ARTICLE

Meta-analysis of epigenome-wide association studies in Alzheimer's disease highlights novel 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], Danielle Brokaw[3], Paul T. Francis[1], Alan J Thomas[13], Seth Love[14], Kevin Morgan[15] Jörn Walter[4], Paul D. Coleman[3], David A. Bennett[16], Philip L. De Jager[17,18], 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] Institute of Neuroscience, Newcastle University, Newcastle Upon Tyne, UK

[14] Dementia Research Group, Institute of Clinical Neurosciences, School of Clinical Sciences, University of Bristol, Bristol, UK

[15] Human Genetics Group, University of Nottingham, Nottingham, UK

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

[17] Center for Translational and Computational Neuroimmunology, Department of Neurology and Taub Institute, Columbia University Medical Center, New York, USA.

[18] 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 3 South, 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 DNA 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. We identified 236 CpGs in the prefrontal cortex, 95 CpGs in the temporal gyrus and ten CpGs in the entorhinal cortex at Bonferroni significance, 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 84 genes had not been previously reported at this significance threshold. We have replicated our findings using two further DNA methylomic datasets consisting of a > 600 further unique donors. 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 protein and intracellular neurofibrillary tangles 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[5-12]. Independently these studies have identified 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 existing EWAS of AD brain[5-8,10,12], totalling 1,453 independent donors, to identify robust and consistent differentially methylated loci associated with Braak stage, used as a measure of neurofibrillary tangle spread through the brain, before replicating these signatures in two further independent DNA methylation datasets. 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 identified in the smaller independent EWAS.

RESULTS

Pathology-associated DNA methylation signatures in discrete cortical brain regions

We identified six EWAS of DNA methylation in AD that had been generated using the 450K array and had 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 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: three cohorts, N = 959; temporal gyrus: four cohorts, N = 608, entorhinal cortex: two cohorts, N = 189 cerebellum: four cohorts, N = 533) (**Supplementary Figure 1**).

The prefrontal cortex represented our largest dataset (N = 959 samples) and we identified 236 Bonferroni significant differentially methylated positions (DMPs), of which 193 were annotated to 137 genes, with 43 unannotated loci based on Illumina UCSC annotation (**Figure 1a**, **Supplementary Figure 2, Supplementary Table 1**). Previous EWAS of the prefrontal cortex have consistently reported the *HOXA* gene cluster as a region that is hypermethylated in AD[6,7], with a cell-type specific EWAS demonstrating this is neuronal-derived[11]. Indeed, the most significant DMP in the prefrontal cortex 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 x 10^{-15}), with a further 16 of the Bonferroni significant DMPs also annotated to this gene. This locus appeared to be particularly hypermethylated with higher Braak stage in the prefrontal cortex, and to a slightly lesser extent in the temporal gyrus (**Supplementary Figure 3**). There was no significant difference in methylation at this locus in the entorhinal cortex (P = 0.864), which is interesting given that the entorhinal cortex may succumb to pathology early in the disease process (Braak stage III). Of the 236 prefrontal cortex DMPs, 92% (217 probes) were nominally significant (P < 0.05) in the temporal gyrus, of which 12% (28 probes) were Bonferroni significant, whilst 9% (22 probes) were nominally significant in the entorhinal cortex, with 1% (3 probes) reaching Bonferroni significance (**Figure 1b**). The effect sizes for the 236 Bonferroni significant prefrontal cortex DMPs were correlated with the effect sizes for the same probes in both the temporal gyrus (Pearson's correlation coefficient (r) = 0.94, P = 6.17 x 10^{-112}) and entorhinal cortex (r = 0.58, P = 1.80 x 10^{-22}) and were enriched for probes with the same direction of effect (sign test: temporal gyrus P = 5.07 x 10^{-67}, entorhinal cortex P = 6.88 x 10^{-26})

(**Supplementary Figure 4**)**.** For the 236 Bonferroni significant prefrontal cortex DMPs these had the largest effect sizes in the prefrontal cortex, with a smaller effect size in the temporal gyrus and entorhinal cortex (**Figure 1c**). Of these 236 DMPs, 29 of these had being previously reported at Bonferroni significance in previous publications on the individual cohorts[5-7,12], including one probe annotated to *ANK1*, one probe annotated to *HOXA3*, one probe annotated to *PPT2/PRRT1* and two probes annotated to *RHBDF2*, amongst others. However, our approach has identified 207 novel Bonferroni significant DMPs (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 identified (from another probe) in earlier studies, for example a further 16 probes in *HOXA3* and two probes in *PPT2/PRRT1*. Interestingly, we also identified a number of novel genes, including some which featured multiple Bonferroni significant DMPs including for example seven probes in *AGAP2* and five probes in *SLC44A2*, amongst others. One other noteworthy novel Bonferroni significant DMP in the prefrontal cortex was cg08898775 (ES = 0.019, P = 4.03×10^{-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 ≥ 2 DMPs, revealed 262 significant DMRs in the prefrontal cortex (**Supplementary Table 2**), the most significant containing 20 probes and located in *HOXA3* (chr7:27,153,212-27,155,234: Sidak-corrected p = 8.21×10^{-50}, **Supplementary Figure 5**), as well as several other DMRs in the *HOXA* gene cluster.

A meta-analysis of temporal gyrus EWAS datasets (N = 608 samples) identified 95 Bonferroni significant probes, of which 75 were annotated to 53 genes, with 20 unannotated probes using Illumina UCSC annotation (**Figure 1a**, **Supplementary Figure 6**, **Supplementary Table 3**). The most significant probe was cg11823178 (ES = 0.029, P = 3.97×10^{-16}, **Supplementary Figure 7**), which is annotated to the *ANK1* gene, with the fifth (cg05066959: ES = 0.042, P = 4.58×10^{-13}) and 82nd (cg16140558: ES = 0.013, P = 8.44×10^{-8}) most significant probes in the temporal gyrus also being annotated to nearby CpGs in that gene. This locus has been widely reported to be hypermethylated in AD from prior EWAS[5,6,8,12], as well as in other neurodegenerative diseases such as Huntington's disease and Parkinson's disease[13]. Another noteworthy gene is *RHBDF2*, where five Bonferroni significant DMPs in the temporal gyrus were annotated to (cg05810363: ES = 0.029, P = 2.25×10^{-11}; cg13076843: ES = 0.031, P = 2.97×10^{-11}; cg09123026: ES = 0.012, P = 3.46×10^{-9}; cg12163800: ES = 0.025, P = 5.85×10^{-9}; cg12309456: ES = 0.016, P = 1.33×10^{-8}); and

which has been highlighted in previous EWAS in AD in the individual cohorts[5,6,12]. Of the 95 Bonferroni significant DMPs in the temporal gyrus, 88% (84 probes) were nominally significant in the prefrontal cortex, of which 29% (28 probes) were Bonferroni significant, whilst 54% (51 probes) were nominally significant in the entorhinal cortex, of which 6% (6 probes) were Bonferroni significant (**Figure 1b**). Given the high degree of overlapping significant loci between the temporal gyrus and other cortical regions, it was not surprising that the ES of the 95 Bonferroni significant temporal gyrus probes were highly correlated with the ES of the same loci in both the prefrontal cortex ($r = 0.91$, $P = 5.09 \times 10^{-38}$) and entorhinal cortex ($r = 0.77$, $P = 4.02 \times 10^{-20}$) and were enriched for the same direction of effect (sign test: prefrontal cortex $P = 5.05 \times 10^{-29}$, entorhinal cortex $= 2.30 \times 10^{-25}$) (**Supplementary Figure 8**). The majority of the 95 Bonferroni significant DMPs in the temporal gyrus were hypermethylated, and the mean ES was greater in the temporal gyrus than the prefrontal cortex or entorhinal cortex (**Figure 1c**). Thirty-two of the 95 Bonferroni significant DMPs in the temporal gyrus have been previously reported to be significantly differentially methylated in published EWAS, including for example three probes in *ANK1* and the five probes in *RHBDF2*. Our meta-analysis approach in the temporal gyrus has identified 63 novel DMPs (at Bonferroni significance), including some novel genes with multiple DMPs, for example four probes in *RGMA* and two probes in *CCND1*, 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 $P = 1.72 \times 10^{-21}$) (**Supplementary Figure 9**). The five DMPs in *RHBDF2* that we already highlighted also represented a significant DMR (Sidak-corrected $P = 8.47 \times 10^{-21}$), with three other genomic regions containing large, significant DMRs consisting of $\geq$ 10 probes, such as *MCF2L* (chr13:113698408-113699016 [10 probes], Sidak-corrected $P = 1.16 \times 10^{-19}$), *PRRT1/PPT2* (chr6:32120773-32121261 [17 probes], Sidak-corrected $P = 4.90 \times 10^{-15}$) and *HOXA5* (chr7:27184264-27184521 [10 probes], Sidak-corrected $P = 1.60 \times 10^{-7}$).

The final cortical region we had available was the entorhinal cortex (N = 189), where we identified ten Bonferroni significant 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 (Illumina UCSC annotation), with two Bonferroni significant probes residing in each of the *ANK1* and *SLC15A4* genes. As with the temporal gyrus, the most significant DMP was cg11823178 ($ES = 0.045$, $P = 5.22 \times 10^{-10}$, **Supplementary Figure 7**), located within the *ANK1* gene, with the

fourth most significant DMP being located within 100bp of that CpG (cg05066959: ES = 0.062, P = 2.93 x 10^{-9}). In total, eight of the ten DMPs in the entorhinal cortex had been reported previously at Bonferroni significance, including the two probes in *ANK1*. Two of the Bonferroni significant DMPs we identified in the entorhinal cortex were novel CpGs (cg11563844: *STARD13*, cg04523589: *CAMP*), having not been reported as Bonferroni significant in previous EWAS. Of the ten entorhinal cortex probes, 90% (9 probes) were nominally significant in the temporal gyrus, of which 60% (6 probes) were Bonferroni significant, whilst 70% (7 probes) were nominally significant in the prefrontal cortex, of which 30% (3 probes) were Bonferroni significant (**Figure 1b**). Of the four DMPs that were Bonferroni significant in only the entorhinal cortex, three of these were nominally significant in at least one other tissue, with just one probe unique to the entorhinal cortex, annotated to *STARD13* (cg11563844, ES = 0.027, P = 1.07 x 10^{-8}). The effect sizes of the ten Bonferroni significant DMPs in the entorhinal cortex were significantly correlated with the effect size of the same probes in the prefrontal cortex (r = 0.74, P = 0.01) and temporal gyrus (r = 0.85, P = 1.52 x 10^{-3}) and were enriched for the same direction of effect (sign test: prefrontal cortex P = 0.021, temporal gyrus P = 1.95 x 10^{-3}) (**Supplementary Figure 11**). The ten DMPs were hypermethylated in all three cortical regions, with the greatest Braak-associated ES in the entorhinal cortex (**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 cerebellum and none of these have reported any Bonferroni significant DMPs. In our meta-analysis we identified no Bonferroni significant DMPs, nor any DMRs in the cerebellum (**Supplementary Figure 13**), despite this analysis including 533 independent samples. There was no correlation of the ES for the Bonferroni significant DMPs we had identified in the meta-analyses of the three cortical regions with the ES of the same probes in the cerebellum (prefrontal cortex: r = 0.11, P = 0.08; temporal gyrus: r = 0.14, P = 0.17; entorhinal cortex: 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. As multiple cortical tissues were available for some cohorts, a mixed-effects model was utilized. In this analysis we controlled for age, sex and neuron/glia proportion, with surrogate variables added as appropriate to reduce inflation. Using this approach, we identified 220 Bonferroni significant probes, of which 168 were annotated to 121 genes, with 52 DMPs unannotated using Illumina UCSC annotation. **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 probes with 219 of the 220 DMPs showing the same direction of effect in at least five cohorts. In total, 154 of the DMPs were hypermethylated, with 66 hypomethylated, representing an enrichment for hypermethylation ($P = 4.85 \times 10^{-10}$). This pattern of methylation was evident across all cortical tissues but was not seen in the cerebellum (**Supplementary Figure 17**). Of the 220 DMPs we identified, 46 of these have been previously reported at Bonferroni significance in published EWAS, including multiple previously identified 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, $P = 4.48 \times 10^{-16}$), which is intergenic and found at chr3:188664632, located between the *TPRG1* and *LPP* genes and had been previously reported at Bonferroni significance by De Jager and colleagues with respect to neuritic plaque burden[6] and by Brokaw and colleagues with respect to post-mortem diagnosis[12]. Our cross-cortex meta-analysis approach has identified 174 novel DMPs (at Bonferroni significance), annotated to 102 genes. Although 11 of these genes had previously been reported at Bonferroni significance (another probe within that gene), the remaining 96 genes represent robust novel loci in AD. Many of these novel differentially methylated genes had multiple Bonferroni significant 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 these were at Bonferroni significance and so here represent robust novel loci.

