nature neuroscience

Methylomic profiling implicates cortical deregulation of *ANK1* in Alzheimer's disease

Katie Lunnon[1], Rebecca Smith[2], Eilis Hannon[1], Philip L De Jager[3–5], Gyan Srivastava[3,5], Manuela Volta[2], Claire Troakes[2], Safa Al-Sarraj[2], Joe Burrage[1], Ruby Macdonald[1], Daniel Condliffe[2], Lorna W Harries[1], Pavel Katsel[6], Vahram Haroutunian[6–8], Zachary Kaminsky[9,10], Catharine Joachim[11], John Powell[2], Simon Lovestone[2,12], David A Bennett[13], Leonard C Schalkwyk[2,14] & Jonathan Mill[1,2,14]

Alzheimer's disease (AD) is a chronic neurodegenerative disorder that is characterized by progressive neuropathology and cognitive decline. We performed a cross-tissue analysis of methylomic variation in AD using samples from four independent human post-mortem brain cohorts. We identified a differentially methylated region in the ankyrin 1 (*ANK1*) gene that was associated with neuropathology in the entorhinal cortex, a primary site of AD manifestation. This region was confirmed as being substantially hypermethylated in two other cortical regions (superior temporal gyrus and prefrontal cortex), but not in the cerebellum, a region largely protected from neurodegeneration in AD, or whole blood obtained pre-mortem from the same individuals. Neuropathology-associated *ANK1* hypermethylation was subsequently confirmed in cortical samples from three independent brain cohorts. This study represents, to the best of our knowledge, the first epigenome-wide association study of AD employing a sequential replication design across multiple tissues and highlights the power of this approach for identifying methylomic variation associated with complex disease.

AD contributes substantially to the global burden of disease, affecting in excess of 26 million people worldwide[1,2]. The pathology associated with AD is characterized by the accumulation of amyloid plaques, tangles of intracellular hyperphosphorylated tau, gliosis, synaptic dysfunction and eventually neuronal cell death[3,4]. Although the neuropathological manifestation of AD is well characterized in post-mortem brain, little is known about the underlying risk factors or mechanism(s) involved in disease progression. Of note, different parts of the brain show differential vulnerability to AD; although there is progressive neurodegeneration across the cortex, with areas such as the entorhinal cortex (EC) being characterized by considerable and early neuropathology, regions such as the cerebellum (CER) are relatively resistant to neuronal damage, with little or no plaque or neurofibrillary tangle pathology[5].

Contemporary research aimed at exploring the etiology of AD has focused primarily on DNA sequence variation, with some notable success[6]. Increasing knowledge about the biology of the genome[7] has suggested an important role for epigenetic variation in human health and disease, and recent methodological advances mean that epigenome-wide association studies (EWAS) are now feasible for complex disease phenotypes, including AD[8]. Epigenetic epidemiology is a relatively new endeavor, however, and there are important considerations regarding study design, tissue type, analysis strategy and data interpretation[9,10]. We carried out a systematic cross-tissue EWAS analysis of DNA methylation in AD using a powerful sequential replication design, with the goal of identifying disease-associated methylomic variation across pathologically relevant regions of the brain.

RESULTS

For the first (discovery) stage of our analysis, we used multiple tissues from donors (n = 122) archived in the MRC London Brainbank for Neurodegenerative Disease. From each donor, we isolated genomic DNA from four brain regions (EC, n = 104; superior temporal gyrus (STC), n = 113; prefrontal cortex (PFC), n = 110; CER, n = 108) and, where available, from whole blood obtained pre-mortem (n = 57) (**Supplementary Tables 1** and **2**). DNA methylation was quantified using the Illumina 450K HumanMethylation array, with pre-processing, normalization and stringent quality control undertaken as previously described[11] (Online Methods). Our analyses focused on identifying differentially methylated positions (DMPs) associated with Braak staging, a standardized measure of neurofibrillary

[1]University of Exeter Medical School, Exeter University, Exeter, UK. [2]Institute of Psychiatry, King's College London, London, UK. [3]Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Departments of Neurology and Psychiatry, Brigham and Women's Hospital, Boston, Massachusetts, USA. [4]Harvard Medical School, Boston, Massachusetts, USA. [5]Program in Medical and Population Genetics, Broad Institute, Cambridge, USA. [6]Department of Psychiatry, The Icahn School of Medicine at Mount Sinai, New York, USA. [7]Department of Neuroscience, The Icahn School of Medicine at Mount Sinai, New York, USA. [8]JJ Peters Virginia Medical Center, Bronx, New York, USA. [9]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [10]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA. [11]Department of Neuropathology, John Radcliffe Hospital, University of Oxford, Oxford, UK. [12]Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford, UK. [13]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, USA. [14]These authors contributed equally to this work. Correspondence should be addressed to J.M. (j.mill@exeter.ac.uk).

Received 26 April; accepted 7 July; published online 17 August 2014; doi:10.1038/nn.3782

