ARTICLE

Meta-analysis of epigenome-wide association studies in Alzheimer's disease highlights 220 differentially methylated loci across cortex

Rebecca G. Smith[1,ϕ], Ehsan Pishva[1,2,ϕ], Gemma Shireby[1], Adam R. Smith[1], Janou A.Y. Roubroeks[1,2], Eilis Hannon[1], Gregory Wheildon[1], Diego Mastroeni[3], Gilles Gasparoni[4], Matthias Riemenschneider[5], Armin Giese[6], Andrew J. Sharp[7], Leonard Schalkwyk[8], Vahram Haroutunian[9,10,11], Wolfgang Viechtbauer[2], Daniel L.A. van den Hove[2,12], Michael Weedon[1], Jörn Walter[4], Paul D. Coleman[3], David A. Bennett[13], Philip L. De Jager[14,15], Jonathan Mill[1], Katie Lunnon[1,*]

[1] University of Exeter Medical School, College of Medicine and Health, University of Exeter, Exeter, UK.

[2] Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNS), Maastricht University, Maastricht, The Netherlands.

[3] Banner ASU Neurodegenerative Research Center, Biodesign Institute, Arizona State University, Tempe, Arizona, USA.

[4] Department of Genetics, University of Saarland (UdS), Saarbruecken, Germany

[5] Department of Psychiatry and Psychotherapy, Saarland University Hospital (UKS), Homburg, Germany

[6] Center for Neuropathology and Prion Research, Ludwig-Maximilians-University (LMU), Munich, Germany

[7] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, USA.

[8] School of Biological Sciences, University of Essex, Colchester, UK.

[9] Department of Psychiatry, The Icahn School of Medicine at Mount Sinai, New York, USA.

[10] Department of Neuroscience, The Icahn School of Medicine at Mount Sinai, New York, USA.

[11] JJ Peters VA Medical Center, Bronx, New York, USA.

[12] Laboratory of Translational Neuroscience, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Würzburg, Germany.

[13] Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA

[14] Center for Translational and Computational Neuroimmunology, Department of Neurology and Taub Institute, Columbia University Medical Center, New York, USA.
[15] The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

[Φ] These authors contributed equally to the study
* Corresponding author: 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

ABSTRACT

Epigenome-wide association studies of Alzheimer's disease have highlighted neuropathology-associated DNA methylation differences, although existing studies have been limited in sample size and utilized different brain regions. Here, we combine data from six methylomic studies of Alzheimer's disease (N=1,453 unique individuals) to identify differential methylation associated with Braak stage in different brain regions and across cortex. At an experiment-wide significance threshold ($P<1.238 \times 10^{-7}$) we identified 236 CpGs in the prefrontal cortex, 95 CpGs in the temporal gyrus and ten CpGs in the entorhinal cortex, with none in the cerebellum. Our cross-cortex meta-analysis (N=1,408 donors) identified 220 CpGs associated with neuropathology, annotated to 121 genes, of which 96 genes had not been previously reported at experiment-wide significance. Polyepigenic scores derived from these 220 CpGs explain 24.7% of neuropathological variance, whilst polygenic scores accounted for 20.2% of variance in these samples. The meta-analysis summary statistics are available in our online data resource (www.epigenomicslab.com/ad-meta-analysis/).

INTRODUCTION

Alzheimer's disease (AD) is a chronic neurodegenerative disease that is accompanied by memory problems, confusion and changes in mood, behavior and personality. AD accounts for ~60% of dementia cases, which affected 43.8 million people worldwide in 2016[1]. The disease is defined by two key pathological hallmarks in the brain: extracellular plaques comprised of amyloid-beta (Aβ) protein and intracellular neurofibrillary tangles (NFTs) of hyperphosphorylated tau protein[2-4]. These neuropathological changes are thought to occur perhaps decades before clinical symptoms manifest and the disease is diagnosed[4]. AD is a multi-factorial and complex disease, with the risk of developing disease still largely unknown despite numerous genetic and epidemiological studies over recent years.

Several studies have suggested that epigenetic mechanisms may play a role in disease etiology. In recent years a number of epigenome-wide association studies (EWAS) have been performed in AD brain samples, which have predominantly utilized the Illumina Infinium HumanMethylation450K BeadChip (450K array) in conjunction with bisulfite-treated DNA to assess levels of DNA methylation in cortical brain tissue from donors with varying degrees of AD pathology[5-11]. Independently these studies have nominated a number of loci that show robust differential DNA methylation in disease, and many of these overlap between studies, for example loci annotated to *ANK1*, *RHBDF2*, *HOXA3*, *CDH23* and *RPL13* have been consistently reported. Here we have performed a meta-analysis of six independent EWAS of AD brain, totalling 1,453 independent donors, to identify robust and consistent differentially methylated loci associated with Braak stage, used as a measure of NFT spread through the brain. Five of these studies have been previously published, which together had identified 112 unique CpGs (annotated to 66 genes using UCSC annotation) that reached Bonferroni significance. Our meta-analysis approach provides additional power to detect DNA methylomic variation associated with AD pathology at novel loci, in addition to providing further replication of loci that have been previously nominated in the smaller independent EWAS.

RESULTS

Pathology-associated DNA methylation signatures in discreet cortical brain regions

We identified six EWAS of DNA methylation in AD with a cohort size of > 50 unique donors. All had data on Braak stage available, which we used as a standardized measure of tau pathology spread through the brain (**Table 1**). We were interested in identifying epigenomic

profiles associated with Braak stage in specific brain regions, leveraging additional power by meta-analysing multiple studies to identify novel loci. To this end, we performed an EWAS in each available tissue and cohort separately, looking for an association between DNA methylation and Braak stage, whilst controlling for age and sex (all tissues) and neuron/glia proportion (cortical bulk tissues only), with surrogate variables (SVs) added as appropriate to reduce inflation. For discovery, we then used the estimated effect size (ES) and standard errors (SEs) from these six studies (N = 1,453 unique donors) for a fixed-effect inverse variance weighted meta-analysis separately for each tissue (prefrontal cortex [PFC]: N = 959; temporal gyrus [TG]: N = 608, entorhinal cortex [EC]: N = 189 cerebellum [CER]: N = 533) (**Supplementary Figure 1**).

The PFC represented our largest dataset (N = 959 samples) and we identified 236 experiment-wide significant (EWS) differentially methylated positions (DMPs), of which 193 were annotated to 137 genes, with 43 unannotated loci (**Figure 1a**, **Supplementary Figure 2, Supplementary Table 1**). Previous EWAS of the PFC have consistently reported the *HOXA* gene cluster as a region that is hypermethylated in AD PFC[6,7], with a cell-type specific EWAS demonstrating this is neuronal-derived[11]. Indeed, the most significant DMP in the PFC in our meta-analysis resided in *HOXA3* (cg22962123: ES [defined as the methylation difference between Braak 0 and Braak VI] = 0.042, $P = 5.97 \times 10^{-15}$), with a further 16 of the EWS DMPs also annotated to this gene. This locus appeared to be particularly hypermethylated with higher Braak stage in the PFC, and to a slightly lesser extent in the TG (**Supplementary Figure 3**). There was no significant difference in methylation at this locus in the EC (P = 0.864), which is interesting given that the EC may succumb to pathology early in the disease process (Braak stage III). Of the 236 PFC DMPs, 217 were nominally significant ($P < 0.05$) in the TG, of which 28 were EWS, whilst 22 were nominally significant in the EC, with three reaching EWS (**Figure 1b**). The effect sizes for the 236 EWS PFC DMPs were highly correlated with the effect sizes for the same probes in both the TG (Pearson's correlation coefficient (r) = 0.94, $P = 6.17 \times 10^{-112}$) and EC ($r = 0.58$, $P = 1.80 \times 10^{-22}$) and showing the same direction of effect more than expected by chance (sign test: TG $P = 5.07 \times 10^{-67}$, EC $P = 6.88 \times 10^{-26}$) (**Supplementary Figure 4).** For the 236 PFC EWS DMPs these had the largest effect sizes in the PFC, with a smaller effect size in the TG and EC (**Figure 1c**). Of the 236 PFC EWS DMPs, 26 of these had being previously reported at EWS in previous publications on the individual cohorts[5-7], including one probe in *ANK1*, one probe in *HOXA3*, one probe in *PPT2/PRRT1* and

two probes in *RHBDF2*, amongst others. However, our approach has nominated 210 novel DMPs at EWS (although several had been reported in previous studies at a more relaxed significance threshold, or in regional analyses). This included several additional probes residing in genes already nominated (from another probe) in earlier studies, for example a further 16 probes in *HOXA3* and two probes in *PPT2/PRRT1*. Interestingly, we also nominated a number of novel differentially methylated genes, including some which featured multiple EWS DMPs including for example seven probes in *AGAP2* and five probes in *SLC44A2*, amongst others. One other noteworthy novel EWS DMP in the PFC was cg08898775 (ES = 0.019, P = 4.03 x 10^{-9}), annotated to *ADAM10*, which encodes for α-secretase which cleaves APP in the non-amyloidogenic pathway. A differentially methylated region (DMR) analysis, which allowed us to identify areas of the genome consisting of $\geq$ 2 DMPs, revealed 262 significant DMRs in the PFC (**Supplementary Table 2**), the most significant containing 20 probes and located in *HOXA3* (chr7:27,153,212-27,155,234: Sidak-corrected p = 8.21 x 10^{-50}, **Supplementary Figure 5**), as well as several other DMRs in the *HOXA* gene cluster.

A meta-analysis of TG EWAS datasets (N = 608 samples) nominated 95 EWS probes, of which 75 were annotated to 53 genes, with 20 unannotated probes (**Figure 1a**, **Supplementary Figure 6**, **Supplementary Table 3**). The most significant probe was cg11823178 (ES = 0.029, P = 3.97 x 10^{-16}, **Supplementary Figure 7**), which is annotated to the *ANK1* gene, with the fifth (cg05066959: ES = 0.042, P = 4.58 x 10^{-13}) and 82nd (cg16140558: ES = 0.013, P = 8.44 x 10^{-8}) most significant probes in the TG also being annotated to nearby CpGs in that gene. This locus has been widely reported to be hypermethylated in AD from prior EWAS[5,6,8], as well as in other neurodegenerative diseases such as Huntington's disease (HD) and Parkinson's disease (PD)[12]. Another noteworthy gene is *RHBDF2*, where five EWS DMPs in the TG were annotated to (cg05810363: ES = 0.029, P = 2.25 x 10^{-11}; cg13076843: ES = 0.031, P = 2.97 x 10^{-11}; cg09123026: ES = 0.012, P = 3.46 x 10^{-9}; cg12163800: ES = 0.025, P = 5.85 x 10^{-9}; cg12309456: ES = 0.016, P = 1.33 x 10^{-8}); and which has been highlighted in previous EWASs in AD in the individual cohorts[5,6]. Of the 95 TG DMPs, 84 were nominally significant in the PFC, of which 28 were EWS, whilst 51 were nominally significant in the EC, six of which were EWS (**Figure 1b**). Given the high degree of overlapping significant loci between the TG and other cortical regions, it was not surprising that the ES of the 95 EWS TG probes were highly correlated with the ES of the same loci in both the PFC (r = 0.91, P = 5.09 x 10^{-38}) and EC (r = 0.77, P = 4.02 x 10^{-20}) and were enriched for the same direction of effect (sign test: PFC P = 5.05 x 10^{-29}, EC = 2.30 x 10^{-25}) (**Supplementary Figure 8**). The majority of the 95

TG EWS DMPs were hypermethylated, and the mean ES difference was greater in the TG than the PFC or EC (**Figure 1c**). Twenty-two of the 95 TG EWS DMPs have been previously reported to be significantly differentially methylated in published EWAS, including for example three probes in *ANK1* and four of the probes in *RHBDF2*. Our meta-analysis approach in the TG has identified 73 novel DMPs (at EWS), including some novel genes with multiple DMPs, for example four probes in *RGMA* and two probes in the ankyrin repeat gene *ANKFY1*, amongst others. Finally, our regional analysis highlighted 104 DMRs (**Supplementary Table 4**); the top DMR resided in the *ANK1* gene (chr8:41,519,308-41,519,399) and contained two probes (Sidak-corrected P = 1.72 x 10^{-21}) (**Supplementary Figure 9**). The five DMPs in *RHBDF2* that were already highlighted also represented a significant DMR (Sidak-corrected P = 8.47 x 10^{-21}), with three other genomic regions containing large, significant DMRs consisting of $\geq$ 10 probes, such as *MCF2L* (chr13:113698408-113699016 [10 probes], Sidak-corrected P = 1.16 x 10^{-19}), *PRRT1/PPT2* (chr6:32120773-32121261 [17 probes], Sidak-corrected P = 4.90 x 10^{-15}) and *HOXA5* (chr7:27184264-27184521 [10 probes], Sidak-corrected P = 1.60 x 10^{-7}).