We were interested to investigate whether specific functional pathways were differentially methylated in AD cortex and so performed a gene ontology pathway analysis of the 121 genes annotated to the 220 Bonferroni significant cross-cortex DMPs. We highlighted epigenetic dysfunction in numerous pathways, interestingly including a number of developmental pathways, mainly featuring the *HOXA* and *HOXB* gene clusters (**Supplementary Table 8**). Given that we identified multiple DMPs in some genes, we were interested to investigate the correlation structure between probes in close proximity to each other to establish how many independent signals we had identified. Using a method developed to identify single nucleotide polymorphisms (SNPs) in linkage disequilibrium (LD)[14], we collapsed the 220 Bonferroni significant loci into 198 independent (non-highly correlated [$r < 0.8$]) signals. We found that the 18 DMPs in the *HOXA* region represented only seven independent signals, whilst the four DMPs in the *RHBDF2* gene and the three DMPs in the *SLC44A2* gene represented just one signal each. Similarly, the two DMPs in each of the *DUSP27*, CPEB4, *GCNT2*, *ANK1* and *MAMSTR* genes represented a single independent signal each, whilst the five DMPs in *AGAP2* was reduced to four independent signals. Next we undertook a formal regional analysis to identify genomic regions of multiple adjacent DMPs and identified 221 DMRs, with the top DMR containing 11 probes and covering the *HOXA* region (chr7:27,153,212-27,154,305: $P = 3.84 \times 10^{-35}$) (**Figure 2c**, **Supplementary Table 9**). The *HOXA* gene cluster further 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: $P = 4.11 \times 10^{-27}$), 33 probes (chr7:27183133-27184667: $P = 2.22 \times 10^{-20}$) and ten probes (chr7:27143235-27143806: $P = 1.75 \times 10^{-18}$).

Replication of pathology associated DMPs in the cortex

To replicate our findings and to determine the cellular origin of DNA methylomic differences 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 prefrontal cortex (N = 45) and sorted neuronal and non-neuronal nuclei from the occipital cortex (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 Bonferroni significant cross-cortex probes in this replication cohort, with a significantly correlated effect size between the discovery dataset and the replication prefrontal cortex ($r = 0.64$, $P = 5.24 \times 10^{-27}$), neuronal ($r = 0.45$, $P = 1.56 \times 10^{-12}$) and non-neuronal datasets ($r = 0.79$, $P = 1.43 \times 10^{-47}$) with a similar enrichment for the same direction of effect (sign test: prefrontal cortex $P = 4.59 \times 10^{-28}$, neuronal $P = 6.13 \times 10^{-15}$, non-neuronal $P = 1.06 \times 10^{-42}$) (**Figure 3a**). 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 prefrontal cortex 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 3b**). We have developed an online database (www.epigenomicslab.com/ad-meta-analysis/), which can generate a forest plot showing the ES and SE across any of the discovery cohorts and the Munich sample types for any of the 403,763 probes that passed our quality control. This allows researchers to determine the consistency of effects across cohorts for a given CpG site as well as the likely cellular origin of the signature. In addition, our tool can generate mini-Manhattan plots to show DMRs utilizing the summary statistics from the cross-cortex meta-analysis.

Finally, we had access to DNA methylation data generated in an eighth independent ("Brains for Dementia Research [BDR]") cohort. This consisted of Illumina Infinium HumanMethylation EPIC BeadChip (EPIC array) data in the prefrontal cortex in 590 individuals[15]. As this is the successor to the 450K array (which had been used for the other seven cohorts), there are some differences in genome coverage, and for the 220 Bonferroni significant cross-cortex DMPs we had identified in the discovery cohorts, only 208 probes are also present on the EPIC array. For these overlapping 208 probes, we observed a significantly correlated effect size between the discovery dataset and the BDR dataset ($r = 0.53$, $P = 4.13 \times 10^{-16}$) (**Figure 3c**), with all 208 probes showing the same direction of effect (sign test $P = 4.86 \times 10^{-63}$).

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

To identify if the cross-cortex DMPs reside in specific genomic features, we used a Fisher's exact test to look for an enrichment of the 220 DMPs using Slieker annotations[16] (**Supplementary Table 10**, **Supplementary Figure 18**). We observed a significant over representation of Bonferroni significant DMPs in CpG islands of gene bodies (odds ratio [OR] = 3.199, $P = 4.76 \times 10^{-10}$), and in CpG island shelves and non-CpG island 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 CpG islands 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 DMPs in the first exon (OR = 1.80, P = 0.02), with an under enrichment within 1500bp of the transcription start site (OR = 0.49, P = 3.82 x 10^{-3}) (**Supplementary Table 11**, **Supplementary Figure 19**).

DNA methylomic signatures in the cortex can explain variance in the degree of pathology

We were interested to investigate whether the Braak-associated DNA methylation patterns we had identified across the cortex could accurately predict the pathological load of a brain sample and how much variance this explained. To this end we took samples with either low pathology (Braak 0-II "controls": N = 386), or high pathology (Braak V-VI "AD": N = 543) and divided these in to 75% "training" and 25% "testing datasets. We then used elastic net regression to identify 95 probes in the 220 cross-cortex Bonferroni significant loci (**Supplementary Table 12)** that were able to explain the most variance between post-mortem low pathology "control" from high pathology "AD" status in our training dataset (N = 696) (**Supplementary Table 13**, **Figure 4**). In our training data, we achieved an Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) of 94.36% (CI = 92.67-95.88%, variance explained = 71.52%). When this was tested in the testing dataset (N = 233) it achieved an AUC of 87.63% (CI = 82.73-91.89%) and explained 52.39% of the variance. We then tested its performance further in the Munich replication samples (N = 38) and the BDR replication samples (N = 454), where it achieved an AUC of 75.1% (CI = 55.56-90.81%, variance explained = 25.47%) and 70.33% (CI = 65.32-74.93%, variance explained = 15.44%), respectively (**Supplementary Table 13**, **Figure 4**).

DNA methylation signatures in AD cortex are largely independent of genetic effects

DNA methylomic variation can be driven by genetic variation via methylation quantitative trait loci (mQTLs). To explore whether SNPs may be driving the methylation differences we observed (in *cis*) we used the xQTL resource to identify *cis*-mQTLs associated with the 220 Bonferroni significant cross-cortex DMPs[17]. We identified 200 Bonferroni corrected mQTLs, which were associated with DNA methylation at 18 of the 220 cross-cortex DMPs (**Supplementary Table 14**). This suggests that the majority of Braak-associated DMPs are not the result of genetic variation in *cis*. None of these mQTLs overlapped with lead SNPs (or

SNPs in LD) identified in the most recent genome-wide association study (GWAS) of diagnosed late-onset AD from Kunkle *et al*[18]. Next, we were interested in exploring whether DNA methylation is enriched in genes known to harbor AD-associated genomic risk variants. Using the AD variants from Kunkle *et al*[18] we examined the enrichment of Braak-associated DNA methylation in 24 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 published independent sample cohorts with a range of cortical brain regions and cerebellum available as a discovery dataset. Two further independent cortical datasets where then used for replication, including data from sorted nuclei populations. Our data can be explored as part of an online searchable database, which can be found on our website (https://www.epigenomicslab.com/ad-meta-analysis). By performing a meta-analysis within each tissue, we have been able to identify 236, 95 and ten Bonferroni significant DMPs in the prefrontal cortex, temporal gyrus and entorhinal cortex, respectively. Although far fewer loci were identified in the entorhinal cortex compared to the other cortical regions, this is likely due to the reduced sample size in this tissue. In the cerebellum despite meta-analyzing > 500 unique samples, we identified no Braak-associated DNA methylation changes. Furthermore, there was no correlation of the ES of Bonferroni significant DMPs identified in any of the cortical regions with the ES of the same probes in the cerebellum. Taken together, this suggests that DNA methylomic changes in AD are cortex cell type specific. This observation is interesting as the cerebellum is said to be "spared" from AD pathology, with an absence of neurofibrillary tangles, although some diffuse amyloid-beta plaques are reported[19]. Interestingly, a recent spatial proteomics study of AD reported a large number of protein changes in the cerebellum in AD; however, the proteins identified were distinct from other regions examined and thus the authors suggested a potential protective role[20].

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 Bonferroni significant probes associated with Braak stage of which 46 probes had been previously reported at Bonferroni significance 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 174 novel CpGs, corresponding to 102 unique genes, of which 84 genes had not been previously reported at Bonferroni significance 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 Bonferroni significant DMPs each, including five probes in *AGAP2*, three probes in *SLC44A2* and two probes each in *CDH9*, *CPEB4*, *DUSP27*, *GCNT2*, *MAMSTR*, *PTK6*, *RGMA*, *RHOB*, *SMURF1*, *THBS1*, *ZNF238* and *ZNF385A*. These genes had not been identified previously in an AD EWAS at this significance threshold, although a number of these genes had been previously identified from DMR analyses, which have a less stringent threshold. However, we did identify one novel gene (*HOXB3*) with three Bonferroni significant DMPs, which had not been identified at this significance threshold 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 prefrontal cortex[21], an interesting observation given that both AD and Huntington's disease are disorders that feature dementia. Furthermore, we have recently reported AD-associated hypermethylation of the *HOXB6* gene in AD blood samples[22]. Our pathway analysis highlighted differential methylation in a number of developmental pathways, mainly featuring the *HOXA* and *HOXB* gene clusters. Although it is unclear why developmental genes may be changed in a disease that primarily affects the elderly, it has been implied that genes such as these may be involved in neuroprotection after development[23]. A number of the other novel genes with multiple DMPs are also biologically relevant in the context of AD, for example *GCNT2* was recently shown to be differentially expressed in the Putamen between males and females with AD[24]. Interestingly, some of the protein products of genes we identified have also been previously linked with AD; PTK6 is a protein kinase whose activity has been shown to be altered in post-mortem AD brain[25].

