Entorhinal cortex EWAS meta-analysis highlights four novel loci showing differential methylation in Alzheimer's disease

Yasmine Sommerer[1], Valerija Dobricic[1], Marcel Schilling[1], Olena Ohlei[1], Sanaz Sedghpour Sabet[2], Tanja Wesse[2], Janina Fuß[2], Sören Franzenburg[2], Andre Franke[2], Laura Parkkinen[3], Christina M. Lill[1,4], Lars Bertram[1,5]

Affiliations:

1. Lübeck Interdisciplinary Platform for Genome Analytics (LIGA), University of Lübeck, Lübeck, Germany
2. Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany
3. Nuffield Department of Clinical Neurosciences, Oxford Parkinson's Disease Centre, University of Oxford, UK
4. Ageing Epidemiology Unit (AGE), School of Public Health, Imperial College London, London, UK
5. Center for Lifespan Changes in Brain and Cognition (LCBC), Department of Psychology, University of Oslo, Oslo, Norway

Corresponding Author:

Prof. Dr. Lars Bertram
Lübeck Interdisciplinary Platform for Genome Analytics (LIGA)
University of Lübeck
Ratzeburger Allee 160, Haus V50, 1st floor, Room 319
23562 Lübeck, Germany
Telephone: +49 451 3101 7490
Email: lars.bertram@uni-luebeck.de

Running title: Alzheimer's disease EWAS in entorhinal cortex

Keywords: Alzheimer's disease; DNAm; brain; entorhinal cortex; meta-analysis; methylation, EWAS; Braak

Abstract

Studies on DNA methylation (DNAm) in Alzheimer's disease (AD) have recently highlighted several genomic loci showing association with disease onset and progression. Here, we conducted an epigenome-wide association study (EWAS) using DNAm profiles in entorhinal cortex (EC) from 149 AD patients and control brains and combined these with two previously published EC datasets by meta-analysis (total n=337). We identified 12 cytosine-phosphate-guanine (CpG) sites showing epigenome-wide significant association with either case-control status or Braak's tau-staging. Four of these CpGs, located in proximity to *TENT5A, PALD1, PRF1,* and *DIRAS1,* were not reported previously. Integrating DNAm levels with RNA sequencing-based mRNA expression data generated in the same individuals showed significant DNAm-mRNA correlations for 6 of the 12 significant CpGs. By calculating rates of epigenetic age acceleration using two recently proposed "epigenetic clock" estimators we found a significant association with accelerated epigenetic aging in AD patients vs. controls. In summary, our study represents the hitherto most comprehensive EWAS in AD using EC and highlights several novel differentially methylated loci with potential effects on gene expression.

Introduction

Alzheimer's disease (AD) is a progressive, neurodegenerative disease that accounts for 50-60% of all dementia cases [1]. The number of AD cases is increasing and it was recently estimated that nearly 44 million individuals lived with dementia in 2016 world-wide [2]. On a neuropathological level, the hallmarks of AD are accumulations of amyloid-beta (Aβ) plaques and neurofibrillary tangles (NFTs) consisting of hyperphosphorylated tau protein. There is growing evidence that the first neuropathological changes already occur two decades or more prior to the onset of clinical symptoms [3]. In early AD, NFTs are regularly observed without the formation of Aβ, with neuropathological changes typically starting in the transentorhinal followed by the entorhinal cortex (EC) before spreading across most cortical brain regions while the disease progresses [4]. Owing to this spatio-temporal course, the EC represents an interesting and informative brain region to study in molecular AD research, including studies aimed at the epigenome or transcriptome.

There is accumulating evidence that epigenetic factors (in addition to genetic factors) may contribute to the onset and progression of AD [5–7]. One of the most widely studied epigenetic marks is DNA methylation (DNAm) owing to the relative technical ease to generate these data on a(n) (epi)genome-wide scale. Since 2014, this has led to a number of epigenome-wide association studies (EWAS) assessing DNAm profiles in various AD-related phenotypes [6,8–16], culminating in a very recent meta-analysis on differential DNAm across various brain datasets [17] . Taken together, these studies identified several genomic loci (e.g., *ANK1, RPL13, SPG7,* and *MCF2L*) showing consistent changes in DNAm patterns associated with AD related phenotypes across several brain regions (e.g., EC as well as temporal and prefrontal cortex).

In this study, we generated DNAm (using the MethylationEPIC microarray) and mRNA (using RNA sequencing) expression profiles in the same EC slices from 65 AD cases and 84 control brains. These data were used to conduct a DNAm-based EWAS using both case-

control status and Braak's tau-staging (henceforth termed "Braak staging") as predictors. For the EWAS part, we combined our DNAm data with data from two previously published EC studies (both generated using the 450K Methylation microarray) [9,17] increasing our total sample size to n=337. Significantly differentially methylated sites were then correlated with corresponding mRNA levels to probe for potential effects of DNAm on gene expression. Our EWAS identified a total of 12 epigenome-wide significant cytosine-phosphate-guanine (CpG) sites, four of which have not been linked to AD previously. Six of these CpGs show significant correlations with corresponding mRNA levels in the same samples, highlighting their potential downstream effects on gene expression.

Material and Methods

Human samples

Snap-frozen, post-mortem human brain tissue from EC slices (Brodmann area BA28) from 91 AD patients and 92 elderly control individuals was obtained from the Oxford Brain Bank. The AD patients and healthy controls were part of the longitudinal, prospective Oxford Project to Investigate Memory and Aging (OPTIMA) using protocols which have been described in detail elsewhere [18]. All subjects underwent a detailed clinical history, physical examination, assessment of cognitive function (Cambridge Examination of Mental Disorders of the Elderly (CAMDEX) [19] with the Cambridge Cognitive Examination (CAMCOG) and Mini-Mental State Examination (MMSE) biannually. The pathological diagnosis of AD was made using the Consortium to Establish a Registry for Alzheimer's disease (CERAD)/National Institutes of Health (NIH) criteria and Braak staging [20–22]. All included patients were of white European descent by self-report. The Ethics Committees of Oxford University and University of Lübeck approved the use of the human tissues for our study and all participants gave informed consent.

DNA and RNA extraction and processing

For 183 EC samples, genomic DNA and total RNA were extracted from approximately 50mg and 25mg of frozen tissue, respectively. DNA extraction was performed using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany), while for total RNA, we used the mirVana miRNA kit (Thermo Fisher Scientific, USA). All extraction steps were carried out according to manufacturer's instructions. DNA and RNA were quantified using a NanoDrop ONE spectrophotometer (Thermo Fisher Scientific). In addition, RNA integrity was assessed using a Bioanalyzer 2100 instrument and the RNA 6000 Nano LabChip kit (Agilent Technologies, USA).

EPIC array profiling

DNAm profiling was performed using the "Infinium MethylationEPIC" array (Illumina, Inc.) on aliquots of DNA extracts diluted to ~50ng/µl concentration. DNA samples were subjected to bisulfite conversions using the EZ DNA Methylation kit (Zymo Reasearch) following the alternative incubation conditions for the Illumina Infinium MethylationEPIC Array from the supplier, and then hybridized to the EPIC array and scanned on an iScan instrument (Illumina, Inc.) according to the manufacturer's instructions (Document#1000000077299v0). All DNA samples were treated in consecutive laboratory experiments to minimize the potential for batch effects. Raw DNAm intensities were called using the iScan control software (v2.3.0.0; Illumina, Inc.) and exported in .idat format for downstream processing and analysis.

DNA methylation data processing and quality control

DNAm data pre-processing and quality control (QC) was performed in R (v. 3.6.1) using the package bigmelon with default settings [23], unless otherwise noted. Idat files were loaded into R and β-values were calculated according to the following formula, with I_{met} being the intensity of the methylated signal, and I_{ume} being the intensity of the unmethylated signal.

(1) $\beta = I_{met} / (I_{met} + I_{ume} + 100)$

Samples were excluded from the analysis if (a) the bisulfite conversion efficiency of the sample according to the *bscon* function in the bigmelon package was below 65%, (b) the sample had a beadcount < 3 in more than 5% of all probes, (c) the sample had a detection p-value below 0.05 in more than 1% of all probes, (d) the sample was identified as an outlier according to the *outlyx* function in the bigmelon package using a threshold of 0.01, (e) the sample showed a large change in β-values after normalization according to the *qual* function in the bigmelon package with a threshold of 0.1, (f) the sample showed a discrepancy between predicted sex according to the Horvath multi-tissue epigenetic age predictor [24] and reported sex, or (g) there was a greater than 70% discrepancy between genotypes of 42 SNPs determined concurrently from the EPIC and GSA SNP genotyping array (see below). All samples were normalized with the *dasen* function of bigmelon.

CpG-probes were removed from the analysis if (a) they had a detection p-value below 0.05, (b) they had a beadcount < 3, (c) they aligned to multiple locations according to Nordlund et al. [25], (d) they aligned to locations influenced by SNPs according to Zhou et al. [26], or (e) they aligned to locations on the X and Y chromosomes.

The final analyses included 665,796 CpG-probes in 149 samples for the AD case-control EWAS and 142 samples for the Braak stage EWAS. A detailed sample description can be found in Table 1.

Epigenome-wide association study (EWAS) analyses to identify differentially methylated probes (DMPs) and differentially methylated regions (DMRs)

Statistical analyses to identify differentially methylated probes (DMPs) were performed based on linear regression models using the *lm* function in R using case-control status (as dichotomous variable) or Braak stage (as continuous variable) as predictor in the EWAS, respectively:

(2) DNAm(case-control) ~ ADstatus + age + sex + DNAm PCs + genetic PCs

(3) DNAm(Braak staging) ~ Braak stage + age + sex + DNAm PCs + genetic PCs

To account for differences in the DNAm profiles due to technical (e.g. laboratory batch, array) and other (e.g. cell-type composition of samples) factors we performed a principal component analysis (PCA; using the R base function *prcomp*) on a subset of uncorrelated CpGs to capture variation in DNAm irrespective of phenotype. To this end, we created a set of uncorrelated CpGs by dividing the genome into 100kb bins and followed by random selection of one CpG-probe per bin. In the full model, 13 such "DNAm PCs" were included as covariates in the EWAS analyses.