The final cortical region we had available was the EC (N = 189), where we identified ten EWS probes in our meta-analysis, all of which were hypermethylated with higher Braak stage (**Figure 1a**, **Supplementary Figure 10, Supplementary Table 5**). These ten probes were annotated to eight genes, with two EWS probes residing in each of the *ANK1* and *SLC15A4* genes. As with the TG, the most significant DMP was cg11823178 (ES = 0.045, P = 5.22 x 10^{-10}, **Supplementary Figure 7**), located within the *ANK1* gene, with the fourth most significant EWS DMP being located within 100bp of that CpG (cg05066959: ES = 0.062, P = 2.93 x 10^{-9}). In total seven of the ten EWS EC probes have been reported previously to be EWS, including the two probes in *ANK1.* Three of the EC EWS DMPs were novel CpGs, having not been nominated as EWS in previous EWAS, with two of these representing novel genes (*STARD13, CAMP)*. Of the ten EC EWS probes, nine were nominally significant in TG, six of which were EWS, whilst seven were nominally significant in PFC, three of which were EWS (**Figure 1b**). Of the four DMPs that were EWS in only the EC, three of these were nominally significant in at least one other tissue, with just one novel probe, annotated to *STARD13* (cg11563844, ES = 0.027, P = 1.07 x 10^{-8}), only showing hypermethylation in the EC. The effect sizes of the ten EC EWS DMPs were significantly correlated with the effect size of the same probes in the PFC (r = 0.74, P = 0.01) and TG (r = 0.85, P = 1.52 x 10^{-3}) and were enriched for the same direction of effect (sign test: PFC P = 0.021, TG P = 1.95 x 10^{-3})

(**Supplementary Figure 11**). For the ten EC EWS DMPs these were hypermethylated in all three cortical regions, with the greatest Braak-associated ES difference in the EC (**Figure 1c**). A regional analysis identified seven DMRs (**Supplementary Table 6**); the top three DMRs (*RHBDF2:* chr17:74,475,240-74,475,402 [five probes], P = 7.68 x 10^{-14}, **Supplementary Figure 12;** *ANK1*: chr8:41519308-41519399 [two probes], P = 4.89 x 10^{-13}; *SLC15A4*: chr12:129281444-129281546 [three probes], P = 5.24 x 10^{-12}) were significant in at least one of the other cortical regions we meta-analyzed.

To date, a few independent EWAS in AD have been undertaken in the CER and none of these showed any EWS DMPs. In our meta-analysis we identified no EWS DMPs, nor any DMRs in the CER (**Supplementary Figure 13**) despite this analysis including 533 CER samples. There was no correlation of the ES for the EWS DMPs we had identified in the meta-analyses of the three cortical regions with the ES of the same probes in the CER (PFC: r = 0.11, P = 0.08; TG: r = 0.14, P = 0.17; EC: r = 0.48, P = 0.16; **Supplementary Figure 14**).

220 CpGs are differentially methylated across the cortex in AD

We were interested in combining data from across the different cortical tissues to identify common differentially methylated loci across the cortex and also to provide more power by utilizing data from 1,408 unique individuals with cortical EWAS data available. To this end we undertook a cross-cortex meta-analysis, using the correlation structure between tissues where multiple cortical brain regions were available from the same individual so that only one measure was used per individual donor per probe. In this analysis we controlled for age, sex and neuron/glia proportion, with SVs added as appropriate to reduce inflation. Using this approach, we identified 220 EWS probes, of which 168 were annotated (Illumina UCSC annotation) to 121 genes, with 52 DMPs unannotated. **Figure 2a**, **Figure 2b**, **Table 2**, **Supplementary Table 7, Supplementary Figure 15**). All of the 220 probes were nominally significant ($P < 0.05$) in ≥ two cohorts, with ten of these probes being nominally significant in all six cohorts (**Supplementary Figure 16**), which included single probes annotated to *ANK1*, *ABR*, *SPG7* and *WDR81*, two probes in *DUSP27*, three probes in *RHBDF2* and one unannotated probe. We observed similar DNA methylation patterns across all cortical cohorts and tissues for the 220 EWS probes with most showing an increase in methylation with higher Braak stage, which was not seen in the CER (**Supplementary Figure 17**). Of the 220 EWS DMPs we identified, 36 of these have been previously reported at EWS in published EWAS, including multiple previously nominated probes in *ANK1* (cg05066959, cg11823178), *MCF2L*

(cg07883124, cg09448088), *PCNT* (cg00621289, cg04147621, cg23449541) and *RHBDF2* (cg05810363, cg12163800, cg12309456, cg13076843). The most significant probe we identified in our cross-cortex analysis was cg12307200 (**Table 2**, ES = -0.015, P = 4.48 x 10^{-16}), which is intergenic and found at chr3:188664632, located between the *TPRG1* and *LPP* genes and had been previously reported at EWS by De Jager and colleagues[6]. Our cross-cortex meta-analysis approach has identified 184 novel DMPs (at EWS), annotated to 107 genes. Although 11 of these genes had previously been reported at EWS (another probe within that gene), the remaining 96 genes represent novel EWS loci in AD. Many of these novel differentially methylated genes had multiple EWS probes, for example five probes in *AGAP2*, three probes in *HOXB3* and *SLC44A2*, and two probes in *CDH9*, *CPEB4*, *DUSP27*, *GCNT2*, *MAMSTR*, *PTK6*, *RGMA*, *RHOB*, *SMURF1*, *THBS1*, *ZNF238* and *ZNF385A* (**Supplementary Table 7**). Although some of these loci may have been reported in earlier AD EWAS, none of them were at EWS and so here represent novel loci. Interestingly, some of the protein products of these genes have been previously linked with AD; PTK6 is a protein kinase whose activity has been shown to be altered in post-mortem AD brain[13]. Similarly, RGMA has been shown to be increased in AD brain, where it accumulated in Aβ plaques[14].

Having observed several genes with multiple EWS DMPs we undertook a formal regional analysis and identified 221 DMRs, with the top DMR containing 11 probes and covering the *HOXA* region (chr7:27,153,212-27,154,305: P = 3.84 x 10^{-35}) (**Figure 2c**, **Supplementary Table 8**). The *HOXA* gene cluster featured a number of times in our DMR analysis; four of the ten most significant DMRs fell in this genomic region, including DMRs spanning four probes (chr7:27146237-27146445: P = 4.11 x 10^{-27}), 33 probes (chr7:27183133-27184667: P = 2.22 x 10^{-20}) and ten probes (chr7:27143235-27143806: P = 1.75 x 10^{-18}). We were interested to investigate whether specific functional pathways were altered across the cortex; a gene ontology (GO) pathway analysis of the 121 genes annotated to the 220 EWS cross-cortex DMPs highlighted epigenetic dysfunction in numerous pathways, interestingly including those associated with development (**Supplementary Table 9**).

Replication of pathology associated DMPs in the cortex

To replicate our findings we used the estimated coefficients and SEs for these 220 probes generated in a seventh independent ("Munich") cohort, which consisted of 450K data generated in the PFC (N = 45) and sorted neuronal and non-neuronal nuclei from the occipital cortex

(OC: N = 26) (**Table 1**). This cohort had not been used in our discovery analyses as < 50 samples were available. Notably, we identified a similar pattern of Braak-associated DNA methylation changes for the 220 EWS cross-cortex probes in this replication cohort, with a strongly correlated effect size between the discovery dataset and the replication PFC (r = 0.64, P = 5.24 x 10^{-27}), neuronal (r = 0.45, P = 1.56 x 10^{-12}) and non-neuronal datasets (r = 0.79, P= 1.43 x 10^{-47}) with a similar direction of effect (sign test: PFC P = 4.59 x 10^{-28}, neuronal P = 6.13 x 10^{-15}, non-neuronal P = 1.06 x 10^{-42}) (**Figure 2d**). The most significant probe from the cross-cortex meta-analysis (cg12307200) showed consistent hypomethylation in disease in all cohorts in all cortical brain regions, with this direction of effect replicated in the PFC and non-neuronal nuclei samples, but not the neuronal nuclei samples, suggesting that this is primarily driven by non-neuronal cell types, which are likely to be glia (**Figure 2e**). We have developed an interactive tool, which can generate a forest plot showing the ES and SE across any of the discovery and replication tissues and cohorts for any of the 403,763 probes that passed our quality control (QC), in addition to mini-Manhattan plots to show DMRs utilizing the summary statistics from the cross-cortex meta-analysis (www.epigenomicslab.com/ad-meta-analysis/).

Cross-cortex AD-associated DMPs are enriched in specific genomic features

To identify if DMPs reside in specific genomic features, we used a Fisher's exact test to look for an enrichment of the 220 EWS DMPs using Slieker annotations[15] (**Supplementary Table 10**, **Supplementary Figure 18**). We observed a significant over representation of EWS DMPs in CpG islands (CGIs) of gene bodies (odds ratio [OR] = 3.199, P = 4.76 x 10^{-10}), and in CGI shelves and non-CGI areas of proximal promoters (OR = 3.571, P = 9.09 x 10^{-3} and OR = 1.641, P = 0.03, respectively). However, DMPs located in CGIs in the proximal promoter were under-represented (OR = 0.353, P = 2.08 x 10^{-6}). There was a significant over representation of the 220 cross-cortex EWS DMPs in the first exon (OR = 1.80, P = 0.02), with an under enrichment within 1500bp of the transcription start site (TSS) (OR = 0.49, P = 3.82 x 10^{-3}) (**Supplementary Table 11**, **Supplementary Figure 19**).

Polyepigenic scores generated from the 220 cross-cortex EWS DMPs explain 24.7% of variance in individuals with high or low pathology.

We calculated polyepigenic scores (PES) in samples with matched genetic data available by summing together the DNA methylation beta values of the 220 EWS cross-cortex DMPs, weighted by their fixed ES. In the cross-cortex data the PES was able to explain 24.7% of variance in pathology when comparing low pathology samples (Braak 0-II: "controls", N =

214) with high pathology samples (Braak V-VI: "AD cases", N = 298) using the cross-cortex meta-analysis data (area under the curve (AUC) receiver operating curve (ROC) analysis = 74.7%, Bootstrap confidence interval [CI] = 70.3-79.2%, **Figure 3**, **Supplementary Table 12a**). We were interested to compare the performance of the PES to polygenic risk score (PRS), which was able to explain 20.2% of pathology variance (AUC = 72.6%, CI = 68.1-77.2%) in the same samples, with the combined PES and PRS together explaining 38.4% of pathology variance (AUC = 81.2%, CI = 77.2-84.9%) and demonstrates that most of the PES is independent of genetic effects. Furthermore, including age and sex, other known AD risk factors, increased the variance explained by the model to 48.6% (AUC = 86.1%, CI = 82.8-89.1%, **Figure 3**, **Supplementary Table 12a**). To demonstrate the reliability of the PES we randomly selected 220 probes in 1,000 simulations, which only explained on average 0.8% of variance (mean AUC = 51.9%, mean CI = 46.7-57.1%, **Supplementary Table 12a**). Using the PES in CER data we could only explain 0.3% pathology variance in Braak stage in a sample (AUC = 54.4%, CI = 48.4-59.5%, **Supplementary Table 12b**).

Methylomic variation in the cortex can be used to accurately classify degree of pathology

We were interested to investigate whether the DNA methylation patterns within the cross-cortex EWS probes that make up the PES could accurately predict the pathological load of a brain sample. To this end we used elastic net regression to identify 95 probes in the 220 cross-cortex EWS loci (**Supplementary Table 13)** that were able to classify post-mortem low pathology "control" from high pathology "AD" status in our training dataset (N = 696) with 88% accuracy (**Supplementary Table 14**, **Supplementary Figure 20**). When this classifier was tested in the testing dataset (N = 233) it achieved 79% accuracy with 80% sensitivity (CI = 73-87%) and 77% specificity (CI = 68-85%), with a positive predictive value (PPV) of 82% (CI = 75-88%) and a negative predictive value (NPV) of 75% (CI = 65-83%) (**Supplementary Table 14**, **Supplementary Figure 20**).

