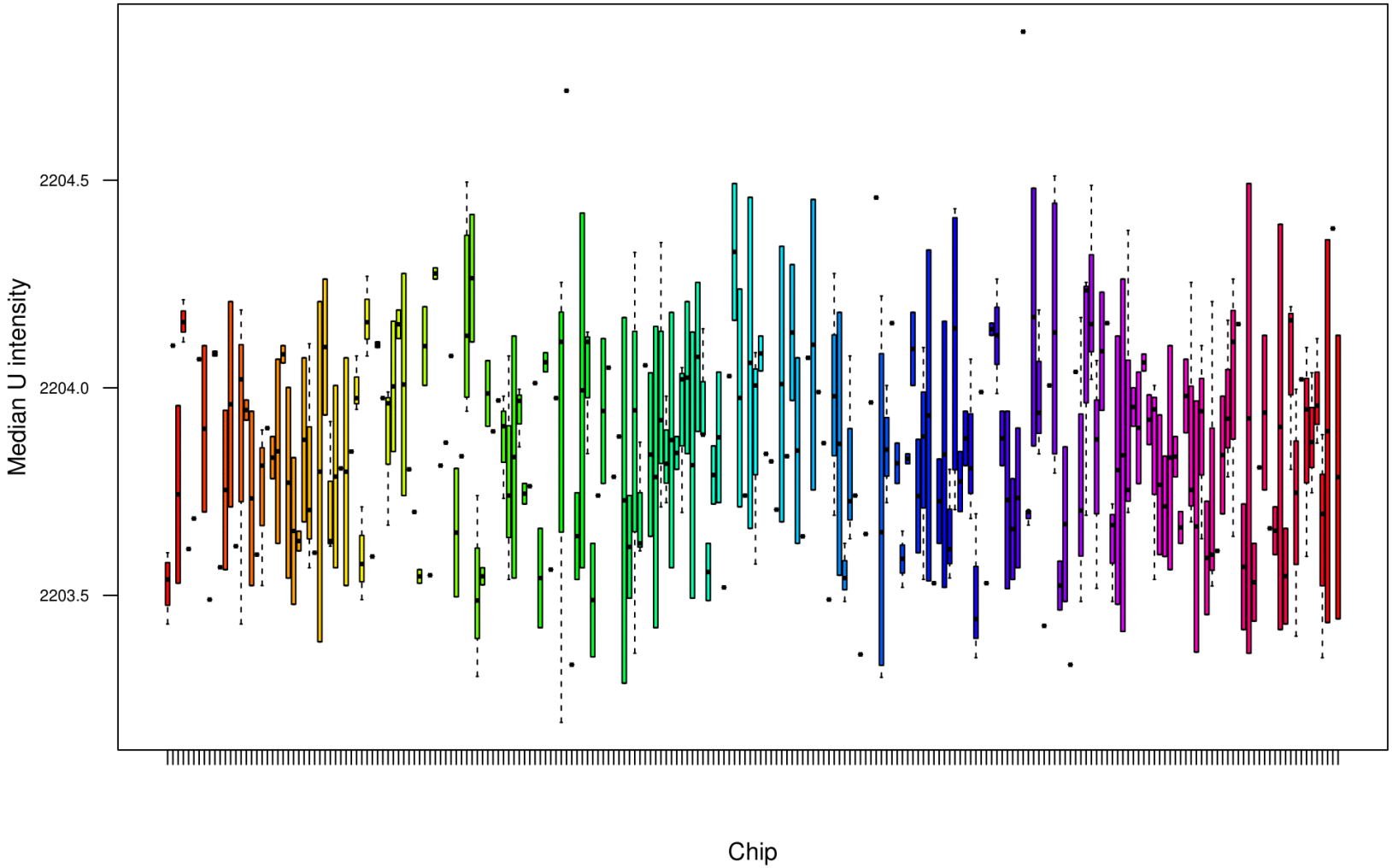

Figure 2.22 Median unmethylated signal intensities in ADNI by EPIC array chip.

Median unmethylated (U) signal intensities are plotted and coloured by EPIC array Chip ID.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Standard deviation	12.512	9.901	5.368	4.967	3.654	3.335	3.171	2.975	2.826	2.702
Proportion of Variance	0.157	0.098	0.029	0.025	0.013	0.011	0.01	0.009	0.008	0.007
Cumulative Proportion	0.157	0.255	0.284	0.309	0.322	0.334	0.344	0.353	0.361	0.368

Table 2.4 Principal components identified by PCA in the ADNI cohort.

PCA was used to identify sources of variance in the data. Shown for the first ten principal components (PC1 - PC10) are the standard deviation, proportion of variance, and the cumulative proportion of variance. As the first two principal components accounted for the arbitrary threshold of more than 5% of variance in the data, these were used to select which covariates would be included in analyses of differential methylation.

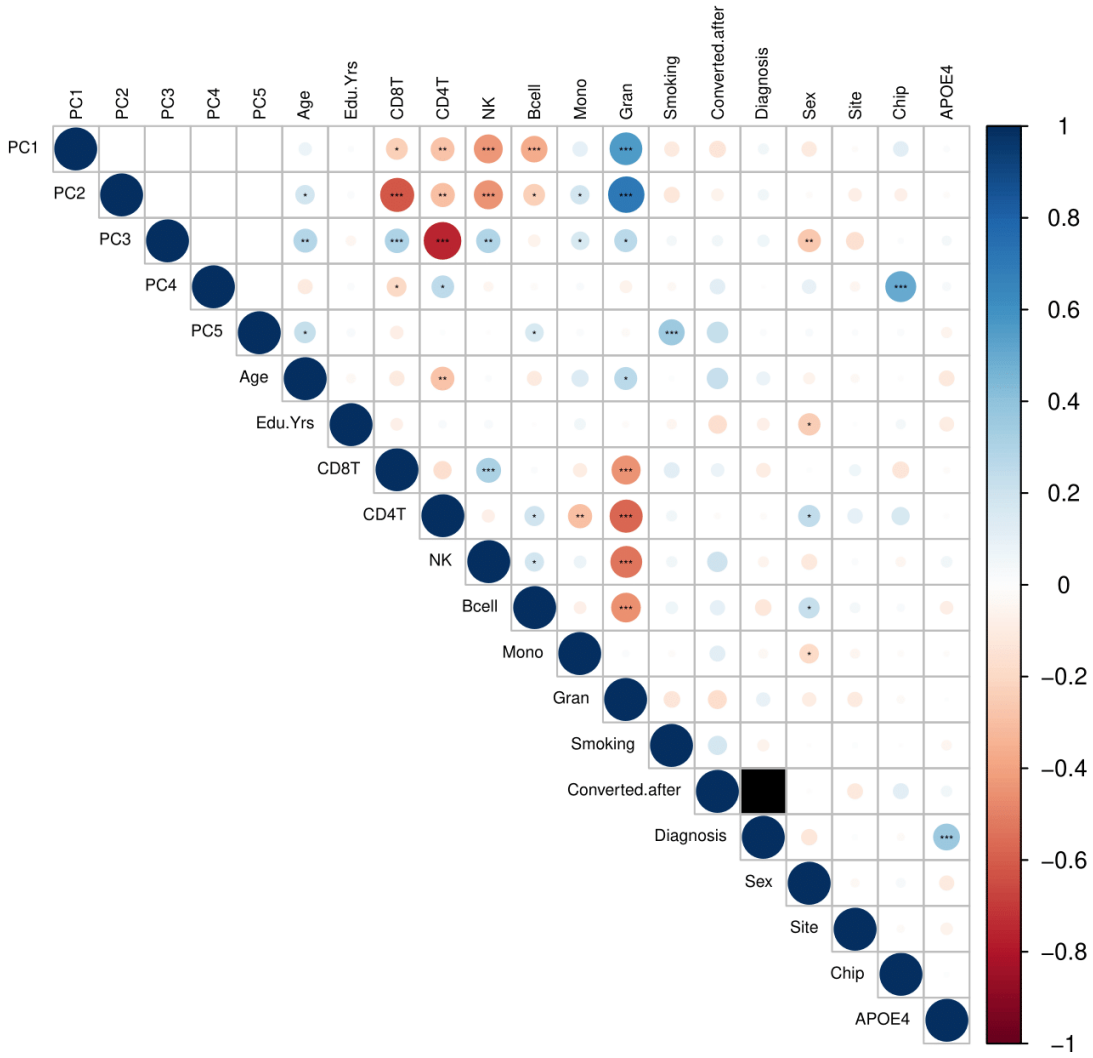

Figure 2.23 Correlation plot of principal components to potential confounders and variables of interest in ADNI.

*The first five principal components resulting from PCA were correlated to diagnosis as well as all potential confounders: age, number of education years (Edu.Yrs), cell types (CD8T, CD4T, natural killer (NK) B cells, monocytes (mono), granulocytes (gran)), smoking score, time to conversion from MCI to AD, sex, source of samples (site), and BeadChip ID. As cell types correlated strongly with the first two principal components (accounting for > 5% of variance), these were selected as covariates in addition to age and sex. Significance: * = 6.67×10^{-4} (Bonferroni-corrected p-value), ** = 1.0×10^{-8}, *** = 1.0×10^{-10}.*

2.4. General data analysis methods

This section describes data analysis methods which apply to multiple chapters in this thesis. Any minor variations to the methods described here are detailed in the relevant chapters.

2.4.1. Differential methylation analysis

Prior to each analysis, covariates as identified in sections 2.2.4 and 2.3.3 were regressed out of the dataset, and the residuals were extracted for subsequent analyses. The exact set of covariates used in each analysis is also specified in the relevant chapters. To identify DMPs within the genome, an analysis of variance (ANOVA) was performed for all analyses examining the baseline diagnostic groups, so as not to assume a linear correlation between CTL, MCI and AD. When examining effects in relation to future progression to AD, a linear regression was applied to compare the two (MCI-MCI, MCI-AD) groups.

To identify DMRs, the python *comb-p* module was run (Pedersen et al., 2012). A 1000bp sliding window was used to identify regions of adjacent, correlated *p*-values, set to a minimum *p*-value threshold of 0.01. Regions which contained only one probe, or regions which did not show a significant *p*-value after multiple testing correction were excluded. The *p*-value resulting from the *comp-b* analysis is calculated following a Stouffer-Liptak-Kechris correction, which adjusts the value for each locus in a weighted manner for the adjacent p-values. A Šidák multiple testing correction is then applied.

2.4.2. Weighted gene correlation network analysis (WGCNA)

WGCNA of the AddNeuroMed data was used in Chapters 3 and 4, to identify biological pathways altered in association with AD and MCI, as well as other variables of interest such as sex, age, or MMSE score.

2.4.2.1. Generation of correlation networks in AddNeuroMed

In order to identify clusters, or 'modules', of highly co-methylated sites in the genome, we made use of the *WGCNA* R package (Langfelder & Horvath, 2008). The hypothesis underlying this method is that genes that highly co-vary, share the same underlying biological processes. Prior to creating the modules, all non-variable probes (variance<median variance) were first removed from the normalised AddNeuroMed data, leaving 200,633 probes for analysis. Samples were then clustered based on their Euclidean distance, and clustering dendrograms were visually inspected to identify outlier samples, which were not detected. Network construction and module detection was then performed in a block-wise manner and constructed irrespective of the direction of correlation between probes (unsigned). In the generation of the network, all probes are connected, though the strength of the connection between two probes is weighted by applying a soft threshold. This threshold suppresses low correlations in a continuous manner, thus emphasising high correlations over low correlations. The soft threshold values were selected using the *pickSoftThreshold* function within the *WGCNA* package. Using this function, the lowest soft thresholding power for which the scale-free topology fit index approached 0.90 was selected. In the generation of modules based on all samples (for Chapters 3 and 4), the soft threshold was set to 9, and in the comparison of MCI converters to MCI non-converters the threshold was set to 8 (Chapter 3, see Figure 2.24). In the resulting

modules, each module is identified by an arbitrarily assigned colour, and the grey module is disregarded from further analyses as it contains unassigned probes. Module eigengenes (MEs) were calculated for each module, as the first principal component across probes assigned to each module. The ME is a single value for each sample and represents the shared methylation profile of the module. Modules were generated twice: once for all samples, to be used in Chapters 3 and 4, and once for the conversion analysis in Chapter 3 using only the subset of MCI-MCI and MCI-AD samples. A total of 16 modules were identified in the full dataset of all samples (Figure 2.25), and 31 modules were identified in the subset of MCI-MCI and MCI-AD samples (Figure 2.26).

Further processing of the modules involved the association of MEs to variables of interest, and subsequent Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis using the *missMethyl* package (Phipson et al., 2016). Details on these methods are provided in the relevant sections in Chapter 3 (section 3.3.4 and 3.3.5) and Chapter 4 (section 4.3.3), as they are analysis-specific.

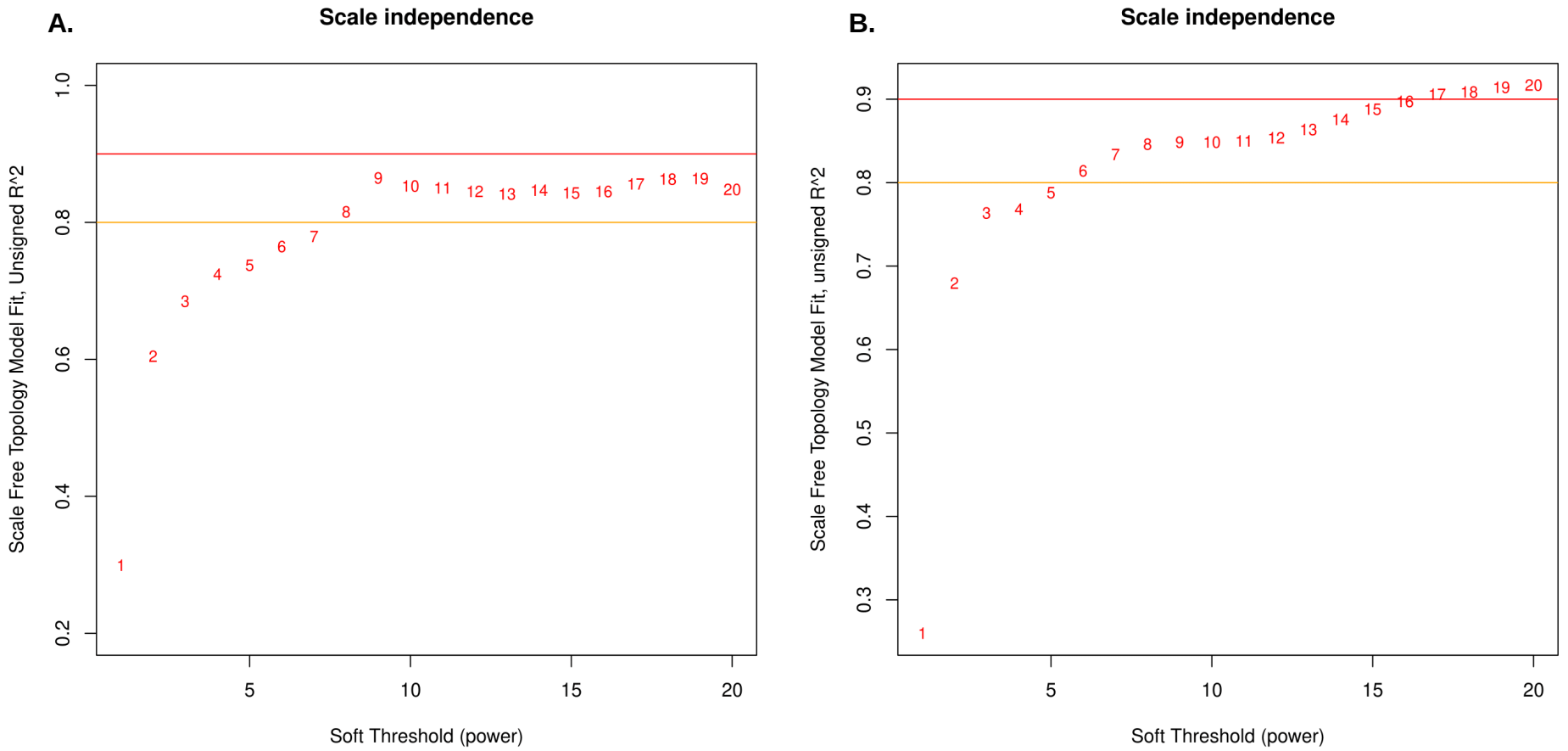

Figure 2.24 Selection of the soft threshold for WGCNA module generation

The function pickSoftThreshold analyses scale free topology for multiple soft thresholding powers. The lowest power for which the scale-free topology fit index curve flattens out upon reaching a high value should be selected, which was 9 for the module generation in all samples (A), and 8 for the module generation in MCI-MCI and MCI-AD samples (B).

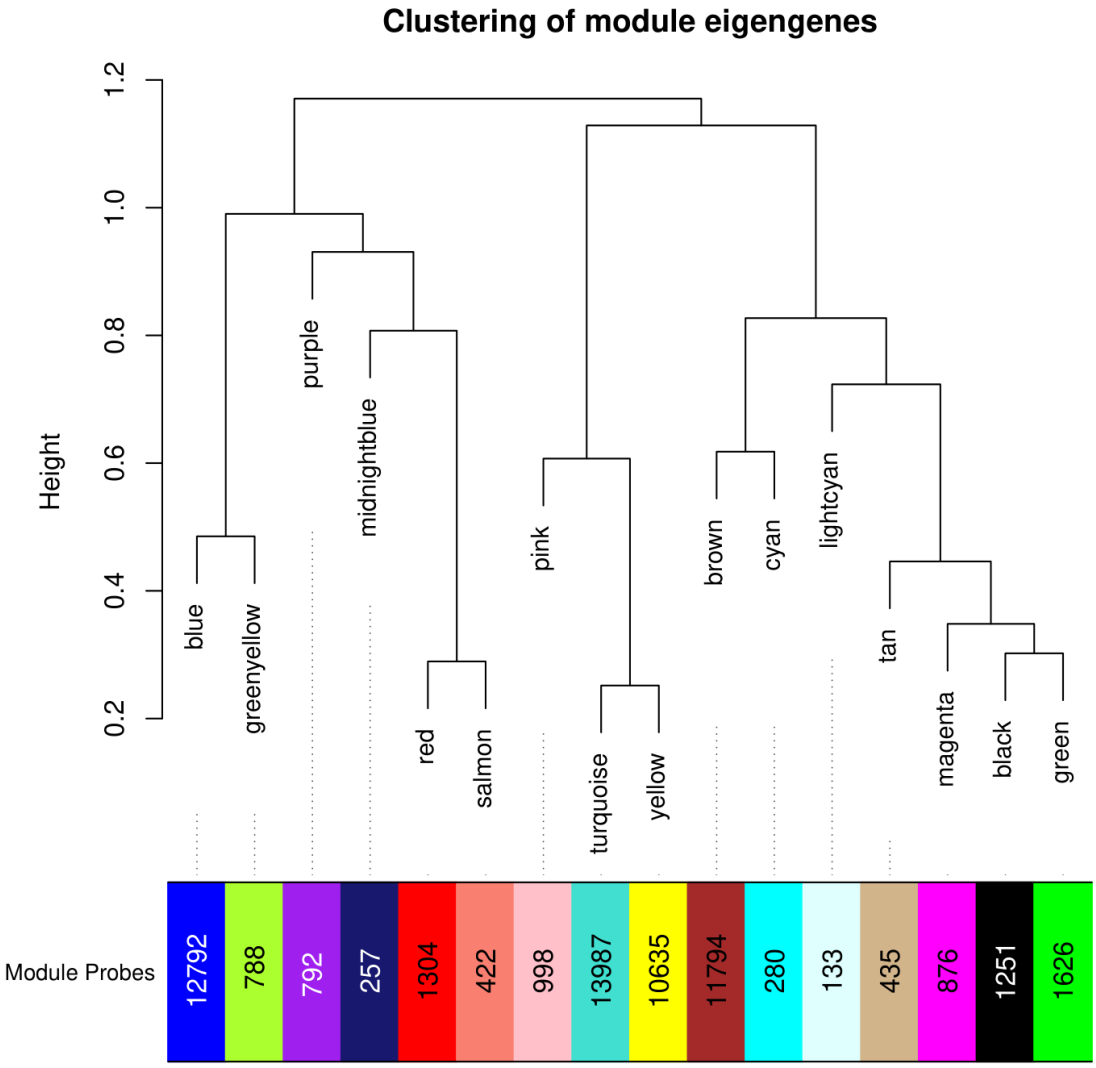

Figure 2.25 Clusters (or 'modules') of highly co-methylated loci identified in the full dataset of 284 samples.

Modules are hierarchically clustered based on calculated module eigengenes (representative of the methylation values within each module), and the number of probes included in each module are indicated along the x-axis. The colour of each module is assigned in an arbitrary manner.

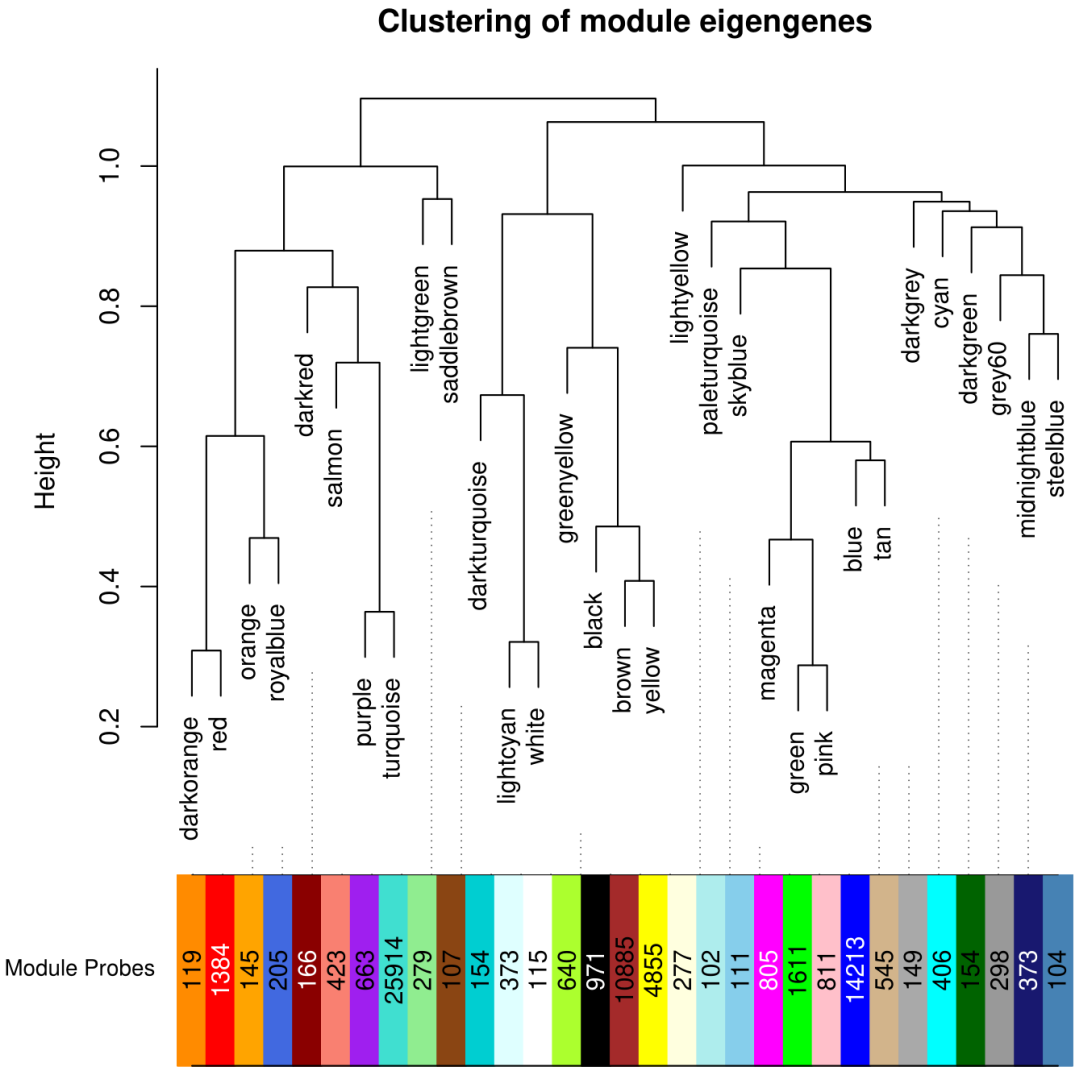

Figure 2.26 Modules identified in the subset of MCI-MCI and MCI-AD subjects.

Modules were generated in the samples of MCI-AD converters within one year of baseline assessment, and stable MCI-MCI individuals. Modules are hierarchically clustered based on calculated module eigengenes, and the number of probes included in each module are indicated along the x-axes. Colours are assigned in an arbitrary manner.

2.5. Pyrosequencing validation

Pyrosequencing was used in Chapter 3 of this thesis, in order to validate DNA methylation changes at single base-pair resolution in the *HOXB6* DMR identified in 450K array data (Chapter 3, section 3.3.7). Pyrosequencing is a "sequencing by synthesis" technique, based on the release of pyrophosphate (PP_i) during the process of DNA elongation (Nyrén, 1987; Nyrén et al., 1993). PP_i is released through the addition of a deoxyribonucleotide triphosphate (dNTP) to a DNA strand, after which it is converted to adenosine triphosphate (ATP) by ATP sulfurylase. In the presence of ATP, a luciferase enzyme then causes the oxidation of luciferin to oxyluciferin, producing a light signal, which is monitored by a pyrosequencer. By performing bisulfite treatment on DNA prior to PCR and pyrosequencing (as described in section 2.2.1), this technique can provide a measure of DNA methylation at the single base level.

A brief description of the methods used is included in Chapter 3; the full protocol is described below. Dr. Adam Smith performed the DNA bisulfite treatment, performed part of the pyrosequencing and assisted at other occasions.

2.5.1. DNA bisulfite treatment

An initial 282 samples were selected from the AddNeuroMed cohort subset for DNA bisulfite treatment using Zymo EZ DNA Methylation-Gold™ Kits (Cambridge Bioscience Cat No.: D5007), according to the manufacturer's instruction (https://files.zymoresearch.com/protocols/_d5007_ez-96_dna_methylation-gold_kit.pdf). A slight modification was made to the last steps of the protocol (0, steps 12 – 15). Instead of adding 30 µL of M-Elution Buffer followed by incubation

and centrifugation, 15 µL was added, followed by incubation, centrifugation, and a repetition of these steps. This had previously been found to give a higher yield of bisulfite-treated DNA.

2.5.1.1. Sample preparation

1. DNA was diluted to 25 ng/µL, and 500 ng of DNA was added to each well of a 96-well PCR plate.

2.5.1.2. Reagent preparation

1. Preparation of the CT conversion reagent (Table 2.5). The reagent is provided in powder form, and must be mixed at room temperature with frequent vortexing or shaking for 15 minutes.
2. M-Wash Buffer preparation (Table 2.6).

Reagent	Volume	Supplier
Ultra-pure H_2O	9 mL	-
M-Dissolving Buffer	500 µL	Zymo
M-Dilution Buffer	3 mL	Zymo
CT Conversion Reagent	Bottle provided in kit	Zymo

Table 2.5 CT conversion reagent.

Reagents as provided by the Zymo EZ DNA Methylation-Gold™ Kits (Cambridge Bioscience Cat No.: D5007), with the exception of Ultra-pure H_2O.

Reagent	Volume	Supplier
M-Wash Buffer	36 mL	Zymo
100% Ethanol	144 mL	-

Table 2.6 M-Wash Buffer reagents.

The M-Wash Buffer is provided by the Zymo EZ DNA Methylation-Gold™ Kits (Cambridge Bioscience Cat No.: D5007).

2.5.1.3. Sodium bisulfite conversion

1. 130 µL of the prepared CT Conversion Reagent was added to each 20 µL DNA sample in each PCR plate well. The samples were mixed by pipetting.
2. The plate was sealed with the film provided in the kit, and transferred to a thermal cycler which was set to run the following steps:
   2.1. 98°C for 10 minutes,
   2.2. 64°C for 2.5 hours,
   2.3. 4°C storage for up to 20 hours.
3. 400 µL of M-Binding Buffer was added to the wells of a Silicon-A™ binding plate mounted on a collection plate.
4. The samples were transferred from the conversion plate (Step 2) to the wells of the Silicon-A™ binding plate, and mixed by pipetting up and down.
5. The plate was centrifuged at 3,500 x *g* for 5 minutes, and the flow-through was discarded.
6. 400 µL of M-Wash Buffer was added to each well, and the plate was centrifuged at 3,500 x *g* for 5 minutes.
7. 200 µL of M-Desulphonation Buffer was added to each well, and the plate was incubated at room temperature (20-30°C) for 20 minutes.
8. The plate was centrifuged at 3,500 x *g* for 5 minutes, and the flow-through was discarded.
9. 400 µL of M-Wash Buffer was added to each well, the plate was centrifuged at 3,500 x *g* for 5 minutes, and the flow-through discarded.
10. Another 400 µL of M-Wash Buffer was added to each well, and the plate was centrifuged at 3,500 x *g* for 10 minutes.
11. The Silicon-A™ binding plate was placed onto an elution plate.
12. 15 µL of M-Elution Buffer was added to each well.

13. The plate was incubated for 5 minutes at room temperature (20-30°C).
14. The plate was centrifuged at 4000 x *g* for 3 minutes to elute the DNA.
15. Steps 12-14 were repeated.

The DNA was stored at or below -20°C for later use. In case of long-term storage, the DNA was stored at or below -70°C.

2.5.2. Pyrosequencing assay design

Biotinylated PCR primers were used to amplify a region of interest within the DNA for pyrosequencing. Additionally, sequencing primers were designed that were specific to the region of interest, and were designed to adhere to the following criteria:

- To ensure sufficient annealing, the sequencing primer was designed to be between 15 and 25 bp long.
- For optimal efficiency in the pyrosequencer, the sequencing primer was designed to have an annealing temperature as close as possible to 40°C.
- The primers did not contain CpG sites to prevent selective attachment to only methylated or unmethylated DNA.
- Primers did not contain large repetitive elements.
- The forward and reverse primers were designed to match in annealing temperature.
- The amplicon size should not greatly exceed 300 bp, and the region of interest to be sequenced did not exceed 100 bp.

The biotinylated primers and sequencing primers were designed using the PyroMark Assay Design software by Qiagen, and the CpG assay designs for the

PyroMark machine were designed using the PyroMark Q24 software from Qiagen, under standard conditions. The primers were designed to cover a region of the *HOXB6* DMR, used in Chapter 3, are displayed in Table 2.7.

Primer	Sequence
HOXB6 Forward	/5Biosg/TTTTTGGTGAGGGGGGAGT
HOXB6 Reverse	CCTACCATCCCTCCCTTATCT
HOXB6 Sequencing	CTCTAACTATTACCCC

Table 2.7 HOXB6 primers for PCR and pyrosequencing.

2.5.2.1. Polymerase chain reaction (PCR)

PCR is used as a method to amplify a specific region of DNA by creating many copies of the same fragment using designed primers. In the PCR, a mixture of DNA, primers targeting the region of interest, free nucleotides, polymerase, and buffers, cycles through a process of heating and cooling, which leads to an amplification of the chosen region. These cycles run as follows:

1. In the first step of the PCR, the mix is heated to a temperature of 95°C for 15 minutes, which activates the heat-sensitive polymerase taq.
2. This is followed by a repetition of three steps, which are repeated in this order a total of 35 times:
    2.1. Denaturation: the mix is heated to 95°C to denature the double-stranded DNA.
    2.2. Annealing: the mix is cooled to 60°C, which allows the primers to anneal to the complimentary sequence within the DNA. This temperature is primer-specific and may vary.
    2.3. Extension: the mix is heated to 72°C which allows the taq polymerase to synthesise the complementary strand of DNA, using the deoxynucleotides (dNTPs) which are present in the mix.
3. A final extension phase at 72°C of 15 minutes is performed after the last cycle of steps 2.1. – 2.3. has finished.

The PCR reagents and optimised cycling conditions are shown in Table 2.8 and Table 2.9, optimisation of the PCR is discussed in 2.5.2.3.

Reagent	Volume	Supplier	Function
10x Buffer B1	3 µL	Solis BioDyne	Solution stabiliser for optimum polymerase reaction
MgCl2 (25mM)	1.98 µL	Solis BioDyne	Cofactor necessary for polymerase
dNTP mix (10mM)	0.3 µL	Fisher Scientific	DNA bases for DNA synthesis
F/R primer mix (10µM)	1.5 µL	IDT	Single-stranded oligonucleotides matching target sequence
HOT FIRE polymerase (5U/µL)	0.3 µL	Solis BioDyne	Catalyses the addition of nucleotides
DNA (25ng/µL)	2 µL		Bisulfite treated DNA
H2O	20.92 µL		Makes total reaction volume 30 µL

Table 2.8 HOXB6 PCR reagents.

Process	Temperature (°C)	Duration	Cycles
Hotstart	95	15 min	1
Denaturation	95	30 sec	
Annealing	60	30 sec	35
Extension	72	1 min	
Final Extension	72	10 min	1
End	15	∞	-

Table 2.9 Optimised PCR cycling conditions.

2.5.2.2. Gel electrophoresis

Agarose gel electrophoresis was used to optimise the efficiency of PCR amplification, verify correct amplification of the region of interest, and assess the presence of primer dimers. Gel electrophoresis was performed as follows:

A 2% gel was created by dissolving 2 g of agarose powder (Sigma Aldrich, cat no.: A9539) in 100 mL 1% tris-borate EDTA (TBE) buffer (Fisher Scientific, cat no.: 10031223). This mixture was heated for approximately two minutes to allow the agarose to fully dissolve. Before the gel had cooled, 10 µL of diluted Syto60 was added for fluorescence (1 µL stock in 99 µL of high purity water (dH_2O), Fisher Scientific, Cat. No.: 10194852), and the solution was mixed gently to prevent formation of bubbles. This solution was then poured into a pre-prepared gel cassette, and combs were added to create wells. The gel was then allowed to cool and set. While the gel was cooling, 5 µL of PCR amplified DNA was added to a PCR plate, and mixed with 2 µL of 6 x Orange G (New England Biolabs; Cat No: B7022S). Once the gel had cooled and set, the combs were removed, and the gel was placed in the gel tank, filled with 1 x TBE buffer.

Each sample was loaded into an individual well, and a 100 base pair ladder (New England Biolabs; Cat No: N3231S) was added to the first well of each row. The gel was then left to run for 30 minutes at 120 V.

After the run had finished, the gel was placed in a Licor Odyssey CLx scanner to visualise and assess PCR product quality. Clear, single bands were indicative of good, high molecular weight DNA. Primer dimers may occur as secondary bands, in the case of pyrosequencing this was only considered to be an issue if the sequencing primers had any overlap with the PCR primers. As this was not the

case, it was certain that the sequencing primers would allow only the sequencing of the region of interest, not the primer dimers.

2.5.2.3. Assay optimisation and quality control

First, bisulfite treated DNA from a select few samples was amplified using the PCR primers in Table 2.7. To select an optimal annealing temperature for these primers, the PCR was run with variable annealing temperatures (Table 2.10), and the results were visually inspected using gel electrophoresis (Figure 2.27). Based on these findings, an annealing temperature of 60°C was chosen for the PCR amplification of all samples.

Prior to pyrosequencing of the *HOXB6* region, correct amplification of the region of interest was assessed using gel electrophoresis. PCRs were performed in three batches (as detailed in Table 2.9), and each PCR run included a methylation positive control and no template (negative) controls (NTC). From each plate, 12 samples were selected (at least one from each column and row), and these, along with the NTCs and positive control, were run on a gel (Figure 2.28).

Reagent	Temperature (°C)	Duration	Cycles
Hotstart	95	15 min	1
Denaturation	95	30 sec	
Annealing	56, 58, 60, 62, 64, 66	30 sec	35
Extension	72	1 min	
Final Extension	72	10 min	1
End	15	∞	-

Table 2.10 PCR cycling conditions for HOXB6 assay optimisation.

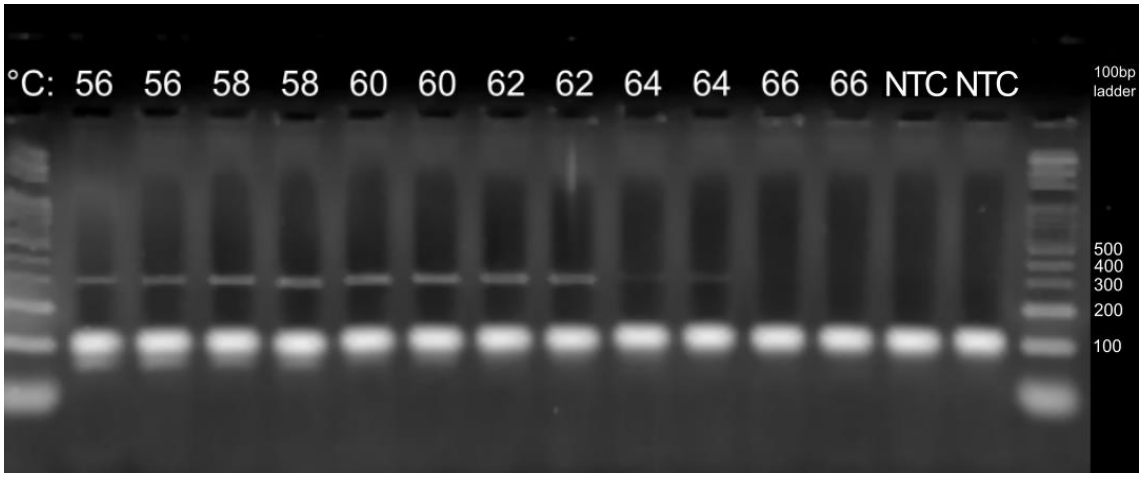

Figure 2.27 Optimisation of the HOXB6 assay.

Gel electrophoresis following a PCR run using a gradient of annealing temperatures (shown in image). A 100 bp ladder is shown on each side, confirming the presence of the desired HOXB6 amplicon of 303 bp. Secondary product or primer dimerisation can be observed at ± 100 bp. However, as the sequencing primer was designed not to overlap the PCR primers, this ensures that only the 303 bp HOXB6 amplicon is sequenced.

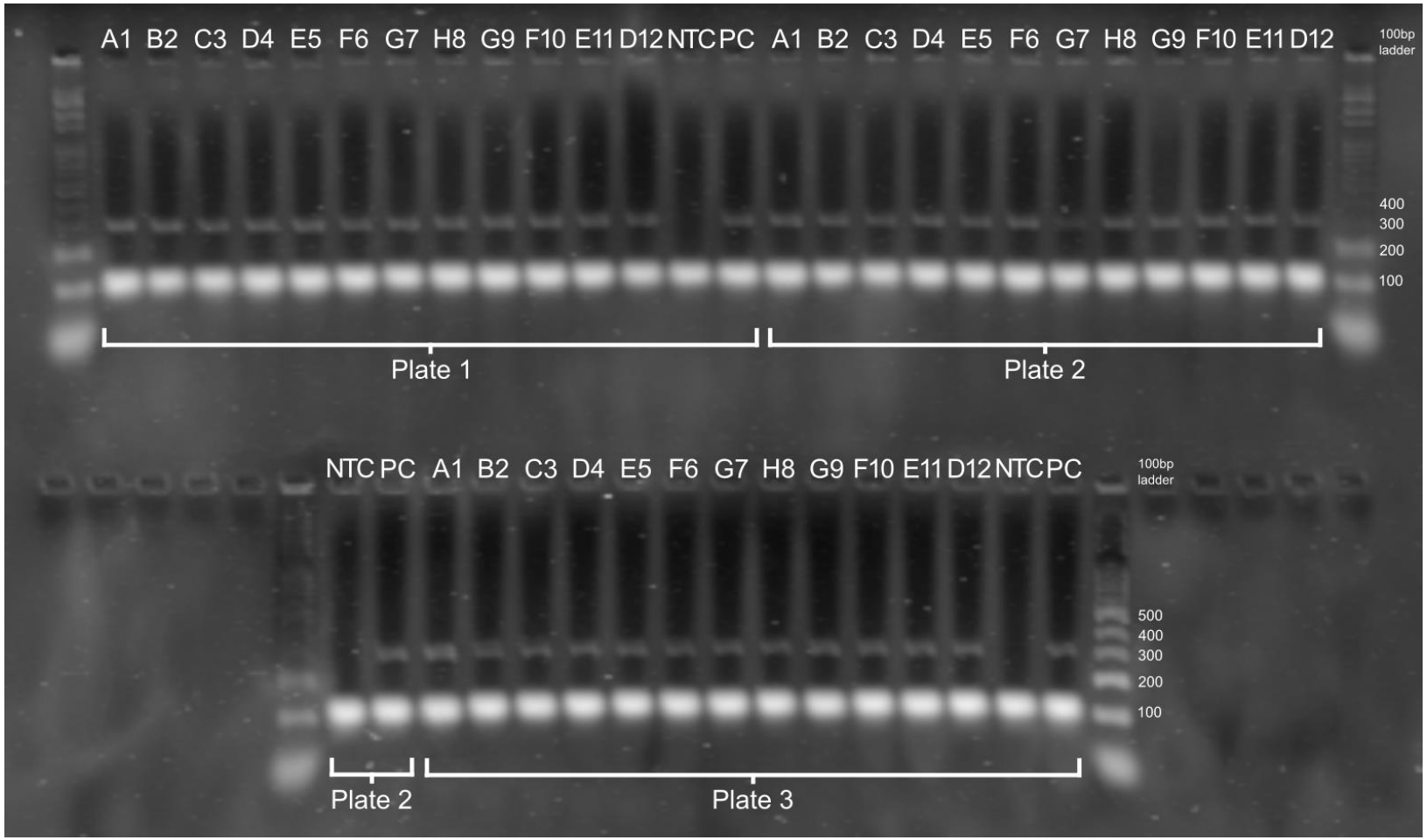

Figure 2.28 Gel electrophoresis quality control of PCR product of the HOXB6 amplicon.

Prior to pyrosequencing of the HOXB6 region, correct amplification of the region of interest (303 bp) was confirmed with gel electrophoresis of one sample from each column, row, and the NTC and positive control (PC), from each plate.

2.5.3. Pyrosequencing

Following *HOXB6* assay optimisation and PCR amplification of the region, pyrosequencing was carried out. The protocol for pyrosequencing is described below, as detailed in: https://www.qiagen.com/gb/resources/resourcedetail?id=59f0275d-e60f-4517-b786-b0e0ca13952e&lang=en.

2.5.3.1. Immobilising the PCR product to beads

Using previously amplified DNA, biotinylated PCR products are immobilised on traptavidin-coated Sepharose beads (Streptavidin Sepharose High Performance, GE Healthcare), as follows:

1. The streptavidin-coated Sepharose beads were gently shaken to re-suspend and obtain a homogenous solution.
2. Per sample, 2 µL of beads was added to a 2.0 mL centrifuge tube, along with 40 µL Binding Buffer (per sample).
3. High-purity water was added to create a total volume of 60 µL per sample.
4. The solution prepared was then added to a 24-well PCR plate, in aliquots of 60 µL.
5. 20 µL of biotinylated PCR product was added to each well of the PCR plate.
6. The PCR plate was sealed and agitated constantly for 5-10 minutes, at 1,400rpm.

2.5.3.2. Preparation of the vacuum workstation:

The troughs in the PyroMark Q24 Vacuum workstation were filled as follows:

- 50 mL ethanol (70%; Sigma Aldrich, cat no.: E7023)
- 40 mL Denaturation Solution (0.2M Sodium Hydroxide (NaOH); Qiagen, cat no.: 979307)
- 50 mL 1x Wash Buffer (Qiagen, cat no.: 979308)
- 50 mL high-purity water
- 70 mL high-purity water

Following this, the vacuum workstation was prepared as follows:

1. With the vacuum filter turned on, the filter probes were lowered into the trough containing 70 mL of high-purity water until all water had been washed through. The filter probes was turned off and returned to the 'parking' position.
2. The trough was then refilled with 70 mL high-purity water.

2.5.3.3. Sequencing primer preparation:

1. The sequencing primer was diluted to 0.3 µM in Annealing Buffer (Qiagen, cat no.: 979309).
2. 25 µL of the diluted sequencing primer was pipetted into each well of a PyroMark Q24 well plate.

2.5.3.4. Combining the PCR products and sequencing primer

1. Once the plate containing the PCR products had been agitated for 5-10 minutes, it was moved to the PyroMark Vacuum Workstation.
2. Immediately following this, the vacuum probe filters were turned on and lowered into the PCR plate in order to capture the beads containing immobilised template.

3. The vacuum tool with the sepharose beads was lowered into the trough containing 70% ethanol and flushed for five seconds.
4. Following this, the vacuum tool was lowered into the trough containing the Denaturing Solution and flushed for five seconds.
5. The vacuum tool was then lowered into the Wash Buffer trough, and flushed for ten seconds.
6. Subsequently, the vacuum tool was raised vertically beyond 90 degrees for 30 seconds, in order to drain all liquid from the filter probes.
7. The vacuum tool was then turned off, and lowered into the PyroMark Q24 plates. The beads were dislodged into the Q24 plates by gently shaking the tool from side to side.
8. The Q24 plate was then heated for 2 minutes at 80°C on a heat block for annealing of the sequencing primer to the samples. Afterwards the plate was left to cool down for 5 minutes at room temperature.

2.5.3.5. Setting up the pyrosequencer

1. The PyroMark Q24 Cartridge was filled with the enzyme and substrate mixtures, as well as the nucleotides, in aliquots described in the pre-run information from the PyroMark Q24 software (Table 2.11 Volumes of PyroMark Gold Q24 reagents for the PyroMark Q24 Cartridge.Table 2.11).

2. The Q24 plate and cartridge were loaded into the sequencer, the sequencing programme was loaded from an external drive, and run.

Reagent	Volume
Enzyme mix	119 µL
Substrate mix	119 µL
G	77 µL
C	105 µL
T	108 µL
A	133 µL

Table 2.11 Volumes of PyroMark Gold Q24 reagents for the PyroMark Q24 Cartridge.

2.5.4. Pyrosequencing data analysis

QC of the pyrosequencing data was performed using the Pyromark Q24 software, in addition to a visual inspection of the data and signal intensities. Methylation percentages at specific CpG sites were then calculated by the software and exported to the R statistical environment. Subsequently, an ANOVA was performed to examine the association of DNA methylation with baseline diagnosis (in the same way that it was performed for the array data as detailed in 2.4.1), controlling for the covariates of age, sex, cell type proportion and batch effects. Pearson correlations were performed between methylation values as determined by pyrosequencing and methylation values for the same CpG locus as measured with the 450K array.

CHAPTER 3. AN EPIGENOME-WIDE ASSOCIATION STUDY OF ALZHEIMER'S DISEASE BLOOD HIGHLIGHTS ROBUST DNA HYPERMETHYLATION IN THE *HOXB6* GENE

This chapter contains work published in Neurobiology of Aging (Roubroeks et al., 2020). The published article has been added to Appendix B of this thesis.

3.1. Introduction

With an increasingly ageing population the prevalence of dementia is expected to almost double in the coming 20 years, with AD being the greatest contributor. AD presents itself as a heterogeneous, multifaceted disease, and this complexity is reflected in the challenges researchers face in elucidating the exact mechanisms underlying this disorder. A number of GWAS have identified susceptibility loci associated with the more common, sporadic form of AD (Lambert et al., 2013). However, these do not account fully for the disease risk, and the exact processes involved in the development and progression of this neurodegenerative disorder remain unknown.

A growing number of studies have investigated the role of epigenetic mechanisms in the aetiology and progression of AD. Recently, numerous EWAS have explored DNA methylomic variation in post-mortem human brain tissue from AD patients and elderly CTL and have highlighted a number of loci that show robust differences in DNA methylation in the cortex across independent cohorts (see Chapter 1.3; Altuna et al., 2019; De Jager et al., 2014; Gasparoni et al., 2018; Lardenoije et al., 2019; Lunnon et al., 2014; A. R. Smith et al., 2019; R. G. Smith et al., 2018, 2020; Watson et al., 2016). In recent years, several studies have identified transcriptomic (Booij et al., 2011; Fehlbaum-Beurdeley et al., 2012; Lunnon et al., 2012, 2013, 2017; Rye et al., 2011) or proteomic (Hye et al., 2006; O'Bryant et al., 2010, 2011, 2016) alterations in the blood early in the disease

and these signatures have been utilised for identifying novel dysfunctional pathways and biomarkers in blood. Although valuable, the use of systemic gene expression or protein markers for this purpose still yields some pitfalls due to the dynamic nature of gene and protein expression. For example, sampling methods can significantly alter expression levels by inducing *ex vivo* mRNA expression (Asare et al., 2008; Thach et al., 2003). Similarly, differences in processing methods between researchers (e.g., handling methods, sample processing method) can affect the quality of mRNA and protein and impinge on downstream analyses (Vartanian et al., 2009; X. Zhao et al., 2012). DNA methylation levels are reported to be more stable than mRNA levels (Paziewska et al., 2014), and as such, studying this in AD blood could be more informative of important biological pathways specifically altered in disease. To date, most blood DNA methylation studies have focused on specifically investigating target genes (da Silva et al., 2014; Furuya, da Silva, Payão, Bertolucci, et al., 2012; Furuya, da Silva, Payão, Rasmussen, et al., 2012; S. C. Wang et al., 2008). Four EWAS of AD blood have been published so far, which have identified a number of disease-associated loci. However, these studies used a limited set of (non-demented) samples and/or did not include any individuals with MCI (Kobayashi et al., 2016; Lardenoije et al., 2019; Lunnon et al., 2014; Madrid et al., 2018). One EWAS of CTL, MCI, and AD subjects was published around the same time as the current study, and was conducted using samples from the ADNI cohort, though this study did not investigate progression to AD (Vasanthakumar et al., 2020).

In order to understand changes in the blood related to the development and progression of AD, it is important to include MCI individuals in addition to AD patients and CTLs. Often viewed as an early stage of AD, MCI is characterised

by memory and other cognitive complaints and impairments. In contrast to the characteristics of AD, the cognitive impairments that individuals with MCI suffer from do not have a significant impact on daily living (Ronald C. Petersen et al., 1999). Although individuals with MCI may remain stable over time or develop another neurodegenerative disorder, MCI subjects, particularly those with aMCI, are at an increased risk of progressing to AD (Jicha et al., 2006). In these cases, the early clinical symptoms reflect the underlying pathological changes related to AD that occur years before the disease fully manifests (J. Hardy, 1997; Jack et al., 2010). Previous research has shown that disease-related changes in gene expression can be detected in peripheral blood from individuals with MCI and AD, with results indicating that some peripheral differences in AD can be detected in MCI subjects (Lunnon et al., 2012). Taken together, the identification of blood methylation patterns related to MCI and AD is of great interest, as it may increase our understanding of peripheral, as well as central changes that occur early in the disease.

In the current study, we have generated genome-wide DNA methylation data from well-characterised CTL, MCI and AD subjects with detailed demographic, clinical, neuroimaging and transcriptomic data previously collected. We have used this dataset to identify differentially methylated loci and epigenetic differences in specific biological processes in blood, which are associated with disease status, or future progression from MCI to AD.

3.2. Aims

The aims of this chapter are as follows:

1. To identify DMPs and DMRs associated with AD, MCI, and progression to AD in blood in the AddNeuroMed cohort;
2. To validate the AD-associated DMR (*HOXB6)* using pyrosequencing in the AddNeuroMed cohort;
3. To investigate correlations between DMRs and the gene expression profiles of the associated genes in the AddNeuroMed cohort;
4. To identify biological pathways related to AD, MCI, and progression to AD in the AddNeuroMed cohort.

3.3. Materials and methods

3.3.1. Subjects

We analysed a subset of 284 blood samples selected for our study from the larger AddNeuroMed cohort, as described in Chapter 2, section 2.1. Briefly, subjects were classified into one of three groups according to their status at baseline sample collection: AD (*n* = 86), MCI (*n* = 109), or elderly CTL (*n* = 89). A subset of MCI subjects progressed to AD within one year of the baseline measurement (MCI-AD, *n* = 38), while others remained stable (MCI-MCI, *n* = 67). A subset of four MCI subjects converted to AD at an unknown time after baseline collection and were excluded from any conversion analysis.

In addition to DNA methylation data, MRI data was collected for 213 individuals as described previously (Simmons et al., 2011). An overview of individuals included in this study can be found in Table 2.1.

3.3.2. DNA methylation analysis

QC and normalisation of the AddNeuroMed DNA methylation data was performed as described in Chapter 2, section 2.2.3. Before identifying DMPs associated with baseline diagnosis, the effect of specific covariates was regressed out. Relevant covariates were identified as described in Chapter 2, section 2.2.4. The covariates included in the regression were age, sex, blood cell type proportion (CD4+ and CD8+ T lymphocytes, natural killer cells, B cells, monocytes, granulocytes) and bisulfite conversion batch. An ANOVA and subsequent post-hoc Tukey's Honest Significant Difference (HSD) test were then performed on the

residuals from the linear regression, to test for DNA methylation differences across all three groups (see section 2.4.1). DMRs were identified using the *comb-p* module in Python, according to the settings described in section 2.4.1.

To identify DMPs and DMRs relating to the future conversion from MCI to AD we performed an analysis comparing the MCI-MCI subjects and the MCI-AD subjects, by first regressing out age, sex, blood cell type proportion, batch and baseline MMSE score. Baseline MMSE score was included as a co-variate as we observed a small, but significant difference in baseline MMSE between the MCI-MCI and MCI-AD groups (Welch t = 2.61, $p = 5.63\times10^{-3}$). We then used a linear regression to compare the two groups and performed *comb-p* analysis as described in section 2.4.1. Quantile-quantile (Q-Q) plots of the *p*-values from both the ANOVA (baseline analysis) and linear regression (conversion analysis) are shown in Figure 3.1.

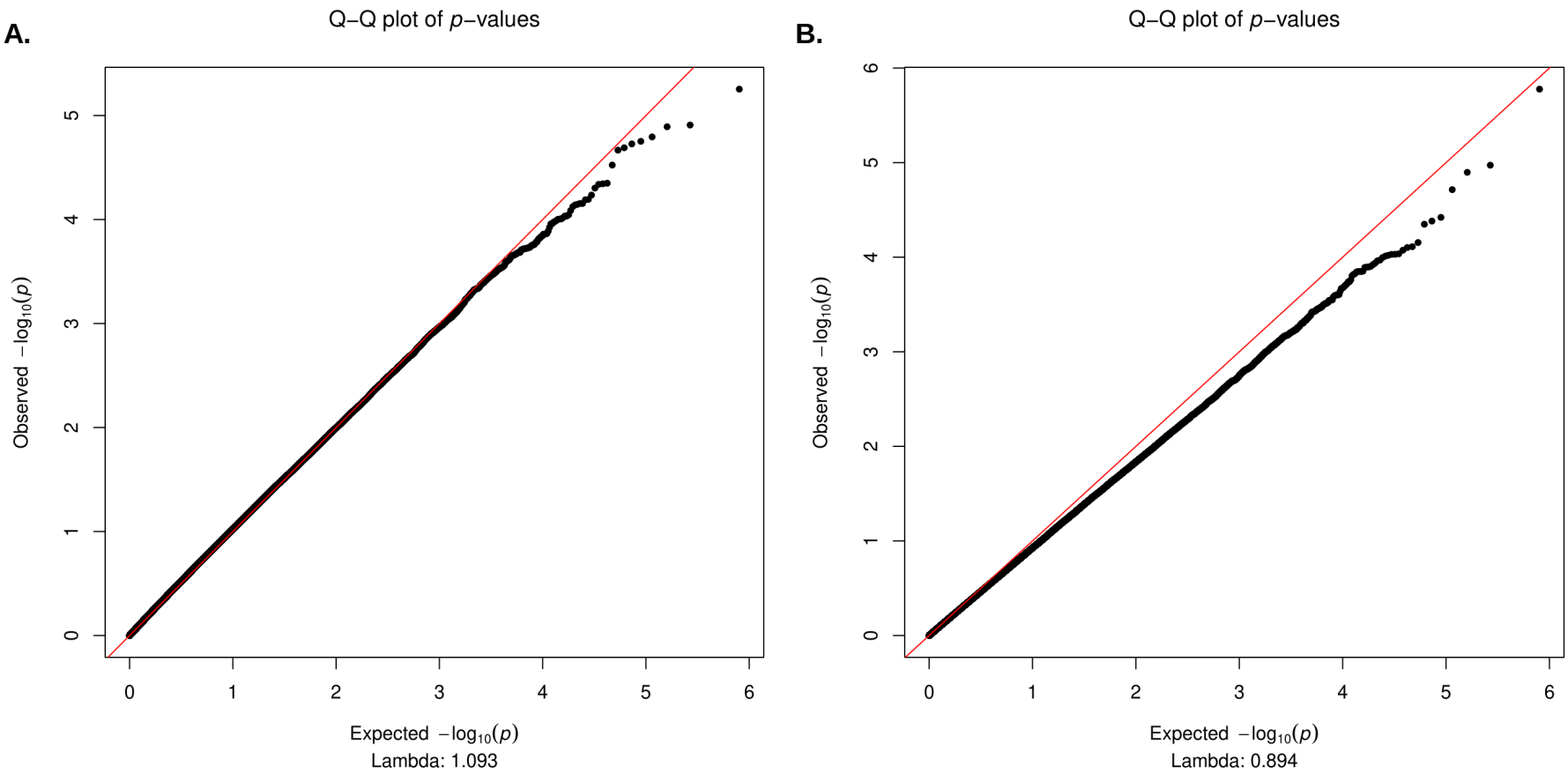

Figure 3.1 Q-Q plots of p-values from the baseline diagnosis analysis and the analysis of conversion to AD.

Q-Q plot of p-values from the ANOVA comparing methylation between CTL, MCI, and AD samples (A), and p-values from the linear regression analysis comparing stable MCI-MCI to MCI-AD who converted within one year of baseline assessment (B).

3.3.3. Generation of weighted gene correlation networks

Clusters of highly co-methylated sites, i.e. modules, were identified using WGCNA as described in Chapter 2.4. Briefly, each generated module is linked to an arbitrarily assigned colour, and within each module MEs were calculated as representative values for the shared DNA methylation profile within a module. Modules were generated twice: once for the baseline group analysis, which compared CTL to MCI to AD, and once for the conversion analysis using only the subset of MCI-MCI and MCI-AD samples.

3.3.4. Association of modules to traits of interest

Following module generation, the relationship between modules and variables of interest was explored. In the baseline analysis, covariates (age, sex, blood cell type proportions, and batch number) were regressed out from the MEs, and extreme outliers (exceeding > 5 standard deviations) were removed. Modules were then associated with baseline diagnosis groups and traits of interest by performing pairwise Pearson or Spearman correlations for continuous or ordinal variables, respectively. Correlations were performed using dummy variables of baseline diagnosis categories to investigate all permutations of comparisons (CTL vs. MCI, CTL vs. AD, and MCI vs. AD), with the group not used in each comparison set to NA. Additional traits of interest included number of education years, number of *APOE ε4* alleles, MMSE score, and the following structural MRI measurements: left, right, and total entorhinal cortex volume (LEV, REV, and TEV, respectively), median entorhinal thickness (MET), left, right, and total hippocampal volume (LHV, RHV, and THV, respectively), ventricular volume (VV), and whole brain volume (WBV). Similarly, regression of the same covariates (with the addition of baseline MMSE score) and outlier removal was

also performed for MEs generated from the MCI-MCI and MCI-AD samples. The residuals from this regression were then used to run a linear regression, comparing non-converters to converters.

3.3.5. Module membership and probe significance

For each of the modules showing significant ($p < 0.05$) associations with one of the three baseline groups, conversion to AD, or traits of interest, we calculated the module membership (MM) and probe significance (PS). MM was calculated as the Pearson correlation between the methylation value of each probe and the ME values, representing the strength of association between a probe and the module it belongs to. PS represents the strength of the correlation between a probe's methylation value and the diagnosis or trait of interest, as performed by Pearson correlations for continuous traits, and Spearman correlations for ordinal traits or diagnosis groups. We correlated and plotted MM to PS for modules of interest and focused on those that showed significant positive correlations (i.e., $r > 0$, $p < 0.05$), which would indicate that probes more integral to the module are mainly driving the association with the trait of interest. Underlying biological processes and pathways were then examined for the modules selected, using GO and KEGG pathway enrichment analyses. For modules containing a large number of probes (> 10,000), we performed these pathway analyses on the probes that were central to the module (i.e., core probes). We set this threshold at 15%, thus selecting the top 15% of probes with the highest MM. Analyses were performed using the *missMethyl* package (Phipson et al., 2016), taking into account the differing number of probes covering each gene on the array.

3.3.6. Analysis of gene expression data and association with methylation data

Normalised gene expression data from Illumina Human HT-12 v3 Expression BeadChip arrays (HT-12 arrays) was obtained from a previous study by Lunnon *et al.* (2012), for 237 individuals included in the current study. Expression data for all genes containing DMRs identified in this study were extracted for analysis if available. For genes nominated from the diagnostic category analysis (CTL, MCI, AD), the covariates of age, sex, and cell type proportions (estimated using Houseman's reference based method) were regressed out of the expression data. An ANOVA and subsequent Tukey's HSD were then performed on the residuals of the regression to identify diagnostic category differences in expression levels of genes containing DMRs. For genes which contained DMRs associated with the progression to AD, only the MCI-MCI and MCI-AD samples were analysed, with age, sex, cell type proportions, and baseline MMSE score regressed out of the expression data, with a subsequent linear regression analysis performed to assess gene expression differences between MCI-MCI and MCI-AD individuals in DMR genes.

Next, methylation values within a DMR were correlated to gene expression values of an annotated gene. Methylation values which had previously been corrected for covariates (i.e., residuals) were extracted for 450K array probes located within each DMR based on genomic location of the DMRs. Pairwise Pearson correlations were then performed between the covariate-adjusted gene expression levels and covariate-adjusted methylation values, for individual 450K probes within a DMR. We also performed correlations of gene expression and mean methylation levels from all 450K probes in the DMR. To determine whether

the association between gene expression and methylation differed between CTL and individuals with MCI or AD, ANOVAs were performed on gene expression levels which included an interaction term between methylation and baseline diagnosis groups (i.e., expression ~ methylation*group). This was performed on the probe most significantly associated with disease for each DMR, and the mean methylation value in the DMR. Similar analyses were performed on the MCI-MCI and MCI-AD subset of individuals, for DMRs associated with progression to AD.

3.3.7. Validation of the *HOXB6* DMR using pyrosequencing

For the purpose of validating our findings, we designed a pyrosequencing assay to quantify DNA methylation at the most significant sites (cg17179862 and cg03803541) within the *HOXB6* region (chr17:46681111-46682414), which was shown to be a DMR in AD relative to CTL. In addition to the two CpG sites the assay was designed for, a further three CpG sites that were not assessed on the 450K array were also covered. The full methods for the pyrosequencing validation carried out for this chapter are detailed in Chapter 2, section 2.5. Briefly, out of the original 284 samples, 264 were used for pyrosequencing. Samples were semi-randomly selected, keeping the group sample number ratios as equal as possible, and randomly distributing samples across plates. A single amplicon of 303 base pairs was amplified using designed primers, and tested for specificity (forward primer = TTTTTGGTGAGGGGGGAGT, reverse primer = CCTACCATCCCTCCCTTATCT, sequencing primer = CTCTAACTATTACCCC). The level of DNA methylation was then quantified using the Pyromark Q24 system (Qiagen), following the standard protocol as provided by the manufacturer and the Pyro Q24 CpG 2.0.6.20 software.

Pyrosequencing data QC was performed using the Pyromark Q24 software, in addition to a visual inspection of the data and signal intensities, with all 264 samples passing QC (CTL: *n* = 83, MCI: *n* = 102, AD: *n* = 79). DNA methylation percentages at specific CpG sites were calculated by the software and exported to the R statistical environment. Subsequently, an ANOVA was performed for each CpG site covered by the assay, as well as the average methylation value across the region. This analysis was identical to the analysis performed on the 450K data, and the covariates of age, sex, cell type proportion and batch were included.

3.4. Results

3.4.1. Identification of differentially methylated loci in MCI and AD blood

The cohort characteristics are shown in Table 2.1. We first investigated whether any individual loci showed DNA methylation differences in either MCI or AD relative to CTL using an ANOVA model after adjusting for the covariates of age, sex, cell proportions and batch (Table 3.1). No DMPs reached the experiment-wide significance threshold that has been established for the 450K array (2.4×10^{-7}; Saffari et al., 2018) with the smallest ANOVA *p*-value being 5.58×10^{-6} for probe cg26146855, of which the closest transcription start site (TSS) according to the GREAT annotation is located in the *TMEM184A* gene. The top 10 most significant probes resulting from the post-hoc Tukey's HSD tests comparing CTL to MCI, MCI to AD and CTL to AD can be found in Table 3.2, Table 3.3, and Table 3.4, respectively. In addition to comparing methylation levels at baseline between the three groups, we were also interested in identifying differences within the MCI population that were predictive of later progression to AD. For this purpose, we compared the MCI-MCI group to the MCI-AD group. While no DMPs passed the experiment-wide significance threshold, the most significant DMP was located in the *TRIM62* gene and showed hypomethylation in converters (probe cg25342005, $p = 1.67\times10^{-6}$; Table 3.5).

Top 10 ANOVA (CTL vs MCI vs AD) DMPs						
ProbeID	**Position**	**F**	***p* (F)**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg26146855	chr7: 1660603	12.632	5.58E-06			*ELFN1* (-88194), *TMEM184A* (-64538)
cg17510385	chr5: 912821	11.768	1.24E-05	*TRIP13*	Body	*NKD2* (-96346), *TRIP13* (19853)
cg05695925	chr2: 206493817	11.729	1.28E-05			*NRP2* (-53406)
cg10140957	chr16: 58768477	11.485	1.60E-05	*GOT2*	TSS1500	*GOT2* (-232)
cg18794145	chr1: 39546949	11.380	1.77E-05	*MACF1*	TSS200	*MACF1* (-249991), *NDUFS5* (54983)
cg06255020	chr2: 46097140	11.318	1.87E-05	*PRKCE*	Body	*EPAS1* (-427400), *PRKCE* (218098)
cg23630878	chr3: 48255260	11.224	2.04E-05			*CDC25A* (-25460), *CAMP* (-9576)
cg16723002	chr17: 27285150	11.166	2.16E-05	*SEZ6*	Body	*PHF12* (-6643), *SEZ6* (47930)
cg13026137	chr6: 31701260	10.810	3.00E-05	*CLIC1*	Body	*DDAH2* (-3222)
cg03456393	chr3: 193310565	10.380	4.47E-05	*OPA1*	TSS1500	*OPA1* (-367)

Table 3.1 The top 10 most significant DMPs for diagnosis.

The 10 most significant probes in an analysis of CTL, MCI and AD blood samples. Shown for each probe is chromosomal position (genome build 37), ANOVA F-statistic and accompanying p-value (p (F)). Methylation data was corrected for the covariates of age, sex, cell type proportion, and batch number prior to this analysis. Probes have been annotated using the Illumina gene annotation (UCSC Gene), genomic region (UCSC Gene Group), and GREAT annotation (distance to closest TSS shown in parentheses).

Top 10 CTL vs MCI DMPs						
ProbeID	**Position**	**Difference (CI)**	***p* CvM**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg17510385	chr5: 912821	-1.45 (-2.19 - -0.71)	1.70E-05	*TRIP13*	Body	*NKD2* (-96346), *TRIP13* (19853)
cg13026137	chr6: 31701260	1.74 (0.84 - 2.65)	2.52E-05	*CLIC1*	Body	*DDAH2* (-3222)
cg18477928	chr12: 111638391	-1.21 (-1.85 - -0.57)	3.28E-05	*CUX2*	Body	*CUX2* (166564), *FAM109A* (168533)
cg00499210	chr11: 65190135	0.67 (0.31 - 1.04)	4.47E-05	*NEAT1*	TSS200	*SCYL1* (-102412), *FRMD8* (36095)
cg18794145	chr1: 39546949	-0.81 (-1.25 - -0.37)	5.27E-05	*MACF1*	TSS200	*MACF1* (-249991), *NDUFS5* (54983)
cg07177123	chr7: 30909162	-0.69 (-1.06 - -0.32)	5.34E-05	*FAM188B*	Body	*GHRHR* (-94473), *FAM188B* (98130)
cg06972724	chr5: 130943742	-1.81 (-2.79 - -0.82)	5.98E-05	*RAPGEF6*	Body	*RAPGEF6* (27186), *CDC42SE2* (344041)
cg00852622	chr1: 112531653	1.68 (0.76 - 2.6)	6.91E-05	*KCND3*	5'UTR,1stExon	*KCND3* (123)
cg05728337	chr21: 43277362	-0.67 (-1.03 - -0.3)	7.84E-05	*PRDM15*	Body	*RIPK4* (-90114), *PRDM15* (22228)
cg01191920	chr7: 158217561	-2.77 (-4.31 - -1.23)	8.63E-05	*PTPRN2*	Body	*PTPRN2* (162920)

Table 3.2 The 10 most significant DMPs associated with MCI relative to CTL.

The 10 most significant probes in an analysis of CTL and MCI blood samples. Shown for each probe is chromosomal position (genome build 37), the group difference in % methylation (MCI – CTL), confidence interval (CI), and p-value as calculated with Tukey's HSD test. Methylation data was corrected for the covariates of age, sex, cell type proportion, and batch number prior to this analysis. Probes have been annotated using the Illumina gene annotation (UCSC Gene), genomic region (UCSC Gene Group), and GREAT annotation (distance to closest TSS shown in parentheses).

Top 10 MCI vs AD DMPs						
ProbeID	**Position**	**Difference (CI)**	***p* MvA**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg23630878	chr3: 48255260	1.63 (0.81 - 2.45)	1.27E-05			*CDC25A* (-25460), *CAMP* (-9576)
cg16518176	chr4: 135248367	-1.7 (-2.6 - -0.8)	3.78E-05			*PABPC4L* (-125465)
cg10140957	chr16: 58768477	-2.67 (-4.11 - -1.23)	5.19E-05	*GOT2*	TSS1500	*GOT2* (-232)
cg01575836	chr11: 15959856	1.06 (0.48 - 1.63)	6.26E-05			*SOX6* (464556), *INSC* (825887)
cg16135995	chr7: 73668739	0.33 (0.15 - 0.52)	7.26E-05	*RFC2*	TSS200	*RFC2* (-2)
cg00114478	chr8: 73658110	-4.81 (-7.45 - -2.16)	7.52E-05	*KCNB2*	Body	*TERF1* (-262986), *KCNB2* (208485)
cg21565421	chr15: 74592665	-3.7 (-5.76 - -1.65)	8.59E-05	*CCDC33*	Body	*CCDC33* (63999), *CYP11A1* (67415)
cg24760753	chr2: 159534030	1.37 (0.61 - 2.13)	8.65E-05	*PKP4*	Body	*DAPL1* (-117798), *PKP4* (220555)
cg26146855	chr7: 1660603	-0.78 (-1.22 - -0.34)	1.07E-04			*ELFN1* (-88194), *TMEM184A* (-64538)
cg16723002	chr17: 27285150	0.69 (0.3 - 1.07)	1.12E-04	*SEZ6*	Body	*PHF12* (-6643), *SEZ6* (47930)

Table 3.3 The 10 most significant DMPs associated with AD relative to MCI.

The 10 most significant probes in an analysis of MCI and AD blood samples. Shown for each probe is chromosomal position (genome build 37), the group difference in % methylation (AD – MCI), confidence interval (CI), and p-value as calculated with Tukey's HSD test. Methylation data has been corrected for the covariates of age, sex, cell type proportion, and batch number prior to this analysis. Probes have been annotated using the Illumina gene annotation (UCSC Gene), genomic region (UCSC Gene Group), and GREAT annotation (distance to closest TSS shown in parentheses).

Top 10 CTL vs AD DMPs						
ProbeID	**Position**	**Difference (CI)**	***p* CvA**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg05695925	chr2: 206493817	-0.99 (-1.49 - -0.49)	1.32E-05			*NRP2* (-53406)
cg26146855	chr7: 1660603	-0.89 (-1.36 - -0.43)	2.11E-05			*ELFN1* (-88194), *TMEM184A* (-64538)
cg03456393	chr3: 193310565	-3.74 (-5.68 - -1.8)	2.51E-05	*OPA1*	TSS1500	*OPA1* (-367)
cg03699566	chr11: 71900652	-0.83 (-1.27 - -0.39)	4.23E-05	*FOLR1*	1stExon,TSS1500,5'UTR	*FOLR1* (51)
cg05827732	chr3: 96533699	-1.37 (-2.11 - -0.64)	4.91E-05	*EPHA6*	1stExon	*EPHA6* (275)
cg13406003	chr6: 127535477	1.74 (0.8 - 2.68)	5.31E-05			*RNF146* (-52349), *RSPO3* (95430)
cg06255020	chr2: 46097140	0.65 (0.3 - 1)	5.40E-05	*PRKCE*	Body	*EPAS1* (-427400), *PRKCE* (218098)
cg03530006	chr13: 38920380	-1.07 (-1.66 - -0.48)	6.99E-05			*UFM1* (-3561)
cg18120975	chr6: 151313327	0.74 (0.33 - 1.15)	8.93E-05	*MTHFD1L*	Body	*AKAP12* (-247806), *MTHFD1L* (126513)
cg22677650	chr19: 3675627	-1.19 (-1.85 - -0.53)	9.06E-05	*PIP5K1C*	Body	*C19orf29* (-48815), *PIP5K1C* (24849)

Table 3.4 The 10 most significant DMPs associated with AD relative to CTL.

The 10 most significant probes in an analysis of CTL and AD blood samples. Shown for each probe is chromosomal position (genome build 37 the group difference in % methylation (AD – CTL), confidence interval (CI), and p-value as calculated with Tukey's HSD test. Methylation data was corrected for the covariates of age, sex, cell type proportion, and batch number prior to this analysis. Probes have been annotated using the Illumina gene annotation (UCSC Gene), genomic region (UCSC Gene Group), and GREAT annotation (distance to closest TSS shown in parentheses).

Top 10 MCI to AD conversion DMPs						
ProbeID	**Position**	**Diagnosis Est**	***p*-value**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg25342005	chr1: 33648656	-0.025	1.67E-06	*TRIM62*	TSS1500	*TRIM62* (-986)
cg17292662	chr8: 54758936	-0.032	1.06E-05			*ATP6V1H* (-3087)
cg19636302	chr19: 46032854	-0.037	1.27E-05	*OPA3*	Body	*VASP* (22167), *OPA3* (55267)
cg20655558	chr22: 41257839	0.013	1.93E-05	*XPNPEP3;DNAJB7*	Body;1stExon	*DNAJB7* (290)
cg07850832	chr5: 37378050	0.008	3.79E-05	*WDR70*	TSS1500	*NUP155* (-6854), *GDNF* (457878)
cg04664179	chr19: 40913800	-0.016	4.15E-05	*PRX*	Body	*HIPK4* (-17707), *PRX* (5470)
cg00723019	chr19: 48867290	0.006	4.49E-05	*TMEM143;SYNGR4*	TSS200;TSS1500	*SYNGR4* (-360), *TMEM143* (-105)
cg16520539	chr6: 28890872	0.008	6.99E-05	*TRIM27*	Body	*TRIM27* (895)
cg22322663	chr6: 30525268	0.008	7.73E-05	*PRR3;GNL1*	Body;5'UTR,1stExon	*GNL1* (102), *PRR3* (783)
cg09907542	chr4: 7033761	-0.033	7.89E-05	*TBC1D14*	3'UTR	*TADA2B* (-11394), *TBC1D14* (122267)

Table 3.5 The 10 most significant DMPs associated with future progression to AD.

The 10 most significant probes in a comparison of MCI-MCI and MCI-AD blood samples. Shown for each probe is chromosomal position (genome build 37), regression coefficient (Diagnosis Est), and accompanying p-value. Methylation data has been corrected for the covariates of age, sex, cell type proportion, batch number, and MMSE score prior to this analysis. Probes have been annotated using the Illumina gene annotation (UCSC Gene), genomic region (UCSC Gene Group), and GREAT annotation (distance to closest TSS shown in parentheses).

3.4.2. A number of significant DMRs can be identified in MCI and AD blood

We next used a sliding window approach to identify regions spanning multiple adjacent DMPs that were significantly different in MCI and AD. We found four DMRs associated with differences across the three baseline groups (CTL, MCI, and AD, Table 3.6A). A ten probe DMR of 574 bp was identified in *MOV10L1* (Figure 3.2), as well as a five-probe (582 bp) intergenic DMR annotated to *CBFA2T3* (Figure 3.3), with probes in both DMRs generally showing hypermethylation in MCI samples, with levels in AD samples similar to CTL. An eight-probe DMR of 301 bp was found in the readthrough transcription region of *TPTEP2-CSNK1E*, which appeared to be mainly driven by hypermethylation in the MCI group (***Figure 3.4***). One of the four identified DMRs was driven by a difference between the CTL and AD groups (Table 3.6B); we identified a 1,303 bp DMR in the *HOXB6* gene, containing 12 probes (Figure 3.5; Figure 3.6). Each of the twelve probes showed hypermethylation in AD.

In the analysis of MCI conversion to AD, nine significant DMRs were identified (Table 3.7). We found DMRs showing decreased methylation in MCI-AD converters relative to MCI-MCI non-converters in the genes *CPT1B* and *CHKB* (932 bp; 14 probes; Figure 3.7), *TMEM184A* (659 bp; six probes; Figure 3.8), *KCNAB3* (558 bp; seven probes; Figure 3.9), *GABBR1* (379 bp; ten probes; Figure 3.10), *PRDM1* (121 bp; five probes; Figure 3.11), *FLJ37453* (568 bp; six probes; Figure 3.12), and *OR56A3* and *TRIM5* (556 bp; five probes; Figure 3.13). Hypermethylation in MCI-AD converters relative to MCI-MCI non-converters was seen in two DMRs located in the genes *SMC1B* and *RIBC2* (725 bp; 15 probes; Figure 3.14), and an intergenic region near the gene *FIGN* (716 bp; six probes; Figure 3.15).

A. ANOVA: CTL vs MCI vs AD								
Gene	**Position**	**Gene Feature**	***n***	***p*-value**	**Šidák-*p***	**Average Methylation %**		
						CTL	**MCI**	**AD**
HOXB-AS3;HOXB6	chr17: 46681111 - 46682414	nc_intron+nc_exon;TSS+intron+exon+utr5	12	2.79E-14	8.58E-12	56.59	58.79	60.81
MOV10L1	chr22: 50528179 - 50528753	TSS+intron+utr5+cds;TSS+exon+utr5	10	2.03E-07	1.42E-04	68.38	70.18	68.23
CBFA2T3	chr16: 88937216 - 88937798	intergenic	5	2.61E-07	1.80E-04	42.57	44.66	42.34
TPTEP2-CSNK1E	chr22: 38714166 - 38714467	intron+utr5	8	1.87E-06	2.49E-03	41.21	42.29	41.69
B. CTL vs AD								
Gene	**Position**	**Gene Feature**	***n***	***p*-value**	**Šidák-*p***	**Average Methylation %**		
						CTL	**MCI**	**AD**
HOXB-AS3;HOXB6	chr17: 46681111 - 46682414	nc_intron+nc_exon;TSS+intron+exon+utr5	12	3.36E-16	1.03E-13	56.59	58.79	60.81

Table 3.6 DMRs associated with baseline diagnosis in blood.

DMRs in a comparison of CTL, MCI, and AD blood samples. Shown are DMRs for (A) the overall three group (ANOVA) comparison, and the post-hoc (B) CTL v AD comparison. Displayed for each region is the UCSC gene name, chromosomal position (genome build 37), gene feature (TSS = transcription start site; utr5 = 5' untranslated region; utr3 = 3' untranslated region; cds = coding sequence), number of probes in region (n), p-value and multiple testing-corrected p-value (Šidák-p), and average β-value per group.

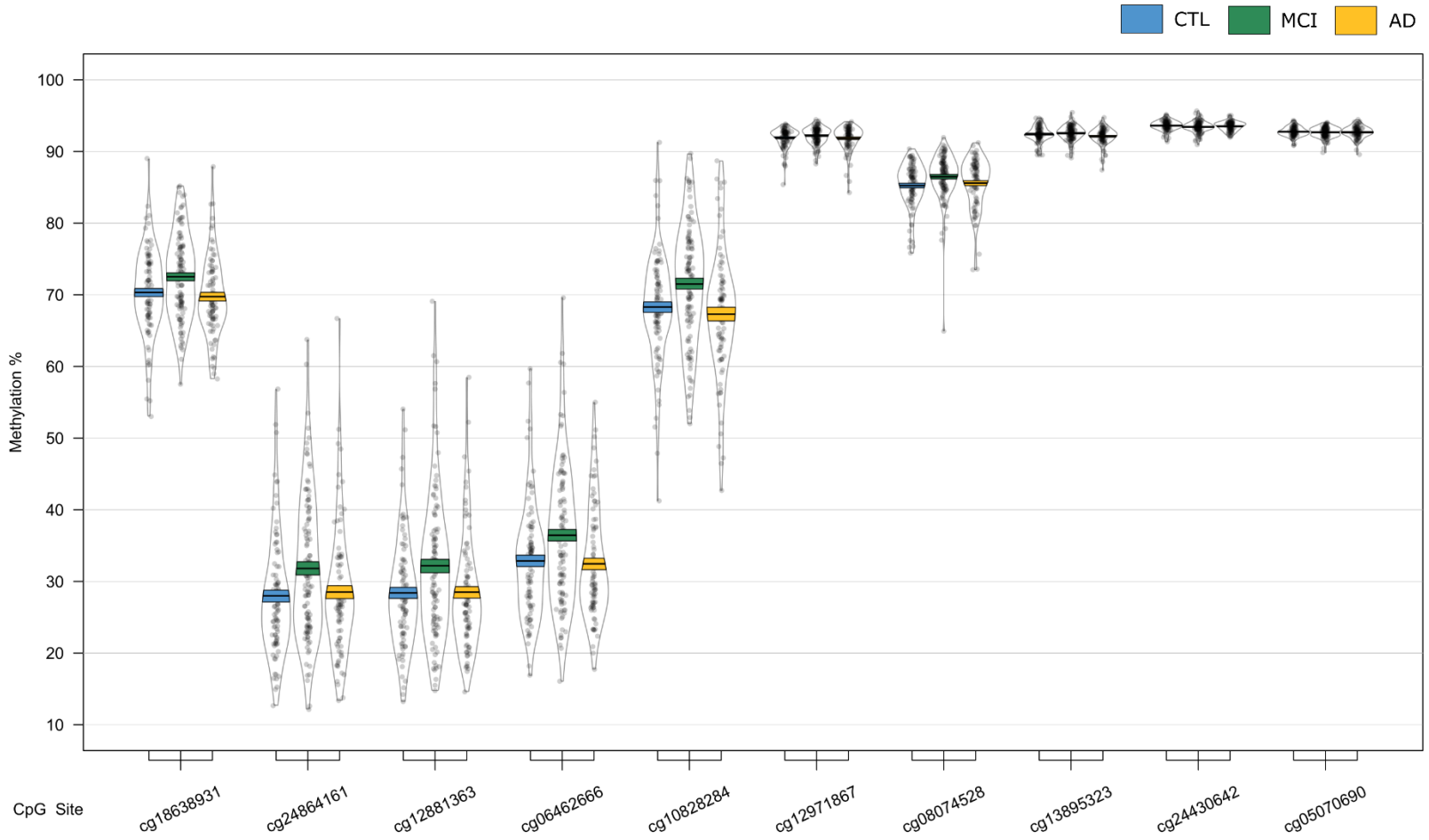

Figure 3.2 The MOV10L1 DMR associated with baseline diagnostic status.

Displayed are the methylation levels of individual probes located within the DMR, ordered by genomic location. Methylation values have been corrected for covariates age, sex, cell type proportion and batch.

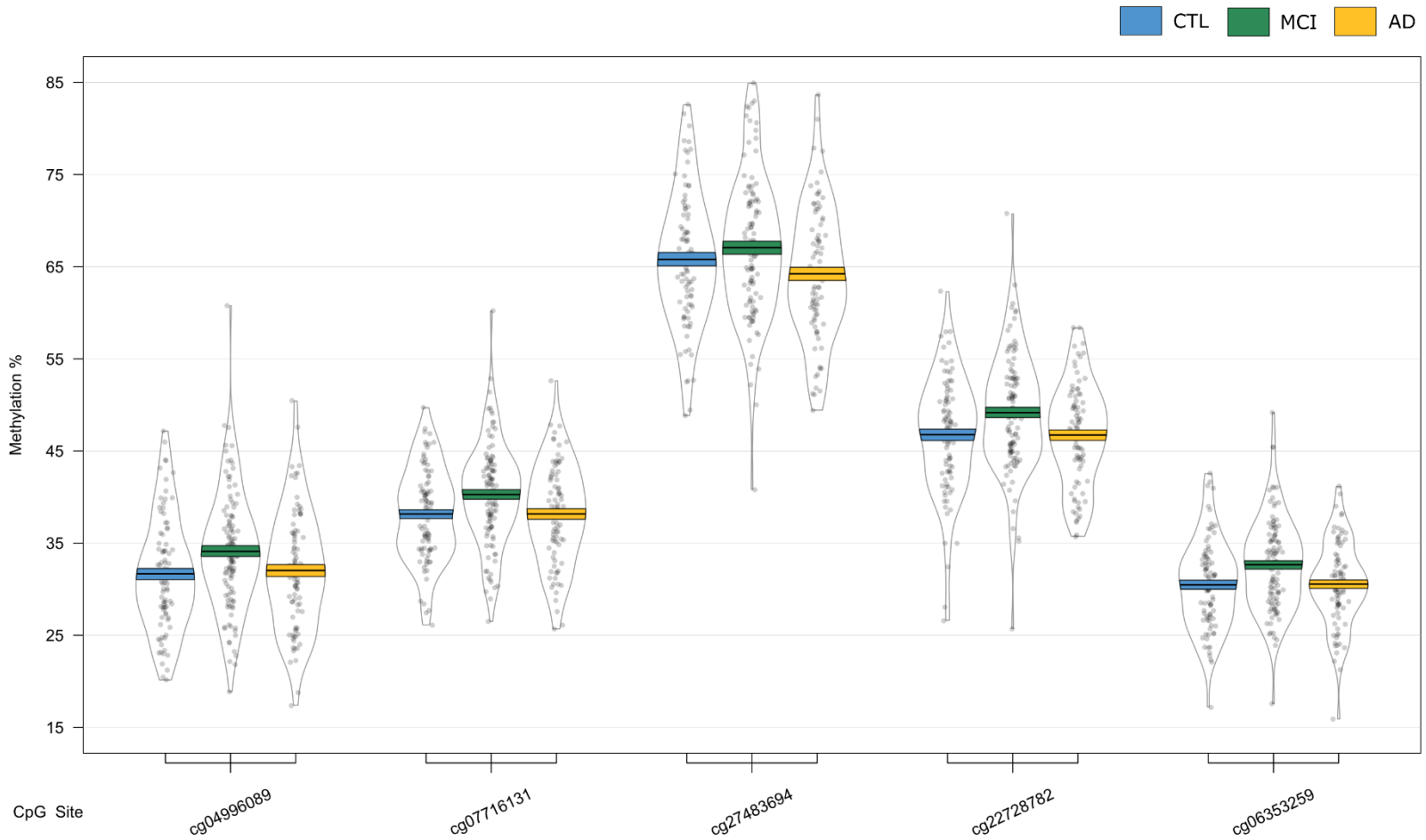

Figure 3.3 The CBFA2T3 DMR associated with baseline diagnostic status.

Displayed are the methylation levels of individual probes located within the DMR, ordered by genomic location. Methylation values have been corrected for covariates age, sex, cell type proportion and batch.

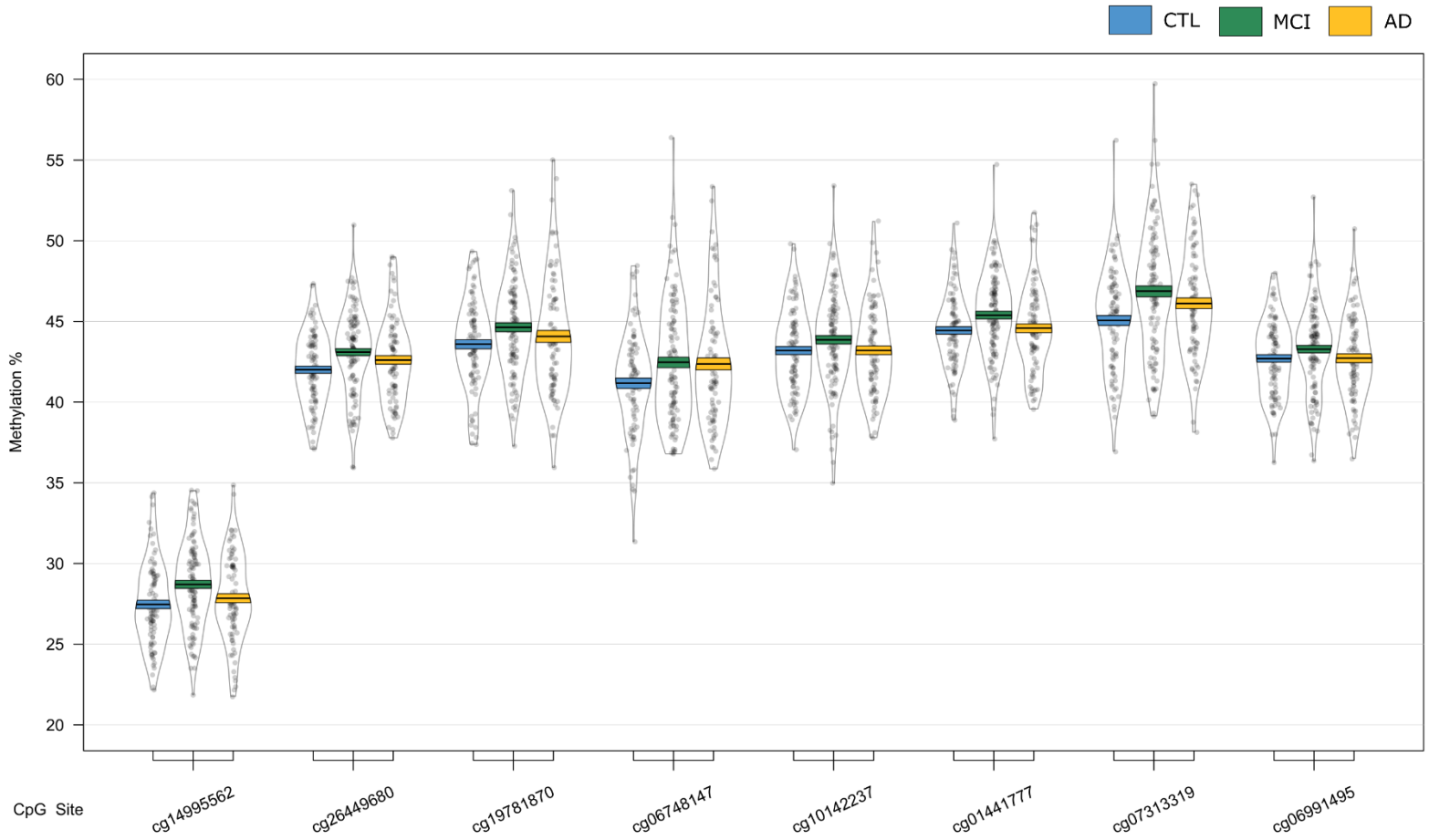

Figure 3.4 The TPTEP2-CSNK1E DMR associated with baseline diagnostic status.

Displayed are the methylation levels of individual probes located within the DMR, ordered by genomic location. Methylation values have been corrected for covariates age, sex, cell type proportion and batch.

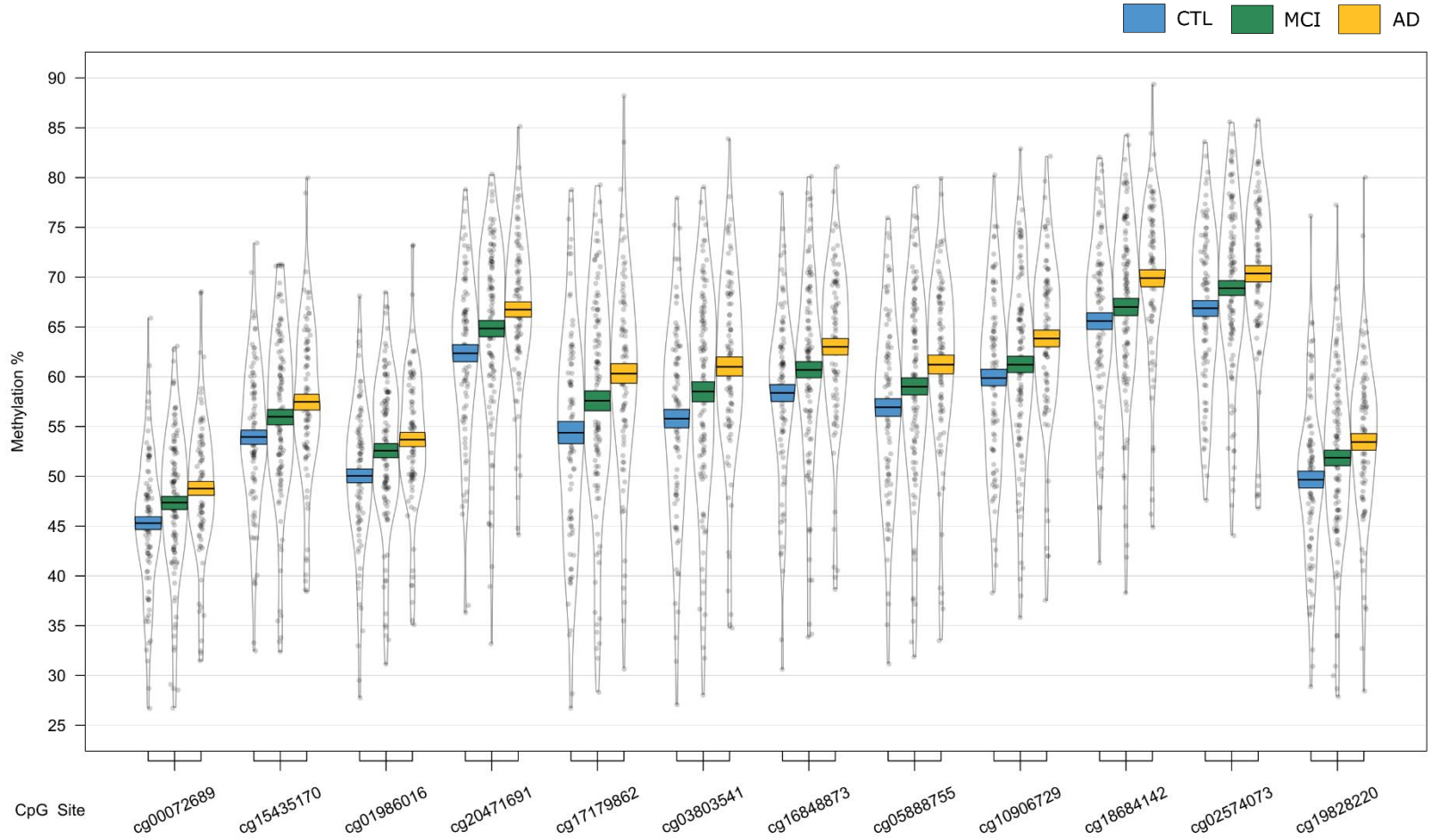

Figure 3.5 The HOXB6 DMR associated with AD relative to CTL.

Displayed are the methylation levels of individual probes located within the DMR, ordered by genomic location. Methylation values have been corrected for covariates age, sex, cell type proportion and batch.

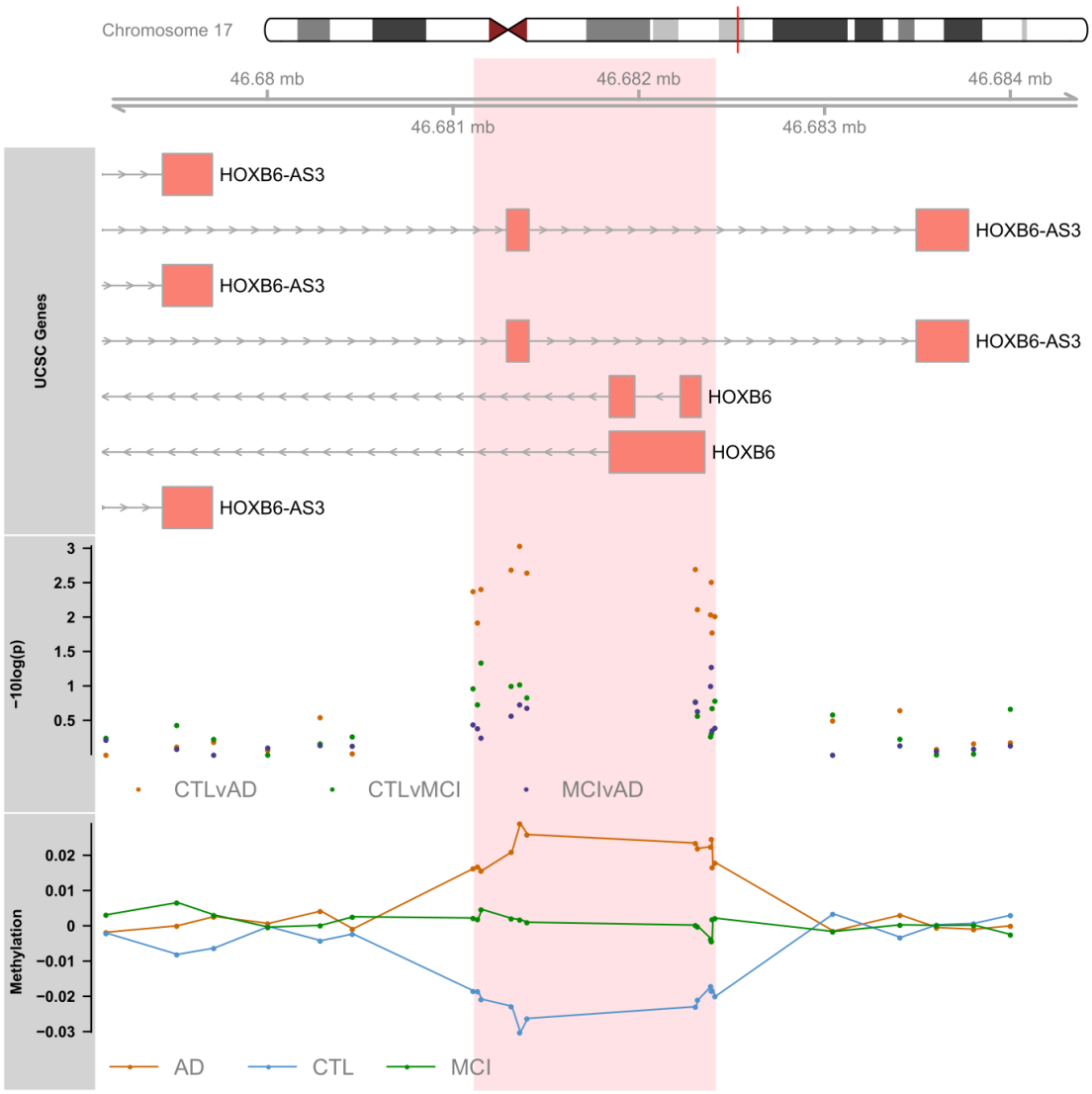

Figure 3.6 The HOXB6 genomic region.

The HOXB6 DMR was shown to be altered in AD relative to CTL. The region spanned by the DMR is highlighted in red, and genomic location and UCSC gene annotations are shown in addition to a mini-manhattan plot of the p-values of probes within and neighbouring the DMR, p-values related to changes in AD relative to CTL are shown in orange, MCI relative to CTL in green, and AD relative to MCI in purple. The bottom panel shows relative methylation levels across the region, with methylation in AD in orange, MCI in green, and CTL in blue.

MCI-MCI vs MCI-AD							
Gene	**Position**	**Gene Feature**	***n***	***p*-value**	**Šidák-*p***	**Average Methylation %**	
						MCI-MCI	**MCI-AD**
CHKB-CPT1B; CPT1B;CHKB	chr22: 51016501 - 51017433	nc_intron;TSS+intron+exon+utr5;exon+utr3	14	2.05E-14	8.84E-12	64.92	61.26
SMC1B;RIBC2	chr22: 45809319 - 45810044	TSS+intron+utr5+cds;TSS+intron+utr5+cds	15	8.26E-09	4.57E-06	24.62	26.62
TMEM184A	chr7: 1595602 - 1596261	TSS+intron+exon+utr5	6	2.41E-08	1.47E-05	45.33	43.31
KCNAB3	chr17: 7832680 - 7833238	TSS+cds	7	8.11E-08	5.83E-05	80.31	76.73
GABBR1	chr6: 29599012 - 29599391	intron+exon+utr5;intron+cds	10	9.72E-08	1.03E-04	63.80	61.48
FIGN	chr2: 164204628 - 164205344	intergenic	6	3.58E-07	2.01E-04	52.84	56.10
PRDM1	chr6: 106546704 - 106546825	TSS+exon+utr5;intron	5	1.04E-07	3.45E-04	62.01	58.89
FLJ37453	chr1: 16163555 - 16164123	nc_intron	6	5.92E-07	4.18E-04	29.67	27.34
OR56A3;TRIM5	chr11: 5959658 - 5960214	intergenic	5	9.68E-07	6.98E-04	81.17	77.54

Table 3.7 DMRs associated with future conversion to AD in blood.

DMRs in a comparison of CTL, MCI, and AD blood samples. Shown are DMRs for the MCI-MCI to MCI-AD comparison. Displayed for each region is the UCSC gene name, chromosomal position (genome build 37), gene feature (TSS = transcription start site; utr5 = 5' untranslated region; utr3 = 3' untranslated region; cds = coding sequence), number of probes in region (n), p-value and multiple testing-corrected p-value (Šidák-p), and average β-value per group.

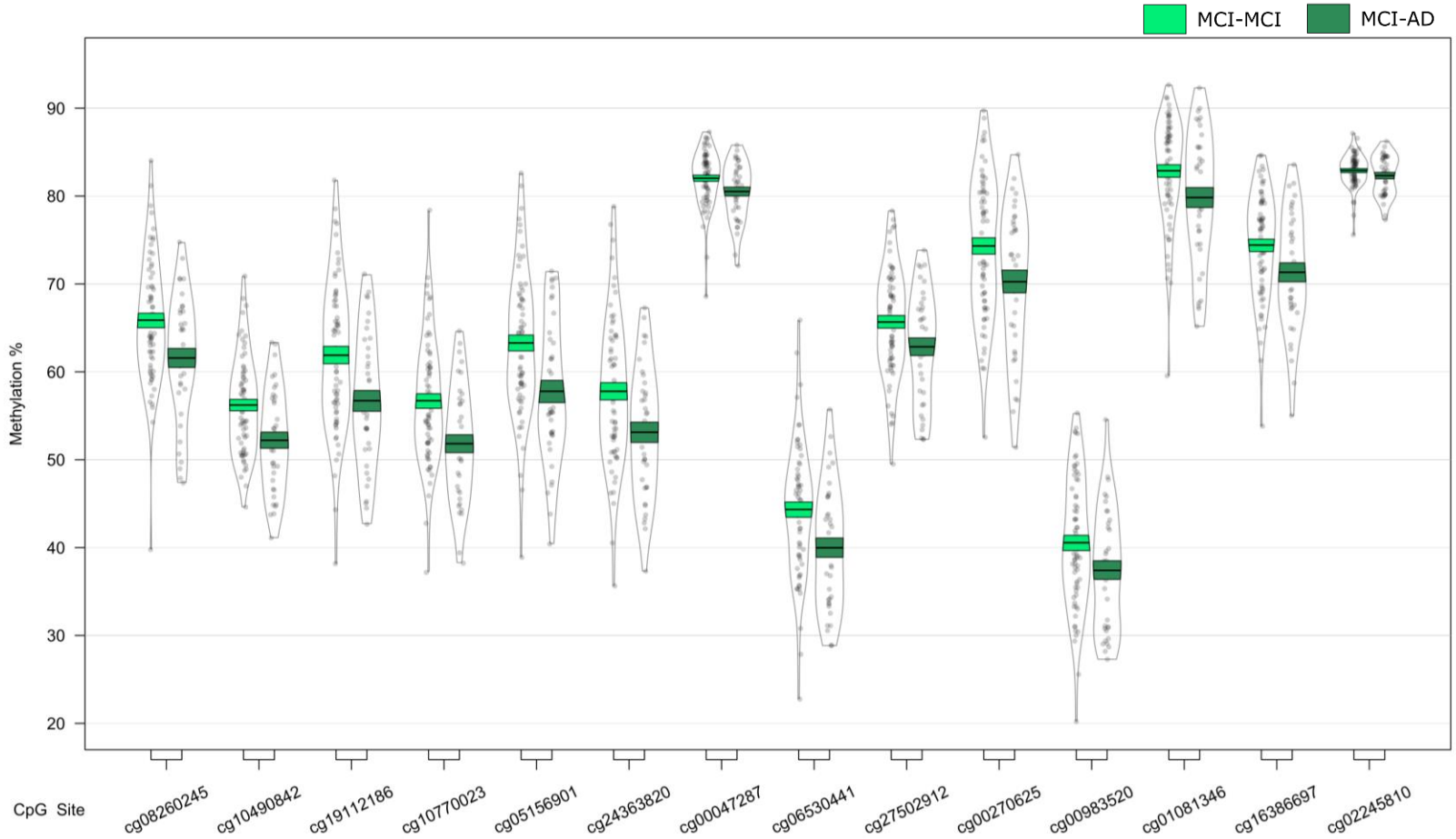

Figure 3.7 The CPT1B-CHKB DMR associated with future conversion to AD.

The CPT1B-CHKB DMR reflected differences between MCI-AD converterd (dark green, shown on the right), and stable MCI-MCI (light green, shown on the left). Displayed are the methylation levels of all probes ($p < 0.05$) within the genomic location covered by the DMR, ordered by genomic location. Methylation values have been corrected for covariates age, sex, cell type proportion, batch, and baseline MMSE score.

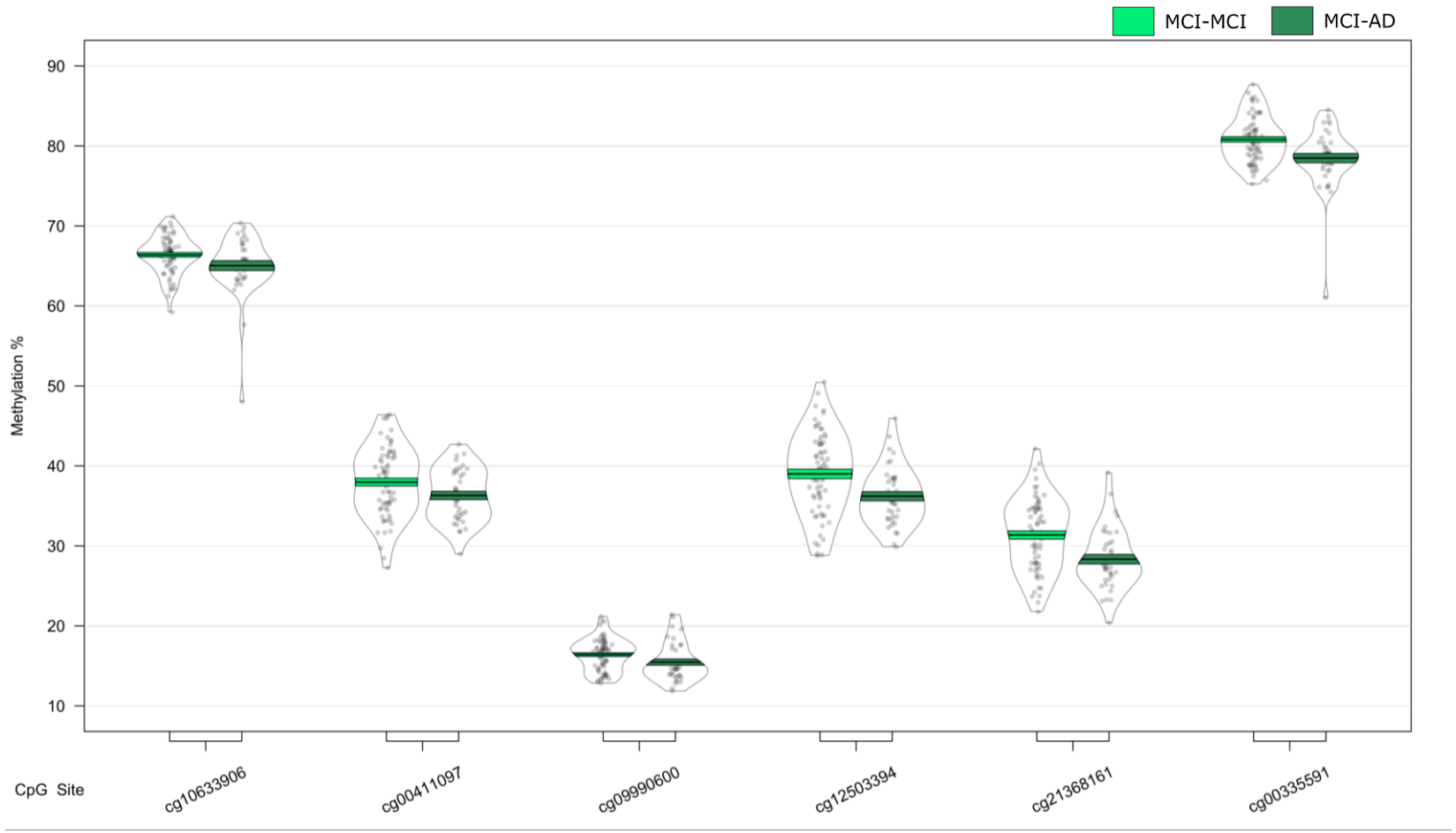

Figure 3.8 The TMEM184A DMR associated with future conversion to AD.

The TMEM184A DMR reflected differences between MCI-AD converterd (dark green, shown on the right), and stable MCI-MCI (light green, shown on the left). Displayed are the methylation levels of all probes ($p < 0.05$) within the genomic location covered by the DMR, ordered by genomic location. Methylation values have been corrected for covariates age, sex, cell type proportion, batch, and baseline MMSE score.

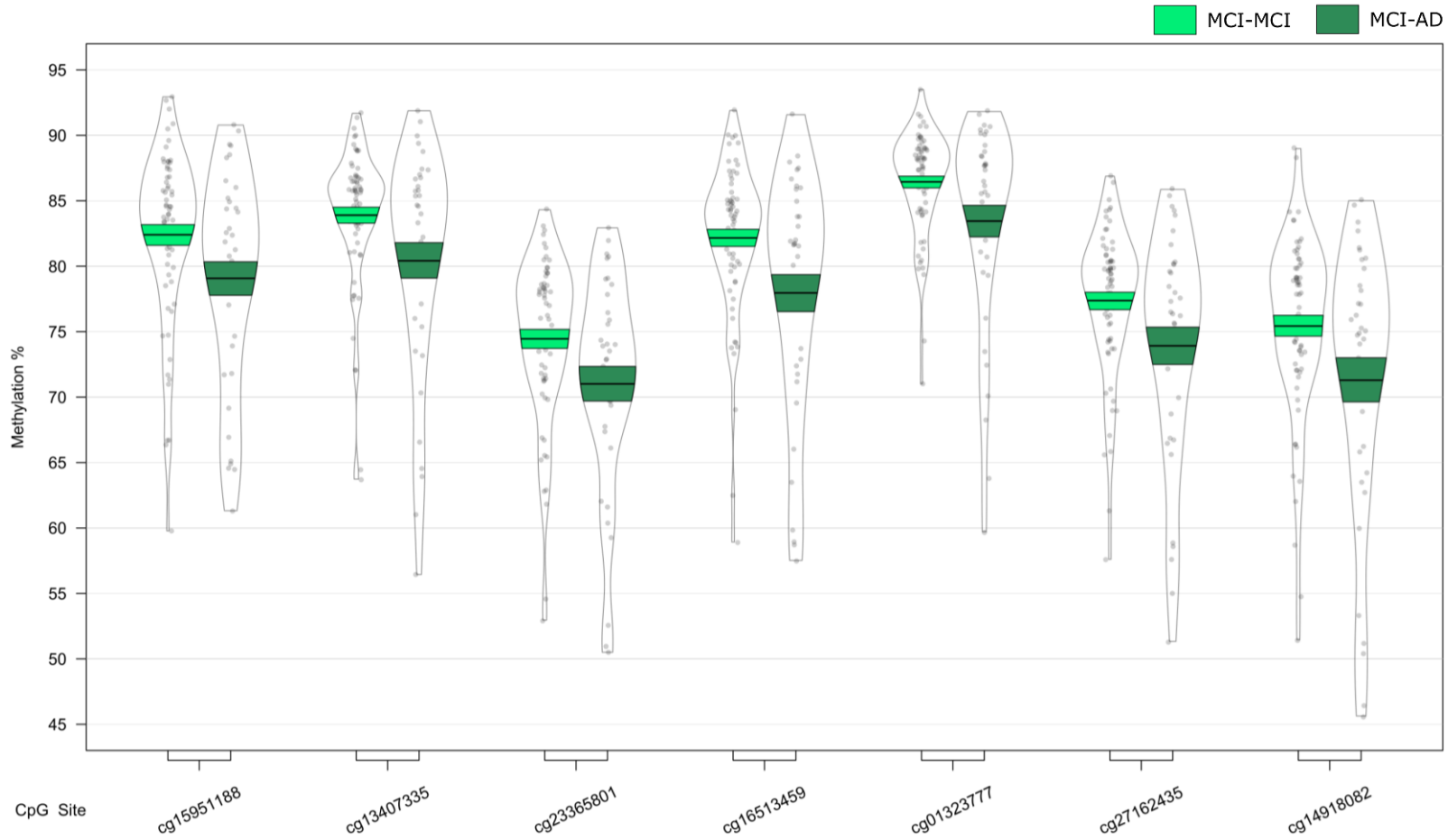

Figure 3.9 The KCNAB3 DMR associated with future conversion to AD.