Table 1 The ten top-ranked Braak-associated DMPs in EC

					London cohort								Mount Sinai cohort			
					EC		STG		PFC		CER		STG		PFC	
Probe	Position	Illumina gene annotation	Genes with TSS within 5 kb upstream	Genes with TSS within 1 kb downstream	*P* value	Δ	CETS-adjusted *P*	*P* value	Δ	*P* value	Δ	*P* value	Δ	*P* value	Δ	*P* value	Δ
cg11823178	8:41519399	*ANK1; MIR486*	*NKX6-3*	*ANK1*	4.59×10^{-7}	4.62	7.09×10^{-7}	6.51×10^{-5}	3.12	2.33×10^{-3}	2.05	–	–	1.63×10^{-4}	3.36	1.07×10^{-3}	2.97
cg22997194	10:72647819	*PCBD1*	*PCBD1*	–	9.99×10^{-7}	−3.89	0.05	–	–	–	–	–	–	–	–	–	–
cg06653632	12:129281444	*SLC15A4*	*SLC15A4*	*TMEM132C*	1.98×10^{-6}	3.51	5.74×10^{-7}	5.02×10^{-3}	2.01	0.0381	1.42	–	–	9.21×10^{-4}	2.40	0.0124	1.89
cg05066959	8:41519308	*ANK1; MIR486*	*NKX6-3*	*ANK1*	1.35×10^{-5}	5.45	6.20×10^{-6}	3.78×10^{-4}	4.03	6.48×10^{-4}	3.45	–	–	5.78×10^{-4}	4.75	4.00×10^{-3}	3.41
cg24152732	19:4180820	*SIRT6*	*SIRT6*	*CREB3L3*	1.37×10^{-5}	−3.11	2.45×10^{-6}	–	–	–	–	0.0429	−1.86	–	–	–	–
cg14972141	13:100217995	–	*CLYBL*	*TM9SF2*	1.48×10^{-5}	−4.01	0.0025	3.41×10^{-4}	−2.25	0.0179	−2.23	7.86×10^{-4}	−2.30	–	–	–	–
cg04029027	7:130125811	*MEST*	*CEP41*	*MEST*	2.29×10^{-5}	3.41	2.06×10^{-5}	0.0292	1.92	–	–	–	–	–	–	–	–
cg05030077	16:2255199	*MLST8*	*MLST8*	–	2.69×10^{-5}	−1.89	0.12	–	–	–	–	–	–	–	–	–	–
cg04151012	2:27806529	*ZNF512*	*ZNF512*	–	2.88×10^{-5}	3.21	1.39×10^{-5}	–	–	–	–	–	–	–	–	–	–
cg18522315	20:8000623	*TMX4*	*TMX4*	–	3.27×10^{-5}	1.75	9.18×10^{-5}	0.0422	0.74	–	–	–	–	–	–	–	–

Shown for each DMP are chromosomal location (hg19), up/downstream genes, *P* value from our quantitative model (Online Methods), difference (Δ) in corrected DNA methylation (%) between individuals with the lowest (score 0) and highest (score VI) Braak score, and CETS *P* value[17] to highlight whether these top-ranked DMPs were mediated by the effect of differential neuronal cell proportions across samples. Also shown are the corresponding statistics across the three other matched brain regions (STG, PFC, CER) in the London cohort and the STG and PFC in the Mount Sinai replication cohort for CpG sites showing a nominally significant difference in the same direction. The 100 top-ranked EC DMPs are given in **Supplementary Table 3**, with data for other brain regions given in **Supplementary Tables 4–6**.

tangle burden determined at autopsy[12], with all analyses controlling for age and sex.

We first assessed DNA methylation differences identified in the EC, given that it is a primary and early site of neuropathology in AD[5]. The top-ranked Braak-associated EC DMPs are shown in **Table 1** and **Supplementary Table 3**, and results for the other brain regions that we profiled (STG, PFC, and CER) are shown in **Supplementary Tables 4–6**. Two of the top-ranked EC DMPs (cg11823178, the top-ranked EC DMP, and cg05066959, the fourth-ranked EC DMP) were located 91 bp away from each other in the ankyrin 1 (*ANK1*) gene on chromosome 8, which encodes a brain-expressed protein[13] involved in compartmentalization of the neuronal plasma membrane[14] (**Fig. 1a**). These DMPs are also located proximal to the *NKX6-3* gene, encoding a homeodomain transcription factor involved in the development of the brain[15,16]. Increased EC DNA methylation at both CpG sites was associated with Braak stage (cg11823178: $r = 0.47$, $t_{102} = 5.39$, nominal $P = 4.59 \times 10^{-7}$; cg05066959: $r = 0.41$, $t_{102} = 5.37$, nominal $P = 1.34 \times 10^{-5}$; **Fig. 1b**). Given that AD is characterized by significant neuronal cell loss, we used an *in silico* algorithm to confirm that the observed association is not confounded by differences in neuronal cell proportions between individuals[17]; both CpG sites remained significantly associated with Braak score after correction for estimated cellular heterogeneity (cg11823178: nominal $P = 7.09 \times 10^{-7}$; cg05066959: nominal $P = 6.20 \times 10^{-6}$; **Table 1**). We used *comb-p* (ref. 18) to identify spatially correlated regions of differential DNA methylation, highlighting a Braak-associated DMR spanning these CpG sites ($P = 6.04 \times 10^{-7}$; **Supplementary Table 7**). Hypermethylation at both DMPs was significantly associated with Braak score in the STG (cg11823178: $r = 0.37$, $t_{111} = 4.15$, nominal $P = 6.51 \times 10^{-5}$; cg05066959: $r = 0.33$, $t_{111} = 3.67$, nominal $P = 3.78 \times 10^{-4}$) and the PFC (cg11823178: $r = 0.29$, $t_{108} = 3.12$, nominal $P = 2.33 \times 10^{-3}$; cg05066959: $r = 0.32$, $t_{108} = 3.52$, nominal $P = 6.48 \times 10^{-4}$) (**Fig. 1c**). In contrast, no significant neuropathology-associated hypermethylation was detected at either CpG site in the CER (cg11823178: $r = 0.01$, $t_{106} = 0.082$, nominal $P = 0.935$; cg05066959: $r = -0.08$, $t_{106} = 0.085$, nominal $P = 0.395$) (**Fig. 1d**), a region largely protected from neurodegeneration in AD, nor was elevated DNA methylation at either site associated with AD diagnosis in whole blood collected pre-mortem (data not shown).