To account for differences in genetic ancestry we performed a PCA (using PLINK v1.9 "--pca") on an LD-pruned set of SNP markers (--indep-pairwise 1500 150 0.2) derived from genome-wide SNP genotyping data generated in parallel on the same DNA samples of each individual using the Global Screening Array (GSA; Illumina, Inc). The first 20 genetic ancestry PCs were used as covariates in the EWAS analyses. More details on the genotyping and QC procedures can be found in a previous publication [27].

Differentially methylated regions (DMRs, i.e. combinations of consecutive DMPs) were assessed with the comb-p tool [28] with the maximal gap within a region set to 500 base pairs and the seed P-value set to 1.00E-03. Only regions including at least three CpGs were considered. Significance of the DMR results was determined using the Sidak method (as implemented in comb-p).

Annotation of CpGs to specific gene regions was based on the Illumina manifest (v1.0 B5) for the EPIC array and the GREAT annotation tool [29].

Meta-analysis of epigenome-wide association study (EWAS) results

To increase power of our EWAS, we combined our EPIC array-based results with those from two publicly available AD EC datasets (GEO accession numbers GSE59685 with 58 AD cases and 21 controls; and GSE105109 with 68 AD cases and 28 controls). The descriptions of these datasets can be found in the primary publications [6,9]. Here, we downloaded the

processed DNAm values and repeated EWAS analyses for Braak stage and AD case-control status using the same linear regression models as described above. These regression models included 13 DNAm PCs for the Braak stage analysis of GSE59685, and 15 DNAm PCs for the remaining analyses. The meta-analysis was conducted with a fixed-effect inverse-variance approach using the function *metagen* in the R package "meta" [30].

Alzheimer's disease poly-epigenetic scores

Poly-epigenetic scores (PES) for each individual were calculated based on the test statistics from a Braak stage meta-analysis using two publicly available AD EC datasets (GEO accession numbers GSE59685; GSE105109). The descriptions of these datasets can be found in the primary publications [6,9]. For the PES calculations, test statistics from the meta-analysis with varying P-value thresholds (P < 1, P < 1.00E-04, P < 1.00E-05, P < 1.00E-06, P < 1.36E-07) were used. For each P-value threshold, we created a set of uncorrelated CpGs by dividing the genome into 100kb bins and followed by the selection of the CpG with the most extreme effect size per bin. Here, the effect sizes from the meta-analysis for London-1 and London-2 (β) and the normalized DNAm values (CpG) for each of the n uncorrelated CpGs for each P-value threshold were used and combined as follows:

(4) $\sum_{i=1}^{n} \beta_i * CpG_i$

PES-based association analyses used linear regression models to predict Braak stage as outcome and PES as predictor, adjusting for the same covariates as in the primary EWAS analyses.

Epigenetic age estimation

Two epigenetic age predictors were used in our analyses: 1. the "Horvath multi-tissue predictor" (HMTP) [24] and 2. the "cortex clock" (CorCl) [31]. Since most other popular epigenetic clocks (Hannum [32], PhenoAge [33], GrimAge [34]) were calibrated for blood tissues, we did not include analyses of these age estimators in this study. DNAm age using HMTP was calculated with the R script provided in Horvath [24], using the DNAm raw data (with prior

removal of probes failing QC). Prior to the age estimation, this algorithm uses BMIQ as data normalization method. For CorCl, the normalized data from the processing pipeline described above were used directly in code provided by the authors of ref. [31] (https://github.com/gemmashireby/CorticalClock).

Age acceleration was defined as residual from a linear regression of epigenetic age on chronological age.

(5) $age_{epigenetic} \sim age_{chronological}$

RNA sequencing and quality control

Total RNA was extracted from EC samples as described above. Libraries were generated using the TruSeq Stranded Total RNA kit (Illumina) and sequenced at a target depth of 30 M reads per samples in 2x100 bp paired-end mode on a NovaSeq 6000 device (Illumina). All transcripts annotated in the Ensembl human genes annotation (version 100) were quantified using kallisto [35] using the –bias and –rf-stranded options. The resulting trancriptome profiles were filtered for protein coding and lncRNA genes only and summarized to the gene level (summing up read counts of all isoforms of each gene) and normalized using DESeq2 [36].

DNAm-mRNA correlation analyses

The normalized RNA-seq data for the selected CpG candidate genes (see above) were correlated to their corresponding DNAm signal: Outliers (>1.5 * inter-quartile-range below/above 1st/3rd quartile) were removed and subjected to R's *cor.test* function using the Spearman method. The resulting p-values were corrected for multiple testing using the Benjamini-Hochberg procedure (as implemented in R's *p.adjust* function.) Additionally, linear models were fit (using R's *glm* function with family = "gaussian") to predict the (log-transformed) gene expression estimates from the corresponding DNAm values. The following potential confounders were included as co-variables in these models: age at death, sex, RNA integrity (RIN) and post-mortem interval (PMI). All continuous variables were scaled and

centred before fitting the model. Finally, additional models were fit adding disease status (AD case vs. control) or Braak stage (as a continuous variable), respectively. The corresponding F-statistic was used to assess the significance of the (additional) amount of gene expression variance by the predictors of interest. As for the correlation analysis above, multiple-testing was addressed by false-discovery-rate (FDR) estimation as per Benjamini-Hochberg.

Results

EWAS of case-control status and Braak staging highlights five DMRs

We did not observe any inflation for the AD case-control and Braak stage EWAS ($\lambda_{ADstatus}$ = 1.0005, λ_{Braak} = 1.0175, Supplementary Figure 1) in our dataset. In the EWAS analyses of all 665,796 CpG-probes that passed QC on the EPIC array, none of the CpGs reached the experiment-wide Bonferroni-corrected significance threshold α = 7.51E-08, neither in the analyses of AD case-control status (Supplementary Figure 2) nor Braak stage (Supplementary Figure 3). However, we note that several CpGs reached at least suggestive evidence of association with either phenotype (α = 1.00E-05; Supplementary Tables 1 and 2). Among these, cg25191519 showed a particularly strong association signal (P = 8.90E-06). This CpG is annotated to the genes *SPG7* and *RPL13*, which were already reported in previous AD DNAm studies [8,15–17]. Among the other suggestive association signals were also the following genes which are highlighted here owing to their moderate to high expression levels in brain tissues according to GTEx V8: *MGAT3*, *MIEF1*, *VRK1*, *PAPOLA*, *PLD5*, and *CEP170* (Supplementary Tables 1 and 2). However, we emphasize that more data using the EPIC array or other methods in independent samples are needed to validate these findings.

Next, we used the DMP test statistic to assess the presence of DMRs, i.e., consecutive runs of differentially methylated probes which are aggregated into "regions". After adjustment for multiple testing, five DMRs (near genes *PRKCZ*, *CYFIP1*, *ACOT7*, *COL4A1*, *IBA57*, and *C1orf69*) showed significant association with AD case-control status (Table 2). In contrast, no

DMRs were found when combining DMPs from the Braak stage EWAS using p-comb. Two of the six genes highlighted by the significant case-control DMRs were previously described in the context of AD DNAm profiling studies. First, a DMR near *CYFIP1* was reported by Li et al. to show association with Braak stage [15]. Second, the gene *IBA57* is annotated to CpG-probe cg12461930, which also showed association with Braak stage in the cross-cortex meta-analysis by Smith et al. [17]. In addition to the significant DMR result in *IBA57*, we also observed a moderate, but at best nominally significant, association of cg12461930 with case-control status ($effect_{casectrl}$=-0.0175, $P_{case/ctrl}$=0.07) and Braak stage ($effect_{Braak}$=-0.0060, P_{Braak}=0.03) in our analyses.

To further evaluate the degree of correspondence of our novel EWAS data with those from the recent EWAS meta-analysis by Smith et al. [17], we repeated the Braak stage meta-analysis from two publicly available EC DNAm datasets used in previous AD EWAS ("London-1" and "London-2", with GEO accession numbers GSE59685 and GSE105109, respectively) [9,10], which were also included in the Smith et al. study [17], and used these resulting test statistics with varying P-value thresholds to calculate the PES. We then tested the PES for association with Braak stage in our dataset using linear models equivalent to the primary EWAS. These analyses revealed that the PES was, indeed, significantly associated with Braak stage in our dataset (1.88E-01 ≤ P ≤ 7.83E-06, Supplementary Table 3) and explained up to 12% of the phenotypic variance in our dataset. Overall, these results suggest that our dataset is equivalent in terms of data quality when compared to those previously published in the field.

EWAS meta-analysis highlights 12 DMPs showing experiment-wide significant association with AD

To increase power, we combined the EWAS results from our samples with those from the two publicly available EC DNAm datasets London-1 and London-2. Of note, the DNAm data of these prior studies were generated on the predecessor array (i.e. "450K Methylation Array")

which has a substantially lower resolution leading to a smaller number of meta-analysed CpGs. The meta-analysis across all three datasets comprised 320 samples for the AD case-control analysis, and 337 samples for the Braak stage analysis. Overall, there were 304,996 overlapping CpGs available for this meta-analysis, resulting in an experiment-wide significance threshold of $\alpha = 1.64E-07$ for this arm of our study. Using this threshold, five CpGs in the AD case-control meta-analysis (Figure 1, Table 3), and nine CpGs in the Braak stage meta-analysis (Figure 2, Table 4), reached experiment-wide significance, four of which were not previously reported as AD EWAS signals. Two CpGs (cg03169557 and cg05066959) were significantly associated with both AD case-control status and AD Braak stage and both were already described in previous DNAm AD studies [8,15–17]. While none of the analyses in the individual datasets analyses showed notable inflation in the test statistics (Supplementary Table 4), both meta-analyses displayed slightly increased inflation ($\lambda_{case\text{-}control} = 1.16$; $\lambda_{Braak} = 1.24$; Supplementary Figure 4), a common observation of EWAS as already noted in Smith et al. [17].

The four newly associated CpG-probes are located in or near the genes *CNFN*, *LIPE*, *TENT5A*, *PALD1*, *PRF1*, and *DIRAS1*. The implicated CpGs all displayed a reduction of DNAm associated with AD, i.e., a negative effect size estimate. Of these genes, *DIRAS1* shows the most pronounced expression in brain (GTEx V8) followed by *LIPE*, *TENT5A*, and *PALD1*. In contrast, *CNFN* and *PRF1* do not show any noteworthy expression in the brain tissues analysed in GTEx. Using a PubMed search (using "{gene name} AND alzheimer*} as search terms) revealed that no publication exists to date directly linking any of these genes to AD. Look-up of the CpG IDs on the EWAS Catalog (http://www.ewascatalog.org/) [37] revealed that cg03073402 and cg20648333 were previously reported to be associated with aging from birth to late adolescence in blood samples [38].