The KCNAB3 DMR reflected differences between MCI-AD converterd (dark green, shown on the right), and stable MCI-MCI (light green, shown on the left). Displayed are the methylation levels of all probes ($p < 0.05$) within the genomic location covered by the DMR, ordered by genomic location. Methylation values have been corrected for covariates age, sex, cell type proportion, batch, and baseline MMSE score.

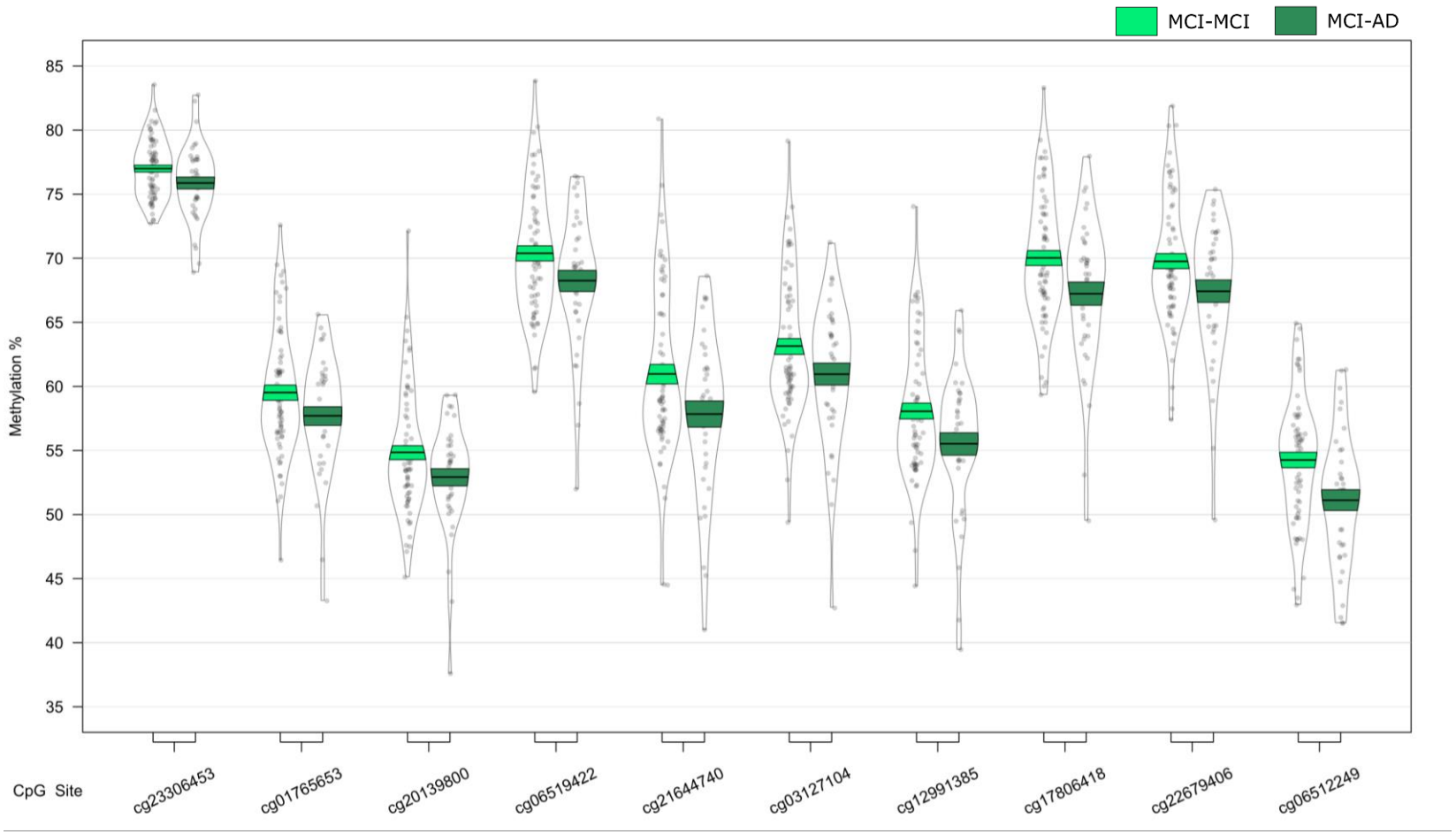

Figure 3.10 The GABBR1 DMR associated with future conversion to AD.

The GABBR1 DMR reflected differences between MCI-AD converterd (dark green, shown on the right), and stable MCI-MCI (light green, shown on the left). Displayed are the methylation levels of all probes ($p < 0.05$) within the genomic location covered by the DMR, ordered by genomic location. Methylation values have been corrected for covariates age, sex, cell type proportion, batch, and baseline MMSE score.

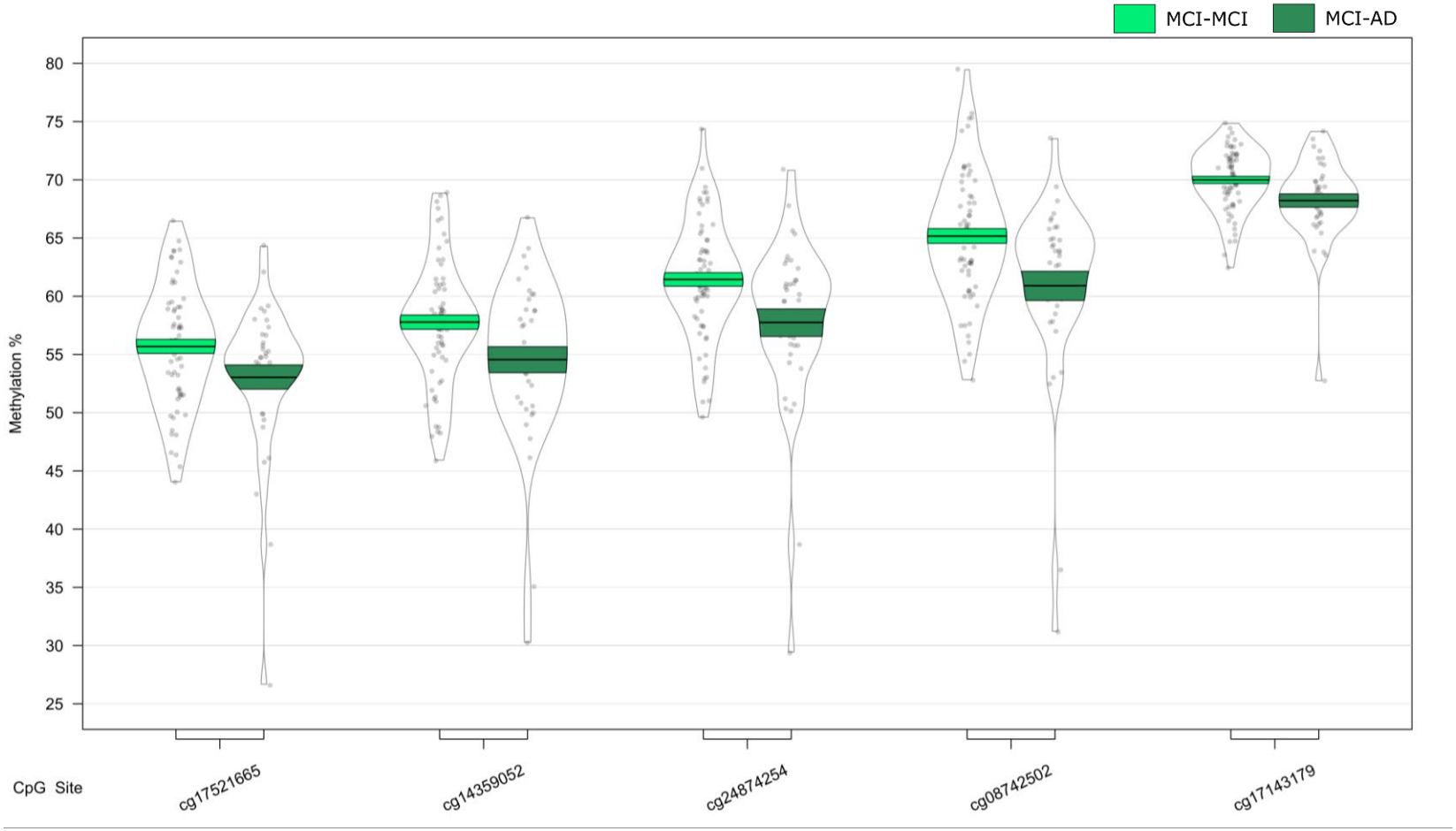

Figure 3.11 The PRDM1 DMR associated with future conversion to AD.

The PRDM1 DMR reflected differences between MCI-AD converterd (dark green, shown on the right), and stable MCI-MCI (light green, shown on the left). Displayed are the methylation levels of all probes ($p < 0.05$) within the genomic location covered by the DMR, ordered by genomic location. Methylation values have been corrected for covariates age, sex, cell type proportion, batch, and baseline MMSE score.

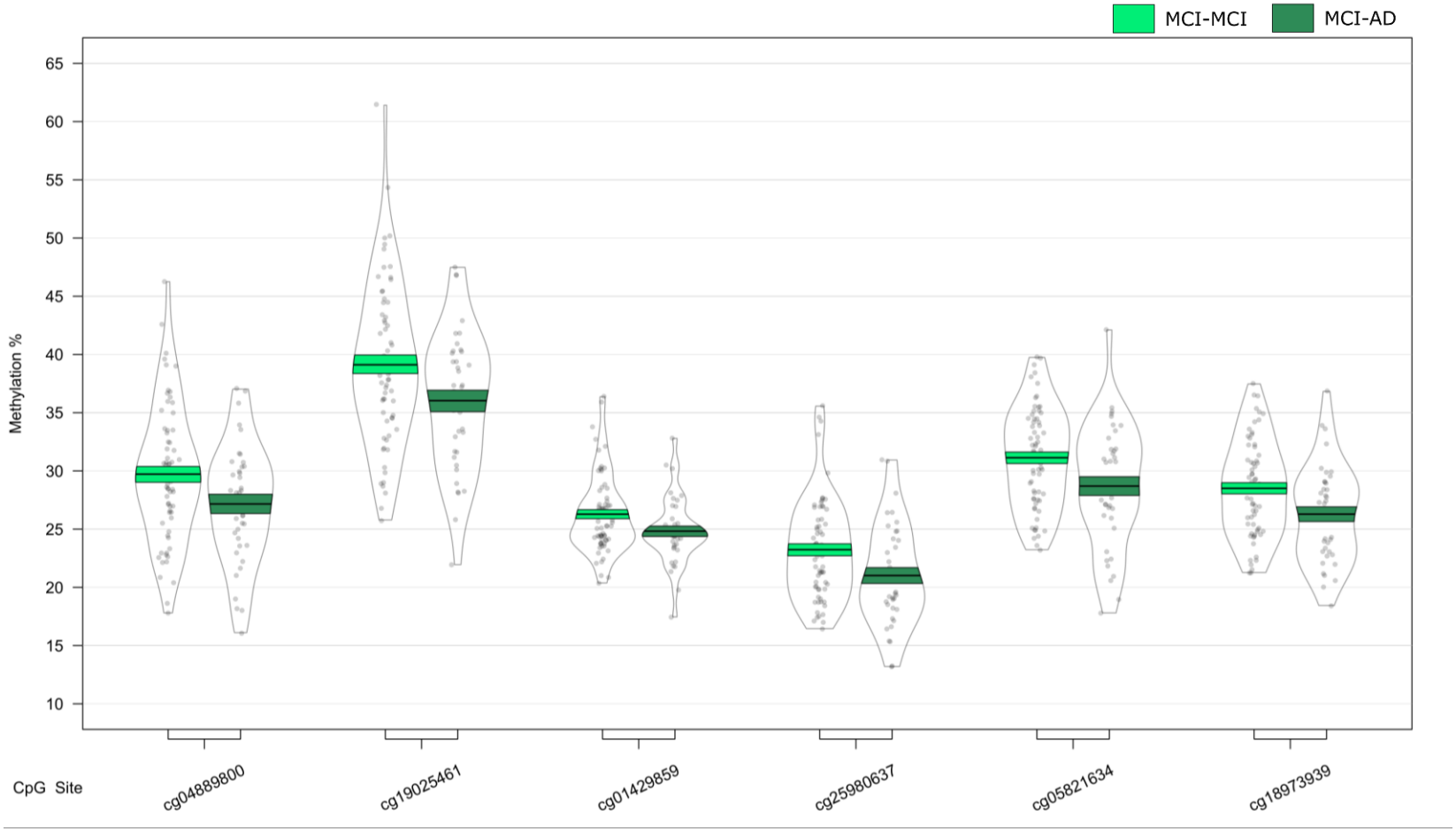

Figure 3.12 The FLJ37453 DMR associated with future conversion to AD.

The FLJ37453 DMR reflected differences between MCI-AD converterd (dark green, shown on the right), and stable MCI-MCI (light green, shown on the left). Displayed are the methylation levels of all probes ($p < 0.05$) within the genomic location covered by the DMR, ordered by genomic location. Methylation values have been corrected for covariates age, sex, cell type proportion, batch, and baseline MMSE score.

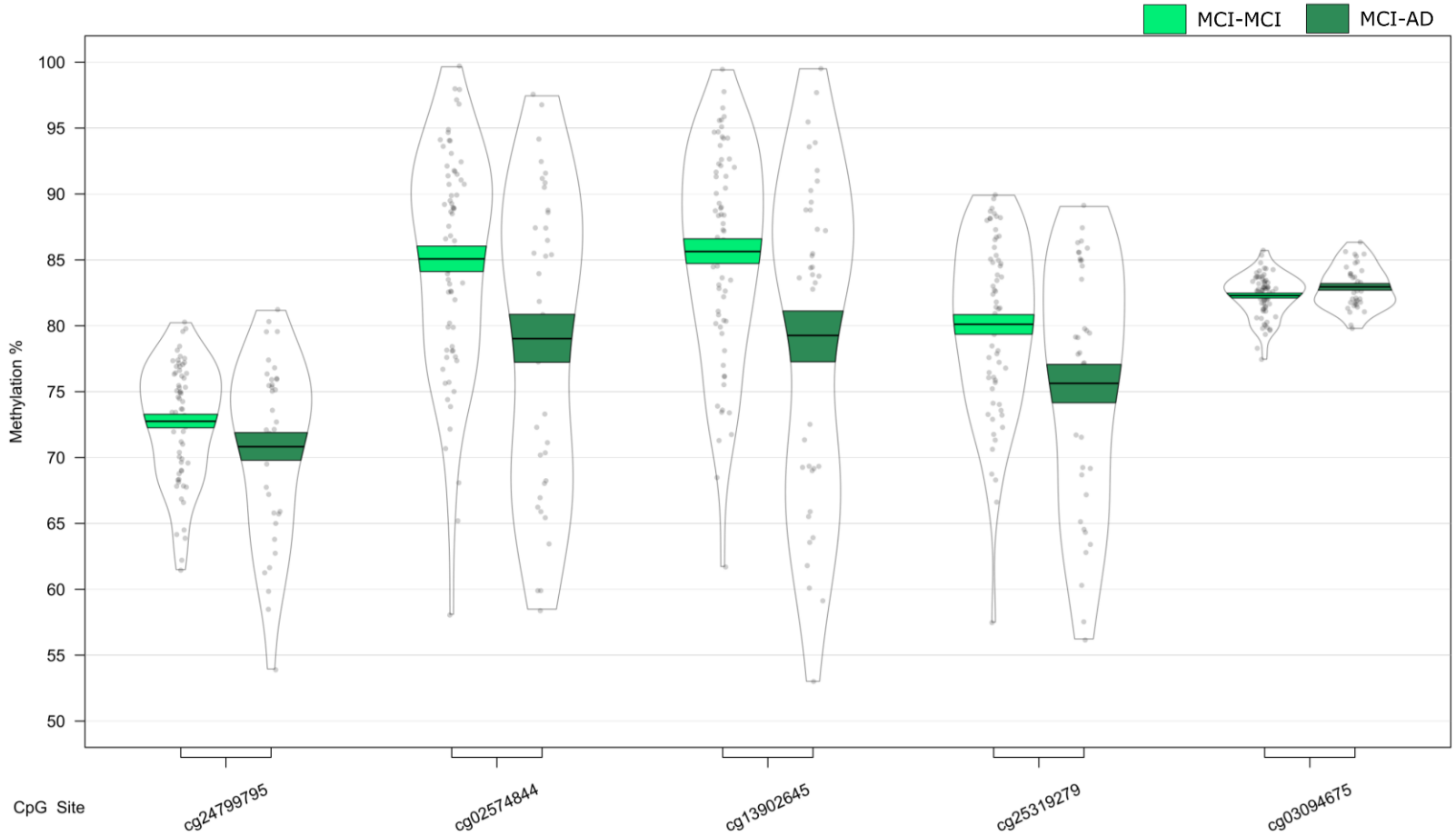

Figure 3.13 The OR56A3 and TRIM5 DMR associated with future conversion to AD.

The OR56A3 and TRIM5 DMR reflected differences between MCI-AD converterd (dark green, shown on the right), and stable MCI-MCI (light green, shown on the left). Displayed are the methylation levels of all probes ($p < 0.05$) within the genomic location covered by the DMR, ordered by genomic location. Methylation values have been corrected for covariates age, sex, cell type proportion, batch, and baseline MMSE score.

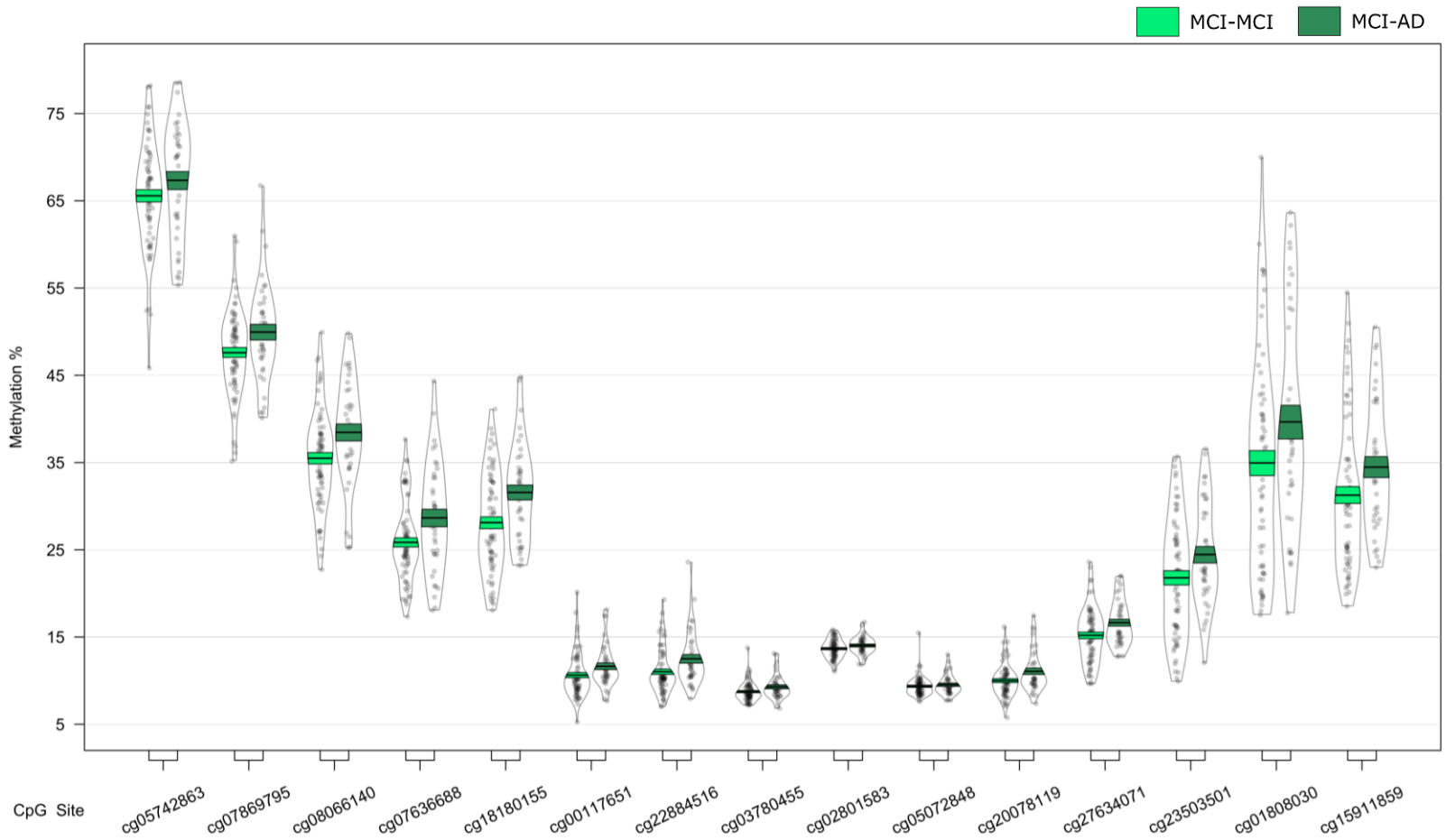

Figure 3.14 The SMC1B and RIBC2 DMR associated with future conversion to AD.

The SMC1B and RIBC2 DMR reflected differences between MCI-AD converterd (dark green, shown on the right), and stable MCI-MCI (light green, shown on the left). Displayed are the methylation levels of all probes ($p < 0.05$) within the genomic location covered by the DMR, ordered by genomic location. Methylation values have been corrected for covariates age, sex, cell type proportion, batch, and baseline MMSE score.

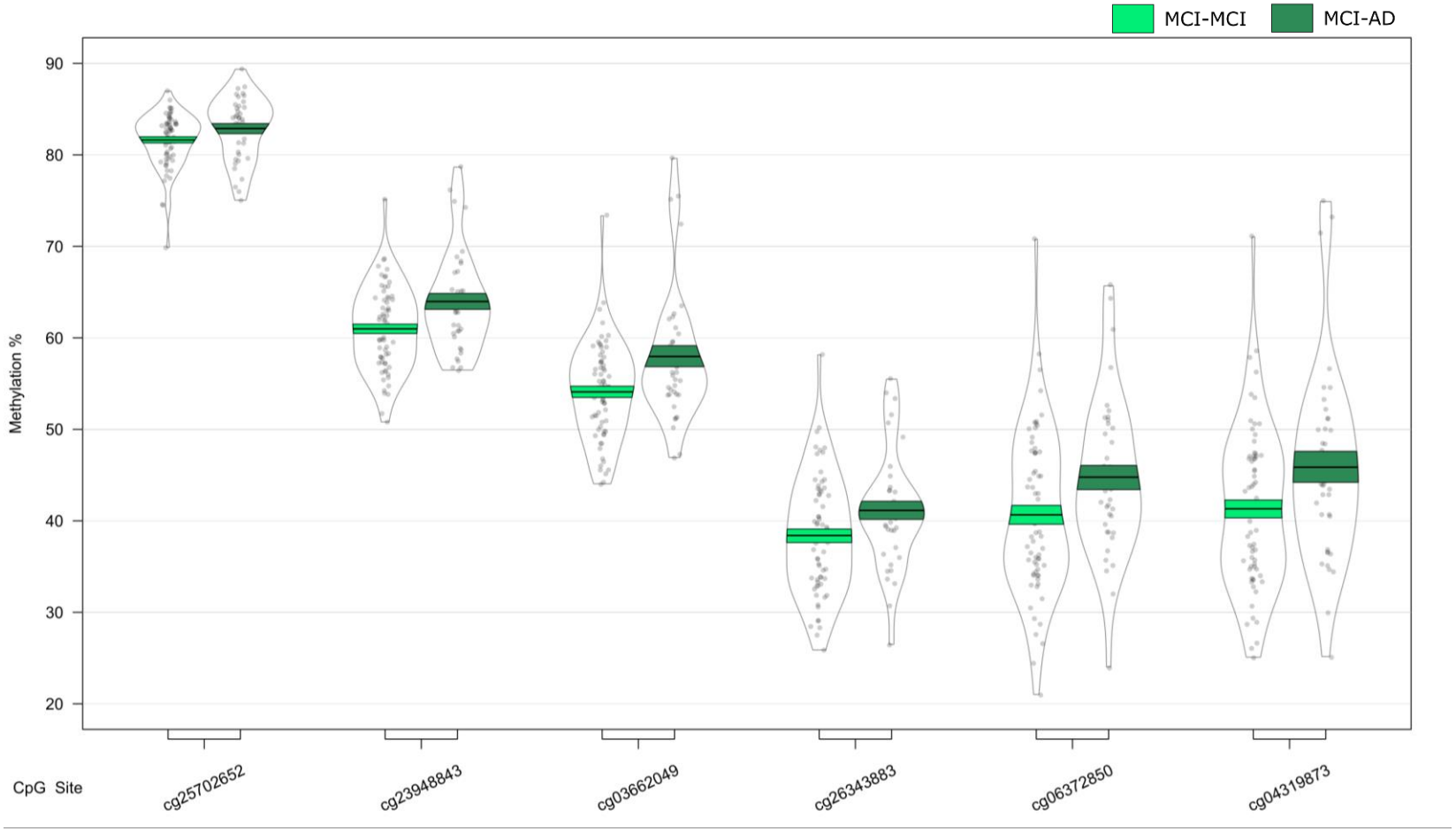

Figure 3.15 The FIGN DMR associated with future conversion to AD.

The FIGN DMR reflected differences between MCI-AD converterd (dark green, shown on the right), and stable MCI-MCI (light green, shown on the left). Displayed are the methylation levels of all probes ($p < 0.05$) within the genomic location covered by the DMR, ordered by genomic location. Methylation values have been corrected for covariates age, sex, cell type proportion, batch, and baseline MMSE score.

3.4.3. Validation of the AD-associated DMR in *HOXB6* by pyrosequencing

Interestingly, differential DNA methylation at the most significant locus within the *HOXB6* DMR (cg17179862) has been previously reported in AD hippocampus (Altuna et al., 2019). In order to further explore AD-associated hypermethylation in this gene we used pyrosequencing to validate our *HOXB6* DMR, covering two CpG sites on the array (cg17179862, cg03803541) as well as three neighbouring CpG sites that were not covered by the 450K array (chr17:46681421, chr17:46681394, and chr17:46681383). We found significant differences between groups at all five CpG sites (Table 3.8, Figure 3.16A), and when averaged over the full five probes (Figure 3.16B), demonstrating hypermethylation in AD samples relative to CTL. The pattern of DNA methylation quantified by the 450K array and pyrosequencing was similar for both cg03803541 (Figure 3.17A) and cg17179862 (Figure 3.17B), with a significant correlation of the methylation values estimated by the two technologies for both cg03803541 (Figure 3.18A; $r = 0.957$, $p = 2.69\times10^{-142}$) and cg17179862 (Figure 3.18B: $r = 0.934$, $p = 5.03\times10^{-68}$).

Genomic position	F	*p* (F)	CTL vs MCI		CTL vs AD		MCI vs AD	
			Difference	Tukey's *p*	Difference	Tukey's *p*	Difference	Tukey's *p*
chr17: 46681421	4.972	7.61E-03	3.378	0.16	6.246	5.26E-03	2.868	0.29
chr17: 46681401 (cg03803541)	4.650	1.04E-02	3.837	0.10	6.012	8.50E-03	2.175	0.49
chr17: 46681394	4.315	1.44E-02	3.294	0.16	5.580	1.09E-02	2.286	0.42
chr17: 46681383	4.207	1.59E-02	3.613	0.11	5.407	1.42E-02	1.794	0.59
chr17: 46681362 (cg17179862)	3.625	2.91E-02	3.012	0.37	6.573	2.17E-02	3.561	0.25
Region Average	4.519	1.18E-02	3.484	0.13	5.717	9.15E-03	2.233	0.45

Table 3.8 Replication of AD-associated hypermethylation in HOXB6 by pyrosequencing.

An analysis of DNA methylation levels in the HOXB6 DMR in CTL, MCI, and AD samples. Each CpG site is indicated by its genomic location (genome build 37), with the 450K probe ID shown in parentheses if covered by the array. Results are displayed as the ANOVA F-statistics, accompanying p-values, and per two-group comparison the difference in methylation, and the p-value corrected for multiple comparisons by Tukey's HSD. Methylation data has been corrected for the covariates of age, sex, cell type proportion, and batch number prior to this analysis.

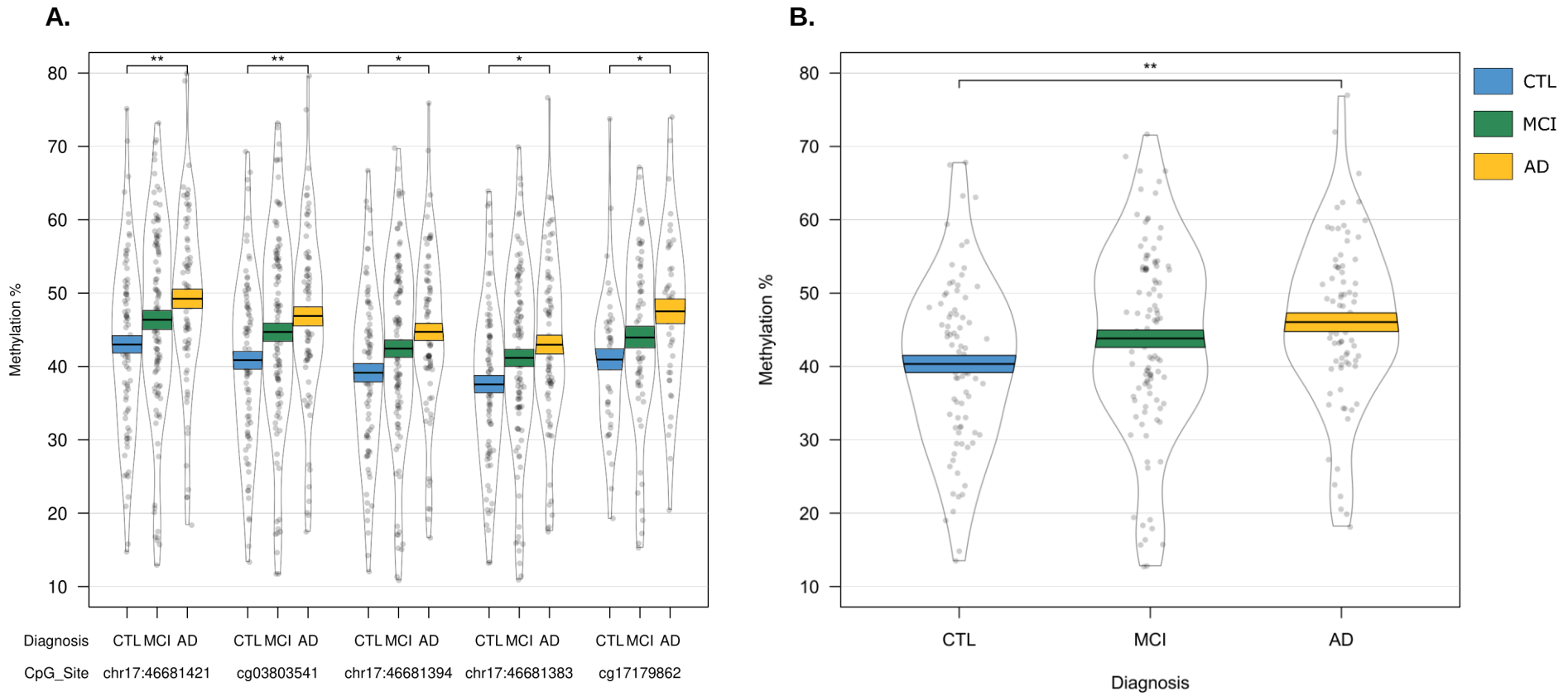

Figure 3.16 Validation of the HOXB6 differentially methylated region (DMR).

DNA methylation was assessed via pyrosequencing and significant changes were found between CTL and individuals with AD, but not MCI at all five CpG sites assessed (A) and when averaged over all five probes (B).

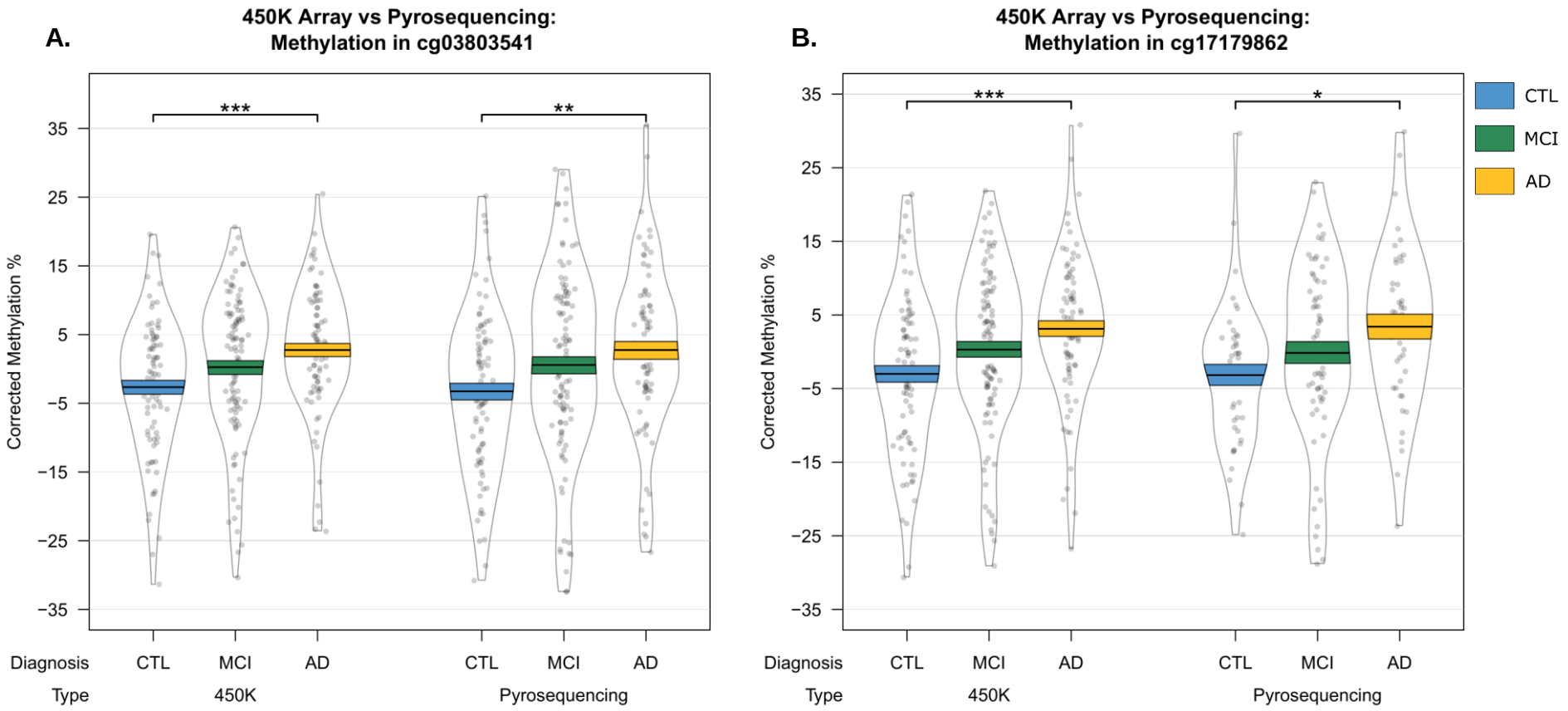

Figure 3.17 Comparison of DNA methylation patterns quantified by the 450K array and pyrosequencing.

DNA methylation levels within the AD-associated HOXB6 DMR were validated using pyrosequencing. Patterns of DNA methylation were similar for cg03803541 (C) and cg17179862 (D), the two probes located within the HOXB6 DMR.

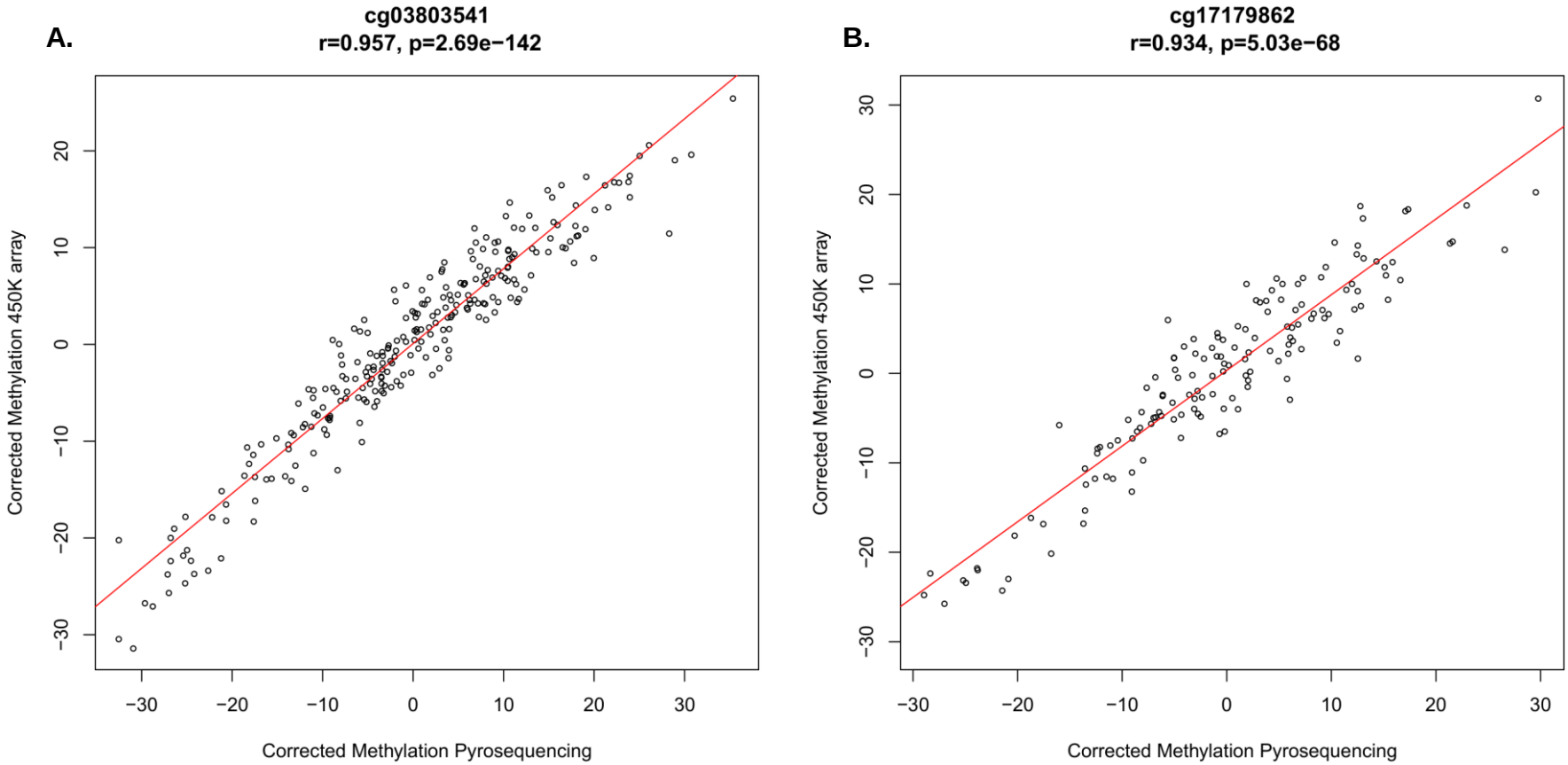

Figure 3.18 Correlations of DNA methylation patterns quantified by the 450K array and pyrosequencing.

DNA methylation levels within the AD-associated HOXB6 DMR were validated using pyrosequencing. The DNA methylation levels quantified by the 450K array and pyrosequencing were significantly correlated for cg03803541 (A) and cg17179862 (B), the two probes located within the HOXB6 DMR.

3.4.4. Transcriptional differences in genes containing DMRs

In order to explore the relationship between DNA methylation and expression, we first assessed whether the expression levels of genes containing the four baseline diagnosis-associated DMRs or the nine conversion DMRs were different in the diagnostic groups. Expression data was only available for two of the six genes we identified as containing DMRs in the baseline group analysis (*HOXB6,* and *CSNK1E* associated with the readthrough transcription region of *TPTEP2-CSNK1E*; Table 3.9). Of these two genes, *CSNK1E,* which had shown increased DNA methylation in MCI samples, also showed a significant difference in gene expression between groups (F = 15.94, p = 3.25×10^{-7}). More specifically, we observed significantly increased mRNA expression in both MCI and AD subjects relative to CTL (Tukey's p = 1.46×10^{-7} and p = 2.66×10^{-3}, respectively; Figure 3.19A). Although there was significantly higher gene expression and DNA methylation (across the DMR), there was no correlation of expression and methylation across all samples, or when we performed correlations separately in the three diagnostic groups (Figure 3.19C; Table 3.10). Although we did not observe any significant differences in gene expression for *HOXB6* (Figure 3.20A), we did find a correlation of expression and methylation when performing correlations in the AD group only (r = -0.24, p = 4.07×10^{-2}; Figure 3.20C; Table 3.10).

Expression data was also available for five of the nine significant DMRs we identified in our analysis of progression from MCI to AD (*GABBR1, PRDM1, FLJ37453, TRIM5,* and *CPT1B/CHKB*). The *CPT1B/CHKB* DMR was covered by three probes on the gene expression microarray, one probe measuring *CPT1B* expression and two probes measuring *CHKB* expression (ILMN_2331205 and

ILMN_1659054). Although none of these genes showed differential expression in MCI subjects who progressed to AD (Table 3.11), *CPT1B/CHKB* showed a significant positive correlation of methylation across the DMR and *CPT1B* gene expression (Table 3.12). The average methylation level across the *CPT1B/CHKB* DMR was significantly correlated with gene expression across all samples (r = 0.40, p = 8.62×10^{-5}, Figure 3.21), which appeared to be primarily driven by a correlation observed in the MCI-MCI samples (r = 0.49, p = 7.27×10^{-5}) and not the MCI-AD samples.

Expression Probe ID	Gene	Chr	ANOVA		CTL vs MCI		CTL vs AD		MCI vs AD	
			F	*p* (F)	Difference (CI)	*p* CvM	Difference (CI)	*p* CvA	Difference (CI)	*p* MvA
ILMN_2415235	*CSNK1E*	22	15.937	3.25E-07	0.18 (0.1 - 0.25)	1.46E-07	0.11 (0.03 - 0.19)	2.66E-03	-0.07 (-0.14 - 0)	0.07
ILMN_1667831	*HOXB6*	17	0.882	0.415	-0.01(-0.05 - 0.03)	0.84	-0.03 (-0.07 - 0.02)	0.39	-0.01 (-0.06 - 0.03)	0.68

Table 3.9 Expression of genes containing DMRs associated with baseline diagnosis.

Changes in the expression of genes containing DMRs associated with a baseline diagnosis of CTL, MCI, or AD. An ANOVA was performed to analyse changes in gene expression levels in these transcripts, whilst controlling for the covariates of age, sex, and cell type proportions. Shown for each probe is the Illumina transcript ID, chromosome (genome build 37), ANOVA F-statistic, accompanying p-value (p (F)), and per two-group comparison the difference in expression, lower and upper confidence interval (CI) and the p-value corrected for multiple comparisons by Tukey's HSD. For the TPTEP2-CSNK1E DMR, expression of CSNK1E was examined, as this locus is a readthrough transcription which encodes the same protein as CSNK1E.

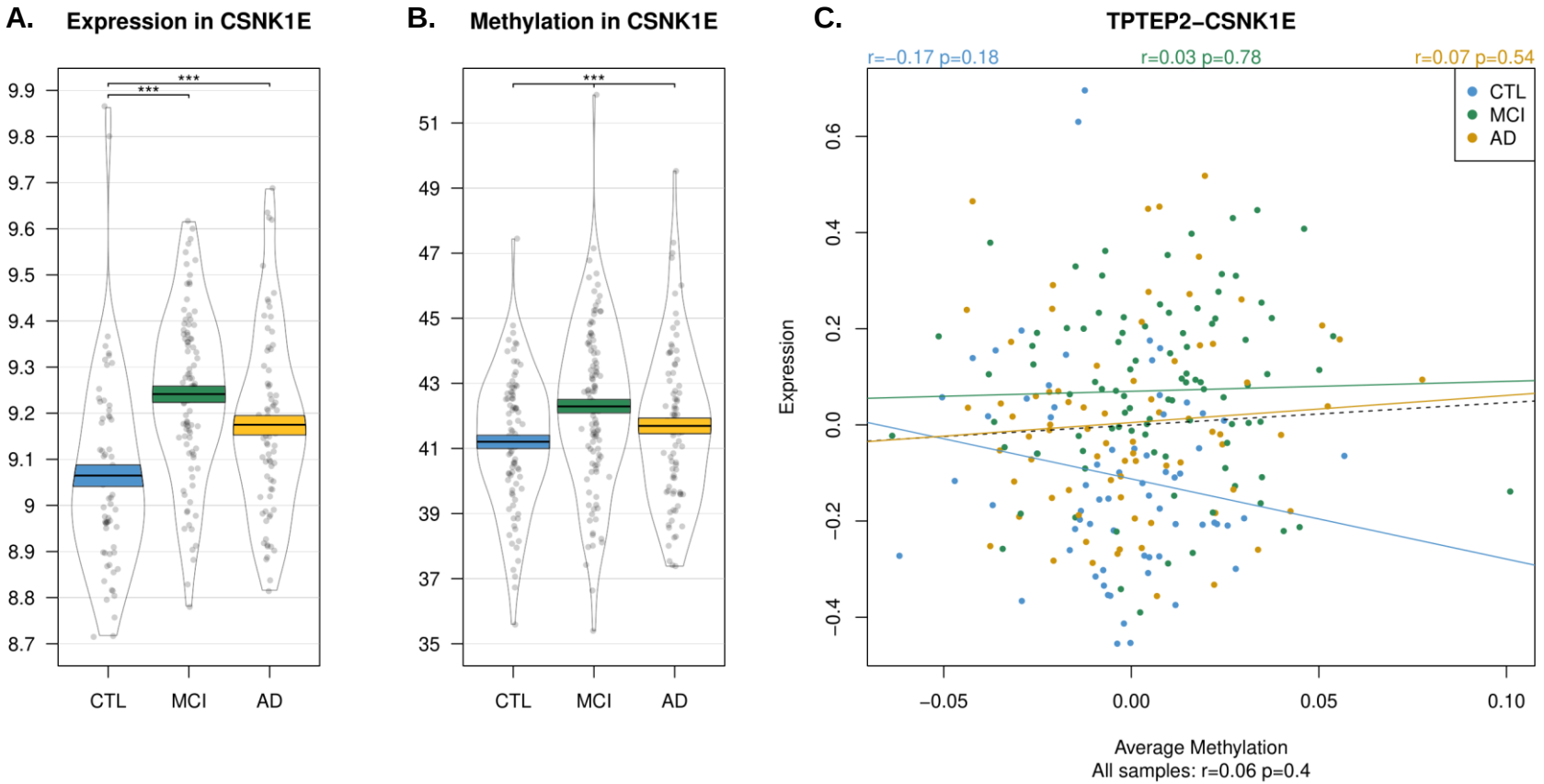

Figure 3.19 Associations between gene expression and DNA methylation in CSNK1E.

(A) *Average gene expression levels of CSNK1E in individuals with AD (yellow), MCI (green), or CTL (blue), controlling for covariates. (B) Average DNA methylation levels of CSNK1E across its associated DMR in AD, MCI and CTL subjects, controlling for covariates. (C) Correlations of DNA methylation and gene expression in CSNK1E, shown by disease group (blue: CTL, green: MCI, yellow: AD), and across all samples (black dotted line).*

Gene	All samples		CTL only		MCI only		AD only	
	r	***p*-value**	***r***	***p*-value**	***r***	***p*-value**	***r***	***p*-value**
HOXB6	-0.090	0.17	-0.170	0.17	0.108	0.30	-0.235	4.07E-02
TPTEP2-CSNK1E; CSNK1E	0.055	0.40	-0.166	0.18	0.029	0.78	0.072	0.54

Table 3.10 Correlation of methylation and expression for HOXB6 and TPTEP2-CSNK1E.

Pearson correlations of gene expression levels to mean DMR methylation values. Genes containing DMRs associated with with a baseline diagnosis of CTL, MCI, or AD were analysed across all three groups. Covariates were regressed out of methylation and expression data prior to performing these correlations. The methylation levels of the TPTEP2-CSNK1E region were correlated to expression of CSNK1E, as this locus is a readthrough transcription which encodes the same protein as CSNK1E.

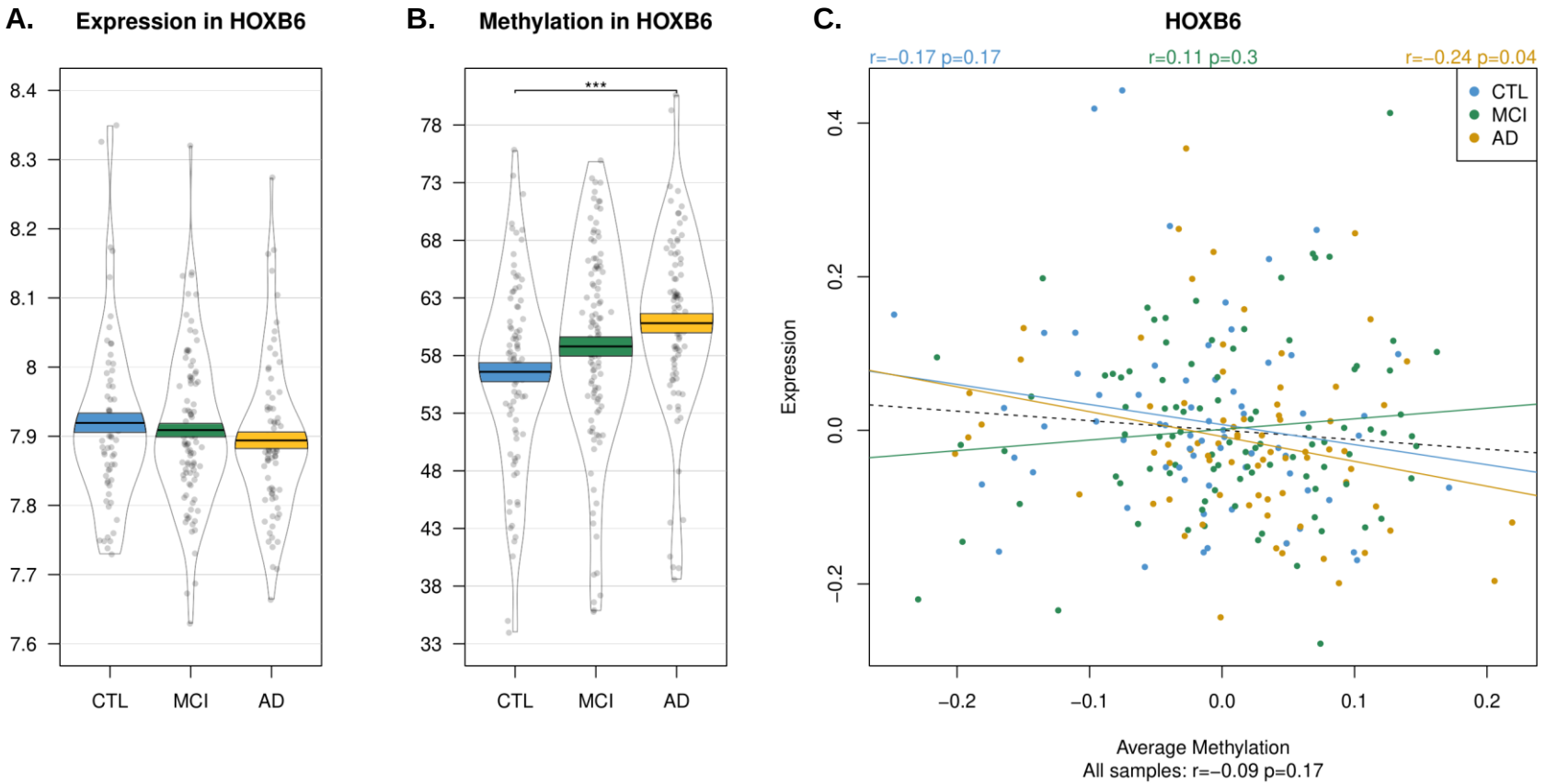

Figure 3.20 Associations between gene expression and DNA methylation in HOXB6.

(A) Average gene expression levels of HOXB6 in individuals with AD (yellow), MCI (green), or CTL (blue), controlling for covariates. (B) Average DNA methylation levels of HOXB6 across its DMR in AD, MCI and CTL subjects, controlling for covariates. (C) Correlations of DNA methylation and gene expression in HOXB6 shown by disease group (blue: CTL, green: MCI, yellow: AD), and across all samples (black dotted line).

ProbeID	Gene	Chromosome	Regression coefficient	*p*-value
ILMN_1791754	*CPT1B*	22	-0.068	0.10
ILMN_1659054	*CHKB*	22	-0.006	0.84
ILMN_2331205	*CHKB*	22	0.002	0.97
ILMN_2404665	*TRIM5*	11	-0.026	0.30
ILMN_2395375	*GABBR1*	6	0.041	0.37
ILMN_1655077	*PRDM1*	6	0.015	0.77
ILMN_1769571	*FLJ37453*	1	-0.006	0.83

Table 3.11 Expression of genes containing DMRs associated with conversion from MCI to AD.

Changes in the expression of genes containing DMRs associated with progression of MCI to AD. A linear regression was performed to analyse changes in gene expression levels in these transcripts, whilst controlling for the covariates of age, sex, cell type proportions, and baseline MMSE score. Shown for each probe is the Illumina transcript ID, chromosome (genome build 37), regression coefficient, and accompanying p-value.

Gene	All samples		MCI-MCI		MCI-AD	
	r	*p*-value	*r*	*p*-value	*r*	*p*-value
CPT1B	0.400	8.62E-05	0.486	7.27E-05	0.122	0.52
CHKB (ILMN_2331205)	-0.194	0.07	-0.221	0.09	-0.163	0.39
CHKB (ILMN_1659054)	-0.179	0.09	-0.139	0.29	-0.302	0.10
GABBR1	-0.114	0.28	-0.184	0.15	0.083	0.66
TRIM5	0.060	0.57	-0.108	0.41	0.225	0.23
PRDM1	0.018	0.87	0.107	0.41	-0.134	0.48
FLJ37453	-0.002	0.99	-0.098	0.45	0.170	0.37

Table 3.12 Correlation of methylation and expression of DMRs associated with progression of MCI to AD.

Pearson correlations of gene expression levels to mean DMR methylation values. Genes containing DMRs associated with conversion from MCI to AD, were analysed across the subset of MCI-MCI and MCI-AD samples. Covariates were regressed out of methylation and expression data prior to performing these correlations.

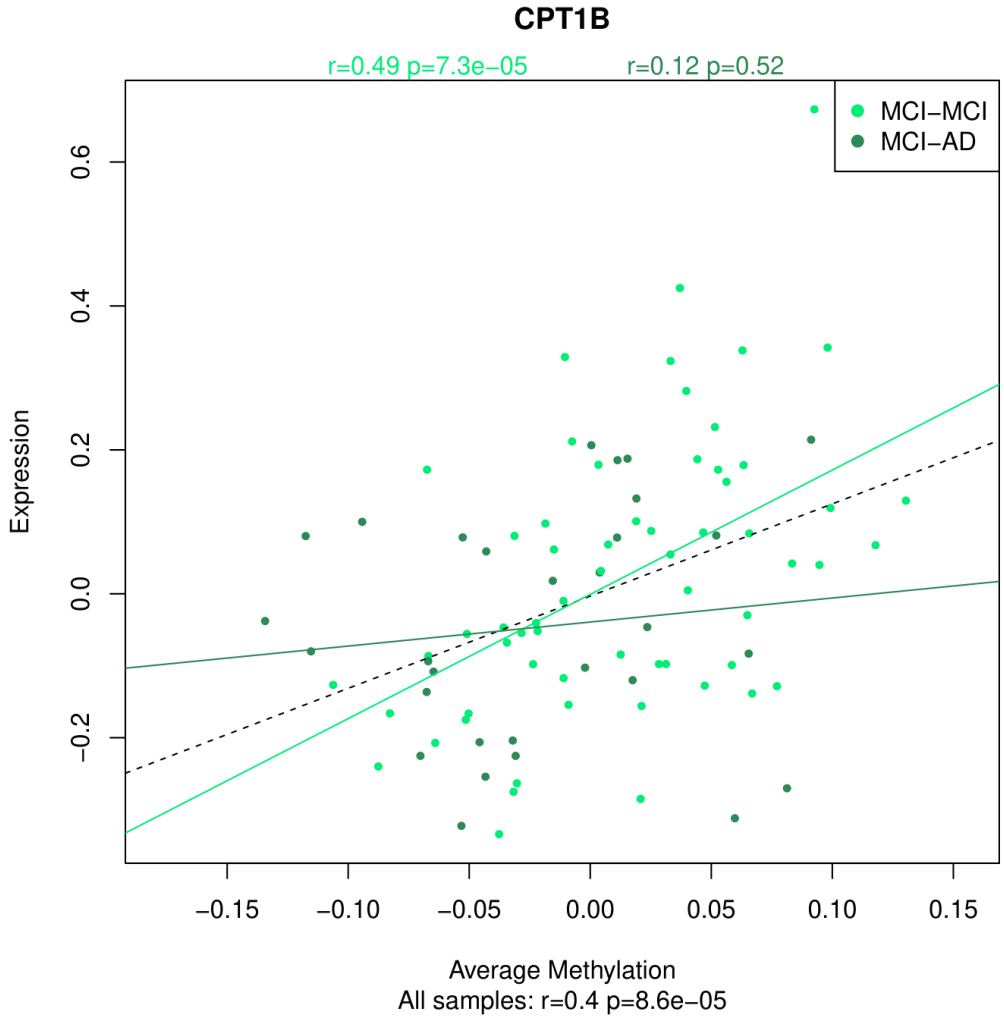

Figure 3.21 Correlation of DNA methylation and gene expression in CPT1B.

Correlations are shown by disease group (light green: MCI-MCI, dark green: MCI-AD), and across all samples (black dotted line). A significant correlation was observed across all samples ($r = 0.4$, $p = 8.62\times10^{-5}$), which was primarily driven by a correlation observed in the MCI-MCI samples ($r = 0.49$, $p = 7.27\times10^{-5}$) and not the MCI-AD samples.

3.4.5. Clusters of methylated loci associated with MCI and AD

To identify clusters of probes that are co-methylated and are therefore hypothesised to share a common function, we performed WGCNA and classified the entire filtered dataset of 200,633 probes into 16 modules (as described in Chapter 2 section 2.4.2.1). These modules were correlated to the group comparisons of diagnostic status at baseline, as well as to several other traits of interest (Figure 3.22), after controlling for covariates. The brown module, which consists of 11,794 probes, was negatively correlated with differences between CTL and MCI (ρ = -0.16, $p = 2.31\times10^{-2}$), and correlated positively with an individual's number of education years (r = 0.13, $p = 3.59\times10^{-2}$). Three more modules also showed a correlation with MCI versus CTL; the lightcyan module consisting of 133 probes (ρ = 0.18, $p = 1.2\times10^{-2}$), and the yellow module which consists of 10,635 probes (ρ = 0.17, $p = 1.51\times10^{-2}$). The yellow module further correlates to the structural imaging variable MET (r = -0.14, $p = 4.26\times10^{-2}$). The purple module (792 probes), also correlates to MCI versus CTL (ρ = -0.17, $p = 1.98\times10^{-2}$), as well as the majority of structural imaging variables: REV (r = 0.21, $p = 3.16\times10^{-3}$), TEV (r = 0.18, $p = 9.85\times10^{-3}$), MET (r = 0.25, $p = 3.22\times10^{-4}$), VV (r = -0.18, $p = 9.25\times10^{-3}$), LHV (r = 0.22, $p = 1.19\times10^{-3}$), RHV (r = 0.20, $p = 3.31\times10^{-3}$), THV (r = 0.22, $p = 1.46\times10^{-3}$), and WBV (r = 0.20, $p = 3.55\times10^{-3}$). Finally, the cyan module (280 probes) correlates to an individual's number of *APOE ε4* alleles (ρ = -0.14, $p = 1.75\times10^{-2}$).

Subsequently, we investigated whether the probes that are integral to a specific module are also the probes that are driving the association with the relevant diagnosis or trait. We did this by correlating and plotting MM and PSvalues, and focusing on those modules that showed positive ($r > 0$, $p < 0.05$) correlations

between MM and PS (Table 3.13). Significant positive MM to PS correlations were found in the brown (r = 0.26, p = 5.93×10^{-179}), purple (r = 0.19, p = 9.18×10^{-8}), and yellow (r = 0.25, p = 6.64×10^{-153}) modules in association with CTL vs MCI. The brown module further showed significant positive MM to PS correlations in relation to education years (r = 0.11, p = 2.40×10^{-32}). The yellow module displayed a positive MM to PS correlation (r = 0.22, p = 6.42×10^{-117}) in association with MET, and the cyan module showed a positive MM to PS correlation in association with the number of *APOE* ε4 alleles (r = 0.20, p = 6.42×10^{-4}). These modules were the primary focus of our pathway analyses. MM and PS plots for these modules are shown in Figure 3.23; for a full overview of all MM and PS correlations, see Table 3.13.

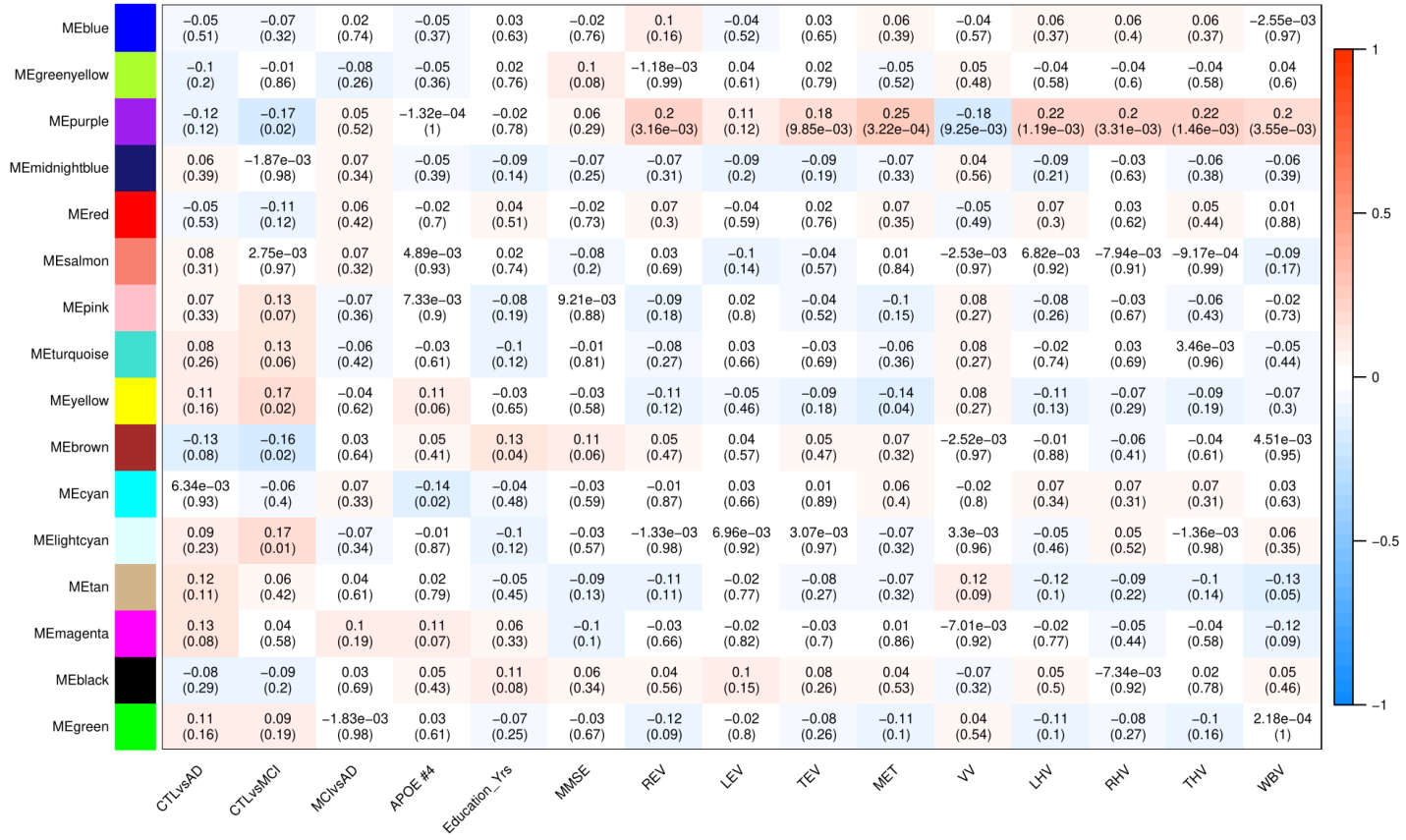

Figure 3.22 Correlations between module eigengenes and traits of interest.

Module names are shown along the y-axis. Correlation estimates are reported, with p-values in parentheses. Spearman correlations were performed for the CTL versus AD comparison, CTL versus MCI comparison, the MCI versus AD comparison, and the number of APOE-ε4 alleles (APOE #4). Pearson correlations were calculated for the number of education years (Education_Yrs), MMSE scores, and structural imaging measurements. Modules with a correlation p-value <0.05 were selected for further analysis

MM-PS correlation	*r*	*p*-value
Trait: CTL vs MCI		
Lightcyan module	0.026	0.76
Yellow module	0.251	6.64E-153
Purple module	0.188	9.18E-08
Brown module	0.258	5.93E-179
Trait: APOE #ε4 alleles		
Cyan module	0.203	6.42E-04
Trait: Education years		
Brown module	0.109	2.40E-32
Trait: REV		
Purple module	-0.222	2.58E-10
Trait: TEV		
Purple module	-0.031	0.38
Trait: MET		
Purple module	-0.249	1.22E-12
Yellow module	0.220	6.42E-117
Trait: WBV		
Purple module	-0.019	0.59
Trait: VV		
Purple module	0.024	0.50
Trait: LHV		
Purple module	-0.075	3.44E-02
Trait: RHV		
Purple module	0.006	0.87
Trait: THV		
Purple module	-0.024	0.51

Table 3.13 Correlations of MM and PS.

Pearson correlations of MM and PS values for each module that showed a significant correlation to a specific trait. Abbreviations: number of APOE ε4 alleles (APOE#4), number of education years (Education years), and the following structural imaging measures: right and total entorhinal volume (REV and TEV, respectively), mean entorhinal thickness (MET),whole brain volume (WBV), ventricular volume (VV), and left, right and total hippocampal volume (LHV, RHV, and THV, respectively

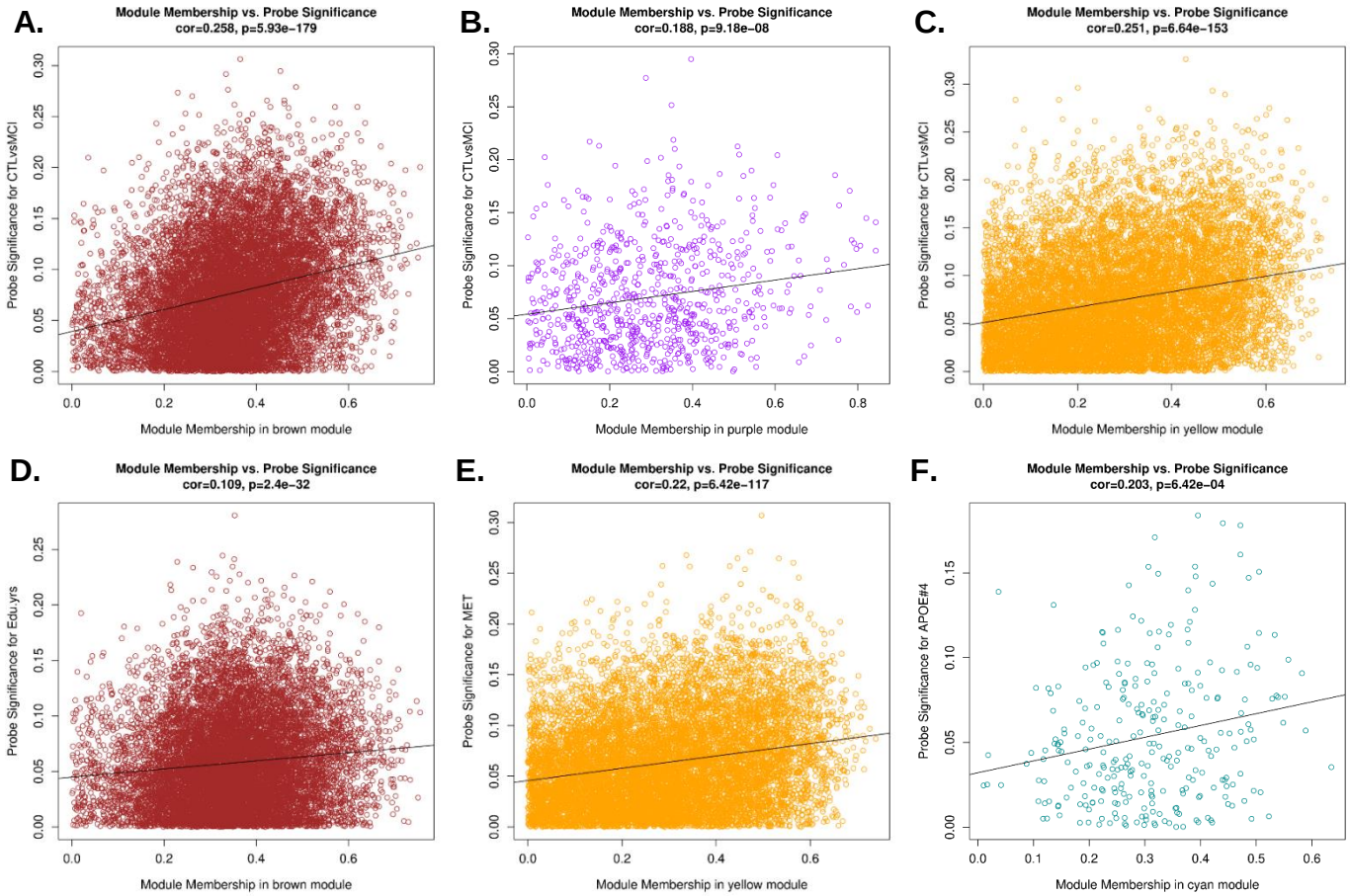

Figure 3.23 Modules showing positive correlations between MM and PS.

Shown are modules that have positive MM to PS correlations, and that were associated with diagnosis or a trait. The (A) brown, (B) purple, and (C) yellow modules were associated with changes MCI subjects relative to CTL. (D) The brown module was associated with the number of education years. The (E) yellow module was associated with MET, and (F) the cyan module associated with the number of APOE ε4 alleles.

3.4.6. Functional role of modules associated with MCI and AD

We sought to identify pathways that were enriched in modules that were affected in disease or were associated with certain traits. For this purpose, we performed GO and KEGG enrichment analyses, with for large (i.e., yellow and brown) modules only the core probes being used for the enrichment analyses. Pathways related to the brown (Figure 3.24), purple, yellow (Figure 3.25), and cyan (Figure 3.26) modules all passed false discovery rate (FDR) multiple testing correction. A large number of GO terms were associated with the core of the brown module, which was related to MCI relative to CTL as well as number of education years, among which were 'extracellular matrix' (q = 4.23×10^{-7}), 'channel activity' (q = 3.19×10^{-5}), and 'passive transmembrane transporter activity' (q = 3.19×10^{-5}; Figure 3.24A). Furthermore, KEGG terms related to this module included 'Protein digestion and absorption' (q = 1.06×10^{-2}), 'Oxytocin signalling pathway' (q = 1.06×10^{-2}), and 'Regulation of actin cytoskeleton' (q = 1.10×10^{-2}; Figure 3.24B). The core of the yellow module showed differences related to MCI, relative to CTL, as well as MET, and we found in our enrichment analyses of the core probes that the top GO terms included 'leukocyte activation' (q = 7.46×10^{-13}), 'cell activation' (q = 7.46×10^{-13}), and 'immune response' (q = 5.84×10^{-11}), whilst the top KEGG terms included 'platelet activation' (q = 1.93×10^{-2}), 'adrenergic signalling in cardiomyocytes' (q = 1.93×10^{-2}), and 'sphingolipid signalling pathway' (q = 2.34×10^{-2}; Figure 3.25). The purple module, which was also associated with differences related to MCI relative to CTL, was connected with one GO term; 'vesicle-mediated transport' (q = 4.35×10^{-2}), but no significant KEGG terms. Finally, the cyan module, which was associated with the number of *APOE ε4* alleles, was related to a number of GO terms, including 'cell activation' (q = 3.07×10^{-4}), 'regulation of cell adhesion' (q = 4.51×10^{-4}), 'leukocyte activation' (q

= 4.51×10^{-4}) and ‘regulation of cell death’ (q = 8.32×10^{-4}; Figure 3.26) and one KEGG pathway: ‘T cell receptor signalling pathway’ (q = 3.73×10^{-2}).

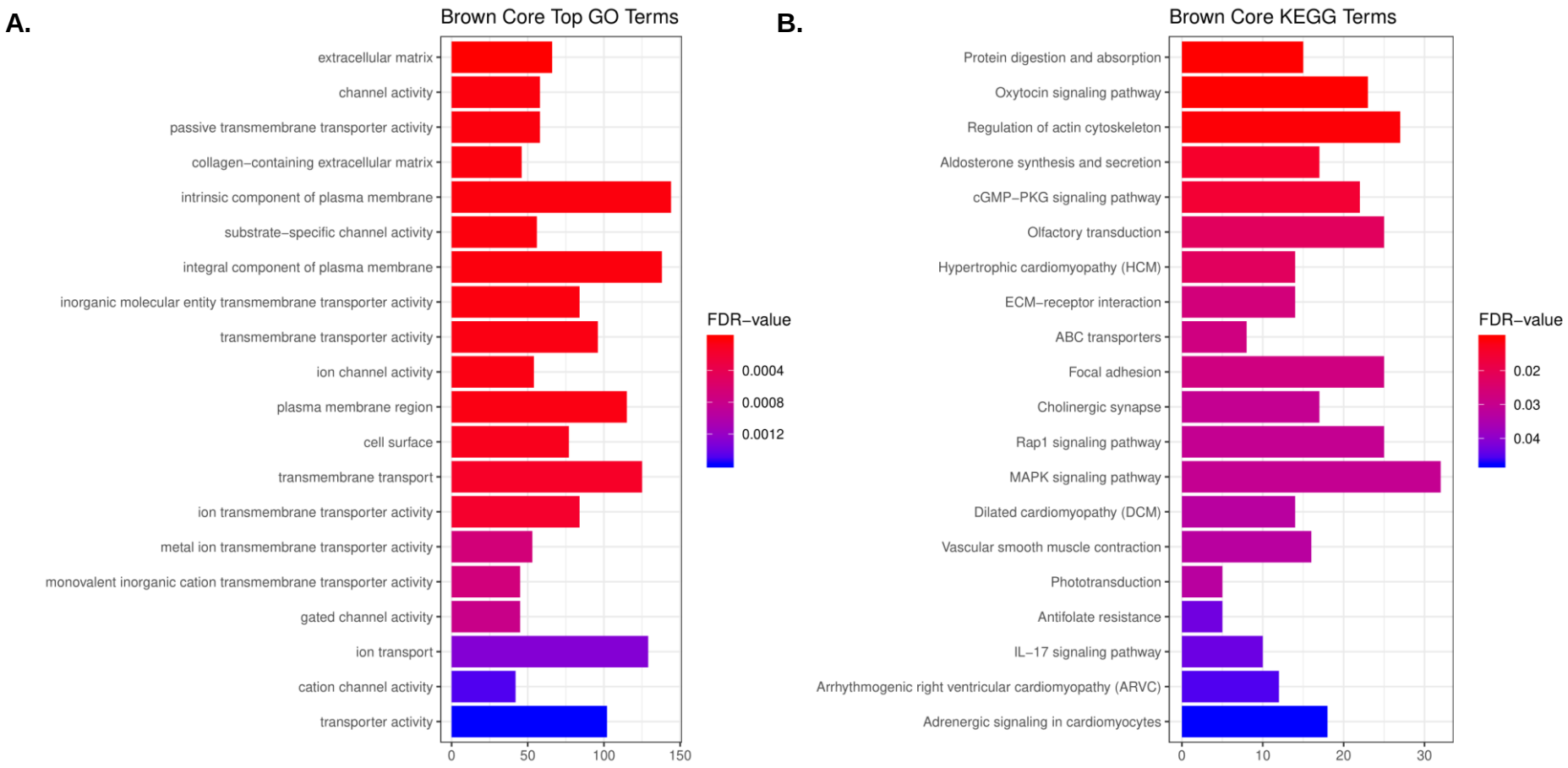

Figure 3.24 Pathways related to the brown MCI- and education years-associated module.

The top 20 significant GO terms (A) and significant KEGG terms (B) related to the core probes within the brown module, which was associated with MCI relative to CTL, and number of education years. The x-axis displays the number of altered genes in the pathway.

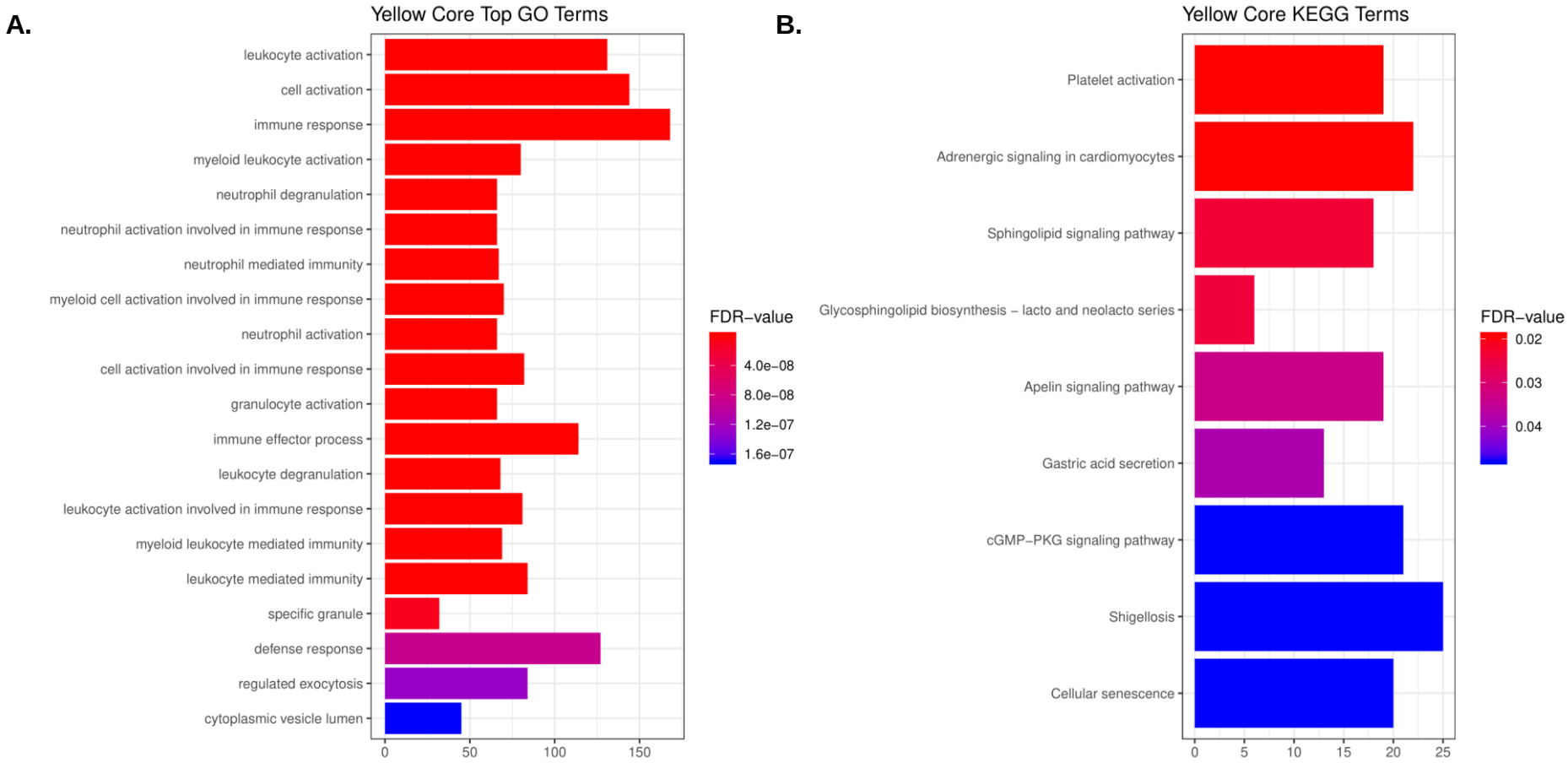

Figure 3.25 Pathways related to the yellow MCI- and MET-associated module.

The top 20 significant GO terms (A) and significant KEGG terms (B) related to the core probes within the yellow module, which was associated with MCI relative to CTL, and the structural imaging measurement of MET. The x-axis displays the number of altered genes in the pathway.

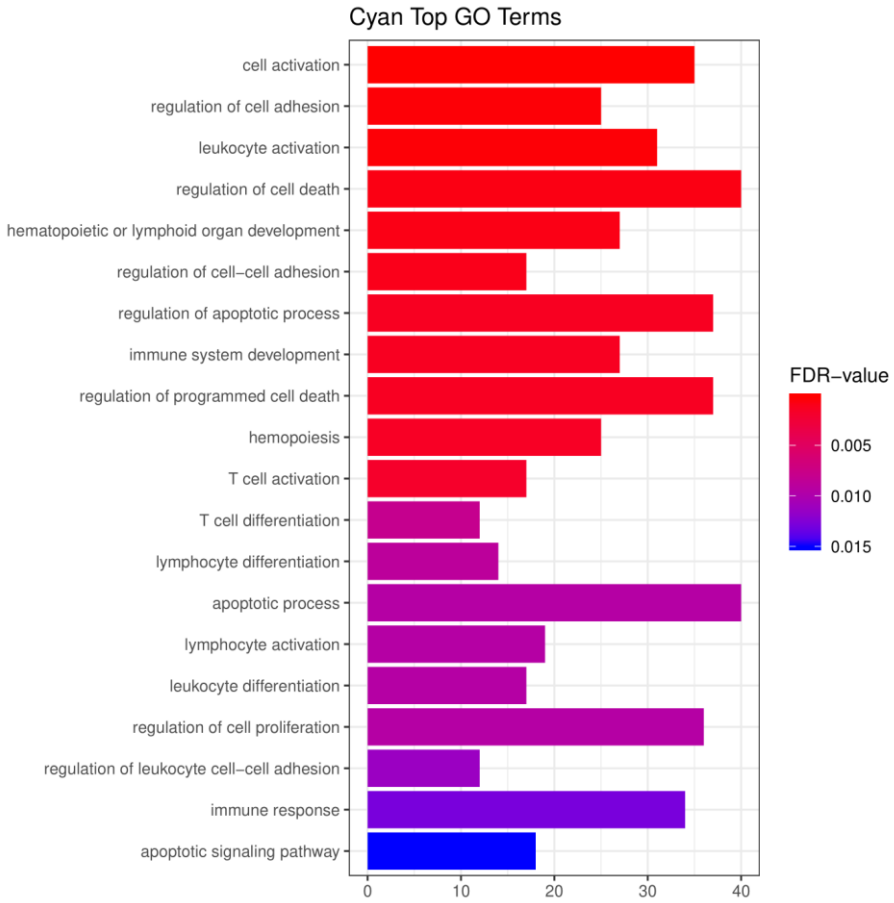

Figure 3.26 Pathways related to the cyan APOE carrier status-associated module.

The top 20 significant GO terms related to the cyan module, which was associated with number of APOE ε4 alleles. The x-axis displays the number of altered genes in the pathway.

3.4.7. Investigating clusters of co-methylated loci associated with progression to AD

In addition to modules associated with baseline diagnosis of MCI and AD, we also identified 31 modules of highly co-methylated loci in the subset of MCI-MCI and MCI-AD samples (as described in Chapter 2 section 2.4.2.1.). Only one of these modules, the orange module, was shown to be significantly associated with future progression to AD (β = -0.04, p = 4.38×10^{-2}). We then correlated the MM to the PS for this module and found a significant positive correlation (r = 0.36, p = 9.40×10^{-6}). Following GO and KEGG pathway analysis, we found no GO terms passing FDR multiple testing correction, but top KEGG terms included 'renal cell carcinoma' (q = 1.21×10^{-2}), 'non-homologous end-joining' (q = 2.00×10^{-2}), and 'ErbB signalling pathway' (q = 2.00×10^{-2}; Figure 3.27).

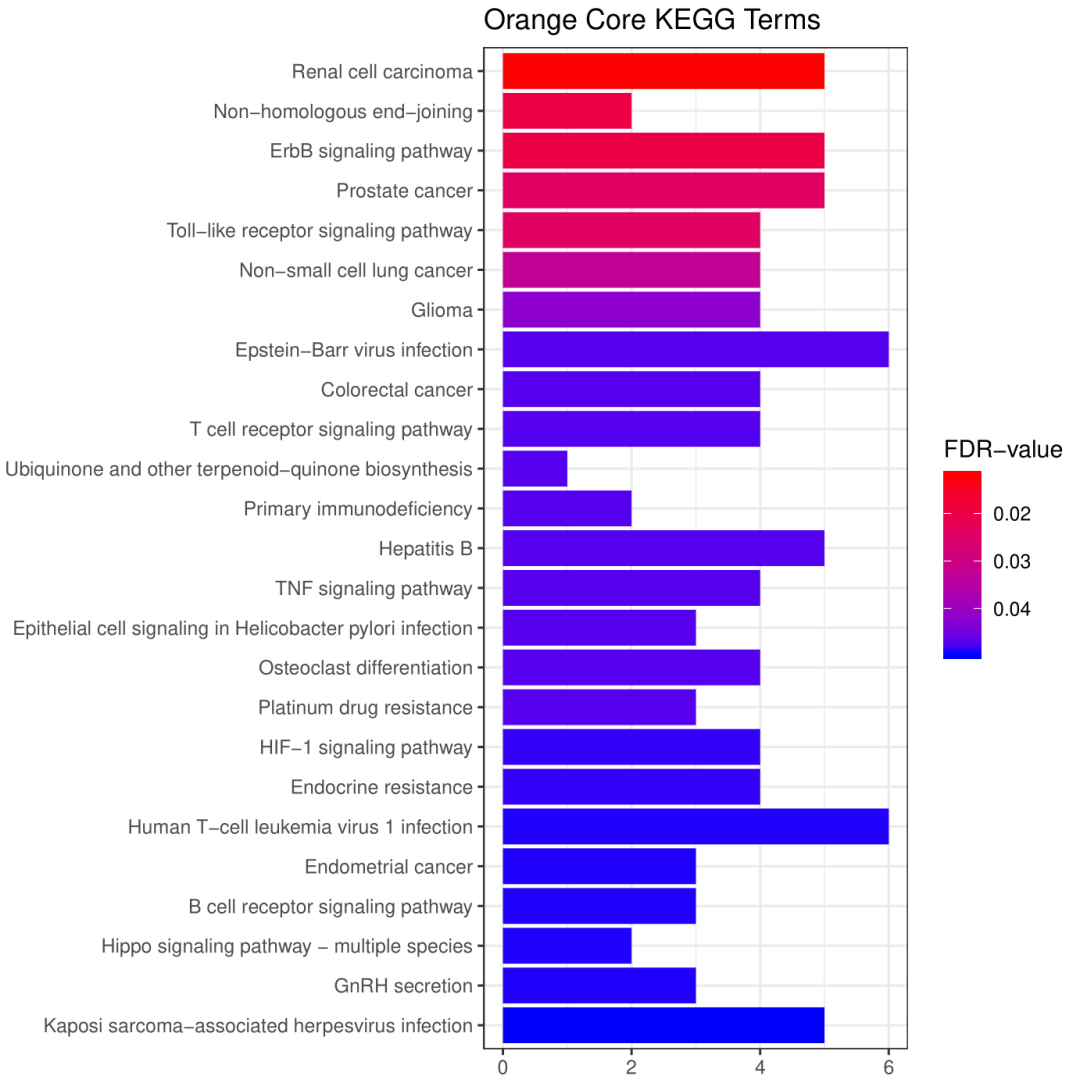

Figure 3.27 Significant KEGG terms related to the MCI to AD conversion-associated orange module.

KEGG terms are related to the probes within the orange module, which was associated with conversion from MCI to AD. The x-axis displays the number of altered genes in the pathway.

3.5. Discussion

The present study, which reflects the first large-scale EWAS of AD blood samples, identified epigenetic signatures related to AD and MCI, as well as signatures associated with future conversion from MCI to AD.

The *HOXB6* gene contained a DMR that reflected differences in methylation in AD relative to CTL, which were validated using pyrosequencing. *HOXB6* encodes the homeobox protein B6, which is part of a larger cluster of homeobox B genes located on chromosome 17. Homeobox genes are DNA binding proteins that have been implicated in early body morphogenesis as well as hematopoietic development. Specifically, *HOXB6* has been shown to be required for normal generation of granulocytes and monocytes (Giampaolo et al., 2002). Interestingly, a recent EWAS of AD hippocampus has shown DNA methylation differences in cg17179862, which was the most significant probe in the DMR we identified and validated (Altuna et al., 2019). The study by Altuna *et al.* further showed that increased methylation at this locus was positively correlated with tau burden.

MOV10L1, which was associated with differences between all three groups, encodes an RNA helicase. This protein was shown to be crucial for the production of Piwi-interacting RNAs (piRNAs) by Vourekas *et al.* (2015). PiRNAs represent small non-coding RNAs involved in epigenetic regulation, which can bind to PIWI proteins and may induce gene silencing via DNA methylation (Aravin et al., 2008; Girard et al., 2006), or RNA-cleavage (for a review see Luteijn and Ketting (2013)). Although initially believed to be mainly present in germline cells, piRNAs

have been shown to be stably expressed in human blood (X. Yang et al., 2015), and have also been shown to be downregulated in tumour tissue and upregulated in blood of renal carcinomas (Iliev et al., 2016). Interestingly, Watson *et al.* (2016) performed an EWAS of AD superior temporal gyrus and identified a DMR spanning 13 probes, including all ten probes we identified in the *MOV10L1* gene in the current study. Of note, where Watson *et al.* detected AD-related hypermethylation in these ten probes, we found hypermethylation in MCI when compared to AD and CTL individuals, while methylation levels of AD subjects were not distinct from CTL individuals.

Of the nine DMRs that were related to future conversion to AD, our most significant region was located in *CPT1B*, which encodes the protein carnitine palmitoyltransferase 1B. Differential DNA methylation in *CPT1B* has been previously identified in blood and foetal cortex of Down syndrome patients (El Hajj et al., 2016; Kerkel et al., 2010). This is interesting as individuals with Down syndrome often develop AD as a result of trisomy of chromosome 21, causing them to have an additional copy of the *APP* gene. The study by El Hajj *et al.* (2016) identified a DMR in *CPT1B* consisting of 18 probes in Down syndrome foetal cortex samples, which spanned the region discovered in the present study. They detected hypermethylation in 13 probes in Down syndrome, whilst we observed hypomethylation in those MCI individuals who convert to AD. Kerkel *et al.* (2010) similarly detected hypermethylation at one CpG site in our *CPT1B* DMR in peripheral blood leukocytes of individuals with Down syndrome, concomitant with significant overexpression of the gene. While we observed hypomethylation of the DMR, the positive relationship found between methylation and expression for this region was replicated in our study. Of note, overexpression of *CPT1B* has

also been found in blood from soldiers with post-traumatic stress disorder (Zhang et al., 2015), a known risk factor for developing AD (Agís-Balboa et al., 2017; Yaffe et al., 2010). Interestingly, no overlap was found between DMRs associated with conversion and DMRs related to diagnosis at baseline. This may reflect limited power in our MCI conversion analysis due to sample size, or could reflect temporal patterns of DNA methylation in the process of conversion from MCI to AD.

In addition to DMRs, by using WGCNA and subsequent pathway analyses, we further identified biological mechanisms affected in disease. The cyan module which was linked to the number of *APOE* ε4 alleles, is involved in GO pathways related to the immune system, which is interesting given that the immune system is known to be activated in AD (Heppner et al., 2015), and as *APOE* ε4 is the strongest genetic risk factor for sporadic AD (Lambert et al., 2013). The core of the brown module which reflects methylomic differences related to an individual's number of education years, and differences in MCI relative to CTL, was shown to be involved in transmembrane processes (GO), as well as oxytocin signalling (KEGG). The oxytocin signalling pathway is linked to social behaviours, as well as several psychiatric disorders (e.g., depression, Feldman et al., 2016) Interestingly, a DMR was recently identified in the oxytocin gene (*OXT*), which was hypomethylated in the AD brain (Lardenoije et al., 2019; Watson et al., 2016) and hypermethylated in the blood in individuals who subsequently converted to AD (Lardenoije et al., 2019). Oxytocin is involved in the modulation of stress, social behaviours, and associative learning (Olff et al., 2013), and altered levels of oxytocin have been reported in AD post-mortem brain tissue (Mazurek et al., 1987) and CSF (North et al., 1992). It is interesting that the sphingolipid signalling

pathway is found in the KEGG results from the core probes of the yellow module related to differences in MCI relative to CTL. Multiple studies have indicated that sphingolipid signalling pathways are implicated in AD (Crivelli et al., 2020; Giovagnoni et al., 2021), and the measurement of lipids in the pathway is being explored as a potential biomarker of AD and neurodegeneration (Mielke & Lyketsos, 2010). Similar to the cyan module, the majority of GO terms in the core of the yellow module are related to various processes of immune activation. In our network analysis examining conversion from MCI to AD, we identified a pathway in the orange module associated with non-homologous end-joining. Non-homologous end-joining activity is involved in repairing double strand DNA breaks, and has been reported to be decreased in AD brain (Kanungo, 2013; Shackelford, 2006).

In summary, this is the first EWAS to identify epigenetic signatures and functional pathways specific to MCI, AD, and conversion to AD in blood. However, there are some limitations to our study. First, we have profiled DNA methylation patterns in whole blood and it is known that there are subtle alterations in the abundance of specific blood cell types in MCI and AD (Lunnon et al., 2012). Although we have controlled for the proportions of these different cells, it will be of interest to investigate disease-associated signatures in individual cell types. Second, individuals were only followed up clinically for up to two years following the baseline assessment and further studies should profile cohorts consisting of CTL and MCI subjects with long term clinical follow up to identify preclinical changes. In addition, biomarkers were not available to support the clinical diagnosis of AD. Third, our comparisons of DNA methylation and gene expression were limited to only those genes with variable expression levels in the previous study (Lunnon

et al., 2012) and did not examine transcript variants. Fourth, we have not replicated our findings in an independent study cohort. Although we did validate our *HOXB6* DMR in the same samples using an alternative technology, in the future it will be interesting to verify the loci we identified in a different set of samples. Finally, although there is some communication between the brain and the blood, not all differences found to be associated with AD in the blood may be functionally related to the processes taking place in the brain. Differences in DNA methylation may be the result of parallel effects or comorbidities, and may not be causally related to disease, but could reflect mediating or downstream effects. It would be interesting for future studies to explore the exact role of the epigenetic signatures identified in this study, and to explore their potential as biomarkers for an early diagnosis of AD and therapeutic targets.

CHAPTER 4. DNA METHYLOMIC PATTERNS IN BLOOD ASSOCIATED WITH AGE, SEX, AND ALZHEIMER'S DISEASE

4.1. Introduction

A multitude of risk factors have been identified and studied in the field of AD research (see Chapter 1.1.5). When discussing risk factors for AD, the focus is often directed at genetic risk factors such as the *APOE* gene, followed by modifiable risk factors (e.g. diet, smoking). However, age poses by far the greatest risk for developing AD ("2020 Alzheimer's Disease Facts and Figures," 2020; J. H. Chen et al., 2009). In a US study conducted in 2010 it was estimated that 3% of all individuals aged 65 - 74 were affected by AD at that time, compared to 17.6% of those aged 75 - 84 years, and 32.3% of those over 85 years old (Hebert et al., 2013). An earlier meta-analysis had found that the incidence rate of AD triples every five years before the age of 64, doubles before the age of 75, and shows a 1.5 times increase around age 86 (Gao et al., 1998). The fact that the disease, even in those genetically predisposed, usually does not develop until middle age, indicates that age is an important factor in the development of AD. It should be noted, however, that AD is not simply a consequence of normal ageing (Nelson et al., 2011).

Ageing is also a major influencer of DNA methylation levels in multiple tissues including blood cells. Early studies found global hypomethylation and region-specific changes related to ageing (Ahuja & Issa, 1986; Bjornsson et al., 2008), though this seemed to be location- and context-dependent (Christensen et al., 2009; Jones et al., 2015). Several cross-sectional studies have shown that patterns of DNA methylation are distinctly different over time and across tissues (Hernandez et al., 2011; Heyn et al., 2012). In this respect, the increased variability of the methylomic landscape as a consequence of ageing is also called

‘epigenetic drift’. Diverging patterns of DNA methylation over time have been found in monozygotic twin studies. Fraga *et al.* (2005) found that while monozygotic twins were very similar in terms of DNA methylation patterns at a young age, they were epigenetically more different at an older age. This indicates that these changes can occur independent of genotype and may be influenced by other factors.

Alterations in DNA methylation with ageing are robust enough to predict chronological age (Hannum et al., 2013; Horvath, 2013), and can be used to derive a surrogate measure of accelerated or reduced biological ageing, where an individual’s predicted methylation age deviates from their chronological age. In the development of the DNA methylation age calculator by Hannum *et al.* (2013), the authors found that the methylome of men aged 4% faster than the methylome of women. Furthermore, researchers have found that an increased DNA methylation age is associated with mortality (B. H. Chen et al., 2016; Christiansen et al., 2016), and that accelerated biological ageing in blood is associated with higher mortality risk, even when controlling for various health and lifestyle influences, including *APOE* genotype (Marioni et al., 2015). While accelerated methylation ageing has been associated with various risk factors for AD (McCartney et al., 2018), a recent study by Sibbett et al. found no evidence for a positive relationship between accelerated methylated ageing and risk for dementia (Sibbett et al., 2020).

In addition to age, an individual’s sex may constitute another risk factor for AD, although this is subject to debate. Sex differences have been found in both pathological changes and clinical symptoms of AD (Mazure & Swendsen, 2016),

and while it is clear that more women suffer from AD than men ("2020 Alzheimer's Disease Facts and Figures," 2020), the cause of this remains unclear. Many factors could influence this difference in prevalence. First and foremost, the slightly longer lifespan of women compared to men has been suggested to play an important role (Austad, 2006). As age is the greatest risk factor for AD, this could explain why the disease is more prevalent in women. However, studies of sex differences in AD prevalence among individuals of the same age have led to mixed findings. Some studies found no sex differences at any given age (Hebert et al., 2001; Kawas et al., 2000), while others have found a higher incidence in women (Fratiglioni et al., 1997; Gao et al., 1998; Letenneur et al., 1999), or even in men (Matthews et al., 2016). Of note, most studies that did not find any differences were conducted in US populations, while other studies that did find differences were conducted in European populations. This has led researchers to suggest that sex differences may be dependent on geographic region and/or time period (Mielke et al., 2018). Therefore, it is especially important to take sex differences into account when studying European populations.

Other explanations for sex differences seen in AD may be an unbalanced distribution of other established risk factors (see Chapter 1.1.5). Limited education constitutes a risk factor for AD and has been hypothesised to influence sex differences (Rocca et al., 2014), since educational inequality has historically affected women more than men. Comparably, a study by Chêne *et al.* (2015) found a selective survival effect in men, with men older than 65 years of age having a healthier cardiovascular risk profile, thus leading to a lower risk for AD. Furthermore, the largest genetic risk factor for AD, the *APOE*- ε4 genotype, has been found to have sex- and age-specific effects. Neu *et al.* (2017) found that,

while there were no sex differences in risk of developing AD overall, women between the ages of 65 and 75 carrying one *APOE* ε4 allele were at an increased risk of developing AD compared to men, whilst women between the ages of 55 and 70 carrying one *APOE* ε4 allele were more at risk of developing MCI than men were. Independently of sex, a recent study has found that the *APOE* genotype affects DNA methylation patterns in blood (Walker et al., 2021). The interplay between these different risk factors and potential sex effects is very interesting, though more research is required to unravel the underlying causal mechanisms.

Sex, like age, is also associated with DNA methylation differences across and in different tissue types (Liu et al., 2010; McCarthy et al., 2014), including blood (Singmann et al., 2015). A recent study by McCartney *et al.* (2019) found age-by-sex interaction effects in DNA methylation patterns in blood on both autosomal and X chromosome loci.

These findings, taken together with the importance of both age and sex in the context of AD, highlight that it is important for AD research to study sex- and age-specific effects, and take into account their interaction effects on DNA methylation patterns.

4.2. Aims

The aims of this chapter are:

1. To identify differential DNA methylation patterns associated with sex, age, and the interactions between sex, age, and diagnosis of AD and MCI in the AddNeuroMed cohort.
2. To replicate the findings from Aim 1 in the ADNI cohort.
3. To identify biological pathways that are associated with sex, age, and the interactions between sex, age, and diagnosis of AD and MCI in the AddNeuroMed cohort.

4.3. Materials and methods

This chapter uses the epigenome-wide DNA methylation data from the AddNeuroMed cohort, which is described in Chapter 2.1, as well as data from the ADNI cohort, described in Chapter 2.3. Analyses have been run to identify differentially methylated positions, regions, and pathways associated with sex, age, and their interaction with disease within the AddNeuroMed cohort. Additionally, a replication analysis for DMPs and DMRs was carried out in the ADNI cohort.

4.3.1. Sex, age, and disease interaction model for identifying DMPs and DMRs

Prior to carrying out an ANOVA for the identification of DMPs, normalised data from the AddNeuroMed cohort was corrected for covariates as described in Chapter 2.4.1. As formal education has been suggested to influence the sex-associated risk for AD, we tested whether the number of education years differed between males and females in the AddNeuroMed dataset. No significant difference was found (t (261) = -0.998, p = 0.32), and therefore education years was not included as a covariate. As sex and age are both variables of interest in this chapter, rather than covariates, only the variables of cell type proportion and bisulfite treatment batch were regressed out of the data. To investigate whether any single locus or region was associated with sex, age, and their interaction with each other and with disease, an interaction model was run on the data as follows. A type III ANOVA was used, which included the main terms of sex, age, and diagnosis, and the interaction effects of sex and age, sex and diagnosis, and age and diagnosis. Q-Q plots were inspected for inflation, and the *comb-p* module

was applied using a 1000 bp sliding window to identify DMRs, as described in 2.4.1.

4.3.2. Replication of DMPs and DMRs in ADNI

With the aim of replicating DMPs and DMRs identified in the AddNeuroMed data, pre-processed and normalised ADNI data (see Chapter 2.3.2) was analysed in a manner comparable to the analysis described in 4.3.1. The cell type proportion covariates were first regressed out of the normalised ADNI data, after which an ANOVA identical to the model described in section 4.3.1 was carried out, and DMRs were identified using identical *comb-p* settings to those described in Chapter 2.4.1. The overlap in DMPs and DMRs was examined, and for each effect of interest (i.e. the main effects of age and sex, or the interaction effects of sex and age, sex and diagnosis, or age and diagnosis), the top 10 most significant DMPs found in AddNeuroMed were extracted from the ADNI results. The probes within the top DMRs found in AddNeuroMed (for each effect of interest) were also extracted from the ADNI results.

4.3.3. Identification of altered biological pathways

To examine whether any pathways were altered in relation to sex, age, diagnosis, or their interactions, the previously generated WGCNA modules were used (that were also utilised in Chapter 3). These modules had been generated on the full DNA methylation data from the AddNeuroMed cohort (see also Chapter 2.4.2 for a description of the WGCNA methods), and represent clusters of co-methylated sites within the blood dataset. First, the MEs which had previously been calculated for these modules were corrected for the covariates of cell type proportion and batch. Following regression, the module eigengenes were

inspected for extreme outliers (MEs exceeding the threshold of mean ME ± 5 × SD). Two extreme outliers were found and set to NA. On these corrected MEs, the same type III ANOVA model was run as described in 4.3.1, in order to identify which modules were associated with sex, age, or the interactions between sex and age, sex and diagnosis, or age and diagnosis. Consistent with the pipeline used in Chapter 3, modules showing a nominally significant ($p < 0.05$) association with any of these traits were considered for further analysis. For each module associated with a trait, the module membership (MM; i.e. the strength of association for a probe to the module) and probe significance (PS; i.e. the association between a probe and the trait of interest) were calculated. MM values were correlated to PS values, and only the modules that showed significant positive correlations ($r > 0$, $p < 0.05$) were taken forward into pathway analyses. Pathway analyses were carried out using the *missMethyl* package, as described in 2.4.2.1.

4.4. Results

4.4.1. The effects of sex and age on DNA methylation

Focusing on the main effects that sex and age have on DNA methylation in blood, there were numerous nominally significant ($p < 0.05$) DMPs associated with both variables in the AddNeuroMed cohort. For the effect of sex, the most significant DMP (probe cg21437028, $p = 6.61\times10^{-8}$) passed the established epigenome-wide significance threshold of 2.4×10^{-7} (Saffari et al., 2018). This DMP was located in an intron of the *SHOX2* gene on chromosome 3, and while significant, it showed a very small effect of hypomethylation in males compared to females (Figure 4.1A). The top 10 most significant DMPs associated with sex are shown in Table 4.1. With regards to age, there were two DMPs that passed the epigenome-wide significance threshold. The top DMP was probe cg02843237 ($p = 9.07\times10^{-8}$), and is annotated to the gene *ADM*. A decrease in DNA methylation was observed at this site with an increasing age of subjects (Figure 4.2 A). The second DMP which showed epigenome-wide significance was located at the cg00664416 probe, in the *MANEAL* gene located on chromosome 1. This site showed hypermethylation with an increasing age of subjects (Figure 4.2, Figure 4.3 A). The top 10 most significant DMPs associated with age are displayed in Table 4.2. It should be noted that an inflation of *p*-values was observed in the Q-Q plots of the main effects, particularly concerning the effects of age (Figure 4.4).

With the *p*-value inflation in mind, as well as the general importance of replication, the replication of these results was explored in the ADNI cohort. As shown in Figure 4.1B, the epigenome-wide significant DMP related to sex did show a similar direction of effect in ADNI, but this effect was not nominally significant (*p*

> 0.05). None of the other top 10 DMPs related to sex in the AddNeuroMed cohort reached nominal significance in the ADNI cohort (Table 4.1). The same holds true for the top 10 DMPs associated with age, i.e. none of those identified in the AddNeuroMed cohort were nominally significant in the ADNI replication cohort, though a small but similar direction of effect was seen for the top two DMPs (Figure 4.2 B and Figure 4.3 B; Table 4.2).

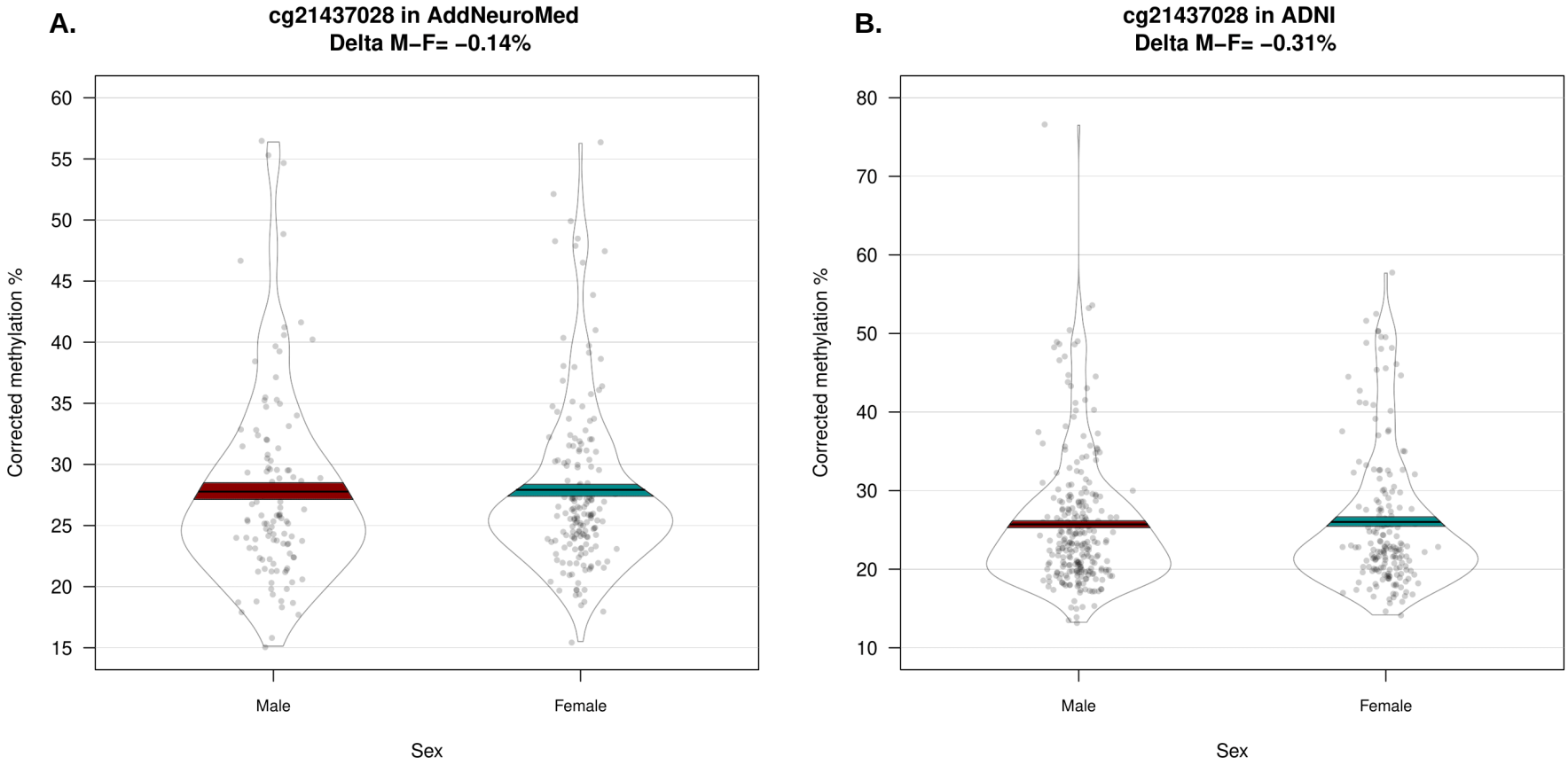

Figure 4.1 The top DMP associated with sex in the AddNeuroMed cohort.

The cg21437028 probe which is annotated to the SHOX2 gene showed epigenome-wide significance in the AddNeuroMed cohort (A) but not in the ADNI cohort (B). Displayed methylation values have been corrected for the covariates of cell type proportions and batch effects, as well as the effects of age and diagnosis.

Top 10 sex DMPs										
ProbeID	**Position**	**AddNeuroMed**			**ADNI**			**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
		F	***p* (F)**	**Δ M-F**	**F**	***p* (F)**	**Δ M-F**			
cg21437028	chr3: 157816403	30.846	6.61E-08	↓	0.973	0.32	↓	*SHOX2*	Body	*VEPH1* (-595268), *SHOX2* (+7548)
cg24560809	chr2: 95939339	23.233	2.38E-06	↓	0.086	0.77	↑	*PROM2*	TSS1500	*PROM2* (-861)
cg03604067	chr7: 1980273	22.166	3.98E-06	↑	0.876	0.35	↑	*MAD1L1*	Body	*ELFN1* (+231476), *MAD1L1* (+292309)
cg02153041	chr12: 131502934	20.428	9.22E-06	↓	0.001	0.98	↓	*GPR133*	Body	*SFSWAP* (-692700), *GPR133* (+64483)
cg12131219	chr1: 25558575	20.194	1.03E-05	↑	0.6	0.44	↑	*SYF2*	Body	*SYF2* (+437)
cg12855310	chr10: 134215035	19.421	1.51E-05	↑	0.1	0.75	↓	*PWWP2B*	Body	*INPP5A* (-136317), *STK32C* (-93559)
cg26298737	chr1: 225964937	19.317	1.58E-05	↓	0.192	0.66	↓	*SRP9*	TSS1500	*SRP9* (-577)
cg25274468	chr7: 114570498	18.718	2.13E-05	↑	1.232	0.27	↓	*MDFIC*	Body	*MDFIC* (+8290)
cg01319471	chr19: 12759546	18.064	2.93E-05	↑	3.176	0.08	↑	*MAN2B1*	Body	*MAN2B1* (+18044), *ZNF791* (+37815)
cg11128045	chr9: 140395219	17.907	3.17E-05	↓	-	-	-	*PNPLA7*	Body	*NELF* (-41434), *PNPLA7* (+49766)

Table 4.1 Top 10 DMPs associated with sex in the AddNeuroMed cohort.