Similarly, RGMA has been shown to be increased in AD brain, where it accumulated in amyloid-beta plaques[26].

Our genomic enrichment analyses identified an over representation of hypermethylated loci in AD and methylation in specific genomic features, for example CpG islands in gene bodies, and shelves and non-CpG island regions in proximal promoters. We demonstrated that the majority of DMPs we identified (N = 202) were not driven by genetic variation as only 18 of the 220 CpG sites have reported mQTLs. However, we did observe 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 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 Bonferroni significant probes, further highlighting that distinct genomic loci reproducibly show epigenetic dysfunction in AD cortex. Although the clinical utility of such a classifier is limited as it is developed in post-mortem cortical brain tissue, it does illustrate that specific robust patterns of DNA methylation differences occur as the disease progresses. These signatures require further investigation as they could represent novel therapeutic targets, particularly given the classifier had an AUC > 70% in all testing, training and replication datasets. However, it is worth noting that the variance explained by the 95 CpG signature was lower in the replication datasets than the discovery samples, which could be due to a low sample number (Munich) or the different Illumina array platform (BDR).

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 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 and used data from sorted cell populations as part of our replication. 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 the Munich replication cohort, or ideally at the level of the single cell. It is important to note that the data

from the sorted nuclei populations in the Munich replication cohort were generated in the occipital cortex, which was not a bulk tissue used for any of the discovery cohorts. In the future it would be interesting to explore whether different disease-associated DNA methylation signatures were observed in neurons and glia isolated from different cortical brain regions. Second, our study has utilized previously generated EWAS data generated on the 450K array or EPIC array. Although the Illumina array platform has been the most widely used platform for EWAS to date, 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 to substantially increase the coverage. In our study we have primarily used the UCSC annotation provided by Illumina to identify the gene relating to each DMP. However, this can lead to the annotation of overlapping genes, or no gene annotation, which can make it difficult to establish the gene of interest in the absence of functional studies. Our study has primarily focused on the results of a fixed-effects meta-analysis, as the majority of Bonferroni-significant DMPs do not display a high degree of heterogeneity. However, ~15% of the cross-cortex DMPs did have a significant heterogeneity P value and in this instance, it is worthwhile also considering the results of the random-effects meta-analysis. Although this heterogeneity could be driven by differences between cohorts, it is also plausible that it may be driven by tissue-specific effects as we used different cortical brain regions in the model. For example cg22962123 annotated to the *HOXA3* gene has a significant heterogeneity P value in the cross-cortex meta-analysis, but we had already shown this loci to be differentially methylated in the prefrontal cortex and temporal gyrus, but not the entorhinal cortex in our intra-tissue meta-analysis. Another limitation of our study is that we have focused our analyses on Braak (neurofibrillary tangle)-associated methylation changes, as this measure was available in all cohorts. Given that amyloid-beta is another neuropathological hallmark of AD, it would also be of interest to identify neuritic plaque-associated DMPs. Unfortunately, this was not feasible in the current study as this measure was not available in all samples. In a similar vein, we did not exclude individuals with mixed pathology, or protein hallmarks of other neurodegenerative diseases, such as the presence of lewy bodies, or TDP-43 pathology. In the future, larger meta-analyses should stratify by the presence of these protein aggregates, particularly given that very few EWAS have been undertaken in other dementias. Indeed, only three DNA methylomic studies have been undertaken in cortical samples of individuals with other dementias to date[27-30], with none of these studies utilizing > 15 individuals for EWAS. Further studies exploring common and

unique DNA methylation signatures and our classifier in other diseases characterized by dementia will be vital for identifying disease-specific epigenetic signatures that could represent novel therapeutic targets. Finally, one key issue for epigenetic studies in post-mortem tissue is the issue of causality, where it is not possible to determine whether disease-associated epigenetic loci are driving disease pathogenesis, or are a consequence of the disease, or even the medication used for treatment. One method that can be used to address this is Mendelian Randomization[31] however, this does require the CpG site to have a strong association with a SNP. Given that we only identified mQTLs at 18 of the 220 Bonferroni significant cross-cortex DMPs, this approach is not suitable for most of the loci we identified. At an experimental level establishing causality is difficult to address in post-mortem human studies, and therefore longitudinal studies in animal models, or modelling methylomic dysfunction through epigenetic editing *in vitro* will be useful approaches 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.

In summary we present the first meta-analyses of AD EWAS, highlighting numerous Bonferroni significant DMPs in the individual cortical regions and across the cortex, but not in the cerebellum, which were replicated in two independent cohorts. A number of these loci are novel and warrant further study to explore their role in disease etiology. We highlight that the nominated epigenetic changes are largely independent of genetic effects, with only 18 of the 220 Bonferroni significant DMPs showing a mQTL. We provide the first evidence that robust epigenomic changes in the cortex can predict the level of pathology in a sample. Looking to the future it will be important 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. As our analyses focused specifically on neuropathology (tau)-associated differential methylation, inclusion criteria for all samples used in the "discovery" or "replication" cohorts was having post-mortem

neurofibrillary tangle Braak stage available. For each discovery sample cohort DNA methylation was quantified using the 450K array. The "London 1" cohort comprised of prefrontal cortex, superior temporal gyrus, entorhinal cortex, and cerebellum 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 entorhinal cortex and cerebellum 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 prefrontal cortex and superior temporal gyrus 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 and cerebellum samples from The Sun Health Research Institute Brain Donation Program[32] published by Brokaw *et al.*[12]. The "Arizona 2" cohort consisted of an additional 88 temporal gyrus and cerebellum 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 two further replication datasets. The "Munich" cohort" from Neurobiobank Munich (NBM), which had bulk prefrontal cortex 450K array data from 45 donors, and 450K array data from fluorescence-activated cell sorted neuronal and non-neuronal (glial) populations from the occipital cortex from 26 donors as described by Gasparoni *et al.*[11]. The "Brains for Dementia Research (BDR)" cohort consisted of bulk prefrontal cortex Illumina Infinium EPIC array data from 590 donors, as described by Shireby *et al*[15]. Demographic information for all eight cohorts is available in **Table 1**.

Data quality control and harmonization

All computations and statistical analyses were performed using R 3.5.2[33] and Bioconductor 3.8[34]. A *MethylumiSet* object was created from iDATs using the methylumi package[35] and *RGChannelSet* object was created using the minfi package[36]. 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 SNP probes on the array show a modest level of correlation (using a cut-off of 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[37], 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. Finally, probes with common (minor allele frequency > 5%) SNPs in the single base extension position or probes that are nonspecific or mis-mapped were excluded[38,39], leaving 403,763 probes for analysis. Samples numbers after quality control are those shown in **Table 1**.

Quantile normalization was applied using the *dasen* function in the wateRmelon package[37]. For the discovery cohorts, 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[40], and were not included as a co-variate for the cerebellum as the neuronal nuclear protein (NeuN) that was used to generate the neuron/glia algorithm is not expressed by some cerebellar neurons[41]. These three variables (age, sex, neuron/glia proportions) were regressed out of the data as we found that they strongly correlated with either of the first two principal components of the DNA methylation data in most of the datasets. Other potential sources of technical and biological variation (post-mortem interval, ancestry, plate, chip, study and bisulfite treatment batch) did not correlate as strongly with methylation in most datasets. We opted to use surrogate variables as a consistent method to control for variation derived from these measured and other unknown variables across all datasets. Surrogate variables were calculated using the *sva* function in the SVA package[42]. Linear regression analyses were then performed with respect to Braak stage (modelled as a continuous variable) using residuals and a variable number of surrogate variables for each study until the inflation index (lambda) fell below 1.2 (see **Supplementary Table 16**). The surrogate variables included for each cohort correlated with the technical and biological variables that we had not regressed out earlier, demonstrating that this method appropriately controlled for variation not driven by Braak stage. Quantile-quantile plots for the four intra-tissue and the cross-cortex meta-analyses are shown in **Supplementary Figure 20**. Although it appears from these 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 GWAS are not suitable for EWAS data[43].

Intra-tissue meta-analysis

We used the estimated coefficients and SEs from the six "discovery" cohorts to undertake an inverse variance intra-tissue meta-analysis independently in each available tissue using the *metagen* function within the Meta package[44], which applies inverse variance weighting. The estimates and SEs from individual cohort Braak linear regression analyses were added to the model for each tissue. The prefrontal cortex analyses included three cohorts (N = 959: London 1, Mount Sinai, ROS/MAP), the temporal gyrus analyses included four cohorts (N = 608: London 1, Mount Sinai, Arizona 1, Arizona 2) and the entorhinal cortex analyses included two cohorts (N = 189: London 1, London 2). The cerebellum analyses included data from four cohorts (N = 533: London 1, London 2, Arizona 1 and Arizona 2) although the cerebellum 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. Bonferroni significance was defined as $P < 1.238 \times 10^{-7}$ to account for 403,763 tests. A fixed effects meta-analysis are the results primarily reported as it is the most appropriate model for our study as it can more reliably estimate the pooled effect and therefore the standard error and P value. However, in the supplementary tables we do also report the results of the random effects meta-analysis as ~10% of Bonferroni significant DMPs in the intra-tissue meta-analysis had high heterogeneity and in which case the results from the random-effects model should also be considered.

Cross-cortex meta-analysis

As multiple cortical brain regions were available for the "London 1" and "Mount Sinai" cohorts, a mixed model was performed using the *lme* function within the nlme package[45]. Estimate coefficients and SEs from each EWAS were extracted and were subjected to bacon[43] 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. A fixed effects model was selected in this instance for consistency with the intra-tissue meta-analysis, although the random effects meta-analysis results also shown in the supplementary tables.

Replication analyses

For the Munich replication cohort, we extracted the beta values for the 220 cross-cortex Bonferroni significant DMPs. This DNA methylation data was then corrected for age, sex and neuron/glia proportions (bulk tissue only) prior to performing a linear regression analysis

with respect to Braak stage. For the BDR replication cohort, we were provided with beta values for the 208 cross-cortex Bonferroni significant DMPs that were present on the EPIC array. This data had been corrected for age, sex, neuron/glia proportions, batch and principal component 1, before the linear regression analysis was performed with respect to Braak stage, with Bacon used to control for inflation. Additional information on the BDR dataset can be found in Shireby *et al*[15].