Notably, we observe a significant overlap in Braak-associated DMPs across the three cortical regions profiled in the London discovery cohort: 38 (permuted $P < 0.005$) and 30 (permuted $P < 0.005$) of the 100 top-ranked EC probes were significantly differentially methylated in the same direction in the STG and PFC, respectively (**Supplementary Table 8**), with a highly significant correlation of top-ranked Braak-associated DNA methylation scores across these sites (EC versus STG: r = 0.88, $P = 6.73 \times 10^{-14}$; EC versus PFC: r = 0.83, $P = 8.77 \times 10^{-13}$). There was, however, a clear distinction between cortical regions and CER, with the top-ranked CER DMPs appearing to be more tissue specific and not differentially methylated in cortical regions (permuted *P* values for enrichment all > 0.05), although ~15% of the top-ranked cortical DMPs were differentially methylated in CER (permuted *P* values ≤ 0.01), indicating that these represent relatively pervasive AD-associated changes that are observed across multiple tissues. We subsequently used a meta-analysis method (Online Methods) to highlight consistent Braak-associated DNA methylation differences across all three cortical regions in the discovery cohort. The top-ranked cross-cortex DMPs are shown in **Table 2** and **Supplementary Table 9**, and DMRs identified using *comb-p* are listed in **Supplementary Table 10**. Of note, cg11823178 was the most significant cross-cortex DMP ($\Delta = 3.20$, Fisher's $P = 3.42 \times 10^{-11}$, Brown's $P = 1.00 \times 10^{-6}$), with cg05066959 again being ranked fourth ($\Delta = 4.26$, Fisher's $P = 1.24 \times 10^{-9}$,

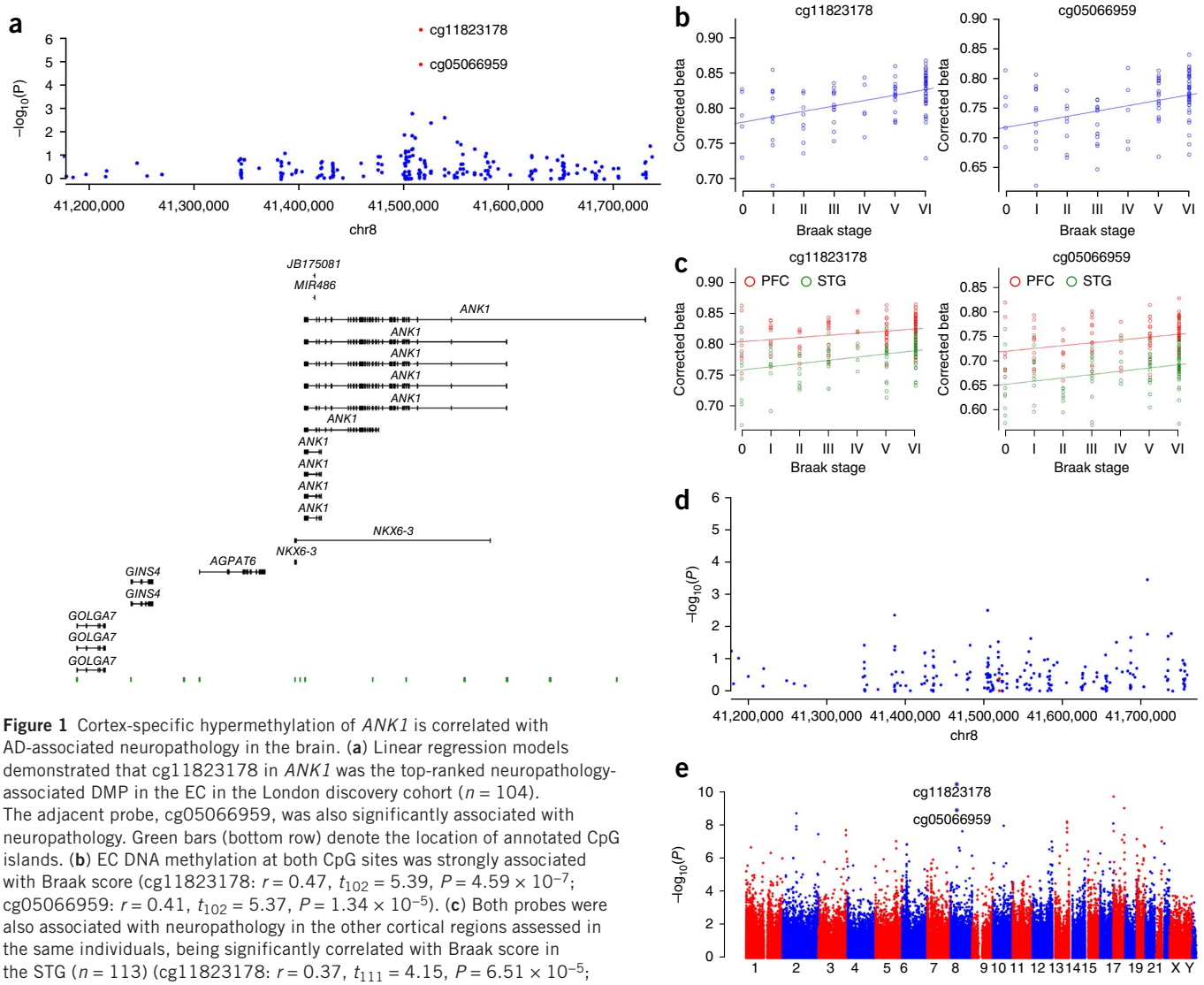

Figure 1 Cortex-specific hypermethylation of *ANK1* is correlated with AD-associated neuropathology in the brain. (**a**) Linear regression models demonstrated that cg11823178 in *ANK1* was the top-ranked neuropathology-associated DMP in the EC in the London discovery cohort ($n = 104$). The adjacent probe, cg05066959, was also significantly associated with neuropathology. Green bars (bottom row) denote the location of annotated CpG islands. (**b**) EC DNA methylation at both CpG sites was strongly associated with Braak score (cg11823178: $r = 0.47$, $t_{102} = 5.39$, $P = 4.59 \times 10^{-7}$; cg05066959: $r = 0.41$, $t_{102} = 5.37$, $P = 1.34 \times 10^{-5}$). (**c**) Both probes were also associated with neuropathology in the other cortical regions assessed in the same individuals, being significantly correlated with Braak score in the STG ($n = 113$) (cg11823178: $r = 0.37$, $t_{111} = 4.15$, $P = 6.51 \times 10^{-5}$; cg05066959: $r = 0.33$, $t_{111} = 3.67$, $P = 3.78 \times 10^{-4}$) and the PFC ($n = 110$) (cg11823178: $r = 0.29$, $t_{108} = 3.12$, $P = 2.33 \times 10^{-3}$; cg05066959: $r = 0.32$, $t_{108} = 3.52$, $P = 6.48 \times 10^{-4}$). (**d**) There was no association between DNA methylation and Braak score at either *ANK1* probe in the CER ($n = 108$) (cg11823178: $r = 0.01$, $t_{106} = 0.082$, $P = 0.935$; cg05066959: $r = -0.08$, $t_{106} = 0.085$, $P = 0.395$), a region largely protected against AD-related neuropathology. (**e**) cg11823178 was the top-ranked cross-cortex DMP (Fisher's $\chi^2(6) = 60.6$, $P = 3.42 \times 10^{-11}$), with cg05066959 also being strongly associated with Braak score (Fisher's $\chi^2(6) = 52.9$, $P = 1.24 \times 10^{-9}$).