Probes in the vicinity of the AD genetic risk variants HLADRB1, SPI1, SORL1, ABCA7 and ADAM10 are differentially methylated in AD cortex

Using the most recent list of (late onset) AD genomic risk variants from Kunkle *et al*[16] we examined the enrichment of Braak-associated cross-cortex DMPs in 24 linkage disequilibrium (LD) blocks harboring risk variants. Twenty of these LD blocks contained > 1 CpG site on the 450K array and using Brown's method we combined P values within each of these 20 genomic regions. We observed Bonferroni-adjusted significant enrichment in the cross-cortex data in

the *HLADRB1* (Chr6: 32395036-32636434: adjusted P = 1.20 x 10^{-3}), *SPI1* (Chr11: 47372377-47466790, adjusted P = 5.76 x 10^{-3}), *SORL1* (Chr11: 121433926- 121461593, adjusted P = 0.019), *ABCA7* (Chr19: 1050130- 1075979, adjusted P = 0.022) and *ADAM10* (Chr15: 58873555- 59120077, adjusted P = 0.022) LD regions (**Supplementary Table 15**).

DISCUSSION

This study represents the first meta-analysis of AD EWAS utilizing six independent sample cohorts with a range of cortical brain regions and CER available as a discovery dataset. All of our data can be explored as part of an interactive search tool, which can be found on our website (https://www.epigenomicslab.com/ad-meta-analysis). Five of these discovery cohorts have been previously published and had collectively identified 112 unique CpG sites (at EWS), corresponding to 66 genes (using UCSC annotation) that were associated with pathology or diagnosis at Bonferroni significance. By performing a meta-analysis within each tissue, we have been able to identify substantially more loci in the individual tissues, nominating 236, 95 and ten EWS DMPs in the PFC, TG and EC, respectively. Although far fewer loci were identified in the EC compared to the other cortical regions, this is likely due to the reduced sample size in this tissue. In the CER despite meta-analyzing > 500 unique samples, we identified no Braak-associated DNA methylation changes. Furthermore, there was no correlation of the EWS DMPs identified in any of the cortical regions with the effect size of the same probes in the CER. Taken together this suggests that DNA methylomic changes in AD are cortex cell type specific. This observation is interesting as the CER is said to be "spared" from AD pathology, with an absence of NFTs, although some diffuse Aβ plaques are reported[17]. A recent spatial proteomics study of AD reported a large number of protein changes in the CER in AD; however, the proteins identified were distinct from other regions examined and thus the authors suggested a potential protective role[18].

Although many loci showed similar patterns of Braak-associated DNA methylation across the different cortical regions, some loci did show some regional specificity. In order to identify CpG sites that showed common DNA methylation changes in disease we performed a cross-cortex meta-analysis. Using this approach we identified 220 EWS probes associated with Braak stage of which 36 probes (corresponding to 20 genes) had been previously reported at EWS in the individual cohort studies that we had used for our meta-analysis, for example two probes in *ANK1*, four probes in *RHBDF2* and one probe in *HOXA3*, amongst others. Interestingly, our

approach did identify 184 novel CpGs, corresponding to 107 unique genes, of which 96 genes had not been previously reported at EWS in any of the previously published AD brain EWAS, highlighting the power of our meta-analysis approach for nominating new loci. This included 15 novel genes with at least two EWS DMPs each, which had not been nominated previously at EWS in AD, although a number of these genes had been previously nominated from DMR analyses, which have a less stringent threshold. However, we did identify one novel gene (*HOXB3*) with three EWS DMPS, which had not been nominated at EWS in previous EWAS DMP or DMR analyses in AD brain. The nomination of loci in the *HOXB* gene cluster is interesting; a recent study of human Huntington's disease brain samples also highlighted significantly increased *HOXB3* gene expression in the PFC[19], an interesting observation given that both AD and HD are disorders that feature dementia. Furthermore, we have recently reported AD-associated hypermethylation of the *HOXB6* gene in AD blood samples[20].

Our genomic enrichment analyses identified an over representation of methylation in AD in specific genomic features, for example CGIs in gene bodies, and shelves and non-CGI regions in proximal promoters. We observed a significant enrichment of cross-cortex loci in the LD regions surrounding the AD-associated genetic variants *HLADRB1*, *SPI1*, *SORL1*, *ABCA7* and *ADAM10* after controlling for multiple testing. In recent years it has been reported that common genetic variation can explain ~33% of AD phenotypic variance[21]. Indeed, when we generated a PRS in our cross-cortex samples this was able to explain 20.2% of variance in our data. Interestingly, by generating PES in the same samples using the 220 EWS cross-cortex DMPs we were able to explain a slightly higher amount of variance in pathology (24.7%) in the same samples, with the PRS and PES together able to explain 38.4% of variance, and suggesting that that most of the PES is largely independent of genetic effects. We were able to increase the amount of phenotypic variance explained to 48.6% when we included age and sex, which are other known AD risk factors. Finally, we have developed a classifier that could accurately predict control samples with low pathology, from those with a post-mortem AD diagnosis due to high pathology using methylation values for 95 of the 220 EWS probes, further highlighting that distinct genomic loci reproducibly show epigenetic dysfunction in AD cortex.

There are some limitations with our study. First, as we have largely utilized methylation data generated in bulk tissue, this will contain a mixture of different cell types. Furthermore, it is known that the proportions of the major CNS brain cell types are altered in AD, with reduced numbers of neurons and increased glia. As such, it is possible that the identified DNA

methylation changes represent a change in cell proportions. To address this, we have included neuron/glia proportions as a co-variate in our models to minimize bias. Although this is the optimal strategy for the current study given the EWAS data had already been generated, future EWAS should be undertaken on sorted cell populations with larger sample numbers than our replication cohort, or ideally at the level of the single cell. Second, our study has utilized previously generated EWAS data generated on the 450K array. Although the Illumina array platform is the current workhorse for EWAS, it is limited to only analyzing a relatively small proportion of the potential methylation sites in the genome (~400,000 on the 450K array) and given the falling cost of sequencing, future studies could exploit this by performing reduced representation bisulfite sequencing (RRBS) to substantially increase the coverage. Finally, one key issue for epigenetic studies in post-mortem tissue is the issue of causality, where it is not possible to determine whether nominated epigenetic loci are driving disease pathogenesis, or are a consequence of the disease, or even the medication used for treatment. Although this is difficult to answer in post-mortem human studies, longitudinal studies in animal models, or modelling methylomic dysfunction through epigenetic editing *in vitro* will begin to address these issues. In addition, examining DNA methylation signatures in brain samples in pre-clinical individuals (*i.e.* during midlife) will be important for establishing the temporal pattern of epigenetic changes relative to the pathology. There were some limitations to our PES analyses, for example the PES was generated in the same samples that we had identified the 220 EWS cross-cortex DMPs and so the data could be over-fitted. Similarly, the ROSMAP samples are part of the Alzheimer's Disease Genetics Consortium (ADGC) collection[22], which was utilized for generating the PRS[23]. We were unable to test the score in our independent replication (Munich) dataset as there were too few (< 10) individuals with a Braak stage < II available and thus in the future it will be important to replicate the PES and compare to PRS in other large independent AD EWAS cortical datasets once they become available. Similarly, it will be important to explore the PES and classifier for their performance in predicting non-NFT pathology in individuals with other dementias. It will also be of interest to perform these phenotypic variance analyses in cohorts with data available on other known risk factors, for example education and inflammation.

In summary we present the first meta-analyses of AD EWAS, highlighting numerous EWS DMPs in the individual cortical regions and across the cortex, but not in the CER. A number of these loci are novel and warrant further study to explore their role in disease etiology. We provide the first evidence that epigenomic changes in the cortex can explain a similar

proportion of disease variance as the PRS. Looking to the future it will be important to validate the PES and classifier analyses in an independent dataset, and to explore the relationship between DNA methylation and gene expression in AD brain.

METHODS

Cohorts

Six sample cohorts were used for "discovery" in this study as they all had DNA methylation data generated on the 450K array for > 50 donors, enabling us to take a powerful meta-analysis approach to identify DNA methylation differences in AD. For each sample cohort DNA methylation was quantified using the 450K array. The "London 1" cohort comprised of entorhinal cortex (EC), superior temporal gyrus (STG), prefrontal cortex (PFC), and cerebellum (CER) tissue obtained from 113 individuals archived in the MRC London Neurodegenerative Disease Brain Bank and published by Lunnon *et al.*[5]. The "London 2" cohort comprised EC and CER samples obtained from an additional 95 individuals from the MRC London Neurodegenerative Disease Brain Bank published by Smith and colleagues[8]. The "Mount Sinai" cohort comprised of STG and PFC tissue obtained from 146 individuals archived in the Mount Sinai Alzheimer's Disease and Schizophrenia Brain Bank published by Smith and colleagues[7]. The "Arizona 1" cohort consisted of 302 middle temporal gyrus (MTG) and CER samples from The Sun Health Research Institute Brain Donation Program[24], which has not been published to date. The "Arizona 2" cohort consisted of an additional 88 TG and CER samples from Lardonije *et al.*[10]. The "ROSMAP" cohort consisted of 709 samples from the Rush University Medical Center: Religious Order Study (ROS) and the Memory and Aging Project (MAP), which were previously published by De Jager and colleagues[6]. For replication purposes we used a seventh dataset, the "Munich" cohort" from Neurobiobank Munich (NBM), which had bulk PFC 450K array data from 45 donors, and 450K array data from fluorescence-activated cell sorting (FACS) isolated neuronal and non-neuronal (glial) populations from the occipital cortex (OC) from 26 donors as described by Gasparoni *et al.*[11]. Demographic information for all seven cohorts is available in **Table 1**.

Data quality control and harmonization

All computations and statistical analyses were performed using R 3.5.2[25] and Bioconductor 3.8[26]. A *MethylumiSet* object was created from iDATs using the methylumi package[27] and *RGChannelSet* object was created using the minfi package[28]. Samples were excluded from

further steps if (a) the mean background intensity of negative probes < 1,000, (b) the mean detection P values > 0.005, (c) the mean intensity of methylated or unmethylated signals were three standard deviations above or below the mean, (d) the bisulfite conversion efficiency < 80%, (e) there was a mismatch between reported and predicted sex, or (f) the 65 single nucleotide polymorphism (SNP) probes on the array show a modest level of correlation (0.65) between two samples (whereby the sample with the higher Braak score was retained). Sample and probe exclusion was performed using the *pfilter* function within the wateRmelon package[29], with the following criteria used for exclusion: samples with a detection P > 0.05 in more than 5% of probes, probes with > three beadcount in 5% of samples and probes having 1% of samples with a detection P value > 0.05). Samples numbers after QC are those shown in **Table 1**.

Quantile normalization was applied using the *dasen* function in the wateRmelon package[29]. DNA methylation data was corrected by regressing out the effects of age and sex in all samples in each cohort and tissue separately, with neuron/glia proportions included as an additional covariate in cortical regions. The neuron/glia proportions were calculated using the CETS package[30], and were not included as a co-variate for the CER as the neuronal nuclear protein (NeuN) that was used to generate the neuron/glia algorithm is not expressed by some cerebellar neurons[31]. Surrogate variables (SVs) were calculated using the *sva* function in the SVA package[32]. Linear regression analyses were then performed with respect to Braak stage (modelled as a continuous variable) using residuals and a variable number of SVs for each study until the inflation index (lambda) fell below 1.2 (see **Supplementary Table 16**). Quantile-quantile (QQ) plots for the five meta-analyses are shown in **Supplementary Figure 21**. Although it appears from the QQ plots that there is P value inflation, it is worth noting that (a) lambda for all meta-analyses < 1.2 and (b) P value inflation is commonly observed in many DNA methylation studies and standard methods to control for this in genome-wide association studies (GWAS) are not suitable for EWAS data[33].