Shown for each probe is the genomic location, ANOVA F-statistic, associated p-value, and the direction of change (Δ M-F; ↓ = lower levels of methylation in males, ↑ = higher methylation levels in males) in the AddNeuroMed cohort as well as the ADNI replication cohort. Probes are annotated using UCSC gene name and gene group (genome build 37), and GREAT annotation genes with the TSS distance indicated in parentheses.

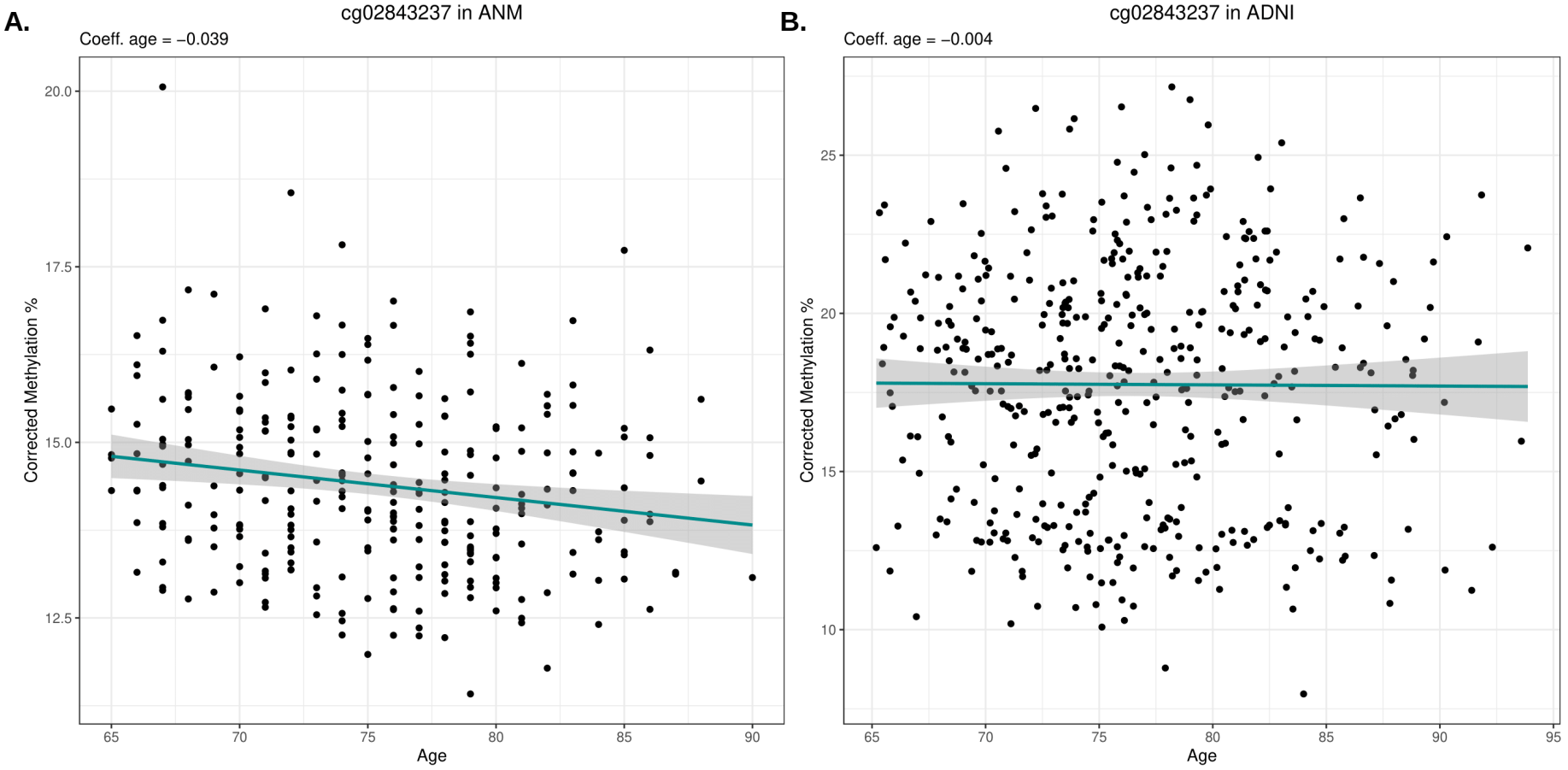

Figure 4.2 The cg02843237 DMP associated with age in the AddNeuroMed cohort.

The cg02843237 probe showed hypomethylation with increasing age in AddNeuroMed (A), and is annotated to the ADM gene. This effect was not seen in the ADNI replication cohort (B). Displayed methylation values have been corrected for the covariates of cell type proportions and batch effects, as well as the effects of sex and diagnosis.

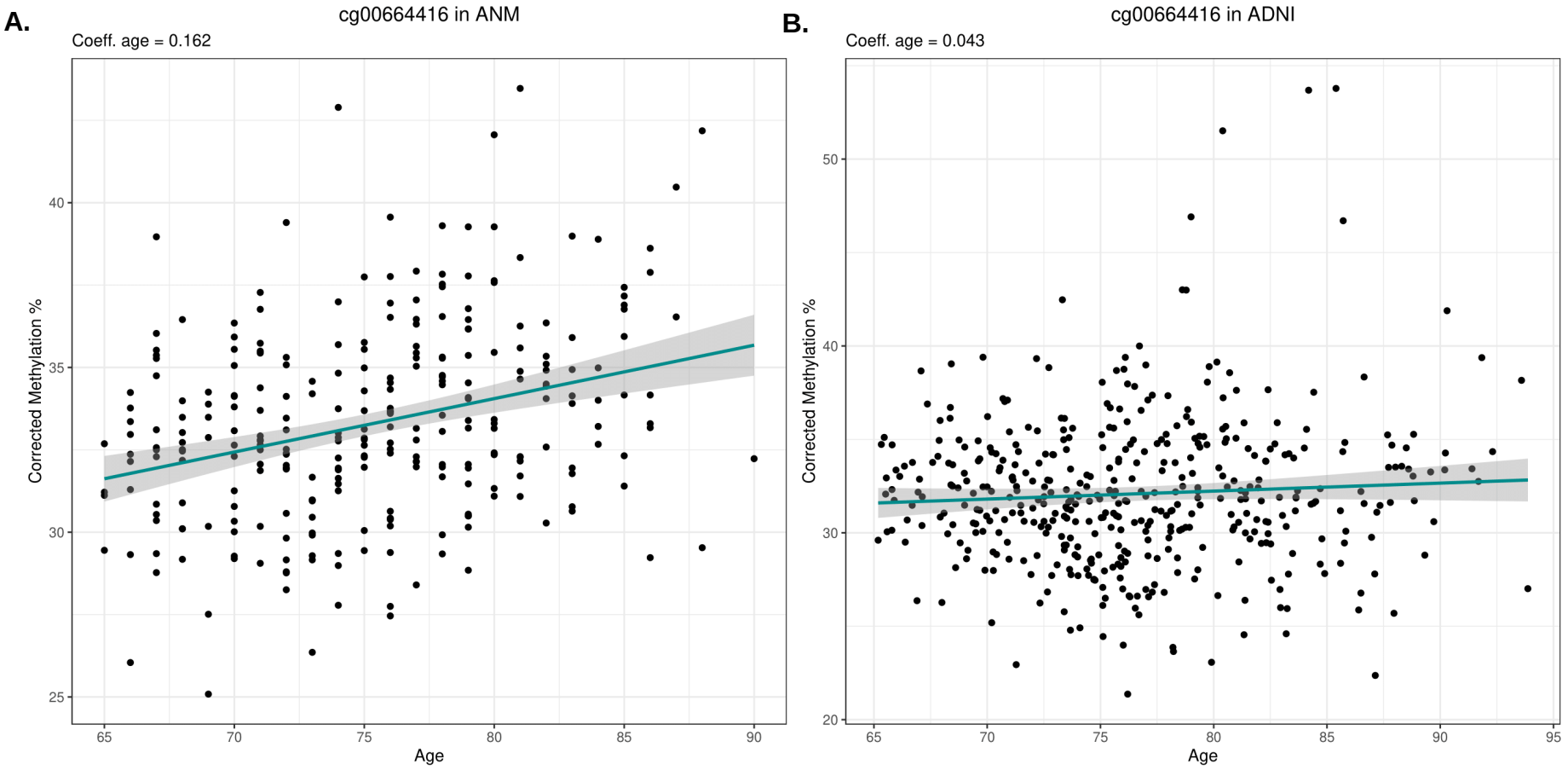

Figure 4.3 The cg00664416 DMP associated with age in the AddNeuroMed cohort.

Probe cg00664416, annotated to the gene MANEAL, displayed hypermethylation with increasing age in the AddNeuroMed cohort (A). This methylation pattern was not reflected in the ADNI replication cohort (B). Displayed methylation values have been corrected for the covariates of cell type proportions and batch effects, as well as the effects of sex and diagnosis.

Top 10 age DMPs										
		AddNeuroMed			ADNI			UCSC Gene	UCSC Gene Group	GREAT Annotation
ProbeID	Position	F	*p* (F)	Δ	F	*p* (F)	Δ			
cg02843237	chr11: 10326289	30.163	9.07E-08	↓	0.002	0.97	↓	*ADM*	TSS1500	*ADM* (-352)
cg00664416	chr1: 38261192	28.402	2.06E-07	↑	0.402	0.53	↑	*MANEAL*	Body,5'UTR,1st Exon	*MANEAL* (+1419), *YRDC* (+12672)
cg18714712	chr19: 49866917	27.617	2.98E-07	↓	0.001	0.98	↑	*DKKL1*	TSS200;TSS1500	*TEAD2* (-1204), *DKKL1* (-124)
cg14382215	chr1: 161169007	27.272	3.50E-07	↑	0.147	0.70	↓	*NDUFS2;ADAMTS4*	TSS200;TSS200	*NDUFS2* (-2929), *ADAMTS4* (-163)
cg00248903	chr19: 3834516	25.338	8.73E-07	↑	0.456	0.50	↓	*ZFR2*	Body	*MATK* (-48102), *ZFR2* (+34510)
cg03669268	chr20: 23549106	23.777	1.84E-06	↑	0.051	0.82	↓	*CST9L*	1stExon,5'UTR	*CST9L* (+279)
cg22095582	chr15: 90131512	22.683	3.10E-06	↑	0.174	0.68	↑	*C15orf42*	Body	*C15orf42* (+12695), *KIF7* (+67169)
cg08108321	chr17: 40713991	21.72	4.93E-06	↓	0.659	0.42	↓	*COASY*	TSS200	*COASY* (-100)
cg00583492	chr11: 66610708	21.534	5.39E-06	↓	0.001	0.97	↑	*RCE1;C11orf80*	TSS200;3'UTR	*RCE1* (-174)
cg00959725	chr1: 33282965	21.294	6.06E-06	↓	0.549	0.46	-	*S100PBP;YARS*	TSS1500;1stExon,5'UTR	*S100PBP* (-210), *YARS* (+667)

Table 4.2 Top 10 DMPs associated with age in the AddNeuroMed cohort.

Shown for each probe is the genomic location, ANOVA F-statistic, associated p-value, and direction of effect (Δ: ↓ = decreased methylation with ageing, ↑ = increased methylation with ageing) in the AddNeuroMed cohort as well as the ADNI replication cohort. Probes are annotated using UCSC gene name and gene group (genome build 37), and GREAT annotation genes with the TSS distance indicated in parentheses.

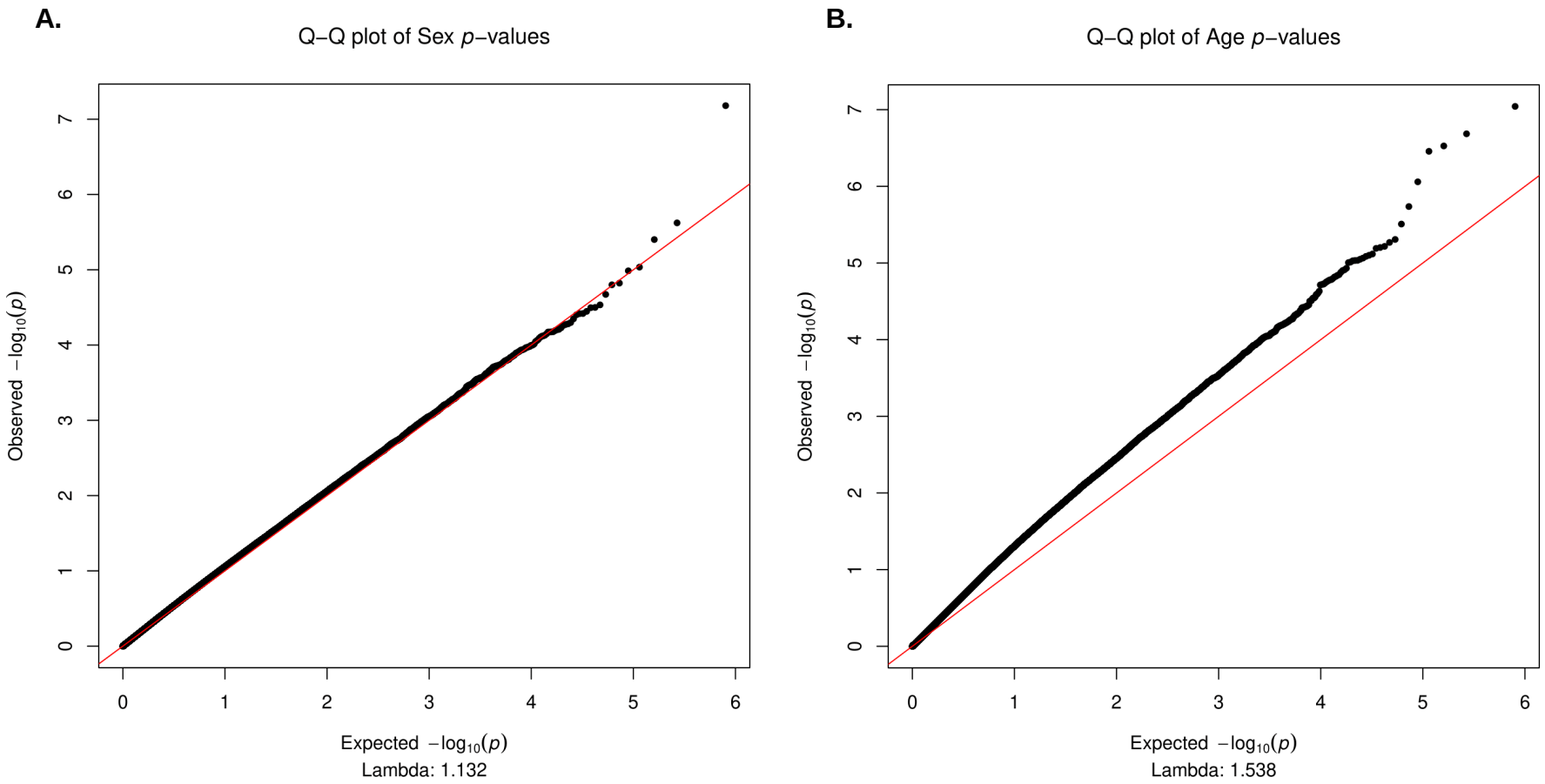

Figure 4.4 Q-Q plots of p-values of the main effects of sex (A) and age (B).

Inflation is observed, particularly in the p-values related to the effects of age.

4.4.2. Interactions between sex, age, and diagnosis

Three interaction effects were of interest in this study: the interaction between sex and age, between sex and diagnosis, and between age and diagnosis. One epigenome-wide significant DMP was found to show an interaction effect for sex and age (cg21437028, $p = 9.03\times10^{-8}$). This DMP is annotated to *SHOX2* and is the same CpG that was the top DMP associated with sex, and should therefore be interpreted in the context of both sex and its interaction with age. In female subjects, this DMP showed hypermethylation with increasing age, whereas male subjects displayed hypomethylation with increasing age (Figure 4.5A). The top 10 DMPs associated with sex and age interaction are shown in Table 4.3. Neither the interaction between sex and diagnosis, nor the interaction between age and diagnosis had any probes passing the epigenome-wide significance threshold. The top probe associated with a sex and diagnosis interaction was cg02429945 ($p = 9.76\times10^{-7}$), which is located on chromosome 5, but does not currently have a UCSC gene annotation. This probe showed opposite methylation patterns in males and females across diagnosis, with male participants tending to show hypermethylation in the MCI stage compared to CTL and AD, and female participants showed hypomethylation in MCI relative to CTL and AD (Figure 4.6A). Table 4.4 displays the top 10 probes associated with an interaction between sex and diagnosis. In the age by diagnosis assessment, the top probe was cg16415340 ($p = 8.88\times10^{-7}$). This probe is annotated to the *INS-IGF2* and *IGF2* genes on chromosome 11, and showed an interesting pattern of methylation where individuals with MCI or AD displayed hypermethylation with increased age, yet in CTL individuals, methylation levels decreased with age (Figure 4.7A). An overview of the top 10 probes associated with interactions between age and diagnosis can be found in Table 4.5.

As also seen in the assessment of the main effects of sex and age, the Q-Q plots for the interaction terms showed slight inflation of the *p*-values (Figure 4.8) and thus the findings were explored for replication in an independent dataset. The replication analysis in ADNI showed that the top DMP associated with an interaction between sex and age did not display a similar effect (Figure 4.5 B). However, the probe cg14437986 was found to be nominally significant in both cohorts (Table 4.3). This site is located on chromosome 6 and is annotated to *C6orf25*. None of the top probes associated with the interaction of sex and diagnosis in AddNeuroMed were found to be nominally significant in ADNI, whereas the one probe associated with an age by diagnosis interaction was replicated. The cg21484956 probe was nominally significant in both datasets and was annotated to *CSRP2* on chromosome 12.

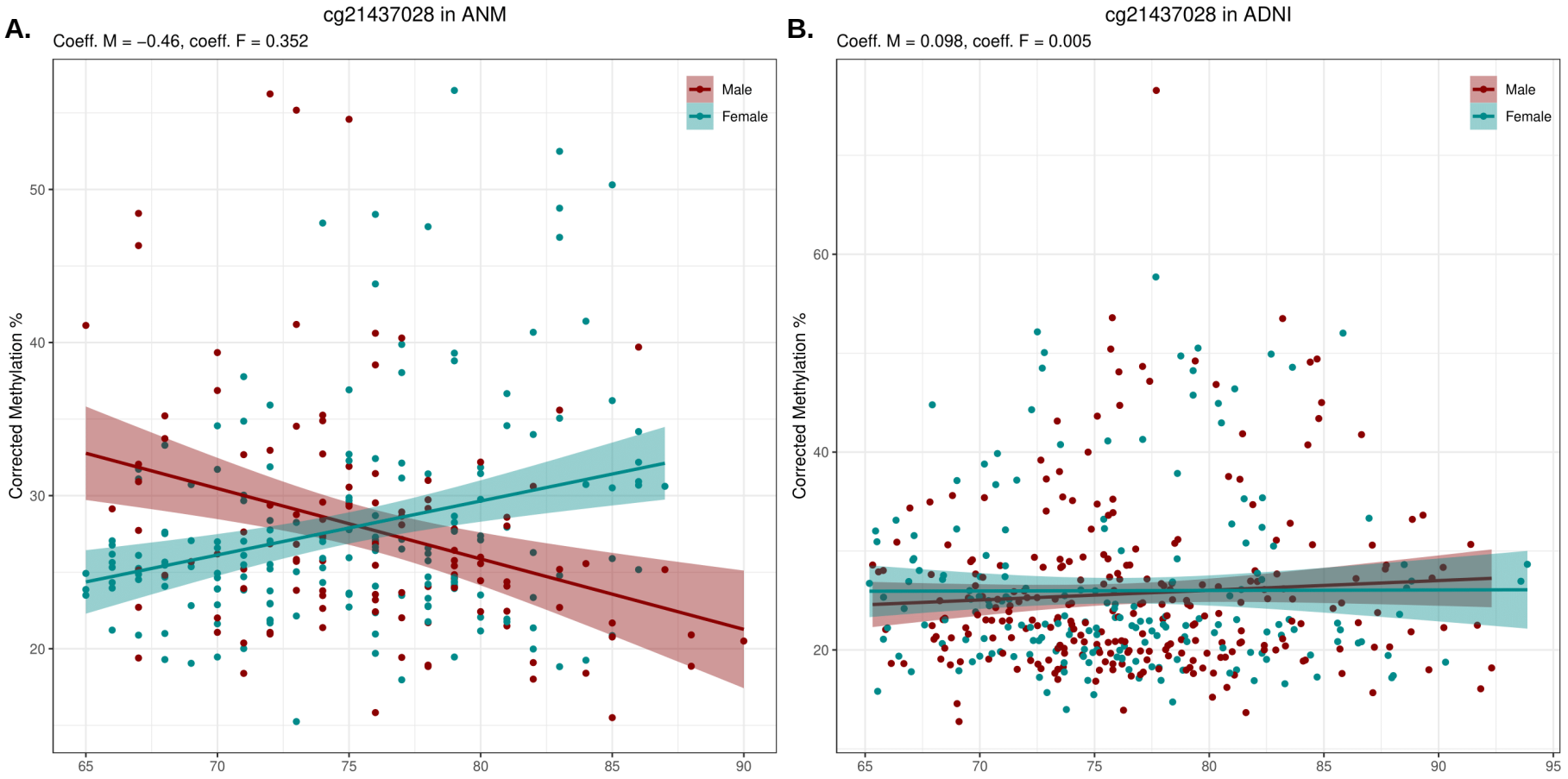

Figure 4.5 The SHOX2 DMP showing a sex and age interaction in the AddNeuroMed cohort.

The cg21437028 probe showed hypermethylation in female subjects with increasing age, and hypomethylation in male participants in the AddNeuroMed cohort, and is annotated to the SHOX2 gene (A). No interaction effect is seen in the ADNI replication cohort (B). Displayed methylation values have been corrected for the covariates of cell type proportions and batch effect, as well as the effects of diagnosis.

Top 10 sex × age DMPs								
ProbeID	**Position**	**AddNeuroMed**		**ADNI**		**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
		F (Sex × Age)	***p* (Sex × Age)**	**F (Sex × Age)**	***p* (Sex × Age)**			
cg21437028	chr3: 157816403	30.172	9.03E-08	1.070	0.30	*SHOX2*	Body	*VEPH1* (-595268), *SHOX2* (+7548)
cg02153041	chr12: 131502934	21.844	4.64E-06	0.006	0.94	*GPR133*	Body	*SFSWAP* (-692700), *GPR133* (+64483)
cg03604067	chr7: 1980273	21.466	5.57E-06	0.948	0.33	*MAD1L1*	Body	*ELFN1* (+231476), *MAD1L1* (+292309)
cg24560809	chr2: 95939339	21.378	5.82E-06	0.239	0.63	*PROM2*	TSS1500	*PROM2* (-861)
cg12855310	chr10: 134215035	20.268	9.97E-06	0.125	0.72	*PWWP2B*	Body	*INPP5A* (-136317), *STK32C* (-93559)
cg25274468	chr7: 114570498	19.004	1.85E-05	1.266	0.26	*MDFIC*	Body	*MDFIC* (+8290)
cg26298737	chr1: 225964937	18.853	1.99E-05	0.130	0.72	*SRP9*	TSS1500	*SRP9* (-577)
cg14437986	chr6: 31691035	18.609	2.24E-05	4.296	3.90E-02	*C6orf25*	TSS200	*LY6G6C* (-1525), *C6orf25* (-85)
cg11128045	chr9: 140395219	18.112	2.86E-05			*PNPLA7*	Body	*NELF* (-41434), *PNPLA7* (+49766)
cg12131219	chr1: 25558575	17.96	3.09E-05	0.664	0.42	*SYF2*	Body	*SYF2* (+437)

Table 4.3 Top 10 DMPs associated with the sex and age interaction in the AddNeuroMed cohort.

Shown for each probe is the genomic location, ANOVA F-statistic and associated p-value for each variable in the AddNeuroMed cohort and the ADNI replication cohort, UCSC gene name and group (genome build 37), and GREAT annotation genes with the TSS distance indicated in parentheses.

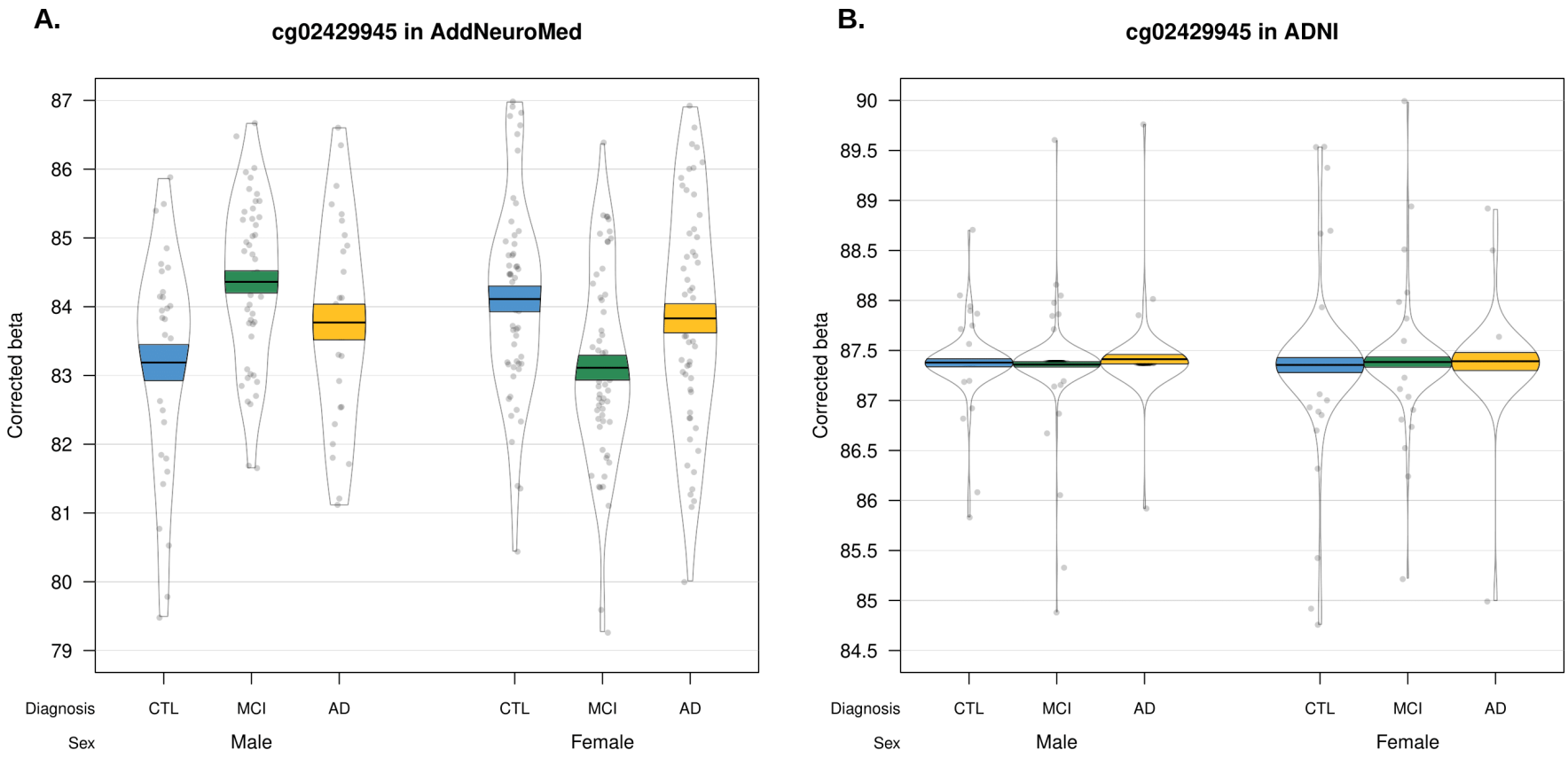

Figure 4.6 The top DMP showing an interaction between sex and diagnosis in the AddNeuroMed cohort.

The cg02429945 probe showed hypermethylation in male participants with MCI, and hypomethylation in female subjects in AddNeuroMed (A), but not ADNI (B). Displayed methylation values have been corrected for the covariates of cell type proportions and batch effects, as well as the effect of age.

Top 10 sex × diagnosis DMPs								
ProbeID	**Position**	**AddNeuroMed**		**ADNI**		**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
		F (Sex × Diagnosis)	***p* (Sex × Diagnosis)**	**F (Sex × Diagnosis)**	***p* (Sex × Diagnosis)**			
cg02429945	chr5: 29150410	14.563	9.76E-07	1.648	0.19			
cg21163392	chr12: 112681365	13.731	2.07E-06	0.995	0.37	*C12orf51*	Body	*TRAFD1* (+118017), *C12orf51* (+138530)
cg23205949	chr10: 91140307	12.583	5.92E-06	0.277	0.76	*IFIT1L*	Body	*IFIT1* (-12014), *IFIT1B* (+2495)
cg10920302	chr3: 45430089	12.211	8.31E-06	1.020	0.36	*LARS2*	5'UTR,1stExon	*LARS2* (+15)
cg07607921	chr17: 4047574	12.088	9.31E-06	0.503	0.61	*ZZEF1;CYB5D2*	TSS1500;Body	*ZZEF1* (-1322), *CYB5D2* (+1113)
cg02596175	chr7: 151574762	12.079	9.39E-06	0.281	0.76	*PRKAG2*	TSS1500	*PRKAG2* (-447)
cg10658516	chr2: 98612531	11.927	1.08E-05	0.385	0.68	*TMEM131*	TSS200	*TMEM131* (-178)
cg20310759	chr1: 87618313	11.818	1.19E-05	0.291	0.75	*LOC339524*	Body	*LMO4* (-175837), *LOC339524* (+20634)
cg16532399	chr1: 38031726	11.644	1.40E-05	1.221	0.30	*DNALI1*	3'UTR	*DNALI1* (+9207), *GNL2* (+29859)
cg02098565	chr16: 89168508	11.549	1.53E-05	0.918	0.40	*ACSF3*	Body	*CDH15* (-69654), *ACSF3* (+8292)

Table 4.4 Top 10 DMPs associated with the sex and diagnosis interaction in the AddNeuroMed cohort.

Shown for each probe is the genomic location, ANOVA F-statistic and associated p-value for each variable in the AddNeuroMed cohort and the ADNI replication cohort, UCSC gene name and group (genome build 37), and GREAT annotation genes with the TSS distance indicated in parentheses.

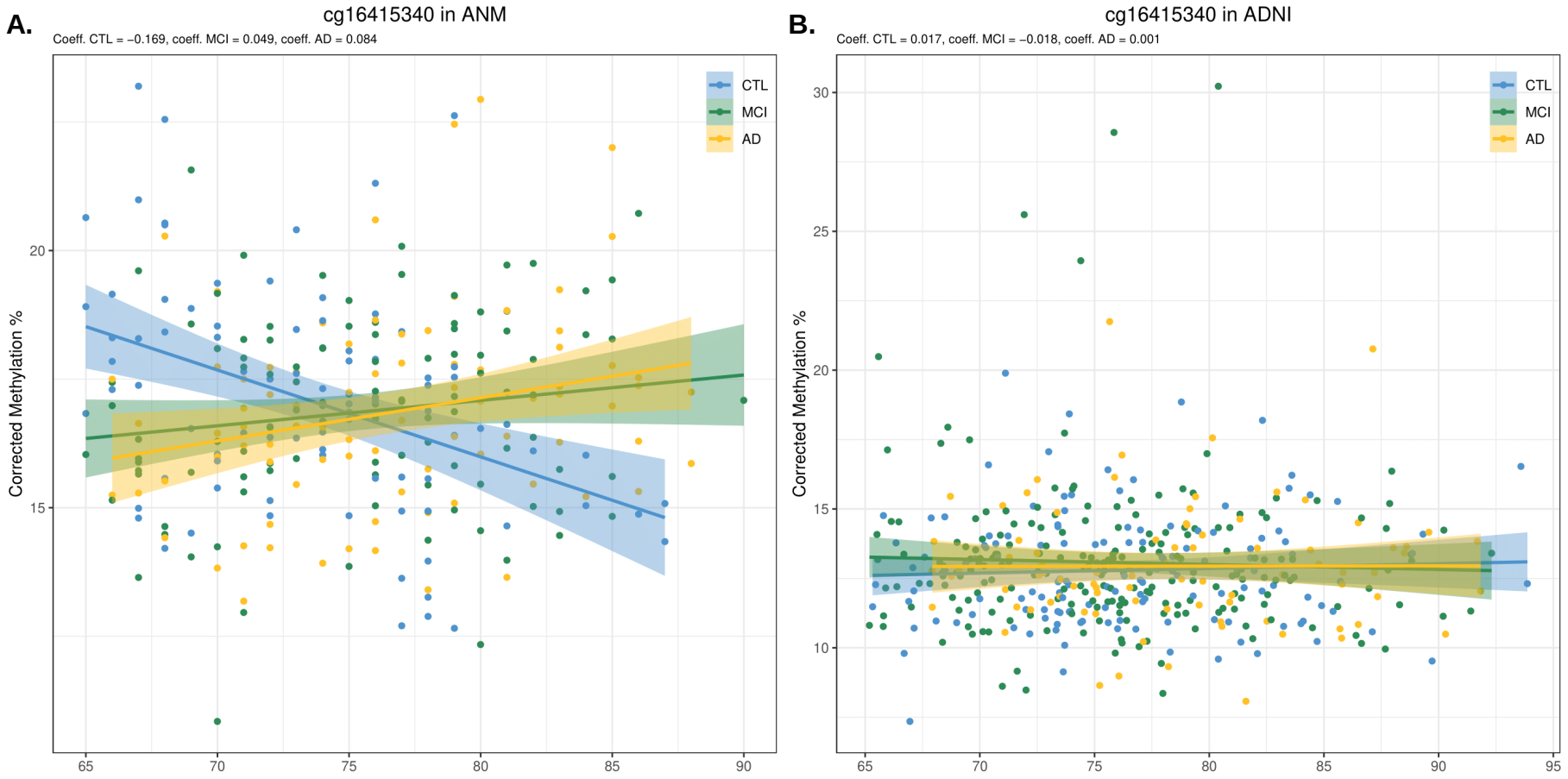

Figure 4.7 The top DMP showing an interaction between age and diagnosis in the AddNeuroMed cohort.

The cg16415340 probe showed hypermethylation in MCI and AD with increasing age, and hypomethylation in CTL in AddNeuroMed (A), but not in ADNI (B). Displayed methylation values have been corrected for the covariates of cell type proportions and batch effects, as well as the effect of sex.

Top 10 age × diagnosis DMPs								
ProbeID	**Position**	**AddNeuroMed**		**ADNI**		**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
		F (Age × Diagnosis)	***p* (Age × Diagnosis)**	**F (Age × Diagnosis)**	***p* (Age × Diagnosis)**			
cg16415340	chr11: 2160964	14.668	8.88E-07	0.237	0.79	*INS-IGF2;IGF2AS; IGF2*	Body;TSS1500; Body,5'UTR,TSS1500	*IGF2* (+1376), *MRPL23* (+192463)
cg16068383	chr15: 56201496	14.498	1.04E-06	0.199	0.82	*NEDD4*	Body	*PRTG* (-166320), *NEDD4* (+7832)
cg15988010	chr4: 166033722	14.108	1.47E-06	-	-	*TMEM192*	Body	*TMEM192* (+301)
cg06214716	chr4: 10118677	13.809	1.93E-06	0.054	0.95	*WDR1*	TSS200	*WDR1* (-105)
cg14277392	chr19: 57018933	13.403	2.80E-06	0.425	0.65	*ZNF471*	TSS1500	*ZNF471* (-278)
cg20303995	chr10: 104159500	13.389	2.83E-06	0.827	0.44	*NFKB2*	Body	*NFKB2* (+5162), *PSD* (+19400)
cg21484956	chr12: 77273469	13.366	2.89E-06	4.663	1.00E-02	*CSRP2*	TSS1500	*CSRP2* (-671)
cg21875080	chr1: 46598774	13.247	3.22E-06	0.007	0.99	*PIK3R3*	TSS200,TSS1500	*PIK3R3* (-395)
cg00060606	chr1: 45140440	13.205	3.35E-06	-	-	*C1orf228;TMEM53*	5'UTR,1stExon;TSS1500	*TMEM53* (-342), *C1orf228* (+47)
cg10953131	chr1: 17997665	13.042	3.89E-06	4.483	1.20E-02	*ARHGEF10L*	Body	*ACTL8* (-84142), *ARHGEF10L* (+131336)

Table 4.5 Top 10 DMPs associated with an interaction between age and diagnosis in the AddNeuroMed cohort.

Shown for each probe is the genomic location, ANOVA F-statistic and associated p-value for each variable in the AddNeuroMed cohort and the ADNI replication cohort, UCSC gene name and group (genome build 37), and GREAT annotation genes with the TSS distance indicated in parentheses.

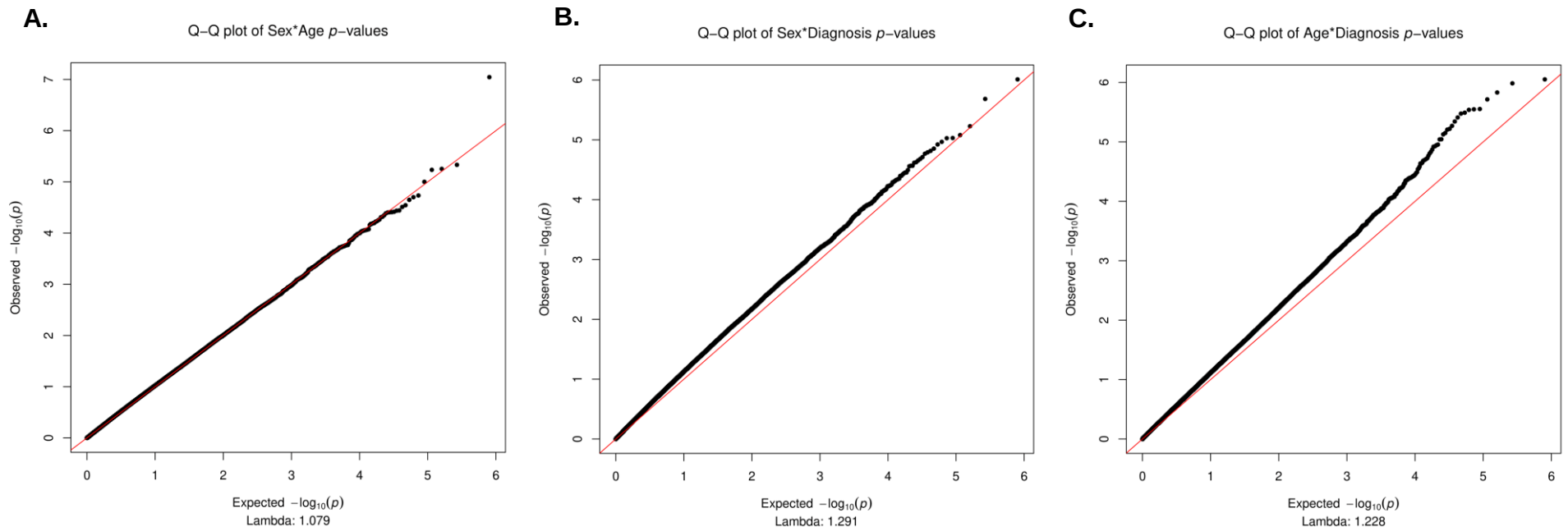

Figure 4.8 Q-Q plots of sex, age, and diagnosis interaction effects.

Q-Q plots of the p-values associated with interaction effects of sex and age (A), sex and diagnosis (B), and age and diagnosis (C) are shown.

4.4.3. Sex, age, and interaction DMRs

Following the identification of DMPs associated with sex, age, and their interactions with diagnosis, *comb-p* was used to identify DMRs. Twelve significant DMRs were found with regard to sex, with the most significant region being annotated to the genes *GNA12* and *AMZ1* (904bp; Table 4.6). The other identified DMRs were annotated to the genes *MPIG6B* (909 bp), *SLC17A8* (579 bp), *CIB1* and *TTLL13P* (757 bp), *PIGZ* (270 bp), *CFAP161* (474 bp), *ANXA2R* and *LOC648987* (825 bp), *CNKSR1* (397 bp), *NRXN2* (566 bp), *STING1* (610 bp), *ALLC* (608 bp), and *RASSF4* (258 bp). There were 25 DMRs identified that showed significant methylation differences with age. The most significant region spanned 1,743 bp, and is annotated to the genes *ANXA2R* and *LOC648987*. Of note, this region fully overlapped the DMR associated with sex that is annotated to the same genes. Several other genomic regions also contained DMRs that were associated with both sex and age. The *PIGZ* gene on chromosome 3 contained a DMR associated with both sex and age spanning the exact same 270 bp region. The sex-associated DMR in *NRXN2* on chromosome 11 also overlapped with an extended DMR associated with age (1113 bp). Additionally, the gene *ALLC* contains a DMR associated with both sex and age, which starts at the same site, but the age-associated DMR spans 210 probes further (818 bp in total). A table of all DMRs related to age is shown in Table 4.7.

In addition to DMRs associated with the main effects, there were a number of regions that were associated with the interactions between sex, age, and diagnosis. Regions that reflected significant interactions between sex and age (Table 4.8) included a DMR in *MPIG6B* (909 bp), which exactly matched the DMR associated with sex. In fact, of the nine DMRs displaying a sex by age interaction,

the majority exactly matched or overlapped with DMRs related to sex. In addition to the DMR in *MPIG6B*, this included DMRs in the genes *CIB1* and *TTLL13P, SLC17A8, PIGZ, CFAP161*, *NRXN2*, *CNKSR1,* and *ANXA2R* and *LOC648987.* Furthermore, the sex by age associated DMR in *PIGZ* exactly overlapped the DMR identified in association with both sex and age. The 448 bp DMR annotated to *ANXA2R* and *LOC648987* was fully spanned by DMRs identified in both age and sex.

A total of 27 significant DMRs were found in the analysis of the interaction between sex and diagnosis (Table 4.9). The most significant DMR was located on chromosome 7, spanned 1,728 bp, and was annotated to the *HOXA4* gene. One gene located on chromosome 16 contained two regions: the *ACSF3* gene contained one DMR spanning 365 bp, and, separated by 3,608 bp, another DMR spanning only 38 bp. In addition to these two DMRs displaying a sex by diagnosis interaction, the same gene contained a DMR (156 bp) displaying a sex by age interaction, though this region did not overlap with either of the other DMRs annotated to this gene. Notably, none of the sex by diagnosis DMRs exactly matched with the sex DMRs, but one DMR overlapped with an age-related DMR. This DMR is annotated to *C10orf90* on chromosome 10, and spans 173 bp.

With regard to the interaction between age and diagnosis, nine significant DMRs were found (Table 4.10). The most significant DMR was located in *MAD1L1* on chromosome 7 (925 bp). This interaction effect also has an associated DMR in the *ANXA2R* and *LOC648987* genomic region, though this 755 bp region only partially overlaps the DMRs associated with either sex, or age. The age and diagnosis interaction DMR did fully overlap the sex and age interaction DMR.

DMRs associated with sex						
Gene	**Position**	**Gene Feature**	***n***	***p*-value**	**Šidák-*p***	**Overlap DMR**
GNA12;AMZ1	chr7: 2802073 - 2802977	intron+TSS+utr5+cds;intron	7	6.10E-13	2.71E-10	Sex × Age
MPIG6B	chr6: 31690904 - 31691813	TSS+intron+utr5+cds	22	3.12E-12	1.38E-09	Sex × Age†
SLC17A8	chr12: 100750473 - 100751052	TSS+exon+utr5	8	1.16E-08	8.03E-06	Sex × Age†
CIB1;TTLL13P	chr15: 90792300 - 90793057	nc_intron;nc_intron+nc_exon	9	2.19E-08	1.16E-05	Sex × Age†
PIGZ	chr3: 196705629 - 196705899	intergenic	5	8.85E-09	1.32E-05	Age†, Sex × Age†
CFAP161	chr15: 81426347 - 81426821	TSS+intron+utr5+cds	9	3.59E-08	3.04E-05	Sex × Age†
ANXA2R;LOC648987	chr5: 43040174 - 43040999	TSS+exon;nc_intron	8	3.85E-07	1.87E-04	Age, Sex × Age, Age × Diag
CNKSR1	chr1: 26503623 - 26504020	TSS+exon+utr5	8	2.78E-07	2.80E-04	Sex × Age†
NRXN2	chr11: 64374653 - 64375219	intron+utr3+cds;utr3+cds	3	1.29E-06	9.15E-04	Age, Sex × Age†
STING1	chr5: 138861832 - 138862442	TSS+intron+exon+utr5	6	2.16E-06	1.42E-03	
ALLC	chr2: 3698746 - 3699354	intergenic	5	3.17E-06	2.09E-03	Age
RASSF4	chr10: 45477550 - 45477808	intron	3	2.54E-06	3.94E-03	

Table 4.6 DMRs associated with sex in AddNeuroMed.

Displayed are the DMRs associated with the main effects of sex in blood. For each DMR the UCSC annotated gene is shown, its chromosomal position (genome build 37), gene feature (cds = coding sequence; TSS = transcription start site; utr5 = 5' untranslated region; utr3 = 3' untranslated region, nc_ = non-coding), number of probes in region (n), p-value and multiple testing-corrected p-value (Šidák-p). The column 'Overlap DMR' indicates whether regions partially overlap, or fully match (indicated by †) another DMR identified in this chapter, and which main- or interaction effect that DMR is associated with (Age, Sex × Age, Sex × Diagnosis (Diag), or Age × Diagnosis).

DMRs associated with age						
Gene	**Position**	**Gene Feature**	***n***	***p*-value**	**Šidák-*p***	**Overlap DMR**
ANXA2R;LOC648987	chr5: 43039410 - 43041153	TSS+cds;nc_intron	12	2.03E-18	4.67E-16	Sex, Sex × Age, Age × Diag
ZNF580;ZNF581	chr19: 56154382 - 56155152	utr3+cds;TSS+intron+exon+utr5	9	7.45E-11	3.88E-08	
FUT11	chr10: 75533025 - 75533432	cds	5	3.79E-09	3.73E-06	
FAAP20	chr1: 2120985 - 2121725	nc_intron+nc_exon+intron+ exon+utr5+utr3+cds	6	1.62E-08	8.78E-06	
ALLC	chr2: 3698746 - 3699564	intergenic	6	2.10E-08	1.03E-05	Sex
OXT	chr20: 3051954 - 3052484	TSS+intron+utr5+cds	10	1.83E-08	1.39E-05	
NRXN2	chr11: 64374653 - 64375766	intron+utr3+cds	4	5.62E-08	2.03E-05	Sex, Sex × Age
MMACHC;CCDC163	chr1: 45965343 - 45966116	TSS+intron+utr5+cds+exon; TSS+exon+utr5+nc_intron+nc_exon	12	5.15E-08	2.67E-05	
INPP5A	chr10: 134361949 - 134362171	intron	4	5.81E-08	1.05E-04	
PIGZ	chr3: 196705629 - 196705899	intergenic	5	9.23E-08	1.37E-04	Sex†, Sex × Age†
CYP2E1	chr10: 135341933 - 135343194	intron+cds	7	7.48E-07	2.38E-04	
GTPBP10	chr7: 89975649 - 89976217	TSS+intron+utr5+cds	9	6.98E-07	4.93E-04	
ZNF471	chr19: 57018614 - 57019374	TSS+intron+exon+utr5	8	1.26E-06	6.67E-04	
PRSS16	chr6: 27173574 - 27173992	intergenic	5	1.43E-06	1.37E-03	
DSCR9	chr21: 38593042 - 38593285	nc_exon	2	1.12E-06	1.85E-03	

DMRs associated with age (Continued)						
Gene	**Position**	**Gene Feature**	***n***	***p*-value**	**Šidák-*p***	**Overlap DMR**
PARD3	chr10: 34408530 - 34408655	intron+cds	3	5.99E-07	1.92E-03	
LAMB2	chr3: 49170496 - 49170669	intergenic	3	1.15E-06	2.67E-03	
SNX33	chr15: 75940301 - 75940556	exon+utr5	3	1.70E-06	2.67E-03	
VPS26C	chr21: 38630234 - 38630729	intron	4	3.47E-06	2.81E-03	
SUN3	chr7: 48031468 - 48031801	intron	3	2.56E-06	3.08E-03	
HLA-E	chr6: 30457711 - 30458602	intron+cds	12	9.54E-06	4.29E-03	
C10orf90	chr10: 128152075 - 128152248	nc_intron;intron	4	2.03E-06	4.70E-03	Sex × Diag†
MANEAL	chr1: 38261057 - 38261402	intron+TSS+exon+utr5	3	7.12E-06	8.25E-03	
TUBA4A	chr2: 220116179 - 220116295	intron+cds	2	2.94E-06	1.01E-02	
NDUFC1	chr4: 140216957 - 140217099	exon+utr5	5	8.65E-06	2.41E-02	

Table 4.7 DMRs associated with age in AddNeuroMed.

Displayed are the DMRs associated with the main effects of age in blood. For each DMR the UCSC annotated gene is shown, its chromosomal position (genome build 37), gene feature (cds = coding sequence; TSS = transcription start site; utr5 = 5' untranslated region; utr3 = 3' untranslated region, nc_ = non-coding), number of probes in region (n), p-value and multiple testing-corrected p-value (Šidák-p). The column 'Overlap DMR' indicates whether regions partially overlap, or fully match (indicated by †) another DMR identified in this chapter, and which main- or interaction effect that DMR is associated with (Sex, Sex × Age, Sex × Diagnosis (Diag), or Age × Diagnosis).

DMRs associated with sex × age						
Gene	**Position**	**Gene Feature**	***n***	***p*-value**	**Šidák-*p***	**Overlap DMR**
MPIG6B	chr6: 31690904 - 31691813	TSS+intron+utr5+cds	22	1.55E-12	6.86E-10	Sex†
GNA12;AMZ1	chr7: 2802374 - 2802977	Intron,TSS+utr5+cds;intron	6	2.11E-10	1.41E-07	Sex
CIB1;TTLL13P	chr15: 90792300 - 90793057	nc_intron;nc_intron+nc_exon	9	6.22E-09	3.30E-06	Sex†
SLC17A8	chr12: 100750473 - 100751052	TSS+exon+utr5	8	6.24E-09	4.33E-06	Sex†
PIGZ	chr3: 196705629 - 196705899	intergenic	5	3.61E-09	5.37E-06	Sex†, Age†
CFAP161	chr15: 81426347 - 81426821	TSS+intron+utr5+cds	9	1.24E-08	1.05E-05	Sex†
NRXN2	chr11: 64374653 - 64375219	intron+utr3+cds	3	3.31E-07	2.35E-04	Sex†, Age
CNKSR1	chr1: 26503623 - 26504020	TSS+exon+utr5	8	1.65E-06	1.67E-03	Sex†
ANXA2R;LOC648987	chr5: 43040174 - 43040622	TSS+exon;nc_intron	4	2.47E-06	2.21E-03	Sex, Age, Age × Diag

Table 4.8 DMRs associated with a sex by age interaction in AddNeuroMed.

Displayed are the DMRs associated with the interaction effect of sex and age in blood. For each DMR the UCSC annotated gene is shown, its chromosomal position (genome build 37), gene feature (cds = coding sequence; TSS = transcription start site; utr5 = 5' untranslated region; utr3 = 3' untranslated region, nc_ = non-coding), number of probes in region (n), p-value and multiple testing-corrected p-value (Šidák-p). The column 'Overlap DMR' indicates whether regions partially overlap, or fully match (indicated by †) another DMR identified in this chapter, and which main- or interaction effect that DMR is associated with (Sex, Age, Sex × Diagnosis (Diag), or Age × Diagnosis).

DMRs associated with sex × diagnosis						
Gene	**Position**	**Gene Feature**	***n***	***p*-value**	**Šidák-*p***	**Overlap DMR**
HOXA4	chr7: 27169674 - 27171402	TSS+intron+utr5+cds	23	2.42E-17	5.62E-15	
ZNF649-AS1;ZNF577	chr19: 52390810 - 52391790	nc_intron+nc_exon;nc_exon, TSS+intron+exon+utr5,TSS+exon+utr5	13	4.84E-10	1.98E-07	
SVIL-AS1	chr10: 29697905 - 29699044	nc_intron+nc_exon	9	7.28E-10	2.56E-07	
CDK11B;SLC35E2B	chr1: 1609388 - 1609973	intron,intron+utr5;intron+utr5	3	2.30E-09	1.58E-06	
C5orf63	chr5: 126408756 - 126409554	TSS+intron+exon+utr5	13	4.51E-09	2.27E-06	
CD247	chr1: 167408509 - 167409199	intron+cds	7	1.66E-08	9.64E-06	
PM20D1	chr1: 205818956 - 205819610	nc_intron+nc_exon,TSS+intron+utr5+cds	8	1.89E-08	1.16E-05	
TMEM72-AS1;TMEM72	chr10: 45406187 - 45406848	nc_intron;TSS+exon+utr5	6	3.18E-08	1.93E-05	
ACSF2;CHAD	chr17: 48545805 - 48546621	nc_intron,intron;TSS+utr5+cds	11	1.31E-07	6.44E-05	
KHDC3L	chr6: 74071738 - 74072821	TSS+intron+utr5+cds	9	4.48E-07	1.66E-04	
ADGRD1	chr12: 131488390 - 131488927	intron+cds	6	2.33E-07	1.74E-04	
LINC00533	chr6: 28601269 - 28601520	intergenic	11	1.18E-07	1.88E-04	
ACSF3	chr16: 89168599 - 89168964	intron+utr5,intron,nc_intron	2	1.77E-07	1.94E-04	
LINC02470	chr12: 10095902 - 10096153	nc_exon	5	1.30E-07	2.08E-04	
GLIPR1L2;CAPS2	chr12: 75784617 - 75785296	TSS+intron+utr5+cds, nc_intron+nc_exon; TSS+intron+utr5+cds,TSS+intron+exon+utr5, nc_intron+nc_exon	10	5.69E-07	3.36E-04	
PTCHD3	chr10: 27702774 - 27703548	TSS+utr5+cds	8	6.67E-07	3.46E-04	
PTGDR	chr14: 52734156 - 52734530	TSS+exon+utr5	8	6.23E-07	6.68E-04	

DMRs associated with sex × diagnosis (Continued)						
Gene	**Position**	**Gene Feature**	***n***	***p*-value**	**Šidák-*p***	**Overlap DMR**
ARID5B	chr10: 63809073 - 63809150	TSS+exon+utr5,intron	5	1.57E-07	8.20E-04	
PTPRN2	chr7: 158045980 - 158046359	intron	6	1.96E-06	2.07E-03	
C10orf90	chr10: 128152075 - 128152248	nc_intron,intron	4	1.22E-06	2.84E-03	Age†
LEP	chr7: 127911258 - 127911968	intergenic	4	5.26E-06	2.97E-03	
ACSF3	chr16: 89164953 - 89164991	intron+utr5,nc_intron	2	3.24E-07	3.41E-03	
VSTM5	chr11: 93583253 - 93583764	intron+utr5+cds	7	5.28E-06	4.13E-03	
AKTIP	chr16: 53543985 - 53544099	intergenic	3	1.40E-06	4.91E-03	
LOC105377162	chr3: 72704324 - 72704702	intergenic	5	4.79E-06	5.07E-03	
NEAT1	chr11: 65194933 - 65195040	nc_exon	4	2.87E-06	1.07E-02	
NCALD	chr8: 103129227 - 103129371	intron+utr5	4	5.08E-06	1.41E-02	

Table 4.9 DMRs associated with a sex by diagnosis interaction in AddNeuroMed.

Displayed are the DMRs associated with the interaction effect of sex and diagnosis in blood. For each DMR the UCSC annotated gene is shown, its chromosomal position (genome build 37), gene feature (cds = coding sequence; TSS = transcription start site; utr5 = 5' untranslated region; utr3 = 3' untranslated region, nc_ = non-coding), number of probes in region (n), p-value and multiple testing-corrected p-value (Šidák-p). The column 'Overlap DMR' indicates whether regions partially overlap, or fully match (indicated by †) another DMR identified in this chapter, and which main- or interaction effect that DMR is associated with (Sex, Age, Sex × Age, or Age × Diagnosis).

DMRs associated with age × diagnosis						
Gene	**Position**	**Gene Feature**	***n***	***p*-value**	**Šidák-*p***	**Overlap DMR**
MAD1L1	chr7: 1895903 - 1896828	intron	6	7.90E-13	3.43E-10	
ANXA2R;LOC648987	chr5: 43039867 - 43040622	TSS+cds;nc_intron	7	7.03E-12	3.74E-09	Sex, Age, Sex × Age
CHI3L1	chr1: 203155737 - 203156626	TSS+utr5+cds	6	3.46E-08	1.56E-05	
VTRNA2-1	chr5: 135415693 - 135416530	nc_gene	14	4.18E-08	2.00E-05	
LOC100131289	chr6: 27730016 - 27730564	nc_intron+nc_exon	4	3.67E-08	2.69E-05	
KCNMB3	chr3: 178978849 - 178979014	intron	5	8.70E-08	2.12E-04	
LINC00424	chr13: 22615049 - 22615335	intergenic	5	5.18E-07	7.27E-04	
HCG15	chr6: 28945182 - 28945508	intergenic	6	6.55E-07	8.06E-04	
MAST4	chr5: 66461884 - 66462472	cds	3	1.44E-06	9.85E-04	

Table 4.10 DMRs associated with an age by diagnosis interaction in AddNeuroMed.

Displayed are the DMRs associated with the interaction effect of age and diagnosis in blood. For each DMR the UCSC annotated gene is shown, its chromosomal position (genome build 37), gene feature (cds = coding sequence; TSS = transcription start site; utr5 = 5' untranslated region; utr3 = 3' untranslated region, nc_ = non-coding), number of probes in region (n), p-value and multiple testing-corrected p-value (Šidák-p). The column 'Overlap DMR' indicates whether regions partially overlap, or fully match (indicated by †) another DMR identified in this chapter, and which main- or interaction effect that DMR is associated with (Sex, Age, Sex × Age, or Sex × Diagnosis (Diag)).

4.4.4. Replication of DMRs in ADNI

For each effect investigated in the AddNeuroMed cohort, probes located within these DMRs were extracted from the ADNI EWAS results. For example, probes within the *GNA12*- and *AMZ1*-annotated DMR associated with sex and the interaction of sex and age in AddNeuroMed were extracted from the ADNI data and inspected for significance in association with sex and sex by age. The majority of probes within AddNeuroMed DMRs were not significant in the ADNI dataset, and several DMRs only contained one nominally significant probe in ADNI. There were a few notable exceptions, and the focus of this section is placed on an arbitrarily set 'level of replication': DMRs where the significance of more than one third of the probes in the region was replicated in ADNI. Of the overlapping *ALLC* DMRs that were associated with both sex and age in AddNeuroMed, three out of six age-associated probes were nominally significant in ADNI for age (min. p = 0.006), and one other probe showed a nominally significant effect of sex (p = 0.024). Similarly, the *ZNF471* age-associated DMR in AddNeuroMed consisted of eight probes in total, of which three showed a nominally significant effect of age in ADNI (min. p = 0.03). Also associated with age, the DMR located in *VPS26C* contained four probes in the AddNeuroMed dataset, and the age effect was replicated in two of these probes in ADNI (min. p = 0.024). Of the DMRs associated with interaction effects, none passed the selected level of replication for the relevant interaction effects in ADNI, though a few interaction DMRs contained probes that were nominally significant for an associated main effect in ADNI. For example, the DMR annotated to *CDK11B* and *SLC35E2B* that reflected an interaction of sex and diagnosis in AddNeuroMed, contained three probes of which two showed a significant effect of sex in ADNI (min. p = 0.002). Similarly, the sex by diagnosis interaction region

in *PTPRN2* contained six probes of which five were associated with sex in ADNI (min. $p = 0.006$). Lastly, two of the three probes in the *MAST4* DMR that reflected an age by diagnosis interaction in AddNeuroMed were nominally significant for age in ADNI (min. $p = 0.006$).

4.4.5. Pathways altered in association with sex, age and their interactions with disease

Using the 16 WGCNA modules generated as described in Chapter 2.4.2 for the AddNeuroMed data (Figure 2.25), it was investigated whether any biological pathways were altered in association with sex, age, or the interactions of sex and age, sex and diagnosis, and age and diagnosis. The covariates of cell type proportions and batch were regressed out of the MEs, and extreme outliers (exceeding > 5 SD) were removed. The association between the variables of interest and the modules was tested by running an ANOVA interaction model as described in 4.3.1. Several modules were found to show a nominally significant (p <0.05) association with age or the interaction between sex and diagnosis (Figure 4.9). The modules greenyellow ($F = 5.1$, $p = 0.02$), pink ($F = 4.6$, $p = 0.03$), and green ($F = 4.5$, $p = 0.03$) reflected differences in age, and the salmon module showed an interaction effect between sex and diagnosis ($F = 4.8$, $p = 0.009$). For these modules, MM to PS correlations and plots were generated, to test whether the probes central to these modules also showed the strongest association with age or a sex by diagnosis interaction. Only the modules showing significant positive MM to PS correlations (i.e. $r > 0$, $p < 0.05$) were selected for further pathway analysis. For age, the greenyellow and green modules passed this selection criterion (Figure 4.10), whereas the salmon module linked to a sex and diagnosis interaction did not pass the criterion (Figure 4.11). Using the

missMethyl package, GO and KEGG enrichment analyses were run on the greenyellow and green modules. A FDR multiple testing correction was carried out for all terms, and only terms with $q < 0.05$ were considered significant. A total of 26 GO terms were found to be significant in the green module related to age, and the top GO terms were 'mRNA metabolic process' ($q = 5.18\times10^{-5}$), and 'RNA processing' ($q = 5.18\times10^{-5}$). Only three significant KEGG terms were identified, i.e. 'spliceosome' ($q = 2.93\times10^{-3}$), 'systemic lupus erythematosus' ($q = 8.52\times10^{-3}$), and 'alcoholism' ($q = 0.02$). The significant GO terms and the significant KEGG terms are displayed in Figure 4.12. The greenyellow module, which also reflected differences in age, had one significant GO term, i.e. 'homophilic cell adhesion via plasma membrane adhesion molecules' ($q = 0.022$), and two significant KEGG terms, i.e. 'human papillomavirus infection' ($q = 1.05\times10^{-3}$) and 'basal cell carcinoma' ($q = 1.61\times10^{-3}$).

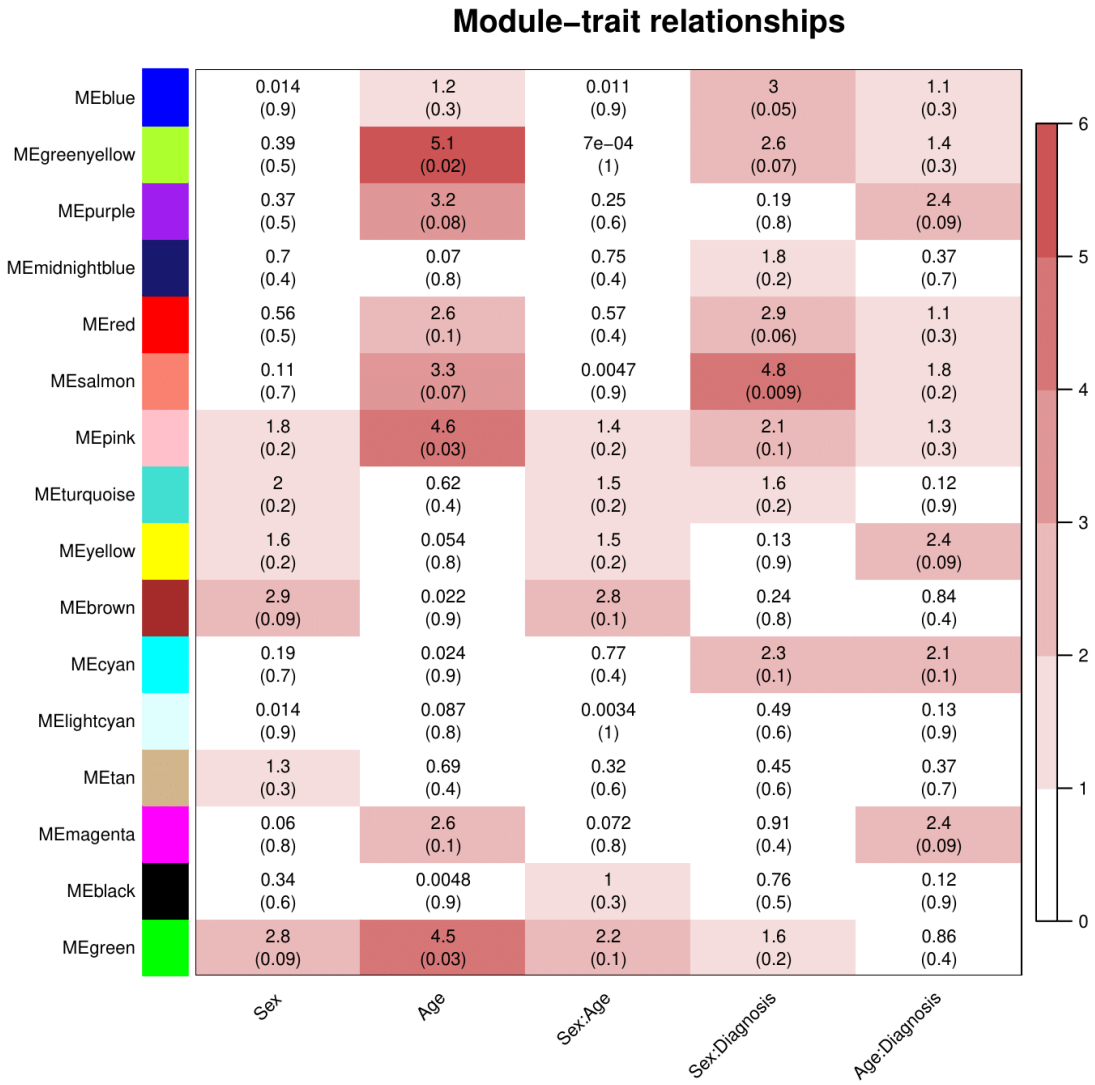

Figure 4.9 Heatmap of module-trait relationships.

An interaction model consisting of the main effects of sex, age, and diagnosis, and the interactions of sex × age, sex × diagnosis, and age × diagnosis, was run on the MEs for each module. Displayed are ANOVA F-statistics for each association (also indicated by colour), and p-values in parentheses.

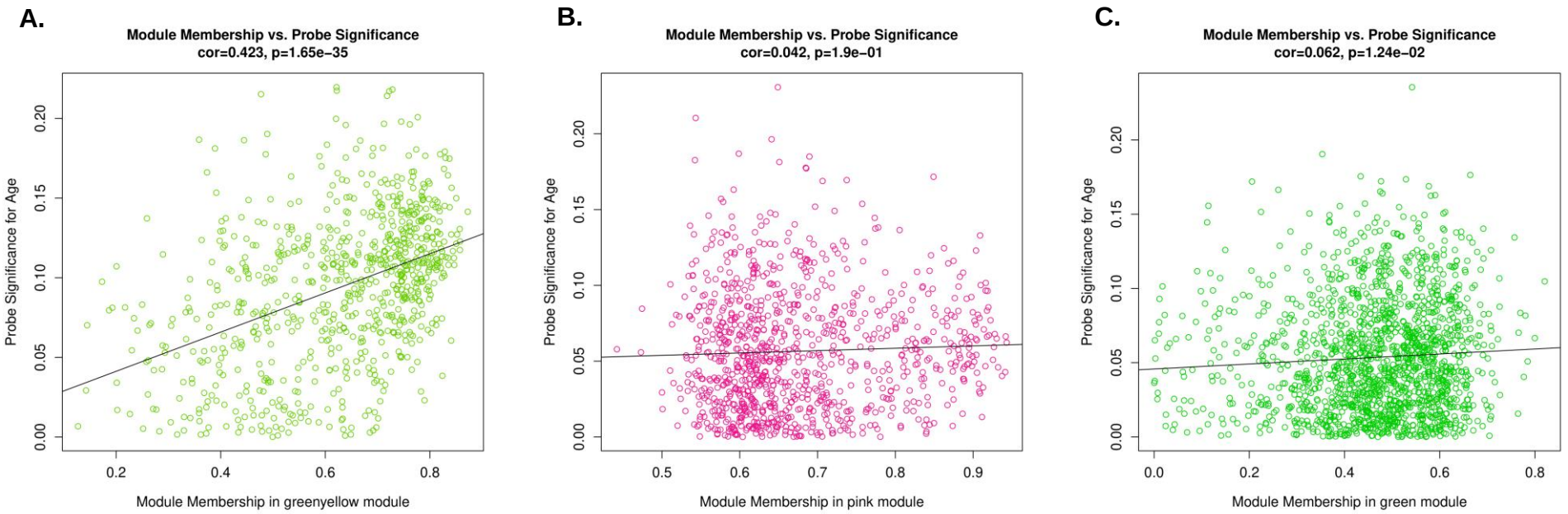

Figure 4.10 Module membership (MM) to probe significance (PS) correlations for modules associated with age in AddNeuroMed.

Only those modules showing positive and significant (i.e. $r > 0$, $p < 0.05$) MM to PS correlations were selected for pathway analyses.

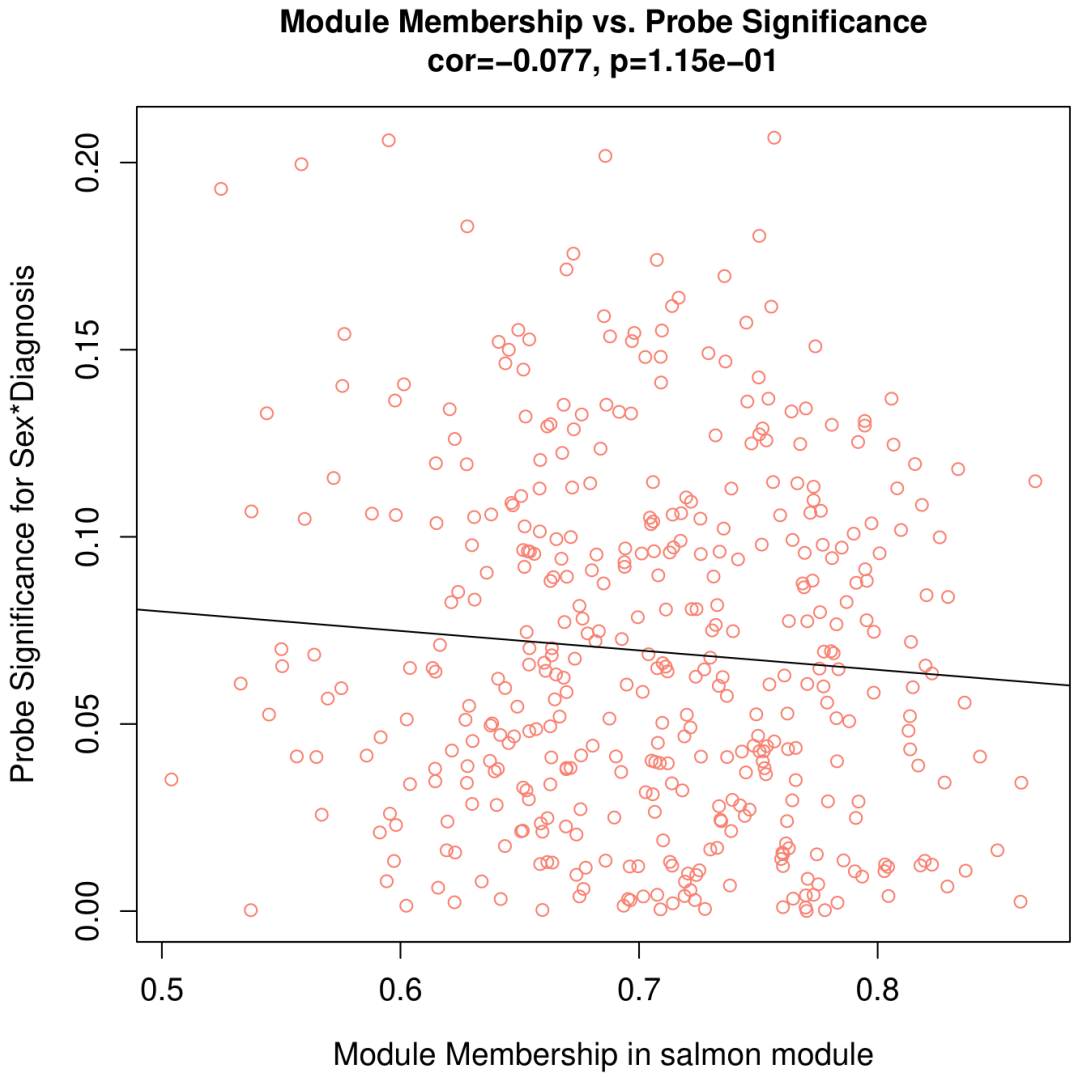

Figure 4.11 Module membership (MM) to probe significance (PS) correlations for the salmon module associated with a sex and diagnosis interaction in AddNeuroMed.

As this module did not show a positive and significant (i.e. $r > 0$, $p < 0.05$) MM to PS correlation, this module was not selected for pathway analyses.

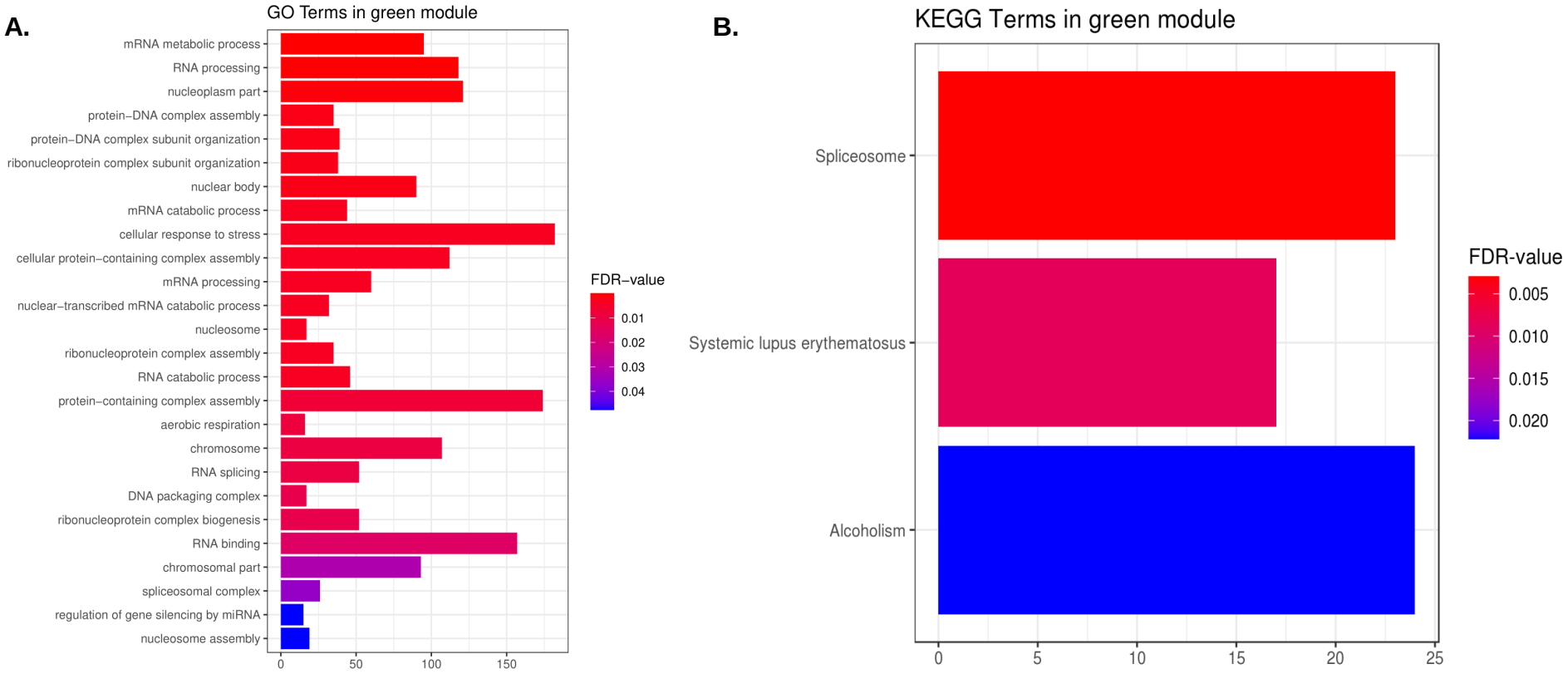

Figure 4.12 GO and KEGG terms enriched in the green module associated with age.

GO and KEGG terms are shown with the number of altered genes in the pathway shown on the x-axis, and terms are coloured by q-value (FDR-value). All significant GO terms are shown in A, and the three significant KEGG terms are shown in B.

4.5. Discussion

The current study identified several epigenome-wide significant positions, regions, and pathways that are associated with sex, age, and their interactions with each other and diagnosis.

A DMP was found in the short stature homeobox 2 gene (*SHOX2*), which showed a pattern of decreased DNA methylation with ageing in male participants, and the opposite pattern in female participants. The *SHOX2* gene is part of the homeobox family of genes, which is involved in hematopoietic development as well as early body morphogenesis (Krumlauf, 1994). The role of *SHOX2* in later ageing and sex differences is unknown, though methylation of this gene has been suggested as a diagnostic biomarker for lung cancer in plasma and lung tissue (Kneip et al., 2011; Q.-T. Zhao et al., 2015).

The first of the two epigenome-wide significant DMPs associated with age (cg02843237) was located in the proadrenomedullin gene (*ADM*), which encodes the peptides adrenomedullin (ADM), proadrenomedullin N-terminal 20 (PAMP), and adrenotensin, in addition to mid-regional pro-adrenomedullin (MR-proADM), which currently has no known effect (Valenzuela-Sánchez et al., 2016). Initial studies discovered that ADM plays a role in hypotension (Kitamura et al., 1993), though it has been shown to have pleiotropic effects (Kato & Kitamura, 2015). In the context of ageing, it has been observed levels of ADM increase with age in plasma (Kato et al., 2002), as well as in the brain of healthy human subjects (Larrayoz et al., 2017). ADM and MR-proADM are produced in equal amounts, and as MR-proADM is more stable, this has often been used to assess ADM

levels and similar effects of age have been found in blood (Lorubbio et al., 2018; J. G. Smith et al., 2009). Furthermore, the study by Larrayoz *et al.* (2017) discovered that aged mice lacking ADM and PAMP expression performed better on a memory task, and that this effect was more evident in female mice. Additionally, these female mice showed lower phosphorylated tau accumulation than wild type (WT) mice. This suggests not only a link between ADM and ageing, but also sex and AD, though the current study only found an effect of age on ADM DNA methylation. Of note, CpG sites annotated to the *ADM* gene have been used as surrogate markers for plasma levels of ADM, and have been used in the 'GrimAge' DNA methylation biomarker of epigenetic acceleration of ageing (Lu et al., 2019).

The second age-associated DMP (cg00664416) was also located within an age-associated DMR and is annotated to the mannosidase endo-alpha like (*MANEAL*) gene. A study of long-lived individuals (> 90 years of age) found that *MANEAL* gene expression was upregulated in these individuals in comparison to younger CTLs (20 - 55 years of age). These results were confirmed using the same dataset by Wierczeiko *et al.* (2018), who also showed an association of methylation levels in the promoter region of *MANEAL* with age.

Of the top 10 most significant DMPs for each effect, two were nominally significant in both AddNeuroMed and ADNI: cg14437986, which reflects a sex by age effect and is annotated to *C6orf25*, and cg21484956, which reflects an age by diagnosis effect and is annotated to *CSRP2*. *C6orf25* is an alias for the megakaryocyte and platelet inhibitory receptor G6b (*MPIG6B*) gene, and the cg14437986 probe is located within the sex and sex by age DMR annotated to

MPIG6B. The *MPIG6B* gene is located in the major histocompatibility complex (MHC) class III region, and is part of the immunoglobulin superfamily (de Vet et al., 2001). Although the link between sex, age, and *MPIG6B* is unclear, a recent study of osteosclerosis caused by a lack of *MPIG6B* expression in mice found that the effect of the Mpig6b deficiency was modulated by female sex hormones (Stavnichuk et al., 2021). *CSRP2*, or the 'cysteine and glycine rich protein 2' gene, has not previously been associated with AD.

Two other notable DMRs are those located in the *OXT* and *HOXA4* genes, associated with age and a sex by diagnosis effect, respectively. The ten-probe *OXT* region has previously been identified as a DMR associated with future progression to AD in blood as well as diagnosis of AD in the middle temporal gyrus (MTG, Lardenoije et al., 2019), and an overlapping region containing one additional probe in the superior temporal gyrus (STG, Watson et al., 2016). In both studies, the *OXT* DMR was found to be hypomethylated in AD brain, though Lardenoije *et al*. (2019) found hypermethylation in blood. The *HOXA4* DMR was first identified in an EWAS of AD neuropathology, which detected hypermethylation across an extended *HOXA* gene region in the prefrontal cortex (PFC) and the superior temporal gyrus (STG; R. G. Smith et al., 2018). The region identified in the current study in blood consists of 23 probes, which completely spans the 21-probe region identified in Smith *et al.* (2018). The results from the current study show that in blood, hypermethylation is observed in female subjects at the MCI stage in comparison to CTL and AD, while hypomethylation is observed at this stage in male subjects.

Of note, the top 10 DMPs for each effect, and all identified DMRs were compared to results from the study of McCartney *et al.* (2019), who identified DMPs associated with sex, age, and sex by age interactions in two large cohorts of participants. While both studies found several methylomic differences annotated to the same genes, none of the regions or probes overlapped. For example, McCartney et al. identified four DMPs annotated to the *PIGZ* gene (chr3: 196693980 – 196694932) which were associated with sex, though this did not overlap the sex-, age- and sex by age DMR identified in the current study (chr3: 196705629 – 196705899). One possible explanation for a lack of overlap in results may be that the McCartney *et al.* study included a much wider age range of participants, and did not contain individuals with MCI or AD.

To conclude, the current study has identified several epigenetic signatures in blood, which are related to age, sex, and the interactions of these risk factors with each other and with disease. The importance of not only correcting for risk factors, but examining their interactions with disease has been highlighted, as this may mask findings of interest. However, some limitations of this study should be noted. As mentioned in Chapter 3, while the proportion of cell types is controlled for within this study, it is known that the abundance of specific cell types differs slightly in MCI and AD (Lunnon et al., 2012), which may require the investigation of individual cell types. A major limitation of this study is the sample size: while the sample size of the AddNeuroMed cohort is sufficient to detect methylomic differences in AD (Chapter 3), and sex and age in the current chapter, the addition of interaction effects may require investigation of these signatures in larger cohorts. Additionally, inflation was observed in the *p*-values of the main effects of age and age by diagnosis. Although the influence of confounders

cannot be ruled out, previous studies have shown that many loci are associated with age (Hannum et al., 2013; Peters et al., 2015). In view of the inflation and the limited sample size, a replication analysis was carried out in ADNI, though little overlap between the two cohorts was found. As the AddNeuroMed cohort is European, and the ADNI cohort North-American, the lack of overlap may be related to geographic region. This may be related to the discrepancies found in studies of sex differences in the incidence of AD, where differences were reported in European studies, but no differences were found in studies conducted in the United States (Mielke et al., 2018). Additionally, while the samples in the ADNI cohort were selected to match those in the AddNeuroMed cohort, a significant difference in age between the CTL groups was found (see Chapter 2.3), with ADNI CTLs being slightly older than the CTL subjects in AddNeuroMed. This may also be related to the lack of overlap between the two cohorts in the results related to age and age interactions. Finally, as the AddNeuroMed cohort is targeted at investigating AD, the subjects included in this study were 65 years of age or older, which provides a limited age range for the investigation of age-associated effects on DNA methylation. It would be interesting for future studies to further explore the interactions between sex, age, and diagnosis in cohorts that include participants with a wider age range. This, in combination with an increased sample size, would also enable the investigation of the interactions of age and sex in the context of future conversion to AD.

CHAPTER 5. DNA METHYLATION ON THE X AND Y CHROMOSOME IN ALZHEIMER'S DISEASE AND MILD COGNITIVE IMPAIRMENT

5.1. Introduction

Due to different sex chromosome pairs in human karyotypes, and the resulting inequality of gene content, the epigenetic landscape of the sex chromosomes requires additional regulation in comparison to autosomes. In order to balance the gene dosage between XX females and XY males, X chromosome inactivation (XCI) has evolved as a compensatory mechanism to partially silence transcription. In each cell containing two or more X-chromosomes, one of the two will be inactivated (Xi) while the other remains active (Xa), and this silencing is carried out by various epigenetic mechanisms, including DNA methylation. However, this process does not affect all genes on the Xi chromosome; around 23% of genes escape inactivation, and this may vary by individual, tissue type, and cell type (Tukiainen et al., 2017). In some instances, the selection of which X chromosome is inactivated is not entirely random. This skewed XCI may be caused by structural abnormalities of the X chromosome, or by stochastic variation (Brown & Robinson, 2000). Skewing may occur in healthy females, and has been found to increase with age (Wong et al., 2011), particularly in blood cells (Sharp et al., 2000). In some cases, skewing may lead to disease phenotypes or increased severity of disease, or inversely, X-linked syndrome carriers where the X chromosome containing the mutation is almost fully silenced (Ørstavik, 2006; Vacca et al., 2016). Of note, severely skewed XCI has been reported in peripheral blood samples from women with AD (Bajic et al., 2015). Furthermore, altered gene dosage of the X chromosome has been shown to influence not only gene expression of the X and Y chromosomes, but also influence expression of genes on autosomal chromosomes (Raznahan et al., 2018).

Another X chromosome-related irregularity, which has been found in both lymphocytes and neurons of female AD patients, is premature centromere division (PCD; Spremo-Potparevic et al., 2015; Spremo-Potparević et al., 2008). A study by Spremo-Potparevic *et al*. (2015) found that PCD of a single X chromosome in lymphocytes occurred more often in AD patients compared to CTL. Early studies of PCD of the X chromosome have found that this phenomenon predominantly occurs on partially inactive X chromosomes (Abruzzo et al., 1985; Fitzgerald & McEwan, 1977). PCD is an occurrence that may lead to aneuploidy, an imbalance of chromosomes, which has been found to have a higher occurrence in brains of AD patients, and which has been hypothesised as a cause of neuronal loss (Mosch et al., 2007; Y. Yang et al., 2001). Although the exact mechanisms underlying centromere regulation and chromosome instability are not yet fully determined, DNA methylation has been suggested to play a role (for a review see Herrera et al., 2008; Scelfo & Fachinetti, 2019).

Although DNA methylation assessment at the centromere is very challenging due to its repetitive sequence, and the methodology used in this thesis cannot assess centromere methylation or skewed inactivation, these various irregularities found in X chromosome regulation may indicate a role for this particular chromosome in AD. This is especially interesting, as in most EWAS studies, the X and Y genes are excluded from analyses, as they cannot be analysed in the same way as autosomal chromosomes. As a result, little is known about the DNA methylation profile of sex chromosomes and the Xi chromosome in AD blood. In this chapter, an EWAS on DNA methylation data from the AddNeuroMed cohort is carried out on the X, and the X and Y chromosomes for females and males, respectively, in

a stratified manner. Additionally, the DNA methylation profile of the Xi chromosome in females with AD or MCI was assessed using the method described by Cotton *et al.* (2015).

5.2. Aims

The aims of this chapter are as follows:

1. Identify AD or MCI-associated DMPs and DMRs on the X chromosome in females, and the X and Y chromosomes in males in the AddNeuroMed cohort.
2. Identify AD or MCI-associated DMPs and DMRs on the Xi chromosome in females in the AddNeuroMed cohort.
3. Replicate the DMPs and DMRs identified in AddNeuroMed in the independent ADNI cohort.

5.3. Materials and methods

This chapter uses the epigenome-wide DNA methylation data from the AddNeuroMed cohort, as described in Chapter 2.1. In the current chapter, two analyses were carried out: (1) a stratified analysis on sex chromosomes for males and females separately, and (2) an analysis of X chromosome inactivation in females.

5.3.1. Stratified analysis of the X and Y chromosomes

The normalised data from the AddNeuroMed cohort of AD, MCI, and CTL samples (as described in Chapter 2.2) was split into a set of female-only samples, and a set of male-only samples. The datasets were filtered to include only X chromosome probes, or X and Y chromosome probes for females and males, respectively. With the exception of sex, the same covariates as described in Chapter 2.2.4 (age, cell type proportion, and batch) were regressed out from the sets of male data and female data. For each dataset, an ANOVA and subsequent Tukey's HSD test were carried out to identify DMPs (see Chapter 2.4.1), and *p*-value inflation was inspected using Q-Q plots (Figure 5.1). A *comb-p* regional analysis was carried out to identify DMRs (described in Chapter 2.4.1).

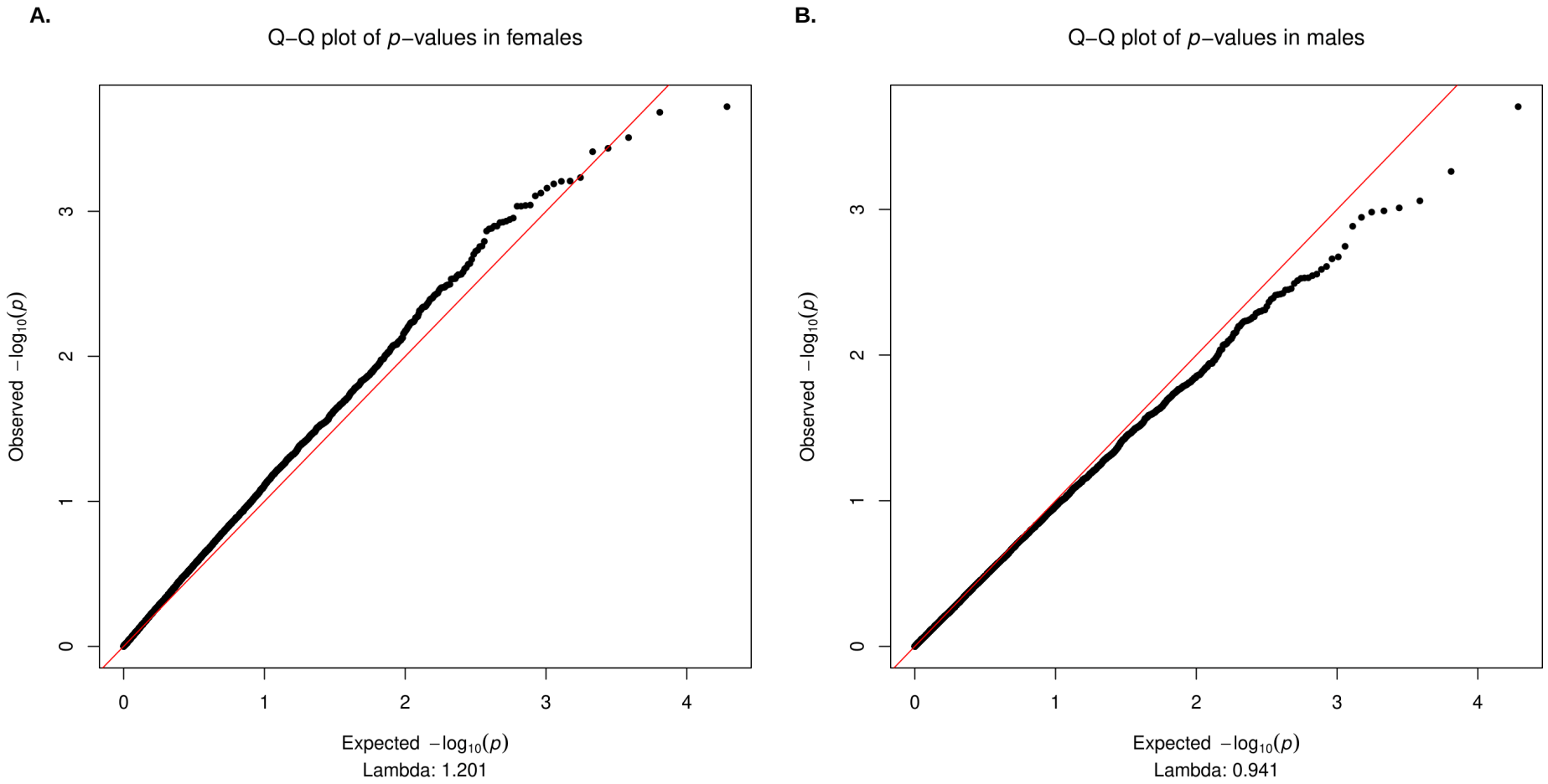

Figure 5.1 Q-Q plots of ANOVA p-values for the X and Y chromosomes in AddNeuroMed.

Shown are the p-values resulting from the analysis of the X chromosome in females (A) and the X and Y chromosomes in males (B).

5.3.2. Analysis of the Xi DNA methylation profile

Cotton *et al.* (2015) have described a method to impute methylation at the Xi by subtracting the mean male DNA methylation at X chromosome probes (representing the Xa methylation) from total DNA methylation levels found at X chromosome probes in females (which contains both Xa and Xi). This gives:

$$X_i = (X_a + X_i) - X_a$$

As the aim of this chapter is to identify DNA methylation changes between individuals with AD, individuals with MCI, and CTLs, this equation was applied to each group separately. The datasets were handled as described in 5.3.1, with the data being split into sets of male- or female-only samples, and the covariates of age, cell type proportion and batch were regressed out from each set. Subsequently, for each diagnostic group an average methylation value was calculated for each X chromosome probe in males. This average was then subtracted from each X chromosome probe in females, separated by diagnostic group. An ANOVA and Tukey's HSD were then run on the dataset of Xi data containing all three groups (CTL, MCI, and AD), as described in Chapter 2.4.1. The Q-Q plot of ANOVA *p*-values showed a large effect of inflation (Figure 5.2). A DMR analysis was carried out as detailed in Chapter 2.4.1.

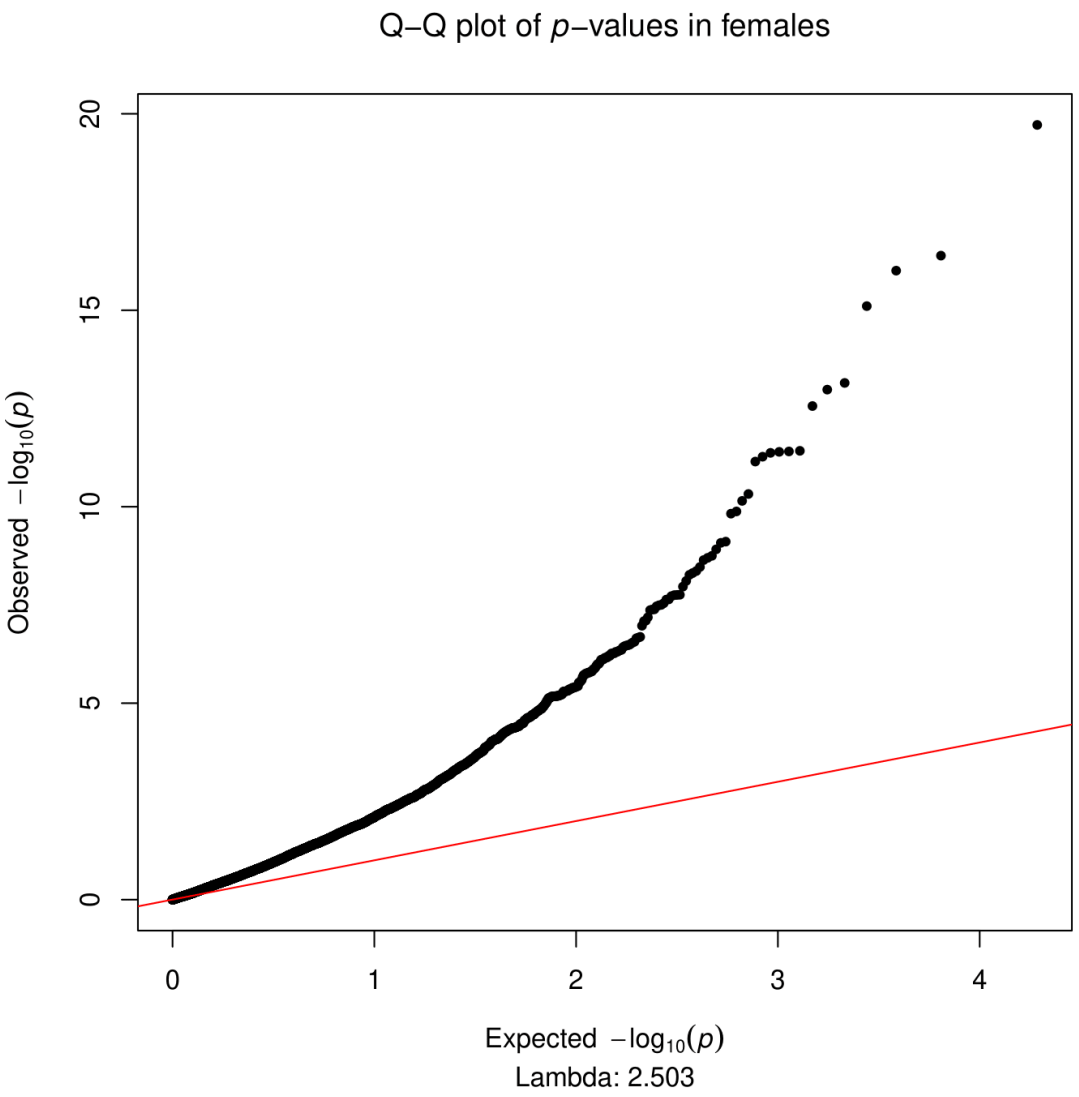

Figure 5.2 Q-Q plots of Xi ANOVA p-values in AddNeuroMed.

Shown are the p-values resulting from the analysis of Xi methylation in females.

5.3.3. Correction of Xi inflation

The large inflation effect seen in the *p*-values of the Xi data may be driven by large biological effects, technical effects and/or unknown variables. To account and adjust for effects of the latter type, two methods were employed in separate models. The first method consisted of a correction using principal components as covariates in the ANOVA model, whereas the second correction model was run using surrogate variables (SVs). For the first method, principal components of the Xi data were generated using the *prcomp* function in R, and the components accounting for the largest amount of variance in the data were selected for further analysis. To ensure only unwanted variation would be removed, each principal component was correlated to diagnosis, and those showing significant correlations were removed from the analysis. Models were then run in an iterative manner, starting with the largest principal component regressed out prior to running the ANOVA analysis of diagnosis, and adding the next largest component in a subsequent model. A total of five models were run, regressing out one to five principal components.

An alternative method of inflation correction was performed using the *sva* package, which was specifically designed for the removal of unwanted sources of variation from high-throughput data (Leek et al., 2012). A total of 43 SVs were generated after the removal of SVs that correlated to diagnosis. In a similar manner to the models run with principal components, ANOVA models were run with one to five SVs regressed out, as well as 10 and all 43 SVs regressed out. Reports of the results and Q-Q plots for each of the principal component and SV models can be found at https://github.com/JanouR/PhD_Thesis. As none of

these methods or models greatly impacted *p*-value inflation, the original model as described in 5.3.2 was selected for this thesis.

5.3.4. Replication analysis

To test the validity of the findings in this chapter, and in view of the inflation found in the Xi results, the stratified and Xi analyses were replicated in the ADNI cohort. The data from this cohort was pre-processed and normalised as described in Chapter 2.3.2, and the analyses were carried out as described in 5.3.1 for the stratified analysis, and 5.3.2 for the analysis of Xi DNA methylation. As no bisulfite batch information was available for the ADNI cohort, only the covariates of age and cell type were regressed out. Similar to the results from the AddNeuroMed cohort, the Q-Q plots of ANOVA *p*-values in ADNI from the stratified analysis show no large inflation of *p*-values (Figure 5.3), whereas the *p*-values from the Xi analysis again show a considerable degree of inflation (Figure 5.4).

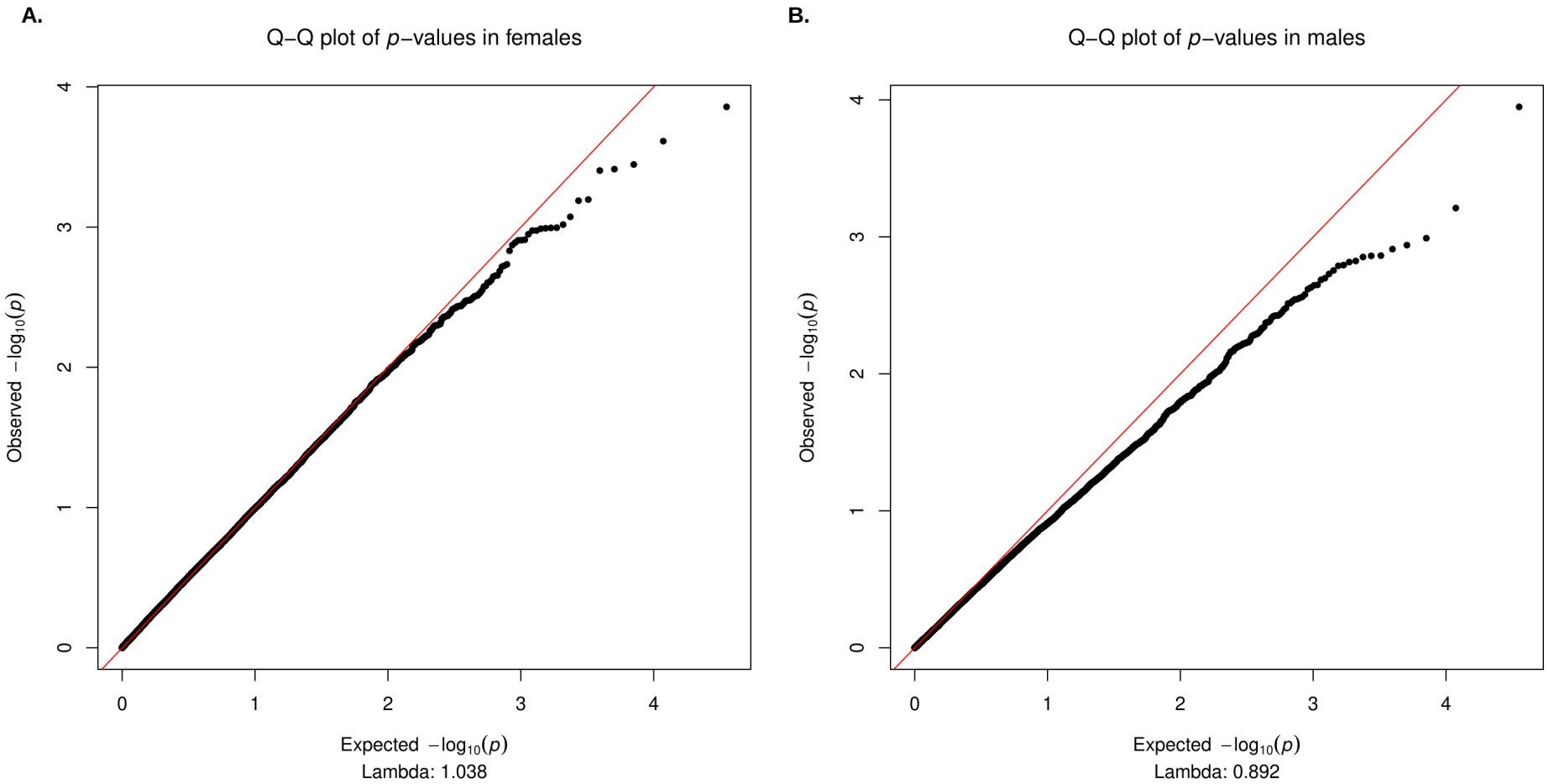

Figure 5.3 Q-Q plots of ANOVA p-values in ADNI.

Shown are the p-values resulting from the analysis of the X chromosome in females (A) and the X and Y chromosomes in males (B) in the ADNI replication cohort.

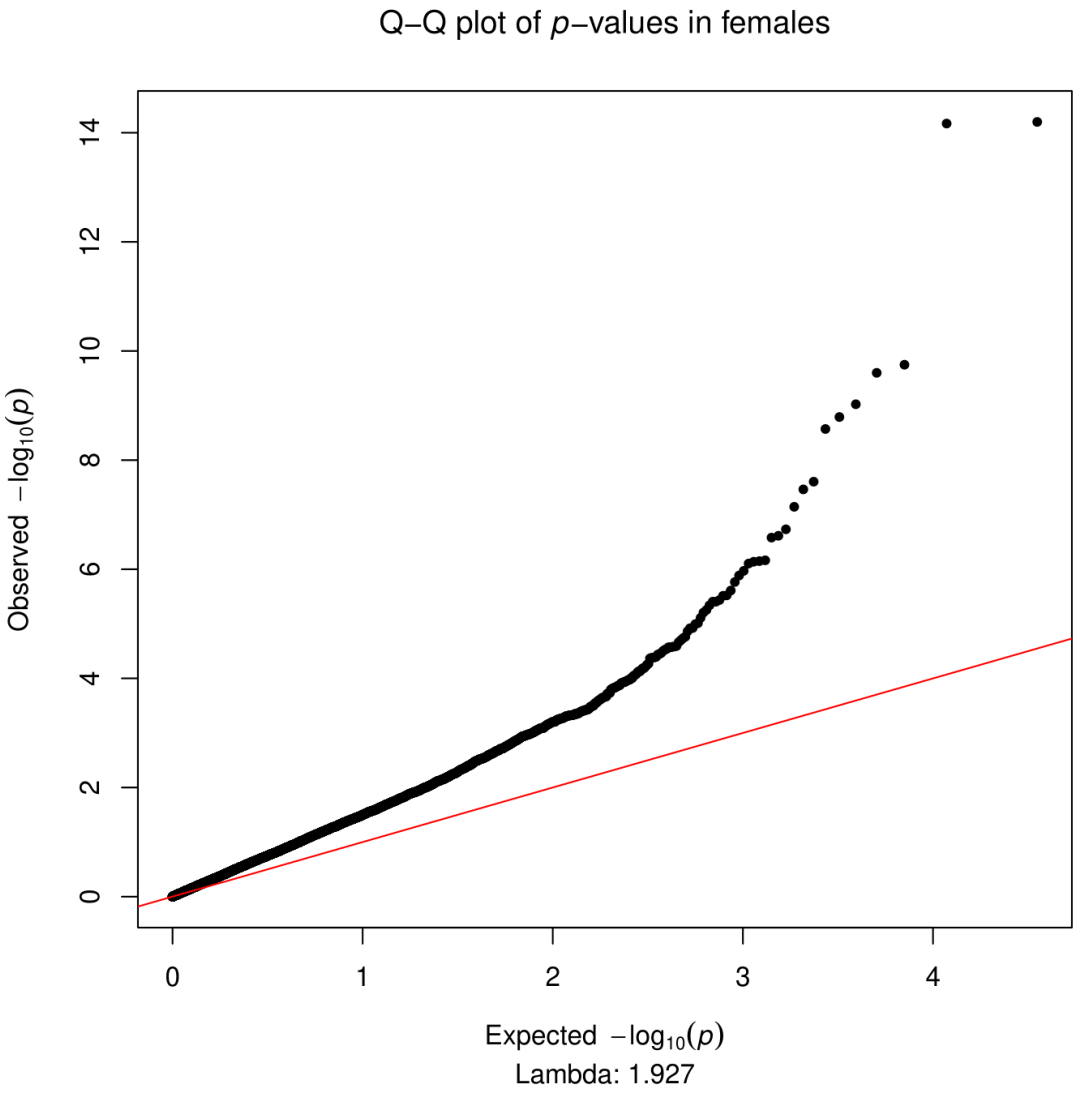

Figure 5.4 Q-Q plots of Xi ANOVA p-values in ADNI.

Shown are the p-values resulting from the analysis of Xi methylation in females in the ADNI cohort.

5.4. Results

5.4.1. Differentially methylated positions on the X and Y chromosomes in AddNeuroMed

In the stratified analysis of the association of DNA methylation and diagnosis in males and females separately, a number of nominally significant sites were found in the AddNeuroMed data. However, none of these sites passed the 450K array epigenome-wide significance threshold of $p < 2.4\times10^{-7}$, or the Bonferroni *p*-value threshold of $p < 5.18\times10^{-6}$ for females, and $p < 5.17\times10^{-6}$ for males (Figure 5.5; Saffari et al., 2018). In the female-only dataset, 644 nominally significant probes ($p < 0.05$) were found, of which the top 10 most significant probes are shown in Table 5.1. The most significant DMP, cg17460003, is located near the *EDA* gene and its ANOVA F-statistic significance was primarily driven by differences between the CTL and MCI group (difference (CI): -0.03 (-0.049 - -0.011), $p = 8.85\times10^{-4}$) and the MCI and AD group (difference (CI): 0.029 (0.01 - 0.048), $p = 1.09\times10^{-3}$). The pattern of methylation at the cg17460003 locus showed hypomethylation in MCI in comparison to CTL and AD.

In males, none of the probes located on the Y chromosome differed in methylation levels between the diagnostic groups ($p < 0.05$). Of the probes located on the X chromosome, there were 411 nominally significant sites associated with diagnosis, of which the most significant probes are shown in Table 5.2. The methylation changes in the most significant probe from the ANOVA, cg26747413, appeared to be driven by differences between the CTL group and the MCI (difference (CI): 0.023 (0.003 - 0.042), $p = 1.72\times10^{-2}$) and between the CTL and AD groups (difference (CI): 0.039 (0.017 - 0.061), $p = 1.33\times10^{-4}$). The MCI and

AD groups showed hypermethylation at the cg26747413 locus in comparison to CTL. This site is annotated to the genes *INE2* and *ZRSR2*. Of the 644 nominally significant DMPs in females, and the 411 nominally significant DMPs in males, 26 probes overlapped, though in only 10 of these probes the methylation differences appeared to be driven by the same diagnostic groups. Out of the 10 overlapping probes, four showed the same direction of effect in the relevant post-hoc comparisons of diagnostic groups. Hypomethylation in AD relative to CTL was found in two of these probes: cg06444329, annotated to *CXorf26* (females: difference (CI): -0.014 (-0.028 - 0), $p = 4.13\times10^{-2}$, males: difference (CI): -0.012 (-0.021 - -0.002), $p = 1.47\times10^{-2}$), and cg17280129, annotated to *NHS* (females: difference (CI): -0.028 (-0.046 - -0.01), $p = 1.08\times10^{-3}$, males: difference (CI): --0.011 (-0.022 - -0.001), $p = 3.27\times10^{-2}$). The probe cg05788681 showed hypermethylation in AD relative to MCI (females: difference (CI): 0.023 (0.006 - 0.04), $p = 5.77\times10^{-3}$, males: difference (CI): 0.018 (0.001 - 0.035), $p = 2.95\times10^{-2}$). Finally, cg12115759 was hypermethylated in MCI relative to CTL (females: difference (CI): 0.04 (0.005 - 0.076), $p = 2.28\times10^{-2}$, males: difference (CI): 0.013 (0.002 - 0.023), $p = 1.26\times10^{-2}$).

.

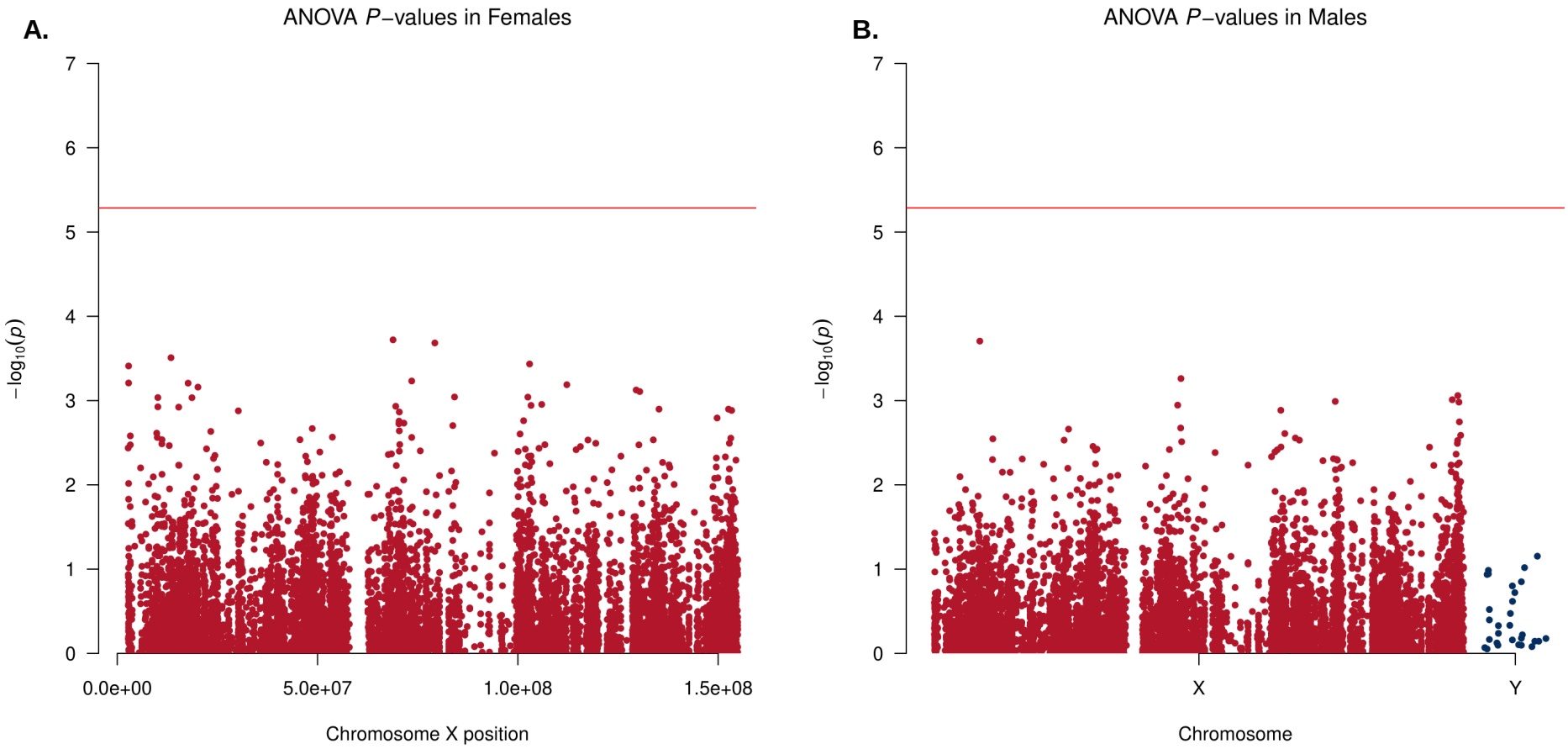

Figure 5.5 Manhattan plots of ANOVA p-values for diagnosis in AddNeuroMed.