Brown's $P = 6.24 \times 10^{-6}$; **Fig. 1e**) and a DMR spanning these probes being associated with neuropathology (Sidak-corrected $P = 3.39 \times 10^{-4}$) (**Supplementary Table 10**). Together, these data suggest that cortical DNA hypermethylation at the *ANK1* locus is robustly associated with AD-related neuropathology.

A second (replication) cortical data set was generated using DNA isolated from two regions (STG and PFC) obtained from a cohort of brains archived in the Mount Sinai Alzheimer's Disease and Schizophrenia Brain Bank ($n = 144$), with detailed neuropathology data including Braak staging and amyloid burden (Online Methods)[19]. Notably, Braak-associated DNA methylation scores for the 100 top-ranked cross-cortex DMPs identified in the London discovery cohort (**Supplementary Table 9**) were strongly correlated with neuropathology-associated differences at the same probes in both cortical regions profiled in the Mount Sinai replication cohort (STG Braak score: $r = 0.63$, $P = 2.66 \times 10^{-12}$; PFC Braak score: $r = 0.64$, $P = 6.03 \times 10^{-13}$; STG amyloid burden: $r = 0.46$, $P = 1.09 \times 10^{-6}$; PFC amyloid burden: $r = 0.65$, $P = 2.87 \times 10^{-13}$; **Fig. 2a**). Furthermore, increased DNA methylation at each of the two *ANK1* CpG sites was significantly associated with elevated Braak score (**Table 1** and **Fig. 2b**) in the STG (cg11823178: $r = 0.28$, $t_{142} = 3.62$, nominal $P = 1.63 \times 10^{-4}$; cg05066959: $r = 0.25$, $t_{142} = 3.29$, nominal $P = 5.78 \times 10^{-4}$) and PFC (cg11823178: $r = 0.24$, $t_{140} = 3.14$, nominal $P = 1.07 \times 10^{-3}$; cg05066959: $r = 0.21$, $t_{140} = 2.75$, nominal $P = 4.00 \times 10^{-3}$), and also amyloid pathology (**Fig. 2c**) in the STG (cg11823178: $r = 0.21$, $t_{142} = 2.81$, nominal $P = 4.99 \times 10^{-4}$; cg05066959: $r = 0.27$, $t_{142} = 3.47$, nominal $P = 5.65 \times 10^{-4}$) and PFC (cg11823178: $r = 0.29$, $t_{140} = 3.69$, nominal $P = 2.35 \times 10^{-4}$; cg05066959: $r = 0.19$, $t_{140} = 2.56$, nominal $P = 9.93 \times 10^{-3}$).

To further confirm the association between cortical *ANK1* hypermethylation and neuropathology, we used bisulfite-pyrosequencing to quantify DNA methylation across an extended region spanning eight CpG sites, including cg11823178 and cg05066959, in DNA extracted from a third independent collection of matched EC, STG

Table 2 The ten top-ranked cross-cortex Braak-associated DMPs

	Probe information																					
								London cohort										Mount Sinai cohort					
						Cross-cortex model			EC			STG			PFC			CER		STG		PFC	
Probe	Position	Illumina annotation	Genes with TSS within 5Kb upstream	Genes with TSS within 1Kb downstream	Fisher's *P* Value	Brown's *P* Value	Δ	*P* value	Δ	*P* value	Δ	*P* value	Δ	*P* value	Δ	*P* value	Δ	*P* value	Δ	*P* value	Δ	*P* value	Δ
cg11823178	8:41519399	*ANK1; MIR486*	*NKX6-3*	*ANK1*	3.42×10^{-11}	1.00×10^{-6}	3.20	4.59×10^{-7}	4.62	6.51×10^{-5}	3.12	2.33×10^{-3}	2.05	–	–	1.63×10^{-4}	3.36	1.07×10^{-3}	2.97				
cg25018458	17:980014	*ABR*	*TIMM22*	*ABR*	1.89×10^{-10}	9.91×10^{-7}	1.34	1.05×10^{-4}	1.29	4.36×10^{-6}	1.59	9.48×10^{-4}	1.12	–	–	4.46×10^{-5}	1.82	2.14×10^{-4}	1.93				
cg05810363	17:74475270	*RHBDF2*	*RHBDF2*	*AANAT*	9.42×10^{-10}	9.15×10^{-6}	3.36	1.37×10^{-4}	3.79	2.26×10^{-5}	3.58	7.93×10^{-4}	2.76	–	–	3.55×10^{-3}	3.65	5.54×10^{-3}	3.19				
cg05066959	8:41519308	*ANK1; MIR486*	*NKX6-3*	*ANK1*	1.24×10^{-9}	6.24×10^{-6}	4.26	1.35×10^{-5}	5.45	3.78×10^{-4}	4.03	6.48×10^{-4}	3.45	–	–	5.78×10^{-4}	4.75	4.00×10^{-3}	3.41				
cg18428542	2:91787196	–	*ACTR3BP2*	–	1.94×10^{-9}	2.53×10^{-5}	–4.00	2.09×10^{-4}	–4.05	1.73×10^{-4}	–3.93	1.47×10^{-4}	–4.04	7.46×10^{-5}	–4.43	–	–	–	–				
cg16665310	13:100218136	–	*CLYBL*	*TM9SF2*	6.11×10^{-9}	1.11×10^{-4}	–5.73	1.39×10^{-4}	–6.71	1.17×10^{-4}	–5.15	1.13×10^{-3}	–5.50	2.32×10^{-4}	–3.84	–	–	–	–				
cg05912299	13:100217962	–	*CLYBL*	*TM9SF2*	7.17×10^{-9}	1.38×10^{-4}	–6.03	6.76×10^{-5}	–7.46	3.67×10^{-3}	–4.64	8.85×10^{-5}	–6.29	9.86×10^{-4}	–5.63	–	–	–	–				
cg03169557	16:89598950	*SPG7*	*RPL13*	*SPG7*	7.95×10^{-9}	7.91×10^{-6}	1.63	3.01×10^{-4}	1.31	1.35×10^{-4}	1.97	6.05×10^{-4}	1.52	–	–	2.85×10^{-3}	2.11	7.98×10^{-4}	2.34				
cg23968456	10:73521631	*CDH23; C10orf54*	*C10orf105*	*C10orf54*	1.09×10^{-8}	1.09×10^{-5}	1.22	1.89×10^{-3}	0.87	2.21×10^{-3}	1.13	8.27×10^{-6}	1.62	–	–	3.07×10^{-3}	1.72	0.0416	1.05				
cg02672452	2:91818189	–	*ACTR3BP2*	–	1.16×10^{-8}	1.27×10^{-4}	–4.42	4.71×10^{-5}	–5.29	2.35×10^{-3}	–3.64	3.33×10^{-4}	–4.49	6.75×10^{-4}	–5.87	–	–	–	–				