Intra-tissue meta-analysis

We used the estimated coefficients and SEs from the six "discovery" cohorts to undertake a fixed-effect inverse variance intra-tissue meta-analysis independently in each available tissue using the *metagen* function within the Meta package[34], which applies inverse variance weighting. The estimates and SEs from individual cohort Braak linear regressions analyses were added to the model for each tissue. The PFC analyses included three cohorts (N = 959:

London 1, Mount Sinai, ROS/MAP), the TG analyses included four cohorts (N = 608: London 1, Mount Sinai, Arizona 1, Arizona 2) and the EC analyses included two cohorts (N = 189: London 1, London 2). The CER analyses included data from four cohorts (N = 533: London 1, London 2, Arizona 1 and Arizona 2) although the CER data for the Arizona 1 and 2 cohorts was generated in the same experiment, and so these were combined together as a single dataset. The ESs and corresponding SEs reported in this study correspond to the corrected DNA methylation (beta) difference between Braak 0 and Braak VI individuals. Experiment-wide significance (EWS) was defined as $P < 1.238 \times 10^{-7}$ to reflect Bonferroni significance (403,763 probes/0.05).

Cross-cortex meta-analysis

As multiple cortical brain regions were available for the "London 1" and "Mount Sinai" cohorts, a mixed model with nested IDs was performed using the *lme* function within the nlme package[35]. Estimate coefficients and SEs from each EWAS were extracted and were subjected to bacon[33] to control for bias and inflation, after which a fixed-effect inverse variance meta-analysis was performed across all discovery cohorts using the *metagen* function.

Annotations, pathway and regional analyses

Probes were annotated for tables using both the Illumina (UCSC) gene annotation (which is derived from the genomic overlap of probes with RefSeq genes or up to 1500bp of the transcription start site of a gene) and "Genomic Regions Enrichment of Annotations Tool" (GREAT)[36] annotation (which annotates a DMP to genes with a TSS within 5kb upstream, or 1kb downstream). Pathway analyses were performed on the Illumina (UCSC) annotated genes corresponding to the 220 EWS cross-cortex DMPs (N = 121 genes). A logistic regression approach, which we have previously described[37,38], was used to test if genes in this list predicted pathway membership, while controlling for the number of probes that passed QC annotated to each gene. Pathways were downloaded from the Gene Ontology (GO) website (http://geneontology.org/) and mapped to genes, including all parent ontology terms. All genes with at least one 450K probe annotated and mapped to at least one GO pathway were considered. Pathways were filtered to those containing between 10 and 2,000 genes. After applying this method to all pathways, significant pathways ($P < 0.05$) were taken and grouped where overlapping genes explained the signal. This was achieved by taking the most significant pathway and retesting all remaining significant pathways while controlling additionally for the best term. If the test genes no longer predicted the pathway, the term was said to be explained

by the more significant pathway, and hence these pathways were grouped together. This algorithm was repeated, taking the next most significant term, until all pathways were considered as the most significant or found to be explained by a more significant term. To identify DMRs consisting of multiple DMPs we used comb-p[39] with a distance of 500bp and a seeded P value of 1.0×10^{-4}.

Genomic enrichment analyses

To test for an enrichment of DMPs in specific genomic features (*i.e.* CGIs, shelves, shores, non-CGI) in certain genomic regions (*i.e.* intergenic, distal promoter, proximal promoter, gene body, downstream) we annotated all DMPs with Slieker annotation[15] and performed a two-sided Fisher's exact test comparing to all probes analyzed (N = 403,763). We also used a Fisher's exact test to test for an enrichment of DMPs in genomic regions related to transcription based on the Illumina annotation (TSS1500, TSS200, 5' UTR, 1st exon, gene body, 3' UTR). To investigate whether Braak-associated methylation was enriched in known AD GWAS variants we used Brown's method to combine together P values from our meta-analyses for probes residing in the LD blocks around the genome-wide significant ($P < 5.0 \times 10^{-8}$) GWAS variants identified by the stage one meta-analysis of Kunkle *et al.*[16] Of the 24 LD blocks reported by Kunkle and colleagues, 20 contained > 1 CpG site on the 450K array and the P values for each CpG in a given block were combined using Brown's method, which accounts for the correlation structure between probes, with the regional P values adjusted to correct for multiple testing.

Quantifying variance explained by PES and PRS

PES were generated in the discovery cohorts for individuals who classified as either low pathology (Braak stages 0-II: "control") or high pathology (Braak stages V-VI: "AD"). PES were calculated by summing together the DNA methylation beta values of the 220 EWS DMPs, weighted by their fixed-effect ES in the meta-analysis using the OSCA tool[40]. The amount of AD variance explained was compared to the PRS, which was calculated using PRSice (v2.0)[41] in matched GWAS data that had been imputed with the most recent 1,000 Genomes reference panel (phase 3, version 5)[42]. Using PRSice a logistic regression including the first ten principal components of the genetic data was performed using the standard PRS protocol[41,43], with control and AD case status defined as post-mortem Braak low (0-II) and high (V-VI), respectively. The GWAS summary statistics were first pruned for LD, then the SNPs in the

target dataset (meta-analysis genetic data) were weighted for their effect size in a recent AD GWAS[16], before being summed into a PRS for every individual analyzed in the current study using the optimal Nagelkerke's pseudo R^2 (NK R^2) threshold, which we calculated to be $P < 1.0 \times 10^{-4}$. The predictive power of seven models: (a) age + sex, (b) PES, (c) PES + age + sex, (d) PRS, (e) PRS + age + sex, (f) PES + PRS and (g) PES + PRS + age + sex, were tested in control (Braak 0-II, N = 214) versus AD cases (Braak V-VI, N = 298) using AUC/ROC analysis as follows. First, logistic regressions were conducted in R using control versus AD status as the outcome in the seven models in the cross-cortex data and then later the CER data. Second, for each of the models, the fitted values from the regressions were used to calculate the AUC/ROC using the R package pROC[44]. The variance explained was then calculated using the Nagelkerke's pseudo R^2 using the R package rsq[45]. In order to demonstrate the reliability of the PES we randomly selected 220 probes, recalculated the PES and tested the effect on explaining Braak variance in the samples. This procedure was performed 1,000 times.

Developing a classifier to predict Braak pathology

For classification analysis, training and testing datasets were randomly assigned in the cross-cortex discovery dataset using control samples (Braak low [0-II]: training N = 283 and testing N = 103) and AD cases (Braak high [V-VI]: training N = 413 and testing N = 130). A penalized regression model was used to select the optimum (N = 95) CpG probes from the 220 cross-cortex EWS DMPs that classified case-control status in the training dataset using the R package GLMnet[46]. Elastic net uses a combination of ridge and lasso regression, in which alpha (α) = 0 corresponds to ridge, whilst $\alpha = 1$ corresponds to lasso, the elastic net α parameter used was 0.5. The lambda value was derived when using 10-fold cross validation on the training dataset. The model was then tested for its sensitivity, specificity, PPV and NPV in the testing dataset.

Acknowledgements

This work was funded by a major project grant from the Alzheimer's Society UK (AS-PG-14-038) to KL, an Alzheimer's Association US New Investigator Research Grant (NIRG-14-320878) to KL and a project grant from the Medical Research Council (MRC) (MR/N027973/1) to KL as part of a larger collaborative project funded to KL and DLAvdH for the EPI-AD consortium through the Joint Programme—Neurodegenerative Disease Research (JPND) initiative. The project was also supported by a number of NIH grants, including P30AG10161, R01AG15819, R01AG17917, R01AG36042, R01AG036039,

R01AG036400 and R01AG067015. We thank the donors and families who have made this research possible.

Contributions

ARS and GW conducted laboratory experiments. RGS, EP, GS, EH, WV and MW undertook data analysis, bioinformatics and/or support with data review. RGS, EP, ARS, JAYR, DM, GG, MR, AG, AJS, LS, VH, DLAvdH, JW, PDC, DAB, PLDJ, JM and KL provided data for the meta-analysis. LS developed the online interactive tool. KL conceived of the idea and directed the project. KL, RGS and EP drafted the manuscript. All authors read and approved the final submission.

Competing interests

The authors declare no competing interests.

Data availability

The data supporting the findings of this study are available within the article, Supplementary Information or from the authors upon request. Some of the datasets are also available on GEO including London 1 data (GSE59685), London 2 data (GSE105109), Mount Sinai data (GSE80970), Arizona 2 TG data (GSE109627) and Munich data (GSE66351). We have developed an interactive tool, which can present summary statistics, which is available from our website: www.epigenomicslab.com/ad-meta-analysis/. All scripts for data analyses performed in this manuscript can be found at: https://github.com/rgs212/Meta-analysis-Smith.

REFERENCES

1. Collaborators, G.B.D.D. 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).
2. Blennow, K., de Leon, M.J. & Zetterberg, H. Alzheimer's disease. *Lancet* **368**, 387-403 (2006).
3. Sperling, R.A. *et al.* Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 280-92 (2011).
4. Jack, C.R., Jr. *et al.* Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol* **9**, 119-28 (2010).

5. Lunnon, K. *et al.* Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat Neurosci* **17**, 1164-70 (2014).
6. 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**, 1156-63 (2014).
7. 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. *Alzheimers Dement* **14**, 1580-1588 (2018).
8. Smith, A.R. *et al.* Parallel profiling of DNA methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in Alzheimer's disease. *Clin Epigenetics* **11**, 52 (2019).
9. Watson, C.T. *et al.* Genome-wide DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease. *Genome Med* **8**, 5 (2016).
10. Lardenoije, R. *et al.* Alzheimer's disease-associated (hydroxy)methylomic changes in the brain and blood. *Clin Epigenetics* **11**, 164 (2019).
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 Chromatin* **11**, 41 (2018).
12. Smith, A.R. *et al.* A cross-brain regions study of ANK1 DNA methylation in different neurodegenerative diseases. *Neurobiol Aging* **74**, 70-76 (2019).
13. Rosenberger, A.F. *et al.* Protein Kinase Activity Decreases with Higher Braak Stages of Alzheimer's Disease Pathology. *J Alzheimers Dis* **49**, 927-43 (2016).
14. Satoh, J., Tabunoki, H., Ishida, T., Saito, Y. & Arima, K. Accumulation of a repulsive axonal guidance molecule RGMa in amyloid plaques: a possible hallmark of regenerative failure in Alzheimer's disease brains. *Neuropathol Appl Neurobiol* **39**, 109-20 (2013).
15. Slieker, R.C. *et al.* Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics Chromatin* **6**, 26 (2013).
16. Kunkle, B.W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* **51**, 414-430 (2019).
17. Selkoe, D.J. The molecular pathology of Alzheimer's disease. *Neuron* **6**, 487-98 (1991).
18. Xu, J. *et al.* Regional protein expression in human Alzheimer's brain correlates with disease severity. *Commun Biol* **2**, 43 (2019).
19. Labadorf, A. *et al.* RNA Sequence Analysis of Human Huntington Disease Brain Reveals an Extensive Increase in Inflammatory and Developmental Gene Expression. *PLoS One* **10**, e0143563 (2015).
20. Roubroeks, J.A.Y. *et al.* An epigenome-wide association study of Alzheimer's disease blood highlights robust DNA hypermethylation in the HOXB6 gene. . *Submitted* (2020).
21. Ridge, P.G., Mukherjee, S., Crane, P.K., Kauwe, J.S. & Alzheimer's Disease Genetics, C. Alzheimer's disease: analyzing the missing heritability. *PLoS One* **8**, e79771 (2013).
22. Naj, A.C. *et al.* Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* **43**, 436-41 (2011).

23. Escott-Price, V. *et al.* Common polygenic variation enhances risk prediction for Alzheimer's disease. *Brain* **138**, 3673-84 (2015).
24. Beach, T.G. *et al.* Arizona Study of Aging and Neurodegenerative Disorders and Brain and Body Donation Program. *Neuropathology* **35**, 354-89 (2015).
25. R Development Core Team. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria 2012* (2012).
26. Gentleman, R.C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80 (2004).
27. Spiers, H. *et al.* Methylomic trajectories across human fetal brain development. *Genome Res* **25**, 338-52 (2015).
28. Aryee, M.J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-9 (2014).
29. Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**, 293 (2013).
30. Guintivano, J., Aryee, M.J. & Kaminsky, Z.A. 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 (2013).
31. Gusel'nikova, V.V. & Korzhevskiy, D.E. NeuN As a Neuronal Nuclear Antigen and Neuron Differentiation Marker. *Acta Naturae* **7**, 42-7 (2015).
32. Chakraborty, S., Datta, S. & Datta, S. Surrogate variable analysis using partial least squares (SVA-PLS) in gene expression studies. *Bioinformatics* **28**, 799-806 (2012).
33. van Iterson, M., van Zwet, E.W., Consortium, B. & Heijmans, B.T. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol* **18**, 19 (2017).
34. Schwarzer, G. meta: A R package for meta-analysis. *R News* **7**, 40-45 (2007).
35. Pinheiro, J., Bates, D., DebRoy, S. & Sarkar, D. nlme: Linear and Nonlinear Mixed Effects Models. (2019).
36. McLean, C.Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501 (2010).
37. Wong, C.C.Y. *et al.* Genome-wide DNA methylation profiling identifies convergent molecular signatures associated with idiopathic and syndromic autism in post-mortem human brain tissue. *Hum Mol Genet* **28**, 2201-2211 (2019).
38. Lunnon, K. *et al.* Variation in 5-hydroxymethylcytosine across human cortex and cerebellum. *Genome Biol* **17**, 27 (2016).
39. 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**, 2986-8 (2012).
40. Zhang, F. *et al.* OSCA: a tool for omic-data-based complex trait analysis. *Genome Biol* **20**, 107 (2019).
41. Choi, S.W. & O'Reilly, P.F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* **8**(2019).
42. Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
43. Euesden, J., Lewis, C.M. & O'Reilly, P.F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466-8 (2015).
44. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).