The GWAS catalog (https://www.ebi.ac.uk/gwas/ [39]) revealed no noteworthy AD-related entries for *CNFN*, *LIPE*, *PALD1*, *PRF1*, or *DIRAS1*. In contrast, genetic variants in *TENT5A* (a.k.a. *FAM46A*) have been found associated by GWAS with a number of traits (https://www.ebi.ac.uk/gwas/genes/TENT5A), some of them with direct relevance for AD, e.g. “Alzheimer's disease, posterior cortical atrophy”, “Alzheimer's disease, cognitive decline measurement”, “PHF-tau measurement”, “neurofibrillary tangles measurement”, and “temporal pole volume measurement”, underlying the potential mechanistic involvement of this gene in AD pathogenesis. The EWAS signal for *TENT5A* was elicited by CpG-probe cg22388948 (located intronically) with a P-value of 5.83E-08 in the AD case-control meta-analysis. Suggestive evidence for association with AD case-control status with the same effect direction could also be observed in the individual London datasets (P_{L1} = 6.73E-05, $Effect_{L1}$ = -0.0375, P_{L2} = 1.72E-02, $Effect_{L2}$ = -0.0174). While the remaining eight epigenome-wide significant CpGs of our meta-analyses were not featured as “top results” in the Smith et al. paper, they were highlighted in other previous AD EWAS (see Table 3 and 4, Supplementary Table 5).

Interestingly, three of the eight CpG-probes that were previously reported to show experiment-wide significant association with Braak stage in EC [17] and were also present in our analysis did not show any evidence of association with Braak stage ($P < 0.05$) in the Oxford dataset (Supplementary Table 6). This related to CpG-probes cg04523589 (annotated to the gene *CAMP*), cg06653632 (annotated to *SLC15A4* and *TMEM132C)*, and cg11563844 (annotated to *STARD13* and *KL)*. The other five (annotated to genes *SPG7*, *ANK1*, *MIR486*, *MYO1C*, *ABR*, *ALDH16A1*, and *FLT3LG*) showed independent evidence of association in our dataset (P-values ranging from 0.05 to 7.38E-04; Supplementary Table 6) with consistent effect directions and can therefore be regarded as independent replication of the results of Smith et al. [17].

Half of the DNAm EWAS signals correlate with gene expression

To elucidate the potential functional implications of the DNAm associations highlighted above, we performed correlation analyses between the DNAm levels and corresponding mRNA expression data generated in the same individuals from the same tissue slices. To this end, we chose all 12 significantly associated CpGs from the EWAS meta-analysis results (i.e. both DMPs [Table 3, Table 4] and DMRs [Table 2]), as well as the eight available AD-associated CpGs in EC from Smith et al. [17] and correlated the DNAm levels with gene expression levels of the annotated gene(s) according to the Illumina manifest (v1.0 B5) for the EPIC array and the GREAT annotation tool. Within DMR "windows", we selected the CpG showing the strongest association with AD case-control status for correlation with mRNA levels. Overall, this led to Spearman rank correlations of 39 DNAm-mRNA pairs (Supplementary Table 7). Ten of these pairs (with eight unique CpGs, two of which were CpGs from the Smith et al. EC meta-analysis [17]) showed evidence for a statistically significant DNAm-mRNA correlation after multiple testing correction (Table 5) accounting for 39 individual genes (Supplementary Table 8 and 9). In addition, using the genes with significant correlations as outcome in differential gene expression analyses performed on the same RNAseq data revealed that all ten were also significantly differentially expressed with respect to both AD case-control status and Braak stage (Table 5).

Generally, the correlation coefficients only indicated moderate (maximal ρ=-0.35), but notwithstanding statistically significant correlations between DNAm and mRNA expression in this dataset. The comparatively moderate extent of the correlations likely reflects the fact that gene expression is regulated by a number of other (epi-)genetic mechanisms beyond DNAm [40]. Another noteworthy observation is that the signs of correlation coefficients of significant DNAm-mRNA pairs were both positive and negative, suggesting a complex relationship between DNAm status at these positions and their effect on mRNA expression. This is likely

due to the fact that the majority of CpGs (nine out of ten) among the significantly correlated DNAm-mRNA pairs was located in gene bodies or distal to the stop-codon, while only one (CpG cg05228284) was located in an CpG-island (CGI) in the 5' untranslated region (UTR). Classically, DNAm is considered a mark of transcriptional repression (here expected to elicit a correlation with a negative sign), however, this only applies to CpGs located in promoter CGIs, not necessarily those located elsewhere [41]. Therefore, the presence of both positive and negative correlations between DNAm and mRNA levels, as it can be seen in our results (Table 5), is to be expected.

Epigenetic age acceleration is associated with AD in EC

In agreement with prior evidence, both DNAm age estimators were highly correlated with chronological age (HMTP: Pearson's r = 0.56, P=1.078E-13; CorCl: r=0.81, P<2.20E-16) in our dataset. However, CorCl showed a (much) stronger correlation with chronological age compared to HMTP, and also did not show the tendency to under-estimate epigenetic ages compared to chronological age (Figure 3).

Using these estimates, we determined the degree of "age acceleration" which was defined as the residual from a linear regression of DNAm age on chronological age [42]. Age acceleration estimates were moderately correlated between both DNAm age estimators, with a Pearson correlation coefficient of 0.36 (P = 8.05E-06). Next, we tested the association between age acceleration and AD case-control status as well as Braak stage, using a linear regression model including sex as covariate. Our expectation was that samples with an advanced disease state (e.g., AD vs. control, or high Braak stage vs. low Braak stage) would also show a more pronounced age acceleration (i.e., older epigenetic age when compared to chronological age). In concordance with this expectation, we found that age acceleration estimates were, indeed, associated with disease state (1.96E-04 ≤ P ≤ 2.72E-05, Table 6), with higher age acceleration being associated with AD cases or higher Braak stages (Table 6). In an additional model, we

included DNAm PCs and genetic PCs, as they were included in the EWAS, to account for genetic ancestry as well as unknown confounders with respect to the DNAm data, e.g. cell type composition or batch effects. After including these variables as covariates, only the CorCl remained significantly associated with AD case-control status (effect = 0.28, P = 4.70E-03), while some other models still showed suggestive evidence of association of age acceleration with disease state (Supplementary Table 10). Of note, DNAm PC1 and PC2 showed consistent associations with disease state, and therefore likely capture variance in DNAm data that reflect unknown confounders in the dataset (Supplementary Table 10), with the true level of association likely located somewhere between both models with and without DNAm PCs and genetic PCs.

Discussion

In this study we performed various EWAS analyses on DNAm profiles generated in human EC tissue samples. Meta-analysis of our novel DNAm data with those from two previous EC-based EWAS provides evidence for four novel loci showing experiment-wide significant association with either AD case-control status or Braak stage. In addition, we confirm several other previously reported EWAS signals. Using RNAseq data generated from the same individuals/tissue samples, we identified significant correlations between DNAm levels and mRNA expression for 10 out of the 39 DNAm-mRNA pairs within these EWAS loci. The most notable of our novel associations was observed with a CpG-probe (CpG cg22388948) in *TENT5A* (a.k.a. *FAM46A*), which not only showed consistent effect directions across all three analysed datasets, but also exhibited a significant (negative) correlation with mRNA levels of the same gene. *TENT5A* represents a promising novel AD candidate gene due to its previous association with several AD-relevant phenotypes by GWAS. Functionally, it belongs to the nucleotidyltransferase (NTase) fold superfamily (FAM46), which serve as non-canonical poly(A) polymerases involved in the modification of cytosolic and/or nuclear RNA 3' ends

and, hence, in the regulation of gene expression [43]. Regarding the other novel signals of our EWAS, we note that both *LIPE* (encoding lipase E [a.k.a. hormone sensitive lipase {*HSL*}], near CpG cg03073402 on chromosome 19p) and *PRF1* (encoding perforin 1; near CpG cg20648333 on chromosome 10q) are involved in molecular pathways, i.e. lipid metabolism [44] and the immune system response [45], respectively, which have both been shown to be relevant in AD pathogenesis based on recent GWAS data [46,47]. The potential involvement in AD pathogenesis of the other novel loci, i.e. *CNFN*, *PALD1*, and *DIRAS1* is less obvious.

Other relevant outcomes of our study are the independent confirmation of some, albeit not all, previous EC-based EWAS signals [17], and the observation that epigenetic age in EC is accelerated with increasing AD progression using two recently proposed estimators of DNAm age, a finding that is consistent with previously published data [48–50]. We also confirm previous results that the HMTP (a.k.a. "Horvath clock"), which was trained on several tissues, may not be ideal to estimate DNAm age in the human brain cortex, and that, instead, the recently proposed "cortex clock" may be better suited for DNAm analyses in this tissue [31,51].