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 from the transcription start site of a gene) and "Genomic Regions Enrichment of Annotations Tool" (GREAT)[46] annotation (which annotates a DMP to genes with a transcription start site within 5kb upstream, or 1kb downstream). Pathway analyses were performed on the Illumina (UCSC) annotated genes corresponding to the 220 Bonferroni significant cross-cortex DMPs (N = 121 genes). A logistic regression approach, which we have previously described[47,48], was used to test if genes in this list predicted pathway membership, while controlling for the number of probes that passed quality control annotated to each gene. Pathways were downloaded from the Gene Ontology 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 Gene Ontology pathway were considered. Pathways were filtered to those containing between 10 and 2,000 genes. After applying this method to all pathways, significant pathways (unadjusted $P < 0.05$) 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[49] 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.* CpG islands, shelves, shores, non-CpG island regions) in certain genomic regions (*i.e.* intergenic, distal promoter,

proximal promoter, gene body, downstream) we annotated all DMPs with Slieker annotation[16] 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 any of the 220 Bonferroni significant cross-cortex DMPs were driven by genetic variation we used the xQTL resource to identify which of these DMPs are established *cis*-mQTLs[17]. To explore 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.*[18] 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 in Braak pathology explained by DNA methylation signatures

For this 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 Bonferroni significant DMPs that determined case-control status in the training dataset using the R package GLMnet[50]. Elastic net uses a combination of ridge and lasso regression, in which alpha (α) = 0 corresponds to ridge, whilst α = 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 AUC ROC value, confidence intervals (CI) and variance explained in the testing dataset as well as the independent replication Munich (Braak 0-II: N = 9, Braak V-VI: N = 29) and BDR (Braak 0-II: N = 196, Braak V-VI: N = 258) prefrontal cortex datasets.

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. Data analysis was undertaken using high-performance computing supported by a Medical Research Council (MRC) Clinical Infrastructure Award (M008924) to JM. The project was also supported through PhD studentships from the Alzheimer's Society (GS), BRACE (Bristol Research into Alzheimer's and Care of the Elderly) (GW) and the MRC GW4 Doctoral Training Program (DTP) (JAYR). BDR is jointly funded by Alzheimer's Research UK and the Alzheimer's Society in association with the MRC. DNA methylation data generated in the Brains for Dementia Research cohort was supported by the Alzheimer's Society and Alzheimer's Research UK (ARUK). 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, DB, PTF, AJT, SL, KM, JW, PDC, DAB, PLDJ, JM and KL provided data for the meta-analysis. LS developed the online database. 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 1 data (GSE134379), Arizona 2 data (GSE109627) and Munich data (GSE66351). We have developed an online database, 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. Brokaw, D.L. *et al.* Cell Death and Survival Pathways in Alzheimer's Disease: An Integrative Hypothesis Testing Approach Utilizing -Omic Datasets. *Neurobiol Aging* **In Press** **https://doi.org/10.1016/j.neurobiolaging.2020.06.022**(2020).
13. Smith, A.R. *et al.* A cross-brain regions study of ANK1 DNA methylation in different neurodegenerative diseases. *Neurobiol Aging* **74**, 70-76 (2019).
14. Nyholt, D.R. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* **74**, 765-9 (2004).
15. Shireby, G.L. *et al.* Recalibrating the Epigenetic Clock: Implications for Assessing Biological Age in the Human Cortex *BioxRiv* **https://doi.org/10.1101/2020.04.27.063719**(2020).
16. 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).
17. Ng, B. *et al.* An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat Neurosci* **20**, 1418-1426 (2017).

18. 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).
19. Selkoe, D.J. The molecular pathology of Alzheimer's disease. *Neuron* **6**, 487-98 (1991).
20. Xu, J. *et al.* Regional protein expression in human Alzheimer's brain correlates with disease severity. *Commun Biol* **2**, 43 (2019).
21. 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).
22. Roubroeks, J.A.Y. *et al.* An epigenome-wide association study of Alzheimer's disease blood highlights robust DNA hypermethylation in the HOXB6 gene. . *Neurobiol Aging* **https://doi.org/10.1016/j.neurobiolaging.2020.06.023**(2020).
23. Friedrich, J. *et al.* Hox Function Is Required for the Development and Maintenance of the Drosophila Feeding Motor Unit. *Cell Rep* **14**, 850-60 (2016).
24. Sun, L.L., Yang, S.L., Sun, H., Li, W.D. & Duan, S.R. Molecular differences in Alzheimer's disease between male and female patients determined by integrative network analysis. *J Cell Mol Med* **23**, 47-58 (2019).
25. 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).
26. 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).
27. Fernandez, A.F. *et al.* A DNA methylation fingerprint of 1628 human samples. *Genome Res* **22**, 407-19 (2012).
28. Masliah, E., Dumaop, W., Galasko, D. & Desplats, P. Distinctive patterns of DNA methylation associated with Parkinson disease: identification of concordant epigenetic changes in brain and peripheral blood leukocytes. *Epigenetics* **8**, 1030-8 (2013).
29. Sanchez-Mut, J.V. *et al.* Human DNA methylomes of neurodegenerative diseases show common epigenomic patterns. *Transl Psychiatry* **6**, e718 (2016).
30. MacBean, L.F., Smith, A.R. & Lunnon, K. Exploring beyond the DNA sequence: A Review of Epigenomic Studies of DNA and Histone Modifications in Dementia. *Current Genetic Medicine Reports* **In Press**(2020).
31. Relton, C.L. & Davey Smith, G. Mendelian randomization: applications and limitations in epigenetic studies. *Epigenomics* **7**, 1239-43 (2015).
32. Beach, T.G. *et al.* Arizona Study of Aging and Neurodegenerative Disorders and Brain and Body Donation Program. *Neuropathology* **35**, 354-89 (2015).
33. R Development Core Team. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria 2012* (2012).
34. Gentleman, R.C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80 (2004).
35. Spiers, H. *et al.* Methylomic trajectories across human fetal brain development. *Genome Res* **25**, 338-52 (2015).
36. 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).
37. Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**, 293 (2013).

38. Chen, Y.A. *et al.* Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* **8**, 203-9 (2013).
39. Price, M.E. *et al.* Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin* **6**, 4 (2013).
40. 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).
41. Gusel'nikova, V.V. & Korzhevskiy, D.E. NeuN As a Neuronal Nuclear Antigen and Neuron Differentiation Marker. *Acta Naturae* **7**, 42-7 (2015).
42. Chakraborty, S., Datta, S. & Datta, S. Surrogate variable analysis using partial least squares (SVA-PLS) in gene expression studies. *Bioinformatics* **28**, 799-806 (2012).
43. 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).
44. Schwarzer, G. meta: A R package for meta-analysis. *R News* **7**, 40-45 (2007).
45. Pinheiro, J., Bates, D., DebRoy, S. & Sarkar, D. nlme: Linear and Nonlinear Mixed Effects Models. (2019).
46. McLean, C.Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501 (2010).
47. 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).
48. Lunnon, K. *et al.* Variation in 5-hydroxymethylcytosine across human cortex and cerebellum. *Genome Biol* **17**, 27 (2016).
49. 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).
50. Friedman, J. *et al.* glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models. (2019).

Figure 1: Intra-tissue meta-analyses of AD methylomic studies highlights Bonferroni significant differentially methylated positions (DMPs) in all cortical tissues. (a) A Manhattan plot for the prefrontal cortex (red), temporal gyrus (green) and entorhinal cortex (blue) meta-analyses, with the ten most significant 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 Bonferroni significance ($P < 1.238 \times 10^{-7}$). (**b**) A Venn diagram highlighting overlapping DMPs at Bonferroni significance across the cortical tissues. (**c**) In each cortical brain region the Bonferroni significant DMPs identified in that region usually had a greater effect size (ES) there, than in any of the other cortical regions. The X-axis represents the methylation (beta) ES between individuals that are Braak stage 0 and VI. Data is separated on the Y-axis by tissue analysis (large text) with the corresponding data at these probes in other tissues (small text). The white dot in the centre represents the median, the dark box represents the interquartile range (IQR), whilst the whisker lines represent the "minimum" (quartile 1 – 1.5 x IQR) and the "maximum" (quartile 3 + 1.5 x IQR). The coloured violin represents all samples including outliers, meaning that the "minimum" and "maximum" may not extend to the end of the violin.

Fig 1. a

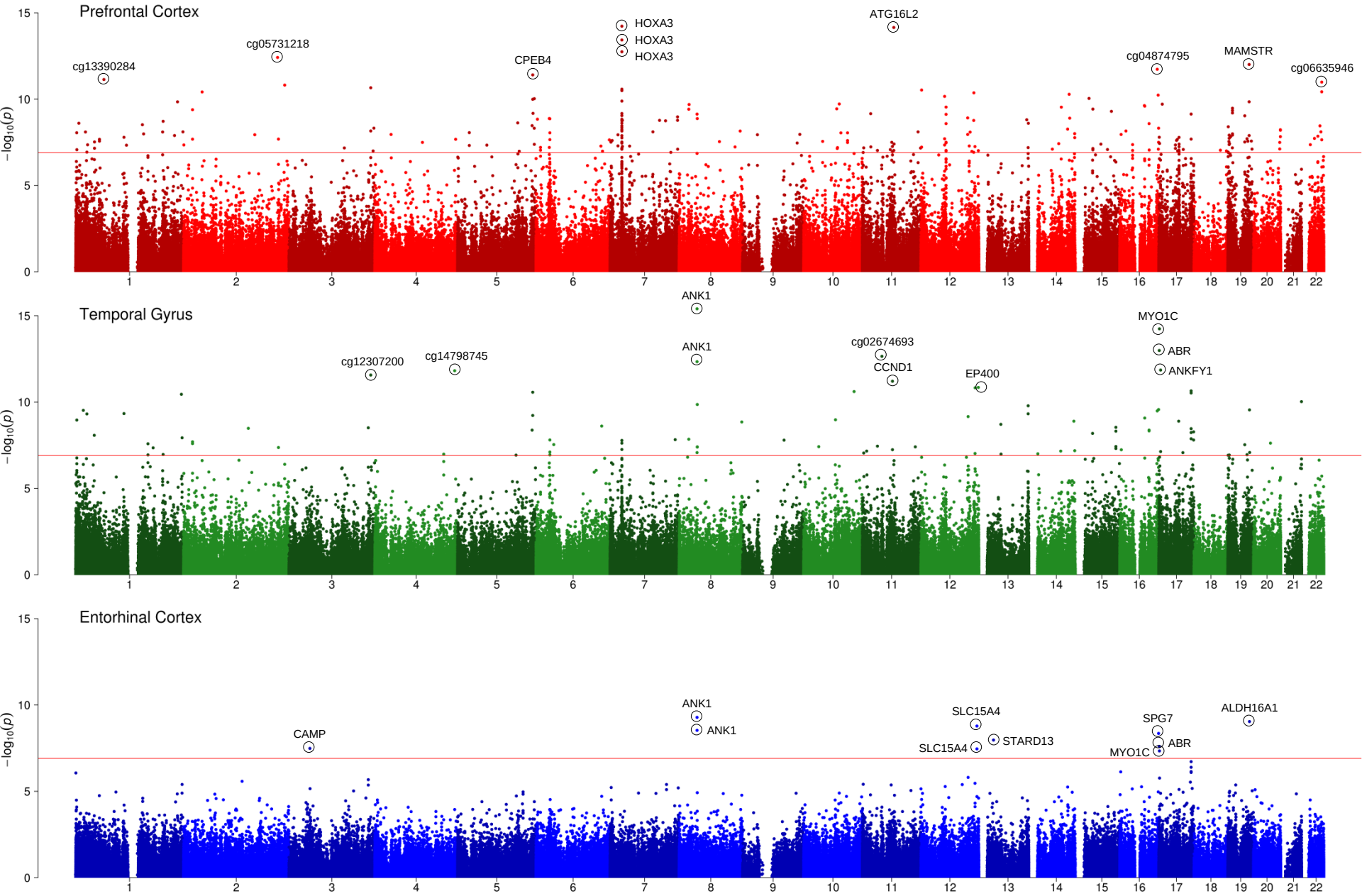
Prefrontal Cortex
cg13390284
cg05731218
CPEB4
HOXA3
HOXA3
HOXA3
ATG16L2
cg04874795
MAMSTR
cg06635946
ANK1
Temporal Gyrus
cg12307200
cg14798745
ANK1
cg02674693
CCND1
EP400
MYO1C
ABR
ANKFY1
Entorhinal Cortex
CAMP
ANK1
ANK1
SLC15A4
SLC15A4
STARD13
SPG7
ABR
MYO1C
ALDH16A1
−log10(p)
0
5
10
15
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