45. Zhang, D. rsq: R-Squared and Related Measures. (2018).
46. Friedman, J. *et al.* glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models. (2019).

Table 1: Demographic information for cohorts included in the meta-analyses. Sample numbers, split of males (M)/females (F) and mean age at death are shown for individuals with low pathology (Braak 0-II), mid-stage pathology (Braak III-IV) and severe pathology (Braak V-VI) in each cohort. Shown are the bulk tissues available from each cohort, which included the cerebellum (CER), entorhinal cortex (EC), middle temporal gyrus (MTG), prefrontal cortex (PFC) and superior temporal gyrus (STG). In the discovery meta-analyses, we used data from six EWAS, which all had > 50 unique donors. In the replication cohort we used the Munich cohort, which also had available sorted neuronal and non-neuronal cell populations from the occipital cortex (OC). For the meta-analyses STG and MTG samples were classed as TG samples. Shown are final numbers for all cohorts after data quality control (QC).

Stage	Cohort	Braak	Number	Gender (M/F)	Age at death in years (± SD)	Tissues analysed
DISCOVERY	London 1	0-II	29	13/16	77.6 (12.8)	PFC, EC, STG, CER (Bulk)
		III-IV	18	7/11	88.5 (5.2)	
		V-VI	66	26/40	85.4 (8.1)	
	London 2	0-II	23	12/11	76.1 (10.0)	EC, CER (Bulk)
		III-IV	16	3/13	87.6 (6.4)	
		V-VI	56	26/30	81.5 (8.6)	
	Mount Sinai	0-II	60	32/28	82 (7.6)	PFC, STG (Bulk)
		III-IV	42	12/30	88.8 (6.6)	
		V-VI	44	12/32	88.0 (7.5)	
	Arizona 1	0-II	61	40/21	80.3 (8.2)	MTG, CER (Bulk)
		III-IV	97	50/47	86.9 (6.9)	
		V-VI	144	63/81	82.3 (8.5)	
	Arizona 2	0-II	16	10/6	82.5 (5.0)	MTG, CER (Bulk)
		III-IV	45	21/24	86.7 (5.1)	
		V-VI	27	12/15	84.6 (7.1)	
	ROS/MAP	0-II	143	70/73	83.2 (6.0)	PFC (Bulk)
		III-IV	409	143/266	86.9 (4.1)	
		V-VI	157	45/112	87.8 (3.5)	
REPLICATION	Munich	0-II	9	5/4	76.7 (10.9)	PFC (Bulk)
		III-IV	7	1/6	82.1 (5.2)	
		V-VI	29	12/17	79.2 (8.5)	
		0-II	11	7/4	75.9 (8.5)	OC (sorted cells)
		III-IV	5	1/4	85.0 (6.5)	
		V-VI	10	4/6	77.9 (6.6)	

Table 2: The 25 most significant experiment-wide significant (EWS) differentially methylated positions (DMPs) associated with Braak stage from the cross-cortex meta-analysis. Probe information is provided corresponding to chromosomal location (hg19/GRCh37 genomic annotation), Illumina gene annotation, genes with a transcription start site (TSS) within 5Kb upstream and genes with TSS within 1Kb downstream (from GREAT annotation). Shown for each DMP is the effect size (ES), standard error (SE) and corresponding unadjusted P value from the inverse variance fixed effects meta-analysis model in the cross cortex data. All ES and SE have been multiplied by six to demonstrate the difference between Braak stage 0 and Braak stage VI samples.

Probe	Position	Illumina Gene Annotation	Genes with TSS within 5Kb upstream (GREAT annotation)	Genes with TSS within 1Kb downstream (GREAT annotation)	ES	SE	P
cg12307200	chr3:188664632		TPRG1 (-225131)	LPP (+733912)	-0.015	0.002	4.48E-16
cg01419713	chr8:42038135	PLAT		PLAT (+27107), AP3M2 (+27672)	0.022	0.003	2.20E-14
cg04874795	chr16:86477638		FOXF1 (-66495)	IRF8 (+545230)	-0.022	0.003	3.95E-14
cg11823178	chr8:41519399	ANK1;MIR486	NKX6-3 (-14522)	ANK1 (+234881)	0.016	0.002	3.24E-13
cg07061298	chr7:27153847	HOXA3	HOXA2 (-11418)	HOXA3 (+5367)	0.018	0.002	4.57E-13
cg13076843	chr17:74475294	RHBDF2		RHBDF2 (+22195), AANAT (+25862)	0.021	0.003	7.57E-13
cg25018458	chr17:980014	ABR		TIMM22 (+79658), ABR (+103154)	0.008	0.001	7.87E-13
cg07883124	chr13:113634042	MCF2L	F7 (-126079)	MCF2L (+10508)	0.017	0.002	9.10E-13
cg03223072	chr10:116398913	ABLIM1	AFAP1L2 (-234670)	ABLIM1 (+19144)	-0.014	0.002	1.10E-12
cg05066959	chr8:41519308	ANK1;MIR486	NKX6-3 (-14431)	ANK1 (+234972)	0.024	0.003	1.45E-12
cg17881200	chr7:27138850		HOXA1 (-3258)		0.017	0.002	1.83E-12
cg19240213	chr7:27163095	HOXA3	HOXA3 (-3882)		0.020	0.003	2.29E-12
cg10045881	chr1:111770291	CHI3L2	CHIA (-63247)	CHI3L2 (+26899)	-0.015	0.002	2.38E-12
cg02674693	chr11:45109122		TP53I11 (-137412), PRDM11 (-59772)		0.018	0.003	3.57E-12
cg06800235	chr1:7692367	CAMTA1	VAMP3 (-138962)	CAMTA1 (+846984)	-0.017	0.002	3.71E-12
cg18264562	chr1:26253412		STMN1 (-20456)	PAFAH2 (+71236)	0.014	0.002	5.46E-12
cg01964852	chr7:27146262	HOXA3	HOXA2 (-3833)		0.016	0.002	5.96E-12
cg01111041	chr6:32121055	PPT2;PRRT1	PRRT1 (-1327), PPT2-EGFL8 (-944), PPT2 (-245)		0.009	0.001	6.83E-12
cg15974867	chr11:69464012	CCND1		CCND1 (+8158), ORAOV1 (+26103)	0.018	0.003	7.46E-12
cg17907520	chr15:31680189			KLF13 (+61132), OTUD7A (+267353)	0.011	0.002	9.65E-12
cg16988611	chr10:82224946	TSPAN14		TSPAN14 (+11025)	0.011	0.002	9.98E-12
cg13579486	chr20:39314091			MAFB (+3789)	-0.012	0.002	1.01E-11
cg01681367	chr16:29676071	SPN	QPRT (-14287)	SPN (+1492)	-0.015	0.002	1.25E-11
cg01301319	chr7:27153580	HOXA3	HOXA2 (-11151)	HOXA3 (+5634)	0.017	0.003	1.54E-11
cg02317313	chr12:122235206	LOC338799	RHOF (-3039)		0.017	0.003	1.69E-11

Figure 1: Intra-tissue meta-analyses of AD methylomic studies highlights experiment-wide significant (EWS) differentially methylated positions (DMPs) in all cortical tissues. (a) A Manhattan plot for the prefrontal cortex (PFC: red), temporal gyrus (TG: green) and entorhinal cortex (EC: blue) meta-analyses, with the ten most significant loci DMPs circled on the plot and Illumina UCSC gene name shown if annotated, or CpG ID if unannotated. The X-axis shows chromosomes 1-22 and the Y-axis shows -log10(p), with the horizontal red line denoting EWS ($P < 1.238 \times 10^{-7}$). (**b**) A Venn diagram highlighting overlapping DMPs at EWS across the cortical tissues. (**c**) In each cortical brain region the EWS DMPs identified in that region usually had a greater effect size there, than in any of the other cortical regions.

Fig 1. a

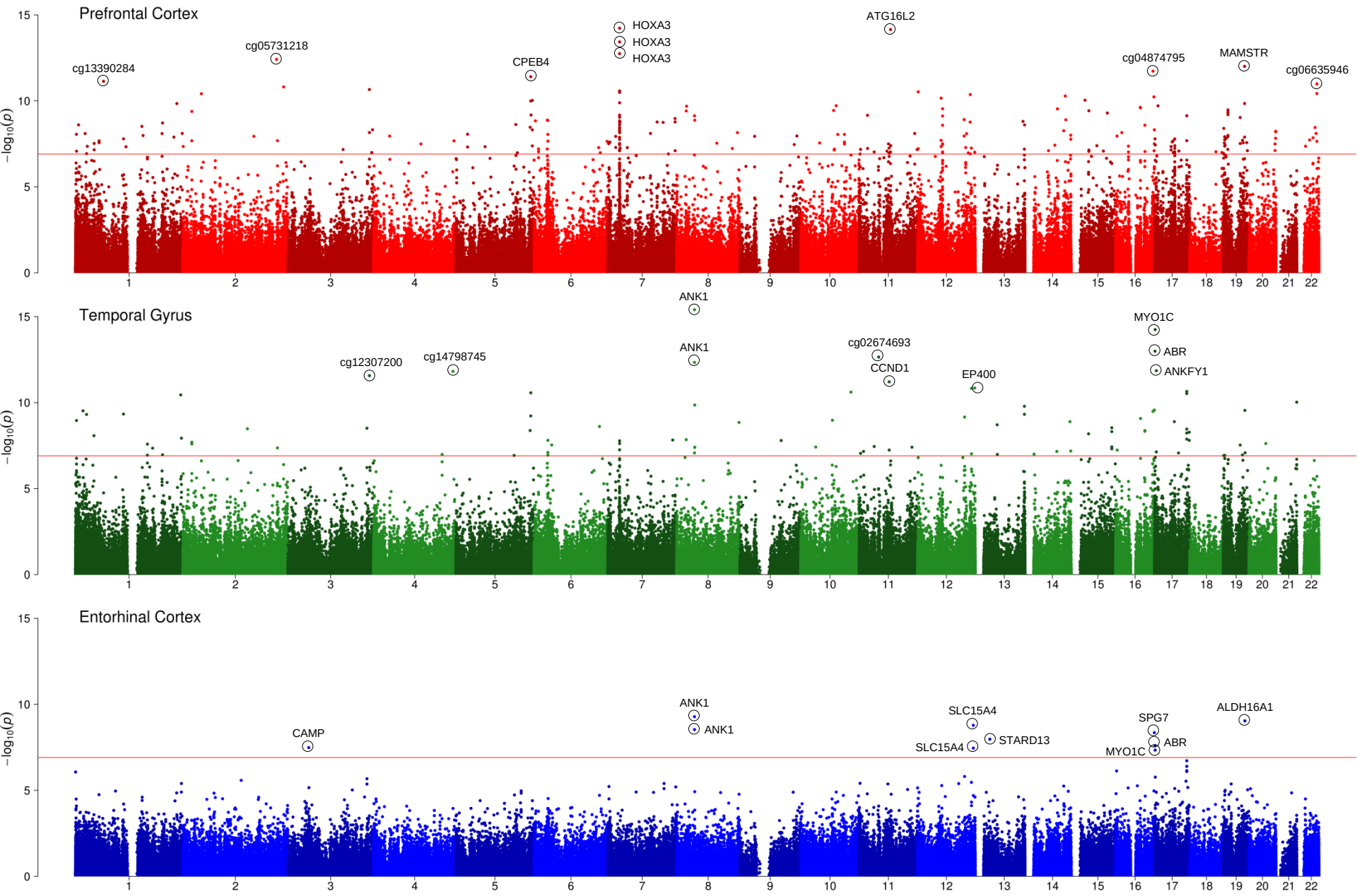

Prefrontal Cortex
cg13390284
cg05731218
CPEB4
HOXA3
HOXA3
HOXA3
ATG16L2
cg04874795
MAMSTR
cg06635946
Temporal Gyrus
cg12307200
cg14798745
ANK1
ANK1
cg02674693
CCND1
EP400
MYO1C
ABR
ANKFY1
Entorhinal Cortex
CAMP
ANK1
ANK1
SLC15A4
SLC15A4
STARD13
SPG7
MYO1C
ABR
ALDH16A1
$-\log_{10}(p)$