The strengths of our study are its comparatively large sample size (n = 149 novel EC samples; n = 337 in the EWAS meta-analyses), the analysis of a brain region highly relevant for AD research (i.e. EC), the use of the hitherto highest resolution DNAm profiling microarray (i.e. the Methylation EPIC array featuring 850K CpG-probes), and the parallel availability of RNAseq-based mRNA expression data (allowing for detailed DNAm-mRNA correlation analyses). Despite these strengths, our study is also subject to a number of limitations which include the following. First and foremost, we used "bulk tissue" for DNAm profiling and RNAseq. Bulk tissue samples represent an agglomerate of different cell-types whose proportions (and DNAm and mRNA profiles) may vary across different samples (e.g. they may change as the disease progresses), a situation that may have affected the outcomes of our study. Single nucleus-based DNAm profiling and RNAseq would be required to fully address this question, which was not feasible in the context of our study. The next-best approach is to

estimate sample-specific cell-type proportions using cellular deconvolution (ideally based on reference data from the same tissue with known cell-type composition) and include these estimates into the analyses. To this end, we estimated cell-type composition based on DNAm reference data obtained from human prefrontal cortex samples (to the best of our knowledge, no DNAm reference data exist for EC) using the *estimateCellCounts* function in the R package Minfi [52] and regressed these out with the *removeBatchEffect* function in the R package limma [53]. Using this approach did not appreciably change our results and none of our top signals changed in the AD case-control or Braak stage EWAS (Supplementary Tables 1 and 2). Second, our reanalysis of publicly available DNAm profiles from EC resulted in slightly increased (i.e. less significant) P-values when compared to those reported in the primary publications [17]. This is likely due to a more conservative covariate adjustment scheme used in our data processing workflow. Repeating all meta-analyses based on a covariate adjustment paradigm similar to that used in Smith et al. [17] (i.e. include DNAm covariates until λ drops to <1.2 in each individual dataset) led to an overall increase in statistical significance of most of the meta-analysis results (Supplementary Table 11), approaching those previously reported. We further note that Smith et al. [17] did not adjust their results for genetic ancestry, which may have also affected their test statistics. Therefore, the "true" level of statistical support likely lies somewhere in between both approaches, and suggests that the EWAS results reported here are likely conservative. Third, our meta-analysis is based on fewer CpGs (n = 304,996) as reported by the overlap of the London-1 and London-2 datasets (up to n = 403,763 [6,9]). This is due to the fact that we used a different DNAm microarray ("MethylationEPIC") here, which does not show perfect overlap in CpG-probes with the predecessor array ("450K") and removed a comparatively large number of probes from all datasets during QC. Fourth, as previously noted [17], EWAS meta-analysis tend to show inflations of the test statistics, which we also observed here. One possible reason is heterogeneity of the effect estimates across studies due to technical reasons. To address this issue, we repeated the meta-analysis using random-effect models, which indeed showed a much

less pronounced degree of inflation than fixed-effect models (Case-control: $\lambda_{fixed\text{-}effect} = 1.16$, $\lambda_{random\text{-}effect} = 0.92$; Braak stage: $\lambda_{fixed\text{-}effect} = 1.24$, $\lambda_{random\text{-}effect} = 1.00$; Supplementary Figure 5). However, we note that for only two out of the twelve experiment-wide significant DMPs, heterogeneity measured by the i^2 statistic exceeded 50% (Supplementary Table 12), supporting the appropriateness of using the fixed-effect models, as was done in other recent EWAS meta-analyses [17]. Fifth, we are not able to discern cause-effect relationships from our data. The observed epigenetic associations may reflect molecular changes preceding (and perhaps modifying) the underlying disease process, but they may also represent a *consequence* of the same. However, the inability to disentangle the sequence of events is not a limitation of our EWAS, but of any epigenetic study. Notwithstanding, our novel results imply several new loci and potential molecular mechanisms that are associated with AD and, as a result, provide important new insights to our understanding of the pathogenetic mechanisms underlying the disease process. Lastly, despite going to great lengths to minimize spurious associations by applying conservative QC thresholds in all steps of our data processing steps, we cannot exclude the possibility that some undetected confounding factors have impacted our results. However, the overall high concordance between our and previous EWAS results argues against a strong systematic bias specific to our data and/or results. Nevertheless, all novel results to emerge from this work should be considered preliminary until independent validation is reported.

In summary, in the largest AD EWAS performed on EC samples to date, we identified a total of 12 epigenome-wide significant CpGs, four of which are novel. Six of these CpGs show significant correlations with corresponding mRNA levels in the same samples, highlighting their potential downstream effects on gene expression. Future work is needed to validate our findings and to clarify the role of the newly implied loci in AD pathogenesis.

Author contributions

Design of the study, supervision, and acquisition of funding: LB. Ascertainment of brain tissue and neuropathological examinations: LP. Handling of tissue samples and generation of molecular data: YS, VD, TW, SSS, AF, JF, SF. Data processing and statistical analyses: YS, MS, OO, CML. First draft of the manuscript: YS, LB. Critical revision and final version of manuscript: all authors.

Acknowledgements

This work was supported by the Cure Alzheimer's Fund (as part of the "CIRCUITS" consortium) to L.B., Deutsche Forschungsgemeinschaft (DFG) and the National Science Foundation China (NSFC) as part of the Joint Sino-German research project ("MiRNet-AD", #391523883) to L.B.. Additional support was provided by the DFG Research Infrastructure NGS_CC (project 407495230) as part of the Next Generation Sequencing Competence Network (#423957469). NGS analyses were carried out at the Competence Centre for Genomic Analysis (Kiel). We acknowledge the Oxford Brain Bank, supported by the Medical Research Council (MRC), Brains for Dementia Research (BDR) (Alzheimer Society and Alzheimer Research UK), Autistica UK and the NIHR Oxford Biomedical Research Centre. Last but not least, we acknowledge the high-performance compute environment ("OmicsCluster") at University of Lübeck where most data processing and analysis steps of this study were run.

Competing Interest Statement

The authors declare no conflict of interest.

References

1. Blennow, K., de Leon, M. J. & Zetterberg, H. Alzheimer's disease. *Lancet* vol. 368 387–403 (2006).
2. Nichols, E. *et al.* Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* **18**, 88–106 (2019).
3. Jack, C. R. *et al.* Hypothetical model of dynamic biomarkers of the Alzheimer's

pathological cascade. *The Lancet Neurology* vol. 9 119–128 (2010).

4. Braak, H. & Braak, E. Frequency of stages of Alzheimer-related lesions in different age categories. *Neurobiol. Aging* **18**, 351–357 (1997).

5. Bertram, L. & Tanzi, R. E. Genomic mechanisms in Alzheimer's disease. *Brain Pathol.* **30**, 966–977 (2020).

6. Smith, A. R. *et al.* Parallel profiling of DNA methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in Alzheimer's disease. *Clin. Epigenetics* **11**, 1–13 (2019).

7. van den Hove, D. L. A., Riemens, R. J. M., Koulousakis, P. & Pishva, E. Epigenome-wide association studies in Alzheimer's disease; achievements and challenges. *Brain Pathol.* **30**, 978–983 (2020).

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

9. Lunnon, K. *et al.* Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat. Neurosci.* **17**, 1164–1170 (2014).

10. Smith, R. G. *et al.* 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 (2018).

11. Gasparoni, G. *et al.* DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. *Epigenetics and Chromatin* **11**, 1–19 (2018).

12. Lardenoije, R. *et al.* Alzheimer's disease-associated (hydroxy)methylomic changes in the brain and blood. *Clin. Epigenetics* **11**, 1–15 (2019).

13. Altuna, M. *et al.* DNA methylation signature of human hippocampus in Alzheimer's disease is linked to neurogenesis. *Clin. Epigenetics* **11**, 1–16 (2019).

14. Semick, S. A. *et al.* Integrated DNA methylation and gene expression profiling across multiple brain regions implicate novel genes in Alzheimer's disease. *Acta Neuropathol.* **137**, 557–569 (2019).

15. Li, Q. S., Sun, Y. & Wang, T. Epigenome-wide association study of Alzheimer's disease replicates 22 differentially methylated positions and 30 differentially methylated regions. *Clin. Epigenetics* **12**, 1–14 (2020).

16. Pellegrini, C. *et al.* A Meta-Analysis of Brain DNA Methylation Across Sex, Age, and Alzheimer's Disease Points for Accelerated Epigenetic Aging in Neurodegeneration. *Front. Aging Neurosci.* **13**, 1–21 (2021).

17. Smith, R. G. *et al.* A meta-analysis of epigenome-wide association studies in Alzheimer's disease highlights novel differentially methylated loci across cortex. *Nat. Commun.* **12**, 3517 (2021).

18. Dobricic, V. *et al.* Differential microRNA expression analyses across two brain regions in Alzheimer's disease. *bioRxiv* 2021.05.31.446406 (2021) doi:10.1101/2021.05.31.446406.

19. Roth, M. *et al.* CAMDEX. A standardised instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. *Br. J.*

Psychiatry **149**, 698–709 (1986).

20. Mirra, S. S. *et al.* The consortium to establish a registry for Alzheimer's disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* **41**, 479–486 (1991).

21. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica* vol. 82 239–259 (1991).

22. Hyman, B. T. *et al.* National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimer's Dement.* **8**, 1–13 (2012).

23. Gorrie-Stone, T. J. *et al.* Bigmelon: Tools for analysing large DNA methylation datasets. *Bioinformatics* **35**, 981–986 (2019).

24. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **16**, 96 (2015).

25. Nordlund, J. *et al.* Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol.* **14**, (2013).

26. Zhou, W., Laird, P. W. & Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res.* **45**, e22 (2017).

27. Hong, S. *et al.* Genome-wide association study of Alzheimer's disease CSF biomarkers in the EMIF-AD Multimodal Biomarker Discovery dataset. *Transl. Psychiatry* **10**, 41 (2020).

28. Pedersen, B. S., Schwartz, D. A., Yang, I. V. & Kechris, K. J. Comb-p: Software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics* **28**, 2986–2988 (2012).

29. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).

30. Balduzzi, S., Rücker, G. & Schwarzer, G. How to perform a meta-analysis with R: A practical tutorial. *Evid. Based. Ment. Health* **22**, 153–160 (2019).

31. Shireby, G. L. *et al.* Recalibrating the epigenetic clock: implications for assessing biological age in the human cortex. *Brain* **143**, 3763–3775 (2020).

32. Hannum, G. *et al.* Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Mol Cell.* **49**, 359–367 (2013).

33. Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *bioRxiv* **10**, 573–591 (2018).

34. Lu, A. T. *et al.* DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany. NY).* **11**, 303–327 (2019).

35. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

36. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

37. Battram, T. *et al.* The EWAS Catalog: a database of epigenome-wide association studies. *OSF Prepr.* 4 (2021) doi:10.31219/osf.io/837wn.

38. Mulder, R. H. *et al.* Epigenome-wide change and variation in DNA methylation in childhood: trajectories from birth to late adolescence. *Hum. Mol. Genet.* **30**, 119–134 (2021).

39. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).

40. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–329 (2015).

41. Jones, P. A. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nature Reviews Genetics* vol. 13 484–492 (2012).

42. Horvath, S. *et al.* The cerebellum ages slowly according to the epigenetic clock. *Aging (Albany. NY).* **7**, 294–306 (2015).

43. Kuchta, K. *et al.* FAM46 proteins are novel eukaryotic non-canonical poly(A) polymerases. *Nucleic Acids Res.* **44**, 3534–3548 (2016).