Displayed are the ANOVA p-values for the X chromosome in females (A), and the X and Y chromosomes in males (B). The Bonferroni-adjusted significance threshold at 5.18×10^{-6} for females and 5.17×10^{-6} for males is highlighted in red.

Top 10 DMPs from the ANOVA in Females						
ProbeID	**Position**	**F**	**p (F)**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg17460003	chrX:68834205	9.025	1.90E-04			*EDA* (-1705)
cg14917571	chrX:79269151	8.927	2.07E-04	*TBX22*	TSS1500	*ITM2A* (-646103), *TBX22* (-8590)
cg00123001	chrX:13395880	8.483	3.10E-04			*EGFL6* (-191813), *ATXN3L* (-57363)
cg09990582	chrX:102941884	8.296	3.67E-04	*MORF4L2*	TSS200,5'UTR	*MORF4L2* (1201), *TCEAL1* (57993)
cg01337545	chrX:2847549	8.236	3.88E-04	*ARSD*	TSS200	*ARSD* (-134)
cg04667267	chrX:73513464	7.787	5.84E-04	*NCRNA00182*	TSS200	*ZCCHC13* (-10560), *CHIC1* (730481)
cg21559133	chrX:2847557	7.725	6.18E-04	*ARSD*	TSS200	*ARSD* (-142)
cg17280129	chrX:17678361	7.72	6.21E-04	*NHS*	Body	*SCML1* (-77230), *NHS* (284819)
ch.X.112118442F	chrX:112231786	7.675	6.47E-04			*AMOT* (-165415)
cg08129292	chrX:20155820	7.602	6.92E-04	*SCARNA9L;EIF1AX*	TSS1500;Body	*CXorf23* (-167439), *EIF1AX* (4145)

Table 5.1 The top 10 most significant DMPs for diagnosis on the female X chromosome in AddNeuroMed.

DNA methylation profiles were assessed on the X chromosome in females, shown are the top 10 DMPs associated with differences in CTL, MCI or AD diagnosis. For each ProbeID, the genomic location is shown (genome build 37), along with the ANOVA F-statistic and associated p-value (p (F)), the UCSC gene annotation and group, and the GREAT annotation with the distance to the nearest transcription start site in parentheses.

Top 10 DMPs from the ANOVA in Males						
ProbeID	**Position**	**F**	**p (F)**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg26747413	chrX:15807140	9.223	1.97E-04	*INE2;ZRSR2*	TSS1500;TSS1500	*ZRSR2* (-1433)
cg23137494	chrX:73639809	8.039	5.49E-04	*SLC16A2*	TSS1500	*SLC16A2* (-1275)
cg15465642	chrX:153218916	7.511	8.72E-04	*HCFC1*	Body	*RENBP* (-8685), *HCFC1* (17902)
cg12720637	chrX:151619769	7.383	9.77E-04	*GABRA3*	1stExon,5'UTR	*GABRA3* (61)
cg02049865	chrX:117973548	7.331	1.02E-03			*LONRF3* (-135164), *ZCCHC12* (15762)
cg02940221	chrX:153536952	7.307	1.04E-03	*TKTL1*	Body	*TKTL1* (12926), *FLNA* (66053)
cg13075926	chrX:72668640	7.214	1.13E-03	*CDX4*	Body	*CHIC1* (-114343), *CDX4* (1551)
cg16821967	chrX:102348005	7.055	1.31E-03	*NXF3*	5'UTR,1stExon	*NXF3* (16)
cg26070134	chrX:153678842	6.7	1.79E-03	*FAM50A*	3'UTR	*PLXNA3* (-7780), *FAM50A* (6370)
cg15878116	chrX:73524046	6.513	2.12E-03	*ZCCHC13*	1stExon,5'UTR	*ZCCHC13* (22)

Table 5.2 The top 10 most significant DMPs for diagnosis on the male X and Y chromosomes in AddNeuroMed.

DNA methylation profiles were assessed on the X and Y chromosomes in males, displayed are the top 10 DMPs associated with differences in CTL, MCI or AD diagnosis. For each ProbeID the genomic location is shown (genome build 37), along with the ANOVA F-statistic and associated p-value (p (F)), the UCSC gene annotation and group, and the GREAT annotation with the distance to the nearest transcription start site in parentheses.

5.4.1.1. Diagnostic group-specific DMPs in AddNeuroMed

Individual diagnostic groups were compared to each other via the application of Tukey's HSD test following the ANOVA. The site cg14917571 was the most significant DMP in the comparison of MCI to CTL in females. This DMP is annotated to the *TBX22* gene, and hypomethylation was found in MCI relative to CTL (difference (CI): -3.39% (-5.33% - -1.44%), $p = 1.76\times10^{-4}$). In males, the top DMP in the MCI to CTL comparison (cg26070134) is annotated to the *FAM50A* gene, and the locus showed hypermethylation in MCI samples (difference (CI): 0.97% (0.32% - 1.62%), $p = 1.62\times10^{-3}$). The top 10 most significant probes in the CTL vs. MCI comparison can be found in Table 5.3 for females, and Table 5.4 for males.

When comparing AD to CTL (Table 5.5), it was found that in females the top DMP was cg15737490, which is annotated to the *NLGN3* gene, and which showed hypermethylation in AD (difference (CI): 2.82% (1% - 4.63%), $p = 9.69\times10^{-4}$). In males, the most significant difference between AD and CTL was in the cg26747413 probe (hypermethylated in AD: difference (CI): 3.94% (1.73% - 6.14%), $p = 1.33\times10^{-4}$; Table 5.6), which was also the most significant site in the ANOVA comparison of all groups (as described above).

Finally, in the comparison of MCI to AD, the cg09990582 probe was found to be the top significant DMP in females (Table 5.7). This probe is annotated to the *MORF4L2* gene, and the site showed hypermethylation in AD compared to MCI (difference (CI): 4.48 (1.88 - 7.08), $p = 2.12\times10^{-4}$). In males, the top probe was cg02940221, which displayed hypomethylation in AD compared to MCI (difference (CI): -0.016 (-0.027 - -0.006), $p = 9.29\times10^{-4}$). This locus is annotated

to the *TKTL1* gene. The top 10 probes associated with differences between MCI and AD in males are shown in Table 5.8.

The *comb-p* analysis for the identification of DMRs did not detect any DMRs associated with baseline diagnosis, or in any of the subgroup comparisons.

Top 10 CTL vs MCI DMPs in Females						
ProbeID	**Position**	**Difference (CI)**	***p* CvM**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg14917571	chrX:79269151	-3.39 (-5.33 - -1.44)	1.76E-04	*TBX22*	TSS1500	*ITM2A* (-646103), *TBX22* (-8590)
cg00123001	chrX:13395880	-1.42 (-2.24 - -0.6)	1.88E-04			*EGFL6* (-191813), *ATXN3L* (-57363)
cg04882894	chrX:130423289	3.61 (1.4 - 5.83)	4.78E-04	*IGSF1*	5'UTR,1stExon,TSS1500	*IGSF1* (113)
cg05554396	chrX:102510120	2.01 (0.76 - 3.25)	5.60E-04	*TCEAL8*	1stExon,5'UTR	*TCEAL8* (0)
cg16888859	chrX:10123962	-1.81 (-2.94 - -0.69)	5.79E-04	*CLCN4*	TSS1500	*CLCN4* (-1022)
cg21559133	chrX:2847557	2.52 (0.91 - 4.13)	8.23E-04	*ARSD*	TSS200	*ARSD* (-142)
cg17460003	chrX:68834205	-2.97 (-4.88 - -1.07)	8.85E-04			*EDA* (-1705)
cg09614279	chrX:84189578	-3.5 (-5.74 - -1.25)	9.26E-04	*UBE2DNL*	Body	*UBE2DNL* (422)
cg08129292	chrX:20155820	-1.61 (-2.65 - -0.57)	9.67E-04	*SCARNA9L;EIF1AX*	TSS1500;Body	*CXorf23* (-167439), *EIF1AX* (4145)
cg01851385	chrX:10125135	4.54 (1.54 - 7.54)	1.29E-03	*CLCN4*	5'UTR,1stExon	*CLCN4* (151)

Table 5.3 The top 10 most significant DMPs in MCI compared to CTL in females in AddNeuroMed.

DNA methylation profiles were assessed on the X chromosome in females; displayed are the top 10 DMPs associated with differences in MCI diagnosis relative to CTL. For each ProbeID the genomic location is shown (genome build 37), along with the DNA methylation percentage difference between the groups (MCI – CTL), with the confidence interval (CI) in parentheses, the associated Tukey's HSD p-value (p CvM), the UCSC gene annotation and group, and the GREAT annotation with the distance to the nearest transcription start site in parentheses.

Top 10 CTL vs MCI DMPs in Males						
ProbeID	**Position**	**Difference (CI)**	***p* CvM**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg26070134	chrX:153678842	0.97 (0.32 - 1.62)	1.62E-03	*FAM50A*	3'UTR	*PLXNA3* (-7780), *FAM50A* (6370)
cg15465642	chrX:153218916	-1.47 (-2.48 - -0.47)	2.02E-03	*HCFC1*	Body	*RENBP* (-8685), *HCFC1* (17902)
cg16248169	chrX:48334308	0.38 (0.12 - 0.64)	2.34E-03	*FTSJ1*	TSS1500	*FTSJ1* (-240)
cg03768687	chrX:70288358	-0.62 (-1.06 - -0.19)	2.60E-03	*SNX12*	TSS200	*SNX12* (-128)
cg00739582	chrX:145077291	-2.1 (-3.56 - -0.63)	2.76E-03	*MIR888;MIR890*	TSS1500;TSS1500	*CXorf1* (168364)
cg11663393	chrX:100663213	-0.46 (-0.78 - -0.13)	3.01E-03	*HNRNPH2;GLA*	1stExon,5'UTR;TSS1500	*GLA* (-213), *HNRNPH2* (93)
cg25832410	chrX:153096039	-0.56 (-0.96 - -0.16)	3.10E-03	*PDZD4*	TSS200	*PDZD4* (-37)
cg15878116	chrX:73524046	1.12 (0.33 - 1.92)	3.10E-03	*ZCCHC13*	1stExon,5'UTR	*ZCCHC13* (22)
cg03944921	chrX:117479980	0.7 (0.2 - 1.2)	3.41E-03	*WDR44*	TSS200	*WDR44* (-55)
cg02049865	chrX:117973548	-0.88 (-1.51 - -0.25)	3.49E-03			*LONRF3* (-135164), *ZCCHC12* (15762)

Table 5.4 The top 10 most significant DMPs in MCI compared to CTL in males in AddNeuroMed.

DNA methylation profiles were assessed on the X and Y chromosomes in males; displayed are the top 10 DMPs associated with differences in MCI diagnosis relative to CTL. For each ProbeID the genomic location is shown (genome build 37), along with the DNA methylation percentage difference between the groups (MCI – CTL), with the confidence interval (CI) in parentheses, the associated Tukey's HSD p-value (p CvM), the UCSC gene annotation and group, and the GREAT annotation with the distance to the nearest transcription start site in parentheses.

Top 10 CTL vs AD DMPs in Females						
ProbeID	**Position**	**Difference (CI)**	***p* CvA**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg15737490	chrX:70390672	2.82 (1 - 4.63)	9.69E-04	*NLGN3*	3'UTR	GJB1 (-44389), NLGN3 (25992)
cg17280129	chrX:17678361	-2.79 (-4.6 - -0.97)	1.08E-03	*NHS*	Body	SCML1 (-77230), NHS (284819)
cg01337545	chrX:2847549	2.57 (0.85 - 4.29)	1.57E-03	*ARSD*	TSS200	ARSD (-134)
cg26481961	chrX:3263474	1.94 (0.63 - 3.26)	1.76E-03	*MXRA5*	5'UTR	MXRA5 (1209), ARSF (305200)
cg04667267	chrX:73513464	3.02 (0.95 - 5.09)	2.08E-03	*NCRNA00182*	TSS200	ZCCHC13 (-10560), CHIC1 (730481)
cg24636657	chrX:105280886	-2.09 (-3.54 - -0.64)	2.42E-03	*SERPINA7*	1stExon	SERPINA7 (1831), NRK (214351)
cg02746718	chrX:103294519	-1.04 (-1.77 - -0.3)	2.91E-03	*H2BFM;MIR1256*	1stExon,5'UTR;Body	H2BFM (4)
cg15322420	chrX:102942967	3.63 (1.05 - 6.21)	3.14E-03	*MORF4L2*	5'UTR,1stExon,TSS1500	MORF4L2 (118)
cg14736837	chrX:15333837	-1.72 (-2.96 - -0.49)	3.43E-03	*ASB11*	TSS1500,TSS200	ASB11 (-92)
cg23726559	chrX:70288735	2.23 (0.61 - 3.85)	3.93E-03	*SNX12*	TSS1500	SNX12 (-505)

Table 5.5 The top 10 most significant DMPs in AD compared to CTL in females in AddNeuroMed.

DNA methylation profiles were assessed on the X chromosome in females; displayed are the top 10 DMPs associated with differences in AD diagnosis relative to CTL. For each ProbeID the genomic location is shown (genome build 37), along with the DNA methylation percentage difference between the groups (AD – CTL), with the confidence interval (CI) in parentheses, the associated Tukey's HSD p-value (p CvA), the UCSC gene annotation and group, and the GREAT annotation with the distance to the nearest transcription start site in parentheses.

Top 10 CTL vs AD DMPs in Males						
ProbeID	**Position**	**Difference (CI)**	***p* CvA**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg26747413	chrX:15807140	3.94 (1.73 - 6.14)	1.33E-04	*INE2;ZRSR2*	TSS1500;TSS1500	*ZRSR2* (-1433)
cg23137494	chrX:73639809	-2.18 (-3.49 - -0.88)	3.74E-04	*SLC16A2*	TSS1500	*SLC16A2* (-1275)
cg12720637	chrX:151619769	3.51 (1.28 - 5.75)	8.78E-04	*GABRA3*	1stExon,5'UTR	*GABRA3* (61)
cg16821967	chrX:102348005	4.16 (1.4 - 6.92)	1.45E-03	*NXF3*	5'UTR,1stExon	*NXF3* (16)
cg13075926	chrX:72668640	-2.56 (-4.32 - -0.81)	2.12E-03	*CDX4*	Body	*CHIC1* (-114343), *CDX4* (1551)
cg00399450	chrX:73834086	-0.54 (-0.91 - -0.17)	2.26E-03	*RLIM*	5'UTR	*RLIM* (374)
cg02049865	chrX:117973548	-1.01 (-1.72 - -0.3)	2.89E-03			*LONRF3* (-135164), *ZCCHC12* (15762)
cg04768884	chrX:19443172	-3.77 (-6.46 - -1.08)	3.34E-03	*MAP3K15*	Body	*PDHA1* (81162), *MAP3K15* (90206)
cg15465642	chrX:153218916	-1.56 (-2.69 - -0.43)	3.94E-03	*HCFC1*	Body	*RENBP* (-8685), *HCFC1* (17902)
cg11864566	chrX:99667214	2.54 (0.7 - 4.38)	4.00E-03			*PCDH19* (-1944)

Table 5.6 The top 10 most significant DMPs in AD compared to CTL in males in AddNeuroMed.

DNA methylation profiles were assessed on the X and Y chromosomes in males; displayed are the top 10 DMPs associated with differences in AD diagnosis relative to CTL. For each ProbeID the genomic location is shown (genome build 37), along with the DNA methylation percentage difference between the groups (AD – CTL), with the confidence interval (CI) in parentheses, the associated Tukey's HSD p-value (p CvA), the UCSC gene annotation and group, and the GREAT annotation with the distance to the nearest transcription start site in parentheses.

Top 10 MCI vs AD DMPs in Females						
ProbeID	**Position**	**Difference (CI)**	***p* MvA**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg09990582	chrX:102941884	4.48 (1.88 - 7.08)	2.12E-04	*MORF4L2*	TSS200,5'UTR	*MORF4L2* (1201), *TCEAL1* (57993)
cg08681413	chrX:105969177	-4.19 (-6.83 - -1.55)	6.86E-04	*RNF128*	TSS1500,Body	*RNF128* (-716)
cg04778337	chrX:135228444	2.14 (0.78 - 3.5)	7.71E-04	*FHL1*	TSS1500	*FHL1* (-1114)
cg23857909	chrX:153362596	2.15 (0.78 - 3.52)	8.00E-04	*MECP2*	Body,5'UTR	*MECP2* (591)
cg17460003	chrX:68834205	2.91 (1.02 - 4.81)	1.09E-03			*EDA* (-1705)
cg26486175	chrX:18658942	-4.15 (-6.87 - -1.43)	1.19E-03	*CDKL5;RS1*	Body;3'UTR	*RS1* (31280), *CDKL5* (215218)
cg00695046	chrX:100545995	-4.54 (-7.61 - -1.47)	1.73E-03	*TAF7L*	Body,5'UTR	*TAF7L* (2063), *DRP2* (71063)
cg12469471	chrX:153151235	3.28 (1.05 - 5.52)	1.83E-03			*AVPR2* (-19192), *L1CAM* (-9837)
cg24401049	chrX:11157158	-4.24 (-7.15 - -1.34)	2.02E-03	*ARHGAP6*	Body	*AMELX* (-154374), *HCCS* (27753)
cg03552552	chrX:73512004	4.23 (1.33 - 7.13)	2.05E-03	*NCRNA00182*	Body	*ZCCHC13* (-12020), *CHIC1* (729021)

Table 5.7 The top 10 most significant DMPs in AD compared to MCI in females in AddNeuroMed.

DNA methylation profiles were assessed on the X chromosome in females; displayed are the top 10 DMPs associated with differences in AD diagnosis relative to MCI diagnosis. For each ProbeID the genomic location is shown (genome build 37), along with the DNA methylation percentage difference between the groups (AD – MCI), with the confidence interval (CI) in parentheses, the associated Tukey's HSD p-value (p MvA), the UCSC gene annotation and group, and the GREAT annotation with the distance to the nearest transcription start site in parentheses.

Top 10 MCI vs AD DMPs in Males						
ProbeID	**Position**	**Difference (CI)**	***p* MvA**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg02940221	chrX:153536952	-1.65 (-2.7 - -0.59)	9.29E-04	*TKTL1*	Body	*TKTL1* (12926), *FLNA* (66053)
cg23737407	chrX:40029094	1.73 (0.55 - 2.91)	2.10E-03	*BCOR*	5'UTR	*ATP6AP2* (-411121), *BCOR* (-72376)
cg23722334	chrX:106515943	3.13 (0.97 - 5.29)	2.33E-03			*KIAA1817* (-327691), *NUP62CL* (-66274)
cg06578851	chrX:101315561	4.28 (1.3 - 7.26)	2.58E-03			*ZMAT1* (-128523), *TCEAL2* (-65098)
cg21893284	chrX:92928508	-0.66 (-1.14 - -0.18)	4.25E-03	*FAM133A;NAP1L3*	TSS1500;1stExon,5'UTR	*NAP1L3* (173)
cg05129588	chrX:69425049	0.98 (0.25 - 1.71)	5.28E-03	*DGAT2L6*	3'UTR	*AWAT1* (-29455), *DGAT2L6* (27714)
cg05483199	chrX:154056033	3.22 (0.81 - 5.63)	5.48E-03	*LOC100132963*	Body	*MPP1* (-22232), *CXorf68* (6903)
cg09673331	chrX:153692567	0.56 (0.12 - 1.01)	8.15E-03	*PLXNA3*	Body	*PLXNA3* (5945), *LAGE3* (15028)
cg00321056	chrX:53311582	-1.78 (-3.18 - -0.38)	8.85E-03	*IQSEC2*	Body,TSS1500	*KDM5C* (-56979), *IQSEC2* (38939)
cg13243388	chrX:129088172	-2.25 (-4.02 - -0.48)	8.91E-03			*BCORL1* (-50991), *UTP14A* (48076)

Table 5.8 The top 10 most significant DMPs in AD compared to MCI in males in AddNeuroMed.

DNA methylation profiles were assessed on the X and Y chromosomes in males; displayed are the top 10 DMPs associated with differences in AD diagnosis relative to MCI diagnosis. For each ProbeID the genomic location is shown (genome build 37), along with the DNA methylation percentage difference between the groups (AD – MCI), with the confidence interval (CI) in parentheses, the associated Tukey's HSD p-value (p MvA), the UCSC gene annotation and group, and the GREAT annotation with the distance to the nearest transcription start site in parentheses.

5.4.2. Replication of X and Y linked DMPs in the ADNI cohort

The stratified assessment of DNA methylation on the X and Y chromosomes was repeated on the ADNI cohort data, which had been pre-processed and normalised (as described in Chapter 2.3.2). As also found in the AddNeuroMed cohort, no epigenome-wide significant DMPs were identified at the epigenome-wide threshold established for the EPIC array ($p < 9\times10^{-8}$; Mansell et al., 2019), or at the Bonferroni-corrected *p*-value threshold for females ($p < 2.83\times10^{-6}$) or males ($p < 2.81\times10^{-6}$; Figure 5.6). For the purpose of replication the results were examined for MCI or AD-related probes that were nominally significant ($p < 0.05$) and showed the same direction of effect in both cohorts.

In the set of female data, there were a total of 34 probes that were nominally significant in both datasets. Looking at the between-group differences, there were seven probes in the CTL vs MCI comparison in females which overlapped between cohorts (Table 5.9). Not only did these probes show the same direction of effect, they also showed a similar magnitude of effect between both cohorts. Hypomethylation in MCI relative to CTL was found in three probes, annotated to the genes *HUWE1* (cg18450499), *WWC3* (cg03449040), and *IQSEC2* (cg03832506). The other four probes showed hypermethylation, and were annotated to the genes *SMARCA1* (cg17420696), *HTR2C* (cg02918903), *SMPX* (cg27013947), and *PRKX* (cg25396787).

In the comparison of CTL to AD in females, there were five loci that fulfilled the criteria of nominal significance and same direction of effect (Table 5.10). All probes showed hypermethylation in AD in comparison to CTL, and these sites were annotated to *NLGN3* (cg15737490), *TSPYL2* (cg23612178), *ARMCX3*

(cg05453458), and *PHKA2* (cg06104510). One site (cg05548952) did not have a UCSC annotated gene, but the closest TSS was located in the gene *UTP14A*. For each of these sites, the effect size was slightly larger in the ADNI cohort compared to the AddNeuroMed cohort.

Two dataset-overlapping loci were identified in the comparison of MCI to AD (Table 5.11). One probe (cg07112779) annotated to *ERAS*, showed hypermethylation in AD, while the other probe (cg10841338) was hypomethylated in AD relative to MCI. The nearest TSS to this probe is annotated to the gene *GPR50*.

Less overlap was found in the DNA methylation profiles of the male X and Y chromosomes: only 10 probes had a nominally significant F-statistic for the overall ANOVA test in both cohorts. In the comparison of between-group differences, there were several probes overlapping, but only in the CTL to MCI comparison did two probes show the same direction of effect (Table 5.12). In both probes hypomethylation was found in MCI relative to CTL, and these probes were annotated to *LOC100132963* (cg05483199), and *DUSP9* (cg26159385).

Finally, a DMR analysis was run on the ADNI cohort data for the separate female and male datasets. However, the analysis did not identify any regions of interest, similarly to the AddNeuroMed data.

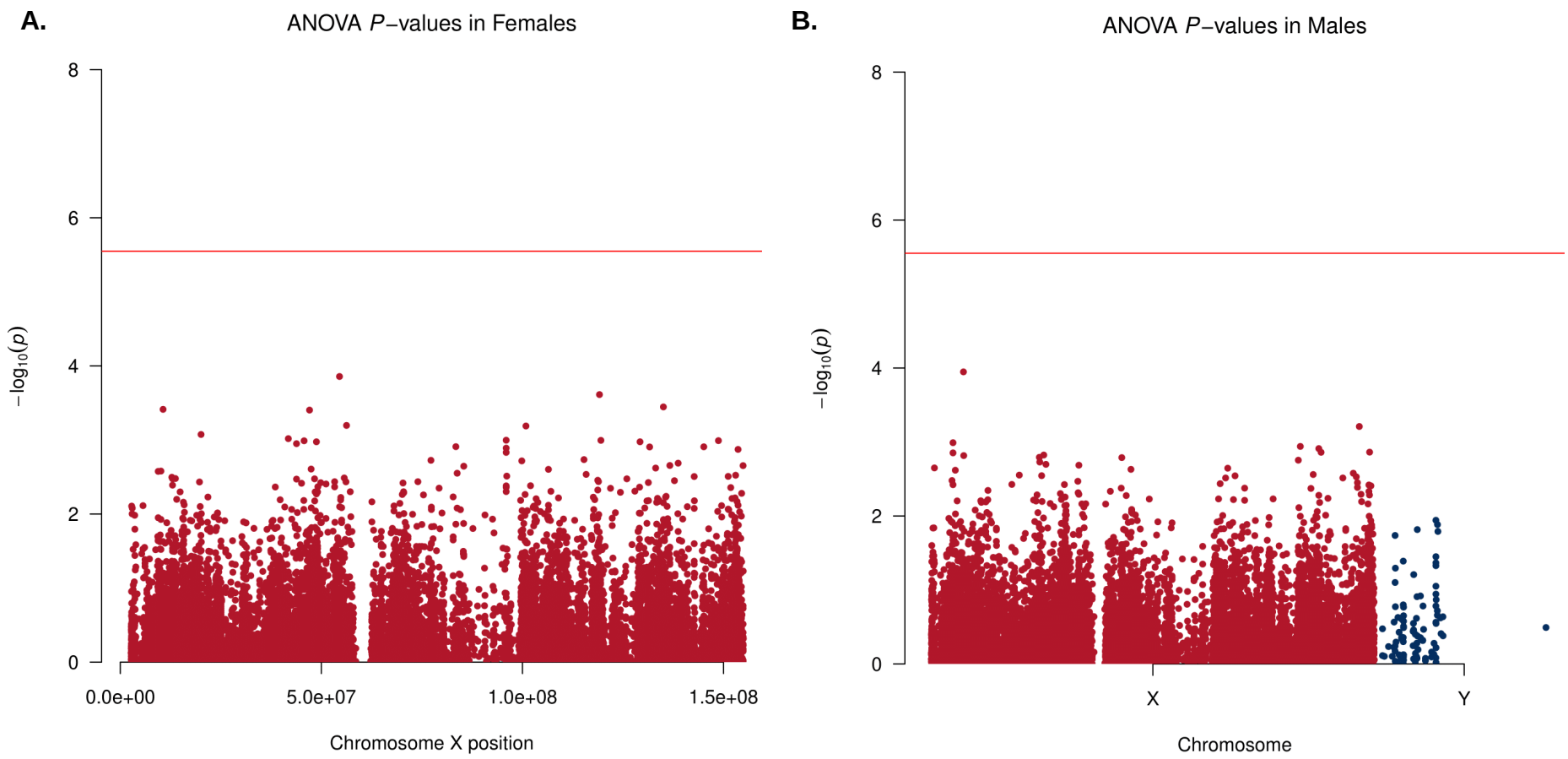

Figure 5.6 Manhattan plots of ANOVA p-values for diagnosis in the ADNI replication cohort.

Displayed are the ANOVA p-values for the X chromosome in females (A), and the X and Y chromosomes in males (B). The Bonferroni-adjusted significance threshold at 2.83×10^{-6} for females and 2.81×10^{-6} for males is highlighted in red.

Female CTL vs MCI DMPs								
ProbeID	**Position**	**AddNeuroMed**		**ADNI**		**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
		Difference	***p* CvM**	**Difference**	***p* CvM**			
cg18450499	chrX:53707967	-2.69 (-4.52 - -0.87)	1.76E-03	-1.99 (-3.76 - -0.21)	2.43E-02	*HUWE1*	5'UTR	*HSD17B10* (-246645), *HUWE1* (5705)
cg17420696	chrX:128657092	3.19 (0.37 - 6.01)	2.22E-02	2.95 (0.03 - 5.87)	4.66E-02	*SMARCA1*	Body	*SMARCA1* (367)
cg25396787	chrX:3733991	1.37 (0.12 - 2.62)	2.78E-02	1.26 (0.02 - 2.5)	4.53E-02			*PRKX* (-102317)
cg02918903	chrX:114141866	3.75 (0.25 - 7.24)	3.25E-02	3.39 (0.29 - 6.5)	2.84E-02	*HTR2C*	Body	*IL13RA2* (110340), *HTR2C* (323316)
cg03449040	chrX:10094363	-0.69 (-1.34 - -0.04)	3.54E-02	-0.7 (-1.17 - -0.23)	1.68E-03	*WWC3*	Body	*CLCN4* (-30621), *SHROOM2* (339868)
cg03832506	chrX:53286187	-2.21 (-4.29 - -0.12)	3.54E-02	-2.38 (-4.06 - -0.7)	2.84E-03	*IQSEC2*	Body	*KDM5C* (-31584), *IQSEC2* (64334)
cg27013947	chrX:21776178	2.29 (0.01 - 4.57)	4.83E-02	2.69 (0.65 - 4.73)	5.96E-03	*SMPX*	1stExon,5'UTR	*SMPX* (99)

Table 5.9. AddNeuroMed DMPs replicated in the ADNI cohort for the female X chromosome in the MCI to CTL comparison.

Shown are probes that were nominally significant ($p < 0.05$) in the comparison of MCI diagnosis to CTL, and showed the same direction of effect in the AddNeuroMed cohort as well as the ADNI cohort. Shown for each ProbeID is the genomic location (genome build 37), the group difference in % methylation (MCI – CTL), confidence interval (CI), and p-value as calculated with Tukey's HSD test in the AddNeuroMed and ADNI cohorts, the UCSC gene annotation and group, and the GREAT annotation with the distance to the nearest transcription start site in parentheses.

Female CTL vs AD DMPs								
ProbeID	**Position**	**AddNeuroMed**		**ADNI**		**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
		Difference	***p* CvA**	**Difference**	***p* CvA**			
cg15737490	chrX:70390672	2.82 (1 - 4.63)	9.69E-04	3.21 (0.49 - 5.94)	1.62E-02	*NLGN3*	3'UTR	*GJB1* (-44389), *NLGN3* (25992)
cg23612178	chrX:53111677	1.74 (0.13 - 3.35)	3.09E-02	2.63 (0.15 - 5.11)	3.51E-02	*TSPYL2*	1stExon,5'UTR	*TSPYL2* (136)
cg05548952	chrX:129068271	2.92 (0.18 - 5.65)	3.40E-02	3.45 (0.17 - 6.74)	3.64E-02			*BCORL1* (-70892), *UTP14A* (28175)
cg05453458	chrX:100878340	2.11 (0.02 - 4.2)	4.68E-02	2.74 (0.04 - 5.44)	4.62E-02	*ARMCX3*	TSS200,5'UTR	*ARMCX3* (221)
cg06104510	chrX:19002482	1.85 (0 - 3.69)	4.96E-02	4.4 (0.99 - 7.81)	7.43E-03	*PHKA2*	TSS200	*PHKA2* (-3)

Table 5.10 AddNeuroMed DMPs replicated in the ADNI cohort for the female X chromosome in the AD to CTL comparison.

Shown are probes that were nominally significant (p < 0.05) in the comparison of AD diagnosis to CTL, and showed the same direction of effect in the AddNeuroMed cohort as well as the ADNI cohort. Shown for each ProbeID is the genomic location (genome build 37), the group difference in % methylation (AD – CTL), confidence interval (CI), and p-value as calculated with Tukey's HSD test in the AddNeuroMed and ADNI cohorts, the UCSC gene annotation and group, and the GREAT annotation with the distance to the nearest transcription start site in parentheses.

Female MCI vs AD DMPs								
ProbeID	**Position**	**AddNeuroMed**		**ADNI**		**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
		Difference	***p* MvA**	**Difference**	***p* MvA**			
cg07112779	chrX:48687187	1.87 (0.53 - 3.2)	3.28E-03	2.16 (0.17 - 4.14)	2.97E-02	*ERAS*	TSS200	*ERAS* (-95)
cg10841338	chrX:150342651	-2.7 (-5.32 - -0.09)	4.10E-02	-4.5 (-8.39 - -0.61)	1.88E-02			*GPR50* (-2404)

Table 5.11 AddNeuroMed DMPs replicated in the ADNI cohort for the female X chromosome in the AD to MCI comparison.

Shown are probes that were nominally significant ($p < 0.05$) in the comparison of AD to MCI diagnosis, and showed the same direction of effect in the AddNeuroMed cohort as well as the ADNI cohort. Shown for each ProbeID is the genomic location (genome build 37), the group difference in % methylation (AD – MCI), confidence interval (CI), and p-value as calculated with Tukey's HSD test in the AddNeuroMed and ADNI cohorts, the UCSC gene annotation and group, and the GREAT annotation with the distance to the nearest transcription start site in parentheses.

Male CTL vs MCI DMPs								
ProbeID	**Position**	**AddNeuroMed**		**ADNI**		**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
		Difference	***p* CvM**	**Difference**	***p* CvM**			
cg05483199	chrX:154056033	-2.63 (-4.95 - -0.31)	2.21E-02	-2.24 (-4.33 - -0.16)	3.15E-02	LOC100132963	Body	MPP1 (-22232), CXorf68 (6903)
cg26159385	chrX:152912611	-0.67 (-1.35 - 0)	5.00E-02	-0.82 (-1.6 - -0.03)	3.85E-02	DUSP9	5'UTR	DUSP9 (4715), PNCK (27204)

Table 5.12 AddNeuroMed DMPs replicated in the ADNI cohort for the male X chromosome.

Shown are probes that are nominally significant ($p < 0.05$) and showed the same direction of effect in the AddNeuroMed cohort as well as the ADNI cohort. In the male dataset, replicated probes were only found in the CTL versus MCI comparison. Shown for each ProbeID is the genomic location (genome build 37), the group difference in % methylation (MCI – CTL), confidence interval (CI), and p-value as calculated with Tukey's HSD test in the AddNeuroMed and ADNI cohorts, the UCSC gene annotation and group, and the GREAT annotation with the distance to the nearest transcription start site in parentheses.

5.4.3. DMPs on Xi in AddNeuroMed

For the second goal of this chapter, the DNA methylation values of the Xi chromosome were calculated as detailed in section 5.3.2, and an ANOVA was performed to assess diagnosis-associated differences. A total of 112 DMPs were found to pass the Bonferroni-adjusted *p*-value threshold of $p < 5.18\times10^{-6}$, of which 49 passed the epigenome-wide threshold ($p < 2.4\times10^{-7}$, Figure 5.7) The 10 most significant DMPs are shown in Table 5.13, a list of all Bonferroni-significant DMPs can be found in Appendix D. As noted, *p*-values were found to be inflated (Figure 5.2). To minimise effects of confounding variables, alternative models were run to correct for inflation (see 5.3.3), though these models showed no improvement in the inflation of the *p*-values. Therefore, the results reported here refer to the initial ANOVA model which included only the covariates of age, cell type, and batch.

The most significant DMP associated with overall differences in diagnostic status was cg22822140, which is annotated to the first exon of the *SRPK3* gene. This significance was driven by differences between the CTL and AD groups and the MCI and AD groups, with hypomethylation observed in the AD group compared to CTL (difference (CI): -9.53% (-11.83% - -7.22%), $p = 4.92\times10^{-14}$) and MCI (difference (CI): -8.86% (-11.12% - -6.59%), $p = 4.98\times10^{-14}$). This DMP was also the top-ranking DMP for the CTL to AD comparison, in which 40 DMPs were found to pass the Bonferroni-adjusted threshold (see Table 5.14 for the top 10 DMPs, and Appendix E for a full list of all 40 significant DMPs). Of note, six of the 40 CTL vs. AD DMPs were annotated to the *SRPK3* gene. Furthermore, the MCI to AD comparison identified 38 DMPs, of which five were located in *SRPK3* and overlapped with those identified in the CTL to AD comparison.

Further examining the between-group comparisons, 55 DMPs were associated with differences in the CTL and MCI groups (see Table 5.15 for the top 10, or Appendix F for the full list). The most significant DMP (cg24605338) is annotated to the gene *NHS*, and was found to be hypomethylated in MCI relative to CTL (difference (CI): -3.68% (-4.8% - -2.56%), $p = 2.63\times10^{-12}$). In the MCI to AD comparison the top DMP is annotated to *TAF7L* (Top 10: Table 5.16, full list: Appendix G), and was hypermethylated in AD in comparison to MCI (difference (CI): 2.84% (2.15% - 3.53%), $p = 4.79\times10^{-14}$).

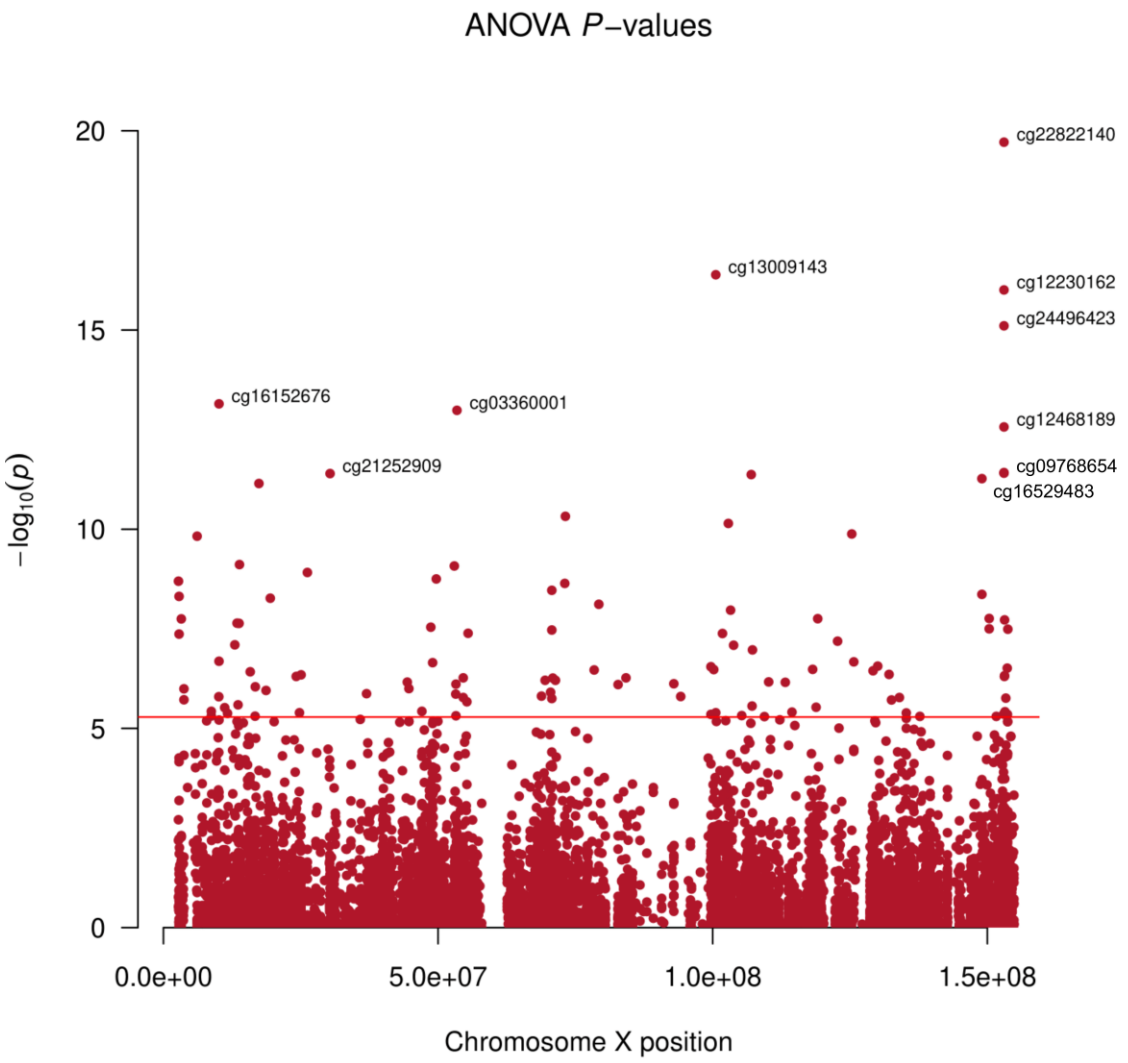

Figure 5.7 Manhattan plot of the Xi ANOVA p-values in AddNeuroMed.

Imputed Xi DNA methylation values were assessed for differences associated with diagnostic status (CTL, MCI, and AD). Shown are the log-inverted p-values from the ANOVA model, with annotations for the top 10 most significant loci. The red line indicates the Bonferroni-adjusted significance threshold of $p < 5.18 \times 10^{-6}$.

Top 10 ANOVA Xi DMPs						
ProbeID	**Position**	**F**	***p* (F)**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg22822140	chrX:153046577	60.317	1.92E-20	*SRPK3*	1stExon	*SRPK3* (122)
cg13009143	chrX:100548223	47.710	4.08E-17	*TAF7L*	TSS200	*TAF7L* (-165)
cg12230162	chrX:153046482	46.334	9.84E-17	*SRPK3*	5'UTR,1stExon	*SRPK3* (27)
cg24496423	chrX:153046480	43.144	7.85E-16	*SRPK3*	5'UTR,1stExon	*SRPK3* (25)
cg16152676	chrX:10126878	36.492	7.10E-14	*CLCN4*	5'UTR	*CLCN4* (1894), *MID1* (724930)
cg03360001	chrX:53453272	35.944	1.04E-13	*RIBC1*	Body	*SMC1A* (-3655)
cg12468189	chrX:153046767	34.577	2.72E-13	*SRPK3*	Body	*SRPK3* (312)
cg09768654	chrX:153046386	30.912	3.79E-12	*SRPK3*	TSS200	*SRPK3* (-69)
cg16529483	chrX:153046451	30.870	3.91E-12	*SRPK3*	TSS200	*SRPK3* (-4)
cg21252909	chrX:30326328	30.829	4.03E-12	*NR0B1*	1stExon	*NR0B1* (1166)

Table 5.13 Top 10 Xi DMPs associated with CTL, MCI, and AD in AddNeuroMed.

Top 10 most significant loci displaying DNA methylation differences on the Xi chromosome related to CTL, MCI, or AD status. All probes are shown by ProbeID, genomic location (genome build 37), ANOVA F-statistic and p-value, UCSC gene and group annotation, and GREAT annotation with distance to the nearest TSS in parentheses. A full list of all DMPs passing the Bonferroni-adjusted p-value threshold of $p < 5.18\times10^{-6}$ is shown in Appendix D.

Top 10 CTL vs AD DMPs						
ProbeID	**Position**	**Difference (CI)**	***p* CvA**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg22822140	chrX:153046577	-9.53 (-11.83 - -7.22)	4.92E-14	*SRPK3*	1stExon	*SRPK3* (122)
cg12468189	chrX:153046767	-5.4 (-6.94 - -3.86)	1.52E-13	*SRPK3*	Body	*SRPK3* (312)
cg12230162	chrX:153046482	-8.72 (-11.28 - -6.16)	4.81E-13	*SRPK3*	5'UTR,1stExon	*SRPK3* (27)
cg24496423	chrX:153046480	-6.66 (-8.68 - -4.63)	2.27E-12	*SRPK3*	5'UTR,1stExon	*SRPK3* (25)
cg19938385	chrX:107020894	-5.93 (-7.81 - -4.06)	1.24E-11			*TSC22D3* (-1878)
cg17405188	chrX:73164102	1.72 (1.17 - 2.27)	1.80E-11	*LOC554203*	TSS200	*ZCCHC13* (-359922), *CHIC1* (381119)
cg16529483	chrX:153046451	-7.68 (-10.23 - -5.12)	1.01E-10	*SRPK3*	TSS200	*SRPK3* (-4)
cg23914849	chrX:52950051	0.68 (0.45 - 0.91)	3.43E-10			*FAM156B* (21967), *FAM156A* (74599)
cg04710661	chrX:2848381	1.87 (1.19 - 2.56)	3.84E-09	*ARSD*	TSS1500	*ARSD* (-966)
cg04768884	chrX:19443172	3.71 (2.3 - 5.13)	1.34E-08	*MAP3K15*	Body	*PDHA1* (81162), *MAP3K15* (90206)

Table 5.14 Top 10 Xi DMPs associated with AD in AddNeuroMed.

DNA methylation differences on the Xi chromosome related to diagnosis of AD in comparison to CTL. The top 10 most significant probes are shown by ProbeID, genomic location (genome build 37), the group difference in % methylation (AD – CTL), confidence interval (CI), and p-value as calculated with Tukey's HSD test, UCSC gene and group annotation, and GREAT annotation with distance to the nearest TSS in parentheses. The full list of all DMPs passing the Bonferroni-adjusted p-value threshold of $p < 5.18 \times 10^{-6}$ *is shown in Appendix E.*

Top 10 CTL vs MCI DMPs						
ProbeID	**Position**	**Difference (CI)**	***p* CvM**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg24605338	chrX:17395856	-3.68 (-4.8 - -2.56)	2.63E-12	*NHS*	Body	*SCML1* (-359735), *NHS* (2314)
cg17003204	chrX:102861456	-3.76 (-5.02 - -2.5)	1.21E-10	*TCEAL3*	TSS1500	*TCEAL3* (-1377)
cg19280671	chrX:149013642	1.76 (1.17 - 2.35)	1.63E-10	*MAGEA8*	Body	*MAMLD1* (-517908), *MAGEA8* (3702)
cg16152676	chrX:10126878	4.18 (2.74 - 5.62)	4.19E-10	*CLCN4*	5'UTR	*CLCN4* (1894), *MID1* (724930)
cg17513789	chrX:73073251	-4.75 (-6.42 - -3.07)	9.34E-10	*XIST*	TSS1500	*ZCCHC13* (-450773), *CHIC1* (290268)
cg15907464	chrX:70712790	-3.53 (-4.8 - -2.27)	1.37E-09	*BCYRN1;INGX*	Body;TSS200	*OGT* (-40121), *TAF1* (126677)
cg14917571	chrX:79269151	-5.32 (-7.26 - -3.37)	3.18E-09	*TBX22*	TSS1500	*ITM2A* (-646103), *TBX22* (-8590)
cg16122592	chrX:26210712	-3.11 (-4.25 - -1.97)	3.57E-09	*MAGEB6*	5'UTR	*MAGEB6* (156)
cg06452970	chrX:13835264	5.1 (3.2 - 7.01)	6.56E-09	*GPM6B*	1stExon,5'UTR,Body	*GPM6B* (49)
cg09953945	chrX:150345040	-2.98 (-4.1 - -1.86)	7.92E-09	*GPR50*	TSS200	*GPR50* (-15)

Table 5.15 Top 10 Xi DMPs associated with MCI in AddNeuroMed.

DNA methylation differences on the Xi chromosome related to diagnosis of MCI in comparison to CTL. The top 10 most significant probes are shown by ProbeID, genomic location (genome build 37), the group difference in % methylation (MCI – CTL), confidence interval (CI), and p-value as calculated with Tukey's HSD test, the UCSC gene annotation and group, and the GREAT annotation with the distance to the nearest transcription start site in parentheses. The full list of all DMPs passing the Bonferroni-adjusted p-value threshold of $p < 5.18 \times 10^{-6}$ is shown in Appendix F.

Top 10 MCI vs AD DMPs						
ProbeID	**Position**	**Difference (CI)**	***p* MvA**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg13009143	chrX:100548223	2.84 (2.15 - 3.53)	4.79E-14	*TAF7L*	TSS200	*TAF7L* (-165)
cg22822140	chrX:153046577	-8.86 (-11.12 - -6.59)	4.98E-14	*SRPK3*	1stExon	*SRPK3* (122)
cg12230162	chrX:153046482	-9.17 (-11.68 - -6.65)	6.34E-14	*SRPK3*	5'UTR,1stExon	*SRPK3* (27)
cg03360001	chrX:53453272	2.12 (1.53 - 2.71)	8.63E-14	*RIBC1*	Body	*SMC1A* (-3655)
cg24496423	chrX:153046480	-7 (-8.99 - -5.01)	1.45E-13	*SRPK3*	5'UTR,1stExon	*SRPK3* (25)
cg21252909	chrX:30326328	-7.69 (-10.01 - -5.37)	1.56E-12	*NR0B1*	1stExon	*NR0B1* (1166)
cg16152676	chrX:10126878	-4.73 (-6.17 - -3.3)	2.08E-12	*CLCN4*	5'UTR	*CLCN4* (1894), *MID1* (724930)
cg09768654	chrX:153046386	-7.07 (-9.35 - -4.8)	2.41E-11	*SRPK3*	TSS200	*SRPK3* (-69)
cg22926378	chrX:6144633	-2.13 (-2.86 - -1.41)	2.60E-10	*NLGN4X*	5'UTR	*NLGN4X* (2072)
cg25896901	chrX:125300481	-7.5 (-10.11 - -4.89)	5.28E-10	*DCAF12L2*	TSS1500	*DCAF12L1* (386360)

Table 5.16 Top 10 Xi DMPs associated with differences in AD relative to MCI in AddNeuroMed.

DNA methylation differences on the Xi chromosome related to diagnosis of AD in comparison to MCI. The 10 most significant probes are shown by ProbeID, genomic location (genome build 37), the group difference in % methylation (AD – MCI), confidence interval (CI), and p-value as calculated with Tukey's HSD test, UCSC gene and group annotation, and GREAT annotation with distance to the nearest TSS in parentheses. The full list of all DMPs passing the Bonferroni-adjusted p-value threshold of $p < 5.18 \times 10^{-6}$ *is shown in Appendix G.*

5.4.4. Replication of Xi DMPs in ADNI

A replication analysis was also carried out for the analysis of Xi DNA methylation in the ADNI cohort. Similar to the results from the AddNeuroMed cohort, a large inflation of ANOVA *p*-values was found in the ADNI cohort (Figure 5.2; Figure 5.4). In the ANOVA comparison of all groups, 21 DMPs were found which passed the Bonferroni-adjusted *p*-value threshold of $p < 2.83\times10^{-6}$, of which 10 passed the epigenome-wide threshold established for the EPIC array (9×10^{-8}; Mansell et al., 2019; Figure 5.8). Of the Bonferroni-significant DMPs identified in the ADNI cohort, two DMPs overlapped with the Bonferroni-significant DMPs discovered in the AddNeuroMed cohort. In ADNI, the most significant ANOVA DMP was found to be probe cg09513996, which does not have an associated UCSC gene annotation, but is annotated to the gene *ERAS* in the GREAT annotation (F = 39.295, $p = 6.34\times10^{-15}$). The changes at this site were driven by differences between the MCI and AD groups (difference (CI): 5.58% (4.09% - 7.07%), $p = 6.49\times10^{-14}$). However, while this site was also identified as a DMP related to MCI in comparison to AD in AddNeuroMed (F = 19.303, $p = 2.87\times10^{-8}$), it was found to have the opposite effect in the discovery cohort. Whereas hypermethylation in AD relative to MCI was observed in the ADNI cohort, hypomethylation in AD was observed in AddNeuroMed (difference (CI): -3.47% (-4.86% - -2.08%), $p = 5.67\times10^{-8}$).

The second replicated DMP associated with overall diagnosis, was cg27558057 (AddNeuroMed: F = 14.353, $p = 1.77\times10^{-6}$, ADNI: F = 24.999, $p = 2.52\times10^{-10}$). In ADNI, the differences at this site were driven by hypomethylation in AD relative to both CTL (difference (CI): -5.22% (-7.15% - -3.29%), $p = 4.12\times10^{-9}$) and MCI (difference (CI): -5.36% (-7.26% - -3.46%), $p = 9.33\times10^{-10}$). Conversely, in

AddNeuroMed the group differences were driven by hypomethylation in MCI relative to CTL (difference (CI): -3% (-4.33% - -1.67%), $p = 9.5\times10^{-7}$). This locus is annotated to the genes *BCYRN1* and *INGX*.

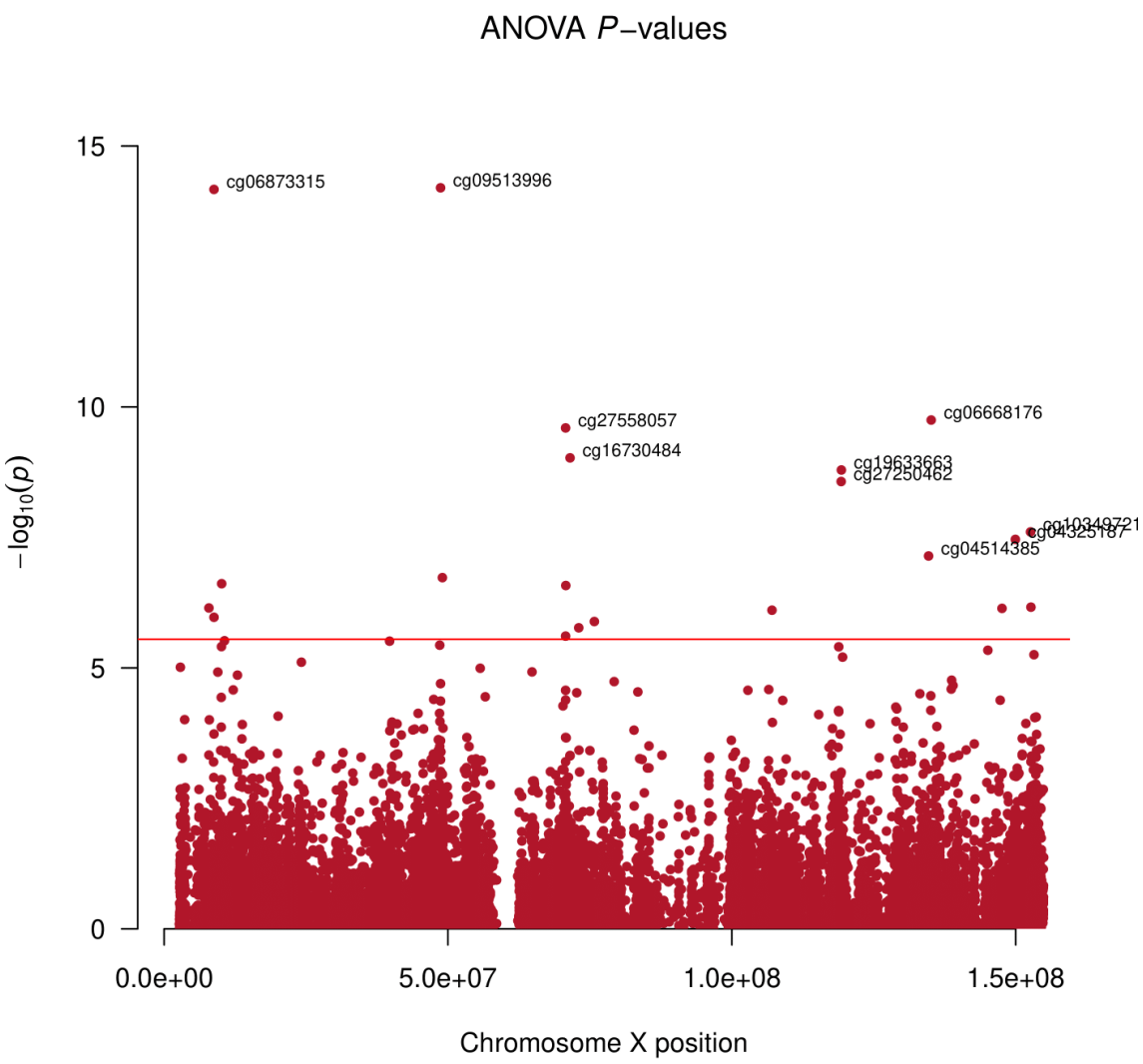

Figure 5.8 Manhattan plot of the Xi ANOVA p-values in the ADNI cohort.

Imputed Xi DNA methylation values were assessed for differences associated with diagnostic status (CTL, MCI, and AD). Shown are the log-inverted p-values from the ANOVA model, with annotations for the top 10 most significant loci. The Bonferroni-adjusted threshold of $p < 2.83 \times 10^{-6}$ is indicated by the red line.

5.4.4.1. Diagnostic group-specific Xi DMPs replicated in ADNI

In view of the large amount of inflation observed in both cohorts, the focus of this section is placed on probes passing the following criteria set for replication: showing nominal significance in both cohorts, as well as the same direction of effect. Of all nominally significant probes in relation to overall diagnostic status in AddNeuroMed, 308 were found to be nominally significant in ADNI, and of the Bonferroni-significant DMPs in AddNeuroMed, 26 were nominally significant in ADNI. Specifically focusing on between-group comparisons, and therefore also the direction of effect, of all nominally significant probes in the comparison of MCI to CTL in AddNeuroMed, 57 probes were nominally significant in ADNI and showed the same direction of effect. Of the Bonferroni-significant DMPs in the same comparison in AddNeuroMed, six probes passed the replication criteria in ADNI (Table 5.17A). The most significant of these DMPs (cg17513789), located near the gene *XIST*, was hypomethylated in MCI relative to CTL in both cohorts, though the effect size in AddNeuromed (difference (CI): -4.75% (-6.42% - -3.07%), $p = 9.34\times10^{-10}$) was larger than the effect size in ADNI (difference (CI): -0.53% (-0.94% - -0.12%), $p = 6.63\times10^{-3}$).

Examining the CTL to AD comparison, 45 probes were found to be nominally significant and showing the same direction of effect in both cohorts. Of the Bonferroni-significant DMPs identified in AddNeuroMed, one DMP was nominally significant in ADNI (Table 5.17B). This DMP is annotated to the gene *ARSD* and showed hypermethylation in AD relative to CTL (AddNeuroMed: difference (CI): 1.87% (1.19% - 2.56%), $p = 3.84\times10^{-9}$, ADNI: difference (CI): 0.94% (0.11% - 1.77%), $p = 2.20\times10^{-2}$).

In the comparison of MCI to AD, 15 probes showed the same direction of effect and were nominally significant in both cohorts. Four Bonferroni-significant DMPs in the AddNeuroMed cohort were also found to be significant in ADNI, though all but one of these sites displayed opposite directions of effect in the two cohorts (Table 5.17C). One of the sites displaying opposite methylation patterns (cg09513996) is the DMP near the *ERAS* gene, which was epigenome-wide significant in both cohorts, as discussed above. The DMP which was fully replicated in ADNI (cg09167861) is annotated to *FHL1*, and showed hypomethylation in AD relative to MCI (AddNeuroMed: difference (CI): -2.1% (-3.07% - -1.12%), $p = 2.79\times10^{-6}$, ADNI: difference (CI): -2.64% (-5.27% - -0.01%), $p = 4.87\times10^{-2}$).

A. CTL vs MCI

ProbeID	Position	AddNeuroMed		ADNI		UCSC Gene	UCSC Gene Group	GREAT Annotation
		Difference (CI)	*p* CvM	Difference (CI)	*p* CvM			
cg17513789	chrX:73073251	-4.75 (-6.42 - -3.07)	9.34E-10	-0.53 (-0.94 - -0.12)	6.63E-03	*XIST*	TSS1500	*ZCCHC13* (-450773), *CHIC1* (290268)
cg25433595	chrX:100183764	4.52 (2.62 - 6.43)	2.42E-07	2.7 (0.86 - 4.55)	1.96E-03	*XKRX*	5'UTR,1stExon	*XKRX* (133)
cg11179997	chrX:119125670	2.7 (1.54 - 3.87)	4.48E-07	1.2 (0.04 - 2.37)	4.14E-02			*NKAP* (-47936), *RHOXF2* (86036)
cg25396787	chrX:3733991	2.88 (1.63 - 4.13)	5.42E-07	1.34 (0.1 - 2.58)	3.06E-02			*PRKX* (-102317)
cg03035653	chrX:99662932	4.01 (2.16 - 5.86)	2.44E-06	2.44 (0.4 - 4.48)	1.44E-02	*PCDH19*	1stExon	*PCDH19* (2338)
cg03449040	chrX:10094363	-1.39 (-2.04 - -0.74)	3.63E-06	-0.94 (-1.41 - -0.47)	1.43E-05	*WWC3*	Body	*CLCN4* (-30621), *SHROOM2* (339868)

B. CTL vs AD

ProbeID	Position	AddNeuroMed		ADNI		UCSC Gene	UCSC Gene Group	GREAT Annotation
		Difference (CI)	*p* CvA	Difference (CI)	*p* CvA			
cg04710661	chrX:2848381	1.87 (1.19 - 2.56)	3.84E-09	0.94 (0.11 - 1.77)	2.20E-02	*ARSD*	TSS1500	*ARSD* (-966)

C. MCI vs AD

ProbeID	Position	AddNeuroMed		ADNI		UCSC Gene	UCSC Gene Group	GREAT Annotation
		Difference (CI)	*p* MvA	Difference (CI)	*p* MvA			
cg09167861	chrX:35251741	-2.1 (-3.07 - -1.12)	2.79E-06	-2.64 (-5.27 - -0.01)	4.87E-02	*FHL1*	Body,1stExon, 5'UTR,TSS200	*FHL1* (22183), *MAP7D3* (81996)

Table 5.17 Overlap of Xi DMPs in the AddNeuroMed and ADNI cohorts.

Shown are the DMPs that were significant after Bonferroni correction ($p < 5.18\times10^{-6}$) in the AddNeuroMed cohort, and were found to be at least nominally significant ($p < 0.05$) in the ADNI cohort for diagnosis-associated differences of imputed Xi DNA methylation. Additionally, only probes which displayed the same direction of effect in both cohorts are shown. For each ProbeID, the genomic location is indicated (genome build 37), along with the group difference in % methylation (MCI – CTL, and AD – CTL), confidence interval (CI), and p-value as calculated with Tukey's HSD test for the AddNeuroMed and ADNI cohorts, the UCSC gene name and group annotation, and the GREAT annotation with the distance to the transcription start site indicated in parentheses.

5.4.5. Regions on the Xi chromosome associated with diagnosis in AddNeuroMed

The imputed DNA methylation profiles of the Xi chromosome were analysed to identify regions of altered methylation in the diagnosis of MCI and AD. In association with diagnostic status across all three groups (the ANOVA comparison of CTL, MCI, and AD), a total of 119 DMRs were identified. The 20 most significant DMRs are shown in Table 5.18, the full list of DMRs can be found in Appendix H. The most significant DMR was annotated to the *SRPK3* gene and spanned nine probes (chrX: 153046175 – 153046895, Šidák-*p* = 4.37×10^{-43}). The same region was also identified as a DMR in two post-hoc group comparisons, with an average relative hypomethylation of -5.36% and -5.15% observed in AD compared to CTL and MCI, respectively (Figure 5.9). In both post-hoc comparisons, the *SRPK3* region was the most significant DMR identified out of 20 DMRs identified in AD versus CTL (Table 5.19), and 10 DMRs identified in AD versus MCI (Table 5.20).

The comparison of CTL to MCI led to the identification of 22 DMRs, of which the most significant DMR is annotated to the genes *TAF1* and *INGX* (chrX: 70712215 – 70713214, Šidák-*p* = 2.9510^{-17}). This nine bp region is hypomethylated in MCI relative to CTL (-2.61% methylation difference). An overview of all DMRs associated with MCI relative to CTL is presented in Table 5.21.

ANOVA Xi DMRs							
Gene	**Position**	***n***	***p*-value**	**Šidák-*p***	**Average Methylation %**		
					CTL	**MCI**	**AD**
SRPK3	chrX:153046175 - 153046895	9	3.27E-44	4.37E-43	1.46	1.25	-3.9
TAF1;INGX	chrX:70712215 - 70713214	9	1.39E-18	1.34E-17	1.29	-1.32	0.57
CLCN4	chrX:10126321 - 10126882	5	8.3E-15	1.43E-13	-0.73	0.56	-0.07
LOC389906	chrX:3732500 - 3734340	4	4.95E-14	2.6E-13	-1.27	0.77	0.37
PCDH19	chrX:99661860 - 99663331	6	2.82E-11	1.85E-10	-1.22	-0.03	1.41
PNCK	chrX:152938698 - 152940275	11	1.03E-10	6.3E-10	0.2	-0.12	-0.15
FHL1	chrX:135228207 - 135228445	3	3.06E-11	1.24E-09	1.95	-1.6	0.03
FGF13	chrX:137950565 - 137951119	3	1.04E-10	1.81E-09	1.59	-1.2	-0.11
L1CAM	chrX:153141451 - 153141906	5	1.46E-10	3.09E-09	0.61	0.05	-0.77
OFD1;TRAPPC2	chrX:13751423 - 13753382	9	1.3E-09	6.41E-09	-0.27	0.12	0.12
MAGIX	chrX:49019804 - 49021149	4	1.92E-09	1.38E-08	-0.13	-0.06	0.18
DUSP21	chrX:44703819 - 44703984	2	2.8E-10	1.64E-08	0.97	-0.04	-1.02
ASB11	chrX:15333348 - 15334256	5	2.17E-09	2.31E-08	1.05	-0.73	-0.22
XIST	chrX:73073251 - 73073712	3	3.62E-09	7.58E-08	1.28	-0.76	-0.41
PRAF2	chrX:48929808 - 48930361	3	7.66E-09	1.34E-07	-0.19	1.73	-2.45
ZRSR2	chrX:15807140 - 15808693	9	3.29E-08	2.04E-07	0.03	0.04	-0.16
MAGEE2	chrX:75004943 - 75005361	8	9.36E-09	2.16E-07	1.82	-0.57	-1.06
FAM156A;FAM156B	chrX:52950051 - 52950333	2	6.8E-09	2.33E-07	-0.21	-0.14	0.38
TEX13B	chrX:107226253 - 107226612	2	9.73E-09	2.61E-07	0.9	-0.63	-0.03
TAF7L	chrX:100547965 - 100548224	5	7.3E-09	2.72E-07	0.31	-0.35	0.17

Table 5.18 Top 20 DMRs on the Xi chromosome associated with diagnostic status in AddNeuroMed.

Shown are DMRs for the overall three group (ANOVA) comparison of CTL to MCI to AD. Displayed for each region is the UCSC gene name, chromosomal position (genome build 37), number of probes in region (n), p-value and multiple testing-corrected p (Šidák-p), and average relative methylation values per group.

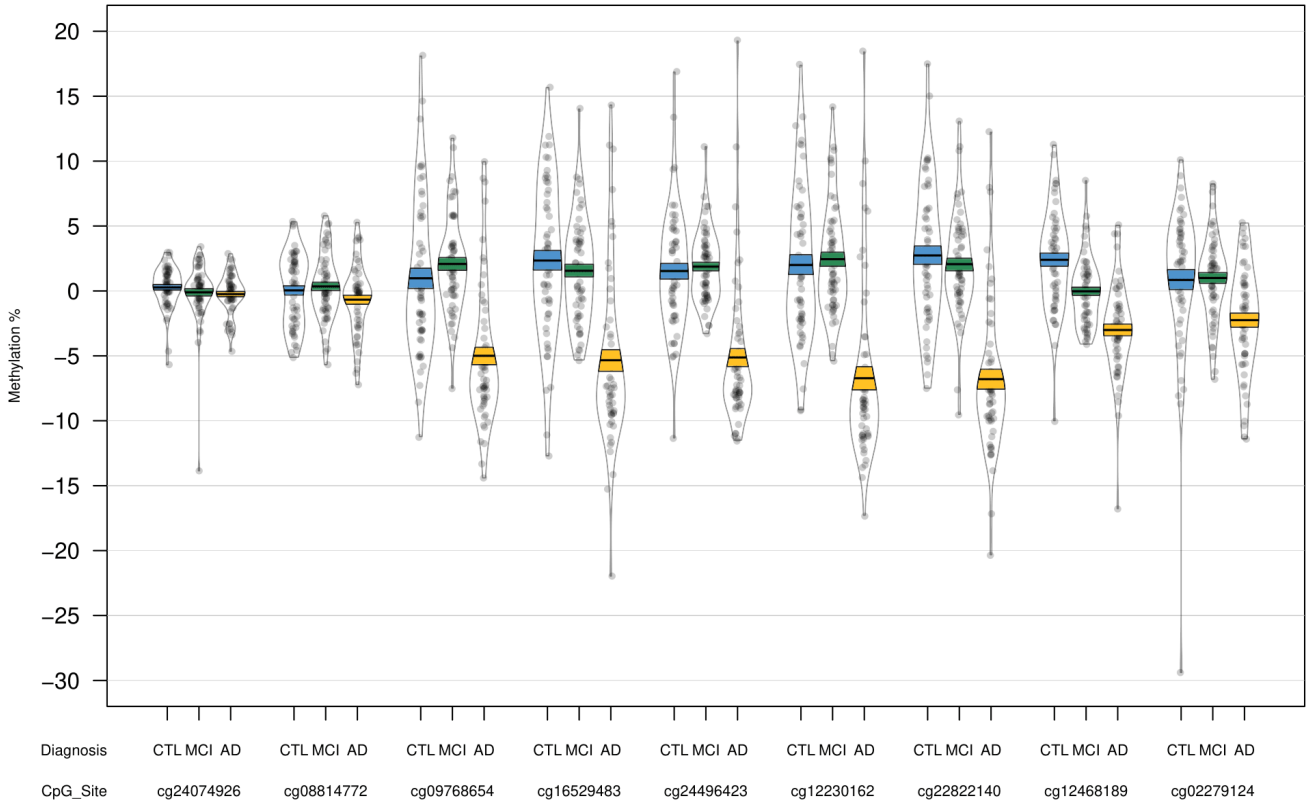

Figure 5.9 The SRPK3 DMR on the Xi chromosome showed hypomethylation in AD in AddNeuroMed.

The DMR located in SRPK3 was identified in a comparison of baseline diagnosis of MCI, AD, and CTL, as well as post-hoc comparisons of AD to CTL and MCI. Displayed are the relative Xi methylation levels of individual probes located within the DMR, ordered by genomic location. Methylation values have been corrected for the covariates of age, cell type proportion and batch.

CTL vs. AD Xi DMRs							
Gene	**Position**	***n***	***p*-value**	**Šidák-*p***	**Average Methylation %**		
					CTL	**MCI**	**AD**
SRPK3	chrX:153046175 - 153046895	9	2.37E-35	3.18E-34	1.46	1.25	-3.9
L1CAM	chrX:153141451 - 153141906	5	1.30E-09	2.75E-08	0.61	0.05	-0.77
DUSP21	chrX:44703819 - 44703984	2	5.39E-09	3.15E-07	0.97	-0.04	-1.02
FAM156A;FAM156B	chrX:52950051 - 52950333	2	8.40E-08	2.87E-06	-0.21	-0.14	0.38
MAGEE2	chrX:75004943 - 75005361	8	2.09E-07	4.83E-06	1.82	-0.57	-1.06
BCOR	chrX:40004432 - 40004482	2	7.36E-07	1.42E-04	1.11	0.72	-2.15
ZNF75D	chrX:134429653 - 134429973	4	5.59E-06	1.68E-04	2.05	-0.78	-0.97
PCDH19	chrX:99661860 - 99661877	2	3.50E-07	1.98E-04	-1.32	-0.98	2.96
CAPN6	chrX:110513791 - 110514232	5	1.47E-05	3.22E-04	-1.83	0.77	0.95
XIST	chrX:73073251 - 73073712	3	1.62E-05	3.39E-04	1.28	-0.76	-0.41
PCDH19	chrX:99194929 - 99195116	2	1.30E-05	6.70E-04	-2.82	1.1	1.86
NXF3	chrX:102347925 - 102348040	3	8.58E-06	7.20E-04	1.29	-0.29	-1.01
DANT1;DANT2	chrX:114957471 - 114957791	2	3.09E-05	9.30E-04	-1.37	0.65	0.57
ARSD	chrX:2848021 - 2848382	2	3.85E-05	1.03E-03	-0.44	-0.16	0.69
TKTL1	chrX:153533256 - 153533552	3	3.48E-05	1.13E-03	1.21	-0.15	-1.12
CDX4	chrX:72668487 - 72668848	3	5.31E-05	1.42E-03	-1.09	0.49	0.41
SMC1A;RIBC1	chrX:53449423 - 53449901	10	7.70E-05	1.55E-03	-1.07	0.22	0.82
PTCHD1-AS	chrX:23016709 - 23017134	2	1.13E-04	2.55E-03	0.79	-0.05	-0.81
MAP3K15	chrX:19443172 - 19443223	2	5.89E-05	1.11E-02	-0.81	-0.6	1.64
JPX	chrX:73164228 - 73164391	2	2.63E-04	1.55E-02	-0.63	0.25	0.38

Table 5.19 DMRs on the Xi chromosome associated with diagnostic status of AD in AddNeuroMed.

Shown are DMRs for the post-hoc AD to CTL comparison. Displayed for each region is the UCSC gene name, chromosomal position (genome build 37), number of probes in region (n), p-value and multiple testing-corrected p (Šidák-p), and average relative methylation values per group.

MCI vs. AD Xi DMRs							
Gene	**Position**	***n***	***p*-value**	**Šidák-*p***	**Average Methylation %**		
					CTL	**MCI**	**AD**
SRPK3	chrX:153046175 - 153046895	9	3.37E-36	4.52E-35	1.46	1.25	-3.9
CLCN4	chrX:10126668 - 10126882	3	7.65E-13	3.45E-11	-1.49	1.73	-0.99
PRAF2	chrX:48929808 - 48930361	3	7.08E-10	1.24E-08	-0.19	1.73	-2.45
TAF1	chrX:70712403 - 70712811	7	1.42E-08	3.36E-07	1.22	-1.41	0.72
ARHGAP6	chrX:11157142 - 11157611	5	1.77E-07	3.65E-06	-0.29	1.8	-1.91
POU3F4	chrX:82764385 - 82764538	2	1.15E-07	7.22E-06	0.16	1.45	-2.23
RHOXF1P1	chrX:119125659 - 119125723	3	5.63E-07	8.49E-05	-0.2	0.98	-1.2
FAM50A	chrX:153674100 - 153674275	2	1.02E-05	5.60E-04	-0.02	0.47	-0.61
BCOR	chrX:40004432 - 40004482	2	1.12E-05	2.16E-03	1.11	0.72	-2.15
PCDH19	chrX:99661860 - 99661877	2	8.40E-06	4.76E-03	-1.32	-0.98	2.96

Table 5.20 DMRs on the Xi chromosome associated with diagnostic status of AD compared to MCI in AddNeuroMed.

Shown are DMRs for the post-hoc AD to MCI comparison. Displayed for each region is the UCSC gene name, chromosomal position (genome build 37), number of probes in region (n), p-value and multiple testing-corrected p (Šidák-p), and average relative methylation values per group.

CTL vs. MCI Xi DMRs							
Gene	**Position**	***n***	***p*-value**	**Šidák-*p***	**Average Methylation %**		
					CTL	**MCI**	**AD**
TAF1;INGX	chrX:70712215 - 70713214	9	3.05E-18	2.95E-17	1.29	-1.32	0.57
LOC389906	chrX:3732500 - 3734340	4	6.95E-13	3.64E-12	-1.27	0.77	0.37
FHL1	chrX:135228207 - 135228445	3	6.02E-12	2.44E-10	1.95	-1.6	0.03
FGF13	chrX:137950565 - 137951119	3	4.89E-11	8.51E-10	1.59	-1.2	-0.11
CT45A1	chrX:134846527 - 134847107	3	1.75E-08	2.90E-07	0.98	-1.01	0.21
MAGEA8-AS1	chrX:149009554 - 149009914	4	3.95E-07	1.06E-05	-1.08	0.89	-0.29
NHS	chrX:17395816 - 17395857	2	8.14E-08	1.91E-05	0.27	-0.49	0.33
TEX13B	chrX:107226253 - 107226612	2	3.28E-06	8.81E-05	0.9	-0.63	-0.03
DANT1;DANT2	chrX:114957471 - 114957791	2	4.76E-06	1.43E-04	-1.37	0.65	0.57
TCEAL7	chrX:102584953 - 102585355	6	8.91E-06	2.14E-04	-1.82	1.26	0.06
ZNF75D	chrX:134429653 - 134429973	4	9.08E-06	2.74E-04	2.05	-0.78	-0.97

CTL vs. MCI Xi DMRs (Continued)							
Gene	**Position**	***n***	***p*-value**	**Šidák-*p***	**Average Methylation %**		
					CTL	**MCI**	**AD**
TAF7L	chrX:100547965 - 100548224	5	8.13E-06	3.03E-04	0.31	-0.35	0.17
XIST	chrX:73073251 - 73073712	3	1.71E-05	3.57E-04	1.28	-0.76	-0.41
CLCN4	chrX:10126668 - 10126882	3	9.63E-06	4.34E-04	-1.49	1.73	-0.99
DENND10P1	chrX:129657974 - 129658062	2	4.06E-06	4.45E-04	-1.73	1.73	-0.65
CAPN6	chrX:110513791 - 110514232	5	2.51E-05	5.50E-04	-1.83	0.77	0.95
ASB11	chrX:15333584 - 15334256	4	3.90E-05	5.60E-04	0.95	-0.83	0.03
SOX3	chrX:139592127 - 139592513	3	2.45E-05	6.12E-04	0.12	0.58	-0.84
CA5B;CA5BP1-CA5B	chrX:15807140 - 15807819	2	8.10E-05	1.15E-03	1.5	-0.66	-0.93
PNCK	chrX:152939615 - 152940158	3	1.22E-04	2.17E-03	-0.07	0.71	-0.95
MAGEE2	chrX:75004943 - 75005361	8	1.41E-04	3.24E-03	1.82	-0.57	-1.06
GPR50-AS1	chrX:150343097 - 150343149	2	3.39E-05	6.26E-03	-1.97	0.99	0.74

Table 5.21 DMRs on the Xi chromosome associated with diagnostic status of MCI in AddNeuroMed.

Shown are DMRs for the post-hoc MCI to CTL comparison. Displayed for each region is the UCSC gene name, chromosomal position (genome build 37), number of probes in region (n), p-value and multiple testing-corrected p (Šidák-p), and average relative methylation values per group.

5.4.6. Replication of Xi DMRs in ADNI

As carried out for all analysis performed on the AddNeuroMed data in this chapter, the identification of DMRs associated with Xi DNA methylation was repeated in the ADNI cohort to examine the replication of results. In the ADNI dataset, 26 DMRs were associated with the overall three-group comparison. Five of these DMRs were annotated to genes that were found to contain diagnosis-associated DMRs in the AddNeuroMed cohort (Table 5.22). These DMRs were located in the genes *CLCN4, XIST, ERAS, GPC3, and AFF2,* although the AddNeuroMed and ADNI DMRs did not overlap. One further DMR of six bp in the ADNI cohort (chrX: 75005028 – 75005361, Šidák-*p* = 6.97×10^{-4}), was spanned by an eight bp DMR identified in AddNeuroMed (chrX: 75004943 – 75005361, Šidák-*p* = 4.83×10^{-6}). This DMR is located in the gene *MAGEE2,* and is hypomethylated in MCI (-2.39%) and AD (-2.88%) relative to CTL in AddNeuroMed (Figure 5.10). However, in ADNI the region is hypermethylated in MCI and AD compared to CTL (2.29% and 3.29%, respectively; Figure 5.11). Within the *ERAS* gene, separate regions were identified as DMRs. In AddNeuroMed, a three bp DMR was associated with overall group differences in AddNeuroMed, which was mainly driven by hypomethylation in the MCI group relative to CTL (-2.93%, Figure 5.12). In ADNI, three DMRs of varying lengths (between five to seven bp) were identified that had the same endpoint, though they did not overlap the AddNeuroMed DMR. The two *ERAS*-annotated DMRs identified in the ANOVA analyses in both cohorts were separated by a 639 bp distance. In ADNI, the group difference also appeared to be mainly driven by hypomethylation in MCI, though particularly in comparison to AD (4.8%, Figure 5.13).

In the post-hoc group comparisons, no DMRs were identified in the comparison of CTL to MCI in ADNI, though three DMRs were associated with differences in AD relative to CTL, and two DMRs were related to MCI versus AD differences. Other than the *ERAS* DMR in both comparisons, a DMR in the gene *CXorf51B* was associated with diagnosis of AD relative to CTL (in the ADNI cohort only), and a DMR in *TAF1* was associated with AD in comparison to the CTL group as well as the MCI group. Interestingly, the *TAF1* DMR exactly matched an MCI versus AD DMR identified in AddNeuroMed. The seven bp DMR was characterised by hypermethylation in AD relative to MCI in AddNeuroMed (2.13%, Figure 5.14), whereas the ADNI data showed hypomethylation in AD relative to CTL (4.09%) and MCI (4.84%, Figure 5.15).

	ADNI							AddNeuroMed			
Gene	**Position**	***n***	***p*-value**	**Šidák-*p***	**Average Methylation %**			**Position**	**Average Methylation %**		
					CTL	**MCI**	**AD**		**CTL**	**MCI**	**AD**
CLCN4	chrX:10134966 - 10135224	2	6.19E-05	4.24E-03	-1.84	0.38	1.89	chrX:10126321 - 10126882	-0.73	0.56	-0.07
XIST	chrX:73051811 - 73051896	2	5.38E-05	0.01	0.31	-0.21	0.08	chrX:73073251 - 73073712	1.28	-0.76	-0.41
MAGEE2	chrX:75005028 - 75005361	6	1.31E-05	6.97E-04	-1.75	0.54	1.54	chrX:75004943 - 75005361	1.82	-0.57	-1.06
ERAS	chrX:48684558 - 48685304	7	2.17E-18	5.13E-17	-0.03	-1.44	3.36	chrX:48685943 - 48686201	1.95	-0.98	-0.28
GPC3	chrX:133119645 - 133119964	8	8.18E-06	4.54E-04	-0.73	0.1	0.99	chrX:133118088 - 133118345	-1.61	0.56	0.92
AFF2	chrX:147581919 - 147582072	6	9.68E-06	1.12E-03	-1.13	0.41	0.87	chrX:147582939 - 147583280	-0.02	-0.1	0.22

Table 5.22 Overlapping genes containing Xi DMRs in the ADNI and AddNeuroMed overall group comparisons.

Displayed are the genes in which DMRs were identified in the ANOVA comparison in the AddNeuroMed cohort as well as the ADNI cohort. Displayed for each region is the UCSC gene name, chromosomal position (genome build 37), number of probes in region (n), p-value and multiple testing-corrected p (Šidák-p), and average relative methylation values per group. Chromosomal position and average methylation values in the DMR within the same gene in AddNeuroMed have been added for comparison.

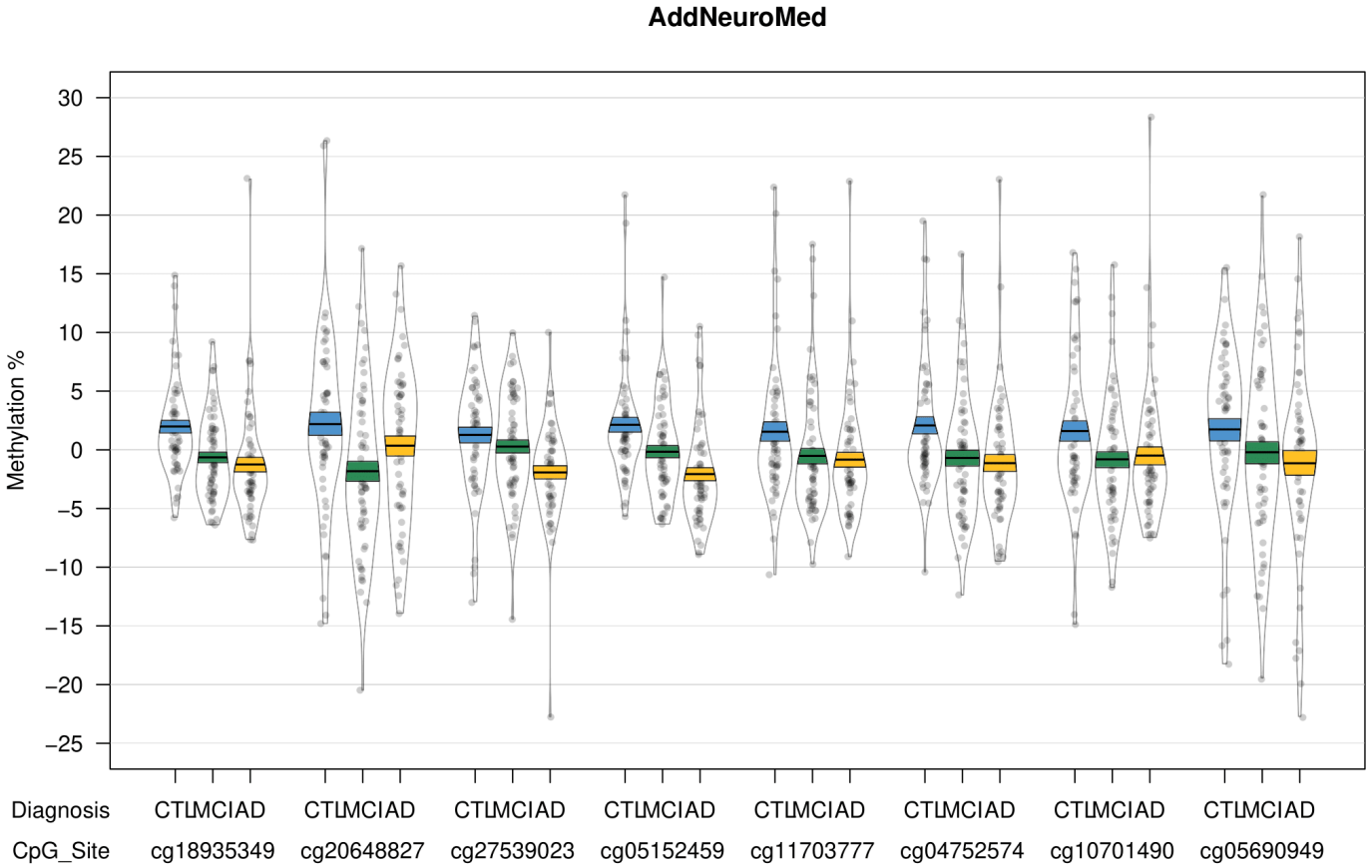

Figure 5.10 The MAGEE2 DMR on the Xi chromosome showed hypomethylation in AD and MCI in AddNeuroMed.

The DMR located in MAGEE2 was identified in a comparison of baseline diagnosis of MCI, AD, and CTL, and post-hoc comparisons of MCI and AD to CTL. Displayed are the relative Xi methylation levels of individual probes located within the DMR, ordered by genomic location. Methylation values have been corrected for the covariates of age, cell type proportion and batch.

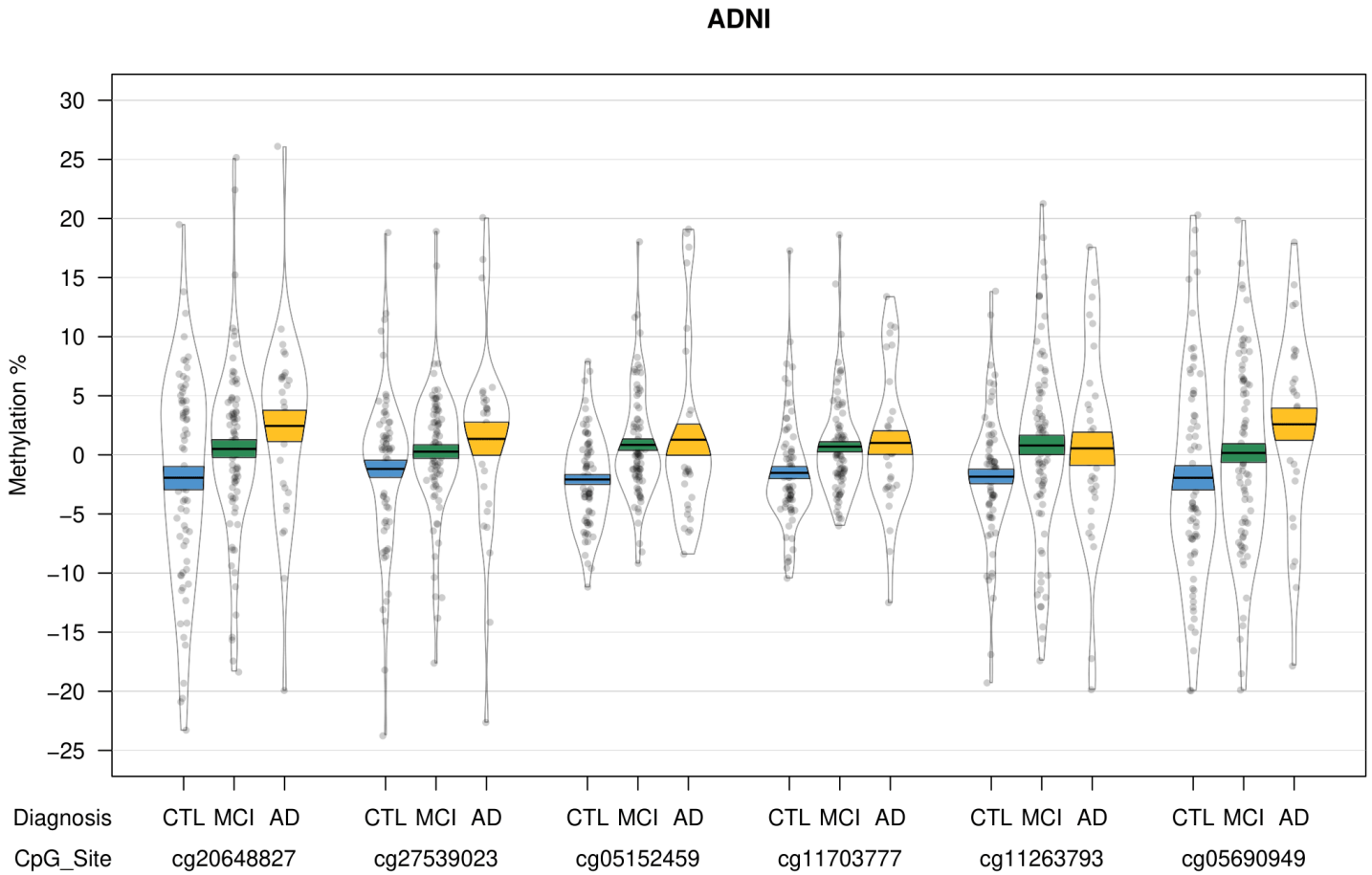

Figure 5.11 The MAGEE2 DMR on the Xi chromosome showed hypermethylation in AD and MCI in ADNI.

The DMR located in MAGEE2 was identified in a comparison of baseline diagnosis of MCI, AD, and CTL. Displayed are the relative Xi methylation levels of individual probes located within the DMR, ordered by genomic location. Methylation values have been corrected for the covariates of age and cell type proportion.

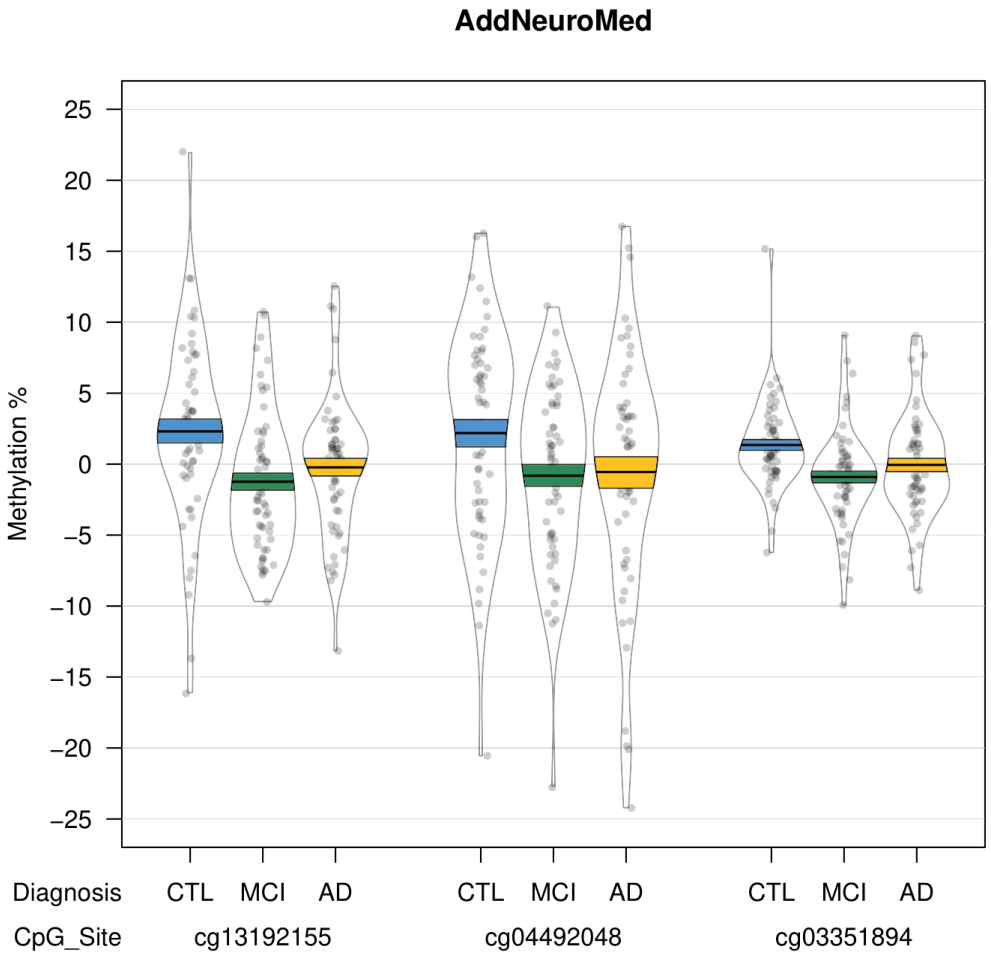

Figure 5.12 The ERAS DMR on the Xi chromosome showed hypomethylation in MCI relative to CTL in AddNeuroMed.

The DMR located in ERAS was identified in a comparison of baseline diagnosis of MCI, AD, and CTL. Displayed are the relative Xi methylation levels of individual probes located within the DMR, ordered by genomic location. Methylation values have been corrected for the covariates of age, cell type proportion and batch.

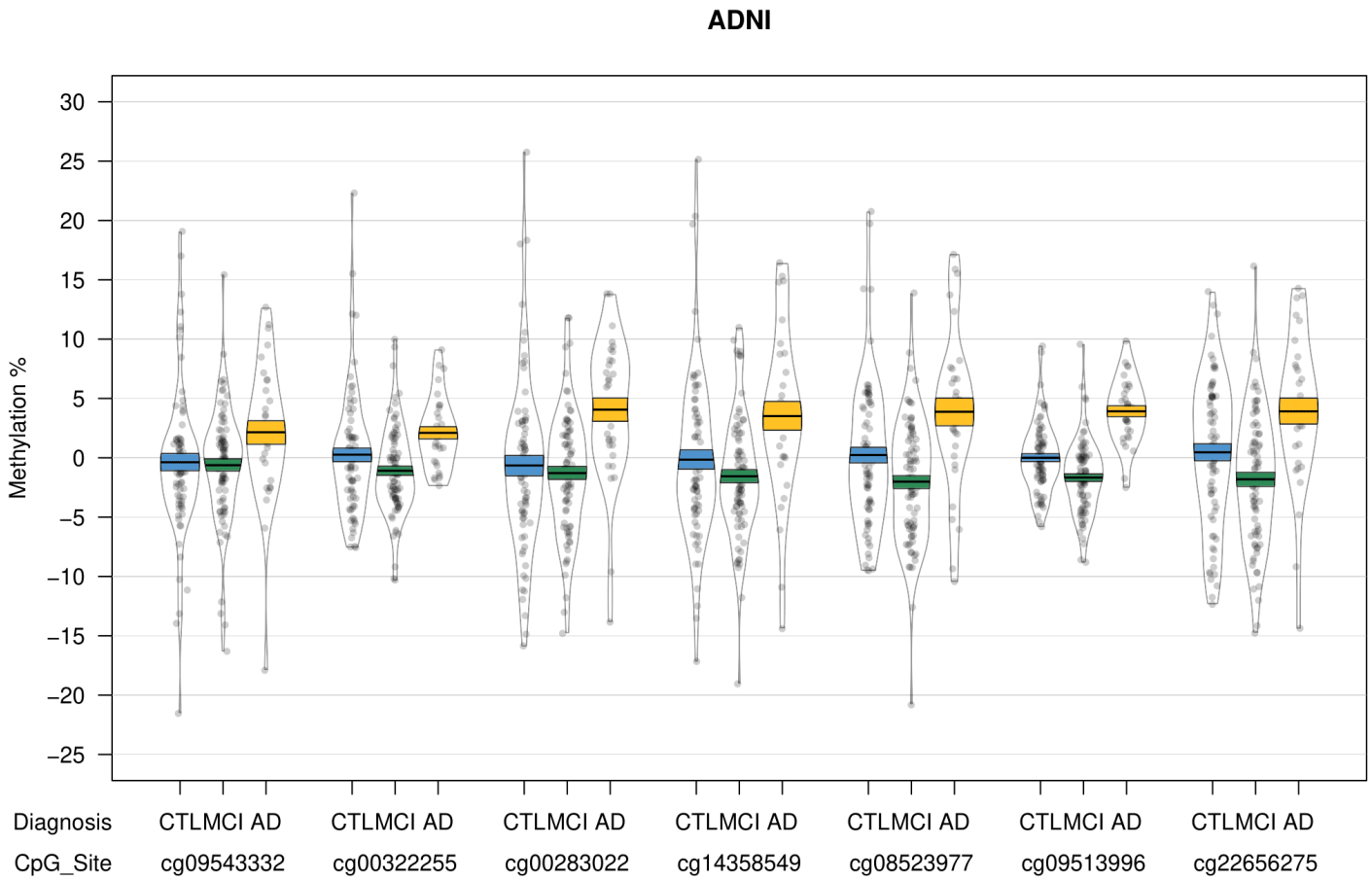

Figure 5.13 The ERAS DMR on the Xi chromosome showed hypomethylation in MCI and CTL relative to AD in ADNI.

The DMR located in ERAS was identified in a comparison of baseline diagnosis of MCI, AD, and CTL, and post-hoc comparisons of AD to CTL and MCI. Displayed are the relative Xi methylation levels of individual probes located within the DMR, ordered by genomic location. Methylation values have been corrected for the covariates of age and cell type proportion.

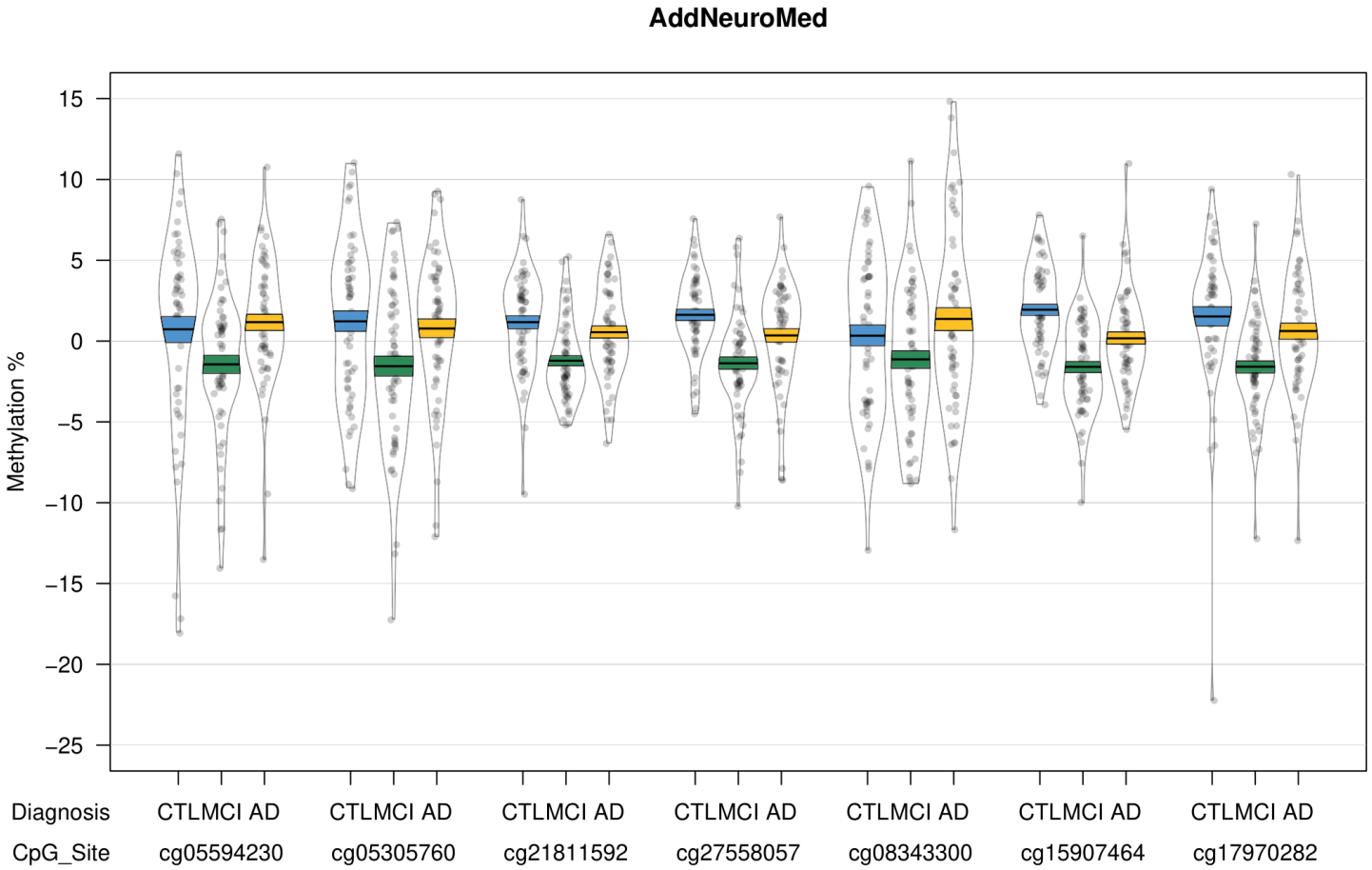

Figure 5.14 The TAF1 DMR on the Xi chromosome showed hypermethylation in AD relative to MCI in AddNeuroMed.

The DMR located in TAF1 was identified in a post-hoc comparison of MCI to AD. Displayed are the relative Xi methylation levels of individual probes located within the DMR, ordered by genomic location. Methylation values have been corrected for the covariates of age, cell type proportion and batch.

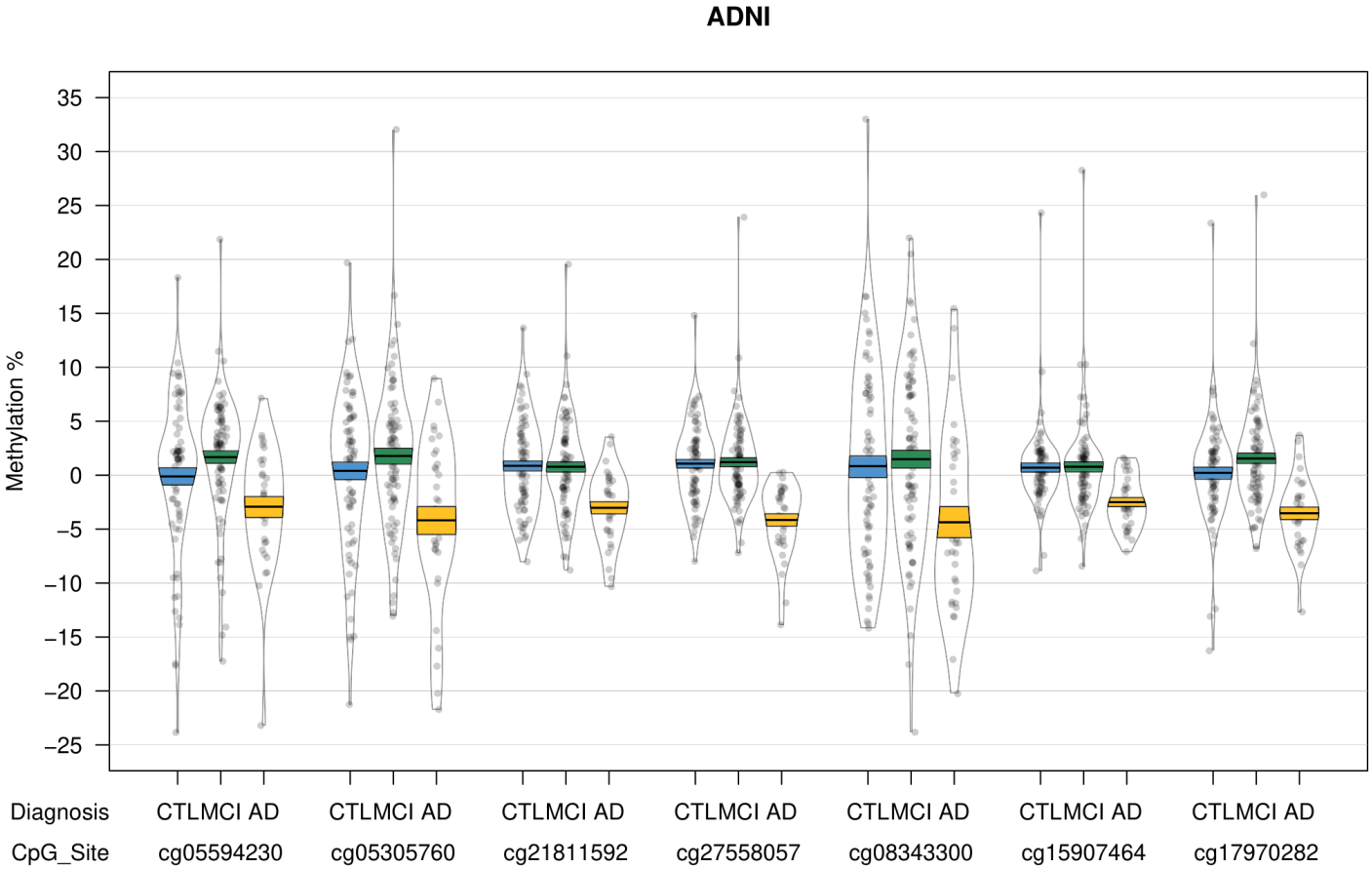

Figure 5.15 The TAF1 DMR on the Xi chromosome showed hypomethylation in AD relative to CTL and MCI in ADNI.

The DMR located in TAF1 was identified in a comparison of baseline diagnosis of MCI, AD, and CTL, and post-hoc comparisons of AD to CTL and MCI. Displayed are the relative Xi methylation levels of individual probes located within the DMR, ordered by genomic location. Methylation values have been corrected for the covariates of age and cell type proportion.

5.5. Discussion

In this chapter, DNA methylation changes associated with diagnosis on the X, Y, and Xi chromosomes were assessed. This was achieved through a stratified analysis of males and females to profile DNA methylation on the X and Y chromosomes, and though an analysis of imputed Xi methylation values in females.

In the analysis of X and Y chromosome methylation, no epigenome-wide significant DMPs or DMRs were identified in the AddNeuroMed cohort. However, several nominally significant DMPs were replicated in the ADNI cohort.

In males, two replicated DMPs on the X chromosome were found to be associated with MCI status. The first, annotated to *LOC100132963,* is a relatively unknown locus, whereas the second DMP identified is annotated to *DUSP9,* of which epigenetic silencing has been associated with progression of various types of cancer (Qiu et al., 2020; F. Wu et al., 2015). Though *DUSP9* has not been linked to MCI or AD yet, it is also a candidate gene for type 2 diabetes mellitus, where sex differences in DNA methylation of this probe in pancreatic islets have been reported previously (Hall et al., 2014). Of note, type 2 diabetes has been identified as a risk factor for AD, and studies have examined common pathological processes between the two diseases (L. Li & Hölscher, 2007).

A total of 14 DMPs on the X chromosome were identified and replicated across the different group comparisons in females, of which seven were associated with DNA methylation differences related to MCI status relative to CTL. Loci in

HUWE1 (cg18450499)*, WWC3* (cg03449040), *IQSEC2* (cg03832506*)* were hypomethylated in MCI, while loci in *SMARCA1* (cg17420696), *HTR2C* (cg02918903)*, SMPX* (cg27013947)*,* and *PRKX* (cg25396787) were hypermethylated. Of particular interest is the DMP in *WWC3,* as gene expression of *WWC3* has been found to be upregulated in the hippocampus of AD patients (Yaping Wang & Wang, 2020). Most importantly, the study by Wang & Wang (2020) found that differential expression of *WWC3* in the blood of AD patients could be used to distinguish AD patients from healthy CTLs, highlighting the biomarker potential of this gene. Although the connection between *WWC3* and AD requires further study, *WWC3* has been found to act on the Wnt/β-Catenin signalling pathway (Yanni Wang et al., 2017) which has been hypothesised to be a causal mechanism in the development and progression of AD (Ferrari et al., 2014).

Associated with diagnosis of AD relative to CTL were five DMPs located in or near *NLGN3* (cg15737490), *TSPYL2* (cg23612178), *ARMCX3* (cg05453458), *PHKA2* (cg06104510), and *UTP14A* (cg05548952), all of which were hypermethylated in AD. Though none of these genes have been linked to AD before, gene expression of *ARMCX3* has been found to display sex differences in early stages of human development (Gonzalez et al., 2018), as does *UTP14* expression in cancers associated with TP53 mutations (Lopes-Ramos et al., 2020).

Finally, in the comparison of MCI to AD in females, two DMPs were replicated, located in *GPR50* (cg10841338) and *ERAS* (cg07112779). The gene *GPR50* is part of the family of G protein-coupled receptors (GPCRs), which have been

implicated in the pathogenesis of AD (Thathiah & De Strooper, 2011). The gene *GPR50* itself is also a sex-specific genetic risk factor for bipolar affective disorder (Thomson et al., 2005). Conversely, a study conducted by Chen *et al.* (2019) found hypomethylation of the *GPR50* promoter in blood of males diagnosed with AD relative to CTL. It should be noted, however, that this study did not correct for the influence of cell types.

The *ERAS* gene contained a (replicated) DMP associated with MCI versus AD differences in females, as well as a separate epigenome-wide significant DMP on the Xi chromosome, though this second DMP was not replicated in ADNI. Furthermore, two DMRs annotated to *ERAS* (separated by 639 bp) were identified in the AddNeuroMed cohort and the ADNI cohort. The *ERAS* gene stands for Embryonic Stem Cell-Expressed Ras, and is expressed in undifferentiated embryonic stem cells as well as several cancer cell lines (Yasuda et al., 2007). Though multiple studies examining its role in different types of cancer exist, this is the first time the gene has been linked to AD and sex differences.

A large number of epigenome-wide significant DMPs associated with diagnosis were identified on the Xi chromosome. Notably, a large amount of inflation was observed in the *p*-values of the statistical model, and various models aimed at correcting for inflation and unwanted variation did not deflate the *p*-values, and the same inflation was also observed in the independent ADNI replication cohort. Although technical artefacts or inflation induced by the imputation of the Xi methylation values are not ruled out, these data lend some confidence that this is not an effect of confounders. Additionally, severely skewed XCI has been found

in peripheral blood of females with AD (Bajic et al., 2015), which may be a biological mechanism contributing to the observed Xi inflation. As this is still rather speculative, the focus is placed on results that could be replicated in the ADNI cohort.

Firstly, eight DMPs were identified in the AddNeuroMed cohort that were replicated in ADNI. In a comparison of CTL to AD, a hypermethylated locus was identified near *ARSD* (cg04710661), which has not been studied yet in the context of AD. A second DMP (cg09167861), showing hypomethylation in AD relative to MCI, was annotated to the gene *FHL1*. Expression of this protein has previously been found to be increased in the hippocampus, in early Braak stages of AD (Hondius et al., 2016). The majority of the replicated DMPs were found in the comparison of MCI to CTL. Of note, one of the replicated Xi DMPs (cg03449040, annotated to *WWC3*) matched the DMP identified on the X chromosome in the stratified analysis of females, and hypomethylation in MCI relative to CTL was found in both analyses. The most significant DMP in the comparison of MCI and CTL, annotated to the gene *XIST* (cg17513789), was hypomethylated in MCI. Interestingly, the gene *XIST* was also found to contain two DMRs associated with differences across all three diagnosis groups in AddNeuroMed and ADNI, although these DMRs did not overlap. The DMR in AddNeuroMed was specifically associated with hypomethylation in the MCI and AD groups relative to CTL. While an overall difference between the diagnosis groups was identified in ADNI, the Xi DNA methylation pattern did not match the pattern seen in AddNeuroMed. Interestingly, the gene *XIST* (X-inactive Specific Transcript) is an indispensable initiator of the XCI process. Its product, the long non-coding RNA Xist, attaches to an X chromosome which triggers the process

of transcriptional silencing (Strehle & Guttman, 2020). Altered gene expression of *XIST* and potential dysregulation of XCI have been suggested to play a role in AD, though its exact involvement remains to be studied (Chanda & Mukhopadhyay, 2020).