Shown for each DMP are chromosomal location (hg19), up/downstream genes, Fisher's *P* value, Brown's *P* value, and cross-cortex estimate (Δ) of methylation difference between individuals with the lowest (score 0) and highest (score VI) Braak stage. Also shown are differences (Δ) between Braak 0 and Braak IV and corresponding *P* values from our individual cortex models in the London discovery cohort (Online Methods) and the STG and PFC in the Mount Sinai replication cohort for CpG sites showing a nominally significant difference in the same direction. The 100 top-ranked cross-cortex Braak-associated DMPs are given in **Supplementary Table 9**.

and PFC tissue (n = 62) obtained from the Thomas Willis Oxford Brain Collection[20] (Online Methods and **Supplementary Table 11a**). Average DNA methylation across this region was significantly elevated in all three cortical regions tested (EC, P = 0.0004; STG, P = 0.0008; PFC, P = 0.014) in affected individuals (**Supplementary Fig. 1**), most notably in the EC, where six of the eight CpG sites assessed were characterized by significant AD-associated hypermethylation (**Fig. 2d**). A meta-analysis of cg11823178 and cg05066959 across all three independent cohorts confirmed consistent neuropathology-associated hypermethylation in each of the cortical regions assessed (**Fig. 2e**,**f**). Further evidence to support our conclusions came from an independent EWAS of AD pathology in 708 cortical samples (De Jager *et al.*)[21]. There was a significant correlation ($r = 0.57$, $P = 1.55 \times 10^{-9}$) between the 100 top-ranked DNA methylation changes identified in our cross-cortex analyses and neuropathology-associated differences at the same probes in the study by De Jager *et al.* (**Fig. 2g**)[21]. Conversely, neuropathology-associated DNA methylation scores for top-ranked DMPs in De Jager *et al.*[21] were strongly correlated ($r = 0.49$, $P = 7.8 \times 10^{-10}$) with those that we observed using the cross-cortex model for the same probes in our discovery cohort (**Supplementary Fig. 2**). In particular, De Jager *et al.*[21] also identified a highly significant association between elevated DNA methylation at cg11823178 and cg05066959 and AD-related neuropathology. Together, these data provide compelling evidence for an association between *ANK1* hypermethylation and the neuropathological features of AD, specifically in the cortical regions associated with disease manifestation. Although not previously implicated in dementia, genetic variation in *ANK1* is associated with diabetic phenotypes[22–24], an interesting observation given the established links between type 2 diabetes and AD[25].

ANK1 is a transcriptionally complex gene, with multiple isoforms and several alternative promoters having been identified (**Supplementary Fig. 3**). Given the established role of DNA methylation in regulating isoform-specific gene expression, we examined whether AD neuropathology was associated with the differential abundance of various *ANK1* isoforms in the EC using quantitative PCR (Online Methods). Briefly, three assays with specificity to *ANK1* isoforms 1–4, 9, and 5,7 and 10 (**Supplementary Table 11b**) were used to profile 36 EC samples from whom high quality RNA was available (**Supplementary Table 2**). Our linear model highlighted a significant association (P = 0.04) between the abundance of isoform 5, 7 and 10 transcripts and AD-associated neuropathology (**Supplementary Fig. 4**). No significant differences in transcript levels ($P > 0.05$) were observed for the other two isoform-specific assays (data not shown).

As a definitive diagnosis of AD can only be made via neuropathological examination at autopsy, there is considerable interest in the identification of clinical biomarkers that may have both diagnostic and prognostic utility during the early stages of the disorder[26,27]. Recent work has identified several transcriptomic blood biomarkers for AD with potential clinical utility for the early diagnosis of the disease[28–32]. In this study, we had access to matched pre-mortem whole-blood DNA for methylomic profiling from a subset of samples in the London discovery cohort (n = 57). Because of the duration elapsed between blood sampling and mortality (average = 4.15 ± 3.00 years), analyses on these data were restricted to the identification of DMPs associated with a clinical diagnosis of AD, rather than Braak score. We identified a number of AD-associated DMPs (**Supplementary Table 12**) in pre-mortem blood, many in the vicinity of genes of relevance to AD, including *DAPK1* (cg14067233), which has been implicated in genetic studies[33,34], *GAS1* (cg14067233), an APP-interacting protein involved in the control of APP maturation and processing[35], and *NDUFS5*