44. Bolsoni-Lopes, A. & Alonso-Vale, M. I. C. Lipolysis and lipases in white adipose tissue – An update. *Archives of Endocrinology and Metabolism* vol. 59 335–342 (2015).

45. Voskoboinik, I. & Trapani, J. A. Perforinopathy: A spectrum of human immune disease caused by defective perforin delivery or function. *Front. Immunol.* **4**, 441 (2013).

46. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).

47. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).

48. Grodstein, F. *et al.* The association of epigenetic clocks in brain tissue with brain pathologies and common aging phenotypes. *Neurobiol. Dis.* **157**, 105428 (2021).

49. Marioni, R. E. *et al.* The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *Int. J. Epidemiol.* **44**, 1388–1396 (2015).

50. Levine, M. E., Lu, A. T., Bennett, D. A. & Horvath, S. Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. *Aging (Albany. NY).* **7**, 1198–1211 (2015).

51. Grodstein, F. *et al.* Characteristics of Epigenetic Clocks Across Blood and Brain Tissue in Older Women and Men. *Front. Neurosci.* **14**, 1402 (2021).

52. Aryee, M. J. *et al.* Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).

53. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

Tables and Figures

Table 1: Demographic data for the entorhinal cortex datasets by case-control status (n=149) and Braak staging (n=142) after QC.

AD Status / Braak Stage	Sample size (% AD)	% Females	Age / years
Control	65	40	80 ± 13
Case	84	51	82 ± 8
Braak stage 0	3 (0)	33	66 ± 4
Braak stage I	8 (0)	25	74 ± 11
Braak stage I/II	27 (0)	37	85 ± 8
Braak stage II	10 (0)	40	81 ± 11
Braak stage III	2 (0)	100	97 ± 5
Braak stage III/IV	8 (13)	50	87 ± 3
Braak stage IV	6 (83)	17	88 ± 4
Braak stage V	24 (100)	29	86 ± 6
Braak stage V/VI	24 (100)	67	81 ± 9
Braak stage VI	30 (100)	60	88 ± 7

Table 2: Results of EWAS using AD case-control status in Oxford sample.

Position	Gene	CpGs	P-value	P_{adj}	Previous studies (ref.)†
chr1:2004968-2005180	*PRKCZ*	3	4.50E-11	1.41E-07	
chr15:22921227-22921426	*CYFIP1*	3	4.75E-10	1.59E-06	DMR in *CYFIP1* [15]
chr1:6445901-6445975	*ACOT7*	3	3.80E-10	3.42E-06	
chr13:110918122-110918331	*COL4A1*	3	1.82E-09	5.80E-06	
chr1:228362233-228362309	*IBA57 / C1orf69*	3	6.03E-08	5.30E-04	*IBA57* previously described [17]

Table 2 legend: AD case-control DMRs identified using p-comb software. CpGs: number of CpGs in each DMR; P_{adj}: P-value after adjustment for multiple testing using false-discovery rate (FDR) adjustment based on the Sidak method. † Evidence for implication of same or largely overlapping locus from studies using DNAm assessments in AD-related phenotypes. More (P<1.00E-05) results from this analysis can be

found in Supplementary Table 1 and 2.

Table 3: Results of EWAS meta-analysis using AD case-control status.

CpG	Position	Genes	P-value	Effect (β)	Previous Studies (ref.)†
cg03169557	chr16:89598950	*RPL13, SPG7*	6.59E-09	0.0214	[8,15–17]
cg05066959	chr8:41519308	*NKX6-3, ANK1, MIR486*	1.30E-08	0.0453	[6,8,15–17]
cg03073402	chr19:42927676	*CNFN, LIPE*	5.59E-08	-0.0086	-
cg22388948	chr6:82460558	*TENT5A*	5.83E-08	-0.0270	-
cg20648333	chr10:72298745	*PALD1, PRF1*	1.44E-07	-0.0183	-

Table 3 legend: Experiment-wide significant CpGs (P<1.64E-07) in the meta-EWAS ascross three datasets (London-1, London-2, Oxford) using AD case-control status.† Evidence for implication of same or largely overlapping locus from studies using DNAm assessments in AD-related phenotypes.

Table 4: Results of EWAS meta-analysis using Braak staging.

CpG	Position	Genes	P-value	Effect (β)	Previous Studies (ref.)†
cg03169557	chr16:89598950	*RPL13, SPG7*	1.03E-09	0.0042	[8,15–17]
cg05066959	chr8:41519308	*NKX6-3, ANK1, MIR486*	2.76E-09	0.0092	[6,8,15–17]
cg20618448	chr19:49962324	*FLT3LG, ALDH16A1*	1.42E-08	0.0043	[17]
cg05030077	chr16:2255199	*MLST8*	5.89E-08	-0.0026	[9]
cg05228284	chr19:2720847	*DIRAS1*	9.25E-08	-0.0023	-
cg05972352	chr13:113663373	*MCF2L, F2*	1.20E-07	0.0058	genes previously reported [8,11,15,17]
cg14761246	chr3:182968758	*MCF2L2, B3GNT5*	1.55E-07	0.0039	[16]
cg22090150	chr17:4098227	*ANKFY1, CYB5D2*	1.57E-07	0.0048	[13,16,17]
cg07571519	chr10:73472315	*C10orf105, SLC29A3*	1.62E-07	0.0050	[16]

Table 4 legend: Experiment-wide significant CpGs (P<1.64E-07) in the meta-EWAS ascross three datasets (London-1, London-2, Oxford) using

Braak stage. † Evidence for implication of same or largely overlapping locus from studies using DNAm assessments in AD-related phenotypes.

Table 5: Experiment-wide significant Spearman rank correlations between CpG DNAm and gene expression levels.

CpG	Gene	Spearman's ρ	$P_{Spearman}$	$\beta_{Case / Control}$	$P_{Case / Control}$	β_{Braak}	P_{Braak}	Location of CpG
cg22090150	*CYB5D2*	-0.35	3.9E-04	-0.28	0.000609	-0.56	0.000512	Intergenic
cg04520340	*PRKCZ*	-0.28	5.59E-03	-0.49	2.83E-11	-0.98	2.83E-11	gene body
cg22388948	*TENT5A*	-0.28	5.59E-03	0.42	1.67E-07	0.79	6.69E-07	gene body
cg05417607	*MYO1C*	0.26	1.17E-02	0.49	2.93E-12	0.93	2.83E-11	gene body
cg22090150	*ANKFY1*	0.24	1.76E-02	0.46	4.68E-09	0.88	2.03E-08	gene body
cg05228284	*DIRAS1*	0.24	1.76E-02	-0.51	7.54E-12	-1.10	5.85E-13	5'-UTR CGI
cg05066959	*ANK1*	-0.23	2.28E-02	-0.39	6.07E-07	-0.76	1.30E-06	gene body island
cg20648333	*PALD1*	0.22	2.97E-02	0.17	0.025071	0.35	0.015889	gene body
cg20648333	*PRF1*	0.22	3.03E-02	0.36	3.90E-06	0.60	0.000114	Intergenic
cg20618448	*FLT3LG*	0.20	4.29E-02	0.36	1.67E-07	0.75	3.64E-08	Intergenic

Table 5 legend: Experiment-wide significant Spearman rank correlations between CpG DNAm and gene expression levels. β: Effect sizes of gene expression association with AD case-control status / Braak stage; P: P-value of gene expression association with AD case-control status / Braak stage after multiple testing adjustment with the Benjamini Hochberg method; Location: Location of the CpG in the genome with respect to the correlated gene; Results of all 39 tested DNAm-mRNA pairs can be found in Supplementary Tables 6, 7, and 8.

Table 6: Results of linear regression analyses testing the association between epigenetic age acceleration and AD using the HMTP and CorCL DNAm age estimators.

Model	Effect (β)	P
AD Status ~ $ageaccel_{HMTP}$ + age + sex	0.10	5.74E-05
Braak stage ~ $ageaccel_{HMTP}$ + age + sex	0.07	1.96E-04
AD Status ~ $ageaccel_{CorCL}$ + age + sex	0.17	2.72E-05
Braak stage ~ $ageaccel_{CorCL}$ + age + sex	0.11	7.51E-05

Table 6 legend: *β: Effect sizes of AD case-control status / Braak stage with HMTP / CorCL age acceleration*

Figure 1: Manhattan plot for the EWAS meta-analysis using AD case-control status across three EC datasets.