Although this chapter has identified novel loci and regions related to AD and MCI, several limitations must be taken into account. First, the stratification of the male and female data for the first goal of identifying disease-associated loci and regions on the X and Y chromosome leads to a limited sample size. In order to detect effect sizes of the magnitude seen in the overall diagnosis-associated DMPs identified in Chapter 3 with a power of 0.8, a sample size of 161 would be required per group. Second, although all processed data was adjusted for cell type proportions, recent studies have highlighted the importance of studying sorted or single-cell populations in the context of AD (L. Guo et al., 2021). An imbalance of cell populations associated with AD pathology may exist between the sexes in blood. This has been detected in female brain cells, where AD pathology-associated populations were enriched compared to male brain cells, and where cell type-specific gene expression was different between males and females (Mathys et al., 2019). Third, this study makes use of imputed Xi chromosome data, which may introduce technical artefacts in the data. This method relies on the assumption that, within a specific diagnostic group, the average Xa methylation pattern in males corresponds to the average Xa methylation pattern in females. Ideally, Xi DNA methylation would be assessed directly, using methods such as allele specific DNA methylation sequencing (Akbari et al., 2021). Additionally, skewed XCI reported in AD (Bajic et al., 2015) could lead to the overrepresentation of one chromosome in any type of Xi

analysis, though the effect this may have on the results is unknown. Of note, while multiple DMRs were identified on the Xi chromosome, no DMRs were identified in the stratified analysis in either of the cohorts. While the results on the Xi may be related to the nature of XCI (i.e. effective gene silencing requiring multiple methylated CpGs; Curradi et al., 2002), the lack of DMRs identified in the stratified analysis may be related to the coverage of probes on the X and Y chromosomes. Although the Y chromosome is relatively short (57,227,415 bp), the X chromosome spans 156,040,895 bp, which is similar to the length of chromosome 7 (159,345,973 bp; Piovesan et al., 2019). In comparison, chromosome 7 is covered by 24,547 probes, while the X chromosome is only covered by 9,647 probes. Although this is highly speculative and requires further confirmation, this may decrease the number of adjacent probes that can be detected in the 1,000 bp sliding window used to identify DMRs. Finally, the 450K array does not allow assessment of DNA methylation at the centromere. If, as suggested in previous studies, centromere (dys)regulation and PCD may play a role in AD, it would be interesting to study this mechanism in other datasets (e.g. whole genome bisulfite sequencing data) which cover DNA methylation across the whole genome.

To conclude, this chapter characterises genome-wide DNA methylation patterns that are present on the X, Y, and Xi chromosomes, and which are associated with a diagnosis of MCI or AD. As many of the loci and regions identified here have not been linked to sex differences or diagnosis of MCI or AD before, this emphasises the importance of accounting for and studying sex differences in autosomes as well as allosomes in AD research.

CHAPTER 6. DISCUSSION

6.1. Introduction

The main aim of this thesis was to characterise DNA methylation changes in blood in AD and MCI, and to investigate the effects of the risk factors sex and age on blood DNA methylation and their interaction with disease. In this chapter, I will summarise the key findings of this thesis, and discuss the strengths and limitations associated with the work in this thesis. Finally, I will outline my future perspective for this field, and discuss how future studies might overcome limitations I was unable to address in this thesis.

6.2. Key Findings from this Thesis

6.2.1. Chapter 3: An Epigenome-Wide Association Study of Alzheimer's Disease Blood Highlights Robust DNA Hypermethylation in the *HOXB6* Gene

The primary aim of Chapter 3 was to carry out an EWAS of AD and MCI blood in the AddNeuroMed cohort, in order to identify DNA methylation profiles associated with MCI, AD, and progression to AD. The second aim was to validate the most interesting AD-associated DMR using an independent method (pyrosequencing), and to examine the relationship between differential methylation and gene expression within the same samples.

The results from Chapter 3 showed that a robust DNA methylation signal can be detected in blood of AD patients within the *HOXB6* gene, a finding that was validated using pyrosequencing. This homeobox gene is involved in hematopoietic development and early body morphogenesis (Krumlauf, 1994), and is required for the normal generation of granulocytes and monocytes (Giampaolo et al., 2002). Furthermore, the most significant locus within the *HOXB6* DMR has been previously identified as a DMP in AD hippocampal tissue, and has been correlated to tau burden (Altuna et al., 2019). In addition to the *HOXB6* DMR, three other regions were associated with overall differences between CTL, MCI, and AD (annotated to the genes *MOV10L1, CBFA2T3,* and *TPTEP2-CSNK1E*), and nine regions were associated with future progression to AD (annotated to the genes *CPT1B* and *CHKB*, *SMC1B* and *RIBC2, TMEM184A, KCNAB3, GABBR1, FIGN, PRDM1, FLJ37453,* and *OR56A3* and *TRIM5*).

In addition to genomic regions associated with disease, pathways altered in association with MCI, AD-associated variables (e.g. *APOE* carrier status), and progression from MCI to AD were also identified in Chapter 3 using WGCNA. Biological pathways were found to be altered in association with MCI diagnosis (relative to CTL), carrier status of *APOE* ε4 alleles, an individual's number of education years, and the MRI measurement of MET. Interestingly, several of these altered pathways were found to be involved in immune system function, which is known to be dysregulated in AD (Heppner et al., 2015). Additionally, this study identified a cluster of co-methylated loci that was associated with future progression to AD. This cluster was involved in pathways related to non-homologous end-joining, which is involved in double-stranded DNA repair and is known to be affected in the AD brain (Kanungo, 2013; Shackelford, 2006).

6.2.2. Chapter 4: DNA Methylomic Patterns in Blood Associated with Age, Sex and Alzheimer's Disease

The aim of Chapter 4 was to characterise DNA methylation profiles associated with the AD-risk factors sex and age, and study whether and, if so, how they interact with each other and with disease.

The first analysis of this study identified DMPs related to sex, age, and the interactions of sex by age, sex by diagnosis, and age by diagnosis. Two DMPs were identified in AddNeuroMed and replicated in the ADNI cohort. The most interesting DMP was located in *C6orf25*, which is an alias for *MPIG6B*, and was associated with an interaction between sex and age. This DMP was also located within a DMR identified in AddNeuroMed, which was related to sex and sex by age differences. *MPIG6B* is located within the MHC class III region, and is part of

the immunoglobulin superfamily (de Vet et al., 2001). Though female sex hormones have been found to modulate effects of the Mpig6b protein (Stavnichuk et al., 2021), the link between sex, age, and *MPIGB6* requires further study. The second DMP that was replicated was annotated to the gene *CSRP2,* and showed effects of an age by diagnosis interaction. Interestingly, this gene has not previously been associated with MCI or AD.

This study further identified DMRs associated with the risk factors and their interactions. Of particular interest are the DMRs identified in *OXT* (associated with age), and *HOXA4* (associated with a sex by diagnosis interaction). Although these DMRs could not be replicated in the ADNI cohort, the *OXT* DMR has previously been identified in both brain and blood in the context of AD (Lardenoije et al., 2019; Watson et al., 2016), and the *HOXA4* DMR has been associated with AD neuropathology in the brain (R. G. Smith et al., 2018).

Finally, this study examined clusters of co-methylated loci through WGCNA, and identified two clusters which were both associated with age. These clusters were involved in the RNA processing and homophilic cell adhesion.

6.2.3. Chapter 5: DNA Methylation on the X And Y Chromosome in Alzheimer's Disease and Mild Cognitive Impairment

The primary aim of Chapter 5 was to characterise AD- and MCI-associated DNA methylation profiles on the X chromosome in females and the X and Y chromosomes in males. Through the imputation of Xi methylation, I was able to fulfil the secondary aim of profiling DNA methylation on the female Xi chromosome in order to investigate patterns associated with disease.

The first analysis conducted in this study, in line with the primary aim, assessed DMPs related to AD and MCI on the sex chromosomes in males and females separately. As the sample size in the AddNeuroMed discovery cohort was too small to identify DMPs, a replication analysis was performed in the ADNI cohort. This led to the identification of two MCI-associated DMPs on the X chromosome in males, located in *LOC100132963* and *DUSP9,* which had not previously been associated with MCI or AD. In females, 14 disease-associated DMPs were identified, of which seven were associated with MCI (in the genes *HUWE1*, *WWC3, IQSEC2, SMARCA1, HTR2C, SMPX,* and *PRKX*), five were associated with AD (in the genes *NLGN3, TSPYL2, ARMCX3, PHKA2,* and *UTP14A*), and two DMPs were associated with differences between MCI and AD (in the genes *GPR50* and *ERAS*). Of particular interest was the DMP located in *WWC3* (cg03449040), which is involved in Wnt/β-Catenin signalling (Yanni Wang et al., 2017), and has been found to be differentially expressed in the AD hippocampus (Yaping Wang & Wang, 2020). Most importantly, gene expression of *WWC3* in blood has been shown to be of use as a diagnostic biomarker for AD (Yaping Wang & Wang, 2020). It would be interesting to further explore the biomarker potential of the *WWC3* gene, and in particular of the DMP identified in this study, as DNA methylation has been found to be more stable than RNA (Gosselt et al., 2021; Huang et al., 2017) and may therefore be a more robust target. In addition to studying the biomarker potential of the *WWC3* DMP, future studies would need to investigate the temporal pattern and sex-specificity of *WWC3* DNA methylation and expression, as hypomethylation in females has been associated with MCI in this study, yet increased expression was associated with AD across both sexes in the study by Yaping Wang & Wang (2020).

In relation to the second aim, the analyses carried out in Chapter 5 identified and replicated eight DMPs on the Xi chromosome. Six of these DMPs were associated with MCI relative to CTL (annotated to the genes *XIST, XKRX, NKAP PRKX, PCCDH19,* and *WWC3*), one DMP was associated with AD relative to CTL (annotated to *ARSD*), and one DMP was associated with AD relative to MCI (annotated to *FHL1*). Of note, the *WWC3* DMP identified in this analysis matched the DMP identified in the stratified analysis described above. It is therefore also important for future studies investigating this locus to take into account that the association with MCI in females may be driven by DNA methylation on the Xi chromosome. The most interesting finding in this analysis was the DMP annotated to the gene *XIST* (cg17513789), which was hypomethylated in MCI relative to CTL. Furthermore, in both the AddNeuroMed cohort and the ADNI cohort, Xi DMRs were identified in this gene. The DNA methylation patterns between the cohorts did not match, though this may be related to the ~21 Kb distance between the DMRs. *XIST* is particularly interesting in this study of the Xi chromosome, as the gene constitutes a crucial part of the XCI process and initiates transcriptional silencing (Strehle & Guttman, 2020). Altered DNA methylation of this gene may either cause or result from altered XCI, and as XCI is established in early development, it would be interesting for future studies to investigate the longitudinal pattern of DNA methylation in this gene in the development of MCI and AD.

6.3. Strengths, limitations and future perspectives

In addition to strengths, several limitations to the studies described in this thesis must be noted. In this section, I will describe the strengths and limitations that are applicable to multiple chapters of this thesis, the mitigating steps I have taken, and how these limitations could be addressed in future studies.

6.3.1. Strengths

This thesis contains the first EWAS of its size of AD, MCI, and future conversion to AD, as well as a characterisation of disease associated DNA methylation profiles on the X, Y, and Xi chromosomes. The use of ANOVA models for all analyses allowed for the exploration of non-linear relationships between the CTL, MCI, and AD groups. Though the limitation of sample size is discussed below, the AddNeuroMed cohort was of a sufficient size to detect a DMR in the *HOXB6* region in association with AD, which was also validated using pyrosequencing. The identification of this DMR provides an interesting target for future studies of diagnostic biomarkers for AD. Additionally, the identification of clusters of loci associated with MCI as well as future progression to AD indicates that these clusters, rather than individual loci, may have potential as diagnostic and prognostic biomarkers. A further strength of the studies described in this thesis is the use of the ADNI cohort for replication analyses in Chapters 4 and 5, which lends confidence to the findings in the AddNeuroMed cohort.

6.3.2. Limitations

6.3.2.1. Sample size

The first limitation that is common to all chapters, and indeed common to many studies, is that of sample size. In order to detect effect sizes similar to those detected in the overall diagnosis-associated DMPs (Chapter 3), a sample size of 161 would be required per diagnostic group with a power set to 0.8. The limited sample size is particularly a concern for the more complex interaction model used in Chapter 4, and the analysis stratified by sex in Chapter 5. I have mitigated this limitation by validating the AD-associated DMR (*HOXB6*) identified in the AddNeuroMed cohort in Chapter 3 using an independent technique (pyrosequencing), and by running replication analyses using the ADNI cohort for Chapters 4 and 5.

6.3.2.2. Cohort limitations

There are several cohort characteristics that pose limitations to the studies reported in this thesis. First, the MCI-AD individuals in the AddNeuroMed cohort converted to AD within one year of baseline DNA sampling, whilst due to the limited follow up period we do not know if any of the MCI-MCI individuals converted to AD at a later time. This may limit the overlap of results found in the AddNeuroMed cohort to results found in other cohorts, as the combined MCI-MCI and MCI-AD individuals may be temporally closer to AD-like profiles than individuals at earlier stages of MCI. In the ADNI replication cohort, this has been taken into account in the selection criteria of MCI-AD individuals, and in the ratio of MCI-AD to MCI-MCI individuals selected.

Limitations that apply to the ADNI cohort include missing data on the bisulfite treatment batch of samples, for which the data could not be corrected. Additionally, the CTL samples in ADNI were significantly older than those in AddNeuroMed (Chapter 2.3).To prevent further loss of samples, and since all diagnostic groups matched in terms of MMSE scores, the distribution of *APOE* ε4 alleles, and the MCI and AD groups matched in age, no CTL samples were removed from the ADNI cohort in order to match the AddNeuroMed age. As the ADNI cohort contains longitudinal data, and samples were selected to match AddNeuroMed samples, some of the selected samples were not taken at baseline. Since repeated exposure to cognitive tests such as MMSE has been shown to facilitate learning effects (Jolles et al., 1998), this may influence performance and therefore results of the MMSE test. While only MMSE scores from the AddNeuroMed cohort were used in Chapter 3 to study associations with biological pathways, the ADNI study used cognitive test scores to assist in diagnostic assessment, which could in theory be influenced by learning effects. These limitations specific to the ADNI cohort may contribute to the lack of overlap seen between the ADNI and AddNeuroMed cohorts in Chapter 4 and Chapter 5.

Finally, there are limitations that apply to both cohorts. Several types of medication may influence DNA methylation levels (Boks et al., 2012), which makes distinguishing these effects from disease effects on DNA methylation challenging, particularly in AD. Additionally, many of the findings in this thesis were related to differences in the MCI group. As there is currently no standard treatment for MCI available, medication use may be less common in this group, which could explain these findings.

The second limitation applicable to both cohorts is related to the exclusion of comorbidities. For example, both cohorts exclude individuals with depression, while depressive symptoms have been associated with a prodromal stage of AD and risk of progression to AD (Amieva et al., 2008; Copeland et al., 2003). While this exclusion criterion eliminates potential confounders from the studies, it may also exclude part of the population that is of interest. In line with the exclusion of comorbidities, it should be noted that the results found in thesis should be investigated for AD-specificity. While I was not able to do this in the current cohorts, it would be interesting for future research to examine whether the *HOXB6* DMR, for example, is specific to AD, or whether it is a marker of general neurodegeneration.

6.3.2.3. DNA methylation array limitations

DNA methylation arrays, unlike sequencing-based approaches, are targeted at specific regions of the genome, and therefore provide less coverage. The 450K array was initially designed to target the promoter region, the 5' untranslated region (UTR), first exon, gene body, and 3'UTR of known genes. It was superseded by the EPIC array, which covered many of the loci included on the 450K array, and added numerous sites targeted at enhancer regions. However, the nature of arrays is such that each genomic element is often only covered by a single probe. In order to increase coverage, and thereby gain a more complete assessment of blood DNA methylation, a sequencing approach such as whole genome bisulfite sequencing (WGBS) could provide a solution, although this may be cost prohibitive.

In addition to limitations regarding coverage, the 450K array and the EPIC array both work with two types of probes (as detailed in Chapter 2.2.2). Due to the nature of the type I probe, which assumes that neighbouring CpGs have the same methylation status, a bias is created. Each probe is 50 bp long, and if one or more CpG sites other than the query site are located within the complementary sequence to the probe, type I probes are more likely to bind to the DNA sequence if the CpG sites have the same methylation status. This limitation was solved in type II probes, through the use of degenerate R [A/G] bases. However, type II probes only have degenerate R bases for up to three CpG sites within the complementary sequence. This is the reason type I probes are enriched in CpG islands, and are known to cover more unmethylated loci than type II probes (as seen in the beta distribution in Figure 2.8 and Figure 2.19).

6.3.2.4. Limitations related to methodology

The DNA methylation profiles described in this thesis are based on β-values. An alternative method of analysis may be based on M-values, which have a logistic relationship to β-values. M-values have been suggested to be statistically preferable to β-values, as β-values show heteroscedasticity for highly methylated or highly unmethylated loci (Du et al., 2010). This may be particularly relevant for the analysis of the Xi chromosome, which is highly methylated, though it is not currently known to what extent this may bias results. However, β-values, unlike M-values, are biologically interpretable as they range from 0 (not methylated) to 1 (fully methylated). Additionally, β-values are used in the majority of EWAS of AD, thereby making results based on β-values more comparable across studies. Taking these factors into account, I opted to use β-values in the studies described in this thesis.

A second limitation that should be noted is the cross-sectional nature of the research conducted for this thesis. Although the primary AddNeuroMed cohort did not include DNA sampling at follow-up measurements, the ADNI cohort does include longitudinal data. Temporal patterns of DNA methylation, and further characterisation of the development and progression of AD for biomarker development, would be interesting to study in the ADNI cohort as well as other longitudinal cohorts of AD.

6.3.2.5. Tissue- and cell-heterogeneity

Closely related to the β-value range described above, is the fact that DNA methylation at a given locus is a dichotomous measure – a site is either methylated or it is not. However, the β-values calculated using the 450K array are continuous, which reflects the fact that samples are comprised of multiple cell types or a population of multiple cells. A methylation value that is between 0 and 1 therefore reflects that a proportion of query loci within the sample are methylated (Birney et al., 2016). As DNA methylation is both tissue-specific as well as cell type-specific, this makes the interpretation of the generated data more complex. Additionally, the use of heterogeneous blood tissue could obscure cell-type specific modifications related to the disease, as changes in one cell type could negate or distort changes in another, as has been previously shown in the context of AD brain (Blalock et al., 2011). In all analyses described in this thesis, the limitation of cell type heterogeneity has been mitigated by correcting for cell type composition imputed from DNA methylation levels (Houseman et al., 2012). In order to disentangle DNA methylation profiles specific to certain cell types, future research would ideally use methods such as fluorescence-activated cell

sorting (FACS), or single-cell genome-wide bisulfite sequencing (Smallwood et al., 2014). However, the implementation of these methods is more expensive and time-consuming than the use of DNA methylation arrays.

6.3.2.6. Blood as a surrogate for brain tissue

Though not necessarily a limitation of the research presented in this thesis, it should be noted that, as AD is a neurodegenerative disease, the DNA methylation profile detected in blood may not match the profile in brain tissue. Blood DNA methylation may reflect parallel effects, comorbidities, mediating effects, or downstream effects, and may therefore not be functionally related to AD-related processes in the brain. This is not considered a limitation as the current studies are not aimed at assessing causality, but aim to characterise patterns of DNA methylation of interest to future biomarker development. With this aim in mind, robustness and AD-specificity of identified markers are of primary importance for future studies.

6.3.2.7. Outlook on future perspectives

To follow up the research presented in this thesis, and to summarise the future research suggested in the sections above, there are a few next steps I would recommend taking. First, I would further investigate the biomarker potential of the *HOXB6* DMR, by assessing AD-specificity and sensitivity for the diagnosis of AD. Second, I would investigate the diagnostic and prognostic biomarker potential of clusters of co-methylated loci, as Chapter 3 has shown that these clusters are altered in association with MCI and progression to AD. I would investigate whether these could be used to reliably diagnose MCI and/or AD, or predict progression to AD, using unsupervised classification methods such as random

forest models. Third, in order to develop biomarkers associated with progression to AD, I would examine temporal DNA methylation patterns in longitudinal cohorts such as ADNI, thereby also extending the limited time to conversion that is studied in this thesis. Fourth, as Chapter 4 has underscored the importance of investigating sex-specificity, ideally I would conduct the research described above across both sexes and in a stratified manner, to examine whether diagnostic or prognostic biomarkers may be different in females and males. However, it may be challenging to obtain a single cohort of sufficient sample size to conduct this research. Fifth, in view of the effects of age and age by diagnosis interactions identified in Chapter 4, it would be interesting to explore the relationship between accelerated DNA methylation ageing (Horvath, 2013) and diagnosis of MCI and AD. Accelerated ageing has been previously linked to risk factors of AD (McCartney et al., 2018), though a recent study found no association with risk of dementia in individuals over 79 years old (Sibbett et al., 2020). It would be interesting to study accelerated ageing in young individuals who later progress to AD. Finally, I would further investigate the validity of the Xi DNA methylation findings presented in Chapter 5, through allele-specific DNA methylation sequencing.

Broadly speaking, I see future studies of AD moving toward cell-sorted or single-cell DNA methylation sequencing studies, which would eliminate the limitations of genomic coverage by arrays as well as cell-type heterogeneity. Furthermore, as AD is a complex disease and the current research in the field does not indicate causality of a single gene or locus, I believe it is important for future research to include multi-omics investigations of AD, in order to study the interplay between different biological systems such as genomics, epigenomics, transcriptomics and

proteomics. However, it may be very challenging to obtain a sizeable, well-defined cohort of samples which addresses the limitations described above, and which has different modalities of data available.

6.4. Conclusions

To summarise, the studies described in this thesis have provided novel insights into DNA methylation profiles in blood of individuals with AD or MCI, the effects of the risk factors age and sex, and disease-associated DNA methylation profiles on the sex chromosomes. The *HOXB6* gene has been presented as an interesting and robust target for future diagnostic biomarker studies of AD. Moreover, future studies on DNA methylation in MCI and AD should not only correct for covariates, but should also take into account sex and age interactions.

APPENDIX A: EPIGENETICS AND DNA METHYLOMIC PROFILING IN ALZHEIMER'S DISEASE AND OTHER NEURODEGENERATIVE DISEASES

Journal of Neurochemistry

JOURNAL OF NEUROCHEMISTRY | 2017 | 143 | 158–170 doi: 10.1111/jnc.14148

REVIEW

Epigenetics and DNA methylomic profiling in Alzheimer's disease and other neurodegenerative diseases

Janou A. Y. Roubroeks,* Rebecca G. Smith,* Daniel L. A. van den Hove†,‡ and Katie Lunnon*

*University of Exeter Medical School, University of Exeter, Devon, UK

†Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University Medical Centre, Maastricht, The Netherlands

‡Laboratory of Translational Neuroscience, Division of Molecular Psychiatry, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Würzburg, Würzburg, Germany

Abstract

Recent studies have suggested a role for epigenetic mechanisms in the complex etiology of various neurodegenerative diseases. In this review, we discuss advances that have been made toward understanding the role of epigenetic processes in neurodegenerative disorders, with a particular focus on Alzheimer's disease, where the most extensive studies have been undertaken to date. We provide a brief overview of DNA modifications, followed by a summarization of studies of DNA modifications in Alzheimer's disease and other neurodegenerative diseases.

Keywords: Alzheimer's disease, DNA hydroxymethylation, DNA methylation, epigenetic, epigenome-wide association study, genetic.

J. Neurochem. (2017) **143**, 158–170.

The complex etiology of Alzheimer's disease

Alzheimer's disease (AD) has the largest burden of disease among all neurodegenerative disorders, with prevalence of dementia expected to rise to over 115 million cases worldwide by 2050 (Prince *et al.* 2013). AD is characterized by severe memory loss and cognitive deterioration, resulting in a complete dependency on others, and a greatly decreased life expectancy (Reisberg *et al.* 1985; Smolarek *et al.* 2010). Pathological hallmarks of the disease, including aggregations of extracellular amyloid-β protein, intracellular neurofibrillary tangles of hyperphosphorylated tau protein and progressive neuronal cell death (Yates and McLoughlin 2008), are thought to occur years before a clinical diagnosis can be made (Amieva *et al.* 2008; Jack *et al.* 2010). In the early stages of disease, these hallmarks are most prevalent in the brainstem, hippocampal and entorhinal areas of the brain, and spread to the temporal, parietal and frontal cortex as the disease progresses (Dubois *et al.* 2010; Iatrou *et al.* 2017). Early onset of AD (< 65 years) is generally associated with Mendelian patterns of inheritance, caused by mutations in the amyloid precursor protein (*APP*), and presenilin 1 and 2

Received June 2, 2017; revised manuscript received August 8, 2017; accepted August 9, 2017.

Address correspondence and reprint requests to Katie Lunnon, University of Exeter Medical School, RILD Building Level 4, Royal Devon and Exeter Hospital, Barrack Rd, Exeter. EX2 5DW. UK. E-mail: k.lunnon@exeter.ac.uk

Abbreviations used: 27K array, illumina infinium human methylation 27K array; 450K array, illumina infinium human methylation 450K array; 5caC, 5-carboxylcytosine; 5fC, 5-formylcytosine; 5hmC, 5-hydroxymethylcytosine; 5mC, 5-methylcytosine; AD, Alzheimer's disease; ALS, amyotrophic lateral sclerosis; *APOE*, apolipoprotein E; *APP*, amyloid precursor protein; Aβ, amyloid-β; C, cytosine; CpG, cytosine-phosphate-guanine; CpH, non-CpG cytosine; CPI, CpG island; CRISPR, clustered regularly interspersed short palindromic repeats; DhMRs, differentially hydroxymethylated regions; DLB, dementia with Lewy bodies; DMPs, differentially methylated positions; DMRs, differentially methylated regions; DNMTs, DNA methyltransferases; EWAS, epigenome-wide association studies; FACS, fluorescence-activated cell sorting; FTD, frontotemporal dementia; GC/MS, gas chromatography/mass spectrometry; GWAS, genome-wide association studies; HD, Huntington's disease; LCM, laser capture microdissection; MACS, magnetic-activated cell sorting; MCI, mild cognitive impairment; NFTs, neurofibrillary tangles; PD, Parkinson's disease; *PSEN1*, presenilin 1; *PSEN2*, presenilin 2; SAM, *S*-adenosyl methionine; TET, ten-eleven translocation; UTR, untranslated region.

genes. However, the exact cause of the far more common, sporadic, late-onset form of the disease (accounting for > 95% of cases) remains unclear (Bertram *et al.* 2010; Reitz and Mayeux 2014).

To elucidate the complex etiology of sporadic AD, a considerable amount of research has focused on identifying DNA sequence variation, mainly through genome-wide association studies (GWAS). Such studies have nominated a number of common variants, although these tend to have a low effect size, with the exception of apolipoprotein E (*APOE*) (Lambert *et al.* 2013). However, a recent study has suggested that genetic variation only accounts for ~30% of phenotypic variance (Ridge *et al.* 2013). Furthermore, observations of monozygotic twins have shown discordance in AD development and onset (Plomin *et al.* 1994; Gatz *et al.* 1997; Mastroeni *et al.* 2009), implicating a role for non-genetic factors in the development of AD, and which has led to a rapidly expanding number of studies investigating a role for epigenetic modifications in AD. Epigenetics refers to the mitotically and meiotically heritable changes in gene expression without alterations in the underlying DNA sequence. This also potentially allows for alterations in gene expression in response to environmental variation, such as stress, diet or exposure to environmental chemicals. Of the various known mechanisms, DNA modifications of the cytosine base are perhaps the most researched and, as such, will be the focus of this review. We will provide a brief overview of DNA modifications, describing the current literature that has examined their role in neurodegenerative diseases, with a particular focus on AD.

Epigenetic DNA modifications and their role in the brain

Currently, the cytosine base is known to exist in five states: as unmodified cytosine (C), 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). The most common state of cytosine in the brain, after the unmodified state, is 5mC, which is mainly located in cytosine-phosphate-guanine (CpG) dinucleotides. Genome-wide distribution studies have found 5mC in a large majority of CpG dinucleotides, with the exception of high-density CpG areas, known as CpG islands, which largely remain unmethylated (Meissner *et al.* 2008; Lister *et al.* 2009; Li *et al.* 2010). Unmodified DNA can be methylated through DNA methyltransferases (DNMTs) that transfer a methyl group from *S*-adenosyl methionine (SAM) to the 5 position of cytosine (Fig. 1). DNMT3A and DNMT3B are *de novo* methyltransferases, acting on unmodified cytosine, while DNMT1 is thought to be largely responsible for the maintenance of methylation through cell divisions (Chen and Riggs 2011). DNA methylation plays a crucial role in several key processes, such as genomic imprinting and X-chromosome inactivation (Bonasio *et al.* 2010). Although it was originally believed to be solely associated with transcriptional repression, recent evidence suggests a differential effect on gene expression depending on location in or around the gene (Ziller *et al.* 2013), with intragenic DNA methylation shown to modulate alternative splicing (Maunakea *et al.* 2013) and gene body methylation associated with increased expression (Varley *et al.* 2013).

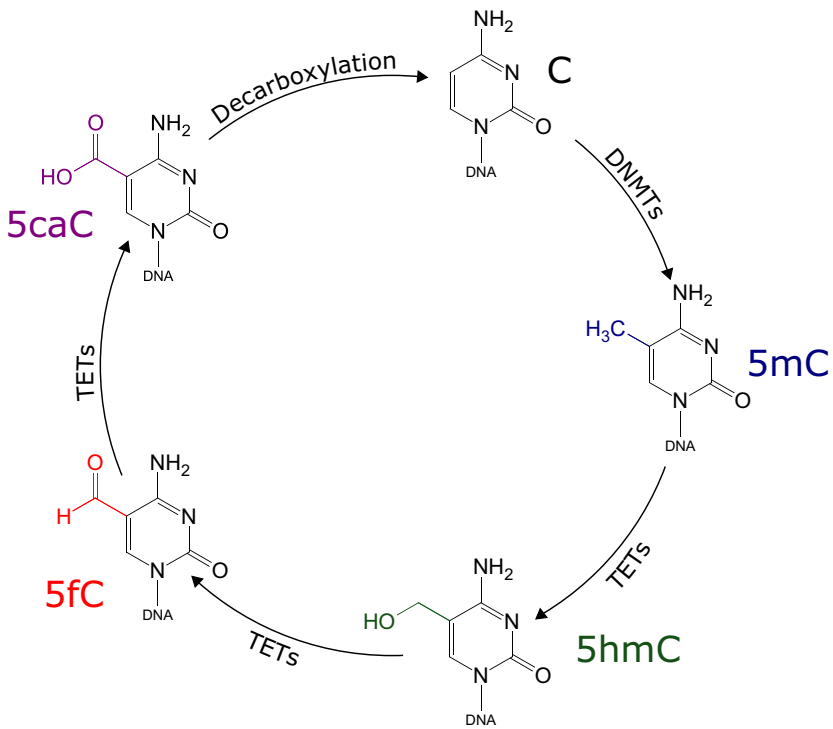

Fig. 1 Cycle of cytosine modifications. Unmethylated cytosine (C) can be converted into 5-methylcytosine (5mC) through DNA methyltransferases (DNMTs). Active demethylation of 5mC, 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) occurs via ten-eleven translocation (TET) enzymes and decarboxylation.

Demethylation, the transition from methylation to unmodified cytosine, can occur either passively or actively (Delatte and Fuks 2013). Passive demethylation takes place over several DNA replication cycles, during which DNMT1 is inhibited or absent from the process, leaving the newly synthesized DNA strand unmethylated. Active demethylation refers to the enzymatic conversion of 5mC to cytosine, which is initiated by oxidation of 5mC into 5hmC by a family of ten-eleven translocation (TET) proteins (Tahiliani *et al.* 2009; Guo *et al.* 2011; Delatte and Fuks 2013). Initially believed to be a transient step in the demethylation process, in recent years, 5hmC has been shown to be a potentially independent and functional epigenetic marker (van den Hove *et al.* 2012; Sun *et al.* 2014). It has been found to be present in most tissues and cell types, but is particularly abundant in the brain (Wen and Tang 2014), where it shows a genomic region-specific distribution (Lunnon *et al.* 2016). 5hmC can be further oxidized into 5fC by TET enzymes, which in turn can be oxidized into 5caC in the demethylation process (He *et al.* 2011; Ito *et al.* 2011). Finally, the demethylation process can be completed through decarboxylation of 5caC into unmodified cytosine (Wu and Zhang 2010; Ito *et al.* 2011).

A role for DNA modifications in AD?

Initially, most human epigenetic studies focused on identifying changes in global DNA methylation in post-mortem brain tissue, employing immunohistochemical techniques with antibodies specific for 5mC or 5hmC (Table 1A). These studies highlight the potential tissue specificity of epigenetic modifications, as some studies of the hippocampus (5mC + 5hmC; (Chouliaras *et al.* 2013)), temporal neocortex (5mC; (Mastroeni *et al.* 2009)) and entorhinal cortex (5mC; (Mastroeni *et al.* 2010), 5hmC; (Condliffe *et al.* 2014)) have found a global decrease in DNA modifications in AD patients compared to controls. However, a study of the middle temporal gyrus and middle frontal gyrus demonstrated elevated levels of 5mC and 5hmC in neurons of AD patients (Coppieters *et al.* 2014). While other studies have published conflicting evidence; Bradley-Whitman and Lovell showed increased 5mC and 5hmC in the hippocampus in AD (Bradley-Whitman and Lovell 2013), and Lashley *et al.* (2015) showed no global changes in either modification in the entorhinal cortex. A recently published study using the more specific technique of gas chromatography/mass spectrometry (GC/MS) examined global levels of methylation and hydroxymethylation in the superior temporal gyrus and middle temporal gyrus, hippocampus/parahippocampal gyrus, cerebellum and the inferior parietal lobe across various stages of AD (Ellison *et al.* 2017). It was found that brain regions affected by AD pathology show global alterations in 5mC and 5hmC in the early stages of AD; yet, in later stages of the disease, these global levels have reversed to levels observed in controls. This disease progression-related change in global DNA methylation and hydroxymethylation levels could contribute to the contradictory results found in the previous immunohistochemical studies, although methodological differences in tissue processing may also play an important role.

Although global DNA modification studies in AD have highlighted a potential role for epigenetic mechanisms in AD, it is, however, important that changes in individual genes are studied, to allow us to better understand mechanisms and pathways. Candidate gene studies have mainly focused on methylation profiling of a vast array of genes, the majority of which were studied based on previous association with either early- or late-onset AD (see Table 1B). Initially, studies focused on *APP* gene methylation in AD, although these have been relatively inconsistent; an early study of *APP* methylation found AD-associated hypomethylation of *APP* in the temporal lobe (West *et al.* 1995), while more recent studies have found no AD-related changes in DNA methylation (Barrachina and Ferrer 2009; Brohede *et al.* 2010), or even increased *APP* DNA methylation (Iwata *et al.* 2014). However, the use of different techniques in each of these studies, the limited sample numbers and the use of heterogeneous tissue could be responsible for conflicting results. Since the initial study of *APP*, a wide range of other candidate genes has been studied. DNA methylation alterations have been found in a number of genes, including *PP2AC* (Sontag *et al.* 2004), *S100A2* and *SORBS3* (Siegmund *et al.* 2007), *BDNF*, *SYP*, *NF-κβ*, and *COX*-2 (Rao *et al.* 2012) and *TREM2* (Celarain *et al.* 2016; Smith *et al.* 2016). Studies of alterations in *MAPT* DNA methylation in AD are currently inconclusive, with reports of both AD-associated hypomethylation (Iwata *et al.* 2014) and no AD-related changes (Barrachina and Ferrer 2009).

In recent years, major technological advances have allowed the first epigenome-wide association studies (EWAS) in AD (Table 1C) (Smith and Lunnon 2017). The first AD EWAS utilized the Illumina Infinium Human Methylation 27K array (27K array) to study > 27,000 CpG sites in 14,475 genes (Bakulski *et al.* 2012). The study identified 948 CpG sites spanning 918 unique genes associated with late-onset AD, in a comparison of frontal cortex tissue from 12 AD cases and 12 age- and gender-matched controls. The most significant AD-associated CpG site showed hypomethylation in the *TMEM59* gene, with a 7.3% difference on the 27K array between AD patients and controls, and a 2.7% difference shown via pyrosequencing validation. Sanchez-Mut *et al.* (2014) also used the 27K array, and examined hippocampal tissue of five Braak stage I–II cases, five Braak stage III–IV cases, five Braak stage V–VI cases and five controls. Braak-associated DNA methylation alterations were found in four loci, of which two resided in *DUSP22*, and one locus each in *CLDN15* and *QSCN6*. The group also reported that hypermethylation of *DUSP22* correlated with its decreased RNA expression.

Table 1 Overview of studies examining a role for DNA modifications in AD

Article	Sample type	Number of samples		Type of modification	Approach	Method	Main findings
		AD	Control				
A. Global measurements of DNA modifications							
Mastroeni *et al.* 2009	TN, CER	1	1	5mC	Global assessment in discordant monozygotic twin pair	Immunohistochemistry	Decreased global 5mC in neurons in AD
Mastroeni *et al.* 2010	EC	20	20	5mC	Global assessment	Immunohistochemistry	Decreased global 5mC in neurons in AD
Chouliaras *et al.* 2013	HIP (CA1, CA3, DG)	10	11	5mC, 5hmC	Global, cell-specific analysis in AD/control samples and a discordant monozygotic twin pair	Immunohistochemistry	Decreased global 5mC and 5hmC in glia and neurons in AD
Bradley-Whitman and Lovell 2013	HIP/PHG, CER	12 (5 Preclinical + 7 late-stage AD)	5	5mC, 5hmC, 5fC, 5caC	Global assessment	Immunohistochemistry	Increased global 5mC and 5hmC, and decreased 5fC and 5caC in all AD subjects
Coppieters *et al.* 2014	MTG, MFG	29	29	5mC, 5hmC	Global assessment	Immunohistochemistry	Increased global 5mC and 5hmC in AD neurons
Condliffe *et al.* 2014	EC, CER	13	8	5mC, 5hmC, 5fC, 5caC	Global assessment	Immunohistochemistry	Decreased global 5hmC in EC and CER in AD
Lashley *et al.* 2015	EC, CER	12	14	5mC, 5hmC	Global assessment	Immunohistochemistry, ELISA	No global neuronal changes in 5mC and 5hmC
Ellison *et al.* 2017	STG, MTG, HIP/PHG, IPL, CER	29 (8 preclinical AD + 8 MCI + 11 late-stage AD)	10	C, 5mC, 5hmC	Global assessment	Gas chromatography/mass spectrometry	Global levels of 5mC and 5hmC altered in early disease stages, return to basal levels at later stages
B. Candidate gene studies							
West *et al.* 1995	TL (BA38)	1	1 C, 1 non-AD dementia	5mC	*APP* targeted	Southern blot	AD-associated decreased *APP* methylation

(continued)

Table 1. (continued)

Article	Sample type	Number of samples		Type of modification	Approach	Method	Main findings
		AD	Control				
Sontag *et al.* 2004	MFG, STG, CER	48	24	5mC	PP2AC targeted methylation analysis	Western blot	AD-associated decreased PP2AC methylation
Siegmund *et al.* 2007	ALTLC	18	39	5mC*	50 target genes	RT-PCR (MethyLight)	AD-associated decreased S100A2 methylation and increased SORBS3 methylation
Wang *et al.* 2008	PFC, lymphocytes	Blood: 6 Brain: 24	Blood: 6 Brain: 10	5mC*	12 target genes previously associated with AD	MALDI-TOF mass spectrometry	Identified epigenetic drift from the norm in late-onset AD
Barrachina and Ferrer 2009	FC, HIP	Stage I–II: 17 Stage III–IV: 15 Stage V–VI: 12	26	5mC*	MAPT, *APP*, PSEN1, RAGE, ADORA2A and UCHL1 targeted analysis of 5mC in AD and other tauopathies	MALDI-TOF mass spectrometry	No AD-associated methylation differences found in any of the target genes
Brohede *et al.* 2010	FC, PC, TC, CER	6	–	5mC*	5mC analysis targeted at CpG island at 5′ end of *APP* gene	Capillary electrophoresis	No methylation at *APP* CpG island in any of the investigated brain regions
Rao *et al.* 2012	FC (BA9)	10	10	5mC	Global 5mC, and 8 target genes known for differential expression in AD	RT-PCR (MethyLight)	Increased methylation of BDNF and synaptophysin, decreased methylation of NF-κβ and COX-2
Furuya *et al.* 2012b	EC, AC, HIP, blood	Blood: 34 Brain: 10	Blood: 22 (young), 23 (elderly) Brain: 10	5mC*	SNAP25 promoter targeted	MALDI-TOF mass spectrometry	No AD-associated methylation differences in SNAP25 promoter in brain or blood
Furuya *et al.* 2012a	EC, AC, HIP, blood	Blood: 36 Brain: 12	Blood: 25 (young), 23 (elderly) Brain: 10	5mC*	SORL1 and SIRT1 promoter targeted	MALDI-TOF mass spectrometry	No AD-associated methylation differences in SORL1 and SIRT1 promoters in brain or blood

(continued)

Table 1. (continued)

Article	Sample type	Number of samples		Type of modification	Approach	Method	Main findings
		AD	Control				
Silva *et al.* 2013	EC, AC, HIP	12	10	5mC*	CNP and DPYSL2 targeted	MALDI-TOF mass spectrometry	No AD-associated methylation differences in *CNP* and *DPYSL2* promoters
Silva *et al.* 2014	EC, AC, HIP, blood	Blood: 34 Brain: 12	Blood: 23 Brain: 10	5mC*	Targeted at promoter regions of HSPA8 and HSPA9	MALDI-TOF mass spectrometry	No AD-associated methylation differences in HSPA 8 and HSPA9 promoters in brain or blood
Iwata *et al.* 2014	Grey matter from ITL, CER, SPL	CER: 45 SPL: 59 ITL: 56	CER: 71 SPL: 76 ITL: 74	5mC*	Genes related to sporadic or familial AD	FACS, pyrosequencing	Increased *APP* methylation and decreased *MAPT* methylation in both neuronal and non-neuronal cells
Celarain *et al.* 2016	HIP	10	6	5mC*, 5hmC	5mC and 5hmC analysis of *TREM2*	Cloning-based Sanger sequencing, 5hMeDIP	AD-associated increase of 5mC+5hmC in *TREM2* TSS-associated region
Smith *et al.* 2016	STG	Cohort 1: 66 Cohort 2: 44 Cohort 3: 117	Cohort 1: 29 Cohort 2: 59 Cohort 3: 75	5mC*	Meta-analysis of 3 cohorts targeting *TREM2*	Illumina 450K array, pyrosequencing	AD-associated hypermethylation of *TREM2*
C. Epigenome-wide association studies							
Bakulski *et al.* 2012	FC	12	12	5mC*	AD-control EWAS	Illumina 27K array	948 AD-associated CpG sites representing 918 unique genes, *TMEM59*
Sanchez-Mut *et al.* 2014	HIP	Braak stage I–II: 5 Braak stage III–IV: 5 Braak stage V–VI: 5	5	5mC*	Braak-stage EWAS	Illumina 27K array	4 Braak-associated CpG sites in *DUSP22, CLDN15, QSCN6 genes*
De Jager *et al.* 2014	DLPFC	708		5mC*	Plaque burden EWAS	Illumina 450K array	71 Plaque burden-associated CpG sites, *BIN1*, *RHBDF2*, *ANK1*
Lunnon *et al.* 2014	EC, STG, PFC, CER, blood	122		5mC*	Braak-stage EWAS	Illumina 450K array	*ANK1, MIR486, RHBDF2*
Watson *et al.* 2016	STG	34	34	5mC*	AD-control EWAS	Illumina 450K array	479 AD-associated DMRs

(continued)

Table 1. (continued)

Article	Sample type	Number of samples		Type of modification	Approach	Method	Main findings
		AD	Control				
Zhao *et al.* 2017	DLPFC	AD: 20 10 MCI: 4	6	5hmC*	Plaque & NFT EWAS	High-throughput sequencing	517 plaque-associated DhMRs, 60 NFT-associated DhMRs

5hmC, 5-hydroxymethylcytosine; 5mC, 5-methylcytosine; AC, auditory cortex; AD, Alzheimer's disease; ALTLC, anterior lateral temporal lobe cortex BA, brodmann area; AP, amyloid precursor protein; CA1, cornu ammonis area 1; CA3, cornu ammonis area 3; CER, cerebellum; CpG, cytosine-phosphate-guanine; DG, dentate gyrus; DhMRs, differentially hydroxymethylated regions; DLPFC, dorsolateral prefrontal cortex; DMRs, differentially methylated regions; EC, entorhinal cortex; EWAS, epigenome-wide association studies; FACS, fluorescence-activated cell sorting; FC, frontal cortex; HIP, hippocampus; IPL, inferior parietal lobe; ITL, inferior temporal lobe; MCI, mild cognitive impairment; MFG, middle frontal gyrus; MTG, middle temporal gyrus; NFT, neurofibrillary tangles; PC, parietal cortex; PFC, prefrontal cortex; PHG, parahippocampal gyrus; PSEN1, presenilin 1; SPL, superior parietal lobe; STG, superior temporal gyrus; TC, temporal cortex; TL, temporal lobe; TN, temporal neocortex.
Studies have investigated (A) global levels of DNA modifications, (B) DNA modifications within candidate genes, and (C) epigenome-wide changes (epigenome-wide association studies-EWAS).
*The methods used in these studies could not distinguish 5mC and 5hmC.

The 27K array was superseded by the Illumina Infinium Human Methylation 450K array (450K array), which is currently the most widely used method for EWAS. It interrogates more than 485,000 CpG sites covering 98.9% of known UCSC RefGenes and 96% of CpG islands (Bibikova *et al.* 2011). Watson *et al.* (2016) recently used this technique to examine superior temporal gyrus tissue from 34 patients with late-onset AD, and an equal number of age-, race-, and gender-matched non-demented controls. They analyzed neighboring AD-associated CpG sites and identified 479 differentially methylated regions (DMRs) with an average size of 927 bp, the majority of which were found to be hypermethylated in AD. Of the 25 most significant DMRs, eight genes (*LOC100507547*, *PRDM16*, *PPT2*, *PPT2-EGFL8*, *PRRT1*, *C10orf105*, *CDH23 and RNF39)* had been previously reported in the first AD EWAS studies (De Jager *et al.* 2014; Lunnon *et al.* 2014). Lunnon *et al.* (2014) studied a cohort of 122 donors in a cross-tissue approach, using tissue from the superior temporal gyrus, entorhinal cortex, prefrontal cortex, cerebellum and pre-mortem blood. These regions are known to be affected differentially by AD, with the entorhinal cortex showing pathology early in the disease process, and the cerebellum remaining relatively unaffected (Wenk 2003). Initially focusing on the entorhinal cortex, the study found two differentially methylated positions located in the *ANK1* gene among the most significant Braak stage-associated sites. These results were replicated in the prefrontal cortex and superior temporal gyrus in the same individuals; yet, no disease-associated changes in these loci were found in the cerebellum or pre-mortem blood. The study also validated AD-associated *ANK1* hypermethylation in a further three validation cohorts, including one that used another technology (bisulfite pyrosequencing). Genetic variation in *ANK1* has been associated with diabetes (Imamura *et al.* 2012; Harder *et al.* 2013) and, interestingly, DNA methylation changes in this gene were implicated in AD in the study by De Jager *et al.* (2014). Their EWAS study determined dorsolateral prefrontal cortex DNA methylomic profiles associated with amyloid-β plaque burden in 708 individuals, and identified 71 plaque burden-associated CpG sites. Twelve of these nominated loci were validated in data taken from the Lunnon *et al.* study, reaching Bonferroni significance. Of the genes identified, *ANK1, CHD23, DIP2A, RHBDF2, RPL13, SERPINF1* and *SERPINF2* were found to show significant AD-associated gene expression changes, indicating a potential functional role for the DNA modifications identified in AD.

One caveat with the EWAS studies described above is that they have all used bisulfite-treated DNA, which means that 5mC and 5hmC are indistinguishable, and results actually represent the sum of the two modifications. A recent adaptation to the protocol, whereby DNA is first oxidized prior to bisulfite treatment, allows measurement of 5mC

alone. Furthermore, by performing an oxidative-bisulfite treatment in parallel with a bisulfite treatment, one can subtract one value from the other to generate a measurement of 5hmC in isolation (Fukuzawa *et al.* 2016). To date, only one EWAS of 5hmC in AD has been published (Zhao *et al.* 2017). This study used high-throughput sequencing to examine the distribution of 5hmC in dorsolateral prefrontal cortex tissue of 20 AD patients, 4 mild cognitive impairment patients and 6 non-demented controls. A total of 517 plaque-associated differentially hydroxymethylated regions were identified, along with 60 neurofibrillary tangles-associated differentially hydroxymethylated regions (at $q < 0.05$). However, because of the low sample size these results will require replication in further studies. It should also be noted that because of low sequencing resolution, this study was not able to differentiate 5mC and 5hmC, and the measures reported are that of the combined signal. Further AD EWAS studies using oxidative-bisulfite-treated DNA are highly anticipated.

DNA modifications in other neurodegenerative diseases

Although not as extensively studied as AD, a number of human studies have examined the role of DNA modifications in other neurodegenerative diseases, particularly in Parkinson's disease (PD) and amyotrophic lateral sclerosis (ALS). Most studies have investigated global DNA methylation within a specific brain region affected by the disease, or focused on candidate genes previously linked to the disease. For instance, in ALS, a disease that causes loss of motor neurons in the brainstem, spinal cord and motor cortex (Robberecht and Philips 2013), global DNA methylation and hydroxymethylation were found to be increased in the spinal cord compared to controls (Chestnut *et al.* 2011; Figueroa-Romero *et al.* 2012). However, in blood, contradictory studies have been published, with one highlighting increased global DNA methylation in ALS (Tremolizzo *et al.* 2014), while another showed no change in global 5mC and 5hmC (Figueroa-Romero *et al.* 2012). In PD and the related disorder, dementia with Lewy bodies (DLB), a global decrease in 5mC of up to 30% was found in the frontal cortex of patients compared to non-demented controls (Desplats *et al.* 2011). This study went on to examine the candidate genes *SNCA*, *SEPW1* and *PRKAR2A*, which had been previously implicated in PD. Intron 1 of *SNCA* was found to be hypomethylated in both DLB and PD, and regulatory regions of *SEPW1* and *PRKAR2A* were found to be hypomethylated in DLB and PD, respectively. Most candidate gene studies of PD have focused on intron 1 of *SNCA*, a gene that has been causally associated with Mendelian forms of PD (Singleton *et al.* 2003; Chartier-Harlin *et al.* 2004) and shown to be a risk factor for non-Mendelian PD (Edwards *et al.* 2010). The majority of these studies have reported hypomethylation of *SNCA* intron 1 in both the brain (substantia nigra, putamen and cortex) and blood of PD patients (Jowaed *et al.* 2010; Ai *et al.* 2014; Tan *et al.* 2014; Pihlstrøm *et al.* 2015), although some other studies found no DNA methylation changes in white blood cells (Richter *et al.* 2012; Song *et al.* 2014) or substantia nigra brain tissue (Guhathakurta *et al.* 2017). Other candidate gene studies of PD have reported hypomethylation of the *NPAS2* promoter in blood (Lin *et al.* 2012), and alterations in *MAPT* in the cerebellum and putamen (Coupland *et al.* 2014).

In studies of ALS, a number of candidate gene studies have found alterations in the ALS and frontotemporal dementia-associated gene *C9orf72*, across several tissues (Xi *et al.* 2013, 2015; Russ *et al.* 2015). Furthermore, promoter regions of other genes previously associated with ALS, such as *SOD1* and *VEGF*, were found to be largely unmethylated in both brain and blood of ALS patients and controls (Oates and Pamphlett 2007). The study of DNA methylation in Huntington's disease (HD) is still in its infancy. However, one study on putamen tissue from HD patients and controls demonstrated an increase in 5mC and a decrease in 5hmC in the 5′ untranslated region of the *ADORA2A* gene (Villar-Menéndez *et al.* 2013), the expression of which is known to be altered in HD (Martinez-Mir *et al.* 1991; Glass *et al.* 2000).

With the exception of AD, a limited number of EWAS studies have been performed in neurodegenerative diseases. A study employing the 27K array to examine cortex and putamen tissue from six PD patients and six controls identified a number of genes altered in PD, including *LOC84245*, *PPP4R2*, *CYP2E1*, *CAT*, *DEFA1* and *CHFR* (Kaut *et al.* 2012). Of these, only *CYP2E1* showed hypomethylation in both the cortex and putamen. Masliah *et al.* (2013) studied frontal cortex tissue and peripheral blood leukocytes of five PD patients and six controls using the 450K array. They identified 2,908 and 3,897 CpG sites showing differential methylation in the brain and blood, respectively. A number of these CpG sites were located in genes previously shown to be deregulated in PD or implicated in genome-wide association studies of PD, among which were *HLA-DQA1*, *GFPT2* and *MAPT*. However, because of the low sample size these results will need further validation. To date, two EWAS studies have been performed in ALS. The first made use of GeneChip Human Tiling 2.0R Arrays to examine epigenome-wide DNA methylation in the frontal cortex of 10 male sporadic ALS patients and 10 age- and gender-matched controls (Morahan *et al.* 2009). A total of 38 DMRs were identified, mapping to 23 genes, of which the most significant resided in *ATRN, MSRA*, *PRDM16* and *SGCZ*. The second EWAS study of ALS examined spinal cord tissue from 12 sporadic ALS patients and 11 age- and gender-matched controls using the 27K array (Figueroa-Romero *et al.* 2012). This

study identified 4,261 differentially methylated CpG sites, which represented 3,574 genes.

Study design caveats for EWAS in neurodegenerative diseases

This review has highlighted that further EWAS are still required in a number of neurodegenerative diseases, particularly with larger sample numbers. However, in order to examine the epigenome in neurodegenerative diseases in the brain, there are a few study design caveats to initially consider (Lunnon and Mill 2013). First, it is known that DNA methylation changes occur with aging, with both hypermethylation and hypomethylation observed depending on tissue and genomic location (Christensen *et al.* 2009; Horvath 2013). This highlights the importance of balancing age across study groups. Second, is the issue of the tissue specificity of epigenetic modifications. Neurodegenerative diseases mainly affect a specific brain region, or are characterized by a specific spatiotemporal pattern of changes, affecting specific regions of the brain as the disease progresses. It is therefore important to carefully select the brain region to be examined, as brain regions could be affected differentially across the different disease stages. Ideally, an investigation of multiple brain regions would allow for the identification of spatiotemporal epigenetic changes related to the disease, although this raises the issue of sample size; it may be challenging to obtain a sizeable, well-defined cohort of samples with tissue available from multiple brain regions. Third, and related to this issue, is cell-type specificity. Neurodegenerative diseases are characterized by neuronal loss, which will alter the cell-type composition of diseased samples, which in turn makes the interpretation of generated data difficult. Additionally, the use of heterogeneous tissues, such as brain, could obscure cell-type specific modifications related to the disease, as changes in one cell type could negate or distort changes in another, as has been previously shown in the context of AD (Blalock *et al.* 2011). A solution to these issues could be the use of laser capture microdissection, fluorescence-activated cell sorting or magnetic-activated cell sorting in order to acquire cell-type specific samples. However, these techniques are often expensive and time-consuming. Another method frequently used is to correct for cell-type composition using bioinformatic approaches. For example, one can estimate neuron-glia cellular proportions in the brain tissue that has been analyzed, based on DNA methylation values in the generated EWAS data, and then control for this in downstream analyses (Guintivano *et al.* 2013). Furthermore, individual cells may also differ in their pattern of epigenetic modifications, and the heritability of different epigenetic modifications across multiple cell divisions may vary, which complicates the examination of epigenetic modifications in bulk tissue, such as brain samples. A potential solution to this problem is presented by the use of advanced single-cell sequencing techniques to investigate cell-specific epigenomes (Bheda and Schneider 2014). Finally, many people with neurodegenerative diseases also suffer from other comorbidities. Particularly, comorbidity with another neurodegenerative disease is not uncommon, for example, AD and PD have been shown to occur together (Poblador-Plou *et al.* 2014). This could hinder the identification of specific DNA methylomic profiles to a particular disease.

Many of the issues associated with studying epigenetic changes in the brain are also applicable to studies in the blood, for example, sample cell-type composition and comorbidity also play a major role. Similar to the cellular proportion corrections that can be performed in brain EWAS data, cell-specific DNA methylation markers in the blood can be used to identify cell-type proportions within samples (Houseman *et al.* 2012; Horvath 2013). In the context of blood studies, environmental factors have also been shown to alter blood DNA methylation profiles, for example, smoking (Zeilinger *et al.* 2013), nutrition (Friso and Choi 2002) and environmental chemicals (Hou *et al.* 2012). Similarly, comorbidity with systemic diseases may affect DNA methylation patterns in blood, as is the case with hypertension, which is commonly associated with AD (Smolarek *et al.* 2010; Gerritsen *et al.* 2016).

Future perspectives

The continued development of genomic technologies will lead to more in-depth analyses of DNA modifications. The recent development of the Illumina Infinium Methylation EPIC BeadChip array allows for the interrogation of over 850,000 CpG sites across the genome. Although, to date, no studies using this technique have been published in neurodegenerative diseases, studies using this approach with bisulfite-treated and oxidative-bisulfite-treated DNA, to specifically analyze 5mC and 5hmC in parallel, are anticipated in the AD research field. Most studies examining DNA modifications in the context of AD have focused on DNA methylation and, more recently, DNA hydroxymethylation. However, 5fC has recently been demonstrated to be a stable modification in mice, showing differential levels to 5hmC which may suggest a functional role. Whether 5fC and possibly 5caC could function as independent epigenetic markers remains to be determined (Bachman *et al.* 2015), although 5fC and 5caC are far less abundant than 5hmC and 5mC (Song and He 2013). Additionally, methylation of non-CpG cytosines has been identified in human brain tissue, and has been shown to be negatively correlated with the associated gene expression in dentate neurons of adult mice (Guo *et al.* 2014). Furthermore, methylation of the adenine base, which was initially believed to occur mainly in prokaryotes (Ratel *et al.* 2006), has recently been detected in mammalian genomes (Koziol

et al. 2016; Wu *et al.* 2016), although its exact regulatory potential remains unclear.

At present, the major limitation of the studies discussed in this review, and epigenetic studies in general, is the issue of causality. Alterations in DNA modifications could be either causal in the disease process, or could themselves arise as a result of the pathological changes associated with disease. It is expected that the use of novel (epi)genetic editing techniques will allow DNA methylation to be altered at specific loci. Clustered regularly interspersed short palindromic repeats (CRISPR) uses RNA-guided Cas9 nucleases to allow the introduction of novel sequences to a strand of DNA. It is now possible to fuse the Cas9 protein with the enzymatic domains of DNMT3A or TET1 to add, or remove, methyl groups to DNA, respectively (Liu *et al.* 2016). The use of this technology will ultimately allow the field to determine the functional consequence of loci nominated from EWAS and to establish whether these represent a cause, or a consequence, of the disease process. Furthermore, as epigenetic changes are potentially reversible, the identification of disease-causative mechanisms may provide realistic targets for future therapeutic interventions.

Acknowledgments and conflict of interest disclosure

This work was funded by an Alzheimer's Society project grant to KL (grant number AS-PG-14-038) and an Alzheimer's Association New Investigator Research Grant to KL (grant number NIRG-14-320878). Janou Roubroeks is supported by the GW4 Biomed MRC Doctoral Training Partnerships. The authors declare no conflicts of interest.

References

Ai S.-X., Xu Q., Hu Y.-C. *et al.* (2014) Hypomethylation of SNCA in blood of patients with sporadic Parkinson's disease. *J. Neurol. Sci.* **337**, 123–128.

Amieva H., Le Goff M., Millet X., Orgogozo J. M., Pérès K., Barberger-Gateau P., Jacqmin-Gadda H. and Dartigues J. F. (2008) Prodromal Alzheimer's disease: successive emergence of the clinical symptoms. *Ann. Neurol.* **64**, 492–498.

Bachman M., Uribe-Lewis S., Yang X., Burgess H. E., Iurlaro M., Reik W., Murrell A. and Balasubramanian S. (2015) 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **11**, 555–557.

Bakulski K. M., Dolinoy D. C., Sartor M. A., Paulson H. L., Konen J. R., Lieberman A. P., Albin R. L., Hu H. and Rozek L. S. (2012) Genome-wide DNA methylation differences between late-onset Alzheimer's disease and cognitively normal controls in human frontal cortex. *J. Alzheimers Dis.* **29**, 571–588.

Barrachina M. and Ferrer I. (2009) DNA methylation of Alzheimer disease and tauopathy-related genes in postmortem brain. *J. Neuropathol. Exp. Neurol.* **68**, 880–891.

Bertram L., Lill C. M. and Tanzi R. E. (2010) The genetics of Alzheimer disease: back to the future. *Neuron* **68**, 270–281.

Bheda P. and Schneider R. (2014) Epigenetics reloaded: the single-cell revolution. *Trends Cell Biol.* **24**, 712–723.

Bibikova M., Barnes B., Tsan C. *et al.* (2011) High density DNA methylation array with single CpG site resolution. *Genomics* **98**, 288–295.

Blalock E. M., Buechel H. M., Popovic J., Geddes J. W. and Landfield P. W. (2011) Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient Alzheimer's disease. *J. Chem. Neuroanat.* **42**, 118–126.

Bonasio R., Tu S. and Reinberg D. (2010) Molecular signals of epigenetic states. *Science* **330**, 612–616.

Bradley-Whitman M. A. and Lovell M. A. (2013) Epigenetic changes in the progression of Alzheimer's disease. *Mech. Ageing Dev.* **134**, 486–495.

Brohede J., Rinde M., Winblad B. and Graff C. (2010) A DNA methylation study of the amyloid precursor protein gene in several brain regions from patients with familial Alzheimer disease. *J. Neurogenet.* **24**, 179–181.

Celarain N., Sanchez-Ruiz de Gordoa J., Zelaya M. V., Roldan M., Larumbe R., Pulido L., Echavarri C. and Mendioroz M. (2016) TREM2 upregulation correlates with 5-hydroxymethycytosine enrichment in Alzheimer's disease hippocampus. *Clin. Epigenetics* **8**, 37.

Chartier-Harlin M.-C., Kachergus J., Roumier C. *et al.* (2004) α-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* **364**, 1167–1169.

Chen Z.-X. and Riggs A. D. (2011) DNA methylation and demethylation in mammals. *J. Biol. Chem.* **286**, 18347–18353.

Chestnut B. A., Chang Q., Price A., Lesuisse C., Wong M. and Martin L. J. (2011) Epigenetic regulation of motor neuron cell death through DNA methylation. *J. Neurosci.* **31**, 16619–16636.

Chouliaras L., Mastroeni D., Delvaux E. *et al.* (2013) Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiol. Aging* **34**, 2091–2099.

Christensen B. C., Houseman E. A., Marsit C. J. *et al.* (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet.* **5**, e1000602.

Condliffe D., Wong A., Troakes C. *et al.* (2014) Cross-region reduction in 5-hydroxymethylcytosine in Alzheimer's disease brain. *Neurobiol. Aging* **35**, 1850–1854.

Coppieters N., Dieriks B. V., Lill C., Faull R. L. M., Curtis M. A. and Dragunow M. (2014) Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. *Neurobiol. Aging* **35**, 1334–1344.

Coupland K. G., Mellick G. D., Silburn P. A. *et al.* (2014) DNA methylation of the *MAPT* gene in Parkinson's disease cohorts and modulation by vitamin E *in vitro*. *Mov. Disord.* **29**, 1606–1614.

De Jager P. L., Srivastava G., Lunnon K. *et al.* (2014) Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.* **17**, 1156–1163.

Delatte B. and Fuks F. (2013) TET proteins: on the frenetic hunt for new cytosine modifications. *Brief. Funct. Genomics* **12**, 191–204.

Desplats P., Spencer B., Coffee E., Patel P., Michael S., Patrick C., Adame A., Rockenstein E. and Masliah E. (2011) Alpha-synuclein sequesters Dnmt1 from the nucleus: a novel mechanism for epigenetic alterations in Lewy body diseases. *J. Biol. Chem.* **286**, 9031–9037.

Dubois B., Feldman H. H., Jacova C. *et al.* (2010) Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurol.* **9**, 1118–1127.

Edwards T. L., Scott W. K., Almonte C. *et al.* (2010) Genome-wide association study confirms SNPs in *SNCA* and the *MAPT* region as common risk factors for Parkinson disease. *Ann. Hum. Genet.* **74**, 97–109.

Ellison E. M., Abner E. L. and Lovell M. A. (2017) Multiregional analysis of global 5-methylcytosine and 5-hydroxymethylcytosine

throughout the progression of Alzheimer's disease. *J. Neurochem.* **140**, 383–394.

Figueroa-Romero C., Hur J., Bender D. E. *et al.* (2012) Identification of epigenetically altered genes in sporadic amyotrophic lateral sclerosis. *PLoS ONE* **7**, e52672.

Friso S. and Choi S.-W. (2002) Gene-nutrient interactions and DNA methylation. *J. Nutr.* **132**, 2382S–2387S.

Fukuzawa S., Takahashi S., Tachibana K., Tajima S. and Suetake I. (2016) Simple and accurate single base resolution analysis of 5-hydroxymethylcytosine by catalytic oxidative bisulfite sequencing using micelle incarcerated oxidants. *Bioorg. Med. Chem.* **24**, 4254–4262.

Furuya T. K., da Silva P. N. O., Payão S. L. M. *et al.* (2012a) SORL1 and SIRT1 mRNA expression and promoter methylation levels in aging and Alzheimer's disease. *Neurochemistry International* **61**, 973–975.

Furuya T. K., Silva P. N. O., Payão S. L. M. *et al.* (2012b) Analysis of SNAP25 mRNA expression and promoter DNA methylation in brain areas of Alzheimer's disease patients. *Neuroscience* **220**, 41–46.

Gatz M., Pedersen N. L., Berg S. *et al.* (1997) Heritability for Alzheimer's disease: the study of dementia in Swedish twins. *J. Gerontol. A Biol. Sci. Med. Sci.* **52A**, M117–M125.

Gerritsen A. A. J., Bakker C., Verhey F. R. J. *et al.* (2016) Prevalence of comorbidity in patients with young-onset Alzheimer disease compared with late-onset: a comparative cohort study. *J. Am. Med. Dir. Assoc.* **17**, 318–323.

Glass M., Dragunow M. and Faull R. L. M. (2000) The pattern of neurodegeneration in Huntington's disease: a comparative study of cannabinoid, dopamine, adenosine and GABAA receptor alterations in the human basal ganglia in Huntington's disease. *Neuroscience* **97**, 505–519.

Guhathakurta S., Evangelista B. A., Ghosh S., Basu S. and Kim Y.-S. (2017) Hypomethylation of intron1 of α-synuclein gene does not correlate with Parkinson's disease. *Mol. Brain* **10**, 6.

Guintivano J., Aryee M. J. and Kaminsky Z. A. (2013) A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. *Epigenetics* **8**, 290–302.

Guo J. U., Su Y., Zhong C., Ming G. L. and Song H. (2011) Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423–434.

Guo J. U., Su Y., Shin J. H. *et al.* (2014) Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat. Neurosci.* **17**, 215–222.

Harder M. N., Ribel-Madsen R., Justesen J. M. *et al.* (2013) Type 2 diabetes risk alleles near BCAR1 and in ANK1 associate with decreased β-cell function whereas risk alleles near ANKRD55 and GRB14 associate with decreased insulin sensitivity in the Danish Inter99 cohort. *J. Clin. Endocrinol. Metab.* **98**, E801–E806.

He Y.-F., Li B.-Z., Li Z. *et al.* (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307.

Horvath S. (2013) DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, 3156.

Hou L., Zhang X., Wang D. and Baccarelli A. (2012) Environmental chemical exposures and human epigenetics. *Int. J. Epidemiol.* **41**, 79–105.

Houseman E. A., Accomando W. P., Koestler D. C., Christensen B. C., Marsit C. J., Nelson H. H., Wiencke J. K. and Kelsey K. T. (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86.

Iatrou A., Kenis G., Rutten B. P., Lunnon K. and van den Hove D. L. (2017) Epigenetic dysregulation of brainstem nuclei in the pathogenesis of Alzheimer's disease: looking in the correct place at the right time? *Cell. Mol. Life Sci.* **74**, 509–523.

Imamura M., Maeda S., Yamauchi T. *et al.* (2012) A single-nucleotide polymorphism in ANK1 is associated with susceptibility to type 2 diabetes in Japanese populations. *Hum. Mol. Genet.* **21**, 3042–3049.

Ito S., Shen L., Dai Q., Wu S. C., Collins L. B., Swenberg J. A., He C. and Zhang Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303.

Iwata A., Nagata K., Hatsuta H. *et al.* (2014) Altered CpG methylation in sporadic Alzheimer's disease is associated with APP and MAPT dysregulation. *Hum. Mol. Genet.* **23**, 648–656.

Jack C. R., Jr, Knopman D. S., Jagust W. J., Shaw L. M., Aisen P. S., Weiner M. W., Petersen R. C. and Trojanowski J. Q. (2010) Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol.* **9**, 119–128.

Jowaed A., Schmitt I., Kaut O. and Wüllner U. (2010) Methylation regulates alpha-synuclein expression and is decreased in Parkinson's disease patients' brains. *J. Neurosci.* **30**, 6355–6359.

Kaut O., Schmitt I. and Wüllner U. (2012) Genome-scale methylation analysis of Parkinson's disease patients' brains reveals DNA hypomethylation and increased mRNA expression of cytochrome P450 2E1. *Neurogenetics*, **13**, 87–91.

Koziol M. J., Bradshaw C. R., Allen G. E., Costa A. S. H., Frezza C. and Gurdon J. B. (2016) Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat. Struct. Mol. Biol.* **23**, 24–30.

Lambert J.-C., Ibrahim-Verbaas C. A., Harold D. *et al.* (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458.

Lashley T., Gami P., Valizadeh N., Li A., Revesz T. and Balazs R. (2015) Alterations in global DNA methylation and hydroxymethylation are not detected in Alzheimer's disease. *Neuropathol. Appl. Neurobiol.* **41**, 497–506.

Li Y., Zhu J., Tian G. *et al.* (2010) The DNA methylome of human peripheral blood mononuclear cells. *PLoS Biol.* **8**, e1000533.

Lin Q., Ding H., Zheng Z., Gu Z., Ma J., Chen L., Chan P. and Cai Y. (2012) Promoter methylation analysis of seven clock genes in Parkinson's disease. *Neurosci. Lett.* **507**, 147–150.

Lister R., Pelizzola M., Dowen R. H. *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322.

Liu X. S., Wu H., Ji X. *et al.* (2016) Editing DNA methylation in the mammalian genome. *Cell* **167**(233–247), e217.

Lunnon K. and Mill J. (2013) Epigenetic studies in Alzheimer's disease: current findings, caveats, and considerations for future studies. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **162B**, 789–799.

Lunnon K., Smith R., Hannon E. *et al.* (2014) Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat. Neurosci.* **17**, 1164–1170.

Lunnon K., Hannon E., Smith R. G. *et al.* (2016) Variation in 5-hydroxymethylcytosine across human cortex and cerebellum. *Genome Biol.* **17**, 27.

Martinez-Mir M. I., Probst A. and Palacios J. M. (1991) Adenosine A2 receptors: selective localization in the human basal ganglia and alterations with disease. *Neuroscience* **42**, 697–706.

Masliah E., Dumaop W., Galasko D. and Desplats P. (2013) Distinctive patterns of DNA methylation associated with Parkinson disease. *Epigenetics* **8**, 1030–1038.

Mastroeni D., McKee A., Grover A., Rogers J. and Coleman P. D. (2009) Epigenetic differences in cortical neurons from a pair of monozygotic twins discordant for Alzheimer's disease. *PLoS ONE* **4**, e6617.

Mastroeni D., Grover A., Delvaux E., Whiteside C., Coleman P. D. and Rogers J. (2010) Epigenetic changes in Alzheimer's disease: decrements in DNA methylation. *Neurobiol. Aging* **31**, 2025–2037.

Maunakea A. K., Chepelev I., Cui K. and Zhao K. (2013) Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res.* **23**, 1256–1269.

Meissner A., Mikkelsen T. S., Gu H. *et al.* (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766–770.

Morahan J. M., Yu B., Trent R. J. and Pamphlett R. (2009) A genome-wide analysis of brain DNA methylation identifies new candidate genes for sporadic amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler.* **10**, 418–429.

Oates N. and Pamphlett R. (2007) An epigenetic analysis of SOD1 and VEGF in ALS. *Amyotroph. Lateral Scler.* **8**, 83–86.

Pihlstrøm L., Berge V., Rengmark A. and Toft M. (2015) Parkinson's disease correlates with promoter methylation in the α-synuclein gene. *Mov. Disord.* **30**, 577–580.

Plomin R., Owen M. J. and McGuffin P. (1994) The genetic basis of complex human behaviors. *Science* **264**, 1733–1739.

Poblador-Plou B., Calderón-Larrañaga A., Marta-Moreno J., Hancco-Saavedra J., Sicras-Mainar A., Soljak M. and Prados-Torres A. (2014) Comorbidity of dementia: a cross-sectional study of primary care older patients. *BMC Psychiatry* **14**, 84.

Prince M., Guerchet M. and Prina M. (2013) Policy Brief for Heads of Government: the Global Impact of Dementia 2013–2050. *Alzheimer's Disease International,* London, pp. 1–8.

Rao J. S., Keleshian V. L., Klein S. and Rapoport S. I. (2012) Epigenetic modifications in frontal cortex from Alzheimer's disease and bipolar disorder patients. *Transl. Psychiat.* **2**, e132.

Ratel D., Ravanat J.-L., Berger F. and Wion D. (2006) N6-methyladenine: the other methylated base of DNA. *BioEssays* **28**, 309–315.

Reisberg B., Gordon B., McCarthy M. and Ferris S. H. (1985) Clinical Symptoms Accompanying Progressive Cognitive Decline and Alzheimer's Disease, in Alzheimer's Dementia: Dilemmas in Clinical Research (Melnick V. L. and Dubler N. N., eds), pp. 19–39. Humana Press, Totowa, NJ.

Reitz C. and Mayeux R. (2014) Alzheimer disease: epidemiology, diagnostic criteria, risk factors and biomarkers. *Biochem. Pharmacol.* **88**, 640–651.

Richter J., Appenzeller S., Ammerpohl O., Deuschl G., Paschen S., Brüggemann N., Klein C. and Kuhlenbäumer G. (2012) No evidence for differential methylation of α-synuclein in leukocyte DNA of Parkinson's disease patients. *Mov. Disord.* **27**, 590–591.

Ridge P. G., Mukherjee S., Crane P. K., Kauwe J. S. K. and Alzheimer's Disease Genetics Consortium. (2013) Alzheimer's disease: analyzing the missing heritability. *PLoS ONE* **8**, e79771.

Robberecht W. and Philips T. (2013) The changing scene of amyotrophic lateral sclerosis. *Nat. Rev. Neurosci.* **14**, 248–264.

Russ J., Liu E. Y., Wu K. *et al.* (2015) Hypermethylation of repeat expanded C9orf72 is a clinical and molecular disease modifier. *Acta Neuropathol.* **129**, 39–52.

Sanchez-Mut J. V., Aso E., Heyn H., Matsuda T., Bock C., Ferrer I. and Esteller M. (2014) Promoter hypermethylation of the phosphatase DUSP22 mediates PKA-dependent TAU phosphorylation and CREB activation in Alzheimer's disease. *Hippocampus* **24**, 363–368.

Siegmund K. D., Connor C. M., Campan M., Long T. I., Weisenberger D. J., Biniszkiewicz D., Jaenisch R., Laird P. W. and Akbarian S. (2007) DNA methylation in the human cerebral cortex is dynamically regulated throughout the life span and involves differentiated neurons. *PLoS ONE* **2**, e895.

Silva P. N., Furuya T. K., Sampaio Braga I. *et al.* (2013) CNP and DPYSL2 mRNA expression and promoter methylation levels in brain of Alzheimer's disease patients. *J. Alzheimers Dis.* **33**, 349–355.

Silva P. N., Furuya T. K., Braga I. L. *et al.* (2014) Analysis of HSPA8 and HSPA9 mRNA expression and promoter methylation in the brain and blood of Alzheimer's disease patients. *J. Alzheimers Dis.* **38**, 165–170.

Singleton A. B., Farrer M., Johnson J. *et al.* (2003) α-Synuclein locus triplication causes Parkinson's disease. *Science* **302**, 841.

Smith R. G. and Lunnon K. (2017) DNA modifications and Alzheimer's disease. *Adv. Exp. Med. Biol.* **978**, 303–319.

Smith A. R., Smith R. G., Condliffe D., Hannon E., Schalkwyk L., Mill J. and Lunnon K. (2016) Increased DNA methylation near TREM2 is consistently seen in the superior temporal gyrus in Alzheimer's disease brain. *Neurobiol. Aging* **47**, 35–40.

Smolarek I., Wyszko E., Barciszewska A. M., Nowak S., Gawronska I., Jablecka A. and Barciszewska M. Z. (2010) Global DNA methylation changes in blood of patients with essential hypertension. *Med. Sci. Monit.* **16**, 149–155.

Song C.-X. and He C. (2013) Potential functional roles of DNA demethylation intermediates. *Trends Biochem. Sci.* **38**, 480–484.

Song Y., Ding H., Yang J., Lin Q., Xue J., Zhang Y., Chan P. and Cai Y. (2014) Pyrosequencing analysis of SNCA methylation levels in leukocytes from Parkinson's disease patients. *Neurosci. Lett.* **569**, 85–88.

Sontag E., Hladik C., Montgomery L., Luangpirom A., Mudrak I., Ogris E. and White Charles L. III (2004) Downregulation of protein phosphatase 2A carboxyl methylation and methyltransferase may contribute to Alzheimer disease pathogenesis. *J. Neuropathol. Exp. Neurol.* **63**, 1080–1091.

Sun W., Zang L., Shu Q. and Li X. (2014) From development to diseases: the role of 5hmC in brain. *Genomics* **104**, 347–351.

Tahiliani M., Koh K. P., Shen Y. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.

Tan Y.-Y., Wu L., Zhao Z.-B., Wang Y., Xiao Q., Liu J., Wang G., Ma J.-F. and Chen S.-D. (2014) Methylation of α-synuclein and leucine-rich repeat kinase 2 in leukocyte DNA of Parkinson's disease patients. *Parkinsonism Relat. Disord.* **20**, 308–313.

Tremolizzo L., Messina P., Conti E. *et al.* (2014) Whole-blood global DNA methylation is increased in amyotrophic lateral sclerosis independently of age of onset. *Amyotroph. Lateral Scler. Frontotemporal Degener.* **15**, 98–105.

van den Hove D. L. A., Chouliaras L. and Rutten B. P. F. (2012) The role of 5-hydroxymethylcytosine in aging and Alzheimer's disease: current status and prospects for future studies. *Curr. Alzheimer Res.* **9**, 545–549.

Varley K. E., Gertz J., Bowling K. M. *et al.* (2013) Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.* **23**, 555–567.

Villar-Menéndez I., Blanch M., Tyebji S., Pereira-Veiga T., Albasanz J. L., Martín M., Ferrer I., Pérez-Navarro E. and Barrachina M. (2013) Increased 5-methylcytosine and decreased 5-hydroxymethylcytosine levels are associated with reduced striatal A2AR levels in Huntington's disease. *NeuroMol. Med.* **15**, 295–309.

Wang S.-C., Oelze B., Schumacher A., Frommer M. and Pereira-Smith O. M. (2008) Age-specific epigenetic drift in late-onset Alzheimer's disease. *PLoS ONE* **3**, e2698.

Watson C. T., Roussos P., Garg P., Ho D. J., Azam N., Katsel P. L., Haroutunian V. and Sharp A. J. (2016) Genome-wide DNA methylation profiling in the superior temporal gyrus reveals

epigenetic signatures associated with Alzheimer's disease. *Genome Med.* **8**, 5.

Wen L. and Tang F. (2014) Genomic distribution and possible functions of DNA hydroxymethylation in the brain. *Genomics* **104**, 341–346.

Wenk G. L. (2003) Neuropathologic changes in Alzheimer's disease. *J. Clin. Psychiatry* **64**, 7–10.

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. *J. Mol. Neurosci.* **6**, 141–146.

Wu S. C. and Zhang Y. (2010) Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* **11**, 607–620.

Wu T. P., Wang T., Seetin M. G. *et al.* (2016) DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nature* **532**, 329–333.

Xi Z., Zinman L., Moreno D. *et al.* (2013) Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion. *Am. J. Hum. Genet.* **92**, 981–989.

Xi Z., Zhang M., Bruni A. C. *et al.* (2015) The C9orf72 repeat expansion itself is methylated in ALS and FTLD patients. *Acta Neuropathol.* **129**, 715–727.

Yates D. and McLoughlin D. M. (2008) The molecular pathology of Alzheimer's disease. *Psychiatry* **7**, 1–5.

Zeilinger S., Kühnel B., Klopp N. *et al.* (2013) Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS ONE* **8**, e63812.

Zhao J., Zhu Y., Yang J., Li L., Wu H., De Jager P. L., Jin P. and Bennett D. A. (2017) A genome-wide profiling of brain DNA hydroxymethylation in Alzheimer's disease. *Alzheimers Dement.* **13**, 674–688.

Ziller M. J., Gu H., Muller F. *et al.* (2013) Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477–481.

APPENDIX B: AN EPIGENOME-WIDE ASSOCIATION STUDY OF ALZHEIMER'S DISEASE BLOOD HIGHLIGHTS ROBUST DNA HYPERMETHYLATION IN THE *HOXB6* GENE

Neurobiology of Aging 95 (2020) 26–45

Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging

An epigenome-wide association study of Alzheimer's disease blood highlights robust DNA hypermethylation in the *HOXB6* gene

Janou A.Y. Roubroeks[a], Adam R. Smith[a], Rebecca G. Smith[a], Ehsan Pishva[a,b], Zina Ibrahim[c,d], Martina Sattlecker[c], Eilis J. Hannon[a], Iwona Kłoszewska[e], Patrizia Mecocci[f], Hilkka Soininen[g,h], Magda Tsolaki[i], Bruno Vellas[j], Lars-Olof Wahlund[k], Dag Aarsland[l,m], Petroula Proitsi[l,n], Angela Hodges[l], Simon Lovestone[o,p], Stephen J. Newhouse[l,n], Richard J.B. Dobson[c,d], Jonathan Mill[a], Daniël L.A. van den Hove[b,q,†], Katie Lunnon[a,†,*]

[a] College of Medicine and Health, University of Exeter, Exeter, UK
[b] School for Mental Health and Neuroscience (MHeNS), Department of Psychiatry and Neuropsychology, Maastricht University, Maastricht, the Netherlands
[c] Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience (IOPPN), King's College London, London, UK
[d] Farr Institute of Health Informatics Research, University College London, London, UK
[e] Medical University of Lodz, Lodz, Poland
[f] Institute of Gerontology and Geriatrics, University of Perugia, Perugia, Italy
[g] Institute of Clinical Medicine, Neurology, University of Eastern Finland, Kuopio, Finland
[h] Neurocenter, Neurology, Kuopio University Hospital, Kuopio, Finland
[i] 1st Department of Neurology, Memory and Dementia Unit, Aristotle University of Thessaloniki, Thessaloniki, Greece
[j] INSERM U 558, University of Toulouse, Toulouse, France
[k] NVS Department, Section for Clinical Geriatrics, Karolinska Institutet, Stockholm, Sweden
[l] King's Health Partners Centre for Neurodegeneration Research, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[m] Centre for Age-Related Diseases, Stavanger University Hospital, Stavanger, Norway
[n] NIHR Biomedical Research Centre for Mental Health at South London and Maudsley NHS Foundation Trust and Institute of Psychiatry, King's College London, London, UK
[o] Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford, UK
[p] Current Affiliation at Janssen-Cilag UK
[q] Department of Psychiatry, Psychosomatics and Psychotherapy, University of Würzburg, Würzburg, Germany

ARTICLE INFO

Article history:
Received 2 June 2020
Received in revised form 27 June 2020
Accepted 27 June 2020
Available online 3 July 2020

Keywords:
Alzheimer's disease (AD)
Biomarker
Blood
DNA methylation
HOXB6
Mild cognitive impairment (MCI)

ABSTRACT

A growing number of epigenome-wide association studies have demonstrated a role for DNA methylation in the brain in Alzheimer's disease. With the aim of exploring peripheral biomarker potential, we have examined DNA methylation patterns in whole blood collected from 284 individuals in the AddNeuroMed study, which included 89 nondemented controls, 86 patients with Alzheimer's disease, and 109 individuals with mild cognitive impairment, including 38 individuals who progressed to Alzheimer's disease within 1 year. We identified significant differentially methylated regions, including 12 adjacent hypermethylated probes in the *HOXB6* gene in Alzheimer's disease, which we validated using pyrosequencing. Using weighted gene correlation network analysis, we identified comethylated modules of genes that were associated with key variables such as *APOE* genotype and diagnosis. In summary, this study represents the first large-scale epigenome-wide association study of Alzheimer's disease and mild cognitive impairment using blood. We highlight the differences in various loci and pathways in early disease, suggesting that these patterns relate to cognitive decline at an early stage.

Declarations of interest: None.

* Corresponding author at: University of Exeter Medical School, RILD Building Level 4, Royal Devon and Exeter Hospital, Barrack Rd, Exeter, EX2 5DW. UK. Tel.: 01392 406758.
E-mail address: k.lunnon@exeter.ac.uk (K. Lunnon).
† These authors contributed equally to this work.

0197-4580/
https://doi.org/10.1016/j.neurobiolaging.2020.06.023

1. Introduction

With an increasingly aging population the prevalence of dementia is expected to almost double in the coming 20 years, with Alzheimer's disease (AD) being the greatest contributor. AD presents itself as a heterogeneous, multifaceted disease, and this complexity is reflected in the challenges researchers face in elucidating the exact mechanisms underlying this disorder. A number of genome-wide association studies (GWAS) have identified susceptibility loci associated with the more common, sporadic form of AD (Lambert et al., 2013). However, these do not account fully for disease risk, and the exact processes involved in the development and progression of this neurodegenerative disorder remain unknown.

A growing number of studies have investigated the role of epigenetic mechanisms in the etiology and progression of AD. Epigenetic mechanisms refer to the reversible regulation of gene expression that occurs independently of the underlying DNA sequence. One such mechanism is DNA methylation, which involves the addition of a methyl group to an unmodified base, most commonly cytosine (yielding 5-methylcytosine: 5mC), and plays a critical role in the regulation of gene expression (Maunakea et al., 2013; Varley et al., 2013; Ziller et al., 2013). Recently, numerous epigenome-wide association studies (EWAS) have explored DNA methylomic variation in postmortem human brain tissue from AD patients and elderly controls and have highlighted a number of loci that show robust differences in DNA methylation in the cortex across independent cohorts (Altuna et al., 2019; De Jager et al., 2014; Gasparoni et al., 2018; Lardenoije et al., 2019; Lunnon et al., 2014; Smith et al., 2018, 2019, 2020; Watson et al., 2016). There is communication between the brain and the blood, especially in disease. In recent years, several studies have identified transcriptomic (Booij et al., 2011; Fehlbaum-Beurdeley et al., 2012; Lunnon et al., 2012, 2013, 2017; Rye et al., 2011) or proteomic (Hye et al., 2006; O'Bryant et al., 2010, 2011, 2016) alterations in the blood early in the disease and these signatures have been used for identifying novel dysfunctional pathways and biomarkers in the blood. Although valuable, the use of systemic gene expression or protein markers for this purpose still yields some pitfalls due to the dynamic nature of gene and protein expression. For example, sampling methods can significantly alter the expression levels by inducing ex vivo mRNA expression (Asare et al., 2008; Thach et al., 2003). Similarly, differences in processing methods between researchers (e.g., handling methods, sample processing methods) can affect the quality of mRNA and protein and impinge on downstream analyses (Vartanian et al., 2009; Zhao et al., 2012). DNA methylation levels are reported to be more stable than mRNA levels (Paziewska et al., 2014) and as such studying this in AD blood could be more informative of important biological pathways specifically altered in disease. To date, most blood DNA methylation studies have focused on specifically investigating candidate genes (da Silva et al., 2014; Furuya et al., 2012b,a; Wang et al., 2008). Four EWAS of AD blood have been published so far, which have identified a number of disease-associated loci. However, these studies used a limited set of (nondemented) samples and/or did not include any individuals with mild cognitive impairment (MCI) (Kobayashi et al., 2016; Lardenoije et al., 2019; Lunnon et al., 2014; Madrid et al., 2018).

To understand changes in the blood related to the development and progression of AD, it is important to include MCI individuals in addition to AD patients and controls. Often viewed as an early stage of AD, MCI is characterized by memory and other cognitive complaints and impairment, although these have no significant impact at this stage on daily living, as is seen in AD (Petersen et al., 1999). Although individuals with MCI may remain stable over time or develop another neurodegenerative disorder, MCI subjects, particularly those with amnestic MCI (aMCI), are at an increased risk of progressing to AD (Jicha et al., 2006). In these cases, the early clinical symptoms reflect the underlying pathological changes related to AD that occur years before the disease fully manifests (Hardy, 1997; Jack et al., 2010). Previous research has shown that disease-related changes in gene expression can be detected in peripheral blood from individuals with MCI and AD, with results indicating that some peripheral differences in AD can be detected in MCI subjects (Lunnon et al., 2012). Taken together, the identification of blood methylation patterns related to MCI and AD is of great interest, as it may increase our understanding of peripheral, as well as central changes that occur early in the disease.

In the current study, we have generated genome-wide DNA methylation data from well-characterized control, MCI and AD subjects with detailed demographic, clinical, neuroimaging, and transcriptomic data previously collected. We have used this dataset to identify differentially methylated loci and epigenetic differences in specific biological processes in blood, which are associated with disease status, or future progression from MCI to AD.

2. Methods

2.1. Subjects

We analyzed a subset of 284 blood samples selected for our study from the larger AddNeuroMed cohort on the basis of additional phenotypic information being available, including genomic (Furney et al., 2011b), transcriptomic (Lunnon et al., 2012, 2013), and magnetic resonance imaging (MRI) data (Furney et al., 2011a,b; Westman et al., 2011). The cross-European AddNeuroMed study is aimed at the identification of biomarkers for AD (Lovestone et al., 2007, 2009) and consists of 6 sites across Europe (Kuopio, Finland; Łódź, Poland; London, United Kingdom; Perugia, Italy; Thessaloniki, Greece; Toulouse, France). Informed consent was obtained from each participant according to the Declaration of Helsinki, and ethical approval was obtained at each site. All sites followed standardized procedures.

Within our subset, the subjects were classified into one of 3 groups according to their status at sample collection: AD ($n = 86$), MCI ($n = 109$), or elderly control (CTL; $n = 89$). The diagnosis of AD was made according to the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria (McKhann et al., 1984), and the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (del Barrio, 2004). Subjects in the MCI group were mainly recruited from memory clinics and scored 0.5 on the total Clinical Dementia Rating Scale (CDR) or 0.5 or 1 on the memory category of the CDR (Morris, 1993). All MCI individuals reported memory problems but showed no significant impairment in daily living according to Petersen's criteria of MCI (Petersen et al., 1999). Further details are provided elsewhere (Liu et al., 2011). A subset of MCI subjects progressed to AD within 1 year of the baseline measurement (MCI-AD, $n = 38$), while others remained stable (MCI-MCI, $n = 67$). A subset of 4 MCI subjects converted to AD at an unknown time after baseline collection and were excluded from any conversion analysis. Elderly CTLs were defined as showing no signs of cognitive impairment. Subjects were excluded from this study if they had any other significant psychiatric or neurological illness, were younger than 65 years of age, or were not white Caucasian. MRI data were collected for 213 individuals as described previously (Simmons et al., 2011). To obtain demographic information and medical data, semistructured interviews were carried out for all subjects. A number of neuropsychological assessments were also performed, including the mini-mental state examination (MMSE) (Folstein et al., 1975). An overview of individuals included in this study can be found in Table 1.

Table 1
Cohort demographics

	Control	MCI		AD
		MCI-MCI	MCI-AD	
N	89	67	42[a]	86
Gender (M/F)	34/55	34/33	16/26	30/56
Age (mean ± SD)	73.8 ± 5.3	75.1 ± 5.6	76.3 ± 5.3	76.8 ± 5.6
MMSE (mean ± SD)	29 ± 1.2	27.3 ± 1.7	26.3 ± 2.2	20.8 ± 4.5
Center (N)				
Kuopio (Finland)	22	17	16	27
Łódź (Poland)	13	5	1	8
London (United Kingdom)	22	6	3	7
Perugia (Italy)	19	19	14	23
Thessaloniki (Greece)	4	16	5	16
Toulouse (France)	9	4	3	5

Subject characteristics of the 284 samples that passed QC. Shown are sample numbers (N), sex (males [M]/females [F]), mean age (± standard deviation [SD]), mean mini-mental state examination (MMSE) ± SD, and sample numbers per source (Center). Of the 109 mild cognitive impairment (MCI) subjects, 67 remained MCI-stable (MCI-MCI) over the 1 year after sample collections, while 42 converted to AD (MCI-AD), with 38 converting within 1 year of sample collections.

[a] Four MCI-AD subjects were excluded from the analysis of MCI to AD conversion (but included in the ANOVA analysis of baseline diagnosis), as the exact time of conversion was not known. All individuals used in this study were white Caucasian.

2.2. DNA methylation analysis

DNA was extracted from the blood samples collected at baseline as described by Furney et al. (2011). The DNA was tested for degradation and purity. 500 ng DNA from each sample was sodium bisulfite–treated using the Zymo EZ-96 DNA methylation kit (Zymo Research, CA, USA) according to the manufacturer's standard protocol. Samples were assessed using the Illumina Infinium Human Methylation 450K BeadChip array (450K array; Illumina, CA, USA) using an Illumina HiScan System (Illumina, CA, USA). All samples were assigned a unique code for the purpose of the experiment and randomized with respect to sex, center, and disease status to avoid batch effects, and processed in batches of 4 BeadChips.

Raw intensity data files were imported into the R statistical environment (version 3.5.2) (R Core Team, 2018) using the *wateRmelon* (Pidsley et al., 2013) package as a methylumi object. Data quality control (QC) and preprocessing was carried out using the packages *wateRmelon* and *minfi* (Aryee et al., 2014). Initial QC checks on the data included labeling checks via sex and genetical identity, and the removal of cross-hybridizing probes, probes located on the sex chromosomes, and probes containing a single nucleotide polymorphism in the probe sequence or within 10 bp (Chen et al., 2013; Price et al., 2013). The *p-filter* function was applied, followed by the *outlyx* function within the *wateRmelon* package, with none of the 284 samples identified as outliers. Quantile normalization was then carried out using the dasen function within wateRmelon, with 401,266 probes taken forward for analysis. Blood cell type proportions were calculated using the Houseman reference–based method (Houseman et al., 2012). DNA methylation data can be found on GEO under the accession number GSE144858.

Before we identified differentially methylated positions (DMPs) associated with diagnosis, we first regressed out the effect of specific covariates that correlated with the first 3 principal components of the normalized data (Supplementary Fig. 1), with these variables being age, sex, blood cell type proportion (CD4 and CD8 T lymphocytes, natural killer cells, B cells, monocytes, granulocytes), and bisulfite conversion batch. An analysis of variance (ANOVA) was then performed on the residuals from the linear regression, to test for DNA methylation differences across all 3 groups. A post-hoc Tukey's honest significant difference (HSD) test (Tukey, 1949) was applied to the results to compare methylation levels between each of the 3 diagnostic groups at baseline (CTL, MCI, AD). For the purpose of these analyses, the MCI group included both MCI-MCI and MCI-AD as both groups were classified as MCI at baseline, which was the time point when blood sampling occurred. To identify differentially methylated regions (DMRs), which represent areas of the DNA containing multiple adjacent DMPs, we used the DMPs from both the ANOVA and the individual between group Tukey's tests and applied the *comb-p* module (Pedersen et al., 2012) in Python (version 2.7.5) (Rossum and Boer, 1991), assessing regions of 1000 base pairs, with a *p*-value threshold of $p < 0.01$. We selected only regions containing ≥2 probes, and that had a multiple testing-corrected $p < 0.05$, which was corrected using the Šidák method (Šidák, 1967). To identify DMPs and DMRs relating to the future conversion from MCI to AD we performed an analysis comparing the MCI-MCI subjects and the MCI-AD subjects, by first regressing out age, sex, blood cell type proportion, batch, and baseline MMSE score. Baseline MMSE score was included as a covariate as we observed a small, but significant difference in baseline MMSE between the MCI-MCI and MCI-AD groups. We then used a linear regression to compare the 2 groups and performed *comb-p* analysis as described above. Quantile-quantile (QQ)-plots of the *p*-values from both the ANOVA and linear regression can be found in Supplementary Fig. 2.

2.3. Generation of weighted gene correlation networks

In order to identify clusters, or "modules", of highly comethylated sites in the genome, we made use of the R package for weighted gene correlation network analysis (*WGCNA*) (Langfelder and Horvath, 2008). The hypothesis underlying this method is that genes that highly co-vary, share the same underlying biological processes. Prior to creating the modules, all nonvariable probes (variance < median variance) were first removed from the normalized data, leaving 200,633 probes for analysis. Samples were then clustered based on their Euclidean distance, and clustering dendrograms were visually inspected to identify outlier samples, which were not detected. Network construction and module detection was then performed in a block-wise manner and constructed irrespective of the direction of correlation between probes (unsigned). The connection strength between 2 probes was weighted using a soft threshold value of 9 in the baseline group analysis and 8 in the comparison of MCI converters to MCI non-converters, which emphasizes high correlations over low correlations. The soft threshold values were selected using the *pickSoftThreshold* function within the *WGCNA* package. In the resulting modules, each module is identified by an arbitrarily assigned color, and the gray module is disregarded from further analyses as it contains unassigned probes. Module eigengenes

(MEs) were calculated for each module, as the first principal component across probes assigned to each module. The ME is a single value for each sample and represents the shared methylation profile of the module. Modules were generated twice: once for the baseline group analysis, which compared CTL to MCI to AD, and once for the conversion analysis using only the subset of MCI-MCI and MCI-AD samples.

2.4. Association of modules to traits of interest

Covariates (age, sex, blood cell type proportions, and batch number) were regressed out from the MEs, and extreme outliers (exceeding >5 standard deviations) were removed. Modules were then associated with baseline diagnosis groups and traits of interest by performing pairwise Pearson or Spearman correlations for continuous or ordinal variables, respectively. Correlations were performed using dummy variables of baseline diagnosis categories to investigate all permutations of comparisons (i.e., CTL versus MCI, CTL versus AD, and MCI versus AD), with the group not used in each comparison set to NA. Additional traits of interest included number of education years, number of *APOE ε4* alleles, MMSE score, and the following structural MRI measurements: left, right, and total entorhinal cortex volume (LEV, REV, and TEV, respectively), left, right, and total hippocampal volume (LHV, RHV, and THV, respectively), ventricular volume (VV), and whole brain volume (WBV). Similarly, regression of the same covariates (with the addition of baseline MMSE score) and outlier removal was also performed for MEs generated from the MCI-MCI and MCI-AD samples. The residuals from this regression were then used to run a linear regression, comparing nonconverters to converters.

2.5. Module membership and probe significance

For each of the modules showing significant ($p < 0.05$) associations with one of the 3 baseline groups, conversion to AD, or traits of interest, we calculated the module membership (MM) and probe significance (PS). MM was calculated as the Pearson correlation between the methylation value of each probe and the ME values, representing the strength of association between a probe and the module it belongs to. PS represents the strength of the correlation between a probe's methylation value and the diagnosis or trait of interest, as performed by Pearson correlations for continuous traits, and Spearman correlations for ordinal traits or diagnostic groups. We correlated and plotted MM to PS for modules of interest and focused on those that showed significant positive correlations (i.e., $r > 0$, $p < 0.05$), which would indicate that probes more integral to the module are mainly driving the association with the trait of interest. Underlying biological processes and pathways were then examined for the modules selected, using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. For modules containing a large number of probes (>10,000), we performed these pathway analyses on the probes that were central to the module (i.e., core probes). We set this threshold at 15%, thus selecting the top 15% of probes with the highest MM. Analyses were performed using the *missMethyl* package (Phipson et al., 2015), taking into account the differing number of probes covering each gene on the array.

2.6. Analysis of gene expression data and association with methylation data

Normalized gene expression data from Illumina Human HT-12 v3 Expression BeadChip arrays (HT-12 arrays) was obtained from a previous study by Lunnon et al. (2012), for 237 individuals included in the current study. Expression data for all genes containing DMRs identified in this study were extracted for analysis if available. For genes nominated from the diagnostic category analysis (CTL, MCI, AD), the covariates of age, sex, and cell type proportions (estimated using Houseman's reference–based method) were regressed out of the expression data and an ANOVA and subsequent Tukey's HSD were then performed on the residuals of the regression to identify diagnostic category differences in expression levels of genes containing DMRs. For genes that contained DMRs associated with the progression to AD, only the MCI-MCI and MCI-AD samples were analyzed, with age, sex, cell type proportions, and baseline MMSE score regressed out of the expression data, with a subsequent linear regression analysis performed to assess gene expression differences between MCI-MCI and MCI-AD individuals in DMR genes.

Next, methylation values within a DMR were correlated to gene expression values of an annotated gene. Methylation values that had previously been corrected for covariates (i.e., residuals) were extracted for 450K array probes located within each DMR based on genomic location of the DMRs. Pairwise Pearson correlations were then performed between the covariate-adjusted gene expression levels and covariate-adjusted methylation values, for individual 450K probes within a DMR. We also performed correlations of gene expression and mean methylation levels from all 450K probes in the DMR. To determine whether the association between gene expression and methylation differed between CTL and individuals with MCI or AD, ANOVAs were performed on gene expression levels which included an interaction term between methylation and baseline diagnostic groups (i.e., expression ~ methylation*group). This was performed on the probe most significantly associated with the disease for each DMR, and the mean methylation value in the DMR. Similar analyses were performed on the MCI-MCI and MCI-AD subset of individuals, for DMRs associated with progression to AD.

2.7. Validation of the HOXB6 differentially methylated region using pyrosequencing

For the purpose of validating our findings, we designed a pyrosequencing assay to quantify DNA methylation at the most significant sites (cg17179862 and cg03803541) within the *HOXB6* region (chr17:46681111–46682414), which was shown to be a DMR in AD relative to CTL. Pyrosequencing assays were designed with the PyroMark Assay Design software 2.0 (Qiagen). In addition to the 2 CpG sites the assay was designed for, further 3 CpG sites that were not assessed on the 450K array were also covered. Out of the original 284 samples, 264 were used for pyrosequencing. Samples were semi-randomly selected, keeping the group sample number ratios as equal as possible, and randomly distributing samples across plates. A single amplicon of 303 base pairs was amplified using designed primers, and tested for specificity (forward primer = TTTTTGGTGAGGGGGGAGT, reverse primer = CCTACCATCCCTCCCTTATCT, sequencing primer = CTCTAACTATTACCCC). The level of DNA methylation was then quantified using the Pyromark Q24 system (Qiagen), following the standard protocol as provided by the manufacturer and the Pyro Q24 CpG 2.0.6.20 software.

Pyrosequencing data QC was performed using the Pyromark Q24 software, in addition to a visual inspection of the data and signal intensities, with all 264 samples passing QC (CTL: n = 83, MCI: n = 102, AD: n = 79). DNA methylation percentages at specific CpG sites were calculated by the software and exported to the R statistical environment. Subsequently, an ANOVA was performed for each CpG site covered by the assay, as well as the average methylation value across the region. This analysis was identical to the analysis performed on the 450K data, and the covariates of age, sex, cell type proportion, and batch were included.

3. Results

3.1. Identification of differentially methylated loci in mild cognitive impairment and Alzheimer's disease blood

The cohort characteristics are shown in Table 1. We first investigated whether any individual loci showed DNA methylation differences in either MCI or AD relative to CTL using an ANOVA model after adjusting for the covariates of age, sex, cell proportions, and batch (Supplementary Table 1). No DMPs reached the experiment-wide significance threshold that has been established for the 450K array (2.4×10^{-7}) (Saffari et al., 2018) with the smallest ANOVA *p*-value being 5.58×10^{-6} for probe cg26146855, of which the closest transcription start site is located in the *TFAMP1* gene. The top 1000 most significant probes resulting from the post-hoc Tukey's HSD tests comparing CTL to MCI, MCI to AD, and CTL to AD can be found in Supplementary Tables 2, 3, and 4, respectively. In addition to comparing methylation levels at baseline between the 3 groups, we were also interested in identifying differences within the MCI population that were predictive of later progression to AD. For this purpose, we compared the MCI-MCI group to the MCI-AD group. While no DMPs passed the experiment-wide significance threshold, the most significant DMP was located in the *TRIM62* gene and showed hypomethylation in converters (probe cg25342005, $p = 1.67 \times 10^{-6}$; Supplementary Table 5).

3.2. A number of significant differentially methylated regions can be identified in mild cognitive impairment and Alzheimer's disease blood

We next used a sliding window approach to identify the regions spanning multiple adjacent DMPs that were significantly different in MCI and AD. We found 4 DMRs associated with differences across the 3 baseline groups (CTL, MCI, and AD) (Table 2A). A 10-probe DMR of 574 bp was identified in *MOV10L1* (Fig. 1A), as well as a 5-probe (582 bp) intergenic DMR annotated to *CBFA2T3* (Fig. 1B), with probes in both DMRs generally showing hypermethylation in MCI samples, with levels in AD samples similar to CTL. An 8-probe DMR of 301 bp was found in the readthrough transcription region of *TPTEP2-CSNK1E*, which appeared to be mainly driven by hypermethylation in the MCI group (Fig. 1C).

One of the 4 identified DMRs was driven by a difference between the CTL and AD groups (Table 2B); we identified a 1303 bp DMR in the *HOXB6* gene, containing 12 probes (Figs. 1D and 2). Each of the 12 probes showed hypermethylation in AD.

In our analysis of MCI conversion to AD, we identified 9 significant DMRs (Table 2C; Fig. 3). We found DMRs showing decreased methylation in MCI-AD converters relative to MCI-MCI nonconverters in the genes *CPT1B* and *CHKB* (932 bp; 14 probes) (Fig. 3A), *TMEM184A* (659 bp; 6 probes) (Fig. 3B), *KCNAB3* (558 bp; 7 probes) (Fig. 3C), *GABBR1* (379 bp; 10 probes) (Fig. 3D), *PRDM1* (121 bp; 5 probes) (Fig. 3E), *FLJ37453* (568 bp; 6 probes) (Fig. 3F), and *OR56A3* and *TRIM5* (556 bp; 5 probes) (Fig. 3G). Hypermethylation in MCI-AD converters relative to MCI-MCI nonconverters was seen in 2 DMRs located in the genes *SMC1B* and *RIBC2* (725 bp; 15 probes) (Fig. 3H), and an intergenic region near the gene *FIGN* (716 bp; 6 probes) (Fig. 3I).

3.3. Validation of the Alzheimer's disease–associated differentially methylated region in HOXB6 by pyrosequencing

Interestingly, differential DNA methylation at the most significant locus within the *HOXB6* DMR (cg17179862) has been previously reported in AD hippocampus (Altuna et al., 2019). To further explore AD-associated hypermethylation in this gene, we used pyrosequencing to validate our *HOXB6* DMR, covering 2 CpG sites

Table 2
Differentially methylated regions in blood

Gene	Position	Gene feature	*n*	*p*-value	Šidák-*P*	Average methylation %		
						CTL	MCI	AD
A. ANOVA: CTL versus MCI versus AD								
HOXB-AS3; HOXB6	chr 17: 46681111 - 46682414	nc_intron+nc_exon; TSS+intron+exon+utr5	12	2.79E-14	8.58E-12	56.59	58.79	60.81
MOV10L1	chr 22: 50528179 - 50528753	TSS+intron+utr5+cds; TSS+exon+utr5	10	2.03E-07	1.42E-04	68.38	70.18	68.23
CBFA2T3	chr 16: 88937216 - 88937798	Intergenic	5	2.61E-07	1.80E-04	42.57	44.66	42.34
TPTEP2-CSNK1E	chr 22: 38714166 - 38714467	intron+utr5	8	1.87E-06	2.49E-03	41.21	42.29	41.69
B. CTL vs. AD								
HOXB-AS3; HOXB6	chr 17: 46681111 - 46682414	nc_intron+nc_exon; TSS+intron+exon+utr5	12	3.36E-16	1.03E-13	56.59	58.79	60.81
Gene	Position	Gene feature	*n*	*p*-value	Šidák-*P*	Average methylation %		
						MCI-MCI	MCI-AD	
C. MCI-MCI vs. MCI-AD								
CHKB-CPT1B; CPT1B; CHKB	chr 22: 51016501 - 51017433	nc_intron; TSS+intron+exon+utr5; exon+utr3	14	2.05E-14	8.84E-12	64.92	61.26	
SMC1B; RIBC2	chr 22: 45809319 - 45810044	TSS+intron+utr5+cds; TSS+intron+utr5+cds	15	8.26E-09	4.57E-06	24.62	26.62	
TMEM184A	chr 7: 1595602 - 1596261	TSS+intron+exon+utr5	6	2.41E-08	1.47E-05	45.33	43.31	
KCNAB3	chr 17: 7832680 - 7833238	TSS+cds	7	8.11E-08	5.83E-05	80.31	76.73	
GABBR1	chr 6: 29599012 - 29599391	intron+exon+utr5; intron+cds	10	9.72E-08	1.03E-04	63.80	61.48	
FIGN	chr 2: 164204628 - 164205344	Intergenic	6	3.58E-07	2.01E-04	52.84	56.10	
PRDM1	chr 6: 106546704 - 106546825	TSS+exon+utr5; intron	5	1.04E-07	3.45E-04	62.01	58.89	
FLJ37453	chr 1: 16163555 - 16164123	nc_intron	6	5.92E-07	4.18E-04	29.67	27.34	
OR56A3; TRIM5	chr 11: 5959658 - 5960214	Intergenic	5	9.68E-07	6.98E-04	81.17	77.54	

Differentially methylated regions (DMRs) in a comparison of control (CTL), mild cognitive impairment (MCI), and Alzheimer's disease (AD) blood samples. Shown are DMRs for (A) the overall three group (ANOVA) comparison, the post-hoc (B) CTL versus AD comparison, and (C) the MCI-stable (MCI-MCI) versus MCI-converter (MCI-AD) comparison. Displayed for each region is the UCSC gene name, chromosomal position (genome build 37), gene feature (TSS = transcription start site; utr5 = 5′ untranslated region; utr3 = 3′ untranslated region; cds = coding sequence), number of probes in region (n), *p*-value and multiple testing-corrected *p*-value (Šidák-P), and average beta per group.