Fig 1. b

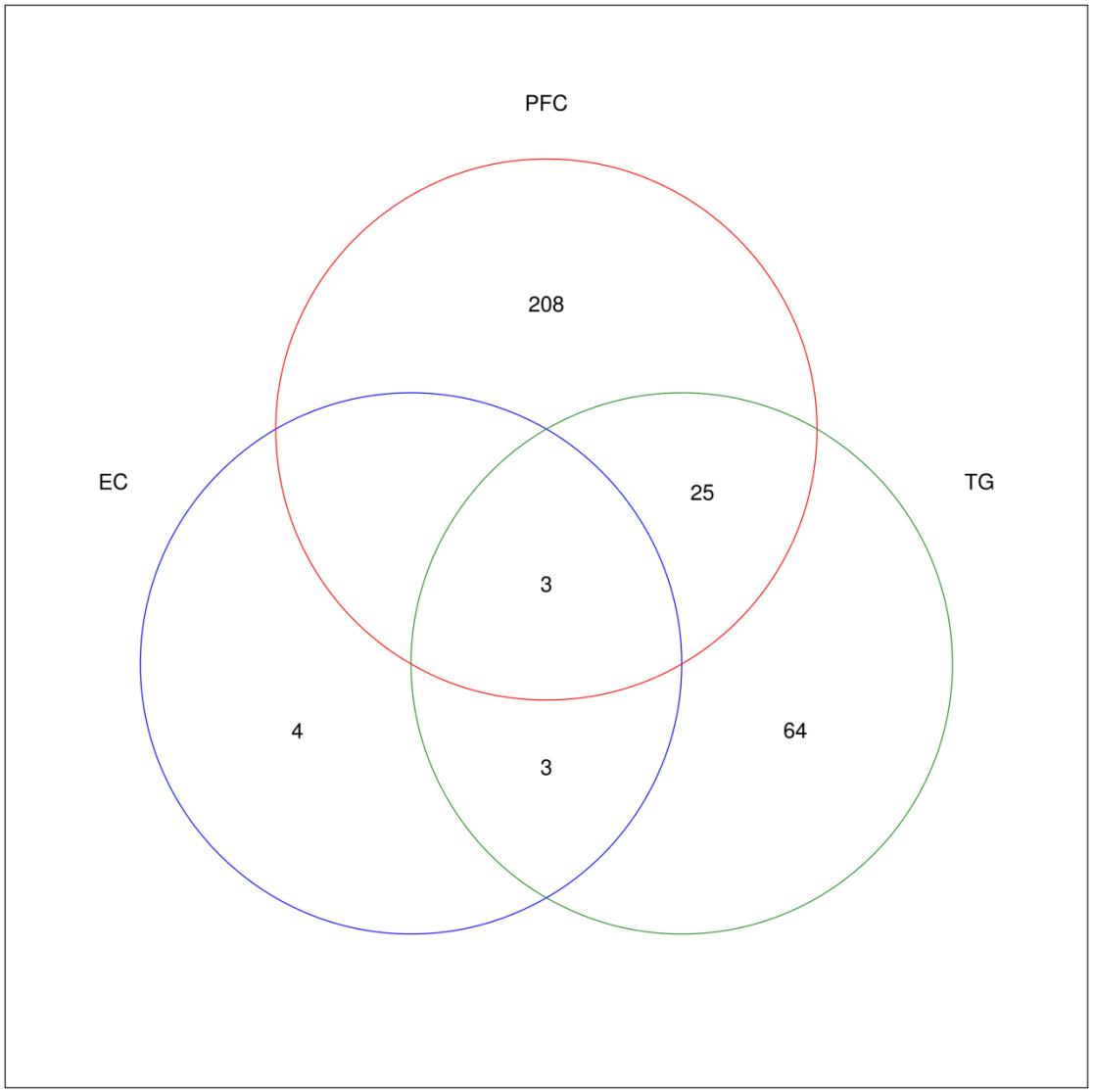


Fig 1. c

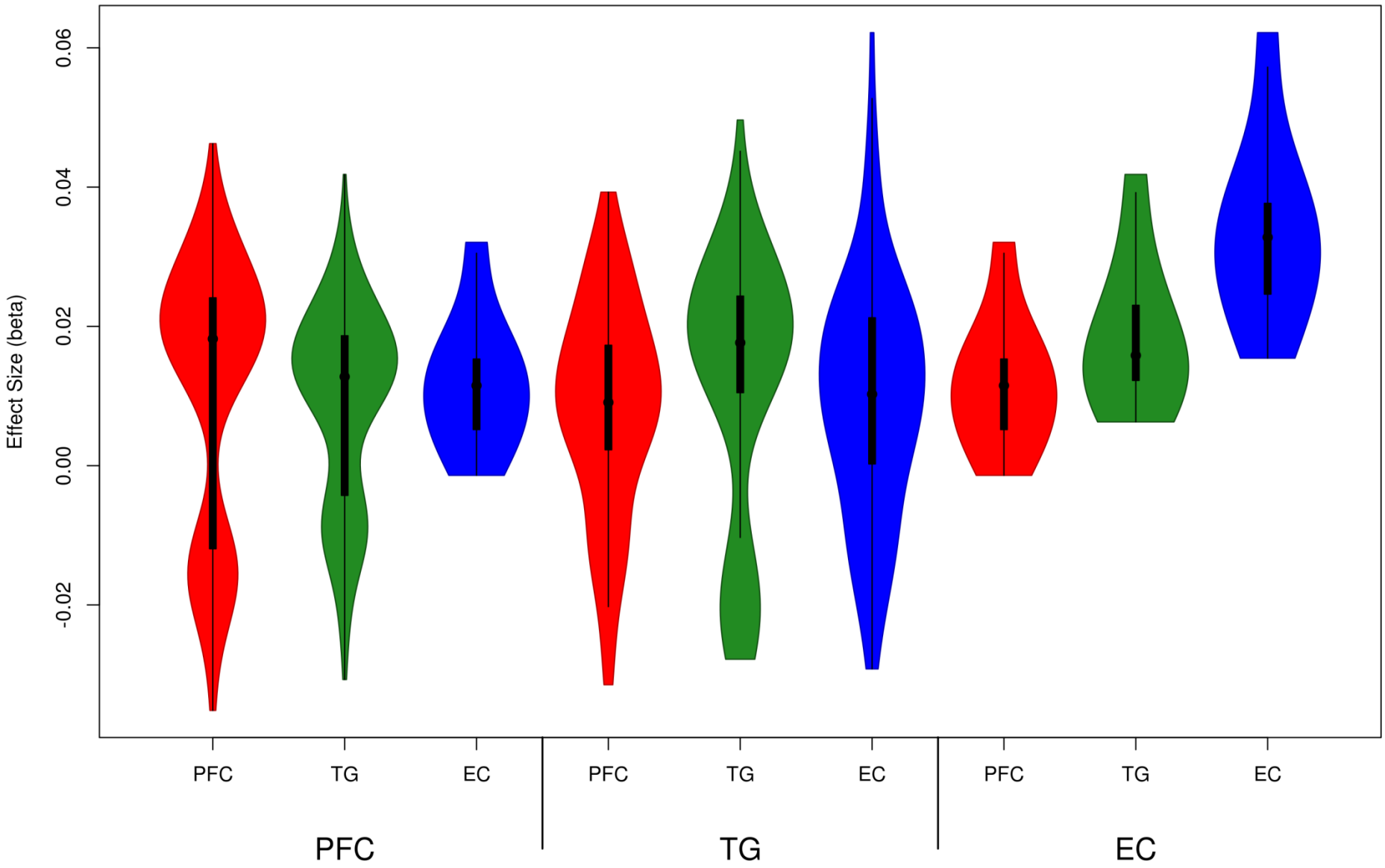

Effect Size (beta)
0.06
0.04
0.02
0.00
-0.02
PFC
TG
EC
PFC
TG
EC
PFC
TG
EC
PFC
TG
EC

Figure 2: A cross-cortex meta-analysis identifies 220 experiment-wide significant (EWS) differentially methylated positions (DMPs) associated with Braak stage. (a) A Miami plot of the cross-cortex meta-analyses. Probes shown above the X-axis indicate hypermethylation with higher Braak stage, whilst probes shown below the X-axis indicate hypomethylation with higher Braak stage. Along the X-axis data are shown the chromosome and genomic position. The Y-axis shows –log10(p). The red horizontal lines indicate EWS level of $P < 1.238 \times 10^{-7}$. Probes with an effect size (ES: difference between Braak 0- Braak VI) ≥ 0.01 and $P < 1.238 \times 10^{-7}$ are shown in blue. The 20 most significant DMPs are circled on the plot and Illumina UCSC gene name is shown if annotated, or CpG ID if unannotated. (**b**) A volcano plot showing the ES (X-axis) and –log10(p) (Y-axis) for the cross-cortical meta-analysis results. Gray probes indicate an ES difference between ≥ 0.01, whilst blue probes indicate an ES difference ≥ 0.01 and $P < 1.238 \times 10^{-7}$. (**c**) The most significant cross-cortex differentially methylated region (DMR) (chr7:27153212-27154305) contained 11 probes and resided in the *HOXA* region. The horizontal red line denotes the EWS level of $P < 1.238 \times 10^{-7}$. Red probes represent a positive effect size (ES) ≥ 0.01, blue probes represent a negative ES ≥ 0.01. Underneath the gene tracks are shown in black with CpG islands in green. (**d**) The ES of the 220 cross-cortex EWS DMPS identified in the discovery cohorts (X-axis) were significantly correlated with the ES in the replication cohort in the PFC (red, $r = 0.64$, $P = 5.24 \times 10^{-27}$), sorted neuronal cells (NeuN+: light blue, $r = 0.45$, $P = 1.56 \times 10^{-12}$) and non-neuronal cells (NeuN-: purple, $r = 0.79$, $P = 1.43 \times 10^{-47}$) (Y-axis). (**e**) A forest plot of the most significant cross-cortex DMP (cg12307200, chr3:188664632, $P = 4.48 \times 10^{-16}$). The effect size is shown in the PFC (red), TG (green) and EC (blue) for the different discovery cohorts. The X-axis shows the beta ES, with dots representing ES and arms indicating standard error (SE). ES from the intra-tissue meta-analysis using all available individual cohorts are represented by polygons in the corresponding tissue color. The black polygon represents the cross-cortex data. Shown in purple on the plot is the ES in the replication cohort in the PFC and sorted neuronal cells and non-neuronal cells, with the direction of effect suggesting the hypomethylation seen in the discovery cohorts is driven by changes in non-neuronal cells.