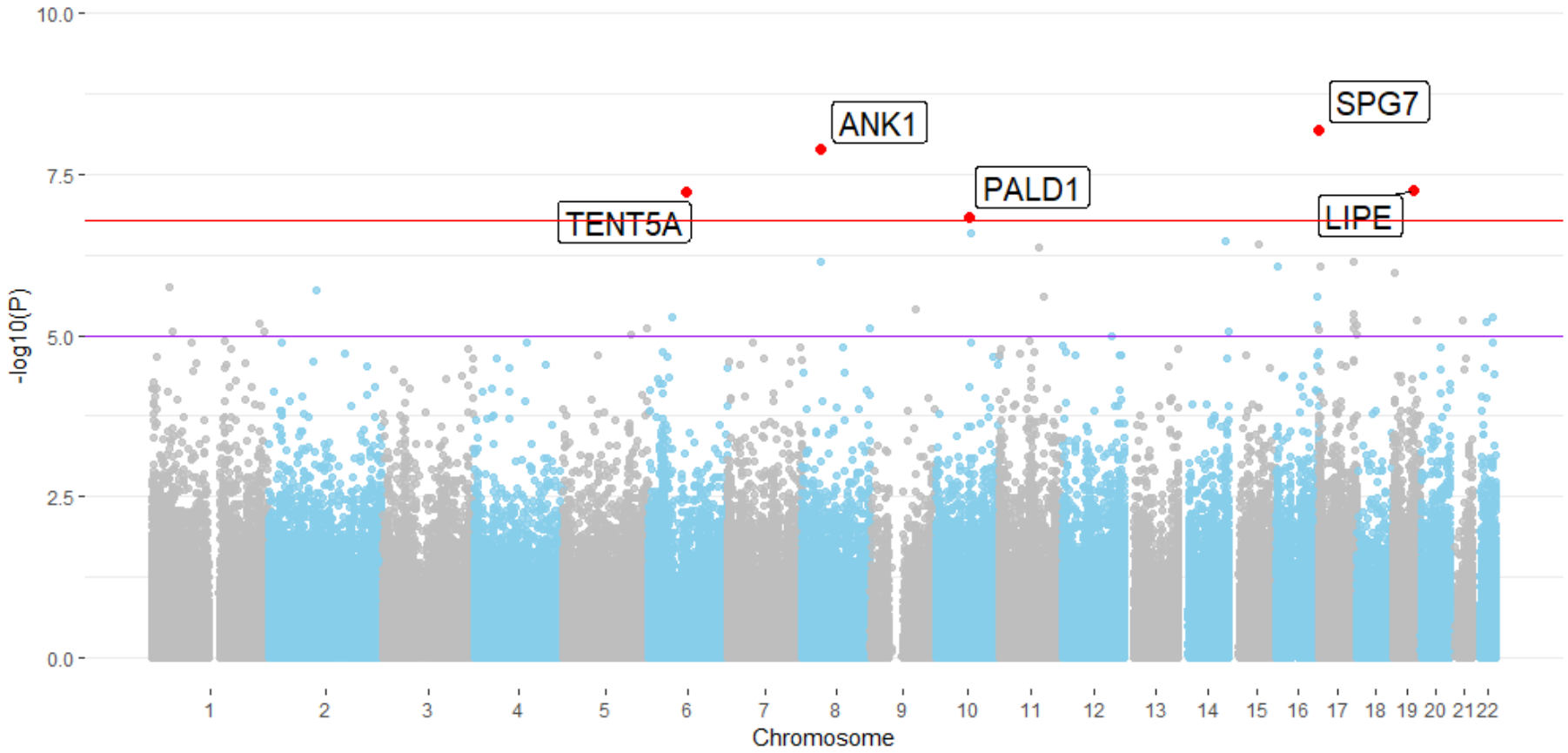


Figure 1 legend: the red line indicates the experiment-wide significance threshold of 1.64E-07, whereas the purple line indicates the suggestive significance threshold of 1.00E-05. CpGs with experiment-wide significant association are marked in red and annotated with the gene name according to the Illumina manifest (v1.0 B5). CpGs around *LIPE, TENT5A, and PALD1* are novel.

Figure 2: Manhattan plot for the EWAS meta-analysis using Braak staging across three EC datasets.

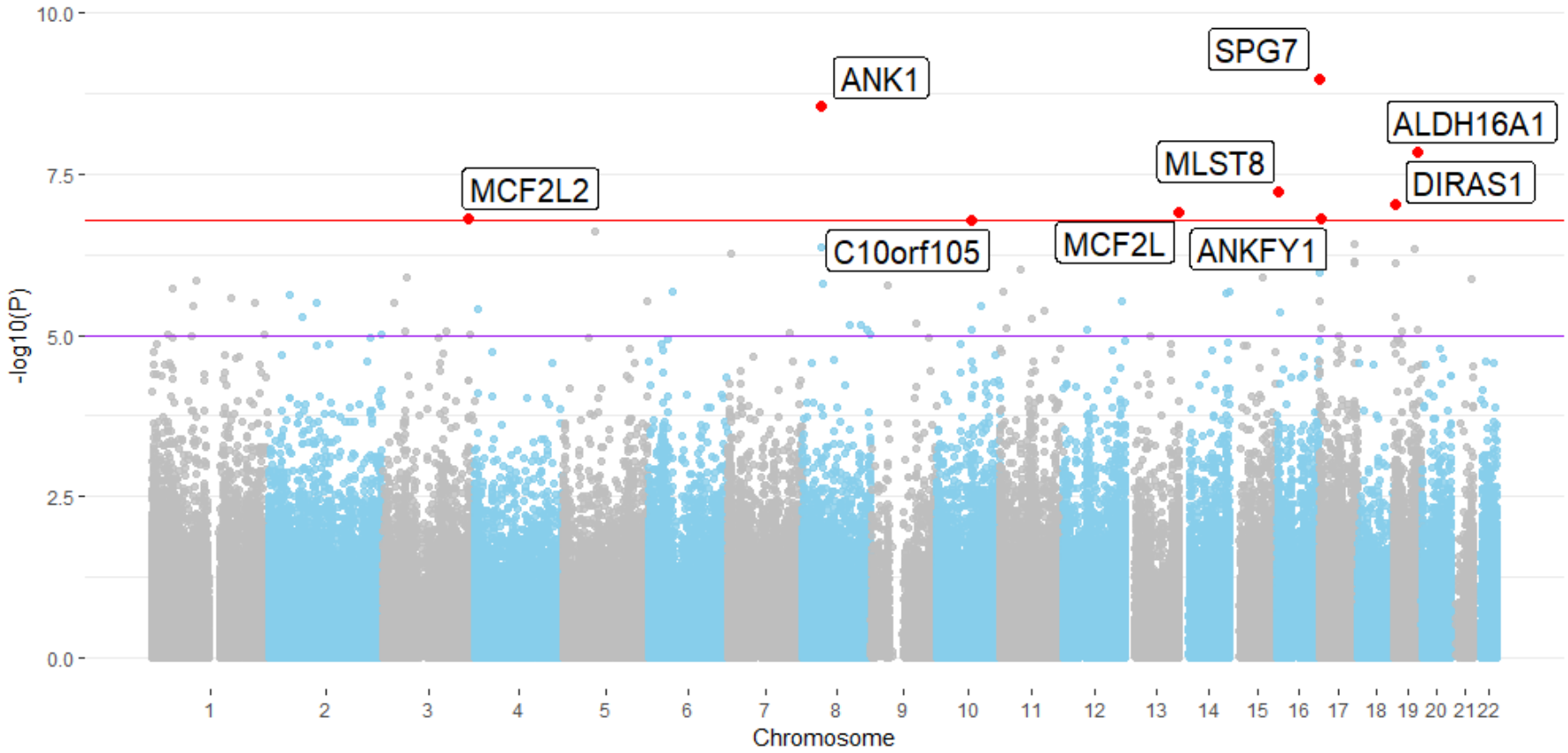


Figure 2 legend: the red line indicates the experiment-wide significance threshold of 1.64E-07, whereas the purple line indicates the suggestive significance threshold of 1.00E-05. CpGs with experiment-wide significant association are marked in red and annotated with the gene name according to the Illumina manifest (v1.0 B5). CpG around *DIRAS1* is novel.

Figure 3: Scatter plots depicting the relation between chronological and epigenetic age colored according to AD case-control status.