Fig 1. b

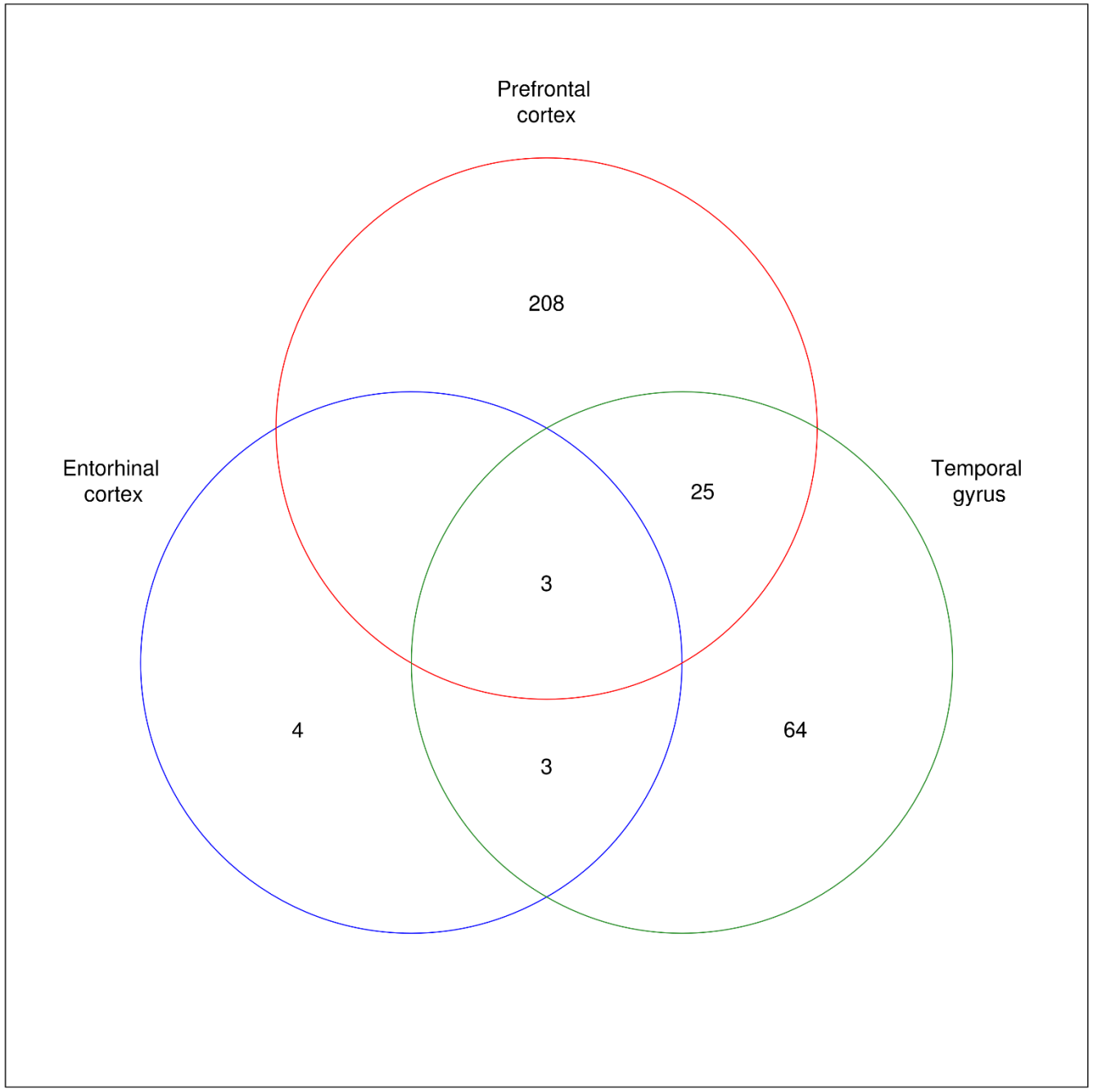
Prefrontal cortex
Entorhinal cortex
Temporal gyrus
208
25
3
4
3
64

Fig 1. c

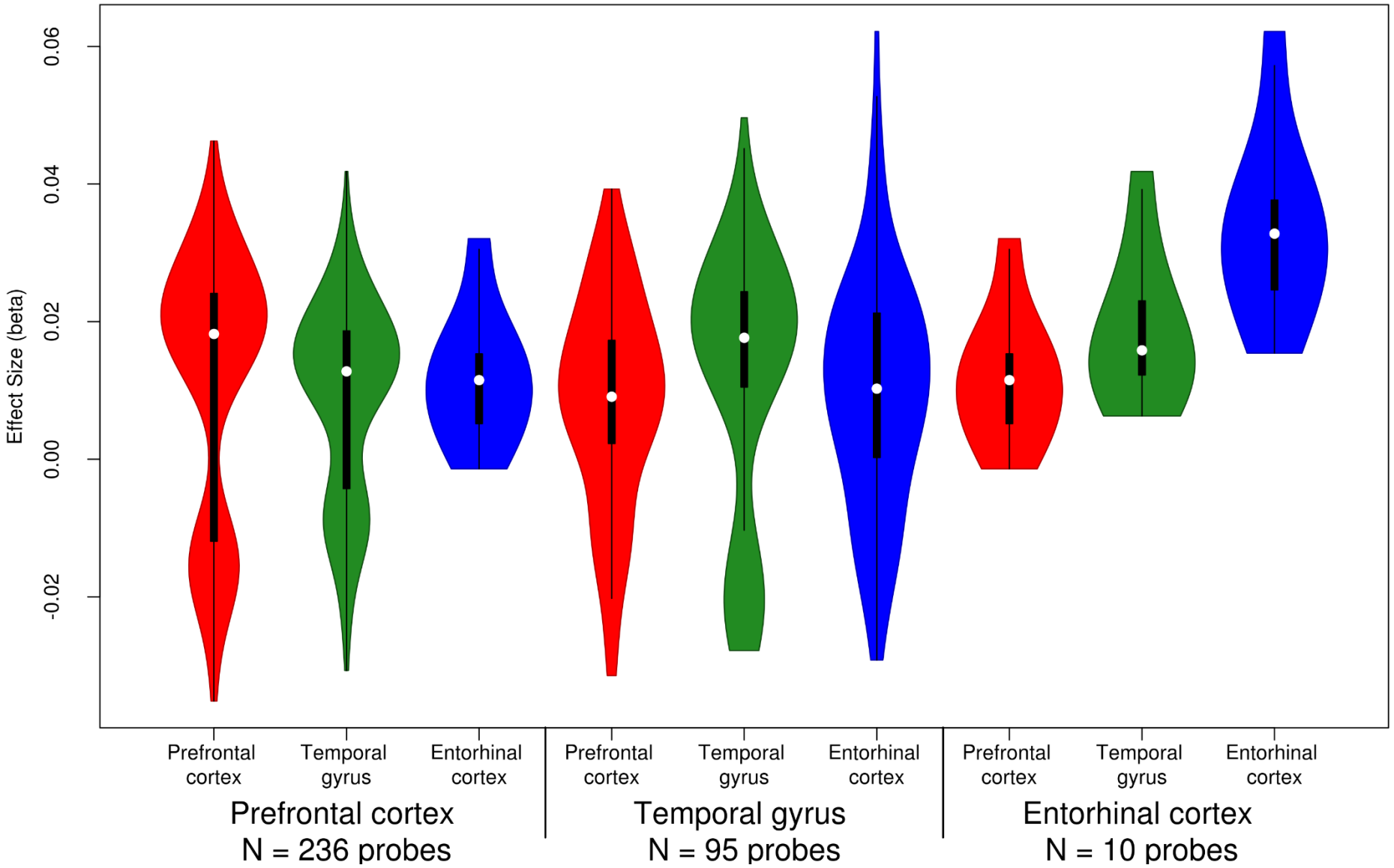
Effect Size (beta)
0.06
0.04
0.02
0.00
-0.02
Prefrontal cortex
Temporal gyrus
Entorhinal cortex
Prefrontal cortex
Temporal gyrus
Entorhinal cortex
Prefrontal cortex
Temporal gyrus
Entorhinal cortex
Prefrontal cortex
N = 236 probes
Temporal gyrus
N = 95 probes
Entorhinal cortex
N = 10 probes

Figure 2: A cross-cortex meta-analysis identifies 220 Bonferroni significant 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. The chromosome and genomic position are shown on the X-axis. The Y-axis shows –log10(p). The red horizontal lines indicate the Bonferroni significance level of $P < 1.238 \times 10^{-7}$. Probes with a methylation (beta) 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 between ≥ 0.01, whilst blue probes indicate an ES ≥ 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 Bonferroni significance level of $P < 1.238 \times 10^{-7}$. Red probes represent a positive ES ≥ 0.01, blue probes represent a negative ES ≥ 0.01. Underneath the gene tracks are shown in black with CpG islands in green.