Fig 2. a

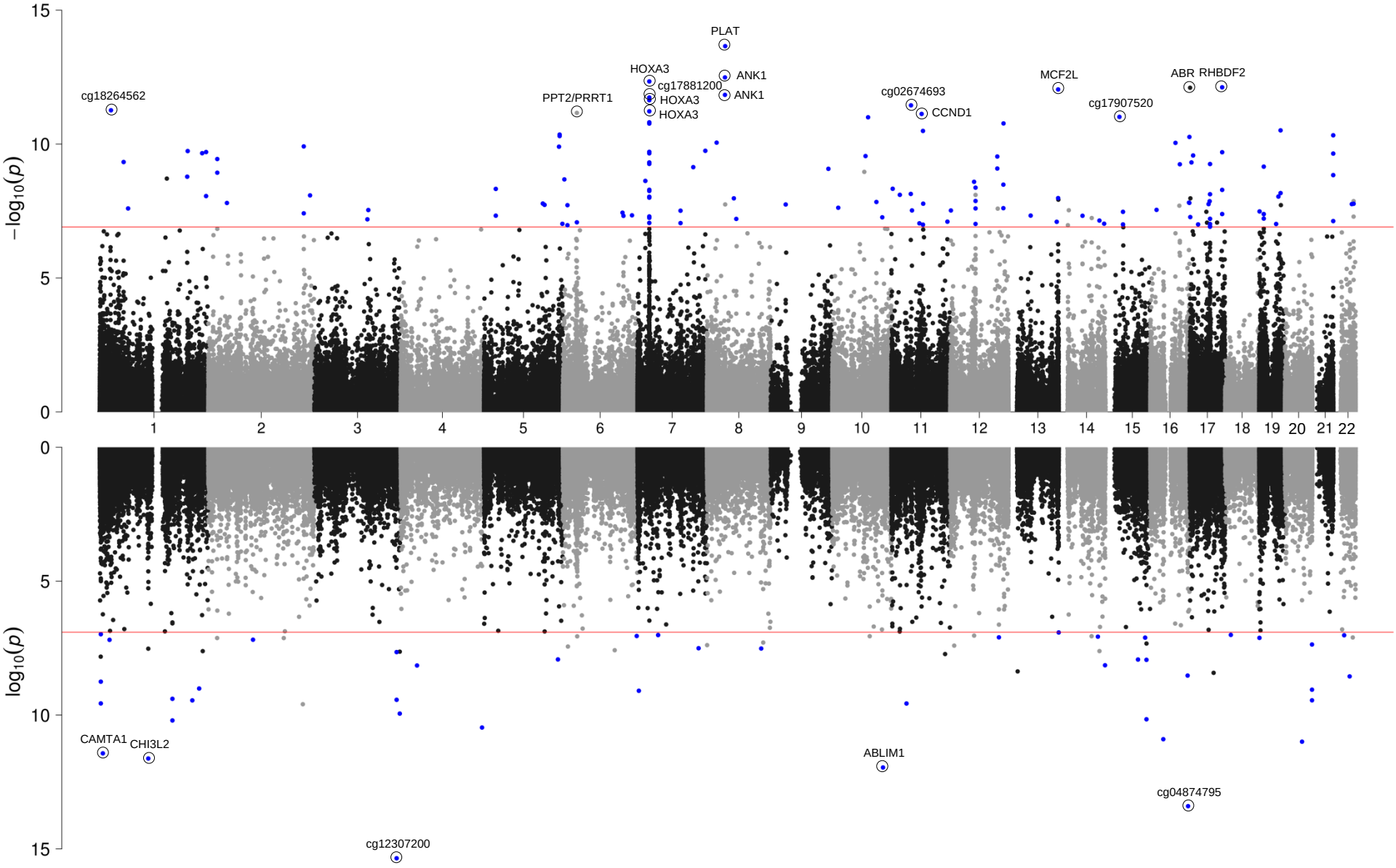

15
10
5
0
$-\log_{10}(p)$
cg18264562
PPT2/PRRT1
HOXA3
cg17881200
HOXA3
HOXA3
PLAT
ANK1
ANK1
cg02674693
CCND1
MCF2L
cg17907520
ABR
RHBDF2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
0
5
10
15
$\log_{10}(p)$
CAMTA1
CHI3L2
cg12307200
ABLIM1
cg04874795

Fig 2. b

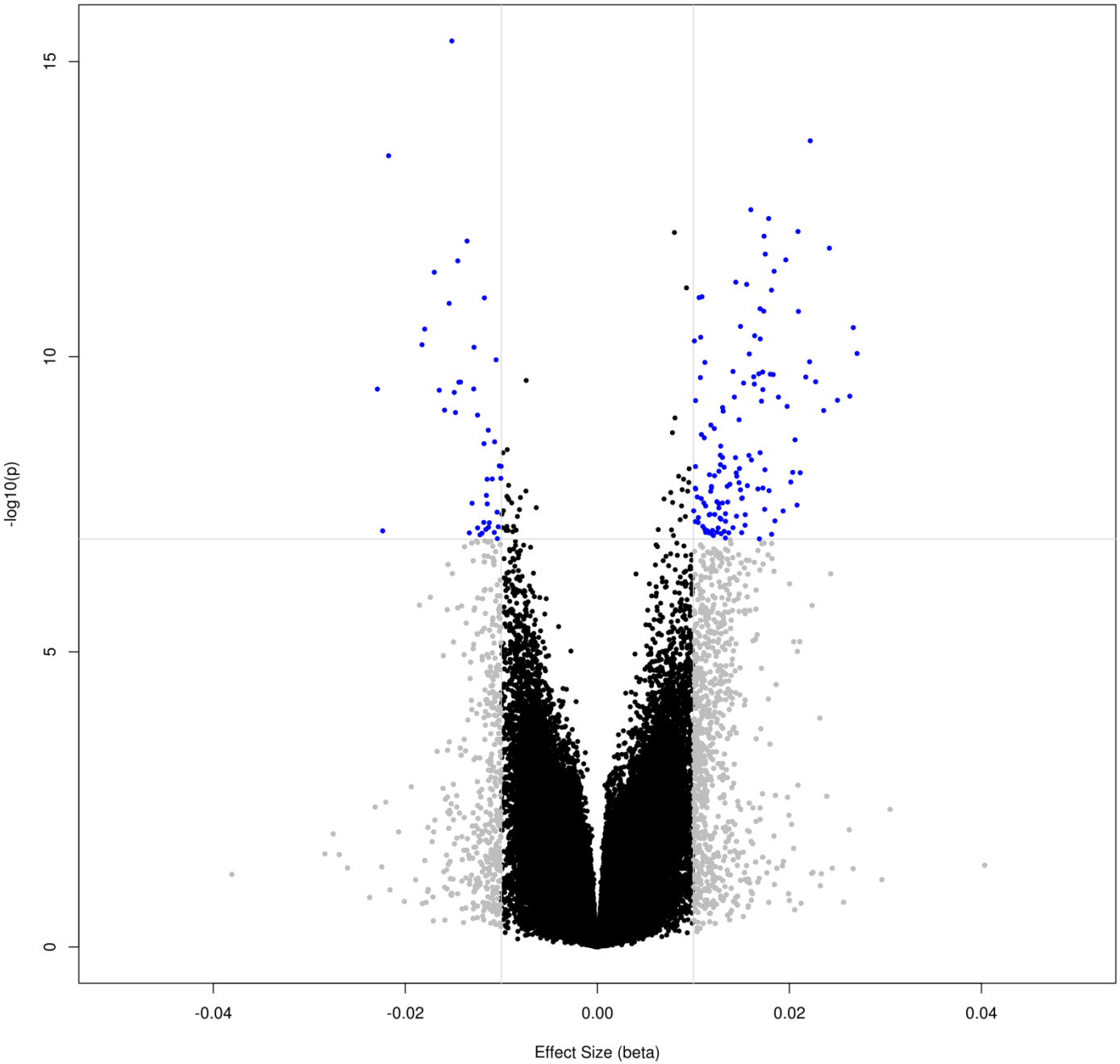


Fig 2. c

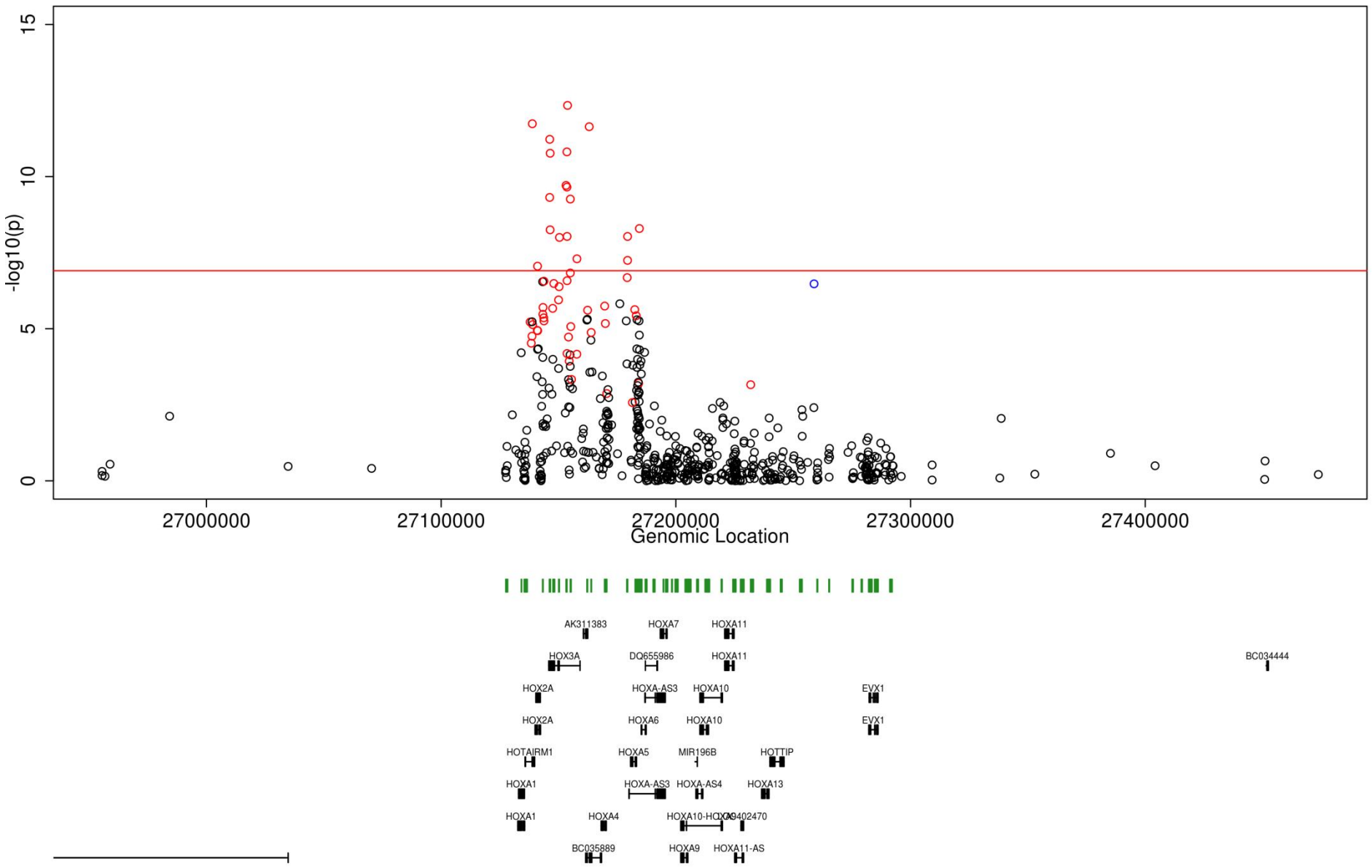

-log10(p)
0
5
10
15
27000000
27100000
27200000
27300000
27400000
Genomic Location
AK311383
HOXA7
HOXA11
HOX3A
DQ655986
HOXA11
BC034444
HOX2A
HOXA-AS3
HOXA10
EVX1
HOX2A
HOXA6
HOXA10
EVX1
HOTAIRM1
HOXA5
MIR196B
HOTTIP
HOXA1
HOXA-AS3
HOXA-AS4
HOXA13
HOXA1
HOXA4
HOXA10-HOXA9
AK402470
BC035889
HOXA9
HOXA11-AS