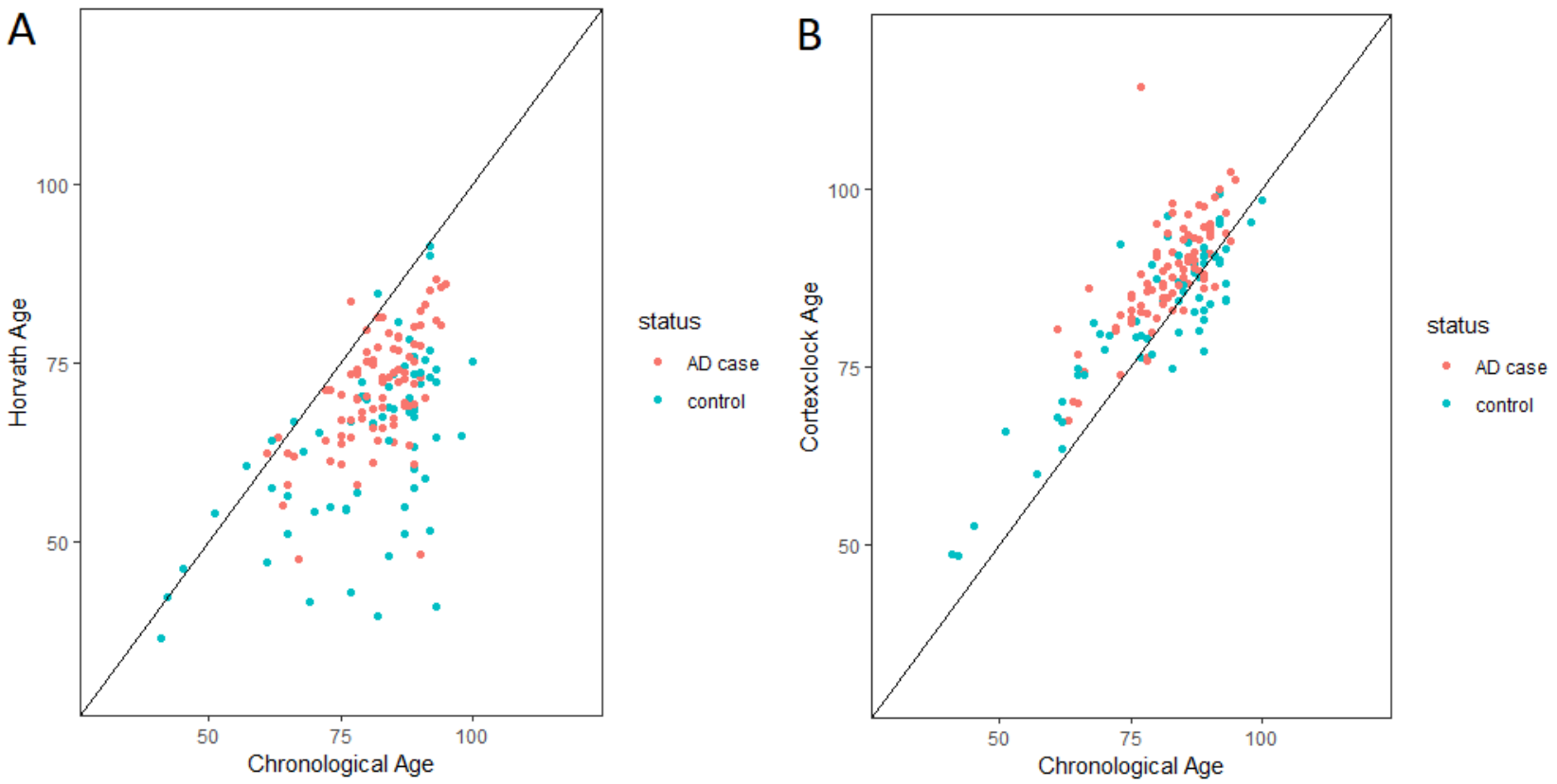
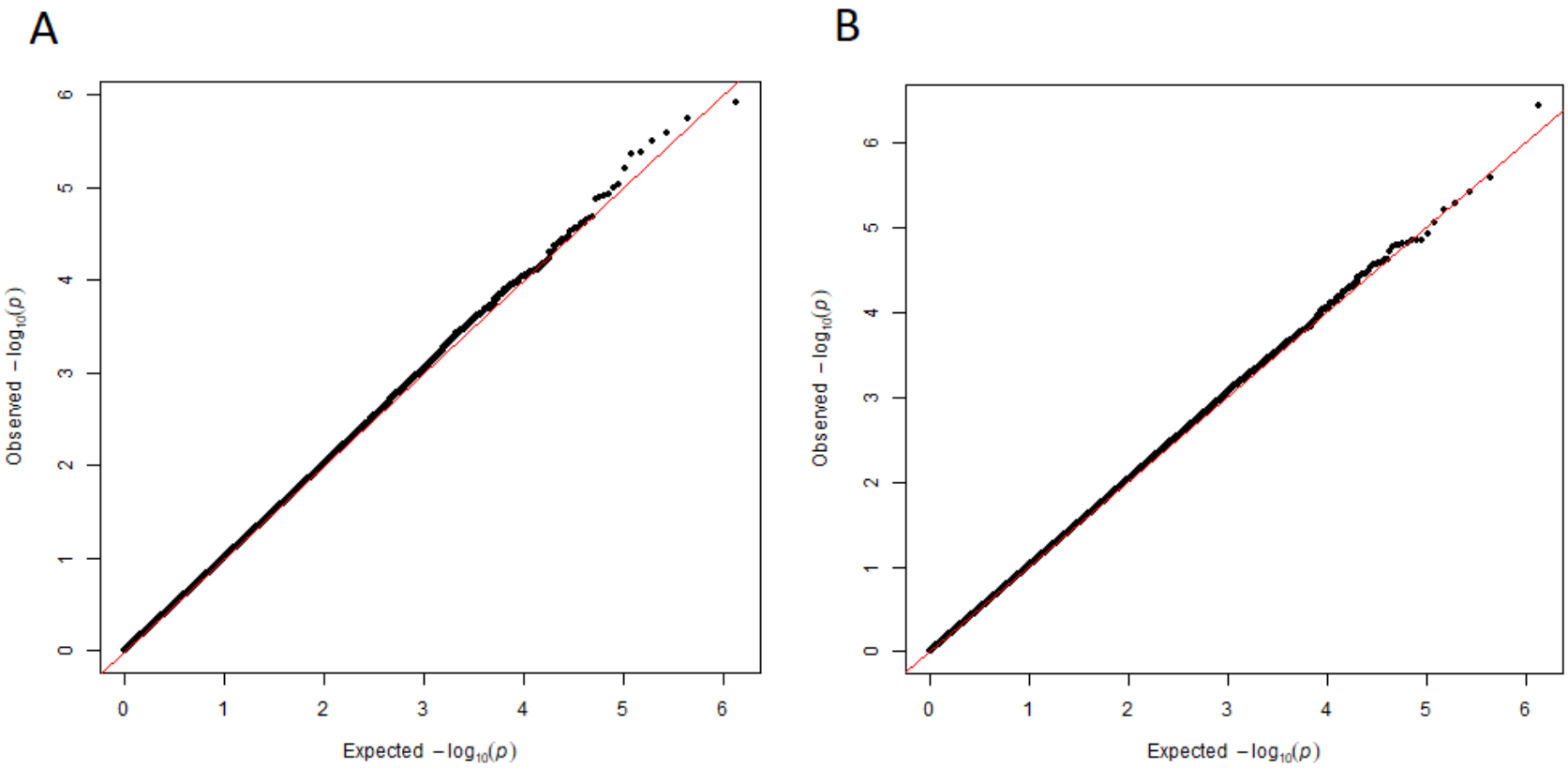
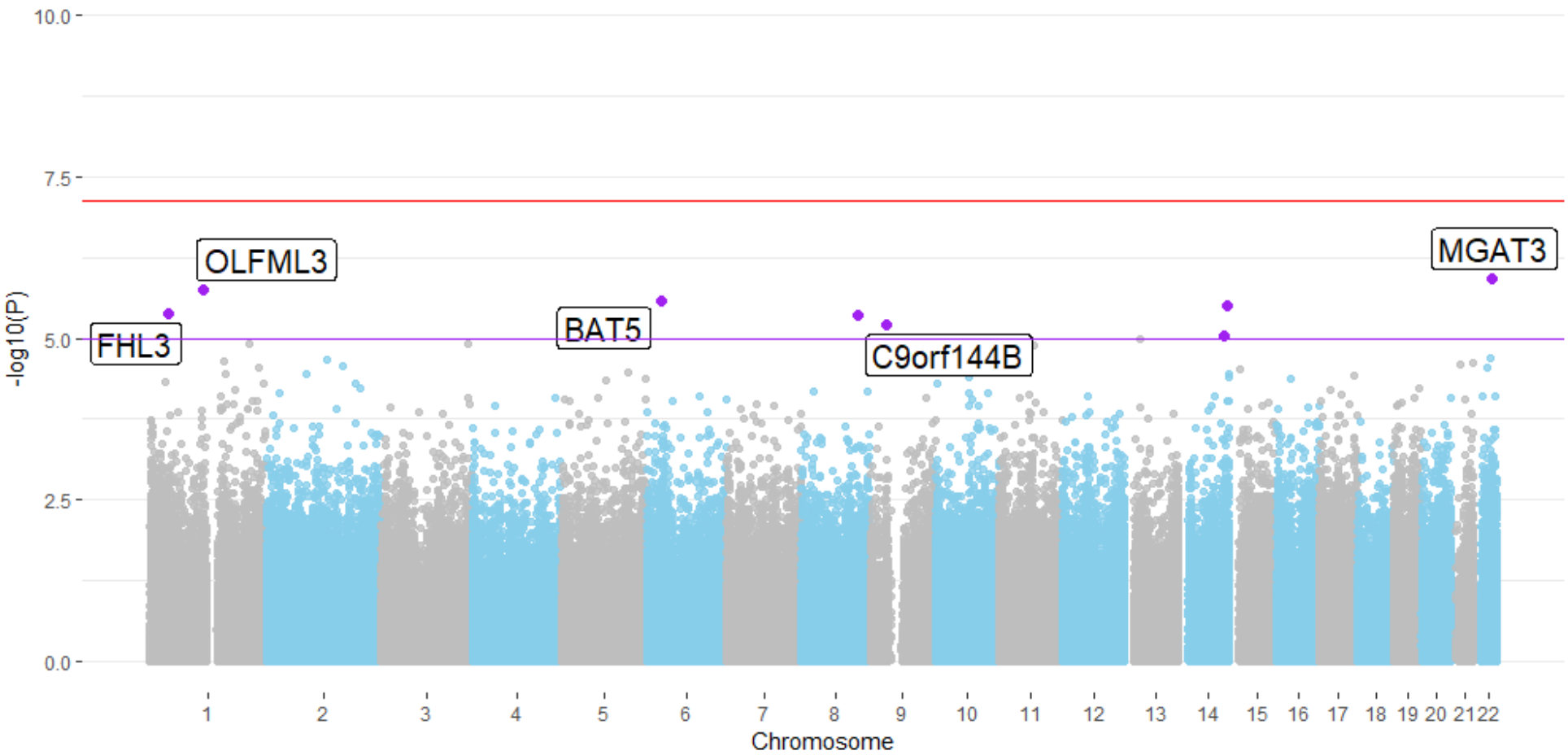


Figure 3 legend: Panel A. Horvath multi-tissue age predictor (left); panel B. cortex clock (right); red: AD case; blue: control.

Supplementary Figures

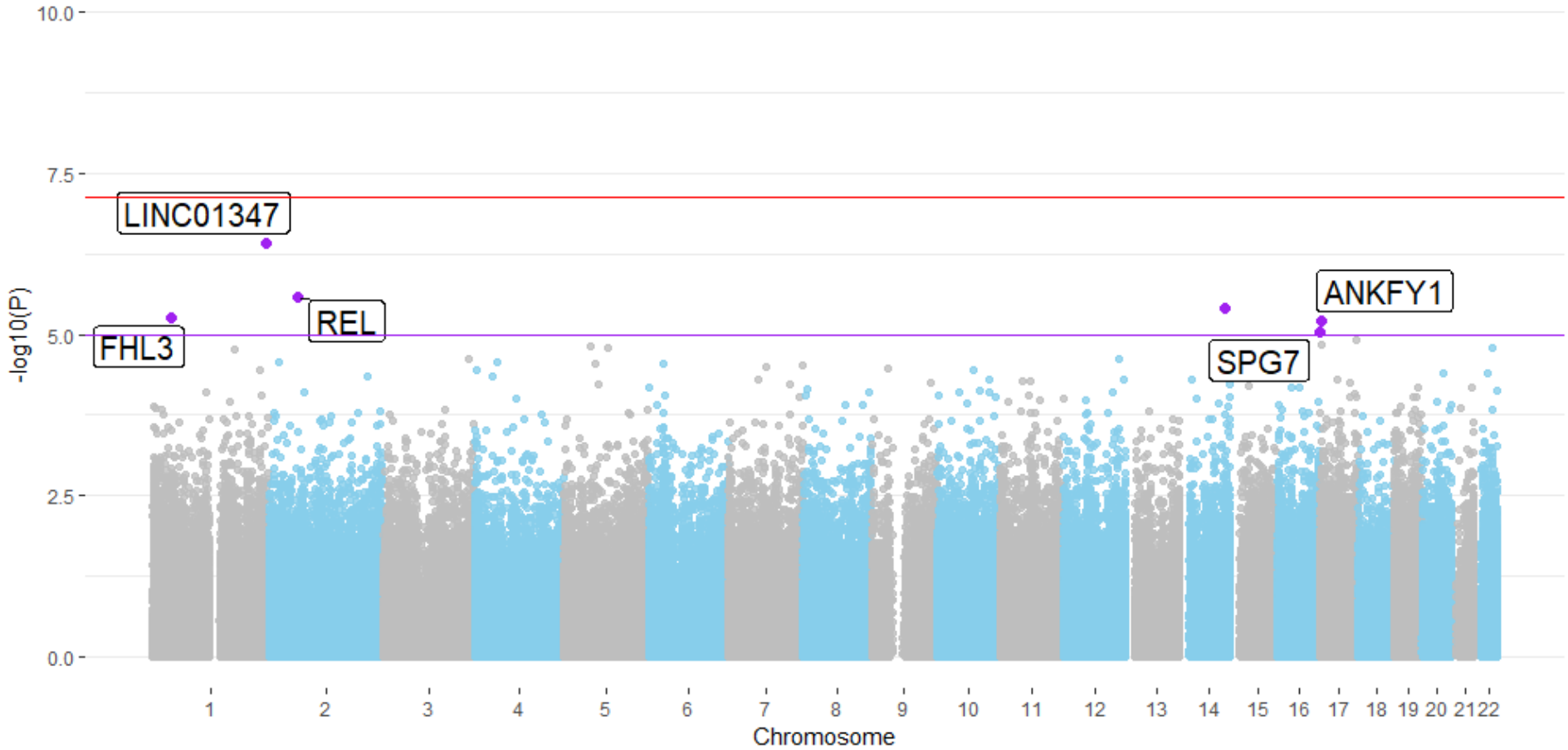


S1: QQ-plots for the results of the AD case-control EWAS (A, λ=1.0005), and the Braak stage EWAS (B, λ=1.0175) in the Oxford datasets.