Fig 2. a

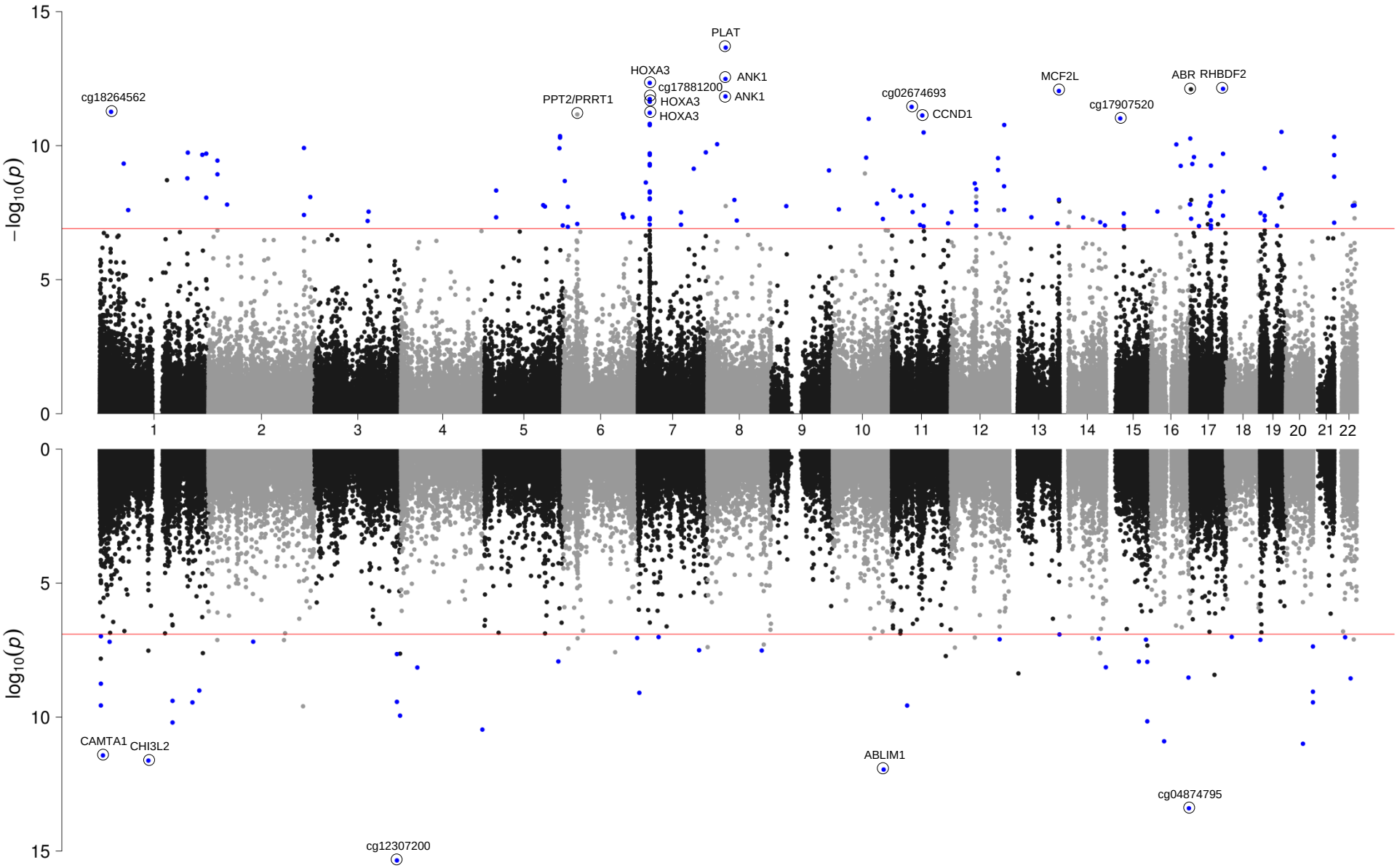
15
10
5
0
$-\log_{10}(p)$
cg18264562
PPT2/PRRT1
PLAT
HOXA3
ANK1
cg17881200
HOXA3
ANK1
HOXA3
cg02674693
CCND1
MCF2L
cg17907520
ABR
RHBDF2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
0
5
10
15
$\log_{10}(p)$
CAMTA1
CHI3L2
cg12307200
ABLIM1
cg04874795

Fig 2. b

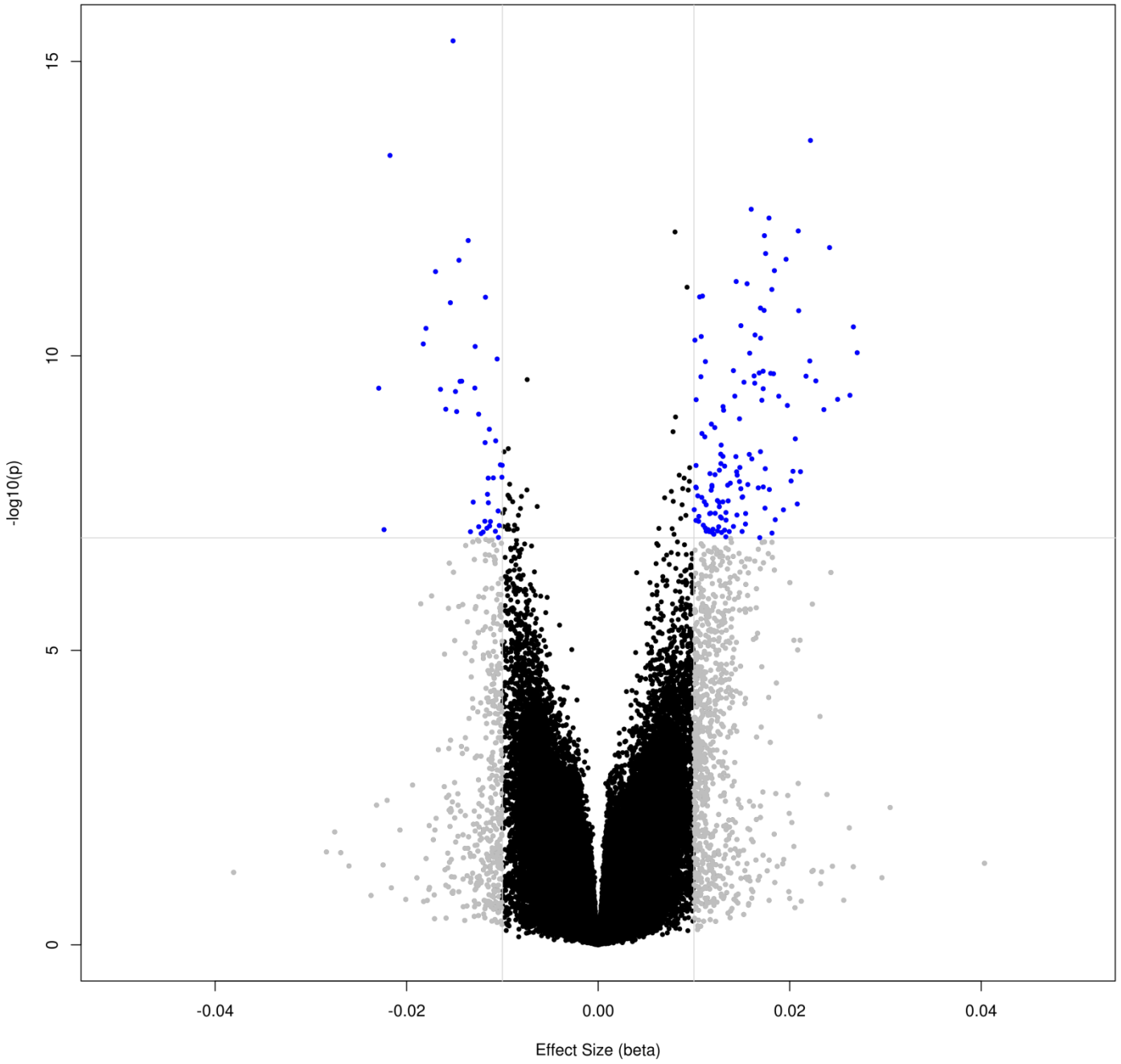

Fig 2. c

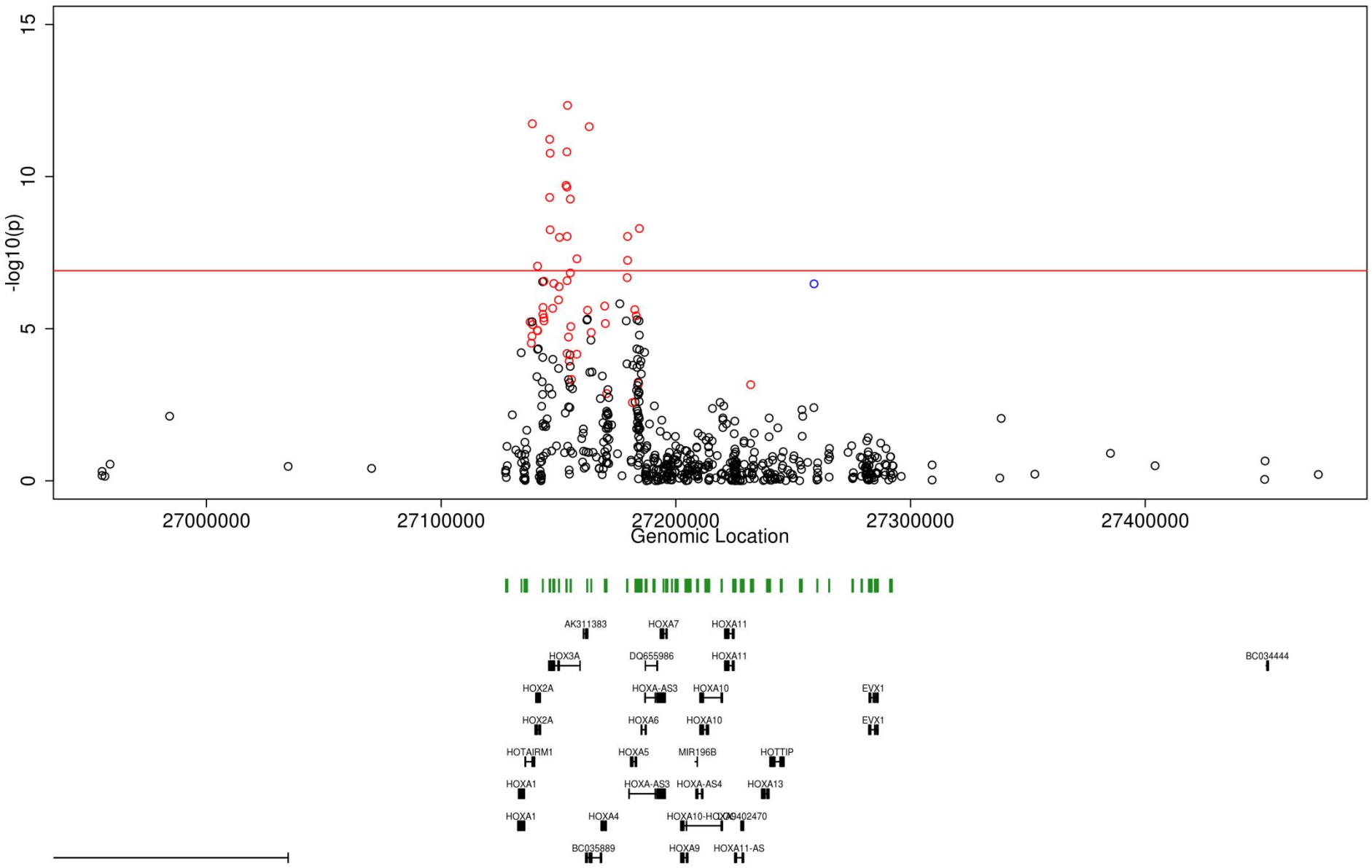
-log10(p)
0
5
10
15
27000000
27100000
27200000
27300000
27400000
Genomic Location
AK311383
HOXA7
HOXA11
HOX3A
DQ655986
HOXA11
BC034444
HOX2A
HOXA-AS3
HOXA10
EVX1
HOX2A
HOXA6
HOXA10
EVX1
HOTAIRM1
HOXA5
MIR196B
HOTTIP
HOXA1
HOXA-AS3
HOXA-AS4
HOXA13
HOXA1
HOXA4
HOXA10-HOXA9
AC402470
BC035889
HOXA9
HOXA11-AS

Figure 3: Independent replication of the Bonferroni significant cross-cortex differentially methylated loci. (**a**) The methylation (beta) effect size (ES) of the 220 cross-cortex differentially methylated positions (DMPs) identified in the discovery cohorts (X-axis) were significantly correlated with the ES in the Munich replication cohort in the prefrontal cortex (red, r = 0.64, P = 5.24×10^{-27}), sorted neuronal cells (light blue, r = 0.45, P = 1.56×10^{-12}) and non-neuronal cells (purple, r = 0.79, P = 1.43×10^{-47}) (Y-axis). (**b**) A forest plot of the most significant cross-cortex DMP (cg12307200, chr3:188664632, P = 4.48×10^{-16}). The effect size is shown in the prefrontal cortex (red), temporal gyrus (green) and entorhinal cortex (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 Munich replication cohort in the prefrontal cortex 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. (**c**) In the BDR replication cohort DNA methylation data was available in the prefrontal cortex for 208 of the 220 Bonferroni significant cross-cortex DMPs. The ES of these 208 cross-cortex DMPs in the discovery cohorts (X-axis) were significantly correlated with the ES in the BDR replication cohort (r = 0.53, P = 4.13×10^{-16}) (Y-axis).