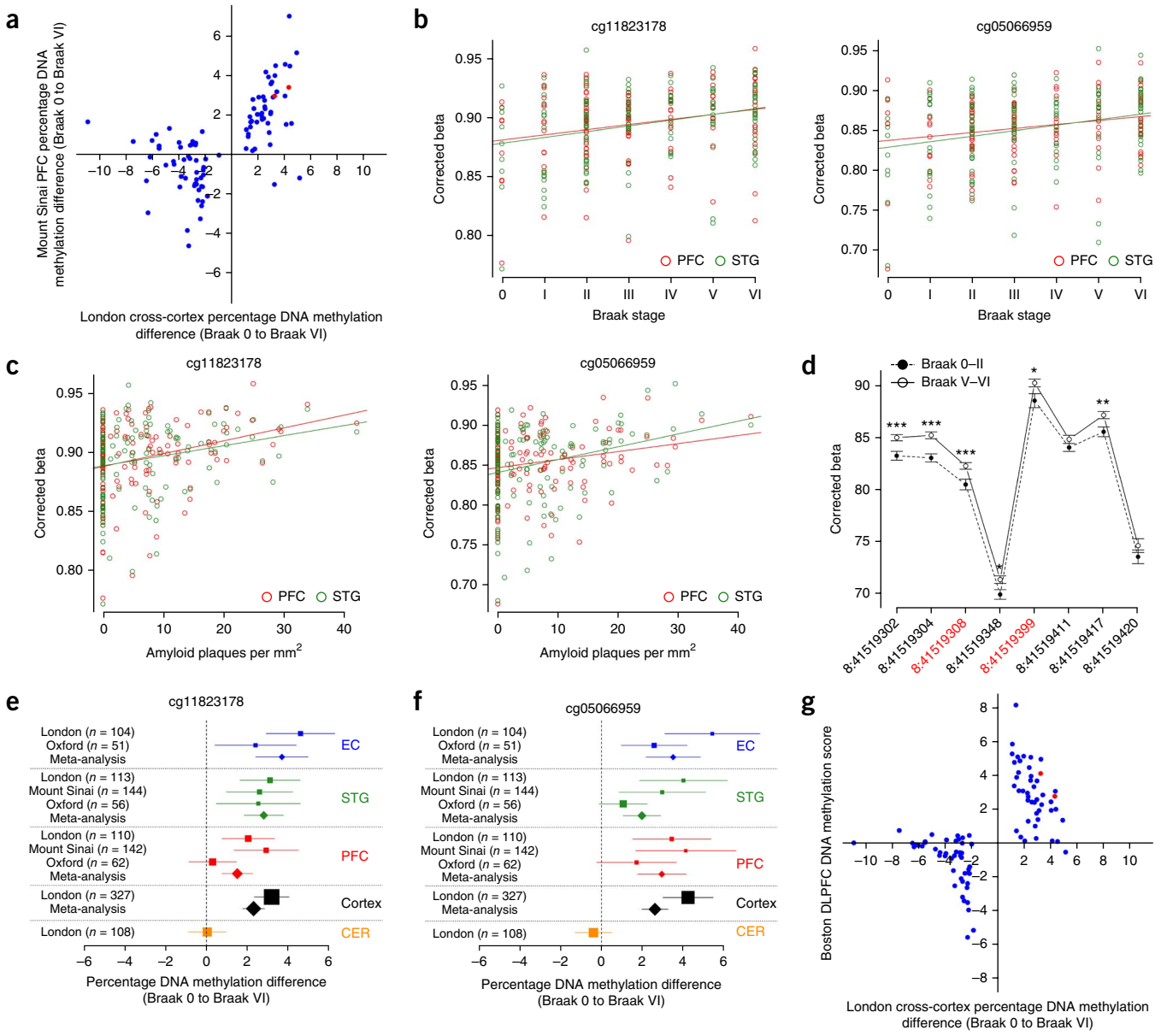

Figure 2 Neuropathology-associated DMPs are consistent across sample cohorts, with replicated evidence for *ANK1* hypermethylation. (**a**) Braak-associated DNA methylation scores for the top-ranked cross-cortex DMPs identified using linear regression models in the London discovery cohort (**Supplementary Table 9**) were significantly correlated with neuropathology-associated differences at the same probes in both cortical regions profiled in the Mount Sinai replication cohort using linear regression models (PFC ($n = 142$) Braak score: $r = 0.64$, $P = 6.03 \times 10^{-13}$; STG ($n = 144$) Braak score: $r = 0.63$, $P = 2.66 \times 10^{-12}$; PFC amyloid burden: $r = 0.65$, $P = 2.87 \times 10^{-13}$; STG amyloid burden: $r = 0.46$, $P = 1.09 \times 10^{-6}$). Shown is data for Mount Sinai PFC Braak score analysis, with the two *ANK1* probes (cg11823178 and cg05066959) highlighted in red. (**b**,**c**) cg11823178 and cg05066959 were significantly associated with Braak score in the STG (cg11823178: $r = 0.28$, $t_{142} = 3.62$, $P = 1.63 \times 10^{-4}$; cg05066959: $r = 0.25$, $t_{142} = 3.29$, $P = 5.78 \times 10^{-4}$) and PFC (cg11823178: $r = 0.24$, $t_{140} = 3.14$, $P = 1.07 \times 10^{-3}$; cg05066959: $r = 0.21$, $t_{140} = 2.75$, $P = 4.00 \times 10^{-3}$) (**b**), and amyloid pathology in the STG (cg11823178: $r = 0.21$, $t_{142} = 2.81$, $P = 4.99 \times 10^{-4}$; cg05066959: $r = 0.27$, $t_{142} = 3.47$, $P = 5.65 \times 10^{-4}$) and PFC (cg11823178: $r = 0.29$, $t_{140} = 3.69$, $P = 2.35 \times 10^{-4}$; cg05066959: $r = 0.19$, $t_{140} = 2.56$, $P = 9.93 \times 10^{-3}$) (**c**). (**d**) In the Oxford replication cohort, bisulfite-pyrosequencing was used to quantify DNA methylation across eight CpG sites spanning an extended *ANK1* region. Linear models, adjusting for age and gender, confirmed significant neuropathology-associated hypermethylation in all three of the cortical regions assessed (**Supplementary Fig. 1**), most notably in the EC ($n = 51$), where six of the eight CpG sites showed a significant (amplicon average $P = 0.0004$) neuropathology-associated increase in DNA methylation (data is represented as mean ± s.e.m., $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.005$). (**e**,**f**) Meta-analyses across the three sample cohorts (London, Mount Sinai and Oxford) confirmed Braak-associated cortex-specific hypermethylation for both cg11823178 (**e**) and cg05066959 (**f**). Finally, there was a marked consistency in neuropathology-associated DMPs identified in our discovery cohort and those identified in De Jager *et al.*[21]. (**g**) Braak-associated DNA methylation scores for the 100 top-ranked cross-cortex DMPs identified in the London discovery cohort were significantly correlated with neuropathology-associated differences (neuritic-plaque load) at the same probes in the dorsolateral prefrontal cortex (DLPFC) identified by De Jager *et al.*[21] in 708 individuals ($r = 0.57$, $P = 1.55 \times 10^{-9}$). The two *ANK1* probes (cg11823178 and cg05066959) are highlighted in red.