Fig 2. d

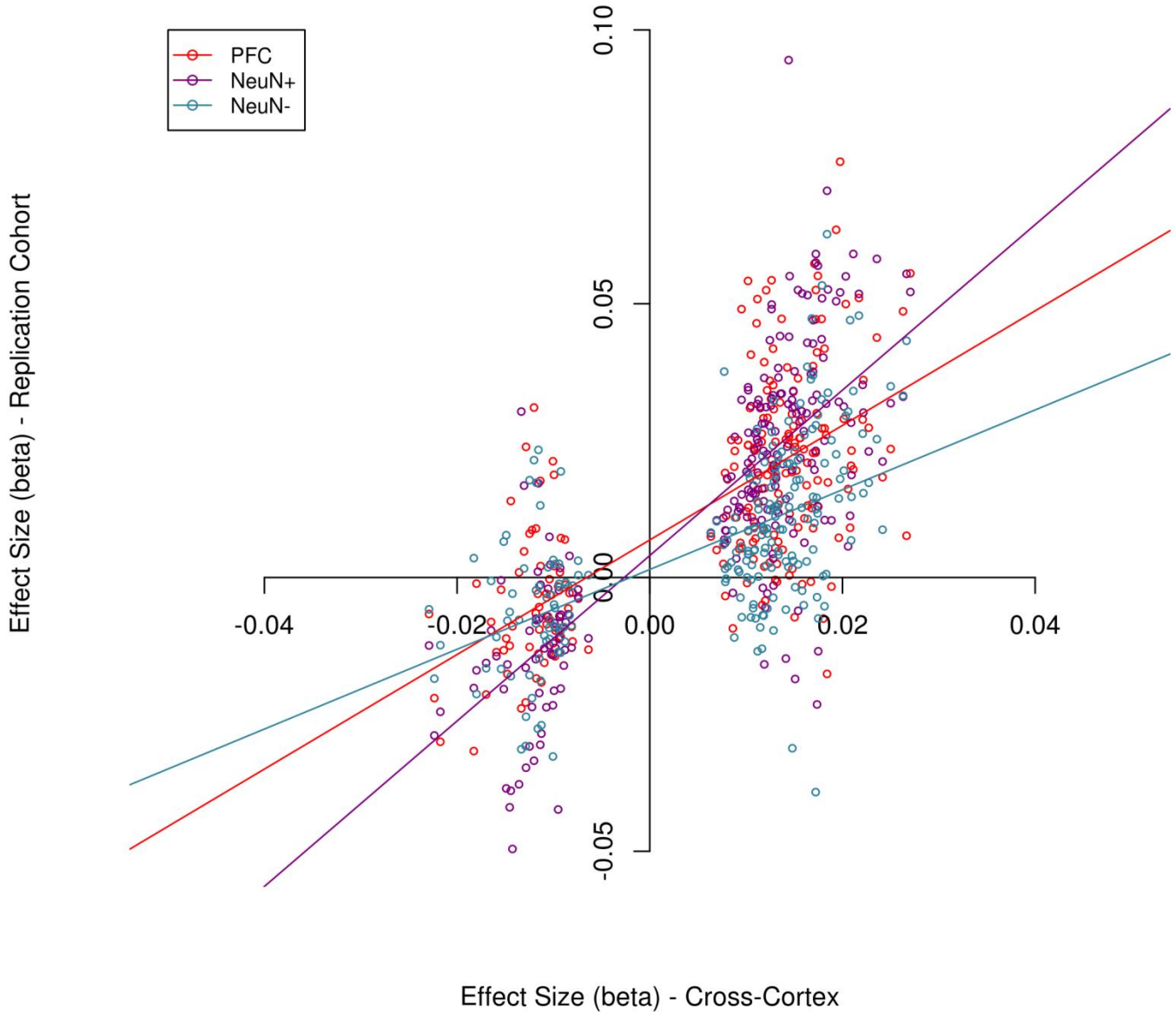

PFC
NeuN+
NeuN-
0.10
0.05
0.00
-0.05
-0.04
-0.02
0.00
0.02
0.04
Effect Size (beta) - Replication Cohort
Effect Size (beta) - Cross-Cortex

Fig 2. e

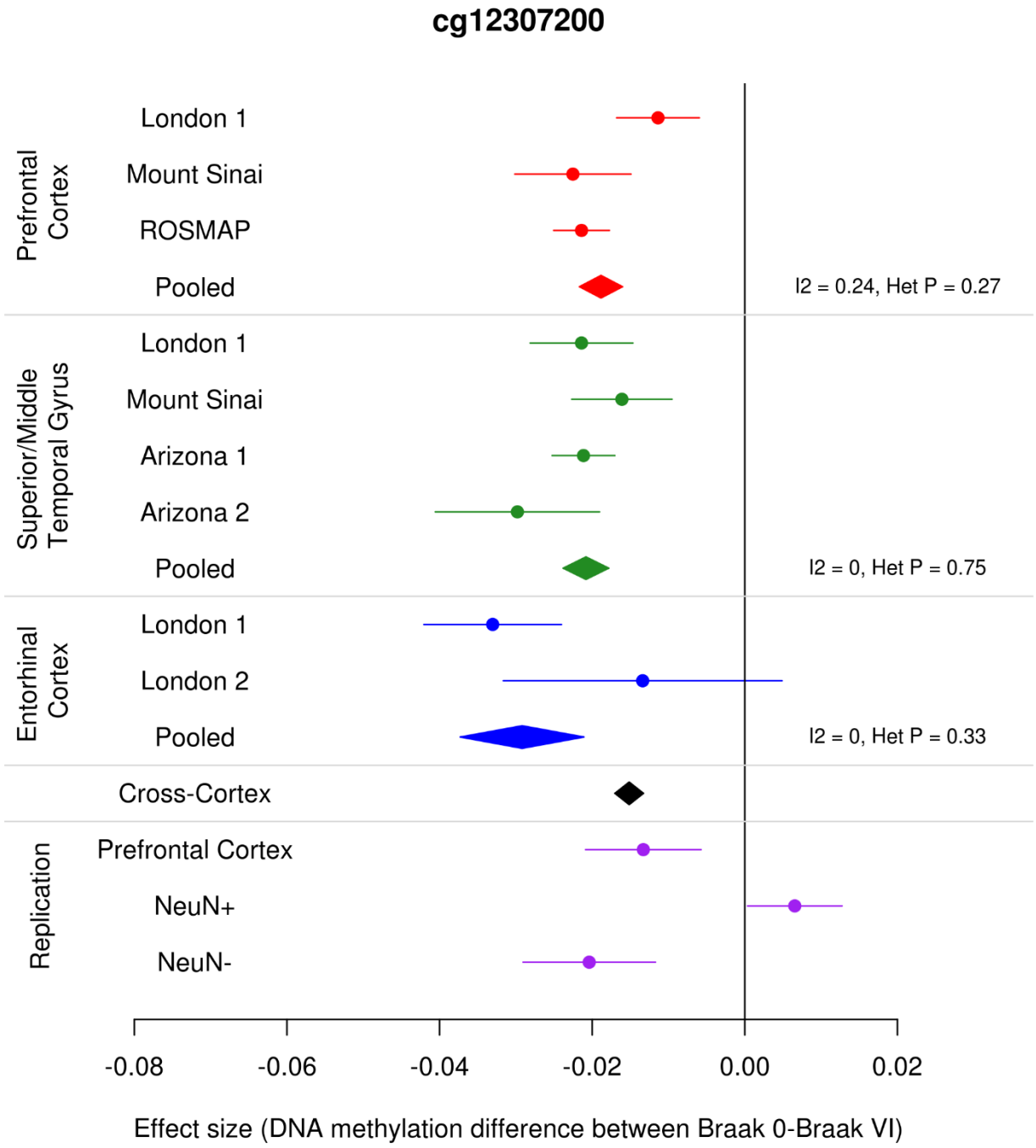
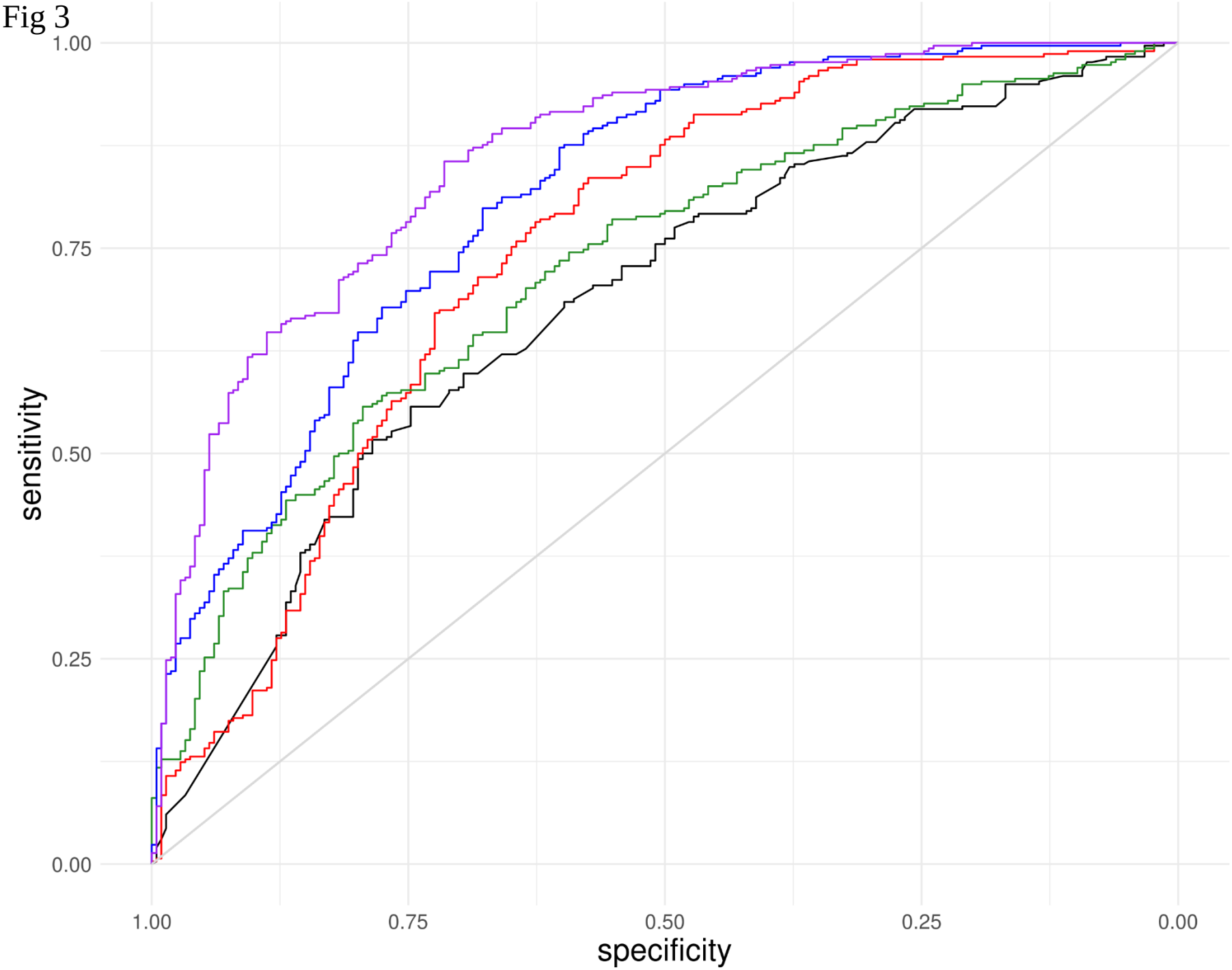

cg12307200
Prefrontal Cortex
London 1
Mount Sinai
ROSMAP
Pooled
I2 = 0.24, Het P = 0.27
Superior/Middle Temporal Gyrus
London 1
Mount Sinai
Arizona 1
Arizona 2
Pooled
I2 = 0, Het P = 0.75
Entorhinal Cortex
London 1
London 2
Pooled
I2 = 0, Het P = 0.33
Cross-Cortex
Replication
Prefrontal Cortex
NeuN+
NeuN-
-0.08
-0.06
-0.04
-0.02
0.00
0.02
Effect size (DNA methylation difference between Braak 0-Braak VI)

Figure 3: Receiver operating curves (ROC) graphs highlighting the predictive power of polyepigenic scores (PES) compared to polygenic risk scores (PRS) for explaining variance in pathology in the cortex. The PES (red line) generated from the 220 EWS probes was able to explain 24.7% (area under the curve [AUC] = 74.7%) of pathological variance in all samples examined with low pathology (Braak 0-II: "control", N = 214) or high pathology (Braak V-VI: "AD", N = 298) samples. In the same samples age + sex (black line) was able to explain 13.1% of pathology variance (AUC = 68.3%), whilst the polygenic risk score (PRS) was able to explain 20.2% of AD variance (green line: AUC = 72.6%) and together the PES and PRS were able to explain 38.4% of AD variance (blue line: AUC = 81.2%), which increased to 48.6% when age and sex were also included in the model (purple line: AUC = 86.1%). Full details on AUC, confidence intervals (CI) and Nagelkerke's pseudo R^2 (NK R^2) can be found in **Supplementary Table 12**.

Fig 3


1.00
0.75
0.50
0.25
0.00
sensitivity
1.00
0.75
0.50
0.25
0.00
specificity