Fig 3. a

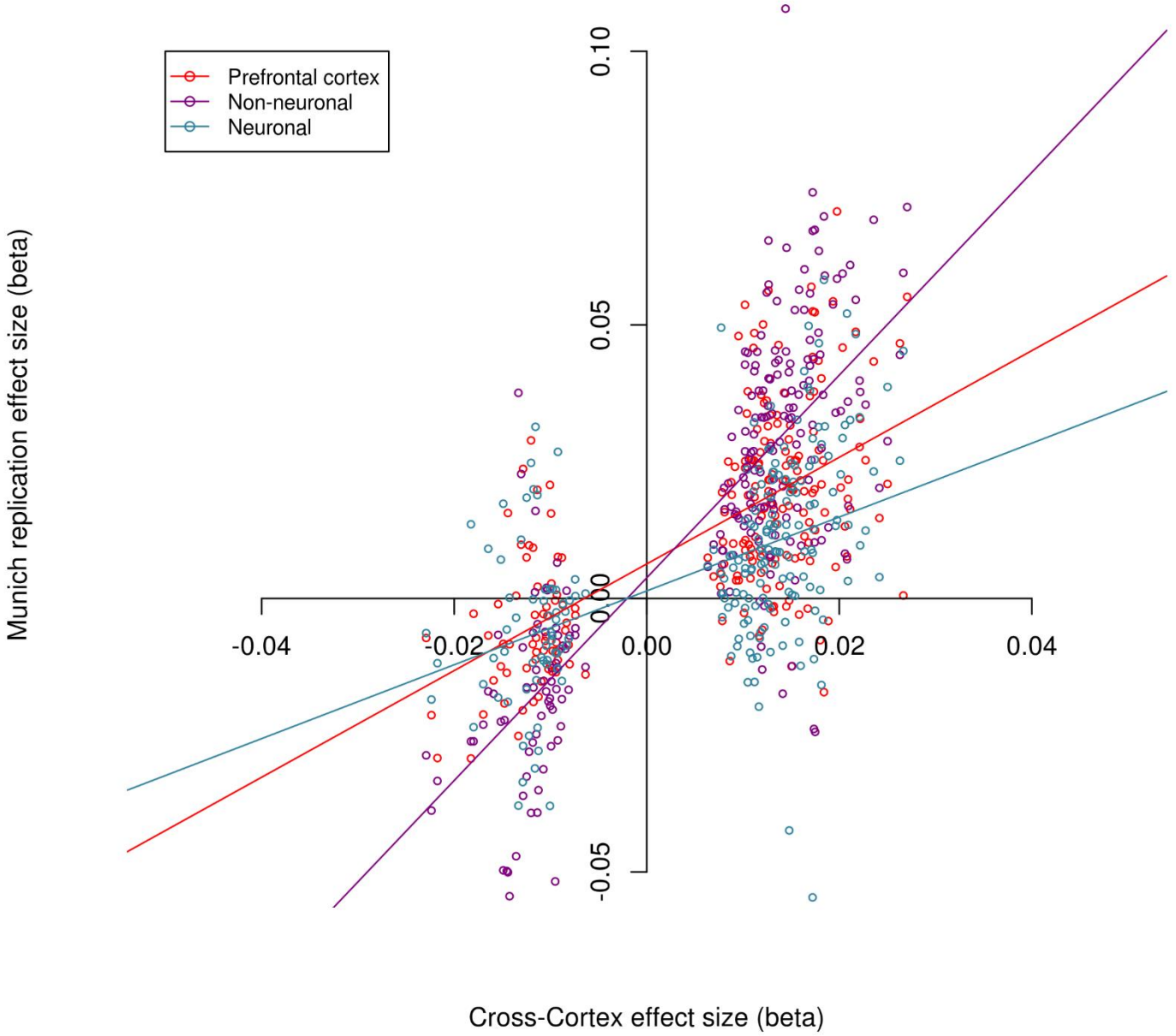
Prefrontal cortex
Non-neuronal
Neuronal
Munich replication effect size (beta)
Cross-Cortex effect size (beta)
0.10
0.05
0.00
-0.05
-0.04
-0.02
0.00
0.02
0.04

Fig 3. b

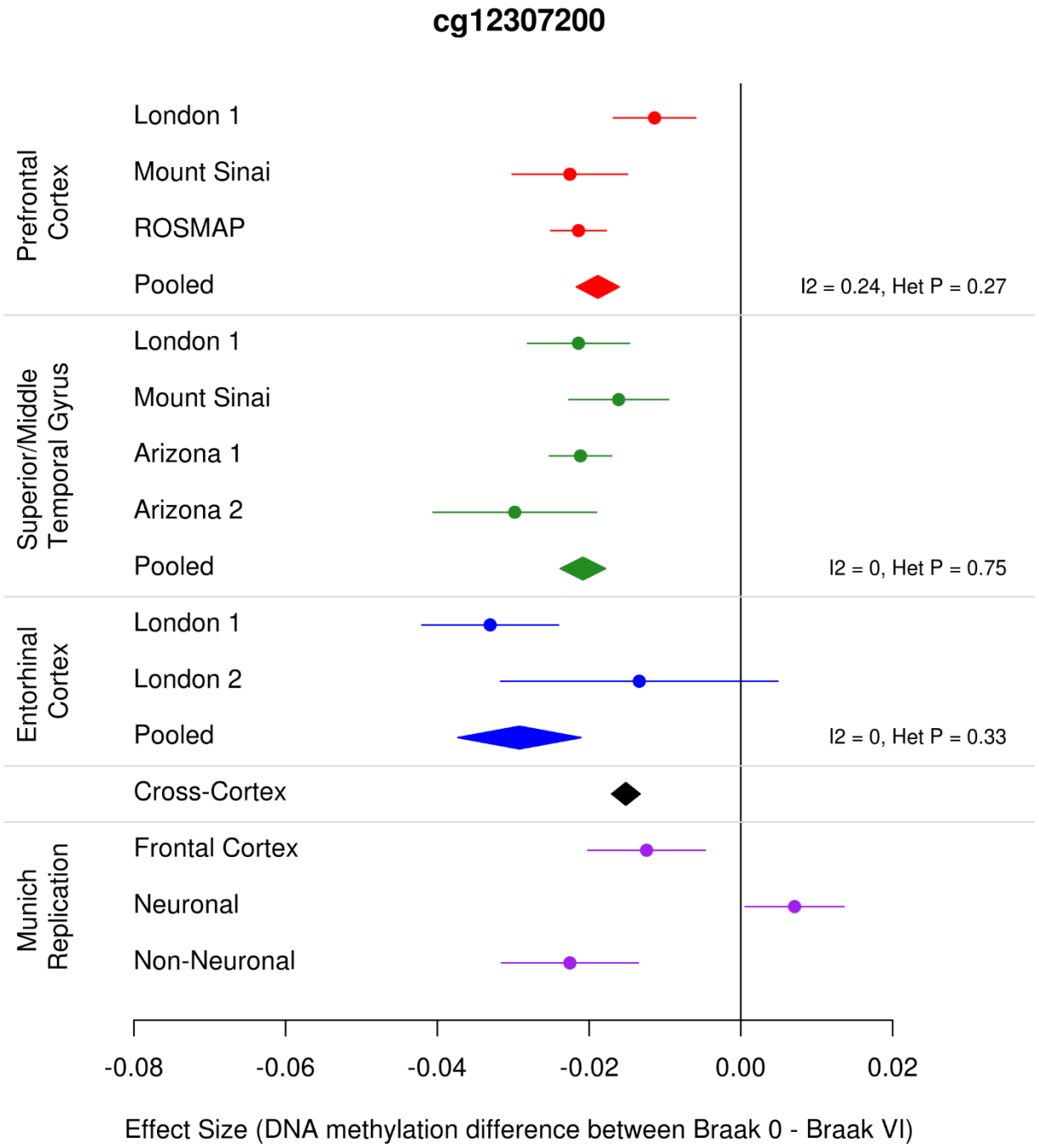
cg12307200
Prefrontal Cortex
London 1
Mount Sinai
ROSMAP
Pooled
I2 = 0.24, Het P = 0.27
Superior/Middle Temporal Gyrus
London 1
Mount Sinai
Arizona 1
Arizona 2
Pooled
I2 = 0, Het P = 0.75
Entorhinal Cortex
London 1
London 2
Pooled
I2 = 0, Het P = 0.33
Cross-Cortex
Munich Replication
Frontal Cortex
Neuronal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

Fig 3. c

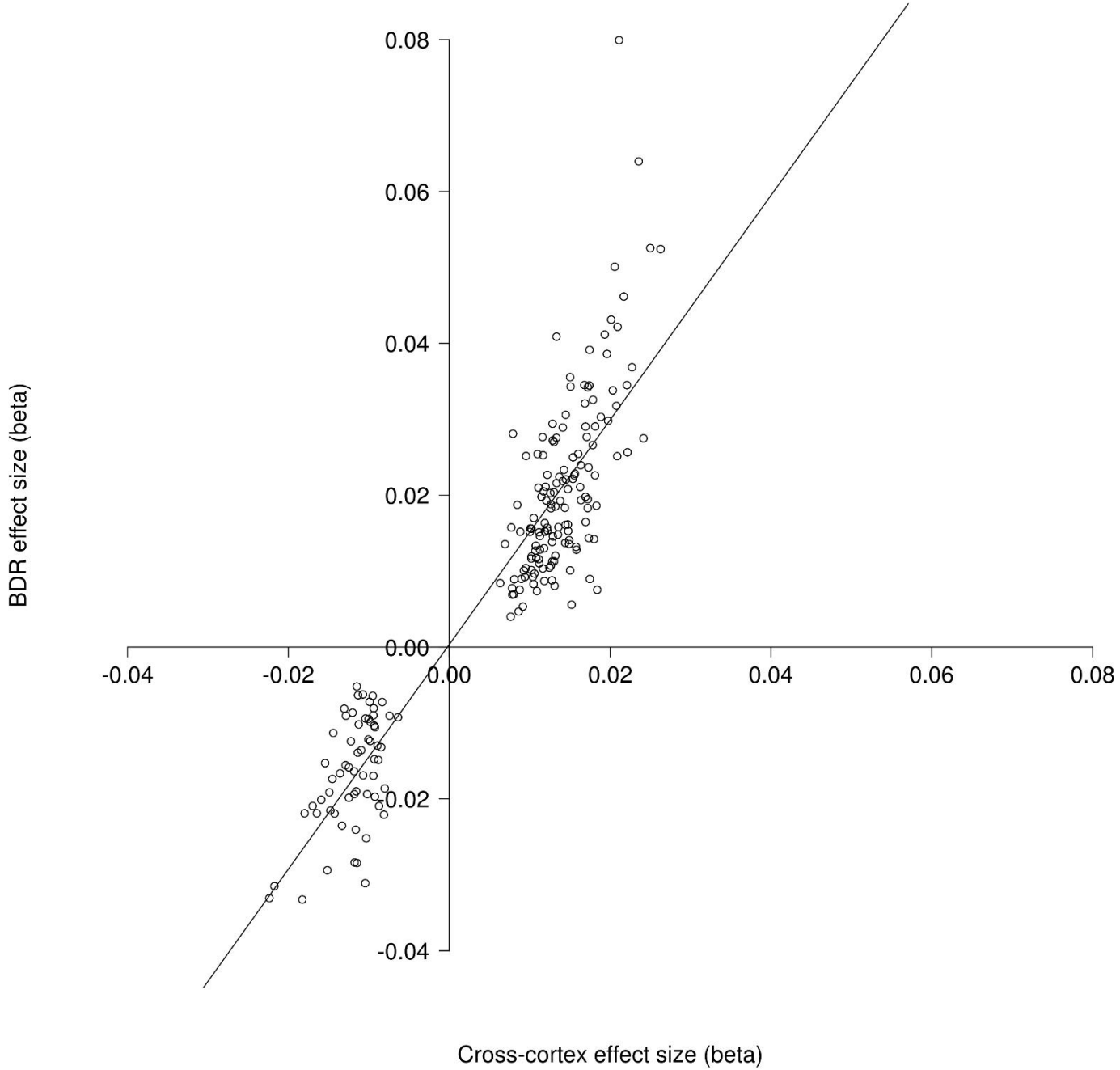
0.08
0.06
0.04
0.02
0.00
-0.02
-0.04
-0.04
-0.02
0.00
0.02
0.04
0.06
0.08
BDR effect size (beta)
Cross-cortex effect size (beta)