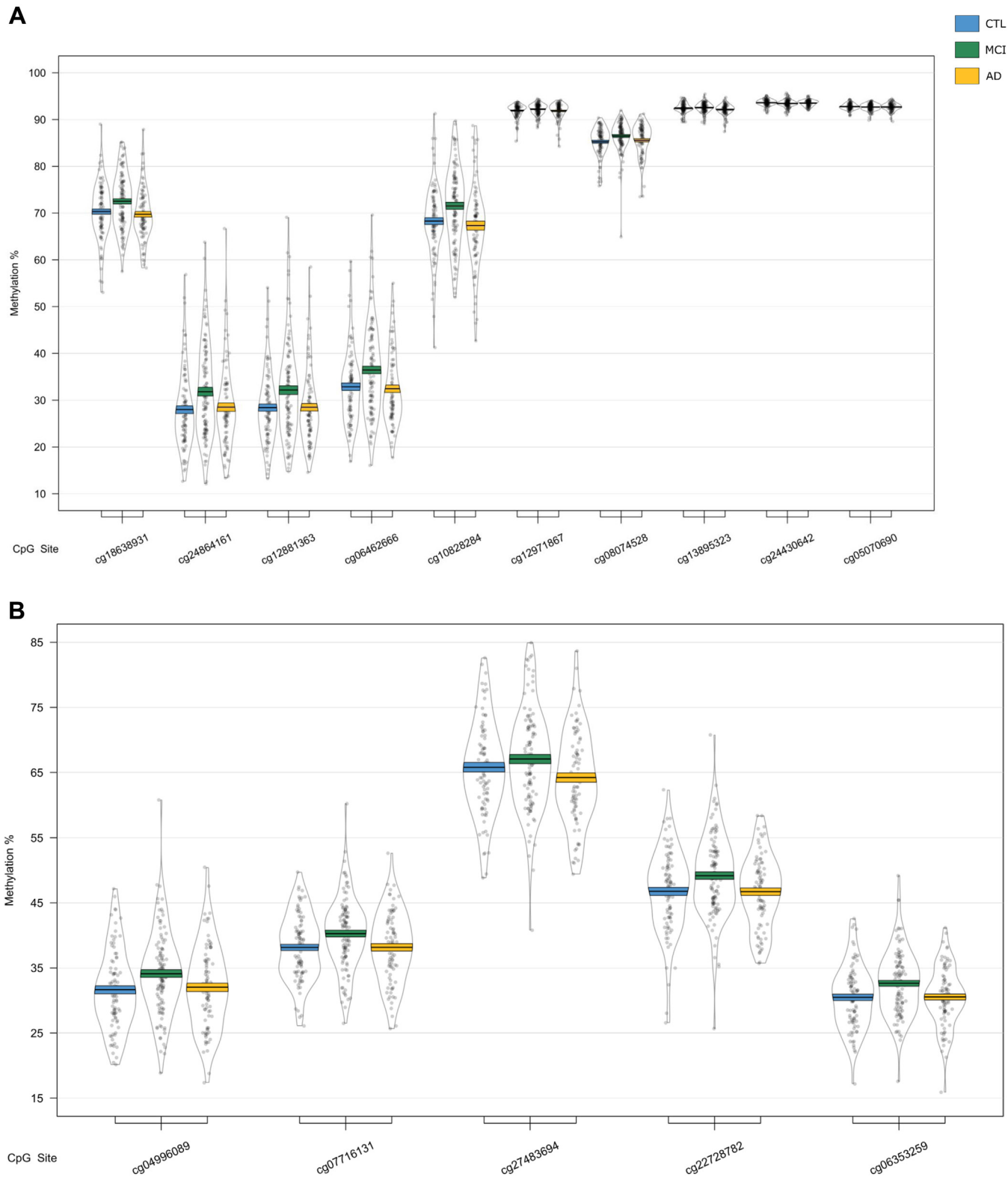

Fig. 1. Differentially methylated regions (DMRs) in a comparison of baseline diagnosis of mild cognitive impairment and Alzheimer's disease relative to controls. DMRs shown are located in or near the genes *MOV10L1* (A), *CBFA2T3* (B), *TPTEP2-CSNK1E* (C), and *HOXB6* (D). Displayed for each DMR are the methylation levels of individual probes located within the DMR, ordered by genomic location. Methylation values have been corrected for covariates age, sex, cell type proportion, and batch.

on the array (cg17179862, cg03803541) as well as 3 neighboring CpG sites that were not covered by the 450K array (chr17:46681421, chr17:46681394, and chr17:46681383). We found significant differences between groups at all 5 CpG sites (Supplementary Table 6, Fig. 4A), and when averaged over the full 5 probes (Fig. 4B), demonstrating hypermethylation in AD samples relative

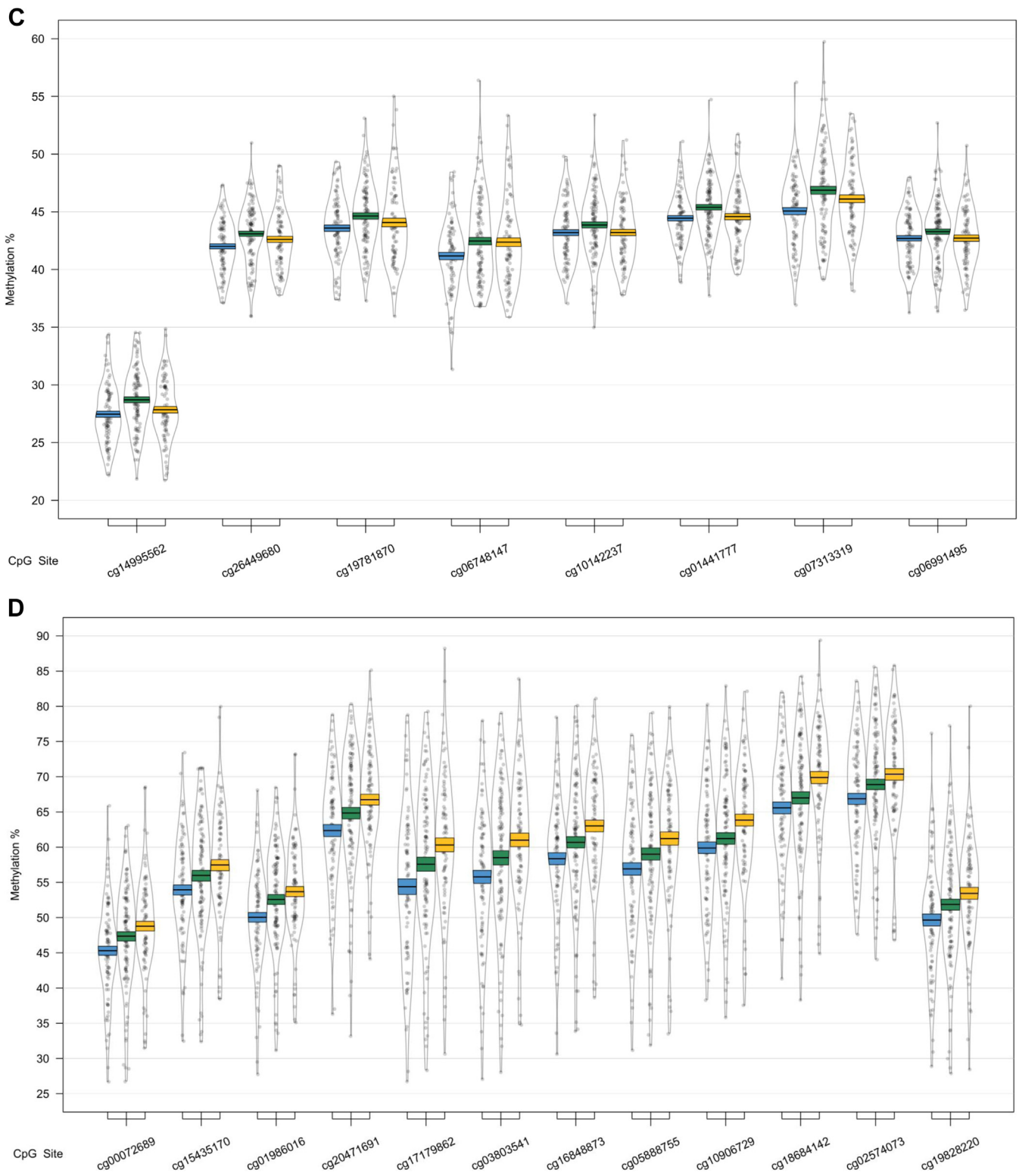

Fig. 1. (*continued*).

to controls. The pattern of DNA methylation quantified by the 450K array and pyrosequencing was similar for both cg03803541 (Fig. 4C) and cg17179862 (Fig. 4D), with a significant correlation of the methylation values estimated by the 2 technologies for both cg03803541 (Fig. 4E: r = 0.957, $p = 2.69 \times 10^{-142}$) and cg17179862 (Fig. 4F: r = 0.934, $p = 5.03 \times 10^{-68}$).

3.4. Transcriptional differences in genes containing differentially methylated regions

To explore the relationship between DNA methylation and expression, we first assessed whether the expression levels of genes containing the 4 baseline diagnosis-associated DMRs or the 9

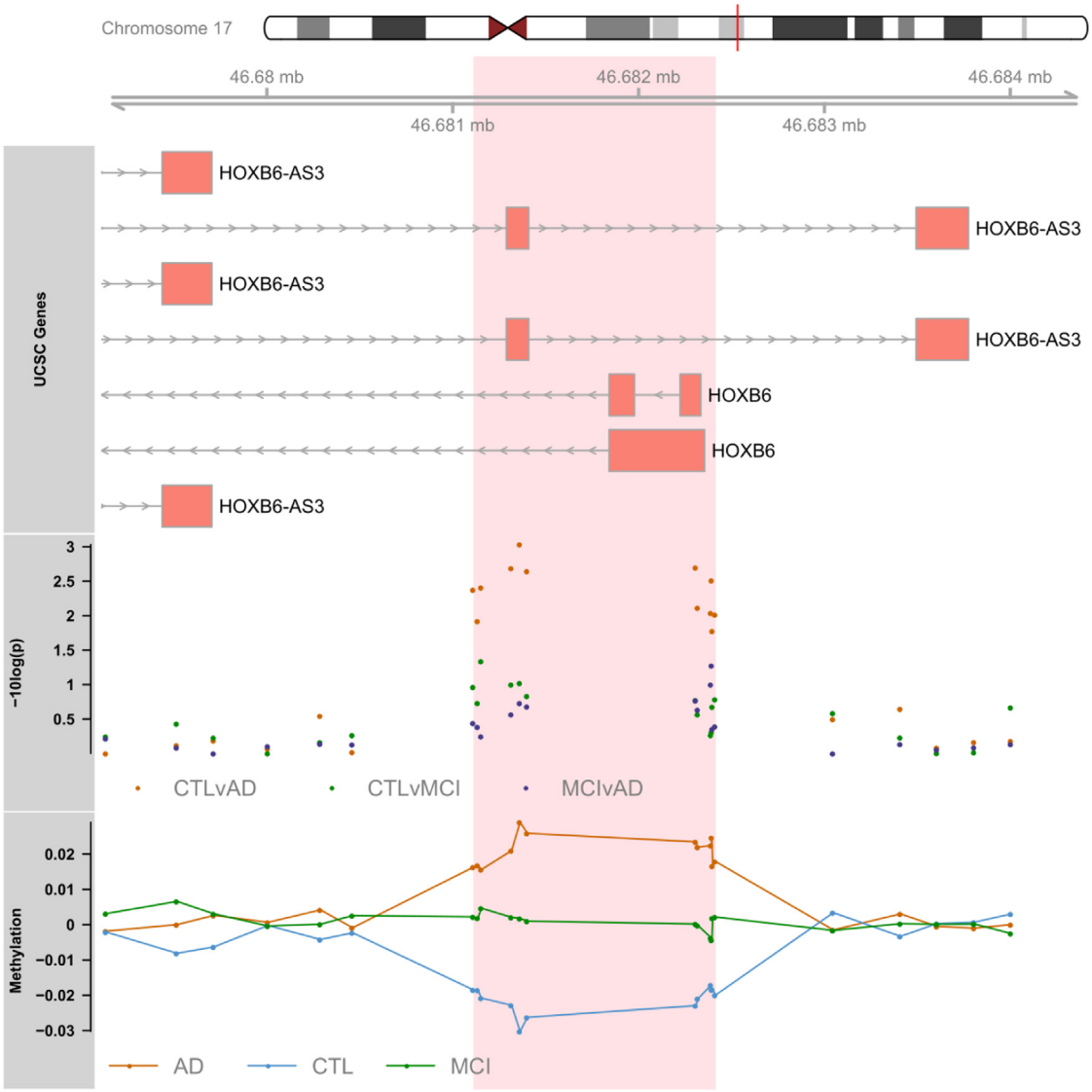

Fig. 2. The *HOXB6* DMRs, shown to be altered in Alzheimer's disease (AD) relative to controls (CTL). The region spanned by the DMRs is highlighted in red, and genomic location and UCSC gene annotations are shown, in addition to a mini-Manhattan plot of the *p*-values of probes within and neighboring the DMR, *p*-values related to changes in AD relative to CTL are shown in orange, mild cognitive impairment (MCI) relative to CTL in green, and AD relative to MCI in purple. The bottom panel shows relative methylation levels across the region, with methylation in AD in orange, MCI in green, and CTL in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

conversion DMRs were different in the diagnostic groups. Expression data were only available for 2 of the 6 genes we identified as containing DMRs in the baseline group analysis (*HOXB6,* and *CSNK1E* associated with the readthrough transcription region of *TPTEP2-CSNK1E*) (Supplementary Table 7). Of these 2 genes, *CSNK1E,* which had shown increased DNA methylation in MCI samples, also showed a significant difference in gene expression between groups ($F = 15.94$, $p = 3.25 \times 10^{-7}$). More specifically, we observed significantly increased mRNA expression in both MCI and AD subjects relative to control (Tukey's $p = 1.46 \times 10^{-7}$ and $p = 0.003$, respectively, Supplementary Fig. 3A). Although there was significantly higher gene expression and DNA methylation (across the DMR), there was no correlation of expression and methylation across all samples, or when we performed correlations separately in the 3 diagnostic groups (Supplementary Fig. 3C, Supplementary Table 8). Although we did not observe any significant differences in gene expression for *HOXB6* (Supplementary Fig. 4A), we did find a correlation of expression and methylation when performing correlations in the AD group only ($r = -0.24$, $p = 0.041$) (Supplementary Fig. 4C, Supplementary Table 8).

Expression data were also available for 5 of the 9 significant DMRs we identified in our analysis of progression from MCI to AD (*GABBR1, PRDM1, FLJ37453, TRIM5,* and *CPT1B/CHKB*). The *CPT1B/CHKB* DMR was covered by 3 probes on the gene expression microarray, one probe measuring *CPT1B* expression and 2 probes measuring *CHKB* expression (ILMN_2331205 and ILMN_1659054). Although none of these genes showed differential expression in MCI subjects who progressed to AD (Supplementary Table 9), *CPT1B/CHKB* showed a significant positive correlation of methylation across the DMR and *CPT1B* gene expression (Supplementary Table 10). The average methylation level across the *CPT1B/CHKB* DMR was significantly correlated with gene expression across all samples ($r = 0.40$, $p = 8.62 \times 10^{-5}$, Supplementary Fig. 5), which appeared to be primarily driven by a correlation observed in the MCI-MCI samples ($r = 0.49$, $p = 7.27 \times 10^{-5}$) and not the MCI-AD samples.

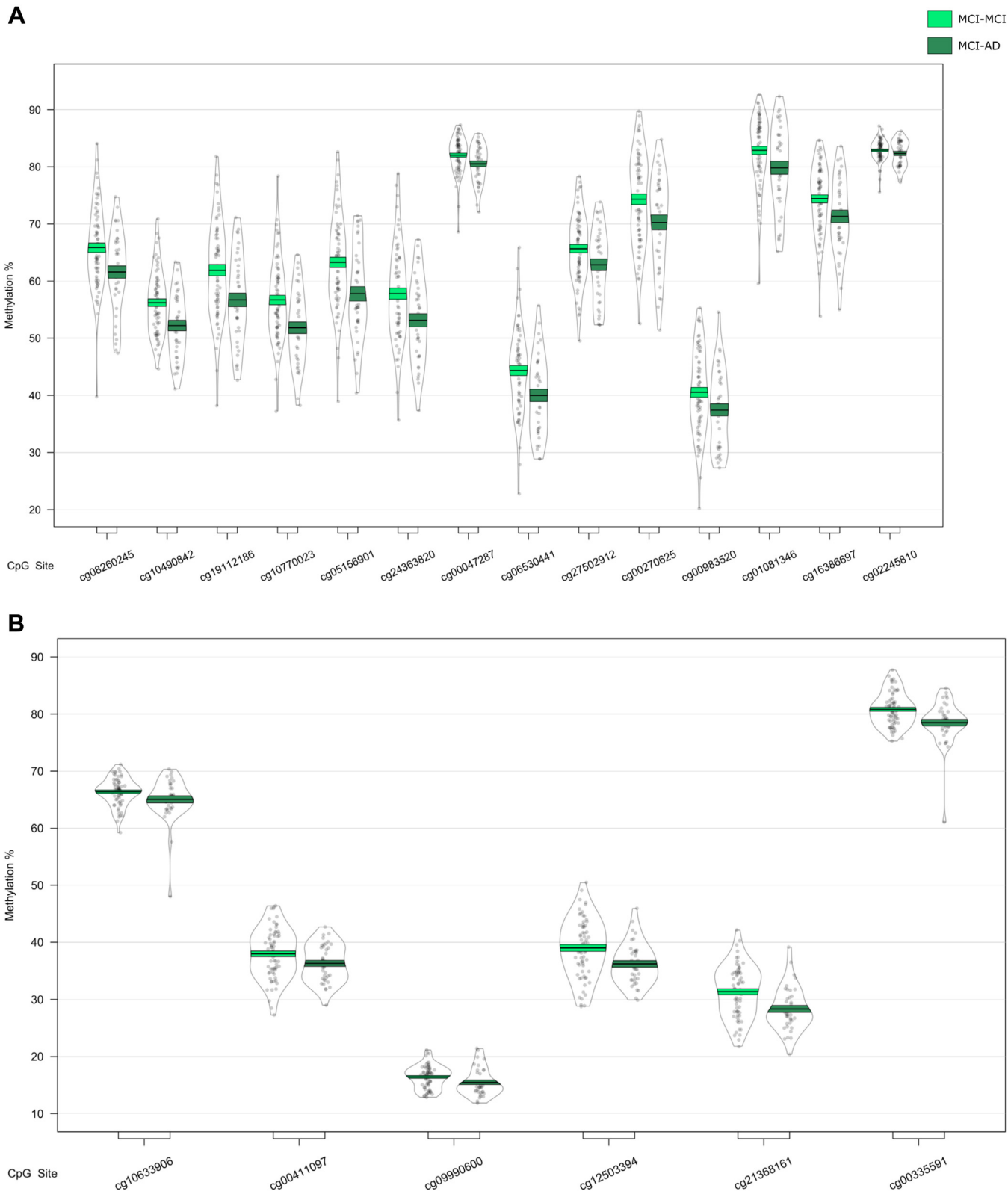

Fig. 3. Differentially methylated regions (DMRs) in a comparison of mild cognitive impairment (MCI) individuals who converted to Alzheimer's disease (AD) within 1 year after baseline assessment (MCI-AD; dark green, shown on the right) and those who remained stable (MCI-MCI; light green, shown on the left). DMRs shown are located in or near the genes *CPT1B* (A), *TMEM184 A* (B), *KCNAB3* (C), *GABBR1* (D), *PRDM1* (E), *FLJ37453* (F), *OR56A3* and *TRIM5* (G), *SMC1B* and *RIBC2* (H), and *FIGN* (I). Displayed for each DMR are the methylation levels of all probes ($p < 0.05$) within the genomic location covered by each DMR, ordered by genomic location. Methylation values have been corrected for covariates age, sex, cell type proportion, batch, and baseline mini-mental state examination score. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

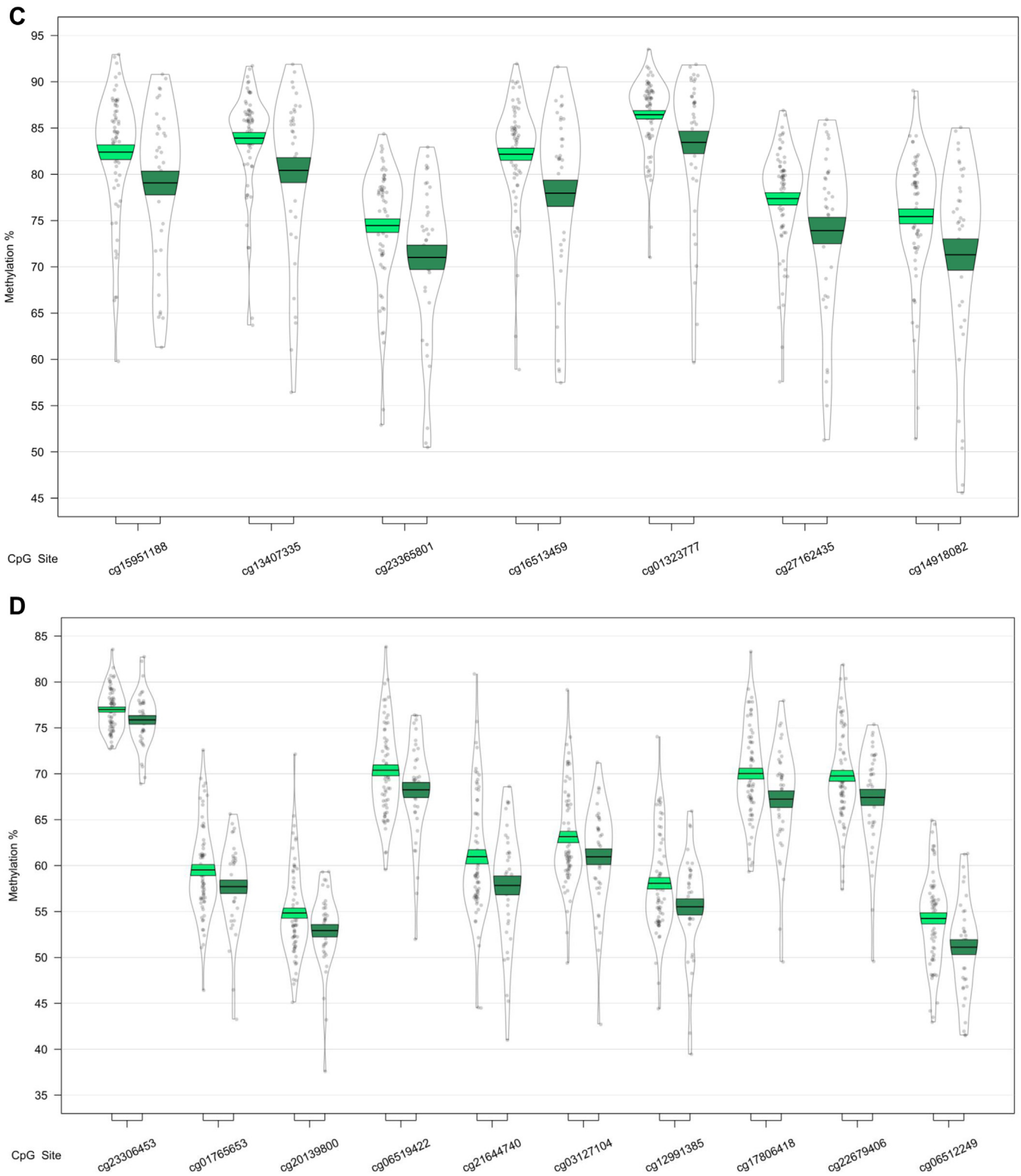

Fig. 3. (*continued*).

3.5. Clusters of methylated loci associated with mild cognitive impairment and Alzheimer's disease

To identify clusters of probes that are comethylated and are therefore hypothesized to share a common function, we performed WGCNA and classified the entire filtered data set of 200,633 probes into 16 modules (Fig. 5A). These modules were correlated to the group comparisons of diagnostic status at baseline, as well as to several other traits of interest (Fig. 5B, Supplementary Table 11), after controlling for covariates. The brown module, which consists of 11,794 probes, was negatively correlated with differences between CTL and MCI ($\rho = -0.16$, $p = 2.31 \times 10^{-2}$) and correlated

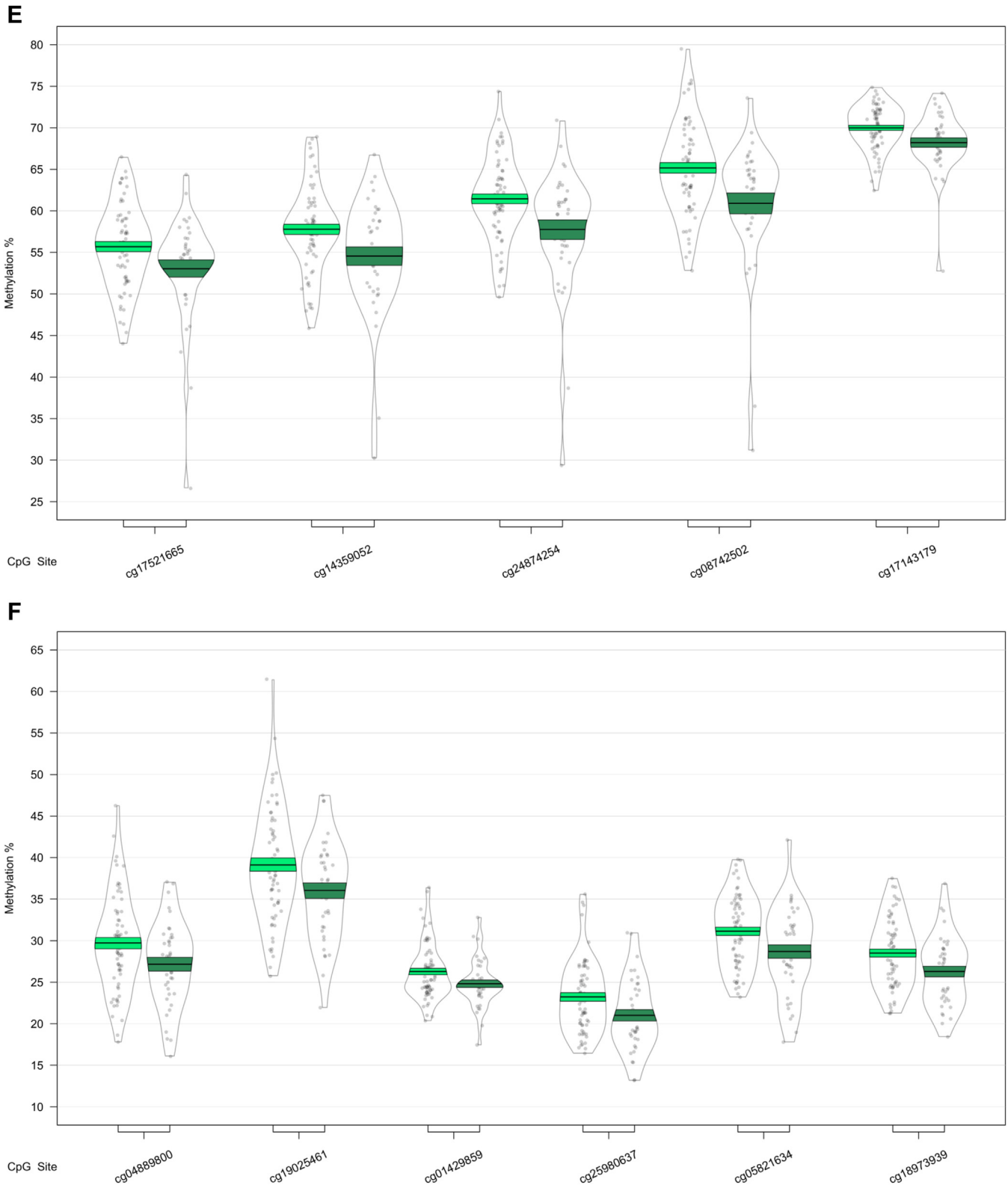

Fig. 3. (*continued*).

positively with an individual's number of education years ($r = 0.13$, $p = 3.59 \times 10^{-2}$). Three more modules also showed a correlation with MCI versus CTL; the light cyan module consisting of 133 probes ($\rho = 0.18$, $p = 1.2 \times 10^{-2}$), and the yellow module which consists of 10,635 probes ($\rho = 0.17$, $p = 1.51 \times 10^{-2}$). The yellow module further correlates to the structural imaging variable MET

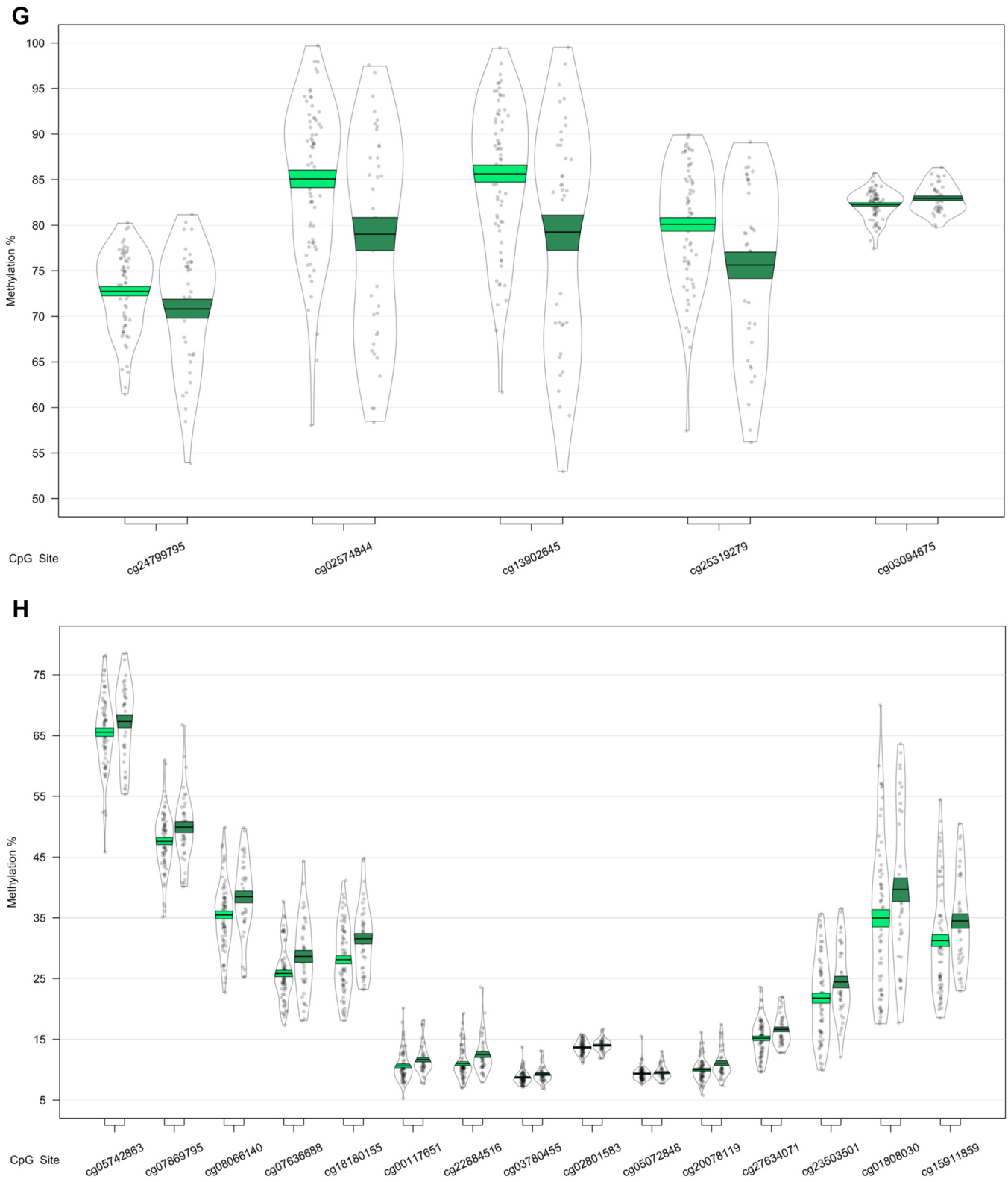

Fig. 3. (*continued*).

($r = -0.14$, $p = 4.26 \times 10^{-2}$). The purple module (792 probes) also correlates to MCI versus CTL ($\rho = -0.17$, $p = 1.98 \times 10^{-2}$), as well as the majority of structural imaging variables: REV ($r = 0.21$, $p = 3.16 \times 10^{-3}$), TEV ($r = 0.18$, $p = 9.85 \times 10^{-3}$), MET ($r = 0.25$, $p = 3.22 \times 10^{-4}$), VV ($r = -0.18$, $p = 9.25 \times 10^{-3}$), LHV ($r = 0.22$, $p = 1.19 \times 10^{-3}$), RHV ($r = 0.20$, $p = 3.31 \times 10^{-3}$), THV ($r = 0.22$, $p =$

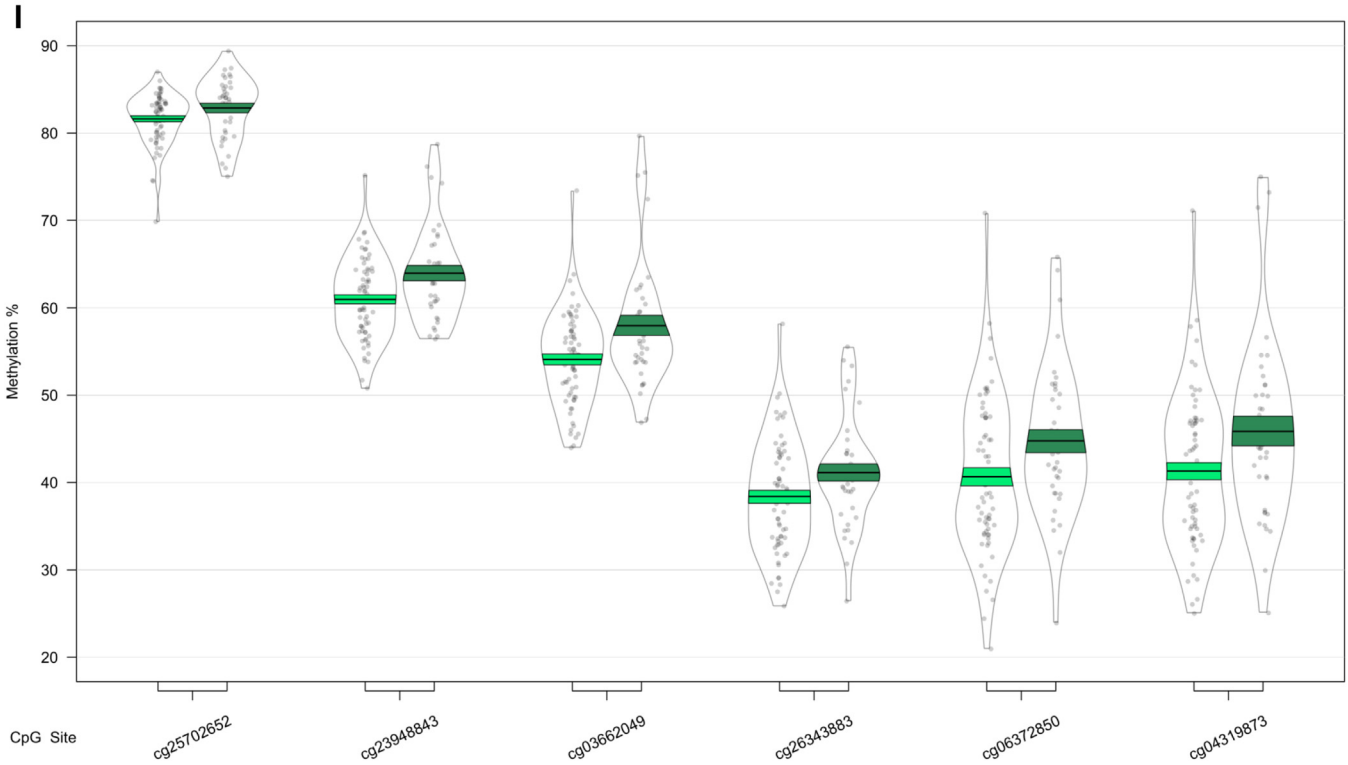

Fig. 3. (*continued*).

1.46×10^{-3}), and WBV ($r = 0.20$, $p = 3.55 \times 10^{-3}$). Finally, the cyan module (280 probes) correlates to an individual's number of *APOE ε4* alleles ($\rho = -0.14$, $p = 1.75 \times 10^{-2}$).

Subsequently, we investigated whether the probes that are integral to a specific module are also the probes that are driving the association with the relevant diagnosis or trait. We did this by correlating and plotting MM and PS values, and focusing on those modules that showed positive ($r > 0, p < 0.05$) correlations between MM and PS (Supplementary Table 12). Significant positive MM to PS correlations were found in the brown ($r = 0.26$, $p = 5.93 \times 10^{-179}$), purple ($r = 0.19$, $p = 9.18 \times 10^{-8}$), and yellow ($r = 0.25$, $p = 6.64 \times 10^{-153}$) modules in association with CTL versus MCI. The brown module further showed significant positive MM to PS correlations in relation to education years ($r = 0.11$, $p = 2.40 \times 10^{-32}$). The yellow module displayed a positive MM to PS correlation ($r = 0.22$, $p = 6.42 \times 10^{-117}$) in association with MET, and the cyan module showed a positive MM to PS correlation in association with the number of *APOE* ε4 alleles ($r = 0.20$, $p = 6.42 \times 10^{-4}$). These modules were the primary focus of our pathway analyses. MM and PS plots for these modules are shown in Supplementary Fig. 6; for a full overview of all MM and PS correlations, see Supplementary Table 12.

3.6. Functional role of modules associated with mild cognitive impairment and Alzheimer's disease

We sought to identify the pathways that were enriched in modules that were affected in disease or were associated with certain traits. For this purpose, we performed GO and KEGG enrichment analyses, with for large (i.e., yellow and brown) modules only the core probes being used for the enrichment analyses. Pathways related to the brown (Supplementary Fig. 7), purple, yellow (Supplementary Fig. 8), and cyan (Supplementary Fig. 9) modules all passed false discovery rate (FDR) multiple testing correction. A large number of GO terms were associated with the core of the brown module, which was related to MCI relative to CTL as well as number of education years, among which were "extracellular matrix" ($q = 4.23 \times 10^{-7}$), "channel activity" ($q = 3.19 \times 10^{-5}$), and "passive transmembrane transporter activity" ($q = 3.19 \times 10^{-5}$) (Supplementary Fig. 7A). Furthermore, KEGG terms related to this module included "Protein digestion and absorption" ($q = 1.06 \times 10^{-2}$), "Oxytocin signaling pathway" ($q = 1.06 \times 10^{-2}$), and "Regulation of actin cytoskeleton" ($q = 1.10 \times 10^{-2}$) (Supplementary Fig. 7B). The core of the yellow module showed differences related to MCI, relative to CTL, as well as MET, and we found in our enrichment analyses of the core probes that the top GO terms included "leukocyte activation" ($q = 7.46 \times 10^{-13}$), "cell activation" ($q = 7.46 \times 10^{-13}$), and "immune response" ($q = 5.84 \times 10^{-11}$), while the top KEGG terms included "platelet activation" ($q = 1.93 \times 10^{-2}$), "adrenergic signaling in cardiomyocytes" ($q = 1.93 \times 10^{-2}$), and "sphingolipid signaling pathway" ($q = 2.34 \times 10^{-2}$) (Supplementary Fig. 8). The purple module, which was also associated with differences related to MCI relative to CTL, was connected with one GO term; "vesicle-mediated transport" ($q = 4.35 \times 10^{-2}$), but no significant KEGG terms. Finally, the cyan module, which was associated with the number of *APOE ε4* alleles, was related to a number of GO terms, including "cell activation" ($q = 3.07 \times 10^{-4}$), "regulation of cell adhesion" ($q = 4.51 \times 10^{-4}$), "leukocyte activation" ($q = 4.51 \times 10^{-4}$), and "regulation of cell death" ($q = 8.32 \times 10^{-4}$) (Supplementary Fig. 9) and one KEGG pathway: "T cell receptor signaling pathway" ($q = 3.73 \times 10^{-2}$).

3.7. Investigating clusters of comethylated loci associated with progression to Alzheimer's disease

In addition to modules associated with baseline diagnosis of MCI and AD, we also identified 31 modules of highly comethylated loci in the subset of MCI-MCI and MCI-AD samples (Supplementary Fig. 10). Only one of these modules, the orange module, was shown to be significantly associated with future progression to AD (β = -0.04, $p = 4.38 \times 10^{-2}$; Supplementary Table 13). We then

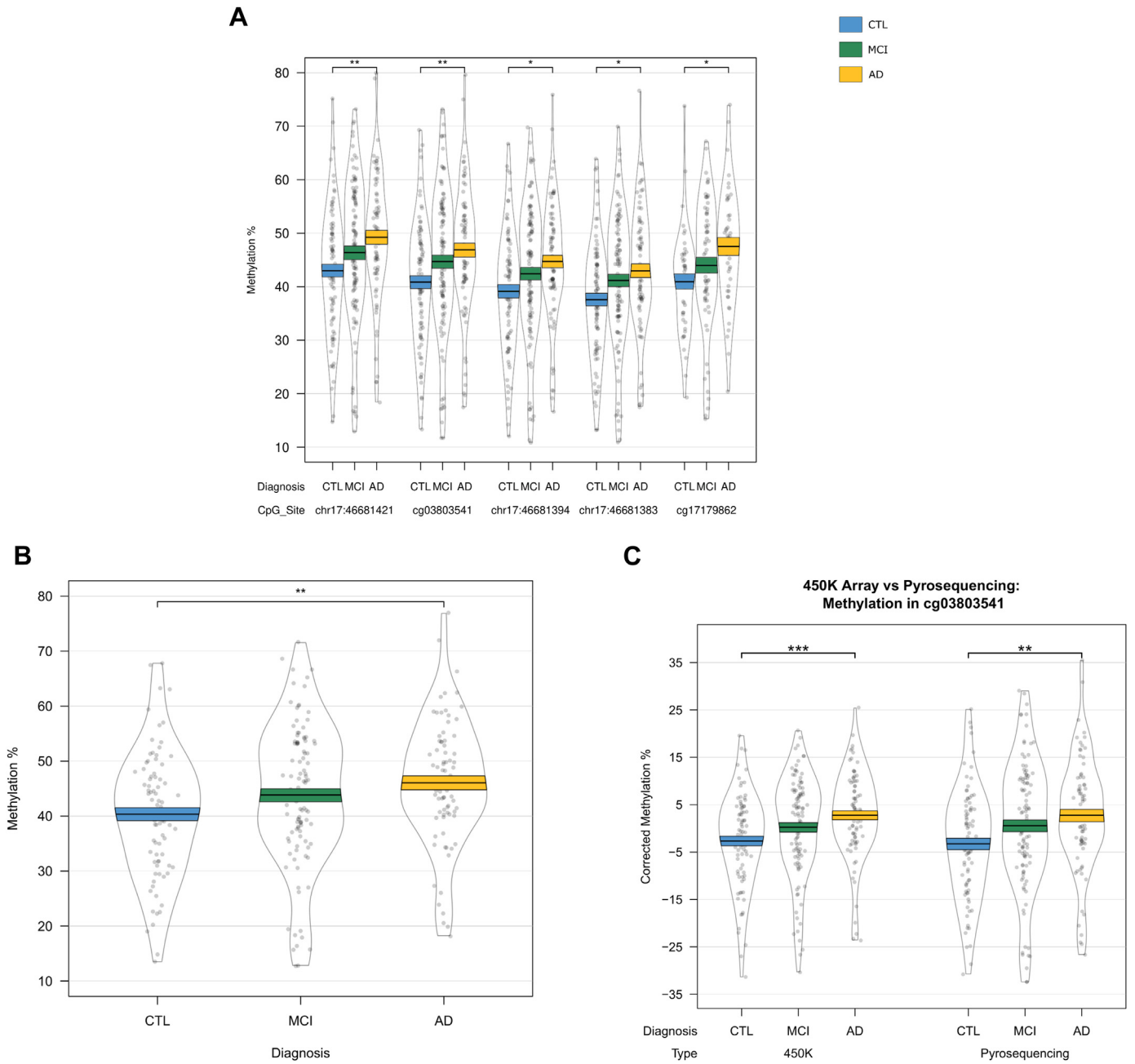

Fig. 4. Validation of the HOXB6 differentially methylated region. DNA methylation was assessed via pyrosequencing and significant changes were found between controls and individuals with Alzheimer's disease, but not mild cognitive impairment at all 5 CpG sites assessed (A) and when averaged over all 5 probes (B). DNA methylation patterns quantified by the Illumina 450K array and pyrosequencing were similar for cg03803541 (C) and cg17179862 (D) and were significantly correlated (E and F, respectively).

correlated the MM to the PS for this module and found a significant positive correlation ($r = 0.36$, $p = 9.40 \times 10^{-6}$). Following GO and KEGG pathway analysis, we found no GO terms passing FDR multiple testing correction, but top KEGG terms included "renal cell carcinoma" ($q = 1.21 \times 10^{-2}$), "nonhomologous end-joining" ($q = 2.00 \times 10^{-2}$), and "ErbB signaling pathway" ($q = 2.00 \times 10^{-2}$) (Supplementary Fig. 11).

4. Discussion

The present study, which reflects the first large-scale EWAS of AD blood samples, identified epigenetic signatures related to AD and MCI, as well as signatures associated with future conversion from MCI to AD.

The *HOXB6* gene contained a DMR that reflected differences in methylation in AD relative to CTL, which were validated using pyrosequencing. *HOXB6* encodes the homeobox protein B6, which is part of a larger cluster of homeobox B genes located on chromosome 17. Homeobox genes are DNA-binding proteins that have been implicated in early body morphogenesis (Krumlauf, 1994) as well as hematopoietic development. Specifically, *HOXB6* has been shown to be required for normal generation of granulocytes and monocytes (Giampaolo et al., 2002). Interestingly, a recent EWAS of AD hippocampus has shown DNA methylation differences in cg17179862, which was the most significant probe in the DMR we identified and validated (Altuna et al., 2019). The study by Altuna et al. further showed that increased methylation at this locus was positively correlated with tau burden.

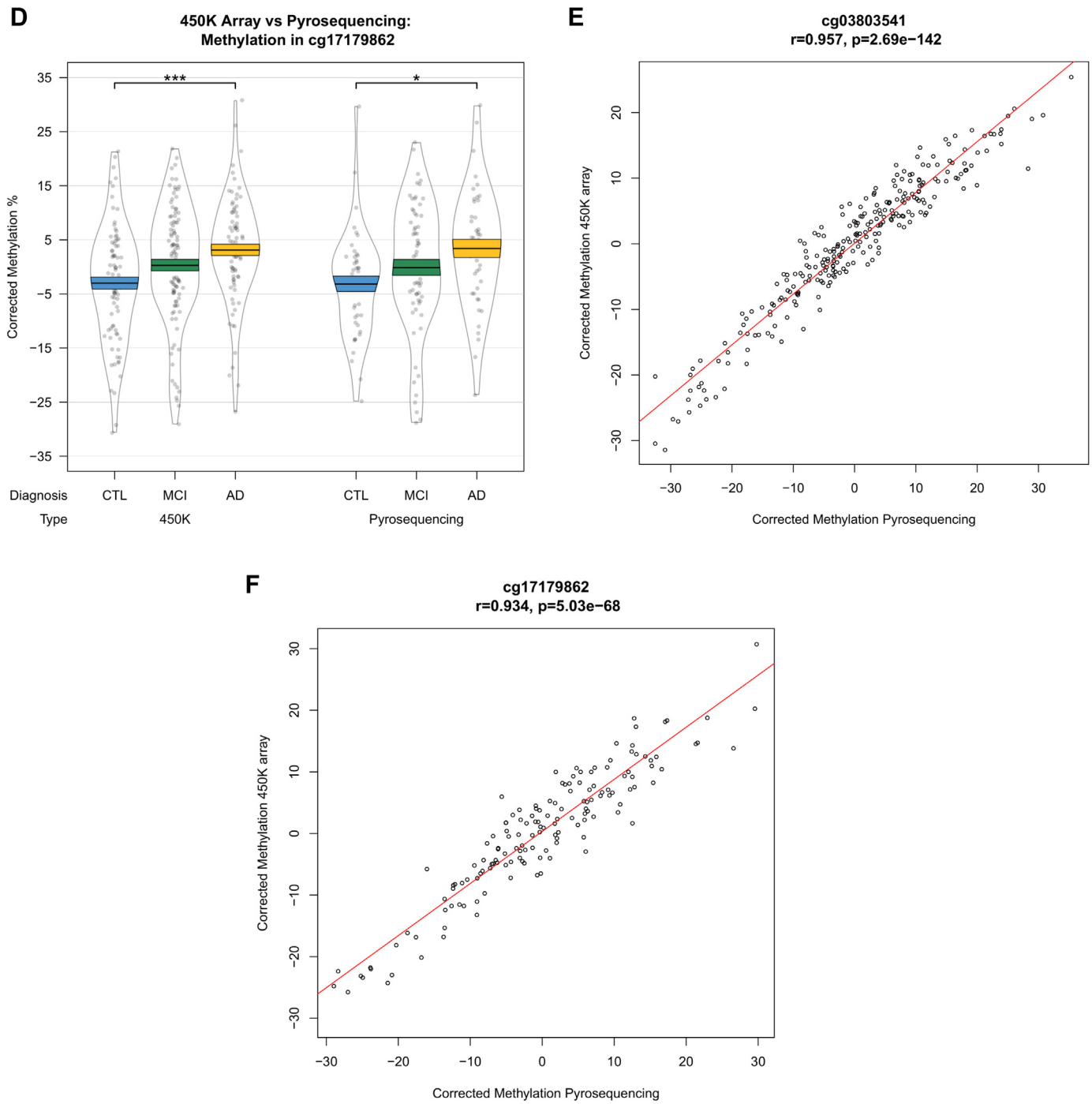

Fig. 4. (*continued*).

MOV10L1, which was associated with differences between all 3 groups, encodes an RNA helicase. This protein was shown to be crucial for the production of Piwi-interacting RNAs (piRNAs) by Vourekas et al. (2015). PiRNAs represent small noncoding RNAs involved in epigenetic regulation, which can bind to PIWI proteins and may induce gene silencing via DNA methylation (Aravin et al., 2008; Girard et al., 2006), or RNA-cleavage (for a review, see Luteijn and Ketting (2013)). Although initially believed to be mainly present in germline cells, piRNAs have been shown to be stably expressed in human blood (Yang et al., 2015) and have also been shown to be downregulated in tumor tissue and upregulated in blood of renal carcinomas (Iliev et al., 2016). Interestingly, Watson et al. (2016) performed an EWAS of AD superior temporal gyrus and identified a DMR spanning 13 probes, including all ten probes we identified in the *MOV10L1* gene in the current study. Of note, where Watson et al. detected AD-related hypermethylation in these 10 probes, we found hypermethylation in MCI when compared with AD and CTL individuals, while methylation levels of AD subjects were not distinct from CTL individuals.

Of the 9 DMRs that were related to future conversion to AD, our most significant region was located in *CPT1B*, which encodes the protein carnitine palmitoyltransferase 1B. Differential DNA

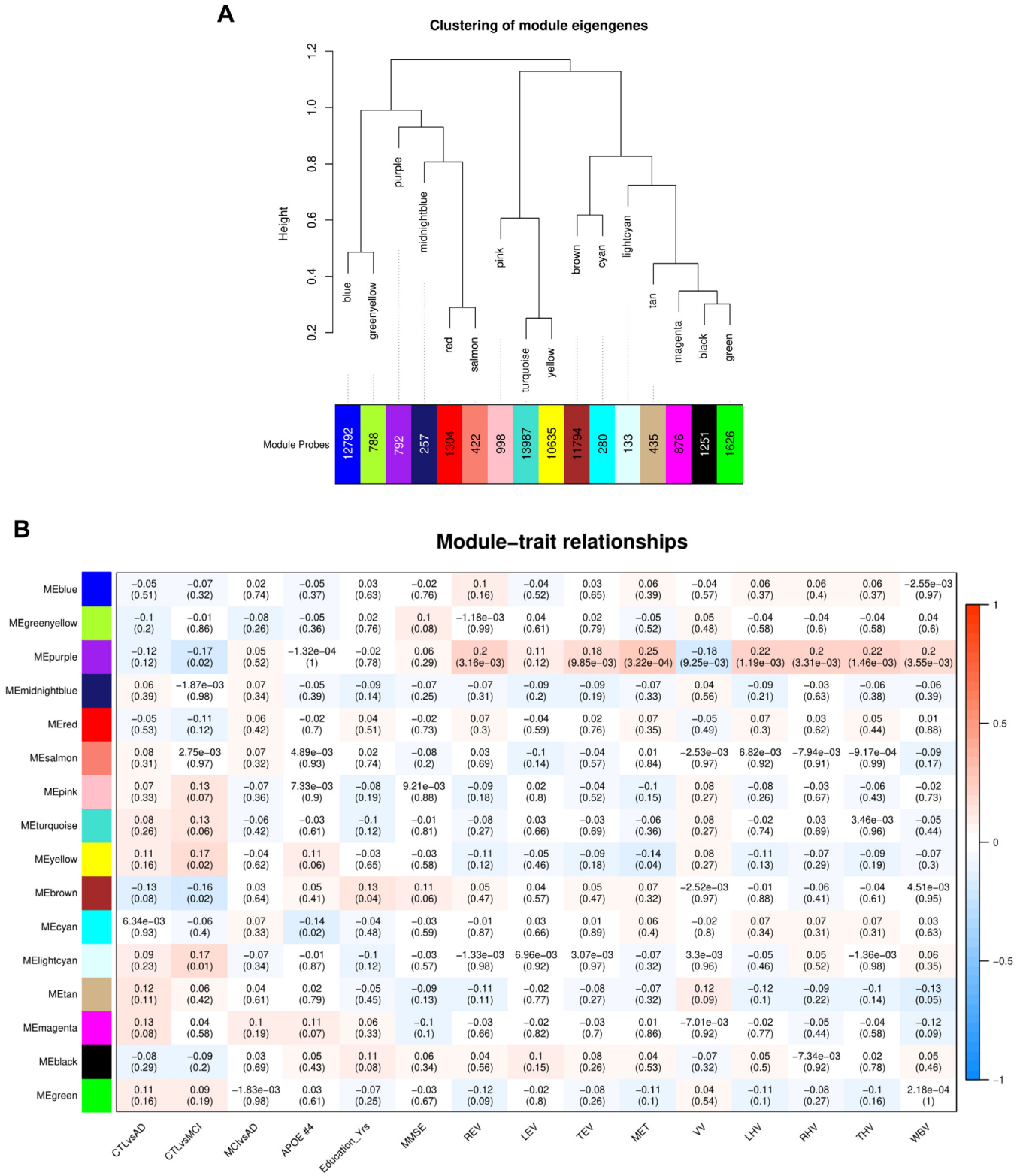

Fig. 5. Clusters (or "modules") of highly comethylated loci identified in the full dataset of 284 samples (A). Modules are hierarchically clustered based on calculated module eigengenes (representative of the methylation values within each module), and the number of probes included in each module are indicated along the x-axes. The color of each module is assigned in an arbitrary manner. (B) Correlations between module eigengenes and traits of interest, with module names shown along the y-axis. Correlation estimates are reported, with *p*-values in parentheses. Spearman correlations were performed for the controls (CTL) versus Alzheimer's disease (AD) comparison, CTL versus mild cognitive impairment (MCI) comparison, the MCI versus AD comparison, and the number of *APOE-ε4* alleles (APOE #4). Pearson correlations were calculated for the number of education years (Education_Yrs); mini mental state examination (MMSE) scores; and the following structural imaging measurements: right, left, and total entorhinal volume (REV, LEV, and TEV, respectively); mean entorhinal thickness (MET); ventricular volume (VV); left, right and total hippocampal volume (LHV, RHV, and THV, respectively); and whole brain volume (WBV). Modules with a correlation *p*-value <0.05 were selected for further analysis. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

methylation in *CPT1B* has been previously identified in blood and fetal cortex of patients with Down syndrome (El Hajj et al., 2016; Kerkel et al., 2010). This is interesting as individuals with Down syndrome often develop AD as a result of trisomy of chromosome 21, causing them to have an additional copy of the amyloid precursor protein (*APP*) gene. The study by El Hajj et al. (2016) identified a DMR in *CPT1B* consisting of 18 probes in Down syndrome fetal cortex samples, which spanned the region discovered in the present study. They detected hypermethylation in 13 probes in Down syndrome, while we observed hypomethylation in those MCI individuals who convert to AD. Kerkel et al. (2010) similarly detected hypermethylation at one CpG site in our *CPT1B* DMR in peripheral blood leukocytes of individuals with Down syndrome, concomitant with significant overexpression of the gene. While we observed hypomethylation of the DMR, the positive relationship found between methylation and expression for this region was validated in our study. Of note, overexpression of *CPT1B* has also been found in blood from soldiers with post-traumatic stress disorder (Zhang et al., 2015), a known risk factor for developing AD (Agís-Balboa et al., 2017; Yaffe et al., 2010). Interestingly, no overlap was found between DMRs associated with conversion and DMRs related to diagnosis at baseline. This may reflect limited power in our MCI conversion analysis due to sample size or could reflect temporal patterns of DNA methylation in the process of conversion from MCI to AD.

In addition to DMRs, by using WGCNA and subsequent pathway analyses we further identified biological mechanisms affected in disease. The cyan module that was linked to the number of *APOE* ε4 alleles is involved in GO pathways related to the immune system, which is interesting given that the immune system is known to be activated in AD (Heppner et al., 2015), and as *APOE* ε4 is the strongest genetic risk factor for sporadic AD (Lambert et al., 2013). The core of the brown module, which reflects methylomic differences related to an individual's number of education years and differences in MCI relative to CTL, was shown to be involved in transmembrane processes (GO), as well as oxytocin signaling (KEGG). The oxytocin signaling pathway is linked to social behaviors, as well as several psychiatric disorders (e.g., depression) (Feldman et al., 2016) Interestingly, a DMR was recently identified in the oxytocin gene (*OXT*), which was hypomethylated in AD brain (Lardenoije et al., 2019; Watson et al., 2016) and hypermethylated in the blood in individuals who subsequently converted to AD (Lardenoije et al., 2019). Oxytocin is involved in the modulation of stress, social behaviors, and associative learning (Olff et al., 2013), and altered levels of oxytocin have been reported in AD postmortem brain tissue (Mazurek et al., 1987) and cerebrospinal fluid (North et al., 1992). It is interesting that the sphingolipid signaling pathway is found in the KEGG results from the core probes of the yellow module related to differences in MCI relative to CTL. Multiple studies have indicated that sphingolipid signaling pathways are implicated in AD (Crivelli et al., 2020), and the measurement of lipids in the pathway is being explored as a potential biomarker of AD and neurodegeneration (Mielke and Lyketsos, 2010). Similar to the cyan module, the majority of GO terms in the core of the yellow module are related to various processes of immune activation. In our network analysis examining conversion from MCI to AD, we identified a pathway in the orange module associated with nonhomologous end-joining. Nonhomologous end-joining activity is involved in repairing the double-strand DNA breaks and has been reported to be decreased in AD brain (Kanungo, 2013; Shackelford, 2006).

In summary, this is the first EWAS to identify epigenetic signatures and functional pathways specific to MCI, AD, and conversion to AD in the blood. However, there are some limitations to our study. First, we have profiled DNA methylation patterns in whole blood, and it is known that there are subtle alterations in the abundance of specific blood cell types in MCI and AD (Lunnon et al., 2012). Although we have controlled for the proportions of these different cells, it will be of interest to investigate disease-associated signatures in individual cell types. Second, individuals were only followed up clinically for up to 2 years following the baseline assessment and further studies should profile cohorts consisting of CTL and MCI subjects with long-term clinical follow-up to identify the preclinical changes. In addition, biomarkers were not available to support the clinical diagnosis of AD. Third, our comparisons of DNA methylation and gene expression were limited to only those genes with variable expression levels in the previous study (Lunnon et al., 2012) and did not examine transcript variants. Fourth, we have not replicated our findings in an independent study cohort. Although we did validate our *HOXB6* DMR in the same samples using an alternative technology, in the future it will be interesting to verify the loci we identified in a different set of samples. Finally, although there is some communication between the brain and the blood, not all differences found to be associated with AD in the blood may be functionally related to the processes taking place in the brain. Differences in DNA methylation may be the result of parallel effects or comorbidities, and may not be causally related to disease, but could reflect mediating or downstream effects. It would be interesting for future studies to explore the exact role of the epigenetic signatures identified in this study, and to explore their potential as biomarkers for an early diagnosis of AD and therapeutic targets.

CRediT authorship contribution statement

Janou A.Y. Roubroeks: Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing, Formal analysis, Investigation. **Adam R. Smith:** Investigation, Validation, Writing - review & editing. **Rebecca G. Smith:** Methodology, Writing - review & editing. **Ehsan Pishva:** Methodology, Formal analysis, Writing - review & editing. **Zina Ibrahim:** Formal analysis, Resources. **Martina Sattlecker:** Formal analysis, Writing - review & editing, Resources. **Eilis J. Hannon:** Methodology, Writing - review & editing. **Iwona Kłoszewska:** Resources. **Patrizia Mecocci:** Resources. **Hilkka Soininen:** Resources. **Magda Tsolaki:** Writing - review & editing, Resources. **Bruno Vellas:** Resources. **Lars-Olof Wahlund:** Writing - review & editing, Resources. **Dag Aarsland:** Writing - review & editing, Resources. **Petroula Proitsi:** Resources. **Angela Hodges:** Resources. **Simon Lovestone:** Writing - review & editing, Resources. **Stephen J. Newhouse:** Resources. **Richard J.B. Dobson:** Writing - review & editing, Resources. **Jonathan Mill:** Funding acquisition, Writing - review & editing, Resources. **Daniël L.A. van den Hove:** Funding acquisition, Conceptualization, Writing - original draft, Writing - review & editing, Supervision, Project administration. **Katie Lunnon:** Funding acquisition, Conceptualization, Writing - original draft, Writing - review & editing, Supervision, Project administration.

Acknowledgements

We thank the AddNeuroMed cohort and the participants who made this research possible. The AddNeuroMed data are from a public-private partnership supported by EFPIA companies and SMEs as part of InnoMed (Innovative Medicines in Europe), an Integrated Project funded by the European Union of the Sixth Framework program priority FP6 2004-LIFESCIHEALTH-5. Clinical leads responsible for data collection are Iwona Kłoszewska (Lodz), Simon Lovestone (London), Patrizia Mecocci (Perugia), Hilkka Soininen (Kuopio), Magda Tsolaki (Thessaloniki), and Bruno Vellas (Toulouse); the imaging leads are Andy Simmons (London), Lars-

Olad Wahlund (Stockholm), and Christian Spenger (Zurich); and the bioinformatics leads are Richard Dobson (London) and Stephen Newhouse (London). This work was further supported by The National Institute for Health Research at the University College London Hospitals Biomedical Research Centre, the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, and Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome Trust.

Support for the current study was provided by a major project grant from the Alzheimer's Society, United Kingdom (AS-PG-14–038) to KL, a project grant from the Medical Research Council (MRC), United Kingdom (MR/N027973/1) to KL as part of the Joint Programme—Neurodegenerative Disease Research (JPND) initiative for the EPI-AD consortium, an NIH, United States, R01 grant (AG036039) to JM, an Equipment Grant from Alzheimer's Research UK (ARTEG2010A-2) to JM, a PhD studentship from the GW4 MRC BioMed DTP (United Kingdom) for JAYR, and an Alzheimer's Research UK South West Network Centre pump priming award to JAYR.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neurobiolaging.2020.06.023.

References

Agís-Balboa, R.C., Pinheiro, P.S., Rebola, N., Kerimoglu, C., Benito, E., Gertig, M., Bahari-Javan, S., Jain, G., Burkhardt, S., Delalle, I., Jatzko, A., Dettenhofer, M., Zunszain, P.A., Schmitt, A., Falkai, P., Pape, J.C., Binder, E.B., Mulle, C., Fischer, A., Sananbenesi, F., 2017. Formin 2 links neuropsychiatric phenotypes at young age to an increased risk for dementia. EMBO J. 36, 2815–2828.

Altuna, M., Urdánoz-Casado, A., Sánchez-Ruiz De Gordoa, J., Zelaya, M.V., Labarga, A., Lepesant, J.M.J., Roldán, M., Blanco-Luquin, I., Perdones, Á., Larumbe, R., Jericó, I., Echavarri, C., Méndez-López, I., Di Stefano, L., Mendioroz, M., 2019. DNA methylation signature of human hippocampus in Alzheimer's disease is linked to neurogenesis. Clin. Epigenetics 11, 91.

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., Hannon, G.J., 2008. A piRNA pathway primed by individual transposons is linked to De Novo DNA methylation in Mice. Mol. Cell 31, 785–799.

Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., Irizarry, R.A., 2014. Minfi: a flexible and comprehensive bioconductor package for the analysis of infinium DNA methylation microarrays. Bioinformatics 30, 1363–1369.

Asare, A.L., Kolchinsky, S.A., Gao, Z., Wang, R., Raddassi, K., Bourcier, K., Seyfert-Margolis, V., 2008. Differential gene expression profiles are dependent upon method of peripheral blood collection and RNA isolation. BMC Genomics 9, 474.

Booij, B.B., Lindahl, T., Wetterberg, P., Skaane, N.V., Sæbø, S., Feten, G., Ryea, P.D., Kristiansen, L.I., Hagen, N., Jensen, M., Bårdsen, K., Winblad, B., Sharma, P., Lönneborga, A., 2011. A gene expression pattern in blood for the early detection of Alzheimer's disease. J. Alzheimers Dis. 23, 109–119.

Chen, Y.A., Lemire, M., Choufani, S., Butcher, D.T., Grafodatskaya, D., Zanke, B.W., Gallinger, S., Hudson, T.J., Weksberg, R., 2013. Discovery of cross-reactive probes and polymorphic CpGs in the illumina infinium Humanmethylation450 microarray. Epigenetics 8, 203–209.

Crivelli, S.M., Giovagnoni, C., Visseren, L., Scheithauer, A.-L., de Wit, N., den Hoedt, S., Losen, M., Mulder, M.T., Walter, J., de Vries, H.E., Bieberich, E., Martinez-Martinez, P., 2020. Sphingolipids in Alzheimer's disease, how can we target them? Adv. Drug Deliv. Rev. https://doi.org/10.1016/j.addr.2019.12.003.

da Silva, P.N.O., Furuya, T.K., Braga, I.L., Rasmussen, L.T., Labio, R.W., Bertolucci, P.H., Chen, E.S., Turecki, G., Mechawar, N., Payão, S.L., Mill, J., Smith, M.C., 2014. Analysis of HSPA8 and HSPA9 mRNA expression and promoter methylation in the brain and blood of Alzheimer's disease patients. J. Alzheimers Dis. 38, 165–170.

De Jager, P.L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L.C., Yu, L., Eaton, M.L., Keenan, B.T., Ernst, J., McCabe, C., Tang, A., Raj, T., Replogle, J., Brodeur, W., Gabriel, S., Chai, H.S., Younkin, C., Younkin, S.G., Zou, F., Szyf, M., Epstein, C.B., Schneider, J.A., Bernstein, B.E., Meissner, A., Ertekin-Taner, N., Chibnik, L.B., Kellis, M., Mill, J., Bennett, D.A., 2014. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat. Neurosci. 17, 1156–1163.

del Barrio, V., 2004. Diagnostic and statistical manual of mental disorders. Encycl. Appl. Psychol. 3, 607–614.

El Hajj, N., Dittrich, M., Böck, J., Kraus, T.F.J., Nanda, I., Müller, T., Seidmann, L., Tralau, T., Galetzka, D., Schneider, E., Haaf, T., 2016. Epigenetic dysregulation in the developing down syndrome cortex. Epigenetics 11, 563–578.

Fehlbaum-Beurdeley, P., Sol, O., Désiré, L., Touchon, J., Dantoine, T., Vercelletto, M., Gabelle, A., Jarrige, A.C., Haddad, R., Lemarié, J.C., Zhou, W., Hampel, H., Einstein, R., Vellas, B., 2012. Validation of AclarusDx™, a blood-based transcriptomic signature for the diagnosis of Alzheimer's disease. J. Alzheimers Dis. 32, 169–181.

Feldman, R., Monakhov, M., Pratt, M., Ebstein, R.P., 2016. Oxytocin pathway genes: evolutionary ancient system impacting on human affiliation, sociality, and psychopathology. Biol. Psychiatry 79, 174–184.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12, 189–198.

Furney, S.J., Kronenberg, D., Simmons, A., Güntert, A., Dobson, R.J., Proitsi, P., Wahlund, L.O., Kloszewska, I., Mecocci, P., Soininen, H., Tsolaki, M., Vellas, B., Spenger, C., Lovestone, S., 2011a. Combinatorial markers of mild cognitive impairment conversion to Alzheimers disease - cytokines and MRI measures together predict disease progression. J. Alzheimers Dis. 26, 395–405.

Furney, S.J., Simmons, A., Breen, G., Pedroso, I., Lunnon, K., Proitsi, P., Hodges, A., Powell, J., Wahlund, L.O., Kloszewska, I., Mecocci, P., Soininen, H., Tsolaki, M., Vellas, B., Spenger, C., Lathrop, M., Shen, L., Kim, S., Saykin, A.J., Weiner, M.W., Lovestone, S., 2011b. Genome-wide association with MRI atrophy measures as a quantitative trait locus for Alzheimer's disease. Mol. Psychiatry 16, 1130–1138.

Furuya, T.K., da Silva, P.N.O., Payão, S.L.M., Bertolucci, P.H.F., Rasmussen, L.T., De Labio, R.W., Braga, I.L.S., Chen, E.S., Turecki, G., Mechawar, N., Mill, J., Smith, M.A.C., 2012a. Analysis of SNAP25 mRNA expression and promoter DNA methylation in brain areas of Alzheimer's disease patients. Neuroscience 220, 41–46.

Furuya, T.K., da Silva, P.N.O., Payão, S.L.M., Rasmussen, L.T., De Labio, R.W., Bertolucci, P.H.F., Braga, I.L.S., Chen, E.S., Turecki, G., Mechawar, N., Mill, J., De Arruda Cardoso Smith, M., 2012b. SORL1 and SIRT1 mRNA expression and promoter methylation levels in aging and Alzheimer's disease. Neurochem. Int. 61, 973–975.

Gasparoni, G., Bultmann, S., Lutsik, P., Kraus, T.F.J., Sordon, S., Vlcek, J., Dietinger, V., Steinmaurer, M., Haider, M., Mulholland, C.B., Arzberger, T., Roeber, S., Riemenschneider, M., Kretzschmar, H.A., Giese, A., Leonhardt, H., Walter, J., 2018. DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. Epigenetics Chromatin 11, 41.

Giampaolo, A., Felli, N., Diverio, D., Morsilli, O., Samoggia, P., Breccia, M., Lo Coco, F., Peschle, C., Testa, U., 2002. Expression pattern of HOXB6 homeobox gene in myelomonocytic differentiation and acute myeloid leukemia. Leukemia 16, 1293–1301.

Girard, A., Sachidanandam, R., Hannon, G.J., Carmell, M.A., 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature 442, 199–202.

Hardy, J., 1997. Amyloid, the presenilins and Alzheimer's disease. Trends Neurosci. 20, 154–159.

Heppner, F.L., Ransohoff, R.M., Becher, B., 2015. Immune attack: the role of inflammation in Alzheimer disease. Nat. Rev. Neurosci. 16, 358–372.

Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K., Kelsey, K.T., 2012. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 13, 86.

Hye, A., Lynham, S., Thambisetty, M., Causevic, M., Campbell, J., Byers, H.L., Hooper, C., Rijsdijk, F., Tabrizi, S.J., Banner, S., Shaw, C.E., Foy, C., Poppe, M., Archer, N., Hamilton, G., Powell, J., Brown, R.G., Sham, P., Ward, M., Lovestone, S., 2006. Proteome-based plasma biomarkers for Alzheimer's disease. Brain 129, 3042–3050.

Iliev, R., Fedorko, M., MacHackova, T., Mlcochova, H., Svoboda, M., Pacik, D., Dolezel, J., Stanik, M., Slaby, O., 2016. Expression levels of PIWI-interacting RNA, piR-823, are deregulated in tumor tissue, blood serum and urine of patients with renal cell carcinoma. Anticancer Res. 36, 6419–6423.

Jack, C.R., Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q., 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol. 9, 119–128.

Jicha, G.A., Parisi, J.E., Dickson, D.W., Johnson, K., Cha, R., Ivnik, R.J., Tangalos, E.G., Boeve, B.F., Knopman, D.S., Braak, H., Petersen, R.C., 2006. Neuropathologic outcome of mild cognitive impairment following progression to clinical dementia. Arch. Neurol. 63, 674–681.

Kanungo, J., 2013. DNA-dependent protein kinase and DNA repair: relevance to Alzheimer's disease. Alzheimers. Res. Ther. 5, 13.

Kerkel, K., Schupf, N., Hatta, K., Pang, D., Salas, M., Kratz, A., Minden, M., Murty, V., Zigman, W.B., Mayeux, R.P., Jenkins, E.C., Torkamani, A., Schork, N.J., Silverman, W., Croy, B.A., Tycko, B., 2010. Altered DNA methylation in leukocytes with trisomy 21. PLoS Genet. 6, e1001212.

Kobayashi, N., Shinagawa, S., Nagata, T., Shimada, K., Shibata, N., Ohnuma, T., Kasanuki, K., Arai, H., Yamada, H., Nakayama, K., Kondo, K., 2016. Development of biomarkers based on DNA methylation in the NCAPH2/LMF2 promoter region for diagnosis of Alzheimer's disease and amnesic mild cognitive impairment. PLoS One 11, e0146449.

Krumlauf, R., 1994. Hox genes in vertebrate development. Cell 78, 191–201.
Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., Jun, G., DeStefano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., Russo, G., Thornton-Wells, T.A., Jones, N., Smith, A.V., Chouraki, V., Thomas, C., Ikram, M.A., Zelenika, D., Vardarajan, B.N., Kamatani, Y., Lin, C.F., Gerrish, A., Schmidt, H., Kunkle, B., Fiévet, N., Amouyel, P., Pasquier, F., Deramecourt, V., De Bruijn, R.F.A.G., Amin, N., Hofman, A., Van Duijn, C.M., Dunstan, M.L., Hollingworth, P., Owen, M.J., O'Donovan, M.C., Jones, L., Holmans, P.A., Moskvina, V., Williams, J., Baldwin, C., Farrer, L.A., Choi, S.H., Lunetta, K.L., Fitzpatrick, A.L., Harris, T.B., Psaty, B.M., Gilbert, J.R., Hamilton-Nelson, K.L., Martin, E.R., Pericak-Vance, M.A., Haines, J.L., Gudnason, V., Jonsson, P.V., Eiriksdottir, G., Bihoreau, M.T., Lathrop, M., Valladares, O., Cantwell, L.B., Wang, L.S., Schellenberg, G.D., Ruiz, A., Boada, M., Reitz, C., Mayeux, R., Ramirez, A., Maier, W., Hanon, O., Kukull, W.A., Buxbaum, J.D., Campion, D., Wallon, D., Hannequin, D., Crane, P.K., Larson, E.B., Becker, T., Cruchaga, C., Goate, A.M., Craig, D., Johnston, J.A., Mc-Guinness, B., Todd, S., Passmore, P., Berr, C., Ritchie, K., Lopez, O.L., De Jager, P.L., Evans, D., Lovestone, S., Proitsi, P., Powell, J.F., Letenneur, L., Barberger-Gateau, P., Dufouil, C., Dartigues, J.F., Morón, F.J., Rubinsztein, D.C., St George-Hyslop, P., Sleegers, K., Bettens, K., Van Broeckhoven, C., Huentelman, M.J., Gill, M., Brown, K., Morgan, K., Kamboh, M.I., Keller, L., Fratiglioni, L., Green, R., Myers, A.J., Love, S., Rogaeva, E., Gallacher, J., Bayer, A., Clarimon, J., Lleo, A., Tsuang, D.W., Yu, L., Bennett, D.A., Tsolaki, M., Bossù, P., Spalletta, G., Collinge, J., Mead, S., Sorbi, S., Nacmias, B., Sanchez-Garcia, F., Deniz Naranjo, M.C., Fox, N.C., Hardy, J., Bosco, P., Clarke, R., Brayne, C., Galimberti, D., Mancuso, M., Matthews, F., Moebus, S., Mecocci, P., Del Zompo, M., Hampel, H., Pilotto, A., Bullido, M., Panza, F., Caffarra, P., Mayhaus, M., Pichler, S., Gu, W., Riemenschneider, M., Lannfelt, L., Ingelsson, M., Hakonarson, H., Carrasquillo, M.M., Zou, F., Younkin, S.G., Beekly, D., Alvarez, V., Coto, E., Razquin, C., Pastor, P., Mateo, I., Combarros, O., Faber, K.M., Foroud, T.M., Soininen, H., Hiltunen, M., Blacker, D., Mosley, T.H., Graff, C., Holmes, C., Montine, T.J., Rotter, J.I., Brice, A., Nalls, M.A., Kauwe, J.S.K., Boerwinkle, E., Schmidt, R., Rujescu, D., Tzourio, C., Nöthen, M.M., Launer, L.J., Seshadri, S., 2013. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat. Genet. 45, 1452–1458.
Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559.
Lardenoije, R., Roubroeks, J.A.Y., Pishva, E., Leber, M., Wagner, H., Iatrou, A., Smith, A.R., Smith, R.G., Eijssen, L.M.T., Kleineidam, L., Kawalia, A., Hoffmann, P., Luck, T., Riedel-Heller, S., Jessen, F., Maier, W., Wagner, M., Hurlemann, R., Kenis, G., Ali, M., Del Sol, A., Mastroeni, D., Delvaux, E., Coleman, P.D., Mill, J., Rutten, B.P.F., Lunnon, K., Ramirez, A., Van Den Hove, D.L.A., 2019. Alzheimer's disease-associated (hydroxy)methylomic changes in the brain and blood. Clin. Epigenetics 11, 164.
Liu, Y., Paajanen, T., Zhang, Y., Westman, E., Wahlund, L.O., Simmons, A., Tunnard, C., Sobow, T., Mecocci, P., Tsolaki, M., Vellas, B., Muehlboeck, S., Evans, A., Spenger, C., Lovestone, S., Soininen, H., 2011. Combination analysis of neuropsychological tests and structural MRI measures in differentiating AD, MCI and control groups-The AddNeuroMed study. Neurobiol. Aging 32, 1198–1206.
Lovestone, S., Francis, P., Kloszewska, I., Mecocci, P., Simmons, A., Soininen, H., Spenger, C., Tsolaki, M., Vellas, B., Wahlund, L.O., Ward, M., 2009. AddNeuroMed - the european collaboration for the discovery of novel biomarkers for alzheimer's disease. Ann. N. Y. Acad. Sci. 1180, 36–46.
Lovestone, S., Francis, P., Strandgaard, K., 2007. Biomarkers for disease modification trials - the innovative medicines initiative and addneuromed. J. Nutr. Heal. Aging 11, 359–361.
Lunnon, K., Ibrahima, Z., Proitsi, P., Lourdusamy, A., Newhouse, S., Sattlecker, M., Furney, S., Saleem, M., Soininen, H., Kłoszewska, I., Mecocci, P., Tsolaki, M., Vellas, B., Coppola, G., Geschwind, D., Simmons, A., Lovestone, S., Dobson, R., Hodges, A., 2012. Mitochondrial dysfunction and immune activation are detectable in early alzheimer's disease blood. J. Alzheimers Dis. 30, 685–710.
Lunnon, K., Keohane, A., Pidsley, R., Newhouse, S., Riddoch-Contreras, J., Thubron, E.B., Devall, M., Soininen, H., Kłoszewska, I., Mecocci, P., Tsolaki, M., Vellas, B., Schalkwyk, L., Dobson, R., Malik, A.N., Powell, J., Lovestone, S., Hodges, A., 2017. Mitochondrial genes are altered in blood early in Alzheimer's disease. Neurobiol. Aging 53, 36–47.
Lunnon, K., Sattlecker, M., Furney, S.J., Coppola, G., Simmons, A., Proitsi, P., Lupton, M.K., Lourdusamy, A., Johnston, C., Soininen, H., Kłoszewska, I., Mecocci, P., Tsolaki, M., Vellas, B., Geschwind, D., Lovestone, S., Dobson, R., Hodges, A., 2013. A blood gene expression marker of early Alzheimer's disease. J. Alzheimers Dis. 33, 737–753.
Lunnon, K., Smith, R., Hannon, E., De Jager, P.L., Srivastava, G., Volta, M., Troakes, C., Al-Sarraj, S., Burrage, J., Macdonald, R., Condliffe, D., Harries, L.W., Katsel, P., Haroutunian, V., Kaminsky, Z., Joachim, C., Powell, J., Lovestone, S., Bennett, D.A., Schalkwyk, L.C., Mill, J., 2014. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. Nat. Neurosci. 17, 1164–1170.
Luteijn, M.J., Ketting, R.F., 2013. PIWI-interacting RNAs: from generation to transgenerational epigenetics. Nat. Rev. Genet. 14, 523–534.
Madrid, A., Hogan, K.J., Papale, L.A., Clark, L.R., Asthana, S., Johnson, S.C., Alisch, R.S., 2018. DNA hypomethylation in blood links B3GALT4 and ZADH2 to alzheimer's disease. J. Alzheimers Dis. 66, 927–934.
Maunakea, A.K., Chepelev, I., Cui, K., Zhao, K., 2013. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res 23, 1256–1269.
Mazurek, M.F., Beal, M.F., Bird, E.D., Martin, J.B., 1987. Oxytocin in Alzheimer's disease. Neurology 37, 1001 LP.
McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M., 1984. Clinical diagnosis of Alzheimer's disease. Neurology 34, 939.
Mielke, M.M., Lyketsos, C.G., 2010. Alterations of the sphingolipid pathway in Alzheimer's disease: new biomarkers and treatment targets? Neuromolecular Med. 12, 331–340.
Morris, J.C., 1993. The clinical dementia rating (cdr): current version and scoring rules. Neurology 43, 2412–2414.
North, W.G., Harbaugh, R., Reeder, T., 1992. An evaluation of human neurophysin production in Alzheimer's disease: Preliminary observations. Neurobiol. Aging 13, 261–265.
O'Bryant, S.E., Edwards, M., Johnson, L., Hall, J., Villarreal, A.E., Britton, G.B., Quiceno, M., Cullum, C.M., Graff-Radford, N.R., 2016. A blood screening test for Alzheimer's disease. Alzheimers Dement. 3, 83–90.
O'Bryant, S.E., Xiao, G., Barber, R., Reisch, J., Doody, R., Fairchild, T., Adams, P., Waring, S., Diaz-Arrastia, R., 2010. A serum protein-based algorithm for the detection of Alzheimer disease. Arch. Neurol. 67, 1077–1081.
O'Bryant, S.E., Xiao, G., Barber, R., Reisch, J., Hall, J., Cullum, C.M., Doody, R., Fairchild, T., Adams, P., Wilhelmsen, K., Diaz-Arrastia, R., 2011. A blood-based algorithm for the detection of Alzheimer's disease. Dement. Geriatr. Cogn. Disord. 32, 55–62.
Olff, M., Frijling, J.L., Kubzansky, L.D., Bradley, B., Ellenbogen, M.A., Cardoso, C., Bartz, J.A., Yee, J.R., van Zuiden, M., 2013. The role of oxytocin in social bonding, stress regulation and mental health: an update on the moderating effects of context and interindividual differences. Psychoneuroendocrinology 38, 1883–1894.
Paziewska, A., Dabrowska, M., Goryca, K., Antoniewicz, A., Dobruch, J., Mikula, M., Jarosz, D., Zapala, L., Borowka, A., Ostrowski, J., 2014. DNA methylation status is more reliable than gene expression at detecting cancer in prostate biopsy. Br. J. Cancer 111, 781–789.
Pedersen, B.S., Schwartz, D.A., Yang, I.V., Kechris, K.J., 2012. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. Bioinformatics 28, 2986–2988.
Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., Kokmen, E., 1999. Mild cognitive impairment: clinical characterization and outcome. Arch. Neurol. 56, 303–308.
Phipson, B., Maksimovic, J., Oshlack, A., 2015. missMethyl: an R package for analyzing data from illumina's HumanMethylation450 platform. Bioinformatics 32, btv560.
Pidsley, R., Wong, C.C.Y., Volta, M., Lunnon, K., Mill, J., Schalkwyk, L.C., 2013. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics 14, 293.
Price, M.E., Cotton, A.M., Lam, L.L., Farré, P., Emberly, E., Brown, C.J., Robinson, W.P., Kobor, M.S., 2013. Additional annotation enhances potential for biologically-relevant analysis of the illumina infinium HumanMethylation450 BeadChip array. Epigenetics Chromatin 6, 4.
R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/.
Rossum, G. van, Boer, J. de, 1991. Interactively testing remote servers using the Python programming language. CWI Q. 4, 283–303.
Rye, P.D., Booij, B.B., Grave, G., Lindahl, T., Kristiansen, L., Andersen, H.M., Horndalsveen, P.O., Nygaard, H.A., Naik, M., Hoprekstad, D., Wetterberg, P., Nilsson, C., Aarsland, D., Sharma, P., Lönneborg, A., 2011. A novel blood test for the early detection of Alzheimer's disease. J. Alzheimers Dis. 23, 121–129.
Saffari, A., Silver, M.J., Zavattari, P., Moi, L., Columbano, A., Meaburn, E.L., Dudbridge, F., 2018. Estimation of a significance threshold for epigenome-wide association studies. Genet. Epidemiol. 42, 20–33.
Shackelford, D.A., 2006. DNA end joining activity is reduced in Alzheimer's disease. Neurobiol. Aging 27, 596–605.
Šidák, Z., 1967. Rectangular confidence regions for the means of multivariate normal distributions. J. Am. Stat. Assoc. 62, 626–633.
Simmons, A., Westman, E., Muehlboeck, S., Mecocci, P., Vellas, B., Tsolaki, M., Kåoszewska, I., Wahlund, L.O., Soininen, H., Lovestone, S., Evans, A., Spenger, C., 2011. The AddNeuroMed framework for multi-centre MRI assessment of Alzheimer's disease: experience from the first 24 months. Int. J. Geriatr. Psychiatry 26, 75–82.
Smith, R.G., Pishva, E., Shireby, G., Smith, A.R., Roubroeks, J.A.Y., Hannon, E., Wheildon, G., Mastroeni, D., Gasparoni, G., Riemenschneider, M., Giese, A., Sharp, A.J., Schalkwyk, L., Haroutunian, V., Viechtbauer, W., van den Hove, D.L.A., Weedon, M., Francis, P.T., Thomas, A.J., Love, S., Morgan, K., Walter, J., Coleman, P.D., Bennett, D.A., De Jager, P.L., Mill, J., Lunnon, K., 2020. Meta-analysis of epigenome-wide association studies in Alzheimer's disease highlights novel differentially methylated loci across cortex. BioRxiv. https://doi.org/10.1101/2020.02.28.957894.
Smith, A.R., Smith, R.G., Pishva, E., Hannon, E., Roubroeks, J.A.Y., Burrage, J., Troakes, C., Al-Sarraj, S., Sloan, C., Mill, J., Van Den Hove, D.L., Lunnon, K., 2019. Parallel profiling of DNA methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in Alzheimer's disease. Clin. Epigenetics 11, 52.
Smith, R.G., Hannon, E., De Jager, P.L., Chibnik, L., Lott, S.J., Condliffe, D., Smith, A.R., Haroutunian, V., Troakes, C., Al-Sarraj, S., Bennett, D.A., Powell, J., Lovestone, S., Schalkwyk, L., Mill, J., Lunnon, K., 2018. Elevated DNA methylation across a 48-kb region spanning the HOXA gene cluster is associated with Alzheimer's disease neuropathology. Alzheimer's Dement 14, 1580–1588.

Thach, D.C., Lin, B., Walter, E., Kruzelock, R., Rowley, R.K., Tibbetts, C., Stenger, D.A., 2003. Assessment of two methods for handling blood in collection tubes with RNA stabilizing agent for surveillance of gene expression profiles with high density microarrays. J. Immunol. Methods 283, 269–279.

Tukey, J.W., 1949. Comparing individual means in the analysis of variance. Biometrics 5, 99–114.

Varley, K.E., Gertz, J., Bowling, K.M., Parker, S.L., Reddy, T.E., Pauli-Behn, F., Cross, M.K., Williams, B.A., Stamatoyannopoulos, J.A., Crawford, G.E., Absher, D.M., Wold, B.J., Myers, R.M., 2013. Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res. 23, 555–567.

Vartanian, K., Slottke, R., Johnstone, T., Casale, A., Planck, S.R., Choi, D., Smith, J.R., Rosenbaum, J.T., Harrington, C.A., 2009. Gene expression profiling of whole blood: comparison of target preparation methods for accurate and reproducible microarray analysis. BMC Genomics 10, 2.

Vourekas, A., Zheng, K., Fu, Q., Maragkakis, M., Alexiou, P., Ma, J., Pillai, R.S., Mourelatos, Z., Jeremy Wang, P., 2015. The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing. Genes Dev. 29, 617–629.

Wang, S.C., Oeize, B., Schumacher, A., 2008. Age-specific epigenetic drift in late-onset Alzheimer's disease. PLoS One 3, e2698.

Watson, C.T., Roussos, P., Garg, P., Ho, D.J., Azam, N., Katsel, P.L., Haroutunian, V., Sharp, A.J., 2016. Genome-wide12 DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease. Genome Med. 8, 5.

Westman, E., Simmons, A., Muehlboeck, J.S., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Weiner, M.W., Lovestone, S., Spenger, C., Wahlund, L.O., 2011. AddNeuroMed and ADNI: similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America. Neuroimage 58, 818–828.

Yaffe, K., Vittinghoff, E., Lindquist, K., Barnes, D., Covinsky, K.E., Neylan, T., Kluse, M., Marmar, C., 2010. Posttraumatic stress disorder and risk of dementia among US veterans. Arch. Gen. Psychiatry 67, 608–613.

Yang, X., Cheng, Y., Lu, Q., Wei, J., Yang, H., Gu, M., 2015. Detection of stably expressed piRNAs in human blood. Int. J. Clin. Exp. Med. 8, 13353–13358.

Zhang, L., Li, H., Hu, X., Benedek, D.M., Fullerton, C.S., Forsten, R.D., Naifeh, J.A., Li, X., Wu, H., Benevides, K.N., Le, T., Smerin, S., Russell, D.W., Ursano, R.J., 2015. Mitochondria-focused gene expression profile reveals common pathways and CPT1B dysregulation in both rodent stress model and human subjects with PTSD. Transl. Psychiatry 5, e580.

Zhao, X., Qureshi, F., Eastman, P.S., Manning, W.C., Alexander, C., Robinson, W.H., Hesterberg, L.K., 2012. Pre-analytical effects of blood sampling and handling in quantitative immunoassays for rheumatoid arthritis. J. Immunol. Methods 378, 72–80.

Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T.Y., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., Gnirke, A., Meissner, A., 2013. Charting a dynamic DNA methylation landscape of the human genome. Nature 500, 477–481.

APPENDIX C: ALZHEIMER'S DISEASE-ASSOCIATED (HYDROXY)METHYLOMIC CHANGES IN THE BRAIN AND BLOOD

Lardenoije et al. Clinical Epigenetics (2019) 11:164
https://doi.org/10.1186/s13148-019-0755-5

Clinical Epigenetics

RESEARCH Open Access

Alzheimer's disease-associated (hydroxy)methylomic changes in the brain and blood

Roy Lardenoije[1,2†], Janou A. Y. Roubroeks[1,3†], Ehsan Pishva[1,3†], Markus Leber[4], Holger Wagner[5], Artemis Iatrou[1], Adam R. Smith[3], Rebecca G. Smith[3], Lars M. T. Eijssen[1,6], Luca Kleineidam[4,5,7], Amit Kawalia[5], Per Hoffmann[8,9,10], Tobias Luck[11], Steffi Riedel-Heller[11], Frank Jessen[7,12], Wolfgang Maier[5,7], Michael Wagner[5,7], René Hurlemann[13], Gunter Kenis[1], Muhammad Ali[1,14], Antonio del Sol[14,15,16,17], Diego Mastroeni[1,18,19], Elaine Delvaux[18,19], Paul D. Coleman[18,19], Jonathan Mill[3,20], Bart P. F. Rutten[1], Katie Lunnon[3], Alfredo Ramirez[4,5†] and Daniël L. A. van den Hove[1,21*†]

Abstract

Background: Late-onset Alzheimer's disease (AD) is a complex multifactorial affliction, the pathogenesis of which is thought to involve gene-environment interactions that might be captured in the epigenome. The present study investigated epigenome-wide patterns of DNA methylation (5-methylcytosine, 5mC) and hydroxymethylation (5-hydroxymethylcytosine, 5hmC), as well as the abundance of unmodified cytosine (UC), in relation to AD.

Results: We identified epigenetic differences in AD patients ($n = 45$) as compared to age-matched controls ($n = 35$) in the middle temporal gyrus, pertaining to genomic regions close to or overlapping with genes such as *OXT* (− 3.76% 5mC, $p_{\check{S}id\acute{a}k} = 1.07E{-}06$), *CHRNB1* (+ 1.46% 5hmC, $p_{\check{S}id\acute{a}k} = 4.01E{-}04$), *RHBDF2* (− 3.45% UC, $p_{\check{S}id\acute{a}k} = 4.85E{-}06$), and *C3* (− 1.20% UC, $p_{\check{S}id\acute{a}k} = 1.57E{-}03$). In parallel, in an independent cohort, we compared the blood methylome of converters to AD dementia ($n = 54$) and non-converters ($n = 42$), at a preclinical stage. DNA methylation in the same region of the *OXT* promoter as found in the brain was found to be associated with subsequent conversion to AD dementia in the blood of elderly, non-demented individuals (+ 3.43% 5mC, $p_{\check{S}id\acute{a}k} = 7.14E{-}04$).

Conclusions: The implication of genome-wide significant differential methylation of *OXT*, encoding oxytocin, in two independent cohorts indicates it is a promising target for future studies on early biomarkers and novel therapeutic strategies in AD.

Keywords: Alzheimer's disease, Epigenetics, DNA methylation, DNA hydroxymethylation, Brain, Middle temporal gyrus, Blood

Background

The neuropathological cascade of the world's leading cause of dementia, late-onset Alzheimer's disease (AD), is characterized by the progressive accumulation of extracellular amyloid plaques and intracellular neurofibrillary tangles, followed by neuronal cell death. The susceptibility to AD is determined by the complex interaction of genetic, environmental, and life-style factors, as well as epigenetic factors. Genetic research has been successful in identifying genetic variants modulating susceptibility to AD, including the first and strongest genetic risk factor for AD in the *APOE* gene. In addition to *APOE*, large-scale genome-wide association studies looking at AD have identified a number of independent common variants with a small-to-modest effect size [1]. Besides genetics, recent studies have suggested an

* Correspondence: d.vandenhove@maastrichtuniversity.nl
†Roy Lardenoije, Janou A. Y. Roubroeks, Ehsan Pishva, Alfredo Ramirez and Daniël L. A. van den Hove contributed equally to this work.
[1]School for Mental Health and Neuroscience (MHeNS), Department of Psychiatry and Neuropsychology, Maastricht University, P.O. Box 616, 6200, MD, Maastricht, the Netherlands
[21]Department of Psychiatry, Psychosomatics and Psychotherapy, University of Würzburg, Würzburg, Germany
Full list of author information is available at the end of the article

important role for epigenetic mechanisms in the etiology of AD [2], with reports of both global and gene-specific alterations in epigenetic modifications [3–6].

Several types of epigenetic DNA modifications have been described, including DNA methylation (5-methylcytosine, 5mC) and DNA hydroxymethylation (5-hydroxymethylcytosine, 5hmC). While the best studied epigenetic DNA modification, 5mC, plays an important gene regulatory role in most tissues, 5hmC seems to have a different impact on gene expression and is particularly enriched in the brain [7, 8], where it may play an important role in learning and memory [9, 10]. Unfortunately, conventional bisulfite (BS) conversion, a widely used procedure when quantifying DNA methylation, does not distinguish between 5mC and 5hmC. However, combining measurements from BS- and oxidative BS (oxBS)-converted DNA now allows for the quantification of both 5mC and 5hmC levels (Fig. 1).

Where genetic factors can identify persons at risk for developing AD from birth, epigenetic markers may offer more dynamic views on trajectories of biological change and may therefore be able to offer an improved, chronological insight into the sequence of events at different stages of AD. As brain tissue cannot be readily sampled in living humans, blood may offer an alternative. Available research on the blood DNA methylome in relation to AD is limited and mainly focuses on the direct comparison of AD cases and healthy controls [3, 11, 12]. Identifying disease-predicting biological profiles at pre-dementia stages of AD may provide improved precision in predicting onset of dementia and give potential treatments a better timeframe to successfully impede, or even halt disease progression [13, 14].

In the present study, we explored the association between AD and epigenetic dysregulation by quantifying 5mC and 5hmC, as well as unmodified cytosine (UC) proportions [15], at a single-site resolution in middle temporal gyrus (MTG) tissue obtained from AD patients ($n = 45$) and elderly, non-demented controls ($n = 35$; see Table 1 and the "Materials and methods" section for detailed demographics) [16]. This brain region was selected as the MTG is known as a site of early AD pathology [17], and differences in global levels of DNA methylation and hydroxymethylation have previously been reported in this brain region in AD [18]. While informative on its own, the inclusion of UC measurements also allows us to better compare our findings with previous studies using conventional BS conversion, since UC is determined by subtracting the BS signal (5mC + 5hmC) from 1 (Fig. 1). Even though the effects will be opposite from directly using the BS signal, incorporating UC in our study represents a crucial legacy analysis that enables the comparison with previous studies solely relying on the BS signal. Moreover, mechanistically, as an example, the affinity of a transcription factor may be different in the presence of UC, 5mC, or 5hmC, implicating that differential levels of UC (in the absence of significantly different 5mC or 5hmC levels) may have direct functional implications on gene expression. We followed up the brain analysis exploring DNA methylation in whole blood in an independent cohort, including samples from AD-converters and non-converters at two time points, before (54 converters, 42 controls) and after (41 converters, 42 controls) conversion to clinical AD (see Table 2 and the "Materials and methods" section for detailed demographics). Blood DNA methylomic markers were measured using only BS-converted DNA, as 5hmC has a very low prevalence in blood [8].

Results

Middle temporal gyrus

Site-specific 5mC, 5hmC, and UC levels were determined for the MTG using Illumina's Infinium HumanMethylation450K microarray (HM 450K array) with BS and oxBS-converted DNA (Fig. 1; see Tables 1 and 2 for cohort demographics). An epigenome-wide association study (EWAS) was performed for each DNA modification to identify the association with AD. The adjusted linear models showed no signs of inflation (all lambda values were between 0.95 and 1.05; see Additional file 2: Figure S1 for QQ plots). None of the AD-associated CpG sites in the MTG passed false discovery rate (FDR) correction (Additional file 1: Tables S1–S3).

A structural and functional genomic annotation enrichment analysis on the 1000 highest ranked sites indicated a significant enrichment of several CpG island features, gene features, and alternative transcription events. This included an enrichment of mainly gene body sites for the 5mC (fold enrichment = 1.42, $p = 1.17E{-}10$) and 5hmC (fold enrichment = 1.17, $p = 3.64E{-}03$) results and mainly intergenic sites for the UC (fold enrichment = 1.59, $p = 1.67E{-}09$) results (Additional file 2: Figure S8; Additional file 1: Table S7).

A regional analysis, looking at the spatial correlation of adjacent modified positions, detected 1 differentially methylated region (DMR), 1 differentially hydroxymethylated region (DHR), and 11 differentially unmodified regions (DURs) that were associated with AD in the MTG (Table 3; Additional file 2: Figure S3). Analysis of MTG expression data of genes annotated to DMRs, DHRs, and DURs showed a significant negative correlation between a DUR associated with *RHBDF2* and *RHBDF2* RNA expression ($\rho = -0.39$, $p_{FDR} = 4.37E{-}03$) (Additional file 1: Table S10). Of note, although the DHR residing in the transcription start site (TSS) of *CHRNB1*, of which all probes show hyperhydroxymethylation in the AD cases, did not correlate with *CHRNB1* mRNA expression ($\rho = -0.09$, $p_{FDR} > 0.05$), a linear regression analysis of regressed MTG expression data of *CHRNB1* showed a significant elevation of *CHRNB1* mRNA levels in AD cases (estimate = 0.13,

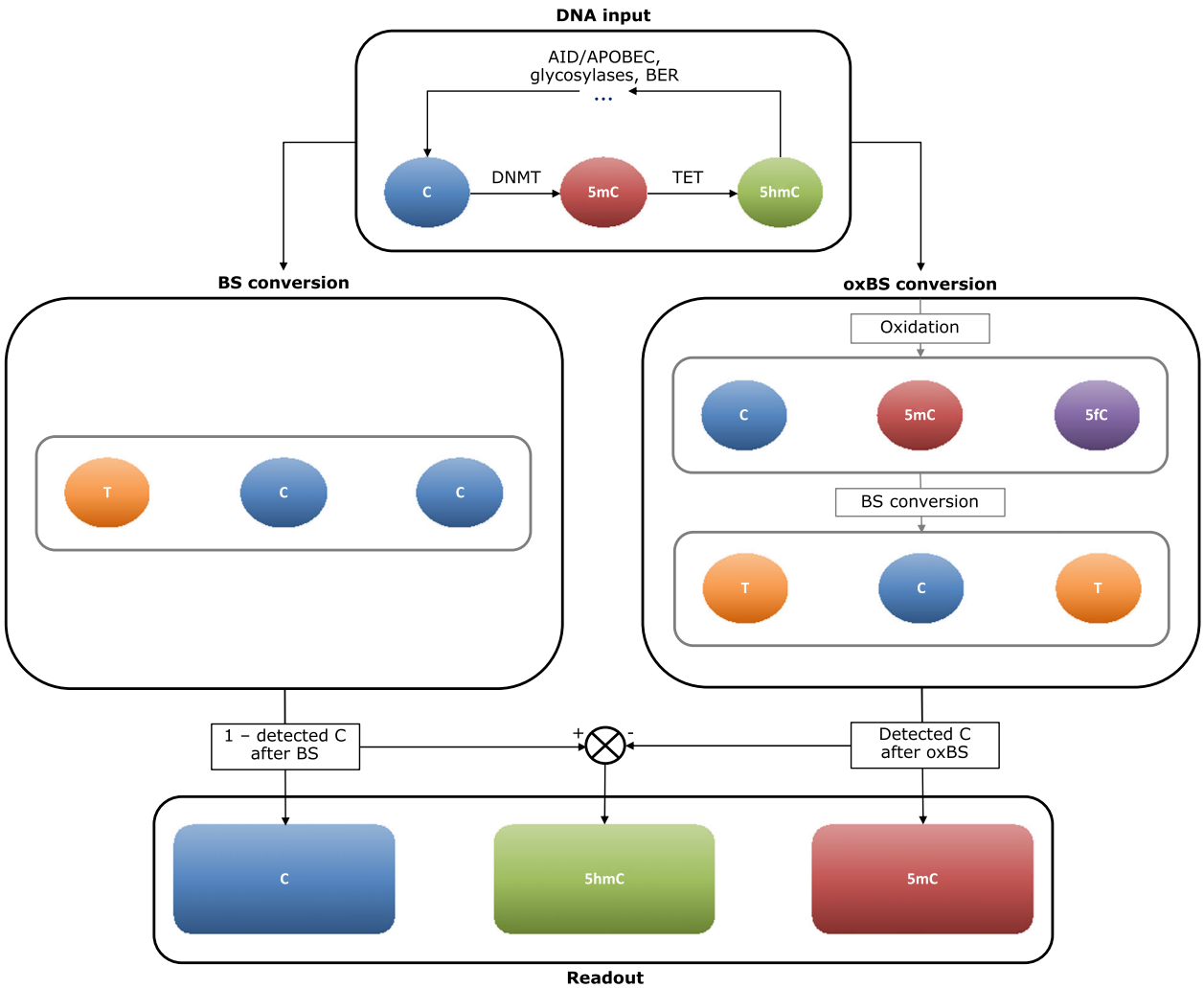

Fig. 1 Overview of the procedure to detect unmodified cytosines (C), 5-methylcytosine (5mC), and 5-hydroxymethylcytosine (5hmC). Naturally, C can be converted to 5mC by DNA methyltransferases (DNMTs) and 5mC can be oxidized by ten-eleven translocation (TET) enzymes, resulting in 5hmC. There are several proposed demethylation pathways through which 5mC and 5hmC can be converted back to C. DNA samples were split in two, one half was only treated with bisulfite (BS), which converts C into thymine (T). 5mC and 5hmC are protected against this conversion, and will be read as a C on the array. The detected C signal after BS conversion is thus actually the combined 5mC and 5hmC signal. As the signals are converted to fractions, with C + 5mC + 5hmC = 1, the fraction of C in the input DNA can be determined by subtracting the C signal after BS conversion (representing the combined 5mC and 5hmC fraction in the input DNA) from 1. The other half of the DNA sample was first oxidized, which converts 5hmC into 5-formylcytosine (5fC), and then treated with BS. 5fC is not protected against the BS conversion, so it also turns into T. C detected on the array after this oxidative BS (oxBS) conversion thus represents the fraction of 5mC in the input DNA. The 5hmC fraction in the input DNA can be determined by subtracting the fraction of 5mC (detect C after oxBS) from the combined 5mC and 5hmC fraction (detected C after BS). This procedure results in three readout signals: unmodified C, 5mC, and 5hmC. Note that 5fC, and probably also 5-carboxylcytosine, are included in the unmodified C fraction.

p = 1.37E−04) (Additional file 2: Figure S4). For a full transcriptomic investigation of the MTG cohort used in the present study, see the recent publication of Piras et al. [19].

Next, a gene regulatory network (GRN) analysis was performed with the unique genes annotated to the 1000 highest ranked probes. Because of different numbers of associated genes from each dataset, we obtained contextualized networks with varying number of interactions. The number of interactions in the contextualized GRNs representing the differential 5mC, 5hmC, and UC MTG states were 325, 398 and 244, respectively. Differential GRN analysis identified several candidate genes highly influential in the simulated transition from a diseased towards a healthy phenotype. Based on a score indicating for each gene, when changed, the number of other genes in the network that were predicted to show altered expression, *IL6* (score = 55), *SIAH1* (score = 78), and *EGF* (score = 55) were found to be the most influential in the 5mC, 5hmC, and UC networks, respectively (Additional file 1: Table S9).

Table 1 Cohort demographics—brain tissue

	AD patients	Non-demented controls
N	45	35
Gender(m/f)	22/23	17/18
Age of death (mean ± SD)	85.09 (6.24)	84.46 (5.50)
PMI (Mean ± SD)	2.77 (0.69)	2.87 (1.03)
Plaque total (mean ± SD)	12.97 (2.25)	4.65 (4.30)
Tangle total (mean ± SD)	11.02 (4.16)	3.96 (2.10)
Braak stage (range (median))	II–VI (V)	I–IV (III)

The brain tissue cohort consisted of 80 middle temporal gyrus (MTG) tissue samples obtained from the Banner Sun Health Research Institute (Sun City, AZ, USA), from which HM 450K array BS and oxBS data was generated. Displayed is the number of samples in each group and the distributions of gender, age, postmortem interval (hours), Braak stage, and plaque and tangle total (the sum of average Aβ plaque densities and tangle densities (resp.) in the entorhinal cortex, hippocampus, parietal lobe cortex, temporal lobe cortex and frontal lobe cortex)

Blood

Since 5hmC is not enriched in the blood, only BS conversion was used to measure site-specific 5mC levels, also with the HM 450K array. A blood EWAS investigating the association between DNA methylation and conversion to AD was performed at baseline and at follow-up, leading to the identification of 3 differentially methylated positions at baseline and 266 at follow-up (Additional file 1: Tables S4–S6). No significant inflation was detected (Additional file 2: Figure S2; see the "Materials and methods" section for details).

Genomic annotation enrichment analysis of the top sites in blood showed enrichment of mainly intergenic sites (fold enrichment = 1.32, $p = 5.80E{-}04$) at baseline and proximal promoters (fold enrichment = 0.79, $p = 1.60E{-}04$) at follow-up (Additional file 2: Figure S9; Additional file 1: Table S8).

Table 2 Cohort demographics—blood samples

	Controls	Converters
Baseline (T1)		
N	42	54
Gender (m/f)	10/32	17/34
Age at baseline (mean ± SD)	81.00 ± 3.11	82.31 ± 3.55
APOE4 carriers	43%	43%
Follow-up (T2)		
N	42	41
Gender (m/f)	10/32	13/28
Age at baseline (mean ± SD)	81.00 ± 3.11	82.01 ± 3.51
APOE4 carriers	43%	41%

Blood samples were obtained from the German Study on Ageing, Cognition and Dementia in Primary Care Patients (AgeCoDe) cohort, and HM 450K array BS data was generated. The cohort includes controls, who showed no signs at baseline or follow-up, and converters who showed no signs of dementia at baseline, but were diagnosed with AD dementia at follow-up. DNA samples were collected at baseline and follow-up for both groups. Displayed is the number of samples in each group, the distributions of Gender and Age at baseline, and the percentage of APOE ε4 allele carriers

The regional analysis found 15 and 21 DMRs associated with conversion to AD at baseline and follow-up, respectively (Table 4; Additional file 2: Figure S5).

GRNs representing the blood baseline and follow-up states contained 475 and 277 interactions, respectively. Differential GRN analysis identified *WNT3A* (score = 50) as the most influential gene in the baseline network, and *SHH* (score = 33) in the follow-up network (Additional file 1: Table S9).

Overlap

Only 1 blood DMR, close to *GLIPR1L2*, showed hypermethylation in relation to AD conversion at both the baseline (+2.72%, $p_{\check{S}id\acute{a}k} = 1.40E{-}04$) and follow-up (+ 1.34%, $p_{\check{S}id\acute{a}k} = 6.94E{-}06$) time points. Extracting the probes located in this blood *GLIPR1L2* DMR from the MTG EWAS for comparison showed, in AD cases, lower UC levels (9/10 probes with negative log2 fold change [logFC]), mixed changes for 5mC (6/10 probes with positive logFC), and lower 5hmC levels for the probes that passed the detection threshold (2/2 probes with negative logFC). Even though the UC observations in the MTG are in line with the blood findings, only for one UC probe (cg07311024) the change was nominally significant (logFC = − 0.01, $p = 3.88E{-}02$). A targeted linear regression analysis of the regressed MTG expression data of *GLIPR1L2* showed a significant decrease in AD cases (estimate = -0.10, $p = 3.12E{-}04$) (Additional file 2: Figure S6).

Interestingly, close to the TSS of *OXT*, we observed a DMR which was detected both in the MTG (− 3.76%, $p_{\check{S}id\acute{a}k} = 1.07E{-}06$), as well as in the blood dataset (at baseline, + 3.43%, $p_{\check{S}id\acute{a}k} = 7.14E{-}04$) (see Additional file 2: Figure S7 for the probe positions of both *OXT* DMRs). MTG *OXT* methylation across Braak stages, as a proxy indicator of disease progression, is displayed in Fig. 2 and suggest *OXT* hypermethylation towards Braak 3-4 stages and *OXT* hypomethylation during later stages. Moreover, in the differential GRN analysis, *OXT* came forward as an influential gene. In case of the 5mC and 5hmC MTG states, a change in *OXT* was predicted to alter the expression of 39 and 54 other genes in the networks, respectively, and in the blood baseline state, *OXT* was predicted to alter 41 genes in the network (Additional file 2: Figures S10–S12; Additional file 1: Table S9).

Discussion

For the current study, we aimed to identify AD-related changes in epigenetic DNA modifications, comparing brain tissue from AD patients and age-matched controls. In addition, we explored DNA methylation in blood samples from AD-converters and non-converters, both at a preclinical stage and after conversion, identifying an AD-associated DMR in *OXT* in both the brain and blood datasets.

Table 3 Differentially methylated, hydroxymethylated, and unmodified regions in the middle temporal gyrus

Gene	Position	Gene feature	*n*	*p* value	Šidák *P*	Average Δ% (range Δ%)
5mC						
OXT	chr20:3051954-3052484	TSS; Intron; 5′UTR; CDS	10 (0 up; 10 down)	1.43E−09	1.07E−06	− 3.76 (− 6.94:− 0.43)
5hmC						
CHRNB1	chr17:7348322-7348439	TSS; Exon; 5′UTR	5 (5 up; 0 down)	2.63E−07	4.01E−04	1.46 (0.70:1.96)
UC						
ACTR3C; *LRRC61*	chr7:150019955-150020946	TSS; Intron; Exon; 5′UTR	17 (1 up; 16 down)	3.54E−12	1.42E−09	− 0.57 (− 1.34:0.02)
RHBDF2	chr17:74475240-74475403	Intron; CDS	5 (0 up; 5 down)	1.99E−09	4.85E−06	− 3.45 (− 4.71:− 1.42)
TMC8	chr17:76128522-76128907	Intron; CDS	8 (0 up; 8 down)	3.29E−09	3.39E−06	− 1.26 (− 2.84:− 0.26)
ASPG	chr14:104551867-104552210	TSS; Intron; 5′UTR; CDS	5 (0 up; 5 down)	1.00E−08	1.16E−05	− 1.21 (− 2.49:− 0.28)
PIEZO1	chr16:88844969-88845205	Intron	3 (0 up; 3 down)	1.87E−07	3.14E−04	− 3.08 (− 3.76:− 2.32)
VWA7	chr6:31734106-31734472	Intron; CDS	10 (10 up; 0 down)	2.04E−07	2.21E−04	3.39 (2.24:4.23)
CLMAT3; *SPARC*	chr5:151066460-151066731	Exon; TSS; 5′UTR	6 (0 up; 6 down)	5.21E−07	7.62E−04	− 0.29 (− 0.64:0.21)
KIAA1522	chr1:33231070-33231314	TSS; Exon; 5′UTR; Intron	6 (0 up; 6 down)	8.48E−07	1.38E−03	− 1.85 (− 2.43:− 1.3)
C3	chr19:6713227-6713460	Intron; CDS	3 (1 up; 2 down)	9.21E−07	1.57E−03	− 1.20 (− 2.1:0.46)
PRSS22	chr16:2908157-2908246	TSS; Exon; 5′UTR	4 (0 up; 4 down)	1.02E−06	4.52E−03	− 1.56 (− 1.91:− 1.39)
FRAT1	chr10:99080756-99081017	Exon	3 (3 up; 0 down)	1.50E−06	2.28E−03	2.34 (1.57:3.03)

Differentially methylated (5mC), hydroxymethylated (5hmC), and unmodified (UC) regions in a comparison of Alzheimer's disease patients ($n = 45$) and controls ($n = 35$). Displayed for each region is the UCSC gene name, chromosomal position (genome build 37), gene feature (TSS, transcription start site; 5′UTR, 5′ untranslated region; CDS, coding sequence), number of probes in region and number of upregulated and downregulated probes (*n*), *p* value and multiple testing-corrected *p* (Šidák-*P*), and average change in beta value (Alzheimer's disease - control), including the range of the probe differences

The DHR identified in the MTG resided in the promotor of *CHRNB1*, which encodes acetylcholine receptor subunit beta and is important for cholinergic neurotransmission. In combination with the observed increased levels of *CHRNB1* mRNA in the MTG, this potentially reflects a compensatory mechanism to maintain acetylcholine signaling in AD. Indeed, the acetylcholine-related pathway is known to be altered in AD and, as such, remains an important target in the development of novel treatment options [20]. Previous epigenomic studies of AD using standard BS-conversion have found associations between AD and *RHBDF2* methylation in multiple cortical regions [3, 4]. We replicated these findings; observing an AD-associated DUR in *RHBDF2*, which included the previously detected CpG sites (cg13076843, cg05810363, and cg12163800) and showed the same direction of effect as previously reported. For instance, using conventional bisulfite (BS) conversion, a 3.36% increase in DNA methylation level of cg05810363 has been observed across cortical regions in association with AD neuropathology [3]. Interestingly, a negative correlation between UC levels within the *RHBDF2* DUR and *RHBDF2* mRNA expression was observed in the MTG. *RHBDF2* is thought to be important for the release of tumor necrosis factor, a major inflammatory cytokine associated with neuroinflammation observed in AD [21, 22]. *C3*, another gene with an AD-associated DUR, encodes a central component of the complement system and mediates developmental synapse elimination by phagocytic microglia. *C3* has previously been implicated in mediating synaptic loss in the early stages of AD [23].