(cg17074958), a mitochondrial gene that has been shown to be differentially expressed in AD blood[36]. Our data suggest, however, that the top-ranked DMPs in blood are distinct from those identified in the brain; there was no significant overlap with either cortex or CER (permuted *P* values for enrichment in EC = 0.89, PFC = 0.40, STG = 0.45, CER = 0.41), suggesting that AD-associated DMPs in blood are unlikely to be directly related to the actual neurodegenerative process itself. Using data from our previous independent blood-based transcriptomic analyses of both AD and mild cognitive impairment (MCI)[36], however, we observed that 18 of our top-ranked blood DMPs are located in the vicinity of known differentially expressed transcripts (**Supplementary Table 13**). These data suggest that, although distinct from AD-associated changes occurring in the brain, many of the AD-associated DMPs identified in blood before death may mediate detectable transcriptomic changes and, given the relative stability and ease of profiling DNA modifications compared to RNA, have potential utility as diagnostic biomarkers of the disorder.

DISCUSSION

We identified evidence for cortex-specific hypermethylation at CpG sites in the *ANK1* gene associated with AD neuropathology. Definitively distinguishing cause from effect in epigenetic epidemiology is difficult, especially for disorders such as AD that manifest in inaccessible tissues such as the brain and are not amenable to longitudinal study[9,10]. However, our observation of highly consistent changes across multiple regions of the cortex in several independent sample cohorts suggests that the identified loci are directly relevant to the pathogenesis of AD. In this regard, the *ANK1* DMR reported here, and subsequently confirmed by De Jager *et al.*[21], represents one of the most robust molecular associations with AD yet identified.

Epigenetic studies must overcome a number of potential confounds (for example, tissue specificity, age, sex) that make study design and replication strategies vital[9]. One issue in EWAS analyses using platforms such as the Illumina 450K array relates to potential technical artifacts caused by genetic variation, although we are confident that the DMPs that we identified do not result from polymorphisms in (or flanking) the assayed CG dinucleotides. We used a stringent two-pronged strategy to exclude these effects: the direct exclusion of probes known to be affected by common single nucleotide polymorphisms (SNPs) and the statistical filtering of extreme sample outliers in individual probe data that are frequently caused by rare SNPs (Online Methods). Although we were unable to explore the extent to which AD-associated variation is driven by genetic variation, the role of genetic-epigenetic interactions in complex disease represents an important area for further study[37]. Power calculations for EWAS analyses are difficult, especially given the paucity of existing data for brain DNA methylation and limited information about the extent of inter-individual variation occurring at individual CpG sites. Conventional methods for multiple-test correction, such as those used in genome-wide association studies, are likely to be overly stringent and inappropriate given the non-independence of DNA methylation across multiple CpG sites and lack of inter-individual variation at many loci; in this study, we therefore report nominal *P* values[9,10,38], with stringent validation in multiple cohorts. Furthermore, studies investigating the role of epigenetic dysfunction in complex brain diseases such as AD are in their infancy, and no real precedents have yet been set about the optimal sample sizes needed to detect them[9]. Our conservative power calculation (Online Methods) suggests that we are well-powered to identify relatively small (~5%) DNA methylation differences between groups for the majority of probes on the Illumina 450K array. More importantly, our study represents the largest cross-tissue study of AD using matched DNA from both affected and unaffected brain regions, and, to the best of our knowledge, the first to employ a sequential replication design incorporating independent study cohorts and two independent technologies (Illumina 450K array and bisulfite-pyrosequencing). The marked overlap between DMPs identified across our sample cohorts (**Fig. 2a**) and those identified by De Jager *et al.*[21] (**Fig. 2g** and **Supplementary Fig. 2**) suggests that our study was adequately powered to detect robust AD-associated differences that can be replicated in other studies.

In summary, our data provide evidence for extensive differences in DNA methylation across brain regions in AD. Our analyses of multiple brain regions obtained from three independent cohorts suggest a role for cortex-specific hypermethylation across a region in *ANK1* in AD-associated neuropathology, with methylomic changes mirroring known patterns of neuropathology and being most substantial in the EC. This finding is strengthened by the independent identification of the same DMR in another large EWAS of AD[21]. Finally, although most brain-identified DMPs, including *ANK1*, are not detected in blood, we did identify multiple AD-associated DNA methylation differences in pre-mortem blood samples, many of which are located in the vicinity of genes that have been found to be transcriptionally altered even in patients with MCI during the early stages of cognitive decline. To conclude, our study represents the first EWAS of AD employing a sequential replication design across multiple tissues and highlights the power of this approach more broadly for the identification of disease-associated DMRs.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Raw data has been deposited in GEO under accession number GSE59685.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank C. Sloan for technical support and I. Bodi and A. King for neuropathological diagnosis of cases. We also thank the Oxford Project to Investigate Memory and Ageing (OPTIMA), the National Institute for Health (NIHR) Biomedical Research Unit in Dementia in the South London and Maudsley NHS Foundation Trust (SLaM), Brains for Dementia Research (Alzheimer Brain Bank, UK), and the donors and families who made this research possible. Blood samples from the London cohort were collected as part of the Alzheimer's Research UK funded study "Biomarkers of AD Neurodegeneration". This work was funded by US National Institutes of Health grant R01 AG036039 to J.M. and an Equipment Grant from Alzheimer's Research UK. The Oxford Brain Bank is supported in part by the NIHR Oxford Biomedical Research Centre based at Oxford University Hospitals NHS Trust and University of Oxford. Brain banking and neuropathology assessments for the Mount Sinai cohort was supported by US National Institutes of Health grants AG02219, AG05138 and MH064673, and the Department of Veterans Affairs VISN3 MIRECC. Replication work in Boston was supported by US National Institutes of Health grants: R01 AG036042, R01AG036836, R01 AG17917, R01 AG15819, R01 AG032990, R01 AG18023, RC2 AG036547, P30 AG10161, P50 AG016574, U01 ES017155, KL2 RR024151 and K25 AG041906-01.