Figure 4: Receiver Operating Characteristic (ROC) graphs highlighting the Area Under the Curve (AUC) for the 95 cross-cortex probes that can best explain the variance in Braak pathology. An elastic net penalized regression model was used to identify a subset of 95 of the Bonferroni significant cross-cortex probes that could best predict whether a sample has low pathology (Braak 0-II: “control”) compared to high pathology (Braak V-VI: “AD”) in a training dataset of 696 (discovery) samples (Braak 0-II: N = 283, Braak V-VI: N = 413). This model had an Area Under the ROC Curve (AUC) of 94.36% (confidence interval [CI] = 92.67-95.88%) and explained 71.52% of the pathological variance (black line). This was then tested in a testing dataset of 233 (discovery) samples (Braak 0-II: N = 103, Braak V-VI: N = 130), where it had an AUC = 87.63% (CI = 82.73-91.89%) and explained 52.39% of the variance (red line). The 95 probe signature was then tested in two independent replication cohorts. In the Munich prefrontal cortex samples (Braak 0-II: N = 9, Braak V-VI: N = 29) the model had an AUC of 75.1% (CI = 55.56-90.81%), explaining 25.47% of the variance (blue line). In the BDR prefrontal cortex samples (Braak 0-II: N = 196, Braak V-VI: N = 258) the model had an AUC = 70.33% (CI = 65.32-74.93%), explaining 15.44% of the variance. A list of the 95 probes and their performance characteristics can be found in Supplementary Tables 12 and 13, respectively.

Fig 4.

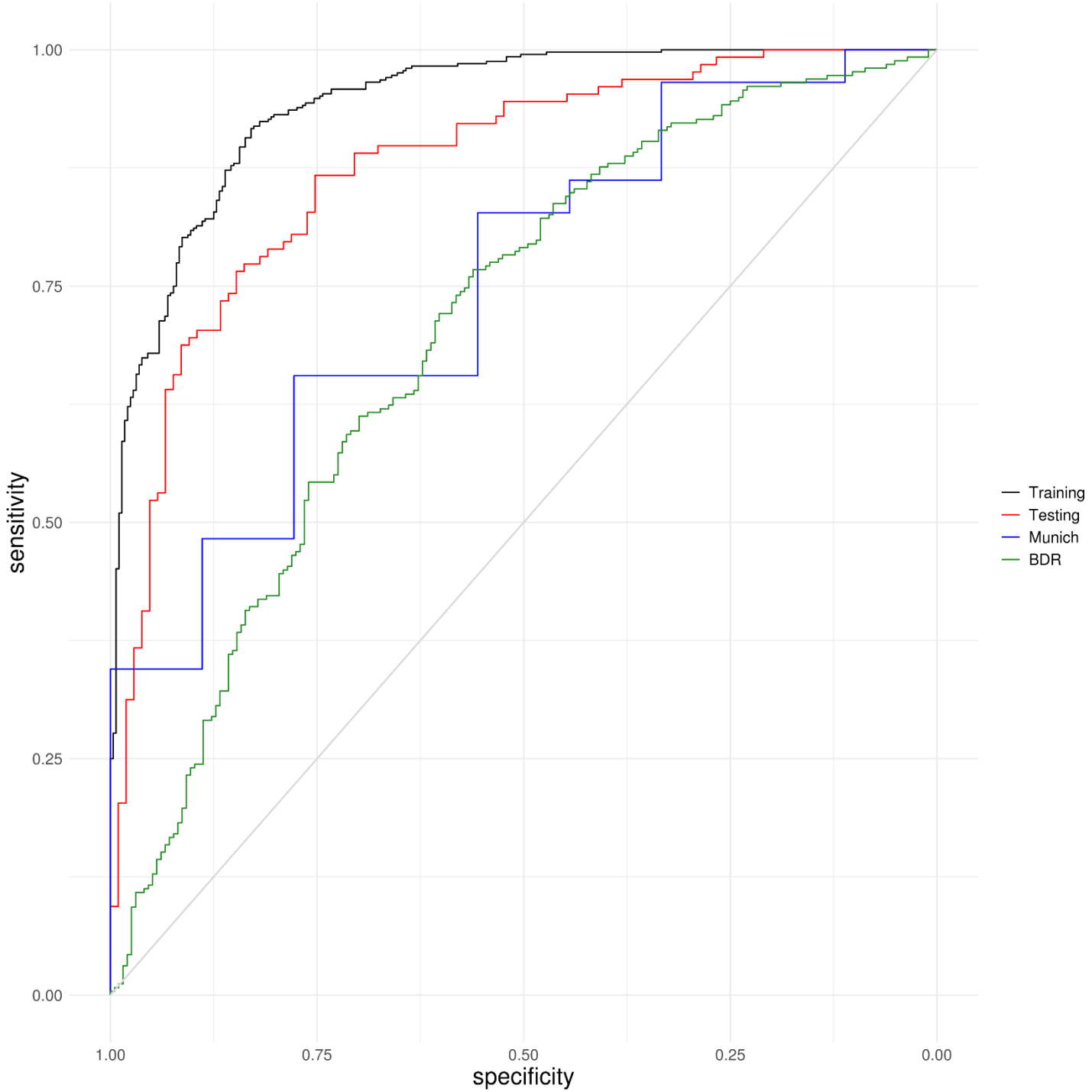
1.00
0.75
0.50
0.25
0.00
sensitivity
1.00
0.75
0.50
0.25
0.00
specificity
Training
Testing
Munich
BDR

Table 1: Demographic information for cohorts included in the meta-analyses. Sample numbers, split of males (M)/females (F) and mean age at death in years (±) standard deviation (SD) 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, entorhinal cortex, middle temporal gyrus, prefrontal cortex and superior temporal gyrus. In the discovery meta-analyses, we used data from six EWAS using the 450K array, which all had > 50 unique donors. For replication we used two cohorts. The Munich cohort had 450K data from bulk prefrontal cortex tissue, as well as data available from sorted neuronal and non-neuronal cell populations from the occipital cortex. The BDR cohort had EPIC array data available from bulk prefrontal cortex samples. For the meta-analyses, superior temporal gyrus and middle temporal gyrus samples were both classed as temporal gyrus samples. Shown are final numbers for all cohorts after data quality control. Ancestry is reported for the discovery cohorts and is the number of unique individuals that had the following inferred ethnicities from the 1000 genomes reference panel: European (Eu), African (Af), American (Am), East Asian (As).

Stage	Cohort	Unique individuals	Ancestry (Eu/Af/Am/As)	Braak	Number	Sex (M/F)	Age at death in (± SD)	Tissues analysed
DISCOVERY	London 1	113	112/0/1/0	0-II	29	13/16	77.6 (12.8)	Prefrontal cortex, entorhinal cortex, superior temporal gyrus, cerebellum (Bulk)
				III-IV	18	7/11	88.5 (5.2)	
				V-VI	66	26/40	85.4 (8.1)	
	London 2	95	92/1/2/0	0-II	23	12/11	76.1 (10.0)	Entorhinal cortex, cerebellum (Bulk)
				III-IV	16	3/13	87.6 (6.4)	
				V-VI	56	26/30	81.5 (8.6)	
	Mount Sinai	146	113/20/11/2	0-II	60	32/28	82 (7.6)	Prefrontal cortex, superior temporal gyrus (Bulk)
				III-IV	42	12/30	88.8 (6.6)	
				V-VI	44	12/32	88.0 (7.5)	
	Arizona 1	302	302/0/0/0	0-II	61	40/21	80.3 (8.2)	Middle temporal gyrus, cerebellum (Bulk)
				III-IV	97	50/47	86.9 (6.9)	
				V-VI	144	63/81	82.3 (8.5)	
	Arizona 2	88	88/0/0/0	0-II	16	10/6	82.5 (5.0)	Middle temporal gyrus, cerebellum (Bulk)
				III-IV	45	21/24	86.7 (5.1)	
				V-VI	27	12/15	84.6 (7.1)	
	ROS/MAP	709	709/0/0/0	0-II	143	70/73	83.2 (6.0)	Prefrontal cortex (Bulk)
				III-IV	409	144/266	86.9 (4.1)	
				V-VI	157	45/113	87.8 (3.5)	
REPLICATION	Munich	45	-	0-II	9	5/4	76.7 (10.9)	Prefrontal cortex (Bulk)
				III-IV	7	1/6	82.1 (5.2)	
				V-VI	29	12/17	79.2 (8.5)	
		26	-	0-II	11	7/4	75.9 (8.5)	Occipital cortex (Sorted cells)
				III-IV	5	1/4	85.0 (6.5)	
				V-VI	10	4/6	77.9 (6.6)	
	BDR	590	-	0-II	196	100/96	83.6 (10.6)	Prefrontal cortex (Bulk)
				III-IV	136	91/65	85.1 (7.45)	
				V-VI	258	128/130	82.5 (8.5)	

Table 2: The 25 most significant 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, closest genes with a transcription start site upstream or downstream (from GREAT annotation). Shown for each DMP is the methylation (beta) 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. A more comprehensive table is provided in Supplementary Table 7.

Probe	Position	Illumina Gene Annotation	GREAT annotation - closest genes with transcription start site upstream (distance to site)	GREAT annotation - closest genes with transcription start site downstream (distance to site)	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