The top DMR from the baseline blood analysis, showing hypermethylation in AD, is close to the *LDLRAD4* gene. This gene has previously been associated with schizophrenia and blood pressure and is thought to suppress transforming growth factor (TGF)-β signaling [24–27]. TGF-β is an inflammatory cytokine playing a role in cell survival and synaptic transmission, and various isoforms have been associated with AD [28]. Additional baseline blood DMRs were close to *TENM3*, involved in neurite growth [29], *SYMPK*, involved in polyadenylation regulation of gene expression and which showed increased expression in AD [30], *SLC44A4*, associated with type 1 diabetes mellitus and human aging [31], *ZMAT2*, which had decreased expression in AD [32], *ULK1*, which may play a role in the autophagic degradation of amyloid beta (Aβ) [33], and *RUNX2*, which links bone health and cognitive function and anxiety-like behavior [34]. The DMR that was found both at baseline and follow-up is associated with *GLIPR1L2*. *GLIPR1L2* also showed decreased expression in the MTG. The function of this gene is not well known, but it may play a role in tumor suppression and immune function [35, 36]. The top AD-associated blood DMR at follow-up, showing hypomethylation, is located in *GSDMD*, which encodes a critical factor in pyroptosis; a form a cell death that may be triggered by Aβ [37, 38]. Other genes with a nearby AD-associated blood DMR at

Table 4 Differentially methylated regions in blood

Gene	Position	Gene feature	*n*	*p* value	Šidák *P*	Average Δ% (range Δ%)
Baseline						
LDLRAD4	chr18:13611370-13611825	TSS; Exon; 5′UTR; Intron	7 (7 up; 0 down)	3.25E−11	2.88E−08	3.64 (2.54:5.14)
ZNF154	chr19:58220080-58220838	TSS; Intron; 5′UTR; CDS; Exon	11 (11 up; 0 down)	1.16E−09	6.16E−07	3.87 (1.75:5.11)
PRRT1	chr6:32116216-32117402	Intron; 3′UTR; CDS	26 (24 up; 2 down)	2.61E−09	8.85E−07	1.77 (− 3.1:5.37)
SYMPK; RSPH6A	chr19:46318514-46319399	Intron; 3′UTR; CDS; TSS; Exon; 5′UTR	7 (7 up; 0 down)	6.18E−09	2.81E−06	3.2 (2.07:4.87)
TENM3	chr4:183728549-183729462	Intergenic	5 (5 up; 0 down)	1.02E−08	4.48E−06	2.69 (− 0.67:4.25)
GLIPR1L2; CAPS2	chr12:75784617-75785296	TSS; Intron; 5′UTR; CDS; Exon	10 (10 up; 0 down)	2.36E−07	1.40E−04	2.72 (1.85:4.5)
GPR35	chr2:241562085-241562758	Intron	6 (6 up; 0 down)	3.16E−07	1.89E−04	3.83 (3.35:4.37)
ZMAT2	chr5:140079591-140080246	TSS; Intron; 5′UTR; CDS	10 (9 up; 1 down)	3.86E−07	2.37E−04	1.21 (− 1.35:2.81)
ZNF649-AS1; ZNF577	chr19:52390810-52391368	Exon; TSS; Intron; 5′UTR	10 (10 up; 0 down)	7.42E−07	5.35E−04	5.11 (2.93:6.61)
ULK1	chr12:132380696-132380904	Intron	3 (3 up; 0 down)	3.57E−07	6.90E−04	2.1 (0.92:3.5)
SLC44A4	chr6:31846769-31847029	TSS; Exon; 5′UTR	8 (0 up; 8 down)	4.61E−07	7.13E−04	− 1.14 (− 3.86:2.41)
OXT	chr20:3051954-3052484	TSS; Intron; 5′UTR; CDS	10 (9 up; 1 down)	9.41E−07	7.14E−04	3.43 (− 0.45:6.79)
FAM222A	chr12:110156245-110156460	Intron; 5′UTR	4 (4 up; 0 down)	2.85E−06	5.33E−03	3.11 (2.66:3.43)
CYBRD1	chr2:172430723-172430817	Intergenic	3 (0 up; 3 down)	1.83E−06	7.78E−03	− 1.75 (− 2.7:− 0.26)
RUNX2	chr6:45391852-45391974	Intron	3 (3 up; 0 down)	5.11E−06	1.67E−02	3.19 (0.74:4.53)
Follow-up						
GSDMD	chr8:144635309-144635611	TSS; Exon; 5′UTR	5 (0 up; 5 down)	7.08E−18	9.43E−15	− 0.68 (− 1.53:− 0.25)
IRGC	chr19:44203583-44203914	Intergenic	3 (0 up; 3 down)	2.40E−12	2.92E−09	0.87 (− 0.32:1.53)
LINC01149	chr6:31409319-31409758	Exon	12 (0 up; 12 down)	3.03E−10	2.78E−07	− 0.96 (− 2.17:− 0.12)
RUFY1	chr5:178986131-178986907	TSS; Exon; 5′UTR; Intron	9 (9 up; 0 down)	3.77E−09	1.96E−06	2.85 (1.3:4.19)
GLIPR1L2; CAPS2	chr12:75784541-75785296	TSS; Intron; 5′UTR; CDS; Exon	11 (11 up; 0 down)	1.30E−08	6.94E−06	1.34 (0.64:2.81)
RAD51B; LOC100996664	chr14:69095057-69095680	Intron; Exon	5 (5 up; 0 down)	3.57E−08	2.31E−05	6.05 (4.89:7.1)
LOC105372397; MAP4K1	chr19:39086733-39087768	Intron; Exon; CDS	5 (4 up; 1 down)	8.14E−08	3.17E−05	− 0.28 (− 1.05:0.3)
KHDRBS2	chr6:62996022-62996703	Exon; TSS; 5′UTR	11 (11 up; 0 down)	6.92E−08	4.09E−05	1.25 (0.69:1.89)
STAG3L5P-PVRIG2P-PILRB; STAG3L5P; PMS2P1	chr7:99933717-99933798	Exon	3 (3 up; 0 down)	1.01E−08	5.02E−05	0.16 (− 0.01:0.45)
ISOC2	chr19:55972646-55973339	TSS; Intron; Exon; 5′UTR	9 (9 up; 0 down)	8.68E−08	5.04E−05	1.52 (0.78:2.51)
RARRES2	chr7:150037988-150038599	Intron; 5′UTR	3 (0 up; 3 down)	1.63E−07	1.07E−04	0.71 (− 0.64:2.19)
LINC01169	chr15:66947171-66947618	Intron	5 (0 up; 5 down)	1.54E−07	1.39E−04	− 2.24 (− 5.41:− 0.83)
TRAM1L1	chr4:118006405-118007226	TSS; CDS	11 (10 up; 1 down)	9.95E−07	4.88E−04	1.14 (− 0.81:2.96)
LDHC	chr11:18433500-18434016	TSS; Intron; Exon; 5′UTR	7 (7 up; 0 down)	6.92E−07	5.40E−04	2.76 (− 0.3:4.7)
ZNF337-AS1; NANP	chr20:25604462-25605179	Intron; Exon; TSS; 5′UTR; CDS	12 (12 up; 0 down)	1.10E−06	6.18E−04	0.4 (− 0.27:1.36)

Table 4 Differentially methylated regions in blood *(Continued)*

Gene	Position	Gene feature	*n*	*p* value	Šidák *P*	Average Δ% (range Δ%)
IFT74; IFT74-AS1	chr9:26956236-26956770	Intron; 5′UTR; TSS; Exon	3 (2 up; 1 down)	8.65E−07	6.52E−04	0.21 (− 0.2:0.52)
MIR3659	chr1:38599626-38600200	Intergenic	4 (0 up; 4 down)	1.34E−06	9.38E−04	− 0.49 (− 1.35:0.64)
LINC01983	chr3:195578011-195578281	Intron	6 (0 up; 6 down)	6.38E−07	9.51E−04	− 0.07 (− 0.76:0.67)
HEXD	chr17:80393124-80393667	Intron; CDS	5 (5 up; 0 down)	1.89E−06	1.40E−03	− 0.52 (− 2.54:0.39)
GNG7	chr19:2543602-2544101	Intron; 5′UTR	5 (5 up; 0 down)	5.29E−06	4.26E−03	1 (0.25:1.95)
F11-AS1	chr4:187422005-187422120	Exon	5 (5 up; 0 down)	1.66E−06	5.81E−03	1.26 (0.4:1.87)

Differentially methylated regions in the comparisons of Alzheimer's disease converters and non-converters, prior to conversion at baseline (converters $n = 54$, non-converters $n = 42$), and after conversion at follow-up (converters $n = 41$, non-converters $n = 42$). Displayed for each region is the UCSC gene name, chromosomal position (genome build 37), gene feature (TSS, transcription start site; 5′UTR, 5′ untranslated region; CDS, coding sequence), number of probes in region and number of upregulated and downregulated probes (*n*), *p* value and multiple testing-corrected *p* (Šidák *P*), and average change in beta value (Alzheimer's disease converter - control), including the range of the probe differences

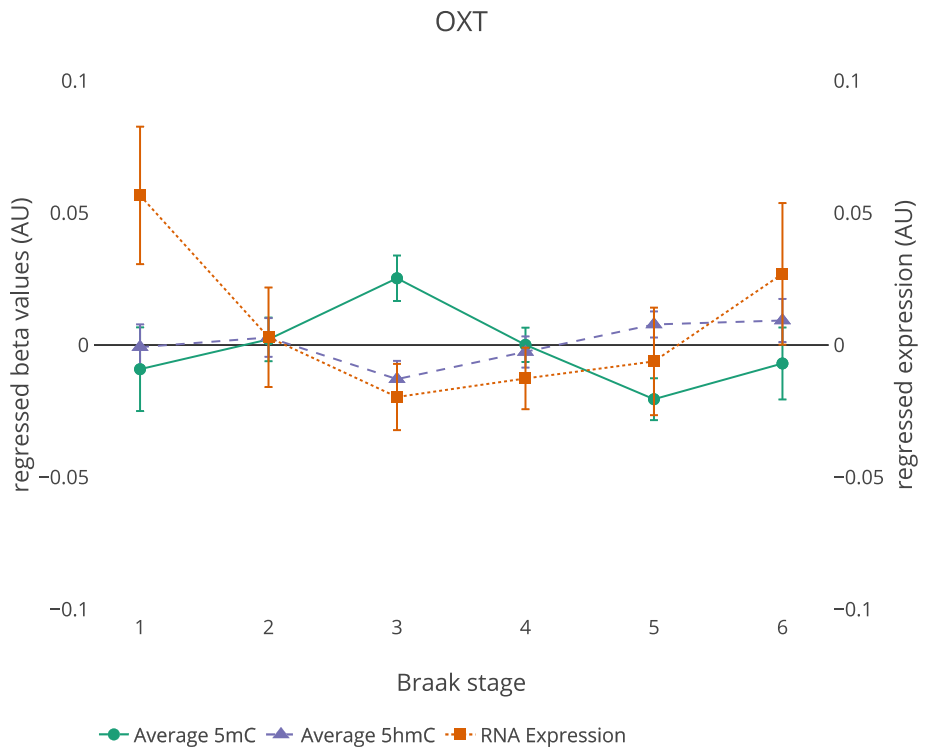

Figure 2. Methylation, hydroxymethylation and expression of *OXT* across Braak staging. Regressed *OXT* expression values and average regressed 5mC and 5hmC values of 10 and 9 overlapping probes within the *OXT* DMR are shown. Regressed values were generated by taking the residuals of a model fitted with the covariates age, gender, and 5 surrogate variables, but excluding the predictor of interest AD diagnosis. Error bars represent mean ± SEM. $N = 76$ for each line.

follow-up include *KHDRBS2*, previously identified in a genome-wide association interaction analysis in relation to AD [39], *RARRES2*, encoding an adipokine that has been linked to inflammation, obesity, diabetes, and cardiovascular diseases [40], and *GNG7*, for which Braak stage-associated differential methylation has been reported in cortical glial cells of AD patients [41].

Taken together, the observation of epigenetic modifications in several inflammation-associated genes in both brain and blood aligns with the amyloid cascade-inflammatory hypothesis of AD [42]. These findings could reflect either downstream effects resulting from the inflammatory activation seen in AD, or, particularly in the brain, reflect mediating effects of DNA modifications on inflammation as a causative factor. Exploring the exact nature of the AD-associated epigenetic modifications in inflammation-associated genes and the potential for blood biomarkers is thus a pivotal aim for future studies.

Strikingly, our methylomic profiling in MTG and whole blood both led to the identification of a common DMR associated with AD, close to the transcription start site of *OXT*. Our design allowed for the disentanglement of specific 5mC and 5hmC signals in the MTG, which, in the case of *OXT*, suggests they change in opposite directions in relation to AD. The detection of a DMR near *OXT* is in line with a recent report of a nearly identical AD-associated *OXT* DMR (containing 1 additional probe) in the superior temporal gyrus (STG) [43]. This area is located directly above the MTG. Furthermore, using GRN analysis addressing the overlap between the top influential genes in the networks and genes with significant differentially modified regions, we observed *OXT* to consistently appear as one of the most influential genes in both brain and blood GRNs. *OXT* encodes oxytocin, a neuropeptide involved in the neuromodulation of social behavior, stress regulation, and associative learning [44]. Interestingly, the functional impact of *OXT* promoter methylation at the same genomic locus has been recently shown [45]. It was linked to several measures of sociability, superior temporal sulcus activity during social cognition tasks, as well as fusiform gyrus gray matter volume, a brain region closely related to the MTG.

The paraventricular nucleus and supraoptic nucleus are thought to be the main sites of central oxytocin production [46], areas which reportedly undergo cell loss during AD [47]. The remaining neurons are thought to undergo a, potentially compensatory, hypertrophy. One might hypothesize that this activation could initially lead to higher than normal oxytocin levels, before synthesis collapses during the final stages of AD. Interestingly, enhanced levels of hippocampal oxytocin have been

associated with memory impairment, and AD-associated elevations of oxytocin have been reported in the hippocampus and temporal cortex [48]. There is also limited evidence oxytocin is reduced in cerebrospinal fluid of manifest AD [49]. Additionally, it has been suggested that co-damage to the locus coeruleus and hypothalamic nuclei could happen early during AD pathogenesis [50], substantiating the hypothesis that oxytocin could serve as an early diagnostic biomarker for AD.

In line with an AD-related increase in temporal cortex oxytocin levels [48], all ten CpG sites within the MTG *OXT* DMR showed decreased levels of methylation in AD cases. Conversely, we observed *OXT* hypermethylation in the DNA from peripheral blood of participants who progressed to dementia. The *OXT* blood DMR was not observed after conversion. Research has shown that independent mechanisms may be involved in peripheral and central regulation of *OXT* expression, supporting this apparent discrepancy observed in blood and brain [51]. Alternatively, these observations suggest there may be a temporal change in *OXT* methylation during AD progression. Looking at MTG *OXT* methylation across Braak stages appears to support the observation of *OXT* hypermethylation at earlier stages, as also seen in the blood, and *OXT* hypomethylation at more advanced stages. Notably, it has recently been reported that oxytocin administration was able to improve social cognition and behavior in frontotemporal dementia patients [52], illustrating the complex modulatory function of oxytocin in different brain regions and its potential use in the treatment of certain manifestations of dementia. Whether oxytocin represents a suitable therapeutic agent for AD remains to be elucidated.

Even though we detect several targets relevant in light of AD, a general lack of overlap between the different analyses presented here might be noted, an observation which is true for EWAS and epigenetics studies in AD in general. Others have discussed a myriad of possible reasons for discrepancies between studies, such as methodological differences, differences in tissue type and processing, study designs, and samples sizes [53]. In view of this, the detection of a common *OXT* DMR in two completely independent cohorts and two different types of tissue, further supported by a recent similar EWAS on the STG [43], makes it an even more promising target for future studies. However, the differences in direction of change and the *OXT* methylation pattern observed over Braak stages indicates these epigenetic changes should be further studied in a longitudinal fashion to establish a clear relationship with AD neuropathology, as well as clinical manifestations of AD.

Given the detection of several regions of interest, it should be noted that the lack of positions significantly associated with AD in the MTG after FDR correction may be the result of a limited sample size. Genome-wide site-specific AD-related epigenetic changes should thus be further investigated using studies with larger sample sizes or meta-analyses. Alternatively, future studies may focus on candidate genes identified in the present work, such as *OXT*.

Conclusions

Our novel approach confirms some previous epigenetic findings identified in the central nervous system, including *RHBDF2*, as well as revealed novel targets, such as in *CHRNB1*, involving dysregulated DNA hydroxymethylation. Furthermore, the nearly identical *OXT* DMRs found in both the blood and brain suggest a systemic epigenetic dysregulation in AD involving *OXT*. The detection of the *OXT* DMR at pre-dementia stages suggests its potential relevance as a novel biomarker and may offer new treatment strategies to be explored in future studies.

Materials and methods

Patients

Informed consent was obtained from all human participants. This includes donors of the Banner Sun Health Research Institute (BSHRI) Brain and Body Donation Program (BBDP), who signed an Institutional Review Board-approved informed consent form, including specific consent to the use of donated tissue for future research [16, 54]. The German Study on Ageing, Cognition and Dementia in Primary Care Patients (AgeCoDe) study protocol was approved by the local ethics committees at the University of Bonn (Bonn, Germany), the University of Hamburg (Hamburg, Germany), the University of Duesseldorf (Duesseldorf, Germany), the University of Heidelberg/Mannheim (Mannheim, Germany), the University of Leipzig (Leipzig, Germany), and the Technical University of Munich (Munich, Germany).

DNA from the MTG was obtained from 82 AD patients and neurologically normal control BBDP donors stored at the Brain and Tissue Bank of the BSHRI (Sun City, AZ, USA) [16, 54] (Table 1). The organization of the BBDP allows for fast tissue recovery after death, resulting in an average post-mortem interval of only 2.8 h for the included samples. Braak staging was carried out for AD neurofibrillary pathology. A consensus diagnosis of AD or non-demented control was reached by following National Institutes of Health AD Center criteria [54]. Comorbidity with any other type of dementia, cerebrovascular disorders, mild cognitive impairment (MCI), and presence of non-microscopic infarcts was applied as exclusion criteria. Although this may limit the generalizability of the current study, these strict exclusion criteria were applied to enhance the detection of AD-specific dysregulation, not confounded by common comorbidities. Detailed information about the BBDP has been reported elsewhere [16, 54].

AgeCoDe is a prospective longitudinal study including 3327 non-demented individuals at baseline, initiated to

investigate the early detection of MCI and dementia in primary care [55]. Participants were randomly selected from the general practice registry in six German cities and cognition was assessed at approximately 18-month intervals and 10-month intervals after visit 7, for up to 11 years after baseline. For this study, whole blood DNA was obtained from a subsample of 99 individuals aged above 75 years from this AgeCoDe cohort (Table 2). Of these, 42 were converters: they had no dementia at baseline, had DNA samples available at baseline and follow-up (after ~ 4.5 years), and had sufficient information available for a diagnosis of AD dementia to be made at the 4.5-year follow-up. There were 44 control subjects, who had to adhere to the same criteria, except that they should have no signs of dementia at neither baseline, nor the 4.5-year follow-up, and all subsequent cognitive assessments up to 11 years after baseline. The remaining 13 participants had not yet converted at the 4.5-year follow-up (when blood was drawn), but were diagnosed during a later follow-up, up to a maximum of 11 years after baseline [56]. These samples were grouped together with the other converters.

The groups were matched for age, gender, and *APOE* genotype. The presence of dementia was assessed in all subjects with the Structured Interview for Diagnosis of Dementia of Alzheimer Type, Multi-infarct Dementia, and Dementia of Other Etiology [57] based on the DSM-IV criteria. The dementia diagnosis in subjects who were not personally interviewed was based on the Global Deterioration Scale [58] (≥ 4) and the Blessed Dementia Rating subscales. The etiological diagnosis of AD was based on the criteria of the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association [59] for probable AD and was only assigned in case of sufficient information provided. All final diagnoses were a consensus between the interviewer and an experienced geriatrician or geriatric psychiatrist. More detailed information about the AgeCoDe cohort has been published previously [55, 56].

(Hydroxy)Methylomic profiling

For the BBDP samples, the TrueMethyl™ 24 Kit version 2.0 by CEGX™ (Cambridge Epigenetix Limited, Cambridge, UK) was used for BS and oxBS conversion of genomic DNA (gDNA) extracted from frozen MTG tissue. All laboratory procedures were performed at GenomeScan (GenomeScan B.V., Leiden, the Netherlands), without knowledge of the phenotypic characteristics of the samples and according to the manufacturer's instructions. Prior to conversion, high molecular weight (HMW) gDNA was quantified using a PicoGreen assay (Invitrogen, Carlsbad, CA, USA), and gel-electrophoresis was performed to assess gDNA quality. All samples were of sufficient quantity and quality. A volume of 1 μg HMW gDNA was used per sample, which, after purification and denaturation, was split into two samples that underwent either DNA oxidation (oxBS samples) or mock DNA oxidation (BS samples). Subsequently, all samples were BS-treated, and the yield of the samples was assessed by a Qubit ssDNA assay (Invitrogen). An additional quality control, using a restriction enzyme only able to cut unconverted cytosines, was performed for a qualitative assessment of 5hmC oxidation and BS conversion. From each BS/oxBS-treated DNA sample, 8 μL was amplified and hybridized on HM 450K arrays (Illumina, Inc., San Diego, CA, USA), and the Illumina iScan was used for imaging of the array. Sample preparation, hybridization, and washing steps for the Illumina Infinium Methylation Assay of the BeadChip arrays were performed according to the manufacturer's protocol.

For the AgeCoDe samples, gDNA was isolated from whole blood and DNA concentration and purity was determined using the NanoDrop ND1000 spectrophotometer (Thermo Fisher Scientific). All samples were of sufficient quantity and quality. Five hundred nanograms of gDNA was used for BS conversion, using a Qiagen EpiTect 96 Bisulfite Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. A total of 200 ng of BS converted DNA was analyzed using HM 450K arrays according to the manufacturer's instructions. The Illumina iScan was used for imaging of the array.

Transcriptomic profiling

Total RNA extracted from frozen MTG, from matched samples as used for the epigenetic MTG analyses, was isolated with the RNeasy Mini Kit (Qiagen) starting with at least 60 mg of tissue. Raw expression data was obtained at the BSHRI, using the HumanHT-12 v4 BeadChip (Illumina).

Statistical analysis

All computational and statistical analyses were performed using the statistical programming language R (version 3.3.2) [60] and RStudio (version 1.0.136) [61], unless otherwise specified. Raw IDAT files from the Illumina iScan were loaded into R using the *minfi* package (version 1.20.2) [62]. To confirm that the longitudinal samples were from the same donor a genetic fingerprinting test was performed based on the 65 SNP probes included on the HM 450K chip, as implemented in the *ewastools* package [63]. Based on this test, 2 donors with mismatching samples were detected and excluded from the blood data. Next, the gender of the samples was predicted based on X chromosome methylation using the *DNAmArray* package (version 0.0.2) [64], compared with the assumed gender, and mismatches were excluded (1 mismatched sample was excluded from the blood data). Cross-hybridizing probes and probes containing a

common SNP in the sequence or within 10 bp of the sequence were removed [65]. The "pfilter" function of the *wateRmelon* package (version 1.18.0) [66] was used for probe filtering (6 969 and 1 437 probes were removed from the MTG and blood data, respectively). The remaining probe data was normalized using the *dasen* method, as implemented in the *wateRmelon* package [66]. Probes on the X and Y chromosomes were excluded from further analyses.

Following normalization, two sets of beta values, from the standard BS arrays (5mC + 5hmC) and from the oxBS arrays (5mC), were generated in case of the MTG. By subtracting oxBS beta values from the BS beta values ($\Delta\beta_{BS\text{-}oxBS}$) for each probe in each sample, 5hmC levels were calculated (Fig. 1). UC values were determined as 1-BS (1-β_{BS}). It should be noted that other DNA demethylation intermediates, such as 5-formylcytosine (5fC) and 5-carboxylcytosine may be represented in the BS or UC levels, as it is currently unclear how these intermediates respond to oxBS conversion [67]. However, these intermediates are present at very low levels and are not enriched in brain tissue like 5hmC is [68]. In order to reduce noise and filter out non-hydroxymethylated sites, outliers deviating more than ± 2SD from the probe mean in the 5hmC dataset were determined and set to the mean ± 2SD first, and subsequently, a threshold of zero was applied to the mean of individual probes (218,009 5hmC values were excluded). Boxplots and density plots of raw and normalized beta values per sample were inspected for clear outliers (2 MTG samples were excluded due to clear deviation from the other samples; data not shown). After data processing, 80 MTG and 96 blood samples remained, with 396,600 remaining probes for MTG 5mC and UC, 178,591 5hmC MTG probes, and 402,480 remaining probes in the blood datasets. The case-control analysis of the blood baseline data included all 96 samples (54 converters, 42 controls), while follow-up data included 83 samples, including the 41 converters that had already converted to AD at the 4.5-year follow-up and excluding those that had converted later. All individuals in the follow-up analysis were also included in the baseline analysis.

An initial model with beta values as outcome, AD diagnosis/conversion as predictor, and age and gender as covariates was used for a surrogate variable (SV) analysis with the *sva* package (version 3.22.0) [69]. The first 5 SVs of this analysis were added to the model to adjust for unobserved confounders, including potential batch effects and differences in cell type composition. As the addition of SVs still resulted in inflation of the regression statistics (lambda = 1.43) of the blood follow-up analysis, and none of the SVs strongly correlated with the HM 450K chip IDs (which was the case for the other analyses), the chip IDs were also added to the model for this analysis. This successfully eliminated the inflation (lambda = 1.00).

Linear regression was performed using the *limma* package (version 3.30.11) [70] to test the association between the beta values and AD diagnosis/conversion. Test statistics were adjusted for bias and inflation with the *bacon* package (version 1.2.0) [71]. An FDR correction for multiple testing was applied to the *p* values to identify differentially (hydroxy)methylated and unmodified positions (probes with $p_{FDR} < 0.05$). Individual probes were annotated using Illumina UCSC annotation.

To examine the distribution of 5mC, 5hmC, and UC levels across genomic regions, we annotated the 1000 highest ranking probes (Additional file 1: Tables S2–S7) using ENCODE annotation data, as described by Slieker et al. [72]. Fisher's exact test was used to assess enrichment in specific genomic regions.

To identify differentially (hydroxy)methylated and unmodified regions (DHRs/DMRs/DURs), spatial correlations between *p* values of the association analysis were determined using *comb-p* [73] with a seeding *p* value of 0.01 and a window size of 1000 bp. Obtained *p* values were Stouffer-Liptak-Kechris corrected for adjacent *p* values and were subsequently corrected for multiple testing using the Šidák correction. Of the regions detected by *comb-p*, only those containing at least 3 CpGs and having a $p_{Šidák} < 0.05$ were accepted as differentially modified regions.

GRNs have been extensively used to achieve deeper understanding of disease related mechanisms [74]. Different topological characteristics of these networks, such as connectivity of nodes [75] or gene-gene interaction tendency in cell/tissue specific contexts [76], have been used to predict disease-related genes. Here, we have employed an in-house developed differential GRNs inference approach [77], which relies on gene expression data to infer GRNs specific to a given gene expression program. The initial set of interactions among the genes of interest was compiled from literature-based database ARIADNE [78] and consists of interactions belonging to the categories of "Direct Regulation," "Expression," and "Promoter Binding." The obtained set of interactions is not context-specific as they are reported to happen in different cell/tissue types and organisms. To obtain context-specific networks from the literature interaction maps, the pruning of interactions incompatible with the gene expression state was carried out, which resulted in contextualized networks compatible with the given gene expression state of the system. As a differential expression setting was used here, we obtained two contextualized GRNs for each state, representing the different network topology of diseased and healthy phenotype. The differential network topology helps us in identifying the set of genes that are regulated by different transcription factors in both networks. These genes formulate an ideal set of candidate perturbagens, as to change their

expression state we have to perturb them individually. The obtained contextualized networks were used to identify genes in the common elementary circuits (positive and negative circuits) that can also serve as a set of candidate genes for perturbation. Genes in elementary circuits have been reported to play a crucial role in maintaining network stability [79] and are considered as a necessary condition for a network to have an attractive cycle [80]. In this regard, genes present in the common elementary circuits are considered to be the backbone of the network and any perturbations in the expression levels of these genes might lead the system to deviate from the normal steady state of the system, which can be described as a transition from healthy to a diseased state. Once we obtained a set of optimal perturbation candidates, we performed single-gene perturbation simulations to see the effect of change in expression of a single gene on all the other genes in the GRN. This measure tells us about the influential capability of the selected gene in the network; the higher the number of downstream genes being affected by perturbing a candidate gene, the more crucial is its role in the regulation of other genes in the GRN.

Positions from the AD association analyses were ranked based on a combined p value and log2 fold change ranking score. The GRN analysis was then conducted separately for the genes annotated to the 1000 highest ranked sites in the MTG (5mC, 5hmC, and UC separately) and blood (baseline and follow-up separately) (Additional file 1: Tables S2–S7). Closest UCSC TSS annotation was used to obtain unique genes. After applying the differential GRN analysis on the contextualized networks, we ranked the key candidate genes based on their scores. This score represents the number of genes whose gene expression is changed (shifted from diseased towards the healthy phenotype) upon perturbation of the candidate gene.

Raw RNA expression data was exported from Illumina's GenomeStudio (version 2011.1) with the Expression Module (v1.9.0) for further analysis in R. Of the 80 subjects used for the epigenetic analyses, 1 case was not included on the expression array, and 3 additional cases were excluded after quality control of the data, due to extreme outlying values or failed reads, leaving 76 subjects for further analyses. Data was quantile-quantile normalized. Using the same model as for the regression analysis, the sva package was used to determine SVs for the epigenetic and expression datasets. The effects of age, gender, and 5 SVs were regressed out of the epigenetic and expression data using *limma* (i.e., "regressed data" refers to the residuals of a model fitted with the covariates, excluding the predictor of interest, being AD diagnosis or conversion in this case). Spearman correlations were determined for the expression data and the average of the regressed beta values of the probes in the DMRs, DHRs, and DURs, as well as correlations between the different epigenetic markers (5mC, 5hmC, and UC) for these probes, using the Hmisc package (version 4.0-2) [81].

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s13148-019-0755-5.

Additional file 1. All Supplementary Tables (1–10), including descriptions.

Additional file 2. All Supplementary Figures (1–12) and descriptions.

Abbreviations

5fC: 5-Formylcytosine; 5hmC: 5-Hydroxymethylcytosine; 5mC: 5-Methylcytosine; AD: Alzheimer's disease; AgeCoDe: Study on Ageing, Cognition and Dementia in Primary Care Patients and Stroke and the Alzheimer's Disease and Related Disorders Association; Aβ: Amyloid beta; BBDP: Brain and Body Donation Program; BS: Bisulfite; BSHRI: Banner Sun Health Research Institute; DHR: Differentially hydroxymethylated region; DMR: Differentially methylated region; DUR: Differentially unmodified region; EWAS: Epigenome-wide association study; FDR: False discovery rate; gDNA: Genomic DNA; GRN: Gene regulatory network; HM 450K array: Illumina's Infinium HumanMethylation450K microarray; HMW: High molecular weight; logFC: log2 fold change; MCI: Mild cognitive impairment; MTG: Middle temporal gyrus; oxBS: Oxidative BS; STG: Superior temporal gyrus; SV: Surrogate variable; TGF: Transforming growth factor; TSS: Transcription start site; UC: Unmodified cytosine

Acknowledgements

We thank the donors and their families who made this research possible.

Authors' contributions

RL, JAYR, and EP were responsible for data analysis, bioinformatics, interpretation of results, and drafting of the manuscript. ML, HW, LK, AK, PH, TL, SRH, FJ, WM, MW, and AR supervised the AgeCoDe cohort, collected samples and patient data, and generated the blood data used for this project. AI, ARS, RGS, AdS, JM, and KL assisted in constructing the analysis pipeline for this project. LMTE was involved in analyzing the expression data and MA did the GRN analysis. RH, GK, and BPFR assisted in interpreting the results and forming the discussion section of the manuscript. DM, ED, and PDC represent the BSHRI-BBDP and supplied the MTG DNA samples and expression data. DvdH and AR conceived and designed the project. DvdH supervised the project and writing of the manuscript. All authors read and approved the final manuscript.

Funding

Funds have been provided by the Internationale Stichting Alzheimer Onderzoek (ISAO)/Alzheimer Netherlands (Award #11532; Funded by the Dorpmans-Wigmans Foundation) (DvdH), the Baeter Laeve foundation, and by the Joint Programme—Neurodegenerative Disease Research (JPND) for the EPI-AD consortium (http://www.neurodegenerationresearch.eu/wp-content/uploads/2015/10/Factsheet_EPI-AD.pdf). The project is supported through the following funding organizations under the aegis of JPND; The Netherlands, The Netherlands Organisation for Health Research and Development (ZonMw); United Kingdom, Medical Research Council; Germany, German Federal ministry of Education and Research (BMBF); Luxembourg, National Research Fund (FNR). This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 643417. Additional funds have been provided by a fellowship as part of NWO grant 022.005.019 (RL) and the GW4 Biomed MRC Doctoral Training Partnership (JR). This research was further made possible by BReIN (Brightlands e-infrastructure for Neurohealth), an initiative which is co-funded by the Province of Limburg, Maastricht University and Maastricht University Medical Centre + in the Netherlands. This publication was also funded in part by the German

Federal Ministry of Education and Research (BMBF) (grants Nr: 01GI0710, 01GI0711, 01GI0712, 01GI0713, 01GI0714, 01GI0715, 01GI0716, 01ET1006B). Analyses were also funded by the German Federal Ministry of Education and Research (BMBF 01EA1410A) within the project "Diet-Body-Brain: from epidemiology to evidence-based communication".

Availability of data and materials

The datasets generated from the BSHRI-BBDP samples and analyzed during the current study are available in the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) repository, under GEO accession numbers GSE109627 and GSE109887 for the epigenetic and expression data, respectively. The datasets generated from the AgeCoDe samples and analyzed during the current study are not publicly available as participants did not provide informed consent for this, but are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Donors of the BBDP signed an Institutional Review Board-approved informed consent form, including specific consent to the use of donated tissue for future research [16, 54]. The AgeCoDe study protocol was approved by the local ethics committees at the University of Bonn (Bonn, Germany), the University of Hamburg (Hamburg, Germany), the University of Duesseldorf (Duesseldorf, Germany), the University of Heidelberg/Mannheim (Mannheim, Germany), the University of Leipzig (Leipzig, Germany), and the Technical University of Munich (Munich, Germany).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests

Author details

[1]School for Mental Health and Neuroscience (MHeNS), Department of Psychiatry and Neuropsychology, Maastricht University, P.O. Box 616, 6200, MD, Maastricht, the Netherlands. [2]Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, 37075 Göttingen, Germany. [3]University of Exeter Medical School, University of Exeter, Exeter, UK. [4]Division of Neurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, University of Cologne, Medical Faculty, 50937 Cologne, Germany. [5]Department of Neurodegeneration and Gerontopsychiatry, University of Bonn, 53127 Bonn, Germany. [6]Department of Bioinformatics—BiGCaT, Maastricht University, Maastricht, The Netherlands. [7]German Center for Neurodegenerative Diseases (DZNE), 53127 Bonn, Germany. [8]Institute of Human Genetics, University of Bonn, 53127 Bonn, Germany. [9]Department of Genomics, Life & Brain Center, University of Bonn, 53127 Bonn, Germany. [10]Division of Medical Genetics, University Hospital and Department of Biomedicine, University of Basel, CH-4058 Basel, Switzerland. [11]Institute of Social Medicine, Occupational Health and Public Health, University of Leipzig, 04103 Leipzig, Germany. [12]Department of Psychiatry and Psychotherapy, University of Cologne, Medical Faculty, 50937 Cologne, Germany. [13]Department of Psychiatry and Division of Medical Psychology, University of Bonn, 53105 Bonn, Germany. [14]Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette, Luxembourg. [15]Moscow Institute of Physics and Technology, Dolgoprudny, Moscow, Russian Federation. [16]CIC bioGUNE, Bizkaia Technology Park, 801 Building, 48160 Derio, Spain. [17]IKERBASQUE, Basque Foundation for Science, Dolgoprudny Bilbao, Spain. [18]L.J. Roberts Center for Alzheimer's Research Banner Sun Health Research Institute, Sun City, AZ, USA. [19]Biodesign Institute, Neurodegenerative Disease Research Center, Arizona State University, Tempe, AZ, USA. [20]Institute of Psychiatry, King's College London, London, UK. [21]Department of Psychiatry, Psychosomatics and Psychotherapy, University of Würzburg, Würzburg, Germany.

Received: 8 April 2019 Accepted: 26 September 2019
Published online: 27 November 2019

References

1. Lambert J-C, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. Nature Res. 2013;45:1452–8.
2. Lardenoije R, Iatrou A, Kenis G, Kompotis K, Steinbusch HWM, Mastroeni D, et al. The epigenetics of aging and neurodegeneration. Prog Neurobiol. 2015;131.
3. Lunnon K, Smith R, Hannon E, De Jager PL, Srivastava G, Volta M, et al. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. Nat Neurosci [Internet]. Nature Research; 2014 [cited 2017 Feb 20];17:1164–1170. Available from: http://www.nature.com/doifinder/10.1038/nn.3782
4. De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat Neurosci [Internet]. 2014 [cited 2017 Feb 20];17:1156–1163. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25129075
5. Chouliaras L, Mastroeni D, Delvaux E, Grover A, Kenis G, Hof PR, et al. Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. Neurobiol Aging [Internet]. 2013 [cited 2017 Feb 20];34:2091–2099. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23582657
6. Zhao J, Zhu Y, Yang J, Li L, Wu H, De Jager PL, et al. A genome-wide profiling of brain DNA hydroxymethylation in Alzheimer's disease. Alzheimer's Dement. 2017;13(6):674–88.
7. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science (80-). 2009;324:929–30 Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 3263819&tool = pmcentrez&rendertype = abstract.
8. Nestor CE, Ottaviano R, Reddington J, Sproul D, Reinhardt D, Dunican D, et al. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. Genome Res [Internet]. Cold Spring Harbor Laboratory Press; 2012 [cited 2017 Oct 31];22:467–477. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22106369
9. Irier H, Street RC, Dave R, Lin L, Cai C, Davis TH, et al. Environmental enrichment modulates 5-hydroxymethylcytosine dynamics in hippocampus. Genomics [Internet]. Academic Press; 2014 [cited 2018 May 14];104:376–382. Available from: https://www.sciencedirect.com/science/article/pii/S088875431400161X
10. Chen R, Zhang Q, Duan X, York P, Chen G-D, Yin P, et al. The 5-Hydroxymethylcytosine (5hmC) Reader UHRF2 Is Required for Normal Levels of 5hmC in Mouse Adult Brain and Spatial Learning and Memory. J Biol Chem [Internet]. 2017 [cited 2018 May 14];292:4533–4543. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28115522
11. Di Francesco A, Arosio B, Falconi A, Micioni Di Bonaventura MV, Karimi M, Mari D, et al. Global changes in DNA methylation in Alzheimer's disease peripheral blood mononuclear cells. Brain Behav Immun [Internet]. 2015 [cited 2017 Apr 18];45:139–144. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25452147
12. Li H, Guo Z, Guo Y, Li M, Yan H, Cheng J, et al. Common DNA methylation alterations of Alzheimer's disease and aging in peripheral whole blood. Oncotarget [Internet]. Impact Journals, LLC; 2016 [cited 2017 Apr 30];7:19089–19098. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26943045
13. St-Amour I, Cicchetti F, Calon F. Immunotherapies in Alzheimer's disease: Too much, too little, too late or off-target? Acta Neuropathol [Internet]. Springer Berlin Heidelberg; 2016 [cited 2017 29];131:481–504. Available from: http://link.springer.com/10.1007/s00401-015-1518-9
14. Sperling R, Mormino E, Johnson K. The Evolution of Preclinical Alzheimer's Disease: Implications for Prevention Trials. Neuron [Internet]. 2014 [cited 2017 Apr 29];84:608–622. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25442939
15. Smith AR, Smith RG, Pishva E, Hannon E, Roubroeks JAY, Burrage J, et al. Parallel profiling of DNA methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in Alzheimer's disease. Clin Epigenetics [Internet]. BioMed Central; 2019 [cited 2019 Sep 9];11:52. Available from: https://clinicalepigeneticsjournal.biomedcentral.com/articles/10.1186/s13148-019-0636-y
16. Beach TG, Adler CH, Sue LI, Serrano G, Shill HA, Walker DG, et al. Arizona Study of Aging and Neurodegenerative Disorders and Brain and Body Donation Program. Neuropathology [Internet]. 2015 [cited 2017 Nov 22];35:354–389. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25619230
17. Ray M, Zhang W, Liang W, Dunckley T, Beach T, Grover A, et al. Analysis of Alzheimer's disease severity across brain regions by topological analysis of gene co-expression networks. BMC Syst Biol. BioMed Central; 2010;4:136.
18. Coppieters N, Dieriks B V, Lill C, Faull RL, Curtis MA, Dragunow M. Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. Neurobiol Aging [Internet]. 2014;35:1334–1344. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24387984
19. Piras IS, Krate J, Delvaux E, Nolz J, Mastroeni DF, Persico AM, et al. Transcriptome Changes in the Alzheimer's Disease Middle Temporal Gyrus:

Importance of RNA Metabolism and Mitochondria-Associated Membrane Genes. Combs C, editor. J Alzheimer's Dis [Internet]. 2019 [cited 2019 Sep 9]; 70:691–713. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31256118
20. Kamkwalala A, Newhouse P. Beyond Acetylcholinesterase Inhibitors: Novel Cholinergic Treatments for Alzheimer's Disease. Curr Alzheimer Res [Internet]. 2016 [cited 2017 Dec 10];13:1. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27697062
21. Adrain C, Zettl M, Christova Y, Taylor N, Freeman M. Tumor Necrosis Factor Signaling Requires iRhom2 to Promote Trafficking and Activation of TACE. Sci Reports. 2012;335:225–8.
22. Bhaskar K, Maphis N, Xu G, Varvel NH, Kokiko-Cochran ON, Weick JP, et al. Microglial derived tumor necrosis factor-α drives Alzheimer's disease-related neuronal cell cycle events. Neurobiol Dis [Internet]. 2014 [cited 2018 Jun 20];62: 273–285. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24141019
23. Zhang K, Programs F, Boe C, Barbieri M, Crimmins EM, Preston SH, et al. Complement and microglia mediate early synapse loss in Alzheimer mouse models. Science. 2016;352(6286):712–6.
24. Meerabux JMA, Ohba H, Iwayama Y, Maekawa M, Detera-Wadleigh SD, DeLisi LE, et al. Analysis of a t(18;21)(p11.1;p11.1) translocation in a family with schizophrenia. J Hum Genet [Internet]. 2009 [cited 2018 May 22];54: 386–391. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19461657
25. Kikuchi M, Yamada K, Toyota T, Yoshikawa T. C18orf1 located on chromosome 18p11.2 may confer susceptibility to schizophrenia. J Med Dent Sci [Internet]. 2003 [cited 2018 May 22];50:225–229. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15074360
26. Nakano N, Maeyama K, Sakata N, Itoh F, Akatsu R, Nakata M, et al. C18 ORF1, a novel negative regulator of transforming growth factor-β Signaling. J Biol Chem [Internet]. 2014 [cited 2018 May 22];289:12680–12692. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24627487
27. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, et al. Genome-wide association study of blood pressure and hypertension. Nat Genet [Internet]. 2009 [cited 2018 May 22];41:677–687. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19430479
28. Chong JR, Chai YL, Lee JH, Howlett D, Attems J, Ballard CG, et al. Increased transforming growth factor β2 in the neocortex of Alzheimer's disease and dementia with Lewy bodies is correlated with disease severity and soluble Aβ42 load. J Alzheimer's Dis [Internet]. 2017 [cited 2018 May 22];56:157–166. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27911312
29. Young TR, Leamey CA. Teneurins: Important regulators of neural circuitry. Int J Biochem Cell Biol [Internet]. 2009 [cited 2018 May 22];41:990–993. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18723111
30. Bai Z, Stamova B, Xu H, Ander BP, Wang J, Jickling GC, et al. Distinctive RNA expression profiles in blood associated with Alzheimer disease after accounting for white matter hyperintensities. Alzheimer Dis Assoc Disord [Internet]. NIH Public Access; 2014 [cited 2018 May 22];28:226–233. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24731980
31. Yao C, Joehanes R, Johnson AD, Huan T, Esko T, Ying S, et al. Sex- and age-interacting eQTLs in human complex diseases. Hum Mol Genet [Internet]. Oxford University Press; 2014 [cited 2018 May 22];23:1947–1956. Available from: https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddt582
32. P Bennett J, M Keeney P. Micro RNA's (mirna's) may help explain expression of multiple genes in Alzheimer's Frontal Cortex. J Syst Integr Neurosci [Internet]. 2017 [cited 2018 May 22];3. Available from: http://www.oatext.com/micro-rnas-mirnas-may-help-explain-expression-of-multiple-genes-in-alzheimers-frontal-cortex.php
33. Tian Y, Bustos V, Flajolet M, Greengard P. A small-molecule enhancer of autophagy decreases levels of Abeta and APP-CTF via Atg5-dependent autophagy pathway. FASEB J [Internet]. The Federation of American Societies for Experimental Biology; 2011 [cited 2018 May 22];25:1934–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21368103
34. Khrimian L, Obri A, Karsenty G. Modulation of cognition and anxiety-like behavior by bone remodeling. Mol Metab [Internet]. Elsevier; 2017 [cited 2018 May 22];6:1610–1615. Available from: https://www.sciencedirect.com/science/article/pii/S2212877817306877
35. Gibbs GM, Roelants K, O'Bryan MK. The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins—roles in reproduction, cancer, and immune defense. Endocr Rev [Internet]. Oxford University Press; 2008 [cited 2018 May 22];29:865–897. Available from: https://academic.oup.com/edrv/article-lookup/doi/10.1210/er.2008-0032
36. Ren C, Ren C-H, Li L, Goltsov AA, Thompson TC. Identification and characterization of RTVP1/GLIPR1-like genes, a novel p53 target gene cluster. Genomics [Internet]. 2006 [cited 2018 May 22];88:163–172. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16714093
37. Song L, Pei L, Yao S, Wu Y, Shang Y. NLRP3 inflammasome in neurological diseases, from functions to therapies. Front Cell Neurosci [Internet]. Frontiers; 2017 [cited 2018 May 22];11:63. Available from: http://journal.frontiersin.org/article/10.3389/fncel.2017.00063/full
38. Tan M-S, Tan L, Jiang T, Zhu X-C, Wang H-F, Jia C-D, et al. Amyloid-β induces NLRP1-dependent neuronal pyroptosis in models of Alzheimer's disease. Cell Death Dis [Internet]. 2014 [cited 2018 May 22];5:e1382. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25144717
39. Gusareva ES, Carrasquillo MM, Bellenguez C, Cuyvers E, Colon S, Graff-Radford NR, et al. Genome-wide association interaction analysis for Alzheimer's disease. Neurobiol Aging [Internet]. NIH Public Access; 2014 [cited 2017 Nov 15];35: 2436–2443. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24958192
40. Fatima SS, Rehman R, Baig M, Khan TA. New roles of the multidimensional adipokine: Chemerin. Peptides [Internet]. Elsevier; 2014 [cited 2018 May 23];62:15–20. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0196978114002885
41. Gasparoni G, Bultmann S, Lutsik P, Kraus TFJ, Sordon S, Vlcek J, et al. DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. Epigenetics Chromatin [Internet]. 2018 [cited 2019 Jun 24];11:41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30045751
42. McGeer PL, McGeer EG. The amyloid cascade-inflammatory hypothesis of Alzheimer disease: implications for therapy. Acta Neuropathol [Internet]. Springer Berlin Heidelberg; 2013 [cited 2019 Mar 22];126:479–497. Available from: http://link.springer.com/10.1007/s00401-013-1177-7
43. Watson CT, Roussos P, Garg P, Ho DJ, Azam N, Katsel PL, et al. Genome-wide DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease. Genome Med [Internet]. BioMed Central; 2016 [cited 2017 Apr 30];8:5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26803900
44. Olff M, Frijling JL, Kubzansky LD, Bradley B, Ellenbogen MA, Cardoso C, et al. The role of oxytocin in social bonding, stress regulation and mental health: An update on the moderating effects of context and interindividual differences. Psychoneuroendocrinology. 2013;38:1883–94.
45. Haas BW, Filkowski MM, Cochran RN, Denison L, Ishak A, Nishitani S, et al. Epigenetic modification of OXT and human sociability. Proc Natl Acad Sci U S A [Internet]. National Academy of Sciences; 2016 [cited 2018 Jun 22];113: E3816–E3823. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27325757
46. Swaab DF, Chapter II. Neurobiology and neuropathology of the human hypothalamus. Handb Chem Neuroanat. Elsevier. 1997;13:39–137.
47. de Lacalle S, Iraizoz I, Gonzalo LM. Cell loss in supraoptic and paraventricular nucleus in Alzheimer's disease. Brain Res [Internet]. 1993 [cited 2018 May 14]; 609:154–158. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8508299
48. Mazurek MF, Beal MF, Bird ED, Martin JB. Oxytocin in Alzheimer's disease: postmortem brain levels. Neurology [Internet]. 1987 [cited 2018 May 14];37: 1001–1003. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3587615
49. North WG, Harbaugh R, Reeder T. An evaluation of human neurophysin production in Alzheimer's disease: preliminary observations. Neurobiol Aging [Internet]. 1992 [cited 2018 May 14];13:261–265. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1522943
50. Mann DM, Yates PO, Marcyniuk B. Changes in Alzheimer's disease in the magnocellular neurones of the supraoptic and paraventricular nuclei of the hypothalamus and their relationship to the noradrenergic deficit. Clin Neuropathol [Internet]. 1985 [cited 2018 May 14];4:127–134. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3160517
51. Torner L, Plotsky PM, Neumann ID, de Jong TR. Forced swimming-induced oxytocin release into blood and brain: Effects of adrenalectomy and corticosterone treatment. Psychoneuroendocrinology [Internet]. 2017 [cited 2017 Jul 6];77:165–174. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0306453016306801
52. Tampi RR, Maksimowski M, Ahmed M, Tampi DJ. Oxytocin for frontotemporal dementia: a systematic review. Ther Adv Psychopharmacol [Internet]. SAGE PublicationsSage UK: London, England; 2017 [cited 2017 Dec 10];7:48–53. Available from: http://journals.sagepub.com/doi/10.1177/2045125316672574
53. Lunnon K, Mill J. Epigenetic studies in Alzheimer's disease: current findings, caveats, and considerations for future studies. Am J Med Genet B Neuropsychiatr Genet [Internet]. 2013 [cited 2015 Dec 7];162B:789–799. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3947441&tool=pmcentrez&rendertype=abstract
54. Beach TG, Sue LI, Walker DG, Roher AE, Lue L, Vedders L, et al. The Sun Health Research Institute Brain Donation Program: Description and Experience, 1987–2007. Cell Tissue Bank. Springer Netherlands; 2008;9:229–245.
55. Luck T, Riedel-Heller SG, Kaduszkiewicz H, Bickel H, Jessen F, Pentzek M, et al. Mild cognitive impairment in general practice: age-specific prevalence and correlate results from the German study on ageing, cognition and

dementia in primary care patients (AgeCoDe). Dement Geriatr Cogn Disord [Internet]. 2007 [cited 2017 Nov 22];24:307–316. Available from: https://www.karger.com/Article/FullText/108099
56. Ramirez A, Wolfsgruber S, Lange C, Kaduszkiewicz H, Weyerer S, Werle J, et al. Elevated HbA1c is associated with increased risk of incident dementia in primary care patients. J Alzheimers Dis [Internet]. 2015 [cited 2017 Nov 22];44:1203–1212. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25524954
57. Zaudig M, Hiller W. Sidam-Handbuch. Strukturiertes Interview für die Diagnose einer Demenz vom Alzheimer Typ, der Multi-Infarkt-(oder vaskulären) Demenzen und Demenzen anderer Ätiologien nach DSM-III-R, DSM-IV und ICD-10. Bern: Huber; 1996.
58. Reisberg B, Ferris SH, de Leon MJ, Crook T. The Global Deterioration Scale for assessment of primary degenerative dementia. Am J Psychiatry [Internet]. 1982 [cited 2017 Apr 20];139:1136–1139. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7114305
59. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology [Internet]. 1984 [cited 2017 Apr 20];34:939–944. Available from: http://www.ncbi.nlm.nih.gov/pubmed/6610841
60. R Core Team. R: A language and environment for statistical computing. [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2016. Available from: https://www.r-project.org/
61. RStudio Team. RStudio: Integrated Development for R [Internet]. Boston, MA: RStudio, Inc.; 2016. Available from: http://www.rstudio.com/
62. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. Oxford University Press; 2014;30:1363–1369.
63. Heiss JA, Just AC. Identifying mislabeled and contaminated DNA methylation microarray data: an extended quality control toolset with examples from GEO. Clin Epigenetics [Internet]. BioMed Central; 2018 [cited 2019 Jun 24];10:73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29881472
64. van Iterson M, Tobi E, Slieker R, den Hollander, Wouter Luijk R, Dekkers K, Heijmans B. DNAmArray [Internet]. Leiden, the Netherlands; 2017. Available from: https://github.com/molepi/DNAmArray
65. Chen Y, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. Taylor & Francis. 2013;8:203–9.
66. Pidsley R, YW CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics. 2013;14:293.
67. Fukuzawa S, Takahashi S, Tachibana K, Tajima S, Suetake I. Simple and accurate single base resolution analysis of 5-hydroxymethylcytosine by catalytic oxidative bisulfite sequencing using micelle incarcerated oxidants. Bioorg Med Chem. 2016;24:4254–62.
68. Song C-X, He C. Potential functional roles of DNA demethylation intermediates. Trends Biochem Sci. 2013;38:480–4.
69. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics [Internet]. Oxford University Press; 2012 [cited 2017 Apr 10];28:882–883. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22257669
70. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res [Internet]. 2015 [cited 2017 Apr 10];43:e47. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv007
71. van Iterson M, van Zwet EW, BIOS Consortium BT, Heijmans BT. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. Genome Biol [Internet]. 2017 [cited 2017 Apr 10];18:19. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-1131-9
72. Slieker RC, Bos SD, Goeman JJ, Bovée JV, Talens RP, van der Breggen R, et al. Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. Epigenetics Chromatin. BioMed Central; 2013;6:26.
73. Pedersen BS, Schwartz DA, Yang IV, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. Bioinformatics. Oxford University Press. 2012;28:2986–8.
74. Schadt EE, Friend SH, Shaywitz DA. A network view of disease and compound screening. Nat Rev Drug Discov [Internet]. 2009 [cited 2017 Dec 6];8:286–295. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19337271
75. Jonsson PF, Bates PA. Global topological features of cancer proteins in the human interactome. Bioinformatics [Internet]. 2006 [cited 2017 Dec 6];22:2291–2297. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16844706
76. Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet [Internet]. 2011 [cited 2017 Dec 6];12:56–68. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21164525
77. Zickenrott S, Angarica VE, Upadhyaya BB, del Sol A. Prediction of disease-gene-drug relationships following a differential network analysis. Cell Death Dis [Internet]. 2016 [cited 2017 Dec 5];7:e2040. Available from: http://www.nature.com/doifinder/10.1038/cddis.2015.393
78. Nikitin A, Egorov S, Daraselia N, Mazo I. Pathway studio–the analysis and navigation of molecular networks. Bioinformatics [Internet]. 2003 [cited 2017 Dec 6];19:2155–2157. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14594725
79. Plahte E, Mestl T, Omholt SW. Feedback loops, stability and multistationarity in dynamical systems. J Biol Syst [Internet]. World Scientific Publishing Company; 1995 [cited 2017 Dec 6];03:409–413. Available from: http://www.worldscientific.com/doi/abs/10.1142/S0218339095000381
80. Thomas R. On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. Springer, Berlin, Heidelberg; 1981 [cited 2017 Dec 6]. p. 180–93. Available from: http://link.springer.com/10.1007/978-3-642-81703-8_24
81. Harrell Jr FE, with contributions from Charles Dupont and many others. Hmisc: Harrell Miscellaneous. R package version 4.0-3 [Internet]. 2017. Available from: https://cran.r-project.org/package = Hmisc

Publisher's Note

APPENDIX D: XI DMPS ASSOCIATED WITH CTL, MCI, AND AD IN ADDNEUROMED

ANOVA Xi DMPs in AddNeuroMed						
ProbeID	**Position**	**F**	***p* (F)**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg22822140	chrX:153046577	60.317	1.92E-20	*SRPK3*	1stExon	*SRPK3* (122)
cg13009143	chrX:100548223	47.71	4.08E-17	*TAF7L*	TSS200	*TAF7L* (-165)
cg12230162	chrX:153046482	46.334	9.84E-17	*SRPK3*	5'UTR,1stExon	*SRPK3* (27)
cg24496423	chrX:153046480	43.144	7.85E-16	*SRPK3*	5'UTR,1stExon	*SRPK3* (25)
cg16152676	chrX:10126878	36.492	7.10E-14	*CLCN4*	5'UTR	*CLCN4* (1894), *MID1* (724930)
cg03360001	chrX:53453272	35.944	1.04E-13	*RIBC1*	Body	*SMC1A* (-3655)
cg12468189	chrX:153046767	34.577	2.72E-13	*SRPK3*	Body	*SRPK3* (312)
cg09768654	chrX:153046386	30.912	3.79E-12	*SRPK3*	TSS200	*SRPK3* (-69)
cg16529483	chrX:153046451	30.87	3.91E-12	*SRPK3*	TSS200	*SRPK3* (-4)
cg21252909	chrX:30326328	30.829	4.03E-12	*NR0B1*	1stExon	*NR0B1* (1166)
cg19938385	chrX:107020894	30.752	4.26E-12			*TSC22D3* (-1878)
cg19280671	chrX:149013642	30.44	5.35E-12	*MAGEA8*	Body	*MAMLD1* (-517908), *MAGEA8* (3702)
cg24605338	chrX:17395856	30.05	7.12E-12	*NHS*	Body	*SCML1* (-359735), *NHS* (2314)
cg17405188	chrX:73164102	27.494	4.77E-11	*LOC554203*	TSS200	*ZCCHC13* (-359922), *CHIC1* (381119)
cg17003204	chrX:102861456	26.96	7.13E-11	*TCEAL3*	TSS1500	*TCEAL3* (-1377)
cg25896901	chrX:125300481	26.155	1.31E-10	*DCAF12L2*	TSS1500	*DCAF12L1* (386360)
cg22926378	chrX:6144633	25.983	1.50E-10	*NLGN4X*	5'UTR	*NLGN4X* (2072)
cg06452970	chrX:13835264	23.847	7.76E-10	*GPM6B*	1stExon,5'UTR,Body	*GPM6B* (49)
cg23914849	chrX:52950051	23.755	8.33E-10			*FAM156B* (21967), *FAM156A* (74599)
cg16122592	chrX:26210712	23.27	1.22E-09	*MAGEB6*	5'UTR	*MAGEB6* (156)
cg07069575	chrX:49688313	22.786	1.78E-09	*CLCN5*	5'UTR	*CLCN5* (1089), *AKAP4* (277350)
cg26484667	chrX:2733164	22.63	2.01E-09	*XG*	3'UTR	*GYG2* (-13698), *XG* (63072)

ANOVA Xi DMPs in AddNeuroMed (Continued)						
ProbeID	**Position**	**F**	***p* (F)**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg17513789	chrX:73073251	22.467	2.29E-09	*XIST*	TSS1500	*ZCCHC13* (-450773), *CHIC1* (290268)
cg15907464	chrX:70712790	21.952	3.43E-09	*BCYRN1;INGX*	Body;TSS200	*OGT* (-40121), *TAF1* (126677)
cg17219209	chrX:149009840	21.658	4.33E-09	*MAGEA8*	TSS200	*MAGEA8* (-100)
cg04710661	chrX:2848381	21.513	4.87E-09	*ARSD*	TSS1500	*ARSD* (-966)
cg04768884	chrX:19443172	21.379	5.41E-09	*MAP3K15*	Body	*PDHA1* (81162), *MAP3K15* (90206)
cg14917571	chrX:79269151	20.935	7.71E-09	*TBX22*	TSS1500	*ITM2A* (-646103), *TBX22* (-8590)
cg00190642	chrX:103268431	20.518	1.08E-08	*H2BFWT;MIR1256*	TSS200;Body	*H2BFWT* (-176)
cg09953945	chrX:150345040	19.926	1.73E-08	*GPR50*	TSS200	*GPR50* (-15)
cg11179997	chrX:119125670	19.905	1.77E-08			*NKAP* (-47936), *RHOXF2* (86036)
cg00805156	chrX:3264341	19.889	1.79E-08	*MXRA5*	5'UTR	*MXRA5* (342)
cg09661041	chrX:153141905	19.818	1.89E-08	*L1CAM*	TSS1500	*L1CAM* (-507)
cg00123001	chrX:13395880	19.591	2.27E-08			*EGFL6* (-191813), *ATXN3L* (-57363)
cg24352688	chrX:13751727	19.572	2.31E-08	*TRAPPC2;OFD1*	5'UTR;TSS1500	*OFD1* (-1104), *TRAPPC2* (1026)
cg09513996	chrX:48685182	19.303	2.87E-08			*ERAS* (-2100)
cg06068202	chrX:150343148	19.185	3.16E-08			*GPR50* (-1907)
cg02329520	chrX:153740317	19.162	3.22E-08	*FAM3A*	Body	*SLC10A3* (-21316), *FAM3A* (4248)
cg08630881	chrX:70713213	19.093	3.41E-08	*BCYRN1;INGX*	Body;TSS1500	*OGT* (-39698), *TAF1* (127100)
cg22574818	chrX:55479616	18.874	4.07E-08	*MAGEH1*	1stExon,3'UTR	*MAGEH1* (1079), *USP51* (36014)
cg01737010	chrX:101772014	18.852	4.15E-08	*TMSB15A*	TSS1500	*TMSB15A* (-316)
cg08464958	chrX:2852845	18.811	4.29E-08	*ARSE*	3'UTR	*ARSD* (-5430), *ARSE* (29648)
cg26101161	chrX:122736584	18.311	6.46E-08	*THOC2*	3'UTR	*THOC2* (130319), *GRIA3* (418489)
cg13788827	chrX:12990209	18.054	7.97E-08			*TMSB4X* (-3016)
cg04817724	chrX:103809991	18.025	8.17E-08	*IL1RAPL2*	TSS1500	*IL1RAPL2* (-1004)

ANOVA Xi DMPs in AddNeuroMed (Continued)						
ProbeID	**Position**	**F**	***p* (F)**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg21364342	chrX:107226611	17.69	1.08E-07	*TEX13B*	TSS1500	*VSIG1* (-61588), *MID2* (157528)
cg08333400	chrX:10126881	16.903	2.06E-07	*CLCN4*	5'UTR	*CLCN4* (1897), *MID1* (724927)
cg07344627	chrX:125687011	16.858	2.14E-07	*DCAF12L1*	TSS200	*DCAF12L1* (-170)
cg14306994	chrX:49019804	16.805	2.24E-07	*MAGIX*	Body,TSS1500	*GPKOW* (-39726), *PLP2* (-8379)
cg08867267	chrX:130033963	16.57	2.73E-07	*ENOX2*	5'UTR	*ENOX2* (3244), *RBMX2* (498021)
cg22090245	chrX:99661860	16.522	2.84E-07	*PCDH19*	1stExon	*PCDH19* (3410)
cg24834461	chrX:153637645	16.433	3.06E-07	*DNASE1L1*	TSS200,5'UTR	*TAZ* (-2231)
cg08850124	chrX:118207749	16.35	3.28E-07			*PGRMC1* (-162461), *LONRF3* (99037)
cg25433595	chrX:100183764	16.329	3.33E-07	*XKRX*	5'UTR,1stExon	*XKRX* (133)
cg00072799	chrX:78427682	16.296	3.43E-07	*GPR174*	1stExon,3'UTR	*GPR174* (1214), *ITM2A* (195366)
cg25177692	chrX:129193893	16.251	3.56E-07			*ELF4* (50794), *BCORL1* (54730)
cg26747413	chrX:15807140	16.179	3.78E-07	*INE2;ZRSR2*	TSS1500;TSS1500	*ZRSR2* (-1433)
cg16508637	chrX:132094977	15.993	4.42E-07	*HS6ST2*	5'UTR	*HS6ST2* (445)
cg07395961	chrX:25041645	15.969	4.51E-07			*ARX* (-7581)
cg12039689	chrX:153141750	15.907	4.75E-07	*L1CAM*	TSS1500	*L1CAM* (-352)
cg06616051	chrX:153094332	15.857	4.95E-07	*PDZD4*	Body	*IDH3G* (-34366), *PDZD4* (1670)
cg00977690	chrX:24167334	15.855	4.96E-07			*ZFX* (-427)
cg17662252	chrX:54560704	15.76	5.37E-07	*GNL3L*	Body	*GNL3L* (4061), *ITIH6* (263968)
cg07892139	chrX:84192519	15.757	5.39E-07			*APOOL* (-66378), *UBE2DNL* (3363)
cg05889382	chrX:70887722	15.74	5.46E-07	*BCYRN1*	Body	*PIN4* (-513803), *CXCR3* (-49356)
cg26468081	chrX:69501565	15.594	6.18E-07	*ARR3;RAB41*	Body;TSS1500	*RAB41* (-456)
cg10482495	chrX:71351314	15.588	6.21E-07	*NHSL2;RGAG4*	Body;1stExon	*CXCR3* (-512948), *PIN4* (-50211)
cg25958561	chrX:110187179	15.482	6.79E-07	*PAK3*	TSS1500	*PAK3* (-179125), *CHRDL1* (-148104)
cg02754763	chrX:44403388	15.468	6.87E-07	*FUNDC1*	TSS1500	*DUSP21* (-299860), *EFHC2* (-200466)

ANOVA Xi DMPs in AddNeuroMed (Continued)						
ProbeID	**Position**	**F**	***p* (F)**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
ch.X.1639000R	chrX:113225430	15.448	6.99E-07			*HTR2C* (-593120)
cg20814574	chrX:92928640	15.348	7.60E-07	*FAM133A;NAP1L3*	TSS1500;TSS200	*NAP1L3* (41)
cg10288121	chrX:53247489	15.33	7.72E-07	*KDM5C*	Body	*KDM5C* (7114), *TSPYL2* (135948)
cg09626131	chrX:82764537	15.301	7.91E-07	*POU3F4*	3'UTR,1stExon	*CYLC1* (-351632), *POU3F4* (1269)
cg09979033	chrX:16726904	15.151	8.98E-07	*CTPS2*	5'UTR	*CTPS2* (4154), *S100G* (58624)
cg20784188	chrX:44703983	15.026	9.98E-07	*DUSP21*	3'UTR,1stExon	*DUSP21* (735)
cg25396787	chrX:3733991	15.013	1.01E-06			*PRKX* (-102317)
cg26486175	chrX:18658942	14.896	1.12E-06	*CDKL5;RS1*	Body;3'UTR	*RS1* (31280), *CDKL5* (215218)
cg04867253	chrX:70472827	14.778	1.23E-06	*ZMYM3;BCYRN1*	Body;Body	*ZMYM3* (1210), *GJB1* (37766)
cg00928753	chrX:36976515	14.676	1.34E-06			*PRRG1* (-232012), *CXorf59* (911463)
cg04751886	chrX:53220515	14.649	1.37E-06	*KDM5C*	3'UTR	*KDM5C* (34088), *TSPYL2* (108974)
cg25880538	chrX:68840309	14.515	1.54E-06	*EDA*	Body	*EDA* (4399), *AWAT2* (429478)
ch.X.94051109R	chrX:94164453	14.486	1.58E-06			
cg03449040	chrX:10094363	14.473	1.60E-06	*WWC3*	Body	*CLCN4* (-30621), *SHROOM2* (339868)
cg16801826	chrX:54587084	14.415	1.68E-06	*GNL3L*	3'UTR	*GNL3L* (30441), *ITIH6* (237588)
cg07089438	chrX:133988190	14.415	1.68E-06	*FAM122C*	3'UTR,Body	*PLAC1* (-195678), *MOSPD1* (61106)
ch.X.153018199F	chrX:153365005	14.369	1.75E-06			*MECP2* (-1818)
cg27558057	chrX:70712724	14.353	1.77E-06	*BCYRN1;INGX*	Body;TSS200	*OGT* (-40187), *TAF1* (126611)
cg13976265	chrX:3732500	14.27	1.90E-06			*PRKX* (-100826)
cg02048654	chrX:132551786	14.255	1.92E-06			*GPC4* (-2582)
cg16975981	chrX:55246829	14.125	2.15E-06	*PAGE5*	5'UTR,1stExon	*MAGEH1* (-231708), *MTRNR2L10* (-37886)
cg12461113	chrX:13587550	13.921	2.56E-06	*EGFL6*	TSS200	*EGFL6* (-143)
cg01067505	chrX:107178820	13.845	2.73E-06			*VSIG1* (-109379), *MID2* (109737)

ANOVA Xi DMPs in AddNeuroMed (Continued)						
ProbeID	**Position**	**F**	***p* (F)**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg08119631	chrX:118822815	13.763	2.93E-06	*SEPT6*	Body	*NKRF* (-82970), *SEPT6* (4517)
cg24401049	chrX:11157158	13.728	3.02E-06	*ARHGAP6*	Body	*AMELX* (-154374), *HCCS* (27753)
cg24743423	chrX:8701002	13.49	3.71E-06	*KAL1*	TSS1500	*KAL1* (-776)
cg22268449	chrX:47053156	13.482	3.73E-06	*UBA1*	TSS200,5'UTR	*CDK16* (-29260), *UBA1* (2958)
cg05148324	chrX:153211181	13.464	3.79E-06	*RENBP*	TSS1500	*RENBP* (-950)
cg05091491	chrX:114423926	13.431	3.90E-06	*RBMXL3;LRCH2*	TSS200;Body	*RBMXL3* (-36)
cg00695046	chrX:100545995	13.394	4.03E-06	*TAF7L*	Body,5'UTR	*TAF7L* (2063), *DRP2* (71063)
cg01975035	chrX:24761808	13.389	4.05E-06	*SCARNA23;POLA1*	TSS1500;Body	*POLA1* (49745), *ARX* (272256)
cg06068891	chrX:135228207	13.38	4.07E-06	*FHL1*	TSS1500	*FHL1* (-1351)
cg05860920	chrX:11679501	13.344	4.20E-06	*ARHGAP6*	Body	*ARHGAP6* (4319), *AMELX* (367969)
cg03198117	chrX:152939967	13.302	4.36E-06	*PNCK*	TSS1500,TSS200	*PNCK* (-152)
cg03670113	chrX:153640589	13.287	4.42E-06	*TAZ;DNASE1L1*	Body;TSS200	*DNASE1L1* (-163), *TAZ* (713)
cg03035653	chrX:99662932	13.277	4.46E-06	*PCDH19*	1stExon	*PCDH19* (2338)
cg24636657	chrX:105280886	13.198	4.77E-06	*SERPINA7*	1stExon	*SERPINA7* (1831), *NRK* (214351)
cg03800724	chrX:53253348	13.186	4.82E-06	*KDM5C*	Body	*KDM5C* (1255), *TSPYL2* (141807)
cg04553232	chrX:8700576	13.168	4.89E-06	*KAL1*	TSS1500	*KAL1* (-350)
cg06138173	chrX:16668923	13.154	4.96E-06	*CTPS2;S100G*	Body;5'UTR	*S100G* (643)
cg12720637	chrX:151619769	13.145	4.99E-06	*GABRA3*	1stExon,5'UTR	*GABRA3* (61)
cg25831435	chrX:137714452	13.144	5.00E-06	*FGF13*	3'UTR	*FGF13* (572732)
cg22964346	chrX:109411307	13.131	5.06E-06	*TMEM164*	Body	*AMMECR1* (150072), *TMEM164* (164966)

Appendix D: Xi DMPs associated with CTL, MCI, and AD in AddNeuroMed.

DNA methylation differences on the Xi chromosome related to CTL, MCI, or AD status. All probes passing the Bonferroni-adjusted p-value threshold of $p < 5.18\times10^{-6}$ are shown by ProbeID, genomic location (genome build 37), ANOVA F-statistic and p-value, UCSC gene and group annotation, and GREAT annotation with distance to the nearest TSS in parentheses.

APPENDIX E: XI DMPS ASSOCIATED WITH AD IN ADDNEUROMED

CTL vs AD Xi DMPs						
ProbeID	**Position**	**Difference (CI)**	***p* CvA**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg22822140	chrX:153046577	-9.53 (-11.83 - -7.22)	4.92E-14	*SRPK3*	1stExon	*SRPK3* (122)
cg12468189	chrX:153046767	-5.4 (-6.94 - -3.86)	1.52E-13	*SRPK3*	Body	*SRPK3* (312)
cg12230162	chrX:153046482	-8.72 (-11.28 - -6.16)	4.81E-13	*SRPK3*	5'UTR,1stExon	*SRPK3* (27)
cg24496423	chrX:153046480	-6.66 (-8.68 - -4.63)	2.27E-12	*SRPK3*	5'UTR,1stExon	*SRPK3* (25)
cg19938385	chrX:107020894	-5.93 (-7.81 - -4.06)	1.24E-11			*TSC22D3* (-1878)
cg17405188	chrX:73164102	1.72 (1.17 - 2.27)	1.80E-11	*LOC554203*	TSS200	*ZCCHC13* (-359922), *CHIC1* (381119)
cg16529483	chrX:153046451	-7.68 (-10.23 - -5.12)	1.01E-10	*SRPK3*	TSS200	*SRPK3* (-4)
cg23914849	chrX:52950051	0.68 (0.45 - 0.91)	3.43E-10			*FAM156B* (21967), *FAM156A* (74599)
cg04710661	chrX:2848381	1.87 (1.19 - 2.56)	3.84E-09	*ARSD*	TSS1500	*ARSD* (-966)
cg04768884	chrX:19443172	3.71 (2.3 - 5.13)	1.34E-08	*MAP3K15*	Body	*PDHA1* (81162), *MAP3K15* (90206)
cg22574818	chrX:55479616	-6.02 (-8.34 - -3.7)	1.82E-08	*MAGEH1*	1stExon,3'UTR	*MAGEH1* (1079), *USP51* (36014)
cg09768654	chrX:153046386	-5.97 (-8.28 - -3.66)	2.11E-08	*SRPK3*	TSS200	*SRPK3* (-69)
cg02329520	chrX:153740317	-1.98 (-2.77 - -1.19)	4.73E-08	*FAM3A*	Body	*SLC10A3* (-21316), *FAM3A* (4248)
cg07344627	chrX:125687011	-6.12 (-8.61 - -3.62)	9.71E-08	*DCAF12L1*	TSS200	*DCAF12L1* (-170)
cg09661041	chrX:153141905	-6.06 (-8.54 - -3.59)	9.82E-08	*L1CAM*	TSS1500	*L1CAM* (-507)
cg00805156	chrX:3264341	4.09 (2.41 - 5.77)	1.27E-07	*MXRA5*	5'UTR	*MXRA5* (342)
cg06068202	chrX:150343148	4.71 (2.76 - 6.67)	1.51E-07			*GPR50* (-1907)
cg06616051	chrX:153094332	4.3 (2.48 - 6.13)	2.99E-07	*PDZD4*	Body	*IDH3G* (-34366), *PDZD4* (1670)
cg00123001	chrX:13395880	-1.96 (-2.79 - -1.13)	3.01E-07			*EGFL6* (-191813), *ATXN3L* (-57363)
cg25896901	chrX:125300481	-6.23 (-8.88 - -3.58)	3.25E-07	*DCAF12L2*	TSS1500	*DCAF12L1* (386360)
ch.X.1639000R	chrX:113225430	1.69 (0.97 - 2.41)	3.37E-07			*HTR2C* (-593120)
cg24352688	chrX:13751727	1.65 (0.94 - 2.35)	3.55E-07	*TRAPPC2; OFD1*	5'UTR;TSS1500	*OFD1* (-1104), *TRAPPC2* (1026)
cg08850124	chrX:118207749	-1.64 (-2.35 - -0.93)	4.60E-07			*PGRMC1* (-162461), *LONRF3* (99037)

CTL vs AD Xi DMPs (Continued)						
ProbeID	**Position**	**Difference (CI)**	***p* CvA**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg00072799	chrX:78427682	-3.38 (-4.83 - -1.92)	4.86E-07	*GPR174*	1stExon,3'UTR	*GPR174* (1214), *ITM2A* (195366)
cg12039689	chrX:153141750	-5.08 (-7.28 - -2.88)	4.94E-07	*L1CAM*	TSS1500	*L1CAM* (-352)
cg20814574	chrX:92928640	1.29 (0.73 - 1.86)	5.43E-07	*FAM133A;NAP1L3*	TSS1500;TSS200	*NAP1L3* (41)
cg00928753	chrX:36976515	5.02 (2.83 - 7.21)	6.26E-07			*PRRG1* (-232012), *CXorf59* (911463)
cg02048654	chrX:132551786	6.34 (3.48 - 9.2)	1.43E-06			*GPC4* (-2582)
cg16122592	chrX:26210712	-2.55 (-3.71 - -1.39)	1.60E-06	*MAGEB6*	5'UTR	*MAGEB6* (156)
cg26101161	chrX:122736584	-2.29 (-3.33 - -1.25)	1.70E-06	*THOC2*	3'UTR	*THOC2* (130319), *GRIA3* (418489)
cg01975035	chrX:24761808	-2.42 (-3.53 - -1.31)	1.95E-06	*SCARNA23;POLA1*	TSS1500;Body	*POLA1* (49745), *ARX* (272256)
cg24636657	chrX:105280886	-3.15 (-4.61 - -1.7)	2.39E-06	*SERPINA7*	1stExon	*SERPINA7* (1831), *NRK* (214351)
cg01961936	chrX:102349348	-2.02 (-2.96 - -1.08)	3.17E-06	*NXF3*	TSS1500	*NXF3* (-1327)
cg16170760	chrX:153708103	-1.81 (-2.65 - -0.96)	3.34E-06	*LAGE3*	TSS1500	*LAGE3* (-508)
cg20784188	chrX:44703983	-2.24 (-3.29 - -1.19)	3.73E-06	*DUSP21*	3'UTR;1stExon	*DUSP21* (735)
cg21364342	chrX:107226611	-2.22 (-3.27 - -1.17)	4.15E-06	*TEX13B*	TSS1500	*VSIG1* (-61588), *MID2* (157528)
cg22964346	chrX:109411307	-1.11 (-1.64 - -0.59)	4.18E-06	*TMEM164*	Body	*AMMECR1* (150072), *TMEM164* (164966)
cg26747413	chrX:15807140	-3.71 (-5.47 - -1.96)	4.23E-06	*INE2;ZRSR2*	TSS1500;TSS1500	*ZRSR2* (-1433)
cg00977690	chrX:24167334	1.14 (0.6 - 1.68)	4.98E-06			*ZFX* (-427)
cg02754763	chrX:44403388	-1.58 (-2.34 - -0.83)	5.16E-06	*FUNDC1*	TSS1500	*DUSP21* (-299860), *EFHC2* (-200466)

Appendix E: Xi DMPs associated with AD in AddNeuroMed.

DNA methylation differences on the Xi chromosome related to diagnosis of AD in comparison to CTL. All DMPs passing the Bonferroni-adjusted p-value threshold of $p < 5.18\times10^{-6}$ are shown by ProbeID, genomic location (genome build 37), the group difference in % methylation (AD – CTL), confidence interval (CI), and p-value as calculated with Tukey's HSD test, UCSC gene and group annotation, and GREAT annotation with distance to the nearest TSS in parentheses.

APPENDIX F: XI DMPS ASSOCIATED WITH MCI IN ADDNEUROMED

CTL vs MCI Xi DMPs						
ProbeID	**Position**	**Difference (CI)**	***p* CvM**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg24605338	chrX:17395856	-3.68 (-4.8 - -2.56)	2.63E-12	*NHS*	Body	*SCML1* (-359735), *NHS* (2314)
cg17003204	chrX:102861456	-3.76 (-5.02 - -2.5)	1.21E-10	*TCEAL3*	TSS1500	*TCEAL3* (-1377)
cg19280671	chrX:149013642	1.76 (1.17 - 2.35)	1.63E-10	*MAGEA8*	Body	*MAMLD1* (-517908), *MAGEA8* (3702)
cg16152676	chrX:10126878	4.18 (2.74 - 5.62)	4.19E-10	*CLCN4*	5'UTR	*CLCN4* (1894), *MID1* (724930)
cg17513789	chrX:73073251	-4.75 (-6.42 - -3.07)	9.34E-10	*XIST*	TSS1500	*ZCCHC13* (-450773), *CHIC1* (290268)
cg15907464	chrX:70712790	-3.53 (-4.8 - -2.27)	1.37E-09	*BCYRN1;ING X*	Body;TSS200	*OGT* (-40121), *TAF1* (126677)
cg14917571	chrX:79269151	-5.32 (-7.26 - -3.37)	3.18E-09	*TBX22*	TSS1500	*ITM2A* (-646103), *TBX22* (-8590)
cg16122592	chrX:26210712	-3.11 (-4.25 - -1.97)	3.57E-09	*MAGEB6*	5'UTR	*MAGEB6* (156)
cg06452970	chrX:13835264	5.1 (3.2 - 7.01)	6.56E-09	*GPM6B*	1stExon,5'UTR,Body	*GPM6B* (49)
cg09953945	chrX:150345040	-2.98 (-4.1 - -1.86)	7.92E-09	*GPR50*	TSS200	*GPR50* (-15)
cg17219209	chrX:149009840	2.15 (1.33 - 2.97)	1.39E-08	*MAGEA8*	TSS200	*MAGEA8* (-100)
cg26484667	chrX:2733164	-3.16 (-4.39 - -1.93)	2.30E-08	*XG*	3'UTR	*GYG2* (-13698), *XG* (63072)
cg08630881	chrX:70713213	-2.83 (-3.93 - -1.73)	2.38E-08	*BCYRN1;ING X*	Body;TSS1500	*OGT* (-39698), *TAF1* (127100)
cg04817724	chrX:103809991	5.25 (3.17 - 7.32)	3.94E-08	*IL1RAPL2*	TSS1500	*IL1RAPL2* (-1004)
cg19938385	chrX:107020894	-4.59 (-6.44 - -2.74)	7.28E-08			*TSC22D3* (-1878)
cg08867267	chrX:130033963	-2.5 (-3.53 - -1.47)	1.22E-07	*ENOX2*	5'UTR	*ENOX2* (3244), *RBMX2* (498021)
cg14306994	chrX:49019804	3.27 (1.92 - 4.62)	1.34E-07	*MAGIX*	Body,TSS1500	*GPKOW* (-39726), *PLP2* (-8379)
cg08464958	chrX:2852845	-2.28 (-3.23 - -1.34)	1.42E-07	*ARSE*	3'UTR	*ARSD* (-5430), *ARSE* (29648)
cg13009143	chrX:100548223	-1.68 (-2.37 - -0.98)	1.57E-07	*TAF7L*	TSS200	*TAF7L* (-165)
cg24834461	chrX:153637645	1.68 (0.98 - 2.38)	2.05E-07	*DNASE1L1*	TSS200,5'UTR	*TAZ* (-2231)
cg25433595	chrX:100183764	4.52 (2.62 - 6.43)	2.42E-07	*XKRX*	5'UTR,1stExon	*XKRX* (133)
cg11179997	chrX:119125670	2.7 (1.54 - 3.87)	4.48E-07			*NKAP* (-47936), *RHOXF2* (86036)
cg25396787	chrX:3733991	2.88 (1.63 - 4.13)	5.42E-07			*PRKX* (-102317)
cg25177692	chrX:129193893	-2.58 (-3.71 - -1.45)	6.61E-07			*ELF4* (50794), *BCORL1* (54730)
cg07089438	chrX:133988190	-2.62 (-3.77 - -1.47)	7.81E-07	*FAM122C*	3'UTR,Body	*PLAC1* (-195678), *MOSPD1* (61106)

CTL vs MCI Xi DMPs (Continued)						
ProbeID	**Position**	**Difference (CI)**	***p* CvM**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg09979033	chrX:16726904	-1.83 (-2.64 - -1.02)	8.24E-07	*CTPS2*	5'UTR	*CTPS2* (4154), *S100G* (58624)
cg26468081	chrX:69501565	-1.51 (-2.18 - -0.84)	8.84E-07	*ARR3;RAB41*	Body;TSS1500	*RAB41* (-456)
cg27558057	chrX:70712724	-3 (-4.33 - -1.67)	9.50E-07	*BCYRN1;INGX*	Body;TSS200	*OGT* (-40187), *TAF1* (126611)
cg21364342	chrX:107226611	-2.33 (-3.37 - -1.29)	9.94E-07	*TEX13B*	TSS1500	*VSIG1* (-61588), *MID2* (157528)
cg26101161	chrX:122736584	-2.31 (-3.34 - -1.28)	1.01E-06	*THOC2*	3'UTR	*THOC2* (130319), *GRIA3* (418489)
cg25958561	chrX:110187179	3.22 (1.79 - 4.65)	1.02E-06	*PAK3*	TSS1500	*PAK3* (-179125), *CHRDL1* (-148104)
cg00123001	chrX:13395880	-1.84 (-2.66 - -1.02)	1.07E-06			*EGFL6* (-191813), *ATXN3L* (-57363)
cg12461113	chrX:13587550	-8.01 (-11.6 - -4.42)	1.21E-06	*EGFL6*	TSS200	*EGFL6* (-143)
cg05889382	chrX:70887722	-2.09 (-3.03 - -1.15)	1.25E-06	*BCYRN1*	Body	*PIN4* (-513803), *CXCR3* (-49356)
cg07892139	chrX:84192519	-4.58 (-6.66 - -2.5)	1.76E-06			*APOOL* (-66378), *UBE2DNL* (3363)
cg16508637	chrX:132094977	2.67 (1.45 - 3.89)	1.93E-06	*HS6ST2*	5'UTR	*HS6ST2* (445)
cg00805156	chrX:3264341	3.64 (1.97 - 5.3)	1.96E-06	*MXRA5*	5'UTR	*MXRA5* (342)
ch.X.153018199F	chrX:153365005	-1.02 (-1.49 - -0.55)	2.18E-06			*MECP2* (-1818)
cg03035653	chrX:99662932	4.01 (2.16 - 5.86)	2.44E-06	*PCDH19*	1stExon	*PCDH19* (2338)
cg06068891	chrX:135228207	-4.97 (-7.27 - -2.68)	2.48E-06	*FHL1*	TSS1500	*FHL1* (-1351)
cg06138173	chrX:16668923	-3.18 (-4.64 - -1.71)	2.50E-06	*CTPS2;S100G*	Body;5'UTR	*S100G* (643)
cg25831435	chrX:137714452	-3.18 (-4.65 - -1.71)	2.70E-06	*FGF13*	3'UTR	*FGF13* (572732)
cg22926378	chrX:6144633	1.58 (0.84 - 2.31)	2.82E-06	*NLGN4X*	5'UTR	*NLGN4X* (2072)
ch.X.112118442F	chrX:112231786	1.28 (0.68 - 1.87)	2.96E-06			*AMOT* (-165415)
cg05860920	chrX:11679501	5.9 (3.14 - 8.65)	3.20E-06	*ARHGAP6*	Body	*ARHGAP6* (4319), *AMELX* (367969)
cg03800724	chrX:53253348	2.56 (1.36 - 3.76)	3.31E-06	*KDM5C*	Body	*KDM5C* (1255), *TSPYL2* (141807)

CTL vs MCI Xi DMPs (Continued)						
ProbeID	**Position**	**Difference (CI)**	***p* CvM**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg03449040	chrX:10094363	-1.39 (-2.04 - -0.74)	3.63E-06	*WWC3*	Body	*CLCN4* (-30621), *SHROOM2* (339868)
cg08333400	chrX:10126881	5.54 (2.93 - 8.14)	3.74E-06	*CLCN4*	5'UTR	*CLCN4* (1897), *MID1* (724927)
cg17397814	chrX:100670855	4.59 (2.43 - 6.76)	3.90E-06			*ARMCX1* (-134658), *HNRNPH2* (7735)
cg09641151	chrX:153693918	2.86 (1.51 - 4.21)	4.09E-06	*PLXNA3*	Body	*PLXNA3* (7296), *LAGE3* (13677)
cg11837390	chrX:106957165	-2.14 (-3.16 - -1.13)	4.49E-06	*TSC22D3*	3'UTR	*TSC22D3* (61851), *PRPS1* (85512)
cg05091491	chrX:114423926	-2.85 (-4.2 - -1.5)	4.57E-06	*RBMXL3;LRCH2*	TSS200;Body	*RBMXL3* (-36)
cg20881335	chrX:14547172	6.61 (3.47 - 9.74)	4.60E-06	*GLRA2*	TSS1500	*GLRA2* (-247)
cg06068202	chrX:150343148	4.06 (2.13 - 5.99)	4.68E-06			*GPR50* (-1907)
cg08119631	chrX:118822815	-2.71 (-4 - -1.42)	5.00E-06	*SEPT6*	Body	*NKRF* (-82970), *SEPT6* (4517)

Appendix F: Xi DMPs associated with MCI in AddNeuroMed.

DNA methylation differences on the Xi chromosome related to diagnosis of MCI in comparison to CTL. All DMPs passing the Bonferroni-adjusted p-value threshold of $p < 5.18\times10^{-6}$ are shown by ProbeID, genomic location (genome build 37), the group difference in % methylation (MCI – CTL), confidence interval (CI), and p-value as calculated with Tukey's HSD test, UCSC gene and group annotation, and GREAT annotation with distance to the nearest TSS in parentheses.

APPENDIX G: XI DMPS ASSOCIATED WITH AD RELATIVE TO MCI IN ADDNEUROMED

MCI vs AD Xi DMPs						
ProbeID	**Position**	**Difference (CI)**	***p* MvA**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg13009143	chrX:100548223	2.84 (2.15 - 3.53)	4.79E-14	*TAF7L*	TSS200	*TAF7L* (-165)
cg22822140	chrX:153046577	-8.86 (-11.12 - -6.59)	4.98E-14	*SRPK3*	1stExon	*SRPK3* (122)
cg12230162	chrX:153046482	-9.17 (-11.68 - -6.65)	6.34E-14	*SRPK3*	5'UTR,1stExon	*SRPK3* (27)
cg03360001	chrX:53453272	2.12 (1.53 - 2.71)	8.63E-14	*RIBC1*	Body	*SMC1A* (-3655)
cg24496423	chrX:153046480	-7 (-8.99 - -5.01)	1.45E-13	*SRPK3*	5'UTR,1stExon	*SRPK3* (25)
cg21252909	chrX:30326328	-7.69 (-10.01 - -5.37)	1.56E-12	*NR0B1*	1stExon	*NR0B1* (1166)
cg16152676	chrX:10126878	-4.73 (-6.17 - -3.3)	2.08E-12	*CLCN4*	5'UTR	*CLCN4* (1894), *MID1* (724930)
cg09768654	chrX:153046386	-7.07 (-9.35 - -4.8)	2.41E-11	*SRPK3*	TSS200	*SRPK3* (-69)
cg22926378	chrX:6144633	-2.13 (-2.86 - -1.41)	2.60E-10	*NLGN4X*	5'UTR	*NLGN4X* (2072)
cg25896901	chrX:125300481	-7.5 (-10.11 - -4.89)	5.28E-10	*DCAF12L2*	TSS1500	*DCAF12L1* (386360)
cg07069575	chrX:49688313	-3.39 (-4.6 - -2.17)	1.87E-09	*CLCN5*	5'UTR	*CLCN5* (1089), *AKAP4* (277350)
cg16529483	chrX:153046451	-6.89 (-9.4 - -4.38)	2.91E-09	*SRPK3*	TSS200	*SRPK3* (-4)
cg19280671	chrX:149013642	-1.59 (-2.18 - -1)	4.85E-09	*MAGEA8*	Body	*MAMLD1* (-517908), *MAGEA8* (3702)
cg00190642	chrX:103268431	4 (2.5 - 5.51)	8.25E-09	*H2BFWT; MIR1256*	TSS200;Body	*H2BFWT* (-176)
cg01737010	chrX:101772014	-2.47 (-3.45 - -1.5)	3.98E-08	*TMSB15A*	TSS1500	*TMSB15A* (-316)
cg13788827	chrX:12990209	2.32 (1.4 - 3.25)	4.96E-08			*TMSB4X* (-3016)
cg09513996	chrX:48685182	-3.47 (-4.86 - -2.08)	5.67E-08			*ERAS* (-2100)
cg06452970	chrX:13835264	-4.4 (-6.3 - -2.5)	4.50E-07	*GPM6B*	1stExon,5'UTR, Body	*GPM6B* (49)
cg26484667	chrX:2733164	2.82 (1.6 - 4.04)	5.13E-07	*XG*	3'UTR	*GYG2* (-13698), *XG* (63072)
cg10482495	chrX:71351314	2.13 (1.19 - 3.06)	7.13E-07	*NHSL2;RGAG4*	Body;1stExon	*CXCR3* (-512948), *PIN4* (-50211)
cg11179997	chrX:119125670	-2.64 (-3.8 - -1.48)	7.26E-07			*NKAP* (-47936), *RHOXF2* (86036)
cg26486175	chrX:18658942	-6.16 (-8.88 - -3.44)	8.42E-07	*CDKL5;RS1*	Body;3'UTR	*RS1* (31280), *CDKL5* (215218)
cg07395961	chrX:25041645	4.71 (2.62 - 6.79)	8.72E-07			*ARX* (-7581)
cg09626131	chrX:82764537	-4.15 (-5.99 - -2.31)	9.57E-07	*POU3F4*	3'UTR,1stExon	*CYLC1* (-351632), *POU3F4* (1269)

MCI vs AD Xi DMPs (Continued)						
ProbeID	**Position**	**Difference (CI)**	***p* MvA**	**UCSC Gene**	**UCSC Gene Group**	**GREAT Annotation**
cg24352688	chrX:13751727	1.56 (0.86 - 2.25)	1.03E-06	*TRAPPC2;OFD1*	5'UTR;TSS1500	*OFD1* (-1104), *TRAPPC2* (1026)
cg24401049	chrX:11157158	-6.43 (-9.34 - -3.53)	1.47E-06	*ARHGAP6*	Body	*AMELX* (-154374), *HCCS* (27753)
cg22090245	chrX:99661860	5.96 (3.26 - 8.67)	1.58E-06	*PCDH19*	1stExon	*PCDH19* (3410)
cg04867253	chrX:70472827	3.62 (1.98 - 5.27)	1.62E-06	*ZMYM3;BCYRN1*	Body;Body	*ZMYM3* (1210), *GJB1* (37766)
cg16801826	chrX:54587084	1.92 (1.04 - 2.79)	1.82E-06	*GNL3L*	3'UTR	*GNL3L* (30441), *ITIH6* (237588)
cg17003204	chrX:102861456	2.73 (1.48 - 3.99)	2.04E-06	*TCEAL3*	TSS1500	*TCEAL3* (-1377)
cg00695046	chrX:100545995	-6.63 (-9.7 - -3.56)	2.68E-06	*TAF7L*	Body,5'UTR	*TAF7L* (2063), *DRP2* (71063)
cg09167861	chrX:135251741	-2.1 (-3.07 - -1.12)	2.79E-06	*FHL1*	Body,1stExon, 5'UTR,TSS200	*FHL1* (22183), *MAP7D3* (81996)
cg24743423	chrX:8701002	-3.75 (-5.51 - -2)	3.36E-06	*KAL1*	TSS1500	*KAL1* (-776)
cg09661041	chrX:153141905	-5.18 (-7.61 - -2.75)	3.54E-06	*L1CAM*	TSS1500	*L1CAM* (-507)
cg03198117	chrX:152939967	-3.91 (-5.75 - -2.07)	4.03E-06	*PNCK*	TSS1500,TSS200	*PNCK* (-152)
cg17662252	chrX:54560704	2.92 (1.54 - 4.29)	4.29E-06	*GNL3L*	Body	*GNL3L* (4061), *ITIH6* (263968)
cg08333400	chrX:10126881	-5.46 (-8.06 - -2.87)	4.52E-06	*CLCN4*	5'UTR	*CLCN4* (1897), *MID1* (724927)
cg17219209	chrX:149009840	-1.72 (-2.54 - -0.9)	4.96E-06	*MAGEA8*	TSS200	*MAGEA8* (-100)

Appendix G: Xi DMPs associated with AD relative to MCI in AddNeuroMed.

DNA methylation differences on the Xi chromosome related to diagnosis of AD in comparison to MCI. All DMPs passing the Bonferroni-adjusted p-value threshold of $p < 5.18\times10^{-6}$ are shown by ProbeID, genomic location (genome build 37), the group difference in % methylation (AD – MCI), confidence interval (CI), and p-value as calculated with Tukey's HSD test, UCSC gene and group annotation, and GREAT annotation with distance to the nearest TSS in parentheses.

APPENDIX H: DMRS ON THE XI CHROMOSOME ASSOCIATED WITH DIAGNOSTIC STATUS OF CTL, MCI, OR AD

ANOVA Xi DMRs							
Gene	**Position**	***n***	***p*-value**	**Šidák-*P***	**Average Methylation %**		
					CTL	**MCI**	**AD**
SRPK3	chrX:153046175 - 153046895	9	3.27E-44	4.37E-43	1.46	1.25	-3.9
TAF1;INGX	chrX:70712215 - 70713214	9	1.39E-18	1.34E-17	1.29	-1.32	0.57
CLCN4	chrX:10126321 - 10126882	5	8.3E-15	1.43E-13	-0.73	0.56	-0.07
LOC389906	chrX:3732500 - 3734340	4	4.95E-14	2.6E-13	-1.27	0.77	0.37
PCDH19	chrX:99661860 - 99663331	6	2.82E-11	1.85E-10	-1.22	-0.03	1.41
PNCK	chrX:152938698 - 152940275	11	1.03E-10	6.3E-10	0.2	-0.12	-0.15
FHL1	chrX:135228207 - 135228445	3	3.06E-11	1.24E-09	1.95	-1.6	0.03
FGF13	chrX:137950565 - 137951119	3	1.04E-10	1.81E-09	1.59	-1.2	-0.11
L1CAM	chrX:153141451 - 153141906	5	1.46E-10	3.09E-09	0.61	0.05	-0.77
OFD1;TRAPPC2	chrX:13751423 - 13753382	9	1.3E-09	6.41E-09	-0.27	0.12	0.12
MAGIX	chrX:49019804 - 49021149	4	1.92E-09	1.38E-08	-0.13	-0.06	0.18
DUSP21	chrX:44703819 - 44703984	2	2.8E-10	1.64E-08	0.97	-0.04	-1.02
ASB11	chrX:15333348 - 15334256	5	2.17E-09	2.31E-08	1.05	-0.73	-0.22
XIST	chrX:73073251 - 73073712	3	3.62E-09	7.58E-08	1.28	-0.76	-0.41
PRAF2	chrX:48929808 - 48930361	3	7.66E-09	1.34E-07	-0.19	1.73	-2.45
ZRSR2	chrX:15807140 - 15808693	9	3.29E-08	2.04E-07	0.03	0.04	-0.16
MAGEE2	chrX:75004943 - 75005361	8	9.36E-09	2.16E-07	1.82	-0.57	-1.06
FAM156A;FAM156B	chrX:52950051 - 52950333	2	6.8E-09	2.33E-07	-0.21	-0.14	0.38
TEX13B	chrX:107226253 - 107226612	2	9.73E-09	2.61E-07	0.9	-0.63	-0.03
TAF7L	chrX:100547965 - 100548224	5	7.3E-09	2.72E-07	0.31	-0.35	0.17
CT45A1	chrX:134846527 - 134847107	3	2.27E-08	3.78E-07	0.98	-1.01	0.21
S100G;CTPS2	chrX:16668107 - 16668981	4	6.65E-08	7.34E-07	1.01	-0.68	-0.2
ARHGAP6	chrX:11157142 - 11157611	5	4.51E-08	9.28E-07	-0.29	1.8	-1.91
MAGEA8-AS1	chrX:149009554 - 149009914	4	3.53E-08	9.45E-07	-1.08	0.89	-0.29
RENBP	chrX:153211181 - 153211712	3	8.07E-08	1.47E-06	0.09	0.45	-0.62
SOX3	chrX:139592127 - 139592513	3	8.94E-08	2.23E-06	0.12	0.58	-0.84

ANOVA Xi DMRs (Continued)

Gene	Position	*n*	*p*-value	Šidák-*P*	Average Methylation %		
					CTL	MCI	AD
ANOS1	chrX:8700172 - 8701003	10	2.95E-07	3.42E-06	-0.26	0.64	-0.55
NAP1L3;FAM133A	chrX:92928229 - 92929039	10	3.4E-07	4.04E-06	-0.44	0.29	0.17
JPX	chrX:73164102 - 73164391	6	1.27E-07	4.22E-06	-0.36	0.04	0.36
ARSD	chrX:2847354 - 2848382	9	5.26E-07	4.93E-06	-0.77	0.33	0.4
RHOXF1P1	chrX:119125659 - 119125723	3	4.77E-08	7.2E-06	-0.2	0.98	-1.2
ZNF75D	chrX:134429653 - 134429973	4	2.95E-07	8.89E-06	2.05	-0.78	-0.97
TXLNG	chrX:16804211 - 16805692	10	1.44E-06	9.39E-06	-0.15	0.14	0.01
TNIP2P1	chrX:40693177 - 40693704	3	6.75E-07	1.24E-05	0.73	-0.13	-0.57
POU3F4	chrX:82764385 - 82764538	2	2.53E-07	1.59E-05	0.16	1.45	-2.23
IL1RAPL2	chrX:103809062 - 103810167	6	1.91E-06	1.66E-05	-0.88	0.56	0.05
CAPN6	chrX:110513791 - 110514232	5	9.97E-07	2.18E-05	-1.83	0.77	0.95
DANT1;DANT2	chrX:114957471 - 114957791	2	8.58E-07	2.59E-05	-1.37	0.65	0.57
BCOR	chrX:40004432 - 40004482	2	1.5E-07	2.89E-05	1.11	0.72	-2.15
NHS	chrX:17395816 - 17395857	2	1.37E-07	3.23E-05	0.27	-0.49	0.33
FAM47A	chrX:34165195 - 34165644	2	3.05E-06	6.54E-05	0.17	-0.17	0
GPM6B	chrX:13835198 - 13835569	5	2.73E-06	7.09E-05	-0.71	0.46	0.15
CA5BP1-CA5B;CA5BP1	chrX:15692851 - 15693957	11	1.14E-05	9.9E-05	-0.4	-0.13	0.57
ATP7A	chrX:77225299 - 77225754	2	6.24E-06	0.000132	0.81	-1.06	0.49
HS6ST2	chrX:132094977 - 132095283	2	4.57E-06	0.000144	-1.1	1.02	-0.17
PAGE5	chrX:55246455 - 55247169	3	1.3E-05	0.000175	0.63	0.25	-1.09
CDX4	chrX:72668487 - 72668848	3	7.62E-06	0.000204	-1.09	0.49	0.41
MAGEH1	chrX:55478539 - 55479617	5	2.53E-05	0.000226	0.25	0.28	-0.56
NXF3	chrX:102347925 - 102348040	3	2.99E-06	0.000251	1.29	-0.29	-1.01
PNMA3	chrX:152224966 - 152225367	2	1.09E-05	0.000263	-1.36	1.9	-1.15
KIAA1210	chrX:118284520 - 118285065	10	1.55E-05	0.000275	0.91	0.43	-1.98
PCDH19	chrX:99194929 - 99195116	2	5.96E-06	0.000308	-2.82	1.1	1.86
PDZD4	chrX:153094195 - 153094595	3	1.34E-05	0.000323	-0.37	-0.8	1.42

ANOVA Xi DMRs (Continued)

Gene	Position	*n*	*p*-value	Šidák-*P*	Average Methylation %		
					CTL	MCI	AD
SMC1A;RIBC1	chrX:53449152 - 53449901	11	3.26E-05	0.00042	-1.05	0.21	0.81
GPR50-AS1	chrX:150343097 - 150343149	2	2.7E-06	0.0005	-1.97	0.99	0.74
PTCHD1-AS	chrX:23016709 - 23017134	2	2.42E-05	0.000549	0.79	-0.05	-0.81
DENND10P1	chrX:129657974 - 129658062	2	6.41E-06	0.000702	-1.73	1.73	-0.65
RTL5;NHSL2	chrX:71350942 - 71351315	2	2.92E-05	0.000755	-0.73	0.33	0.53
GPC4	chrX:132551364 - 132551787	2	3.71E-05	0.000846	-1.62	0.24	1.44
CENPI	chrX:100446496 - 100446737	3	2.13E-05	0.000852	0.92	0.56	-1.69
TCEAL7	chrX:102584953 - 102585355	6	3.97E-05	0.000951	-1.82	1.26	0.06
KLHL34	chrX:21676081 - 21676872	8	9.02E-05	0.0011	0.41	-0.6	0.35
PAK3	chrX:110186893 - 110187601	8	9.05E-05	0.001232	-0.74	0.58	-0.03
STS	chrX:7066487 - 7066774	3	4.27E-05	0.001435	-0.33	0	0.31
MAP3K15	chrX:19443172 - 19443223	2	7.91E-06	0.001496	-0.81	-0.6	1.64
FRMPD3-AS1	chrX:106749745 - 106750343	5	0.000102	0.00164	-0.96	-0.75	1.92
FAM50A	chrX:153674100 - 153674275	2	3.07E-05	0.001689	-0.02	0.47	-0.61
SERPINA7	chrX:105280566 - 105280887	2	5.95E-05	0.001788	1.36	-0.45	-0.83
PRRG1	chrX:37208111 - 37208438	5	6.16E-05	0.001815	0.59	-0.71	0.28
UBA1	chrX:47053017 - 47053605	9	0.000112	0.001827	-0.21	0.1	0.07
TKTL1	chrX:153533256 - 153533552	3	6.11E-05	0.00199	1.21	-0.15	-1.12
BCORL1	chrX:129193893 - 129194323	4	9.03E-05	0.002024	0.63	-0.69	0.22
MXRA5	chrX:3264341 - 3264729	6	8.74E-05	0.002171	-0.88	0.21	0.7
H2BW1	chrX:103268125 - 103268177	2	1.18E-05	0.002189	0.36	-1.59	1.57
ERAS	chrX:48685943 - 48686201	3	6.33E-05	0.002366	1.95	-0.98	-0.28
GPR50-AS1	chrX:150344901 - 150345041	2	3.49E-05	0.002399	0.87	-0.78	0.13
SH2D1A	chrX:123480277 - 123480595	3	8.18E-05	0.002479	-0.09	-0.42	0.57
NR0B1	chrX:30326328 - 30326677	3	9.85E-05	0.002718	-0.21	0.89	-1.09
WDR44	chrX:117479618 - 117480247	9	0.000178	0.00273	-0.28	-0.2	0.5
TCEAL8	chrX:102510111 - 102510315	6	5.79E-05	0.002733	-0.71	0.65	-0.02

ANOVA Xi DMRs (Continued)

Gene	Position	*n*	*p*-value	Šidák-*P*	Average Methylation %		
					CTL	MCI	AD
CLTRN	chrX:15683007 - 15683428	6	0.000125	0.002855	0.54	0.03	-0.61
SHROOM2	chrX:9858470 - 9858994	6	0.000169	0.003114	0.27	-0.07	-0.22
ZIC3	chrX:136649093 - 136649490	2	0.000151	0.003664	-0.38	0.65	-0.34
MTCP1;CMC4	chrX:154293390 - 154293851	2	0.000181	0.003781	0.89	-0.25	-0.48
BCOR	chrX:40017092 - 40017470	3	0.000222	0.005655	-1.08	1.03	-0.19
MAGEA5;MAGEA10-MAGEA5	chrX:151286380 - 151286549	2	9.96E-05	0.00567	0.93	-1.04	0.45
GPC3	chrX:133118088 - 133118345	3	0.000155	0.005785	-1.61	0.56	0.92
IGSF1	chrX:130423014 - 130423459	9	0.000274	0.005917	-0.95	0.17	0.79
EOLA2;LINC00894	chrX:149106299 - 149107030	13	0.000483	0.006354	-0.45	0.09	0.36
PNCK	chrX:152931217 - 152931545	2	0.000221	0.006491	-1.42	0.51	0.85
WWC3	chrX:10094051 - 10094364	3	0.000222	0.006824	0.27	-0.44	0.29
CLCN5	chrX:49776519 - 49776568	2	3.52E-05	0.00691	0.96	-0.54	-0.33
LINC00269	chrX:68441389 - 68441691	2	0.000218	0.006937	-0.01	1.53	-1.68
MIR450A1	chrX:133674329 - 133674439	2	9.03E-05	0.007891	1.15	-0.64	-0.31
ZCCHC12	chrX:117958532 - 117958794	2	0.000222	0.008128	1.02	-0.83	-0.08
TTC3P1	chrX:74963322 - 74963720	2	0.000356	0.008598	0.13	-0.17	-0.03
FIRRE	chrX:130964695 - 130965144	6	0.000441	0.009442	-0.49	0.12	0.45
SLITRK4	chrX:142723457 - 142723827	4	0.000374	0.009711	-0.72	-0.35	1.34
LOC101059915	chrX:70887583 - 70887941	4	0.000372	0.00997	0.45	-0.07	-0.41
PRRG3	chrX:150865927 - 150866111	2	0.000203	0.01059	0.18	-0.24	0.06
AFF2	chrX:147582939 - 147583280	3	0.000402	0.01131	-0.02	-0.1	0.22
GS1-594A7.3	chrX:15621084 - 15621478	3	0.000477	0.0116	0.77	-0.48	-0.18
FHL1	chrX:135251416 - 135251742	4	0.000432	0.01269	-0.28	0.63	-0.59
TSPAN6	chrX:99891789 - 99892068	7	0.000369	0.01269	-0.33	-0.11	0.48
BEX1	chrX:102319998 - 102320381	2	0.000584	0.01461	1.06	-0.48	-0.46
MAGEB6	chrX:26210302 - 26210713	7	0.000684	0.01592	0.37	0.02	-0.42
GLRA2	chrX:14547147 - 14547321	3	0.00029	0.01596	-1.92	1.32	-0.03

ANOVA Xi DMRs (Continued)							
Gene	**Position**	***n***	***p*-value**	**Šidák-*P***	**Average Methylation %**		
					CTL	**MCI**	**AD**
KDM5C	chrX:53254643 - 53254819	7	0.000312	0.01698	-0.26	0.16	0.08
MAMLD1	chrX:149613398 - 149613746	4	0.00063	0.01731	-1.09	0.7	0.33
FRMD7	chrX:131263364 - 131263424	2	0.000125	0.01984	0.6	-0.49	0.02
NDUFB11	chrX:47003160 - 47003363	2	0.000468	0.022	0.62	0.04	-0.73
GAGE10	chrX:49159887 - 49160074	3	0.00047	0.02395	0.24	0.2	-0.54
ZIC3	chrX:136632433 - 136632533	2	0.000255	0.02426	0.43	0.14	-0.36
STS	chrX:7070284 - 7070452	2	0.000453	0.02569	-0.61	0.33	0.2
PAGE2	chrX:55115294 - 55115457	5	0.000458	0.02676	0.9	-0.52	-0.13
DCAF12L1	chrX:125687720 - 125687857	2	0.000486	0.03363	1.21	0.28	-1.87
PCDH19	chrX:99665948 - 99666347	5	0.001655	0.03927	0.39	-0.43	0.15
ARL13A	chrX:100224631 - 100224723	2	0.000397	0.04076	1.26	-0.66	-0.51
MIR507	chrX:146312598 - 146312952	3	0.001874	0.04984	0.68	-0.71	q0.24

Appendix H: DMRs on the Xi chromosome associated with diagnostic status of CTL, MCI, or AD.

Shown are DMRs for the overall three group (ANOVA) comparison. Displayed for each region is the UCSC gene name, chromosomal position (genome build 37), number of probes in region (n), p-value and multiple testing-corrected p (Šidák-p), and average relative methylation values per group.