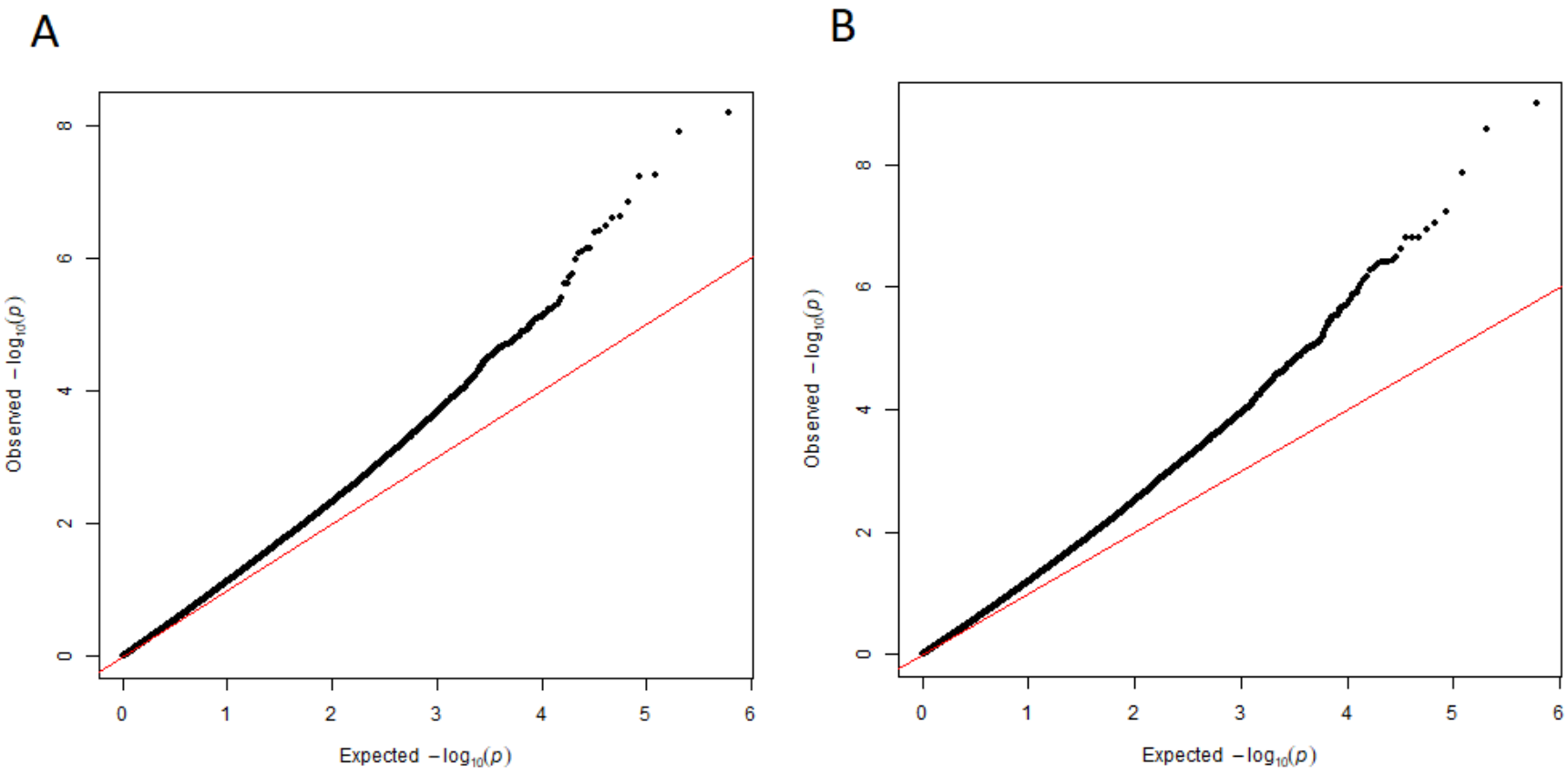


S2: Manhattan plot for the results of the AD case-control EWAS in the Oxford dataset; the red line indicates the experiment-wide significance threshold of 7.51E-08, whereas the purple line indicates the suggestive significance threshold of 1.00E-05. CpGs with suggestively significant association are marked in purple and annotated with the gene name according to the Illumina manifest (v1.0 B5). NB that three CpGs (on chr. 8

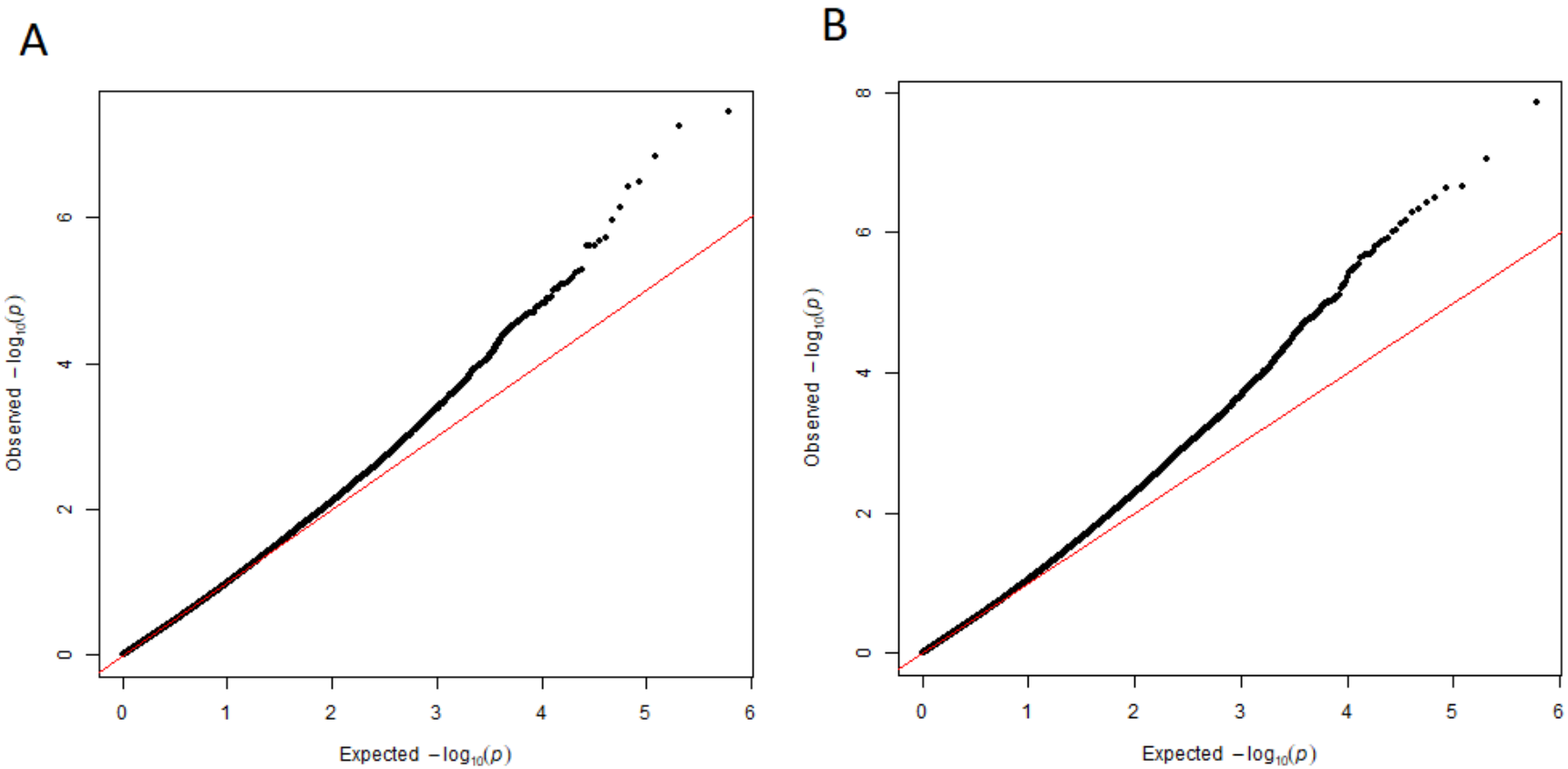
and chr. 14 were not annotated to any genes).



S3: Manhattan plot for the results of the AD Braak stage EWAS in the Oxford dataset; the red line indicates the experiment-wide significance threshold of 7.51E-8, whereas the purple line indicates the suggestive significance threshold of 1.00E-05. CpGs with suggestively significant association are marked in purple and annotated with the gene name according to the Illumina manifest (v1.0 B5). NB that one CpG (on chr. 14 was not annotated to any genes).



S4: QQ-plots for the results of the AD case-control meta-EWAS (A, λ=1.16), and the Braak stage meta-EWAS (B, λ =1.24) with fixed-effect models across three datasets (London-1, London-2, Oxford).



S5: QQ-plots for the results of the AD case-control meta-EWAS (A, λ=0.92), and the Braak stage meta-EWAS (B, λ =1.00) with random-effect models across three datasets (London-1, London-2, Oxford).