REFERENCES

2020 Alzheimer's disease facts and figures. (2020). *Alzheimer's & Dementia*, *16*(3), 391–460. https://doi.org/https://doi.org/10.1002/alz.12068

Abruzzo, M. A., Mayer, M., & Jacobs, P. A. (1985). Aging and aneuploidy: evidence for the preferential involvement of the inactive X chromosome. *Cytogenetic and Genome Research*, *39*(4), 275–278. https://doi.org/10.1159/000132157

Agís-Balboa, R. C., Pinheiro, P. S., Rebola, N., Kerimoglu, C., Benito, E., Gertig, M., Bahari-Javan, S., Jain, G., Burkhardt, S., Delalle, I., Jatzko, A., Dettenhofer, M., Zunszain, P. A., Schmitt, A., Falkai, P., Pape, J. C., Binder, E. B., Mulle, C., Fischer, A., & Sananbenesi, F. (2017). Formin 2 links neuropsychiatric phenotypes at young age to an increased risk for dementia. *The EMBO Journal*, *36*(19), 2815–2828. https://doi.org/10.15252/embj.201796821

Ahuja, N., & Issa, J.-P. J. (1986). *Histology and histopathology*. F. Hernández. https://digitum.um.es/digitum/handle/10201/95166

Akbari, V., Garant, J.-M., O'Neill, K., Pandoh, P., Moore, R., Marra, M. A., Hirst, M., & Jones, S. J. M. (2021). Megabase-scale methylation phasing using nanopore long reads and NanoMethPhase. *Genome Biology*, *22*(1), 68. https://doi.org/10.1186/s13059-021-02283-5

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Gamst, A., Holtzman, D. M., Jagust, W. J., Petersen, R. C., Snyder, P. J., Carrillo, M. C., Thies, B., & Phelps, C. H. (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*, *7*(3), 270–279. https://doi.org/10.1016/j.jalz.2011.03.008

Alford, S., Patel, D., Perakakis, N., & Mantzoros, C. S. (2018). Obesity as a risk factor for Alzheimer's disease: weighing the evidence. *Obesity Reviews*, *19*(2), 269–280. https://doi.org/10.1111/obr.12629

Altuna, M., Urdánoz-Casado, A., Sánchez-Ruiz De Gordoa, J., Zelaya, M. V., Labarga, A., Lepesant, J. M. J., Roldán, M., Blanco-Luquin, I., Perdones, Á., Larumbe, R., Jericó, I., Echavarri, C., Méndez-López, I., Di Stefano, L., & Mendioroz, M. (2019). DNA methylation signature of human hippocampus in

Alzheimer's disease is linked to neurogenesis. *Clinical Epigenetics*, *11*(1), 91. https://doi.org/10.1186/s13148-019-0672-7

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

American Psychiatric Association. (2014). Cautionary Statement for Forensic Use of DSM-5. In *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition*. https://doi.org/10.1176/appi.books.9780890425596.744053

Amieva, H., Le Goff, M., Millet, X., Orgogozo, J. M., Pérès, K., Barberger-Gateau, P., Jacqmin-Gadda, H., & Dartigues, J. F. (2008). Prodromal Alzheimer's disease: Successive emergence of the clinical symptoms. *Annals of Neurology*, *64*(5), 492–498. https://doi.org/10.1002/ana.21509

Aravin, A. A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K. F., Bestor, T., & Hannon, G. J. (2008). A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice. *Molecular Cell*, *31*(6), 785–799. https://doi.org/10.1016/j.molcel.2008.09.003

Arends, Y. M., Duyckaerts, C., Rozemuller, J. M., Eikelenboom, P., & Hauw, J.-J. (2000). Microglia, amyloid and dementia in Alzheimer disease: A correlative study. *Neurobiology of Aging*, *21*(1), 39–47. https://doi.org/https://doi.org/10.1016/S0197-4580(00)00094-4

Arriagada, P. V., Growdon, J. H., Hedley-Whyte, E. T., & Hyman, B. T. (1992). Neurofibrillary tangles but not senile plaques parallel duration and severity of Alzheimer's disease. *Neurology*, *42*(3), 631–639. https://doi.org/10.1212/wnl.42.3.631

Asare, A. L., Kolchinsky, S. A., Gao, Z., Wang, R., Raddassi, K., Bourcier, K., & Seyfert-Margolis, V. (2008). Differential gene expression profiles are dependent upon method of peripheral blood collection and RNA isolation. *BMC Genomics*, *9*(1), 474. https://doi.org/10.1186/1471-2164-9-474

Austad, S. N. (2006). Why women live longer than men: Sex differences in longevity. *Gender Medicine*, *3*(2), 79–92. https://doi.org/https://doi.org/10.1016/S1550-8579(06)80198-1

Bajic, V., Mandusic, V., Stefanova, E., Bozovic, A., Davidovic, R., Zivkovic, L., Cabarkapa, A., & Spremo-Potparevic, B. (2015). Skewed X-Chromosome Inactivation in Women Affected by Alzheimer's Disease. *Journal of Alzheimer's Disease*, *43*(4), 1251–1259. https://doi.org/10.3233/JAD-141674

Bakulski, K. M., Dolinoy, D. C., Sartor, M. A., Paulson, H. L., Konen, J. R., Lieberman, A. P., Albin, R. L., Hu, H., & Rozek, L. S. (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*, *29*(3), 571–588. https://doi.org/10.3233/JAD-2012-111223

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, *21*(3), 381–395. https://doi.org/10.1038/cr.2011.22

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

Beyreuther, K., & Masters, C. L. (1991). Amyloid Precursor Protein (APP) and BZA4 Amyloid in the Etiology of Alzheimer's Disease: Precursor-Product Relationships in the Derangement of Neuronal Function. *Brain Pathology*, *1*(4), 241–251. https://doi.org/10.1111/j.1750-3639.1991.tb00667.x

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J. M., Delano, D., Zhang, L., Schroth, G. P., Gunderson, K. L., Fan, J. B., & Shen, R. (2011). High density DNA methylation array with single CpG site resolution. *Genomics*, *98*(4), 288–295. https://doi.org/10.1016/j.ygeno.2011.07.007

Bibl, M., Gallus, M., Welge, V., Lehmann, S., Sparbier, K., Esselmann, H., & Wiltfang, J. (2012). Characterization of cerebrospinal fluid aminoterminally truncated and oxidized amyloid-β peptides. *PROTEOMICS – Clinical Applications*, *6*(3–4), 163–169. https://doi.org/https://doi.org/10.1002/prca.201100082

Birney, E., Smith, G. D., & Greally, J. M. (2016). Epigenome-wide Association Studies and the Interpretation of Disease -Omics. *PLoS Genetics*, *12*(6), e1006105–e1006105. https://doi.org/10.1371/journal.pgen.1006105

Bjornsson, H. T., Sigurdsson, M. I., Fallin, M. D., Irizarry, R. A., Aspelund, T., Cui, H., Yu, W., Rongione, M. A., Ekström, T. J., Harris, T. B., Launer, L. J., Eiriksdottir, G., Leppert, M. F., Sapienza, C., Gudnason, V., & Feinberg, A. P. (2008). Intra-individual change over time in DNA methylation with familial clustering. *JAMA*, *299*. https://doi.org/10.1001/jama.299.24.2877

Blacker, D., Haines, J. L., Rodes, L., Terwedow, H., Go, R. C. P., Harrell, L. E., Perry, R. T., Bassett, S. S., Chase, G., Meyers, D., Albert, M. S., & Tanzi, R.

(1997). ApoE-4 and age at onset of Alzheimer's disease: The NIMH genetics initiative. *Neurology*, *48*(1), 139–147. https://doi.org/10.1212/WNL.48.1.139

Blalock, E. M., Buechel, H. M., Popovic, J., Geddes, J. W., & Landfield, P. W. (2011). Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient Alzheimer's disease. *Journal of Chemical Neuroanatomy*, *42*(2), 118–126. https://doi.org/10.1016/j.jchemneu.2011.06.007

Bloudek, L. M., Spackman, D. E., Blankenburg, M., & Sullivan, S. D. (2011). Review and Meta-Analysis of Biomarkers and Diagnostic Imaging in Alzheimer's Disease. *Journal of Alzheimer's Disease*, *26*, 627–645. https://doi.org/10.3233/JAD-2011-110458

Boks, M. P., de Jong, N. M., Kas, M. J. H., Vinkers, C. H., Fernandes, C., Kahn, R. S., Mill, J., & Ophoff, R. A. (2012). Current status and future prospects for epigenetic psychopharmacology. *Epigenetics*, *7*(1), 20–28. https://doi.org/10.4161/epi.7.1.18688

Bonasio, R., Tu, S., & Reinberg, D. (2010). Molecular signals of epigenetic states. *Science*, *330*(6004), 612–616. https://doi.org/10.1126/science.1191078

Booij, B. B., Lindahl, T., Wetterberg, P., Skaane, N. V., Sæbø, S., Feten, G., Ryea, P. D., Kristiansen, L. I., Hagen, N., Jensen, M., Bårdsen, K., Winblad, B., Sharma, P., & Lönneborga, A. (2011). A gene expression pattern in blood for the early detection of Alzheimer's disease. *Journal of Alzheimer's Disease*, *23*(1), 109–119. https://doi.org/10.3233/JAD-2010-101518

Braak, H., & Braak, E. (1995). Staging of alzheimer's disease-related neurofibrillary changes. *Neurobiology of Aging*, *16*(3), 271–278. https://doi.org/https://doi.org/10.1016/0197-4580(95)00021-6

Braak, H., & Del Tredici, K. (2011). The pathological process underlying Alzheimer's disease in individuals under thirty. *Acta Neuropathologica*, *121*(2), 171–181. https://doi.org/10.1007/s00401-010-0789-4

Braak, H., & Del Tredici, K. (2014). Are cases with tau pathology occurring in the absence of Aβ deposits part of the AD-related pathological process? *Acta Neuropathologica*, *128*(6), 767–772. https://doi.org/10.1007/s00401-014-1356-1

Braak, H., Thal, D. R., Ghebremedhin, E., & Del Tredici, K. (2011). Stages of the pathologic process in alzheimer disease: Age categories from 1 to 100 years. *Journal of Neuropathology and Experimental Neurology*, *70*(11), 960–969.

https://doi.org/10.1097/NEN.0b013e318232a379

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

Breiling, A., & Lyko, F. (2015). Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics and Chromatin*, *8*(1), 24. https://doi.org/10.1186/s13072-015-0016-6

Brion, J. P., Octave, J. N., & Couck, A. M. (1994). Distribution of the phosphorylated microtubule-associated protein tau in developing cortical neurons. *Neuroscience*, *63*(3), 895–909. https://doi.org/10.1016/0306-4522(94)90533-9

Brohede, J., Rinde, M., Winblad, B., & Graff, C. (2010). A DNA methylation study of the amyloid precursor protein gene in several brain regions from patients with familial alzheimer disease. *Journal of Neurogenetics*, *24*(4), 179–181. https://doi.org/10.3109/01677063.2010.503978

Brown, C. J., & Robinson, W. P. (2000). The causes and consequences of random and non-random X chromosome inactivation in humans. *Clinical Genetics*, *58*(5), 353–363. https://doi.org/https://doi.org/10.1034/j.1399-0004.2000.580504.x

Brunnström, H. R., & Englund, E. M. (2009). Cause of death in patients with dementia disorders. *European Journal of Neurology*, *16*(4), 488–492. https://doi.org/10.1111/j.1468-1331.2008.02503.x

Bryan, K. J., Lee, H., Perry, G., Smith, M. A., & Casadesus, G. (2009). Transgenic Mouse Models of Alzheimer's Disease: Behavioral Testing and Considerations. In *Methods of Behavior Analysis in Neuroscience*. http://www.ncbi.nlm.nih.gov/pubmed/21204338

Burke, S. L., Cadet, T., Alcide, A., O'Driscoll, J., & Maramaldi, P. (2018). Psychosocial risk factors and Alzheimer's disease: the associative effect of depression, sleep disturbance, and anxiety. *Aging and Mental Health*, *22*(12), 1577–1584. https://doi.org/10.1080/13607863.2017.1387760

Cagnin, A., Brooks, D. J., Kennedy, A. M., Gunn, R. N., Myers, R., Turkheimer, F. E., Jones, T., & Banati, R. B. (2001). In-vivo measurement of activated microglia in dementia. *Lancet*, *358*(9280), 461–467. https://doi.org/10.1016/S0140-6736(01)05625-2

Caruso, A., Nicoletti, F., Mango, D., Saidi, A., Orlando, R., & Scaccianoce, S.

(2018). Stress as risk factor for Alzheimer's disease. *Pharmacological Research*, *132*, 130–134. https://doi.org/10.1016/j.phrs.2018.04.017

Celarain, N., Sánchez-Ruiz de Gordoa, J., Zelaya, M. V., Roldán, M., Larumbe, R., Pulido, L., Echavarri, C., & Mendioroz, M. (2016). TREM2 upregulation correlates with 5-hydroxymethycytosine enrichment in Alzheimer's disease hippocampus. *Clinical Epigenetics*, *8*(1), 37. https://doi.org/10.1186/s13148-016-0202-9

Chanda, K., & Mukhopadhyay, D. (2020). LncRNA Xist, X-chromosome Instability and Alzheimer's Disease. *Current Alzheimer Research*, *17*(6), 499–507.

Chartier-Hariln, M.-C., Parfitt, M., Legrain, S., Pérez-Tur, J., Brousseau, T., Evans, A., Berr, C., Vldal, O., Roques, P., Gourlet, V., Fruchart, J.-C., Delacourte, A., Rossor, M., & Amouyel, P. (1994). Apolipoprotein E, ε4 allele as a major risk factor for sporadic early and late-onset forms of Alzheimer's disease: analysis of the 19q13.2 chromosomal region. *Human Molecular Genetics*, *3*(4), 569–574. https://doi.org/10.1093/hmg/3.4.569

Chen, B. H., Marioni, R. E., Colicino, E., Peters, M. J., Ward-Caviness, C. K., Tsai, P. C., Roetker, N. S., Just, A. C., Demerath, E. W., Guan, W., Bressler, J., Fornage, M., Studenski, S., Vandiver, A. R., Moore, A. Z., Tanaka, T., Kiel, D. P., Liang, L., Vokonas, P., … Horvath, S. (2016). DNA methylation-based measures of biological age: Meta-analysis predicting time to death. *Aging*, *8*(9), 1844–1865. https://doi.org/10.18632/aging.101020

Chen, G. F., Xu, T. H., Yan, Y., Zhou, Y. R., Jiang, Y., Melcher, K., & Xu, H. E. (2017). Amyloid beta: Structure, biology and structure-based therapeutic development. *Acta Pharmacologica Sinica*, *38*(9), 1205–1235. https://doi.org/10.1038/aps.2017.28

Chen, J. H., Lin, K. P., & Chen, Y. C. (2009). Risk factors for dementia. *Journal of the Formosan Medical Association*, *108*(10), 754–764. https://doi.org/10.1016/S0929-6646(09)60402-2

Chen, W., Ji, H., Li, L., Xu, C., Zou, T., Cui, W., Xu, S., Zhou, X., Duan, S., & Wang, Q. (2019). Significant association between GPR50 hypomethylation and AD in males. *Mol Med Rep*, *20*(2), 1085–1092. https://doi.org/10.3892/mmr.2019.10366

Chen, Y. A., Lemire, M., Choufani, S., Butcher, D. T., Grafodatskaya, D., Zanke, B. W., Gallinger, S., Hudson, T. J., & Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium

HumanMethylation450 microarray. *Epigenetics*, *8*(2), 203–209. https://doi.org/10.4161/epi.23470

Chen, Z. X., & Riggs, A. D. (2011). DNA methylation and demethylation in mammal. *Journal of Biological Chemistry*, *286*(21), 18347–18353. https://doi.org/10.1074/jbc.R110.205286

Chêne, G., Beiser, A., Au, R., Preis, S. R., Wolf, P. A., Dufouil, C., & Seshadri, S. (2015). Gender and incidence of dementia in the Framingham Heart Study from mid-adult life. *Alzheimer's & Dementia*, *11*(3), 310–320. https://doi.org/https://doi.org/10.1016/j.jalz.2013.10.005

Chételat, G., La Joie, R., Villain, N., Perrotin, A., De La Sayette, V., Eustache, F., & Vandenberghe, R. (2013). Amyloid imaging in cognitively normal individuals, at-risk populations and preclinical Alzheimer's disease. *NeuroImage: Clinical*, *2*(1), 356–365. https://doi.org/10.1016/j.nicl.2013.02.006

Chouliaras, L., Mastroeni, D., Delvaux, E., Grover, A., Kenis, G., Hof, P. R., Steinbusch, H. W. M., Coleman, P. D., Rutten, B. P. F., & van den Hove, D. L. A. (2013). Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiology of Aging*, *34*(9), 2091–2099. https://doi.org/10.1016/j.neurobiolaging.2013.02.021

Christensen, B. C., Houseman, E. A., Marsit, C. J., Zheng, S., Wrensch, M. R., Wiemels, J. L., Nelson, H. H., Karagas, M. R., Padbury, J. F., Bueno, R., Sugarbaker, D. J., Yeh, R. F., Wiencke, J. K., & Kelsey, K. T. (2009). Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CPG island context. *PLoS Genetics*, *5*(8), e1000602. https://doi.org/10.1371/journal.pgen.1000602

Christiansen, L., Lenart, A., Tan, Q., Vaupel, J. W., Aviv, A., McGue, M., & Christensen, K. (2016). DNA methylation age is associated with mortality in a longitudinal Danish twin study. *Aging Cell*, *15*(1), 149–154. https://doi.org/https://doi.org/10.1111/acel.12421

Cross-region reduction in 5-hydroxymethylcytosine in Alzheimer's disease brain, 35 Neurobiology of Aging 1850 (2014). https://doi.org/10.1016/j.neurobiolaging.2014.02.002

Copeland, M. P., Daly, E., Hines, V., Mastromauro, C., Zaitchik, D., Gunther, J., & Albert, M. (2003). Psychiatric Symptomatology and Prodromal Alzheimer's

Disease. *Alzheimer Disease & Associated Disorders*, *17*(1). https://journals.lww.com/alzheimerjournal/Fulltext/2003/01000/Psychiatric_Symptomatology_and_Prodromal.1.aspx

Coppieters, N., Dieriks, B. V., Lill, C., Faull, R. L. M., Curtis, M. A., & Dragunow, M. (2014). Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. *Neurobiology of Aging*, *35*(6), 1334–1344. https://doi.org/10.1016/j.neurobiolaging.2013.11.031

Corder, E. H., Saunders, A. M., Risch, N. J., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Rimmler, J. B., Locke, P. A., Conneally, P. M., Schmader, K. E., Small, G. W., Roses, A. D., Haines, J. L., & Pericak-Vance, M. A. (1994). Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nature Genetics*, *7*(2), 180–184. https://doi.org/10.1038/ng0694-180

Corder, E. H., Saunders, A. M., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Small, G. W., Roses, A. D., Haines, J. L., & Pericak-Vance, M. A. (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*, *261*(5123), 921–923. https://doi.org/10.1126/science.8346443

Cotton, A. M., Price, E. M., Jones, M. J., Balaton, B. P., Kobor, M. S., & Brown, C. J. (2015). Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. *Human Molecular Genetics*, *24*(6), 1528–1539. https://doi.org/10.1093/hmg/ddu564

Crivelli, S. M., Giovagnoni, C., Visseren, L., Scheithauer, A.-L., de Wit, N., den Hoedt, S., Losen, M., Mulder, M. T., Walter, J., de Vries, H. E., Bieberich, E., & Martinez-Martinez, P. (2020). Sphingolipids in Alzheimer's disease, how can we target them? *Advanced Drug Delivery Reviews*. https://doi.org/https://doi.org/10.1016/j.addr.2019.12.003

Curradi, M., Izzo, A., Badaracco, G., & Landsberger, N. (2002). Molecular mechanisms of gene silencing mediated by DNA methylation. *Molecular and Cellular Biology*, *22*(9), 3157–3173. https://doi.org/10.1128/MCB.22.9.3157-3173.2002

da Silva, P. N. O., Furuya, T. K., Braga, I. L., Rasmussen, L. T., Labio, R. W., Bertolucci, P. H., Chen, E. S., Turecki, G., Mechawar, N., Payão, S. L., Mill, J., & Smith, M. C. (2014). Analysis of HSPA8 and HSPA9 mRNA expression

and promoter methylation in the brain and blood of Alzheimer's disease patients. *Journal of Alzheimer's Disease*, *38*(1), 165–170. https://doi.org/10.3233/JAD-130428

da Silva, P. N. O., Furuya, T. K., Sampaio Braga, I., Rasmussen, L. T., De Labio, R. W., Bertolucci, P. H., Chen, E. S., Turecki, G., Mechawar, N., Payão, S. L., Mill, J., & Cardoso Smith, M. (2013). CNP and DPYSL2 mRNA expression and promoter methylation levels in brain of Alzheimer's disease patients. *Journal of Alzheimer's Disease*, *33*(2), 349–355. https://doi.org/10.3233/JAD-2012-121192

Dallaire-Théroux, C., Callahan, B. L., Potvin, O., Saikali, S., & Duchesne, S. (2017). Radiological-Pathological Correlation in Alzheimer's Disease: Systematic Review of Antemortem Magnetic Resonance Imaging Findings. *Journal of Alzheimer's Disease*, *57*, 575–601. https://doi.org/10.3233/JAD-161028

Davis, S., Du, P., Bilke, S., Triche, T., Bootwalla, M. (2015). *methylumi: Handle Illumina methylation data. R package version 2.18.2.*

De Jager, P. L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L. C., Yu, L., Eaton, M. L., Keenan, B. T., Ernst, J., McCabe, C., Tang, A., Raj, T., Replogle, J., Brodeur, W., Gabriel, S., Chai, H. S., Younkin, C., Younkin, S. G., Zou, F., … Bennett, D. A. (2014). Alzheimer's disease: Early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nature Neuroscience*, *17*(9), 1156–1163. https://doi.org/10.1038/nn.3786

de Vet, E. C. J. M., Aguado, B., & Campbell, R. D. (2001). G6b, a Novel Immunoglobulin Superfamily Member Encoded in the Human Major Histocompatibility Complex, Interacts with SHP-1 and SHP-2*. *Journal of Biological Chemistry*, *276*(45), 42070–42076. https://doi.org/https://doi.org/10.1074/jbc.M103214200

Delatte, B., & Fuks, F. (2013). TET proteins: On the frenetic hunt for new cytosine modifications. *Briefings in Functional Genomics*, *12*(3), 191–204. https://doi.org/10.1093/bfgp/elt010

Doody, R. S., Raman, R., Farlow, M., Iwatsubo, T., Vellas, B., Joffe, S., Kieburtz, K., He, F., Sun, X., Thomas, R. G., Aisen, P. S., Siemers, E., Sethuraman, G., & Mohs, R. (2013). A phase 3 trial of semagacestat for treatment of Alzheimer's disease. *New England Journal of Medicine*, *369*(4), 341–350. https://doi.org/10.1056/NEJMoa1210951

Drewes, G., Trinczek, B., Illenberger, S., Biernat, J., Schmitt-Ulms, G., Meyer, H. E., Mandelkow, E. M., & Mandelkow, E. (1995). Microtubule-associated protein/microtubule affinity-regulating kinase (p110(mark)). A novel protein kinase that regulates tau-microtubule interactions and dynamic instability by phosphorylation at the Alzheimer- specific site serine 262. *Journal of Biological Chemistry*, *270*(13), 7679–7688. https://doi.org/10.1074/jbc.270.13.7679

Du, P., Zhang, X., Huang, C. C., Jafari, N., Kibbe, W. A., Hou, L., & Lin, S. M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*, *11*. https://doi.org/10.1186/1471-2105-11-587

Dubois, B., Feldman, H. H., Jacova, C., Cummings, J. L., DeKosky, S. T., Barberger-Gateau, P., Delacourte, A., Frisoni, G., Fox, N. C., Galasko, D., Gauthier, S., Hampel, H., Jicha, G. A., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., … Scheltens, P. (2010). Revising the definition of Alzheimer's disease: A new lexicon. *The Lancet Neurology*, *9*(11), 1118–1127. https://doi.org/10.1016/S1474-4422(10)70223-4

Dubois, B., Feldman, H. H., Jacova, C., Hampel, H., Molinuevo, J. L., Blennow, K., DeKosky, S. T., Gauthier, S., Selkoe, D., Bateman, R., Cappa, S., Crutch, S., Engelborghs, S., Frisoni, G. B., Fox, N. C., Galasko, D., Habert, M.-O., Jicha, G. A., Nordberg, A., … Cummings, J. L. (2014). Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *The Lancet Neurology*, *13*(6), 614–629. https://doi.org/https://doi.org/10.1016/S1474-4422(14)70090-0

El Hajj, N., Dittrich, M., Böck, J., Kraus, T. F. J., Nanda, I., Müller, T., Seidmann, L., Tralau, T., Galetzka, D., Schneider, E., & Haaf, T. (2016). Epigenetic dysregulation in the developing Down syndrome cortex. *Epigenetics*, *11*(8), 563–578. https://doi.org/10.1080/15592294.2016.1192736

Elliott, H. R., Tillin, T., McArdle, W. L., Ho, K., Duggirala, A., Frayling, T. M., Smith, G. D., Hughes, A. D., Chaturvedi, N., & Relton, C. L. (2014). Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clinical Epigenetics*, *6*(1), 4. https://doi.org/10.1186/1868-7083-6-4

Ellison, E. M., Abner, E. L., & Lovell, M. A. (2017). Multiregional analysis of global 5-methylcytosine and 5-hydroxymethylcytosine throughout the progression

of Alzheimer's disease. *Journal of Neurochemistry*, *140*(3), 383–394. https://doi.org/10.1111/jnc.13912

Fagan, A. M., Roe, C. M., Xiong, C., Mintun, M. A., Morris, J. C., & Holtzman, D. M. (2007). Cerebrospinal Fluid tau/β-Amyloid42 Ratio as a Prediction of Cognitive Decline in Nondemented Older Adults. *Archives of Neurology*, *64*(3), 343–349. https://doi.org/10.1001/archneur.64.3.noc60123

Fehlbaum-Beurdeley, P., Sol, O., Désiré, L., Touchon, J., Dantoine, T., Vercelletto, M., Gabelle, A., Jarrige, A. C., Haddad, R., Lemarié, J. C., Zhou, W., Hampel, H., Einstein, R., & Vellas, B. (2012). Validation of AclarusDx™, a blood-based transcriptomic signature for the diagnosis of Alzheimer's disease. *Journal of Alzheimer's Disease*, *32*(1), 169–181. https://doi.org/10.3233/JAD-2012-120637

Feldman, R., Monakhov, M., Pratt, M., & Ebstein, R. P. (2016). Oxytocin Pathway Genes: Evolutionary Ancient System Impacting on Human Affiliation, Sociality, and Psychopathology. *Biological Psychiatry*, *79*(3), 174–184. https://doi.org/https://doi.org/10.1016/j.biopsych.2015.08.008

Ferrari, G., Avila, M., Medina, M., Perez-Palma, E., Bustos, B., & Alarcon, M. (2014). Wnt/β-Catenin Signaling in Alzheimer's Disease. In *CNS & Neurological Disorders - Drug Targets* (Vol. 13, Issue 5, pp. 745–754). https://doi.org/10.2174/1871527312666131223113900

Fitzgerald, P. H., & McEwan, C. M. (1977). Total aneuploidy and age-related sex chromosome aneuploidy in cultured lymphocytes of normal men and women. *Human Genetics*, *39*(3), 329–337.

Fraga, M. F., Ballestar, E., Paz, M. F., Ropero, S., Setien, F., Ballestar, M. L., Heine-Suñer, D., Cigudosa, J. C., Urioste, M., Benitez, J., Boix-Chornet, M., Sanchez-Aguilera, A., Ling, C., Carlsson, E., Poulsen, P., Vaag, A., Stephan, Z., Spector, T. D., Wu, Y.-Z., … Esteller, M. (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), 10604 LP – 10609. https://doi.org/10.1073/pnas.0500398102

Fratiglioni, L., Viitanen, M., von Strauss, E., Tontodonati, V., Herlitz, A., & Winblad, B. (1997). Very Old Women at Highest Risk of Dementia and Alzheimer's Disease. *Neurology*, *48*(1), 132 LP – 138. https://doi.org/10.1212/WNL.48.1.132

Frisoni, G. B., Pievani, M., Testa, C., Sabattoli, F., Bresciani, L., Bonetti, M.,

Beltramello, A., Hayashi, K. M., Toga, A. W., & Thompson, P. M. (2007). The topography of grey matter involvement in early and late onset Alzheimer's disease. *Brain*, *130*(3), 720–730. https://doi.org/10.1093/brain/awl377

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L., & Paul, C. L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences*, *89*(5), 1827 LP – 1831. https://doi.org/10.1073/pnas.89.5.1827

Fukuzawa, S., Takahashi, S., Tachibana, K., Tajima, S., & Suetake, I. (2016). Simple and accurate single base resolution analysis of 5-hydroxymethylcytosine by catalytic oxidative bisulfite sequencing using micelle incarcerated oxidants. *Bioorganic and Medicinal Chemistry*, *24*(18), 4254–4262. https://doi.org/10.1016/j.bmc.2016.07.016

Furney, S. J., Kronenberg, D., Simmons, A., Güntert, A., Dobson, R. J., Proitsi, P., Wahlund, L. O., Kloszewska, I., Mecocci, P., Soininen, H., Tsolaki, M., Vellas, B., Spenger, C., & Lovestone, S. (2011). Combinatorial markers of Mild Cognitive Impairment conversion to Alzheimers Disease - Cytokines and MRI measures together predict disease progression. *Journal of Alzheimer's Disease*, *26*(SUPPL. 3), 395–405. https://doi.org/10.3233/JAD-2011-0044

Furney, S. J., Simmons, A., Breen, G., Pedroso, I., Lunnon, K., Proitsi, P., Hodges, A., Powell, J., Wahlund, L. O., Kloszewska, I., Mecocci, P., Soininen, H., Tsolaki, M., Vellas, B., Spenger, C., Lathrop, M., Shen, L., Kim, S., Saykin, A. J., … Lovestone, S. (2011). Genome-wide association with MRI atrophy measures as a quantitative trait locus for Alzheimer's disease. *Molecular Psychiatry*, *16*(11), 1130–1138. https://doi.org/10.1038/mp.2010.123

Furuya, T. K., da Silva, P. N. O., Payão, S. L. M., Bertolucci, P. H. F., Rasmussen, L. T., De Labio, R. W., Braga, I. L. S., Chen, E. S., Turecki, G., Mechawar, N., Mill, J., & Smith, M. A. C. (2012). Analysis of SNAP25 mRNA expression and promoter DNA methylation in brain areas of Alzheimer's Disease patients. *Neuroscience*, *220*, 41–46. https://doi.org/10.1016/j.neuroscience.2012.06.035

Furuya, T. K., da Silva, P. N. O., Payão, S. L. M., Rasmussen, L. T., De Labio, R. W., Bertolucci, P. H. F., Braga, I. L. S., Chen, E. S., Turecki, G.,

Mechawar, N., Mill, J., & De Arruda Cardoso Smith, M. (2012). SORL1 and SIRT1 mRNA expression and promoter methylation levels in aging and Alzheimer's Disease. In *Neurochemistry International* (Vol. 61, Issue 7). https://doi.org/10.1016/j.neuint.2012.07.014

Gao, S., Hendrie, H. C., Hall, K. S., & Hui, S. (1998). The Relationships Between Age, Sex, and the Incidence of Dementia and Alzheimer Disease: A Meta-analysis. *Archives of General Psychiatry*, *55*(9), 809–815. https://doi.org/10.1001/archpsyc.55.9.809

Gasparoni, G., Bultmann, S., Lutsik, P., Kraus, T. F. J., Sordon, S., Vlcek, J., Dietinger, V., Steinmaurer, M., Haider, M., Mulholland, C. B., Arzberger, T., Roeber, S., Riemenschneider, M., Kretzschmar, H. A., Giese, A., Leonhardt, H., & Walter, J. (2018). DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. *Epigenetics and Chromatin*, *11*(1), 41. https://doi.org/10.1186/s13072-018-0211-3

Ghosal, K., Vogt, D. L., Liang, M., Shen, Y., Lamb, B. T., & Pimplikar, S. W. (2009). Alzheimer's disease-like pathological features in transgenic mice expressing the APP intracellular domain. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(43), 18367–18372. https://doi.org/10.1073/pnas.0907652106

Ghoshal, N., García-Sierra, F., Wuu, J., Leurgans, S., Bennett, D. A., Berry, R. W., & Binder, L. I. (2002). Tau conformational changes correspond to impairments of episodic memory in mild cognitive impairment and Alzheimer's disease. *Experimental Neurology*, *177*(2), 475–493. https://doi.org/10.1006/exnr.2002.8014

Giampaolo, A., Felli, N., Diverio, D., Morsilli, O., Samoggia, P., Breccia, M., Lo Coco, F., Peschle, C., & Testa, U. (2002). Expression pattern of HOXB6 homeobox gene in myelomonocytic differentiation and acute myeloid leukemia. *Leukemia*, *16*(7), 1293–1301. https://doi.org/10.1038/sj.leu.2402532

Giovagnoni, C., Ali, M., Eijssen, L. M. T., Maes, R., Choe, K., Mulder, M., Kleinjans, J., del Sol, A., Glaab, E., Mastroeni, D., Delvaux, E., Coleman, P., Losen, M., Pishva, E., Martinez-Martinez, P., & van den Hove, D. L. A. (2021). Altered sphingolipid function in Alzheimer's disease; a gene regulatory network approach. *Neurobiology of Aging*, *102*, 178–187.

https://doi.org/https://doi.org/10.1016/j.neurobiolaging.2021.02.001

Girard, A., Sachidanandam, R., Hannon, G. J., & Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, *442*(7099), 199–202. https://doi.org/10.1038/nature04917

Goedert, M., & Spillantini, M. G. (2006). A Century of Alzheimer's Disease. *Science*, *314*(5800), 777 LP – 781. https://doi.org/10.1126/science.1132814

Gonzalez, T. L., Sun, T., Koeppel, A. F., Lee, B., Wang, E. T., Farber, C. R., Rich, S. S., Sundheimer, L. W., Buttle, R. A., Chen, Y.-D. I., Rotter, J. I., Turner, S. D., Williams, J., Goodarzi, M. O., & Pisarska, M. D. (2018). Sex differences in the late first trimester human placenta transcriptome. *Biology of Sex Differences*, *9*(1), 4. https://doi.org/10.1186/s13293-018-0165-y

Goodger, Z. V., Rajendran, L., Trutzel, A., Kohli, B. M., Nitsch, R. M., & Konietzko, U. (2009). Nuclear signaling by the APP intracellular domain occurs predominantly through the amyloidogenic processing pathway. *Journal of Cell Science*, *122*(20), 3703–3714. https://doi.org/10.1242/jcs.048090

Gosselt, H. R., Griffioen, P. H., van Zelst, B. D., Oosterom, N., de Jonge, R., & Heil, S. G. (2021). Global DNA (hydroxy)methylation is stable over time under several storage conditions and temperatures. *Epigenetics*, *16*(1), 45–53. https://doi.org/10.1080/15592294.2020.1786318

Götz, J., Chen, F., Van Dorpe, J., & Nitsch, R. M. (2001). Formation of neurofibrillary tangles in P301L tau transgenic mice induced by Aβ42 fibrils. *Science*, *293*(5534), 1491–1495. https://doi.org/10.1126/science.1062097

Grudzien, A., Shaw, P., Weintraub, S., Bigio, E., Mash, D. C., & Mesulam, M. M. (2007). Locus coeruleus neurofibrillary degeneration in aging, mild cognitive impairment and early Alzheimer's disease. *Neurobiology of Aging*, *28*(3), 327–335. https://doi.org/10.1016/j.neurobiolaging.2006.02.007

Guerreiro, R., Wojtas, A., Bras, J., Carrasquillo, M., Rogaeva, E., Majounie, E., Cruchaga, C., Sassi, C., Kauwe, J. S. K., Younkin, S., Hazrati, L., Collinge, J., Pocock, J., Lashley, T., Williams, J., Lambert, J. C., Amouyel, P., Goate, A., Rademakers, R., … Hardy, J. (2013). TREM2 variants in Alzheimer's disease. *New England Journal of Medicine*, *368*(2), 117–127. https://doi.org/10.1056/NEJMoa1211851

Guo, J. U., Su, Y., Zhong, C., Ming, G. L., & Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*, *145*(3), 423–434. https://doi.org/10.1016/j.cell.2011.03.022

Guo, L., Zhong, M. B., Zhang, L., Zhang, B., & Cai, D. (2021). Sex Differences in Alzheimer's Disease: Insights From the Multiomics Landscape. *Biological Psychiatry*. https://doi.org/https://doi.org/10.1016/j.biopsych.2021.02.968

Hall, E., Volkov, P., Dayeh, T., Esguerra, J. L. S., Salö, S., Eliasson, L., Rönn, T., Bacos, K., & Ling, C. (2014). Sex differences in the genome-wide DNA methylation pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets. *Genome Biology*, *15*(12), 522. https://doi.org/10.1186/s13059-014-0522-z

Hampel, H., Blennow, K., Shaw, L. M., Hoessler, Y. C., Zetterberg, H., & Trojanowski, J. Q. (2010). Total and phosphorylated tau protein as biological markers of Alzheimer's disease. *Experimental Gerontology*, *45*(1), 30–40. https://doi.org/https://doi.org/10.1016/j.exger.2009.10.010

Hannon, E., Knox, O., Sugden, K., Burrage, J., Wong, C. C. Y., Belsky, D. W., Corcoran, D. L., Arseneault, L., Moffitt, T. E., Caspi, A., & Mill, J. (2018). Characterizing genetic and environmental influences on variable DNA methylation using monozygotic and dizygotic twins. *PLOS Genetics*, *14*(8), e1007544. https://doi.org/10.1371/journal.pgen.1007544

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S. V., Klotzle, B., Bibikova, M., Fan, J. B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T., & Zhang, K. (2013). Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Molecular Cell*, *49*(2), 359–367. https://doi.org/10.1016/j.molcel.2012.10.016

Harder, M. N., Ribel-Madsen, R., Justesen, J. M., Sparsø, T., Andersson, E. A., Grarup, N., Jørgensen, T., Linneberg, A., Hansen, T., & Pedersen, O. (2013). Type 2 diabetes risk alleles near BCAR1 and in ANK1 associate with decreased β-cell function whereas risk alleles near ANKRD55 and GRB14 associate with decreased insulin sensitivity in the danish Inter99 cohort. *Journal of Clinical Endocrinology and Metabolism*, *98*(4), E801–E806. https://doi.org/10.1210/jc.2012-4169

Hardy, J. (1997). Amyloid, the presenilins and Alzheimer's disease. *Trends in Neurosciences*, *20*(4), 154–159. https://doi.org/10.1016/S0166-2236(96)01030-2

Hardy, J. A., & Higgins, G. A. (1992). Alzheimer's disease: The amyloid cascade hypothesis. *Science*, *256*(5054), 184–185. https://doi.org/10.1126/science.1566067

Hardy, J., & Allsop, D. (1991). Amyloid deposition as the central event in the aetiology of Alzheimer's disease. *Trends in Pharmacological Sciences*, *12*(C), 383–388. https://doi.org/10.1016/0165-6147(91)90609-V

Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M. L., Pahwa, J. S., Moskvina, V., Dowzell, K., Williams, A., Jones, N., Thomas, C., Stretton, A., Morgan, A. R., Lovestone, S., Powell, J., Proitsi, P., Lupton, M. K., Brayne, C., … Williams, J. (2009). Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nature Genetics*, *41*(10), 1088–1093. https://doi.org/10.1038/ng.440

He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, N., Sun, Y., Li, X., Dai, Q., Song, C. X., Zhang, K., He, C., & Xu, G. L. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, *333*(6047), 1303–1307. https://doi.org/10.1126/science.1210944

Hebert, L. E., Scherr, P. A., McCann, J. J., Beckett, L. A., & Evans, D. A. (2001). Is the Risk of Developing Alzheimer's Disease Greater for Women than for Men? *American Journal of Epidemiology*, *153*(2), 132–136. https://doi.org/10.1093/aje/153.2.132

Hebert, L. E., Weuve, J., Scherr, P. A., & Evans, D. A. (2013). Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. *Neurology*, *80*(19), 1778 LP – 1783. https://doi.org/10.1212/WNL.0b013e31828726f5

Heppner, F. L., Ransohoff, R. M., & Becher, B. (2015). Immune attack: The role of inflammation in Alzheimer disease. *Nature Reviews Neuroscience*, *16*(6), 358–372. https://doi.org/10.1038/nrn3880

Hernandez, D. G., Nalls, M. A., Gibbs, J. R., Arepalli, S., van der brug, M., Chong, S., Moore, M., Longo, D. L., Cookson, M. R., Traynor, B. J., & Singleton, A. B. (2011). Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Human Molecular Genetics*, *20*(6), 1164–1172. https://doi.org/10.1093/hmg/ddq561

Herrera, L. A., Prada, D., Andonegui, M. A., & Dueñas-González, A. (2008). The epigenetic origin of aneuploidy. *Current Genomics*, *9*(1), 43–50. https://doi.org/10.2174/138920208783884883

Heyn, H., Li, N., Ferreira, H. J., Moran, S., Pisano, D. G., Gomez, A., Diez, J., Sanchez-Mut, J. V, Setien, F., Carmona, F. J., Puca, A. A., Sayols, S.,

Pujana, M. A., Serra-Musach, J., Iglesias-Platas, I., Formiga, F., Fernandez, A. F., Fraga, M. F., Heath, S. C., … Esteller, M. (2012). Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci U S A*, *109*. https://doi.org/10.1073/pnas.1120658109

Hickman, S. E., Allison, E. K., & El Khoury, J. (2008). Microglial dysfunction and defective β-amyloid clearance pathways in aging alzheimer's disease mice. *Journal of Neuroscience*, *28*(33), 8354–8360. https://doi.org/10.1523/JNEUROSCI.0616-08.2008

Hill, J. M., Clement, C., Pogue, A. I., Bhattacharjee, S., Zhao, Y., & Lukiw, W. J. (2014). Pathogenic microbes, the microbiome, and Alzheimer's disease (AD) . In *Frontiers in Aging Neuroscience* (Vol. 6, p. 127). https://www.frontiersin.org/article/10.3389/fnagi.2014.00127

Hollingworth, P., Harold, D., Sims, R., Gerrish, A., Lambert, J. C., Carrasquillo, M. M., Abraham, R., Hamshere, M. L., Pahwa, J. S., Moskvina, V., Dowzell, K., Jones, N., Stretton, A., Thomas, C., Richards, A., Ivanov, D., Widdowson, C., Chapman, J., Lovestone, S., … Williams, J. (2011). Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nature Genetics*, *43*(5), 429–436. https://doi.org/10.1038/ng.803

Hondius, D. C., van Nierop, P., Li, K. W., Hoozemans, J. J. M., van der Schors, R. C., van Haastert, E. S., van der Vies, S. M., Rozemuller, A. J. M., & Smit, A. B. (2016). Profiling the human hippocampal proteome at all pathologic stages of Alzheimer's disease. *Alzheimer's & Dementia*, *12*(6), 654–668. https://doi.org/https://doi.org/10.1016/j.jalz.2015.11.002

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*(10), 3156. https://doi.org/10.1186/gb-2013-14-10-r115

Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., Wiencke, J. K., & Kelsey, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, *13*(1), 86. https://doi.org/10.1186/1471-2105-13-86

Huang, L.-H., Lin, P.-H., Tsai, K.-W., Wang, L.-J., Huang, Y.-H., Kuo, H.-C., & Li, S.-C. (2017). The effects of storage temperature and duration of blood samples on DNA and RNA qualities. *PLOS ONE*, *12*(9), e0184692. https://doi.org/10.1371/journal.pone.0184692

Hye, A., Lynham, S., Thambisetty, M., Causevic, M., Campbell, J., Byers, H. L.,

Hooper, C., Rijsdijk, F., Tabrizi, S. J., Banner, S., Shaw, C. E., Foy, C., Poppe, M., Archer, N., Hamilton, G., Powell, J., Brown, R. G., Sham, P., Ward, M., & Lovestone, S. (2006). Proteome-based plasma biomarkers for Alzheimer's disease. *Brain*, *129*(11), 3042–3050. https://doi.org/10.1093/brain/awl279

Hyman, B. T., Phelps, C. H., Beach, T. G., Bigio, E. H., Cairns, N. J., Carrillo, M. C., Dickson, D. W., Duyckaerts, C., Frosch, M. P., Masliah, E., Mirra, S. S., Nelson, P. T., Schneider, J. A., Thal, D. R., Thies, B., Trojanowski, J. Q., Vinters, H. V, & Montine, T. J. (2012). National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimer's & Dementia : The Journal of the Alzheimer's Association*, *8*(1), 1–13. https://doi.org/10.1016/j.jalz.2011.10.007

Hyman, B. T., Van Hoesen, G. W., Damasio, A. R., & Barnes, C. L. (1984). Alzheimer's disease: Cell-specific pathology isolates the hippocampal formation. *Science*, *225*(4667), 1168–1170. https://doi.org/10.1126/science.6474172

Iatrou, A., Kenis, G., Rutten, B. P. F., Lunnon, K., & van den Hove, D. L. A. (2017). Epigenetic dysregulation of brainstem nuclei in the pathogenesis of Alzheimer's disease: looking in the correct place at the right time? In *Cellular and Molecular Life Sciences* (Vol. 74, Issue 3, pp. 509–523). Birkhauser Verlag AG. https://doi.org/10.1007/s00018-016-2361-4

Iliev, R., Fedorko, M., MacHackova, T., Mlcochova, H., Svoboda, M., Pacik, D., Dolezel, J., Stanik, M., & Slaby, O. (2016). Expression levels of PIWI-interacting RNA, piR-823, are deregulated in tumor tissue, blood serum and urine of patients with renal cell carcinoma. *Anticancer Research*, *36*(12), 6419–6423. https://doi.org/10.21873/anticanres.11239

Illumina. (n.d.). *Infinium® HumanMethylation450 BeadChip*. Retrieved May 27, 2020, from https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_humanmethylation450.pdf

Imamura, M., Maeda, S., Yamauchi, T., Hara, K., Yasuda, K., Morizono, T., Takahashi, A., Horikoshi, M., Nakamura, M., Fujita, H., Tsunoda, T., Kubo, M., Watada, H., Maegawa, H., Okada-Iwabu, M., Iwabu, M., Shojima, N., Ohshige, T., Omori, S., … Kadowaki, T. (2012). A single-nucleotide polymorphism in ANK1 is associated with susceptibility to type 2 diabetes in

Japanese populations. *Human Molecular Genetics*, *21*(13), 3042–3049. https://doi.org/10.1093/hmg/dds113

Iqbal, K., Liu, F., Gong, C.-X., & Grundke-Iqbal, I. (2010). Tau in Alzheimer Disease and Related Tauopathies. *Current Alzheimer Research*, *7*(8), 656–664. https://doi.org/10.2174/156720510793611592

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., & Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, *333*(6047), 1300–1303. https://doi.org/10.1126/science.1210597

Iwata, A., Nagata, K., Hatsuta, H., Takuma, H., Bundo, M., Iwamoto, K., Tamaoka, A., Murayama, S., Saido, T., & Tsuji, S. (2014). Altered CpG methylation in sporadic alzheimer's disease is associated with APP and MAPT dysregulation. *Human Molecular Genetics*, *23*(3), 648–656. https://doi.org/10.1093/hmg/ddt451

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J. L., Montine, T., Phelps, C., Rankin, K. P., Rowe, C. C., Scheltens, P., Siemers, E., Snyder, H. M., … Silverberg, N. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's and Dementia*, *14*(4), 535–562. https://doi.org/10.1016/j.jalz.2018.02.018

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Feldman, H. H., Frisoni, G. B., Hampel, H., Jagust, W. J., Johnson, K. A., Knopman, D. S., Petersen, R. C., Scheltens, P., Sperling, R. A., & Dubois, B. (2016). A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology*, *87*(5), 539–547. https://doi.org/10.1212/WNL.0000000000002923

Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., Petersen, R. C., & Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, *9*(1), 119–128. https://doi.org/10.1016/S1474-4422(09)70299-6

Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., Sealock, J., Karlsson, I. K., Hägg, S., Athanasiu, L., Voyle, N., Proitsi, P., Witoelar, A., Stringer, S., Aarsland, D., Almdahl, I. S., Andersen, F., Bergh, S., Bettella, F., … Posthuma, D. (2019). Genome-wide meta-analysis

identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature Genetics*, *51*(3), 404–413. https://doi.org/10.1038/s41588-018-0311-9

Jicha, G. A., Parisi, J. E., Dickson, D. W., Johnson, K., Cha, R., Ivnik, R. J., Tangalos, E. G., Boeve, B. F., Knopman, D. S., Braak, H., & Petersen, R. C. (2006). Neuropathologic outcome of mild cognitive impairment following progression to clinical dementia. *Archives of Neurology*, *63*(5), 674–681. https://doi.org/10.1001/archneur.63.5.674

Jolles, J., Boxtel, M. P. J. van, Ponds, R. W. H. M., Metsemakers, J. F. M., & Houx, P. J. (1998). Leeftijdsgeassocieerde cognitieve functiestoornissen: Cognitive aging in a longitudinal perspective: the Maastricht Aging Study (MAAS). *Acta Neuropsychiatrica*, *10*(4), 81–83. https://doi.org/DOI: 10.1017/S0924270800036413

Jones, M. J., Goodman, S. J., & Kobor, M. S. (2015). DNA methylation and healthy human aging. *Aging Cell*, *14*(6), 924–932. https://doi.org/https://doi.org/10.1111/acel.12349

Joshi, A., Ringman, J. M., Lee, A. S., Juarez, K. O., & Mendez, M. F. (2012). Comparison of clinical characteristics between familial and non-familial early onset Alzheimer's disease. In *Journal of Neurology* (Vol. 259, Issue 10, pp. 2182–2188). https://doi.org/10.1007/s00415-012-6481-y

Jun, G., Naj, A. C., Beecham, G. W., Wang, L. S., Buros, J., Gallins, P. J., Buxbaum, J. D., Ertekin-Taner, N., Fallin, M. D., Friedland, R., Inzelberg, R., Kramer, P., Rogaeva, E., St George-Hyslop, P., Cantwell, L. B., Dombroski, B. A., Saykin, A. J., Reiman, E. M., Bennett, D. A., … Schellenberg, G. D. (2010). Meta-analysis confirms CR1, CLU, and PICALM as Alzheimer disease risk loci and reveals interactions with APOE genotypes. *Archives of Neurology*, *67*(12), 1473–1484. https://doi.org/10.1001/archneurol.2010.201

Kaikkonen, M. U., Lam, M. T. Y., & Glass, C. K. (2011). Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovascular Research*, *90*(3), 430–440. https://doi.org/10.1093/cvr/cvr097

Kanungo, J. (2013). DNA-dependent protein kinase and DNA repair: relevance to Alzheimer's disease. *Alzheimer's Research & Therapy*, *5*(2), 13. https://doi.org/10.1186/alzrt167

Kato, J., & Kitamura, K. (2015). Bench-to-bedside pharmacology of adrenomedullin. *European Journal of Pharmacology*, *764*, 140–148.

https://doi.org/https://doi.org/10.1016/j.ejphar.2015.06.061

Kato, J., Kitamura, K., Uemura, T., Kuwasako, K., Kita, T., Kangawa, K., & Eto, T. (2002). Plasma Levels of Adrenomedullin and Atrial and Brain Natriuretic Peptides in the General Population: Their Relations to Age and Pulse Pressure. *Hypertension Research*, *25*(6), 887–892. https://doi.org/10.1291/hypres.25.887

Kawas, C., Gray, S., Brookmeyer, R., Fozard, J., & Zonderman, A. (2000). Age-specific incidence rates of Alzheimer's disease. *Neurology*, *54*(11), 2072 LP – 2077. https://doi.org/10.1212/WNL.54.11.2072

Kerkel, K., Schupf, N., Hatta, K., Pang, D., Salas, M., Kratz, A., Minden, M., Murty, V., Zigman, W. B., Mayeux, R. P., Jenkins, E. C., Torkamani, A., Schork, N. J., Silverman, W., Croy, B. A., & Tycko, B. (2010). Altered DNA methylation in leukocytes with trisomy 21. *PLoS Genetics*, *6*(11), e1001212. https://doi.org/10.1371/journal.pgen.1001212

Kim, J., Chakrabarty, P., Hanna, A., March, A., Dickson, D. W., Borchelt, D. R., Golde, T., & Janus, C. (2013). Normal cognition in transgenic BRI2-Aβ mice. *Molecular Neurodegeneration*, *8*(1), 15. https://doi.org/10.1186/1750-1326-8-15

Kitamura, K., Kangawa, K., Kawamoto, M., Ichiki, Y., Nakamura, S., Matsuo, H., & Eto, T. (1993). Adrenomedullin: A Novel Hypotensive Peptide Isolated from Human Pheochromocytoma. *Biochemical and Biophysical Research Communications*, *192*(2), 553–560. https://doi.org/https://doi.org/10.1006/bbrc.1993.1451

Klunk, W. E., Engler, H., Nordberg, A., Wang, Y., Blomqvist, G., Holt, D. P., Bergström, M., Savitcheva, I., Huang, G. F., Estrada, S., Ausén, B., Debnath, M. L., Barletta, J., Price, J. C., Sandell, J., Lopresti, B. J., Wall, A., Koivisto, P., Antoni, G., … Långström, B. (2004). Imaging Brain Amyloid in Alzheimer's Disease with Pittsburgh Compound-B. *Annals of Neurology*, *55*(3), 306–319. https://doi.org/10.1002/ana.20009

Kneip, C., Schmidt, B., Seegebarth, A., Weickmann, S., Fleischhacker, M., Liebenberg, V., Field, J. K., & Dietrich, D. (2011). SHOX2 DNA Methylation Is a Biomarker for the Diagnosis of Lung Cancer in Plasma. *Journal of Thoracic Oncology*, *6*(10), 1632–1638. https://doi.org/https://doi.org/10.1097/JTO.0b013e318220ef9a

Kobayashi, N., Shinagawa, S., Nagata, T., Shimada, K., Shibata, N., Ohnuma,

T., Kasanuki, K., Arai, H., Yamada, H., Nakayama, K., & Kondo, K. (2016). Development of biomarkers based on DNA methylation in the NCAPH2/LMF2 promoter region for diagnosis of Alzheimer's disease and amnesic mild cognitive impairment. *PLoS ONE*, *11*(1), e0146449. https://doi.org/10.1371/journal.pone.0146449

Koedam, E. L. G. E., Lauffer, V., Van Der Vlies, A. E., Van Der Flier, W. M., Scheltens, P., & Pijnenburg, Y. A. L. (2010). Early-versus late-onset Alzheimer's disease: More than age alone. *Journal of Alzheimer's Disease*, *19*(4), 1401–1408. https://doi.org/10.3233/JAD-2010-1337

Konki, M., Malonzo, M., Karlsson, I. K., Lindgren, N., Ghimire, B., Smolander, J., Scheinin, N. M., Ollikainen, M., Laiho, A., Elo, L. L., Lönnberg, T., Röyttä, M., Pedersen, N. L., Kaprio, J., Lähdesmäki, H., Rinne, J. O., & Lund, R. J. (2019). Peripheral blood DNA methylation differences in twin pairs discordant for Alzheimer's disease. *Clinical Epigenetics*, *11*(1), 130. https://doi.org/10.1186/s13148-019-0729-7

Krabbe, G., Halle, A., Matyash, V., Rinnenthal, J. L., Eom, G. D., Bernhardt, U., Miller, K. R., Prokop, S., Kettenmann, H., & Heppner, F. L. (2013). Functional Impairment of Microglia Coincides with Beta-Amyloid Deposition in Mice with Alzheimer-Like Pathology. *PLoS ONE*, *8*(4), e60921. https://doi.org/10.1371/journal.pone.0060921

Krumlauf, R. (1994). Hox genes in vertebrate development. *Cell*, *78*(2), 191–201. https://doi.org/10.1016/0092-8674(94)90290-9

Ksiezak-Reding, H., Pyo, H. K., Feinstein, B., & Pasinetti, G. M. (2003). Akt/PKB kinase phosphorylates separately Thr212 and Ser214 of tau protein in vitro. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, *1639*(3), 159–168. https://doi.org/10.1016/j.bbadis.2003.09.001

Kummer, M. P., & Heneka, M. T. (2014). Truncated and modified amyloid-beta species. *Alzheimer's Research & Therapy*, *6*(3), 28. https://doi.org/10.1186/alzrt258

Kunkle, B. W., Grenier-Boley, B., Sims, R., Bis, J. C., Damotte, V., Naj, A. C., Boland, A., Vronskaya, M., van der Lee, S. J., Amlie-Wolf, A., Bellenguez, C., Frizatti, A., Chouraki, V., Martin, E. R., Sleegers, K., Badarinarayan, N., Jakobsdottir, J., Hamilton-Nelson, K. L., Moreno-Grau, S., … Lieberman, A. P. (2019). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nature*

Genetics, *51*(3), 414–430. https://doi.org/10.1038/s41588-019-0358-2

Lambert, J. C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., Jun, G., DeStefano, A. L., Bis, J. C., Beecham, G. W., Grenier-Boley, B., Russo, G., Thornton-Wells, T. A., Jones, N., Smith, A. V., Chouraki, V., Thomas, C., Ikram, M. A., Zelenika, D., … Seshadri, S. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics*, *45*(12), 1452–1458. https://doi.org/10.1038/ng.2802

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 559. https://doi.org/10.1186/1471-2105-9-559

Lardenoije, R., Roubroeks, J. A. Y., Pishva, E., Leber, M., Wagner, H., Iatrou, A., Smith, A. R., Smith, R. G., Eijssen, L. M. T., Kleineidam, L., Kawalia, A., Hoffmann, P., Luck, T., Riedel-Heller, S., Jessen, F., Maier, W., Wagner, M., Hurlemann, R., Kenis, G., … Van Den Hove, D. L. A. (2019). Alzheimer's disease-associated (hydroxy)methylomic changes in the brain and blood. *Clinical Epigenetics*, *11*(1), 164. https://doi.org/10.1186/s13148-019-0755-5

Larrayoz, I. M., Ferrero, H., Martisova, E., Gil-Bea, F. J., Ramírez, M. J., & Martínez, A. (2017). Adrenomedullin Contributes to Age-Related Memory Loss in Mice and Is Elevated in Aging Human Brains . In *Frontiers in Molecular Neuroscience* (Vol. 10, p. 384). https://www.frontiersin.org/article/10.3389/fnmol.2017.00384

Larsson, S. C., & Markus, H. S. (2018). Does Treating Vascular Risk Factors Prevent Dementia and Alzheimer's Disease? A Systematic Review and Meta-Analysis. *Journal of Alzheimer's Disease*, *64*(2), 657–668. https://doi.org/10.3233/JAD-180288

Lashley, T., Gami, P., Valizadeh, N., Li, A., Revesz, T., & Balazs, R. (2015). Alterations in global DNA methylation and hydroxymethylation are not detected in Alzheimer's disease. *Neuropathology and Applied Neurobiology*, *41*(4), 497–506. https://doi.org/10.1111/nan.12183

Lee, C. Y. D., & Landreth, G. E. (2010). The role of microglia in amyloid clearance from the AD brain. *Journal of Neural Transmission*, *117*(8), 949–960. https://doi.org/10.1007/s00702-010-0433-4

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., & Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in

high-throughput experiments. *Bioinformatics (Oxford, England)*, *28*(6), 882–883. https://doi.org/10.1093/bioinformatics/bts034

Letenneur, L., Gilleron, V., Commenges, D., Helmer, C., Orgogozo, J. M., & Dartigues, J. F. (1999). Are sex and educational level independent predictors of dementia and Alzheimer's disease? Incidence data from the PAQUID project. *Journal of Neurology, Neurosurgery & Psychiatry*, *66*(2), 177 LP – 183. https://doi.org/10.1136/jnnp.66.2.177

Li, B., Chohan, M. O., Grundke-Iqbal, I., & Iqbal, K. (2007). Disruption of microtubule network by Alzheimer abnormally hyperphosphorylated tau. *Acta Neuropathologica*, *113*(5), 501–511. https://doi.org/10.1007/s00401-007-0207-8

Li, L., & Hölscher, C. (2007). Common pathological processes in Alzheimer disease and type 2 diabetes: A review. *Brain Research Reviews*, *56*(2), 384–402. https://doi.org/https://doi.org/10.1016/j.brainresrev.2007.09.001

Li, Y., Zhu, J., Tian, G., Li, N., Li, Q., Ye, M., Zheng, H., Yu, J., Wu, H., Sun, J., Zhang, H., Chen, Q., Luo, R., Chen, M., He, Y., Jin, X., Zhang, Q., Yu, C., Zhou, G., … Zhang, X. (2010). The DNA methylome of human peripheral blood mononuclear cells. *PLoS Biology*, *8*(11), e1000533. https://doi.org/10.1371/journal.pbio.1000533

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q. M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., & Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*(7271), 315–322. https://doi.org/10.1038/nature08514

Liu, J., Morgan, M., Hutchison, K., & Calhoun, V. D. (2010). A study of the influence of sex on genome wide methylation. *PLoS One*, *5*. https://doi.org/10.1371/journal.pone.0010028

Lopes-Ramos, C. M., Quackenbush, J., & DeMeo, D. L. (2020). Genome-Wide Sex and Gender Differences in Cancer. In *Frontiers in Oncology* (Vol. 10, p. 2486). https://www.frontiersin.org/article/10.3389/fonc.2020.597788

Lorubbio, M., Conti, A. A., & Ognibene, A. (2018). *Midregional pro-adrenomedullin (MR-ProADM) reference values in serum*.

Lovestone, S., Francis, P., Kloszewska, I., Mecocci, P., Simmons, A., Soininen, H., Spenger, C., Tsolaki, M., Vellas, B., Wahlund, L. O., & Ward, M. (2009).

AddNeuroMed - The european collaboration for the discovery of novel biomarkers for alzheimer's disease. *Annals of the New York Academy of Sciences*, *1180*(1), 36–46. https://doi.org/10.1111/j.1749-6632.2009.05064.x

Lovestone, S., Francis, P., & Strandgaard, K. (2007). Biomarkers for disease modification trials - The innovative medicines initiative and addneuromed. *Journal of Nutrition, Health and Aging*, *11*(4), 359–361.

Lu, A. T., Quach, A., Wilson, J. G., Reiner, A. P., Aviv, A., Raj, K., Hou, L., Baccarelli, A. A., Li, Y., Stewart, J. D., Whitsel, E. A., Assimes, T. L., Ferrucci, L., & Horvath, S. (2019). DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging*, *11*(2), 303–327. https://doi.org/10.18632/aging.101684

Lunnon, K., Hannon, E., Smith, R. G., Dempster, E., Wong, C., Burrage, J., Troakes, C., Al-Sarraj, S., Kepa, A., Schalkwyk, L., & Mill, J. (2016). Variation in 5-hydroxymethylcytosine across human cortex and cerebellum. *Genome Biology*, *17*(1), 27. https://doi.org/10.1186/s13059-016-0871-x

Lunnon, K., Ibrahima, Z., Proitsi, P., Lourdusamy, A., Newhouse, S., Sattlecker, M., Furney, S., Saleem, M., Soininen, H., Kłoszewska, I., Mecocci, P., Tsolaki, M., Vellas, B., Coppola, G., Geschwind, D., Simmons, A., Lovestone, S., Dobson, R., & Hodges, A. (2012). Mitochondrial dysfunction and immune activation are detectable in early alzheimer's disease blood. *Journal of Alzheimer's Disease*, *30*(3), 685–710. https://doi.org/10.3233/JAD-2012-111592

Lunnon, K., Keohane, A., Pidsley, R., Newhouse, S., Riddoch-Contreras, J., Thubron, E. B., Devall, M., Soininen, H., Kłoszewska, I., Mecocci, P., Tsolaki, M., Vellas, B., Schalkwyk, L., Dobson, R., Malik, A. N., Powell, J., Lovestone, S., & Hodges, A. (2017). Mitochondrial genes are altered in blood early in Alzheimer's disease. *Neurobiology of Aging*, *53*, 36–47. https://doi.org/10.1016/j.neurobiolaging.2016.12.029

Lunnon, K., Sattlecker, M., Furney, S. J., Coppola, G., Simmons, A., Proitsi, P., Lupton, M. K., Lourdusamy, A., Johnston, C., Soininen, H., Kłoszewska, I., Mecocci, P., Tsolaki, M., Vellas, B., Geschwind, D., Lovestone, S., Dobson, R., & Hodges, A. (2013). A blood gene expression marker of early Alzheimer's disease. *Journal of Alzheimer's Disease*, *33*(3), 737–753. https://doi.org/10.3233/JAD-2012-121363

Lunnon, K., Smith, R. G., Hannon, E., De Jager, P. L., Srivastava, G., Volta, M., Troakes, C., Al-Sarraj, S., Burrage, J., Macdonald, R., Condliffe, D., Harries, L. W., Katsel, P., Haroutunian, V., Kaminsky, Z., Joachim, C., Powell, J., Lovestone, S., Bennett, D. A., … Mill, J. (2014). Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nature Neuroscience*, *17*(9), 1164–1170. https://doi.org/10.1038/nn.3782

Luteijn, M. J., & Ketting, R. F. (2013). PIWI-interacting RNAs: From generation to transgenerational epigenetics. *Nature Reviews Genetics*, *14*(8), 523–534. https://doi.org/10.1038/nrg3495

Lyketsos, C. G., Carrillo, M. C., Ryan, J. M., Khachaturian, A. S., Trzepacz, P., Amatniek, J., Cedarbaum, J., Brashear, R., & Miller, D. S. (2011). Neuropsychiatric symptoms in Alzheimer's disease. *Alzheimer's and Dementia*, *7*(5), 532–539. https://doi.org/10.1016/j.jalz.2011.05.2410

Maass, A., Landau, S., Baker, S. L., Horng, A., Lockhart, S. N., La Joie, R., Rabinovici, G. D., & Jagust, W. J. (2017). Comparison of multiple tau-PET measures as biomarkers in aging and Alzheimer's disease. *NeuroImage*, *157*, 448–463. https://doi.org/https://doi.org/10.1016/j.neuroimage.2017.05.058

Maccioni, R. B., Lavados, M., Guillón, M., Mujica, C., Bosch, R., Farías, G., & Fuentes, P. (2006). Anomalously phosphorylated tau and Aβ fragments in the CSF correlates with cognitive impairment in MCI subjects. *Neurobiology of Aging*, *27*(2), 237–244. https://doi.org/10.1016/j.neurobiolaging.2005.01.011

Madrid, A., Hogan, K. J., Papale, L. A., Clark, L. R., Asthana, S., Johnson, S. C., & Alisch, R. S. (2018). DNA Hypomethylation in Blood Links B3GALT4 and ZADH2 to Alzheimer's Disease. *Journal of Alzheimer's Disease*, *66*(3), 927–934. https://doi.org/10.3233/JAD-180592

Mann, D. M. A., & Esiri, M. M. (1989). The pattern of acquisition of plaques and tangles in the brains of patients under 50 years of age with Down's syndrome. *Journal of the Neurological Sciences*, *89*(2–3), 169–179. https://doi.org/10.1016/0022-510X(89)90019-1

Mansell, G., Gorrie-Stone, T. J., Bao, Y., Kumari, M., Schalkwyk, L. S., Mill, J., & Hannon, E. (2019). Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. *BMC Genomics*, *20*(1), 366. https://doi.org/10.1186/s12864-019-5761-7

Marioni, R. E., Shah, S., McRae, A. F., Chen, B. H., Colicino, E., Harris, S. E., Gibson, J., Henders, A. K., Redmond, P., Cox, S. R., Pattie, A., Corley, J., Murphy, L., Martin, N. G., Montgomery, G. W., Feinberg, A. P., Fallin, M. D., Multhaup, M. L., Jaffe, A. E., … Deary, I. J. (2015). DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biology*, *16*(1), 25. https://doi.org/10.1186/s13059-015-0584-6

Mastroeni, D., Grover, A., Delvaux, E., Whiteside, C., Coleman, P. D., & Rogers, J. (2010). Epigenetic changes in Alzheimer's disease: Decrements in DNA methylation. *Neurobiology of Aging*, *31*(12), 2025–2037. https://doi.org/10.1016/j.neurobiolaging.2008.12.005

Mastroeni, D., McKee, A., Grover, A., Rogers, J., & Coleman, P. D. (2009). Epigenetic differences in cortical neurons from a pair of monozygotic twins discordant for Alzheimer's disease. *PLoS ONE*, *4*(8), e6617. https://doi.org/10.1371/journal.pone.0006617

Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J. Z., Menon, M., He, L., Abdurrob, F., Jiang, X., Martorell, A. J., Ransohoff, R. M., Hafler, B. P., Bennett, D. A., Kellis, M., & Tsai, L.-H. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*, *570*(7761), 332–337. https://doi.org/10.1038/s41586-019-1195-2

Matthews, F. E., Stephan, B. C. M., Robinson, L., Jagger, C., Barnes, L. E., Arthur, A., Brayne, C., Comas-Herrera, A., Wittenberg, R., Dening, T., McCracken, C. F. M., Moody, C., Parry, B., Green, E., Barnes, R., Warwick, J., Gao, L., Mattison, A., Baldwin, C., … Collaboration, C. F. and A. S. (CFAS). (2016). A two decade dementia incidence comparison from the Cognitive Function and Ageing Studies I and II. *Nature Communications*, *7*(1), 11398. https://doi.org/10.1038/ncomms11398

Mattick, J. S. (2011). The central role of RNA in human development and cognition. *FEBS Letters*, *585*(11), 1600–1616. https://doi.org/10.1016/j.febslet.2011.05.001

Mattsson, N., Zetterberg, H., Hansson, O., Andreasen, N., Parnetti, L., Jonsson, M., Herukka, S.-K., van der Flier, W. M., Blankenstein, M. A., Ewers, M., Rich, K., Kaiser, E., Verbeek, M., Tsolaki, M., Mulugeta, E., Rosén, E., Aarsland, D., Visser, P. J., Schröder, J., … Blennow, K. (2009). CSF Biomarkers and Incipient Alzheimer Disease in Patients With Mild Cognitive Impairment. *JAMA*, *302*(4), 385–393.

https://doi.org/10.1001/jama.2009.1064

Maunakea, A. K., Chepelev, I., Cui, K., & Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Research*, *23*(11), 1256–1269. https://doi.org/10.1038/cr.2013.110

Mazure, C. M., & Swendsen, J. (2016). Sex differences in Alzheimer's disease and other dementias. *The Lancet Neurology*, *15*(5), 451–452. https://doi.org/10.1016/S1474-4422(16)00067-3

Mazurek, M. F., Beal, M. F., Bird, E. D., & Martin, J. B. (1987). Oxytocin in Alzheimer's disease. *Neurology*, *37*(6), 1001 LP – 1001. https://doi.org/10.1212/WNL.37.6.1001

McCarthy, N. S., Melton, P. E., Cadby, G., Yazar, S., Franchina, M., Moses, E. K., Mackey, D. A., & Hewitt, A. W. (2014). Meta-analysis of human methylation data for evidence of sex-specific autosomal patterns. *BMC Genomics*, *15*(1), 981. https://doi.org/10.1186/1471-2164-15-981

McCartney, D. L., Stevenson, A. J., Walker, R. M., Gibson, J., Morris, S. W., Campbell, A., Murray, A. D., Whalley, H. C., Porteous, D. J., McIntosh, A. M., Evans, K. L., Deary, I. J., & Marioni, R. E. (2018). Investigating the relationship between DNA methylation age acceleration and risk factors for Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *10*, 429–437. https://doi.org/https://doi.org/10.1016/j.dadm.2018.05.006

McCartney, D. L., Walker, R. M., Morris, S. W., McIntosh, A. M., Porteous, D. J., & Evans, K. L. (2016). Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genomics Data*, *9*, 22–24. https://doi.org/https://doi.org/10.1016/j.gdata.2016.05.012

McCartney, D. L., Zhang, F., Hillary, R. F., Zhang, Q., Stevenson, A. J., Walker, R. M., Bermingham, M. L., Boutin, T., Morris, S. W., Campbell, A., Murray, A. D., Whalley, H. C., Porteous, D. J., Hayward, C., Evans, K. L., Chandra, T., Deary, I. J., McIntosh, A. M., Yang, J., … Marioni, R. E. (2019). An epigenome-wide association study of sex-specific chronological ageing. *Genome Medicine*, *12*(1), 1. https://doi.org/10.1186/s13073-019-0693-z

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., & Kawas, C. H. (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. *Alzheimers Dement*, *7*. https://doi.org/10.1016/j.jalz.2011.03.005

Meda, L., Cassatella, M. A., Szendrei, G. I., Otvos, L., Baron, P., Villalba, M., Ferrari, D., & Rossi, F. (1995). Activation of microglial cells by β-amyloid protein and interferon-γ. *Nature*, *374*(6523), 647–650. https://doi.org/10.1038/374647a0

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R., & Lander, E. S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, *454*(7205), 766–770. https://doi.org/10.1038/nature07107

Mendez, M. F. (2012). Early-onset Alzheimer's Disease: Nonamnestic Subtypes and Type 2 AD. *Archives of Medical Research*, *43*(8), 677–685. https://doi.org/10.1016/j.arcmed.2012.11.009

Meyer, M. R., Tschanz, J. T., Norton, M. C., Welsh-Bohmer, K. A., Steffens, D. C., Wyse, B. W., & Breitner, J. C. S. (1998). APOE genotype predicts when-not whether-one is predisposed to develop Alzheimer disease [2]. *Nature Genetics*, *19*(4), 321–322. https://doi.org/10.1038/1206

Michelucci, A., Heurtaux, T., Grandbarbe, L., Morga, E., & Heuschling, P. (2009). Characterization of the microglial phenotype under specific pro-inflammatory and anti-inflammatory conditions: Effects of oligomeric and fibrillar amyloid-β. *Journal of Neuroimmunology*, *210*(1–2), 3–12. https://doi.org/10.1016/j.jneuroim.2009.02.003

Mielke, M. M., Ferretti, M. T., Iulita, M. F., Hayden, K., & Khachaturian, A. S. (2018). Sex and gender in Alzheimer's disease – Does it matter? *Alzheimer's & Dementia*, *14*(9), 1101–1103. https://doi.org/https://doi.org/10.1016/j.jalz.2018.08.003

Mielke, M. M., & Lyketsos, C. G. (2010). Alterations of the sphingolipid pathway in Alzheimer's disease: new biomarkers and treatment targets? *Neuromolecular Medicine*, *12*(4), 331–340. https://doi.org/10.1007/s12017-010-8121-y

Morley, J. E., & Farr, S. A. (2014). The role of amyloid-beta in the regulation of memory. *Biochemical Pharmacology*, *88*(4), 479–485. https://doi.org/https://doi.org/10.1016/j.bcp.2013.12.018

Morley, J. E., Farr, S. A., Nguyen, A. D., & Xu, F. (2019). *What is the Physiological*

Function of Amyloid-Beta Protein? Springer.

Mosch, B., Morawski, M., Mittag, A., Lenz, D., Tarnok, A., & Arendt, T. (2007). Aneuploidy and DNA Replication in the Normal Human Brain and Alzheimer's Disease. *The Journal of Neuroscience*, *27*(26), 6859 LP – 6867. https://doi.org/10.1523/JNEUROSCI.0379-07.2007

Müller, T., Meyer, H. E., Egensperger, R., & Marcus, K. (2008). The amyloid precursor protein intracellular domain (AICD) as modulator of gene expression, apoptosis, and cytoskeletal dynamics-Relevance for Alzheimer's disease. *Progress in Neurobiology*, *85*(4), 393–406. https://doi.org/10.1016/j.pneurobio.2008.05.002

Naj, A. C., Jun, G., Beecham, G. W., Wang, L. S., Vardarajan, B. N., Buros, J., Gallins, P. J., Buxbaum, J. D., Jarvik, G. P., Crane, P. K., Larson, E. B., Bird, T. D., Boeve, B. F., Graff-Radford, N. R., De Jager, P. L., Evans, D., Schneider, J. A., Carrasquillo, M. M., Ertekin-Taner, N., … Schellenberg, G. D. (2011). Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nature Genetics*, *43*(5), 436–443. https://doi.org/10.1038/ng.801

Nelson, P. T., Head, E., Schmitt, F. A., Davis, P. R., Neltner, J. H., Jicha, G. A., Abner, E. L., Smith, C. D., Van Eldik, L. J., Kryscio, R. J., & Scheff, S. W. (2011). Alzheimer's disease is not "brain aging": neuropathological, genetic, and epidemiological human studies. *Acta Neuropathologica*, *121*(5), 571–587. https://doi.org/10.1007/s00401-011-0826-y

Nestor, C. E., Ottaviano, R., Reddington, J., Sproul, D., Reinhardt, D., & Dunican, D. (2012). Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Res.*, *22*. https://doi.org/10.1101/gr.126417.111

Neu, S. C., Pa, J., Kukull, W., Beekly, D., Kuzma, A., Gangadharan, P., Wang, L.-S., Romero, K., Arneric, S. P., Redolfi, A., Orlandi, D., Frisoni, G. B., Au, R., Devine, S., Auerbach, S., Espinosa, A., Boada, M., Ruiz, A., Johnson, S. C., … Toga, A. W. (2017). Apolipoprotein E Genotype and Sex Risk Factors for Alzheimer Disease: A Meta-analysis. *JAMA Neurology*, *74*(10), 1178–1189. https://doi.org/10.1001/jamaneurol.2017.2188

Norden, D. M., & Godbout, J. P. (2013). Review: Microglia of the aged brain: Primed to be activated and resistant to regulation. *Neuropathology and Applied Neurobiology*, *39*(1), 19–34. https://doi.org/10.1111/j.1365-

2990.2012.01306.x

North, W. G., Harbaugh, R., & Reeder, T. (1992). An evaluation of human neurophysin production in Alzheimer's disease: Preliminary observations. *Neurobiology of Aging*, *13*(2), 261–265. https://doi.org/https://doi.org/10.1016/0197-4580(92)90038-Y

Nyrén, P. (1987). Enzymatic method for continuous monitoring of DNA polymerase activity. *Analytical Biochemistry*, *167*(2), 235–238. https://doi.org/https://doi.org/10.1016/0003-2697(87)90158-8

Nyrén, P., Pettersson, B., & Uhlén, M. (1993). Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. *Analytical Biochemistry*, *208*(1), 171–175. https://doi.org/10.1006/abio.1993.1024

O'Bryant, S. E., Edwards, M., Johnson, L., Hall, J., Villarreal, A. E., Britton, G. B., Quiceno, M., Cullum, C. M., & Graff-Radford, N. R. (2016). A blood screening test for Alzheimer's disease. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, *3*, 83–90. https://doi.org/10.1016/j.dadm.2016.06.004

O'Bryant, S. E., Xiao, G., Barber, R., Reisch, J., Doody, R., Fairchild, T., Adams, P., Waring, S., & Diaz-Arrastia, R. (2010). A serum protein-based algorithm for the detection of Alzheimer disease. *Archives of Neurology*, *67*(9), 1077–1081. https://doi.org/10.1001/archneurol.2010.215

O'Bryant, S. E., Xiao, G., Barber, R., Reisch, J., Hall, J., Cullum, C. M., Doody, R., Fairchild, T., Adams, P., Wilhelmsen, K., & Diaz-Arrastia, R. (2011). A blood-based algorithm for the detection of Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, *32*(1), 55–62. https://doi.org/10.1159/000330750

Olff, M., Frijling, J. L., Kubzansky, L. D., Bradley, B., Ellenbogen, M. A., Cardoso, C., Bartz, J. A., Yee, J. R., & van Zuiden, M. (2013). The role of oxytocin in social bonding, stress regulation and mental health: An update on the moderating effects of context and interindividual differences. *Psychoneuroendocrinology*, *38*(9), 1883–1894. https://doi.org/10.1016/j.psyneuen.2013.06.019

Ørstavik, K. H. (2006). Skewed X inactivation in healthy individuals and in different diseases. *Acta Paediatrica*, *95*(S451), 24–29. https://doi.org/https://doi.org/10.1111/j.1651-2227.2006.tb02385.x

Paziewska, A., Dabrowska, M., Goryca, K., Antoniewicz, A., Dobruch, J., Mikula, M., Jarosz, D., Zapala, L., Borowka, A., & Ostrowski, J. (2014). DNA methylation status is more reliable than gene expression at detecting cancer in prostate biopsy. *British Journal of Cancer*, *111*(4), 781–789. https://doi.org/10.1038/bjc.2014.337

Pedersen, B. S., Schwartz, D. A., Yang, I. V., & Kechris, K. J. (2012). Comb-p: Software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics*, *28*(22), 2986–2988. https://doi.org/10.1093/bioinformatics/bts545

Perry, V. H., & Holmes, C. (2014). Microglial priming in neurodegenerative disease. *Nature Reviews Neurology*, *10*(4), 217–224. https://doi.org/10.1038/nrneurol.2014.38

Peters, M. J., Joehanes, R., Pilling, L. C., Schurmann, C., Conneely, K. N., Powell, J., Reinmaa, E., Sutphin, G. L., Zhernakova, A., Schramm, K., Wilson, Y. A., Kobes, S., Tukiainen, T., Nalls, M. A., Hernandez, D. G., Cookson, M. R., Gibbs, R. J., Hardy, J., Ramasamy, A., … Consortium, N. (2015). The transcriptional landscape of age in human peripheral blood. *Nature Communications*, *6*(1), 8570. https://doi.org/10.1038/ncomms9570

Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., Jack, C. R., Jagust, W. J., Shaw, L. M., Toga, A. W., Trojanowski, J. Q., & Weiner, M. W. (2010). Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization. *Neurology*, *74*(3), 201–209. https://doi.org/10.1212/WNL.0b013e3181cb3e25

Petersen, Ronald C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology*, *56*(3), 303–308. https://doi.org/10.1001/archneur.56.3.303

Petersen, Ronald C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, *256*(3), 183–194. https://doi.org/10.1111/j.1365-2796.2004.01388.x

Phipson, B., Maksimovic, J., & Oshlack, A. (2016). MissMethyl: An R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics*, *32*(2), 286–288. https://doi.org/10.1093/bioinformatics/btv560

Pidsley, R., Y Wong, C. C., Volta, M., Lunnon, K., Mill, J., & Schalkwyk, L. C.

(2013). A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*, *14*(1), 293. https://doi.org/10.1186/1471-2164-14-293

Piovesan, A., Pelleri, M. C., Antonaros, F., Strippoli, P., Caracausi, M., & Vitale, L. (2019). On the length, weight and GC content of the human genome. *BMC Research Notes*, *12*(1), 106. https://doi.org/10.1186/s13104-019-4137-z

Price, M. E., Cotton, A. M., Lam, L. L., Farré, P., Emberly, E., Brown, C. J., Robinson, W. P., & Kobor, M. S. (2013). Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics and Chromatin*, *6*(1), 4. https://doi.org/10.1186/1756-8935-6-4

Prince, M., Wimo, A., Guerchet, M., Gemma-Claire, A., Wu, Y.-T., & Prina, M. (2015). World Alzheimer Report 2015: The Global Impact of Dementia - An analysis of prevalence, incidence, cost and trends. *Alzheimer's Disease International*, 84. www.alz.co.uk

Qiu, Z., Liang, N., Sun, T., Xue, H., Xie, T., Wang, X., & Wang, Q. (2020). Downregulation of DUSP9 promotes tumor progression and contributes to poor prognosis in human colorectal cancer. *Frontiers in Oncology*, *10*, 1976.

R Development Core Team 3.0.1. (2013). A Language and Environment for Statistical Computing. In *R Foundation for Statistical Computing*. http://www.r-project.org

Rajasekhar, K., Chakrabarti, M., & Govindaraju, T. (2015). Function and toxicity of amyloid beta and recent therapeutic interventions targeting amyloid beta in Alzheimer's disease. *Chemical Communications*, *51*(70), 13434–13450.

Rao, J. S., Keleshian, V. L., Klein, S., & Rapoport, S. I. (2012). Epigenetic modifications in frontal cortex from Alzheimer's disease and bipolar disorder patients. *Translational Psychiatry*, *2*(7), e132. https://doi.org/10.1038/tp.2012.55

Raznahan, A., Parikshak, N. N., Chandran, V., Blumenthal, J. D., Clasen, L. S., Alexander-Bloch, A. F., Zinn, A. R., Wangsa, D., Wise, J., Murphy, D. G. M., Bolton, P. F., Ried, T., Ross, J., Giedd, J. N., & Geschwind, D. H. (2018). Sex-chromosome dosage effects on gene expression in humans. *Proceedings of the National Academy of Sciences*, *115*(28), 7398 LP – 7403. https://doi.org/10.1073/pnas.1802889115

Reitz, C., & Mayeux, R. (2014). Alzheimer disease: Epidemiology, diagnostic

criteria, risk factors and biomarkers. *Biochemical Pharmacology*, *88*(4), 640–651. https://doi.org/10.1016/j.bcp.2013.12.024

Rocca, W. A., Mielke, M. M., Vemuri, P., & Miller, V. M. (2014). Sex and gender differences in the causes of dementia: A narrative review. *Maturitas*, *79*(2), 196–201. https://doi.org/https://doi.org/10.1016/j.maturitas.2014.05.008

Roubroeks, J. A. Y., Smith, A. R., Smith, R. G., Pishva, E., Ibrahim, Z., Sattlecker, M., Hannon, E. J., Kłoszewska, I., Mecocci, P., Soininen, H., Tsolaki, M., Vellas, B., Wahlund, L. O., Aarsland, D., Proitsi, P., Hodges, A., Lovestone, S., Newhouse, S. J., Dobson, R. J. B., … Lunnon, K. (2020). An epigenome-wide association study of Alzheimer's disease blood highlights robust DNA hypermethylation in the HOXB6 gene. *Neurobiology of Aging*, *95*, 26–45. https://doi.org/10.1016/j.neurobiolaging.2020.06.023

Roubroeks, J. A. Y., Smith, R. G., van den Hove, D. L. A., & Lunnon, K. (2017). Epigenetics and DNA methylomic profiling in Alzheimer's disease and other neurodegenerative diseases. *Journal of Neurochemistry*, *143*(2), 158–170. https://doi.org/10.1111/jnc.14148

Rye, P. D., Booij, B. B., Grave, G., Lindahl, T., Kristiansen, L., Andersen, H. M., Horndalsveen, P. O., Nygaard, H. A., Naik, M., Hoprekstad, D., Wetterberg, P., Nilsson, C., Aarsland, D., Sharma, P., & Lönneborg, A. (2011). A novel blood test for the early detection of Alzheimer's disease. *Journal of Alzheimer's Disease*, *23*(1), 121–129. https://doi.org/10.3233/JAD-2010-101521

Sadakierska-Chudy, A., & Filip, M. (2014). A Comprehensive View of the Epigenetic Landscape. Part II: Histone Post-translational Modification, Nucleosome Level, and Chromatin Regulation by ncRNAs. *Neurotoxicity Research*, *27*(2), 172–197. https://doi.org/10.1007/s12640-014-9508-6

Saffari, A., Silver, M. J., Zavattari, P., Moi, L., Columbano, A., Meaburn, E. L., & Dudbridge, F. (2018). Estimation of a significance threshold for epigenome-wide association studies. *Genetic Epidemiology*, *42*(1), 20–33. https://doi.org/10.1002/gepi.22086

Sanchez-Mut, J. V., Aso, E., Heyn, H., Matsuda, T., Bock, C., Ferrer, I., & Esteller, M. (2014). Promoter hypermethylation of the phosphatase DUSP22 mediates PKA-dependent TAU phosphorylation and CREB activation in Alzheimer's disease. *Hippocampus*, *24*(4), 363–368. https://doi.org/10.1002/hipo.22245

Scelfo, A., & Fachinetti, D. (2019). Keeping the Centromere under Control: A Promising Role for DNA Methylation. *Cells*, *8*(8), 912. https://doi.org/10.3390/cells8080912

Schalkwyk, L. C., Pidsley, R., & Wong, C. C. Y. (2013). wateRmelon: Illumina 450 methylation array normalization and metrics. In *R package version 1.2.2* (p.).

Schwartzentruber, J., Cooper, S., Liu, J. Z., Barrio-Hernandez, I., Bello, E., Kumasaka, N., Young, A. M. H., Franklin, R. J. M., Johnson, T., Estrada, K., Gaffney, D. J., Beltrao, P., & Bassett, A. (2021). Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nature Genetics*, *53*(3), 392–402. https://doi.org/10.1038/s41588-020-00776-w

Scott, T. J., O'Connor, A. C., Link, A. N., & Beaulieu, T. J. (2014). Economic analysis of opportunities to accelerate Alzheimer's disease research and development. *Annals of the New York Academy of Sciences*, *1313*(1), 17–34. https://doi.org/10.1111/nyas.12417

Selkoe, D. J. (1991). The molecular pathology of Alzheimer's disease. *Neuron*, *6*(4), 487–498.

Sengupta, A., Kabat, J., Novak, M., Wu, Q., Grundke-Iqbal, I., & Iqbal, K. (1998). Phosphorylation of tau at both Thr 231 and Ser 262 is required for maximal inhibition of its binding to microtubules. *Archives of Biochemistry and Biophysics*, *357*(2), 299–309. https://doi.org/10.1006/abbi.1998.0813

Shackelford, D. A. (2006). DNA end joining activity is reduced in Alzheimer's disease. *Neurobiology of Aging*, *27*(4), 596–605. https://doi.org/https://doi.org/10.1016/j.neurobiolaging.2005.03.009

Sharp, A., Robinson, D., & Jacobs, P. (2000). Age- and tissue-specific variation of X chromosome inactivation ratios in normal women. *Human Genetics*, *107*(4), 343–349. https://doi.org/10.1007/s004390000382

Shin, I. S., Carter, M., Masterman, D., Fairbanks, L., & Cummings, J. L. (2005). Neuropsychiatric symptoms and quality of life in Alzheimer disease. *American Journal of Geriatric Psychiatry*, *13*(6), 469–474. https://doi.org/10.1097/00019442-200506000-00005

Sibbett, R. A., Altschul, D. M., Marioni, R. E., Deary, I. J., Starr, J. M., & Russ, T. C. (2020). DNA methylation-based measures of accelerated biological ageing and the risk of dementia in the oldest-old: A study of the Lothian Birth

Cohort 1921. *BMC Psychiatry*, *20*(1), 91. https://doi.org/10.1186/s12888-020-2469-9

Siegmund, K. D., Connor, C. M., Campan, M., Long, T. L., Weisenberger, D. J., Biniszkiewicz, D., Jaenisch, R., Laird, P. W., & Akbarian, S. (2007). DNA methylation in the human cerebral cortex is dynamically regulated throughout the life span and involves differentiated neurons. *PLoS ONE*, *2*(9), e895. https://doi.org/10.1371/journal.pone.0000895

Simic, G., Stanic, G., Mladinov, M., Jovanov-Milosevic, N., Kostovic, I., & Hof, P. R. (2009). Does Alzheimer's disease begin in the brainstem?: Annotation. *Neuropathology and Applied Neurobiology*, *35*(6), 532–554. https://doi.org/10.1111/j.1365-2990.2009.01038.x

Simmons, A., Westman, E., Muehlboeck, S., Mecocci, P., Vellas, B., Tsolaki, M., Kåoszewska, I., Wahlund, L. O., Soininen, H., Lovestone, S., Evans, A., & Spenger, C. (2011). The AddNeuroMed framework for multi-centre MRI assessment of Alzheimer's disease: Experience from the first 24 months. *International Journal of Geriatric Psychiatry*, *26*(1), 75–82. https://doi.org/10.1002/gps.2491

Singmann, P., Shem-Tov, D., Wahl, S., Grallert, H., Fiorito, G., Shin, S.-Y., Schramm, K., Wolf, P., Kunze, S., Baran, Y., Guarrera, S., Vineis, P., Krogh, V., Panico, S., Tumino, R., Kretschmer, A., Gieger, C., Peters, A., Prokisch, H., … Halperin, E. (2015). Characterization of whole-genome autosomal differences of DNA methylation between men and women. *Epigenetics & Chromatin*, *8*(1), 43. https://doi.org/10.1186/s13072-015-0035-3

Smallwood, S. A., Lee, H. J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S. R., Stegle, O., Reik, W., & Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nature Methods*, *11*(8), 817–820. https://doi.org/10.1038/nmeth.3035

Smith, A. K., Kilaru, V., Klengel, T., Mercer, K. B., Bradley, B., Conneely, K. N., Ressler, K. J., & Binder, E. B. (2015). DNA extracted from saliva for methylation studies of psychiatric traits: Evidence tissue specificity and relatedness to brain. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *168*(1), 36–44. https://doi.org/https://doi.org/10.1002/ajmg.b.32278

Smith, A. R., Smith, R. G., Condliffe, D., Hannon, E., Schalkwyk, L., Mill, J., & Lunnon, K. (2016). Increased DNA methylation near TREM2 is consistently

seen in the superior temporal gyrus in Alzheimer's disease brain. In *Neurobiology of Aging* (Vol. 47). https://doi.org/10.1016/j.neurobiolaging.2016.07.008

Smith, A. R., Smith, R. G., Pishva, E., Hannon, E., Roubroeks, J. A. Y., Burrage, J., Troakes, C., Al-Sarraj, S., Sloan, C., Mill, J., Van Den Hove, D. L., & Lunnon, K. (2019). Parallel profiling of DNA methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in Alzheimer's disease. *Clinical Epigenetics*, *11*(1), 52. https://doi.org/10.1186/s13148-019-0636-y

Smith, J. G., Newton-Cheh, C., Hedblad, B., Struck, J., Morgenthaler, N. G., Bergmann, A., Wang, T. J., & Melander, O. (2009). Distribution and Correlates of Midregional Proadrenomedullin in the General Population. *Clinical Chemistry*, *55*(8), 1593–1595. https://doi.org/10.1373/clinchem.2009.126482

Smith, R. G., Hannon, E., De Jager, P. L., Chibnik, L., Lott, S. J., Condliffe, D., Smith, A. R., Haroutunian, V., Troakes, C., Al-Sarraj, S., Bennett, D. A., Powell, J., Lovestone, S., Schalkwyk, L., Mill, J., & Lunnon, K. (2018). Elevated DNA methylation across a 48-kb region spanning the HOXA gene cluster is associated with Alzheimer's disease neuropathology. *Alzheimer's and Dementia*, *14*(12), 1580–1588. https://doi.org/10.1016/j.jalz.2018.01.017

Smith, R. G., & Lunnon, K. (2017). DNA modifications and alzheimer's disease. In *Advances in Experimental Medicine and Biology* (Vol. 978, pp. 303–319). Springer New York LLC. https://doi.org/10.1007/978-3-319-53889-1_16

Smith, R., Pishva, E., Shireby, G., Smith, A., Roubroeks, J., Hannon, E., Wheildon, G., Mastroeni, D., Gasparoni, G., Riemenschneider, M., Giese, A., Sharp, A., Schalkwyk, L., Haroutunian, V., Viechtbauer, W., van den Hove, D., Weedon, M., Brokaw, D., Francis, P., … Lunnon, K. (2020). A meta-analysis of epigenome-wide association studies in Alzheimer's disease highlights novel differentially methylated loci across cortex. *BioRxiv*, 2020.02.28.957894. https://doi.org/10.1101/2020.02.28.957894

Sontag, E., Hladik, C., Montgomery, L., Luangpirom, A., Mudrak, I., Ogris, E., & White, C. L. (2004). Downregulation of protein phosphatase 2A carboxyl methylation and methyltransferase may contribute to Alzheimer disease pathogenesis. *Journal of Neuropathology and Experimental Neurology*,

63(10), 1080–1091. https://doi.org/10.1093/jnen/63.10.1080

Spremo-Potparevic, B., Bajic, V., Perry, G., & Zivkovic, L. (2015). Alterations of the X Chromosome in Lymphocytes of Alzheimer's Disease Patients. In *Current Alzheimer Research* (Vol. 12, Issue 10, pp. 990–996). https://doi.org/10.2174/1567205012666151027124154

Spremo-Potparević, B., Zivković, L., Djelić, N., Plećas-Solarović, B., Smith, M. A., & Bajić, V. (2008). Premature centromere division of the X chromosome in neurons in Alzheimer's disease. *Journal of Neurochemistry*, *106*(5), 2218–2223. https://doi.org/10.1111/j.1471-4159.2008.05555.x

Stavnichuk, M., Tauer, J. T., Nagy, Z., Mazharian, A., Welman, M., Lordkipanidzé, M., Senis, Y. A., & Komarova, S. V. (2021). Severity of megakaryocyte-driven osteosclerosis in Mpig6b-deficient mice is sex-linked. *Journal of Bone and Mineral Research*, *n/a*(n/a). https://doi.org/https://doi.org/10.1002/jbmr.4245

Strehle, M., & Guttman, M. (2020). Xist drives spatial compartmentalization of DNA and protein to orchestrate initiation and maintenance of X inactivation. *Current Opinion in Cell Biology*, *64*, 139–147. https://doi.org/https://doi.org/10.1016/j.ceb.2020.04.009

Sun, W., Zang, L., Shu, Q., & Li, X. (2014). From development to diseases: The role of 5hmC in brain. *Genomics*, *104*(5), 347–351. https://doi.org/10.1016/j.ygeno.2014.08.021

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., & Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, *324*(5929), 930–935. https://doi.org/10.1126/science.1170116

Takizawa, C., Thompson, P. L., Van Walsem, A., Faure, C., & Maier, W. C. (2014). Epidemiological and economic burden of Alzheimer's disease: A systematic literature review of data across Europe and the United States of America. *Journal of Alzheimer's Disease*, *43*(4), 1271–1284. https://doi.org/10.3233/JAD-141134

Tamayev, R., Matsuda, S., Arancio, O., & D'Adamio, L. (2012). β- but not γ-secretase proteolysis of APP causes synaptic and memory deficits in a mouse model of dementia. *EMBO Molecular Medicine*, *4*(3), 171–179. https://doi.org/10.1002/emmm.201100195

Thach, D. C., Lin, B., Walter, E., Kruzelock, R., Rowley, R. K., Tibbetts, C., & Stenger, D. A. (2003). Assessment of two methods for handling blood in collection tubes with RNA stabilizing agent for surveillance of gene expression profiles with high density microarrays. *Journal of Immunological Methods*, *283*(1–2), 269–279. https://doi.org/10.1016/j.jim.2003.10.004

Thathiah, A., & De Strooper, B. (2011). The role of G protein-coupled receptors in the pathology of Alzheimer's disease. *Nature Reviews Neuroscience*, *12*(2), 73–87. https://doi.org/10.1038/nrn2977

Thomson, P. A., Wray, N. R., Thomson, A. M., Dunbar, D. R., Grassie, M. A., Condie, A., Walker, M. T., Smith, D. J., Pulford, D. J., Muir, W., Blackwood, D. H. R., & Porteous, D. J. (2005). Sex-specific association between bipolar affective disorder in women and GPR50, an X-linked orphan G protein-coupled receptor. *Molecular Psychiatry*, *10*(5), 470–478. https://doi.org/10.1038/sj.mp.4001593

Tukiainen, T., Villani, A.-C., Yen, A., Rivas, M. A., Marshall, J. L., Satija, R., Aguirre, M., Gauthier, L., Fleharty, M., Kirby, A., Cummings, B. B., Castel, S. E., Karczewski, K. J., Aguet, F., Byrnes, A., Aguet, F., Ardlie, K. G., Cummings, B. B., Gelfand, E. T., … Genome Browser Data Integration & Visualization—UCSC Genomics Institute, U. of C. S. C. (2017). Landscape of X chromosome inactivation across human tissues. *Nature*, *550*(7675), 244–248. https://doi.org/10.1038/nature24265

Vacca, M., Della Ragione, F., Scalabrì, F., & D'Esposito, M. (2016). X inactivation and reactivation in X-linked diseases. *Seminars in Cell & Developmental Biology*, *56*, 78–87. https://doi.org/https://doi.org/10.1016/j.semcdb.2016.03.009

Valenzuela-Sánchez, F., Valenzuela-Méndez, B., Rodríguez-Gutiérrez, J. F., Estella-García, Á., & González-García, M. Á. (2016). New role of biomarkers: mid-regional pro-adrenomedullin, the biomarker of organ failure. *Annals of Translational Medicine*, *4*(17), 329. https://doi.org/10.21037/atm.2016.08.65

van den Hove, D. L. A., Chouliaras, L., & Rutten, B. P. F. (2014). The Role of 5-Hydroxymethylcytosine in Aging and Alzheimer's Disease: Current Status and Prospects for Future Studies. *Current Alzheimer Research*, *9*(5), 545–549. https://doi.org/10.2174/1567205012800618008

Varley, K. E., Gertz, J., Bowling, K. M., Parker, S. L., Reddy, T. E., Pauli-Behn, F., Cross, M. K., Williams, B. A., Stamatoyannopoulos, J. A., Crawford, G.

E., Absher, D. M., Wold, B. J., & Myers, R. M. (2013). Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Research*, *23*(3), 555–567. https://doi.org/10.1101/gr.147942.112

Vartanian, K., Slottke, R., Johnstone, T., Casale, A., Planck, S. R., Choi, D., Smith, J. R., Rosenbaum, J. T., & Harrington, C. A. (2009). Gene expression profiling of whole blood: Comparison of target preparation methods for accurate and reproducible microarray analysis. *BMC Genomics*, *10*(1), 2. https://doi.org/10.1186/1471-2164-10-2

Vasanthakumar, A., Davis, J. W., Idler, K., Waring, J. F., Asque, E., Riley-Gillis, B., Grosskurth, S., Srivastava, G., Kim, S., Nho, K., Nudelman, K. N. H., Faber, K., Sun, Y., Foroud, T. M., Estrada, K., Apostolova, L. G., Li, Q. S., Saykin, A. J., & (ADNI), for the A. D. N. I. (2020). Harnessing peripheral DNA methylation differences in the Alzheimer's Disease Neuroimaging Initiative (ADNI) to reveal novel biomarkers of disease. *Clinical Epigenetics*, *12*(1), 84. https://doi.org/10.1186/s13148-020-00864-y

Villain, N., Chételat, G., Grassiot, B., Bourgeat, P., Jones, G., Ellis, K. A., Ames, D., Martins, R. N., Eustache, F., Salvado, O., Masters, C. L., Rowe, C. C., Villemagne, V. L., & Group, the A. R. (2012). Regional dynamics of amyloid-β deposition in healthy elderly, mild cognitive impairment and Alzheimer's disease: a voxelwise PiB–PET longitudinal study. *Brain*, *135*(7), 2126–2139. https://doi.org/10.1093/brain/aws125

Visser, P. J., Verhey, F., Knol, D. L., Scheltens, P., Wahlund, L.-O., Freund-Levi, Y., Tsolaki, M., Minthon, L., Wallin, Å. K., Hampel, H., Bürger, K., Pirttila, T., Soininen, H., Rikkert, M. O., Verbeek, M. M., Spiru, L., & Blennow, K. (2009). Prevalence and prognostic value of CSF markers of Alzheimer's disease pathology in patients with subjective cognitive impairment or mild cognitive impairment in the DESCRIPA study: a prospective cohort study. *The Lancet Neurology*, *8*(7), 619–627. https://doi.org/https://doi.org/10.1016/S1474-4422(09)70139-5

Vitiello, M. V., & Borson, S. (2001). Sleep disturbances in patients with alzheimer's disease: Epidemiology, pathophysiology and treatment. *CNS Drugs*, *15*(10), 777–796. https://doi.org/10.2165/00023210-200115100-00004

Volicer, L., Harper, D. G., Manning, B. C., Goldstein, R., & Satlin, A. (2001). Sundowning and circadian rhythms in Alzheimer's disease. *American*

Journal of Psychiatry, *158*(5), 704–711. https://doi.org/10.1176/appi.ajp.158.5.704

von Rotz, R. C., Kohli, B. M., Bosset, J., Meier, M., Suzuki, T., Nitsch, R. M., & Konietzko, U. (2004). The APP intracellular domain forms nuclear multiprotein complexes and regulates the transcription of its own precursor. *Journal of Cell Science*, *117*(19), 4435–4448. https://doi.org/10.1242/jcs.01323

Vourekas, A., Zheng, K., Fu, Q., Maragkakis, M., Alexiou, P., Ma, J., Pillai, R. S., Mourelatos, Z., & Jeremy Wang, P. (2015). The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing. *Genes and Development*, *29*(6), 617–629. https://doi.org/10.1101/gad.254631.114

Waddington, C. H. (2012). The epigenotype. 1942. *International Journal of Epidemiology*, *41*(1), 10–13. https://doi.org/10.1093/ije/dyr184

Walker, R. M., Vaher, K., Bermingham, M. L., Morris, S. W., Bretherick, A. D., Zeng, Y., Rawlik, K., Amador, C., Campbell, A., Haley, C. S., Hayward, C., Porteous, D. J., McIntosh, A. M., Marioni, R. E., & Evans, K. L. (2021). Identification of epigenome-wide DNA methylation differences between carriers of APOE ε4 and APOE ε2 alleles. *Genome Medicine*, *13*(1), 1. https://doi.org/10.1186/s13073-020-00808-4

Wang, S. C., Oeize, B., & Schumacher, A. (2008). Age-specific epigenetic drift in late-onset Alzheimer's disease. *PLoS ONE*, *3*(7), e2698. https://doi.org/10.1371/journal.pone.0002698

Wang, Yanni, Jiang, M., Yao, Y., & Cai, Z. (2017). WWC3 Inhibits Glioma Cell Proliferation Through Suppressing the Wnt/β-Catenin Signaling Pathway. *DNA and Cell Biology*, *37*(1), 31–37. https://doi.org/10.1089/dna.2017.3931

Wang, Yaping, & Wang, Z. (2020). Identification of dysregulated genes and pathways of different brain regions in Alzheimer's disease. *International Journal of Neuroscience*, *130*(11), 1082–1094. https://doi.org/10.1080/00207454.2020.1720677

Watson, C. T., Roussos, P., Garg, P., Ho, D. J., Azam, N., Katsel, P. L., Haroutunian, V., & Sharp, A. J. (2016). Genome-wide12 DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease. *Genome Medicine*, *8*(1), 5. https://doi.org/10.1186/s13073-015-0258-8

Wen, L., & Tang, F. (2014). Genomic distribution and possible functions of DNA

hydroxymethylation in the brain. *Genomics*, *104*(5), 341–346. https://doi.org/10.1016/j.ygeno.2014.08.020

Wenk, G. L. (2003). Neuropathologic changes in Alzheimer's disease. *Journal of Clinical Psychiatry*, *64*(SUPPL. 9), 7–10. https://pdfs.semanticscholar.org/defc/a46743480ba37f4dd832d797c538be3a686e.pdf

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

Westman, E., Simmons, A., Muehlboeck, J. S., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Weiner, M. W., Lovestone, S., Spenger, C., & Wahlund, L. O. (2011). AddNeuroMed and ADNI: Similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America. *NeuroImage*, *58*(3), 818–828. https://doi.org/10.1016/j.neuroimage.2011.06.065

Wierczeiko, A., Fournier, D., Todorov, H., Klingenberg, S., Endres, K., & Gerber, S. (2018). Decoupling of DNA Methylation Status and Gene Expression Levels in Aging Individuals. *Genomics and Computational Biology; Vol 4 No 2 (2018): Special Student IssueDO - 10.18547/Gcb.2018.Vol4.Iss2.E100040* . https://www.genomicscomputbiol.org/ojs3/GCB/article/view/40

Wimo, A., Jönsson, L., Gustavsson, A., McDaid, D., Ersek, K., Georges, J., Gulácsi, L., Karpati, K., Kenigsberg, P., & Valtonen, H. (2011). The economic impact of dementia in Europe in 2008-cost estimates from the Eurocode project. *International Journal of Geriatric Psychiatry*, *26*(8), 825–832. https://doi.org/10.1002/gps.2610

Wimo, Anders, Jönsson, L., Bond, J., Prince, M., & Winblad, B. (2013). The worldwide economic impact of dementia 2010. *Alzheimer's and Dementia*, *9*(1), 1-11.e3. https://doi.org/10.1016/j.jalz.2012.11.006

Wong, C. C. Y., Caspi, A., Williams, B., Houts, R., Craig, I. W., & Mill, J. (2011). A Longitudinal Twin Study of Skewed X Chromosome-Inactivation. *PLOS ONE*, *6*(3), e17873. https://doi.org/10.1371/journal.pone.0017873

Wu, F., Lv, T., Chen, G., Ye, H., Wu, W., Li, G., & Zhi, F.-C. (2015). Epigenetic silencing of DUSP9 induces the proliferation of human gastric cancer by

activating JNK signaling. *Oncol Rep*, *34*(1), 121–128. https://doi.org/10.3892/or.2015.3998

Wu, S. C., & Zhang, Y. (2010). Active DNA demethylation: Many roads lead to Rome. *Nature Reviews Molecular Cell Biology*, *11*(9), 607–620. https://doi.org/10.1038/nrm2950

Yaffe, K., Vittinghoff, E., Lindquist, K., Barnes, D., Covinsky, K. E., Neylan, T., Kluse, M., & Marmar, C. (2010). Posttraumatic stress disorder and risk of dementia among US veterans. *Archives of General Psychiatry*, *67*(6), 608–613. https://doi.org/10.1001/archgenpsychiatry.2010.61

Yang, X., Cheng, Y., Lu, Q., Wei, J., Yang, H., & Gu, M. (2015). Detection of stably expressed piRNAs in human blood. *International Journal of Clinical and Experimental Medicine*, *8*(8), 13353–13358. https://www.ncbi.nlm.nih.gov/pubmed/26550265

Yang, Y., Geldmacher, D. S., & Herrup, K. (2001). DNA replication precedes neuronal cell death in Alzheimer's disease. *Journal of Neuroscience*, *21*(8), 2661–2668. https://doi.org/10.1523/jneurosci.21-08-02661.2001

Yasuda, K., Yashiro, M., Sawada, T., Ohira, M., & Hirakawa, K. (2007). ERas Oncogene Expression and Epigenetic Regulation by Histone Acetylation in Human Cancer Cells. *Anticancer Research*, *27*(6B), 4071 LP – 4075. http://ar.iiarjournals.org/content/27/6B/4071.abstract

Yates, D., & McLoughlin, D. M. (2008). The molecular pathology of Alzheimer's disease. *Psychiatry*, *7*(1), 1–5. https://doi.org/10.1016/j.mppsy.2007.11.009

Yoshikai, S. ichi, Sasaki, H., Doh-ura, K., Furuya, H., & Sakaki, Y. (1990). Genomic organization of the human amyloid beta-protein precursor gene. *Gene*, *87*(2), 257–263. https://doi.org/10.1016/0378-1119(90)90310-N

Yu, L., Chibnik, L. B., Yang, J., McCabe, C., Xu, J., Schneider, J. A., De Jager, P. L., & Bennett, D. A. (2016). Methylation profiles in peripheral blood CD4+ lymphocytes versus brain: The relation to Alzheimer's disease pathology. *Alzheimer's & Dementia*, *12*(9), 942–951. https://doi.org/https://doi.org/10.1016/j.jalz.2016.02.009

Zeilinger, S., Kühnel, B., Klopp, N., Baurecht, H., Kleinschmidt, A., Gieger, C., Weidinger, S., Lattka, E., Adamski, J., Peters, A., Strauch, K., Waldenberger, M., & Illig, T. (2013). Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation. *PLoS ONE*, *8*(5), e63812. https://doi.org/10.1371/journal.pone.0063812

Zhang, L., Li, H., Hu, X., Benedek, D. M., Fullerton, C. S., Forsten, R. D., Naifeh, J. A., Li, X., Wu, H., Benevides, K. N., Le, T., Smerin, S., Russell, D. W., & Ursano, R. J. (2015). Mitochondria-focused gene expression profile reveals common pathways and CPT1B dysregulation in both rodent stress model and human subjects with PTSD. *Translational Psychiatry*, *5*(6), e580. https://doi.org/10.1038/tp.2015.65

Zhang, Lanyu, Silva, T. C., Young, J. I., Gomez, L., Schmidt, M. A., Hamilton-Nelson, K. L., Kunkle, B. W., Chen, X., Martin, E. R., & Wang, L. (2020). Epigenome-wide meta-analysis of DNA methylation differences in prefrontal cortex implicates the immune processes in Alzheimer's disease. *Nature Communications*, *11*(1), 6114. https://doi.org/10.1038/s41467-020-19791-w

Zhao, J., Zhu, Y., Yang, J., Li, L., Wu, H., De Jager, P. L., Jin, P., & Bennett, D. A. (2017). A genome-wide profiling of brain DNA hydroxymethylation in Alzheimer's disease. *Alzheimer's and Dementia*, *13*(6), 674–688. https://doi.org/10.1016/j.jalz.2016.10.004

Zhao, Q.-T., Guo, T., Wang, H.-E., Zhang, X.-P., Zhang, H., Wang, Z.-K., Yuan, Z., & Duan, G.-C. (2015). Diagnostic value of SHOX2 DNA methylation in lung cancer: a meta-analysis. *OncoTargets and Therapy*, *8*, 3433–3439. https://doi.org/10.2147/OTT.S94300

Zhao, X., Qureshi, F., Eastman, P. S., Manning, W. C., Alexander, C., Robinson, W. H., & Hesterberg, L. K. (2012). Pre-analytical effects of blood sampling and handling in quantitative immunoassays for rheumatoid arthritis. *Journal of Immunological Methods*, *378*(1–2), 72–80. https://doi.org/10.1016/j.jim.2012.02.007

Ziller, M. J., Gu, H., Müller, F., Donaghey, J., Tsai, L. T. Y., Kohlbacher, O., De Jager, P. L., Rosen, E. D., Bennett, D. A., Bernstein, B. E., Gnirke, A., & Meissner, A. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature*, *500*(7463), 477–481. https://doi.org/10.1038/nature12433