AUTHOR CONTRIBUTIONS

K.L., R.S., R.M., M.V., D.C. and J.B. conducted laboratory experiments. J.M. conceived and supervised the project and obtained funding. E.H., L.C.S., R.S. and K.L. undertook data analyses and bioinformatics. C.T., S.L., J.P., S.A.-S., P.K., V.H. and C.J. provided samples for analysis. P.L.D.J., G.S. and D.A.B. provided replication data. Z.K. provided help with cellular heterogeneity correction. L.W.H. provided help with the alternative splicing assays. J.M., K.L. and L.C.S. drafted the manuscript. All of the authors read and approved the final submission.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Brookmeyer, R., Johnson, E., Ziegler-Graham, K. & Arrighi, H.M. Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement.* **3**, 186–191 (2007).
2. Knapp, M. & Prince, M. Dementia UK: the Full report. (Alzheimer's Society, 2007).
3. Hardy, J.A. & Higgins, G.A. Alzheimer's disease: the amyloid cascade hypothesis. *Science* **256**, 184–185 (1992).
4. Hardy, J. & Selkoe, D.J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297**, 353–356 (2002).
5. Wenk, G.L. Neuropathologic changes in Alzheimer's disease. *J. Clin. Psychiatry* **64** (suppl. 9), 7–10 (2003).
6. Lambert, J.C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
7. Encode Project Consortium *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
8. Lunnon, K. & Mill, J. Epigenetic studies in Alzheimer's disease: current findings, caveats, and considerations for future studies. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **162B**, 789–799 (2013).
9. Mill, J. & Heijmans, B.T. From promises to practical strategies in epigenetic epidemiology. *Nat. Rev. Genet.* **14**, 585–594 (2013).
10. Murphy, T.M. & Mill, J. Epigenetics in health and disease: heralding the EWAS era. *Lancet* **383**, 1952–1954 (2014).
11. Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**, 293 (2013).
12. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol.* **82**, 239–259 (1991).
13. Kordeli, E. & Bennett, V. Distinct ankyrin isoforms at neuron cell bodies and nodes of Ranvier resolved using erythrocyte ankyrin-deficient mice. *J. Cell Biol.* **114**, 1243–1259 (1991).
14. Boiko, T. *et al.* Ankyrin-dependent and -independent mechanisms orchestrate axonal compartmentalization of L1 family members neurofascin and L1/neuron-glia cell adhesion molecule. *J. Neurosci.* **27**, 590–603 (2007).
15. Nelson, S.B., Janiesch, C. & Sander, M. Expression of Nkx6 genes in the hindbrain and gut of the developing mouse. *J. Histochem. Cytochem.* **53**, 787–790 (2005).
16. Alanentalo, T. *et al.* Cloning and analysis of Nkx6.3 during CNS and gastrointestinal development. *Gene Expr. Patterns* **6**, 162–170 (2006).
17. Guintivano, J., Aryee, M. & Kaminsky, Z. 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).
18. Pedersen, B.S., Schwartz, D.A., Yang, I.V. & Kechris, K.J. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics* **28**, 2986–2988 (2012).
19. Haroutunian, V. *et al.* Regional distribution of neuritic plaques in the nondemented elderly and subjects with very mild Alzheimer disease. *Arch. Neurol.* **55**, 1185–1191 (1998).
20. Esiri, M.M. Brain banks: the Oxford experience. *J. Neural Transm. Suppl.* **39**, 25–30 (1993).
21. De Jager, P.L. *et al.* Alzheimer's disease pathology is associated with early alterations in brain DNA methylation at ANK1, BIN1 and other loci. *Nat. Neurosci.* doi:10.1038/nn.3786 (17 August 2014).
22. Soranzo, N. *et al.* Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemic and nonglycemic pathways. *Diabetes* **59**, 3229–3239 (2010).
23. Imamura, M. *et al.* A single-nucleotide polymorphism in ANK1 is associated with susceptibility to type 2 diabetes in Japanese populations. *Hum. Mol. Genet.* **21**, 3042–3049 (2012).
24. Harder, M.N. *et al.* Type 2 diabetes risk alleles near BCAR1 and in ANK1 associate with decreased beta-cell function whereas risk alleles near ANKRD55 and GRB14 associate with decreased insulin sensitivity in the Danish Inter99 cohort. *J. Clin. Endocrinol. Metab.* **98**, E801–E806 (2013).
25. Barnes, D.E. & Yaffe, K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. *Lancet Neurol.* **10**, 819–828 (2011).
26. Lovestone, S. Searching for biomarkers in neurodegeneration. *Nat. Med.* **16**, 1371–1372 (2010).
27. Hooper, C., Lovestone, S. & Sainz-Fuertes, R. Alzheimer's disease, diagnosis and the need for biomarkers. *Biomark. Insights* **3**, 317–323 (2008).
28. Fehlbaum-Beurdeley, P. *et al.* Toward an Alzheimer's disease diagnosis via high-resolution blood gene expression. *Alzheimers Dement.* **6**, 25–38 (2010).
29. Rye, P.D. *et al.* A novel blood test for the early detection of Alzheimer's disease. *J. Alzheimers Dis.* **23**, 121–129 (2011).
30. Booij, B.B. *et al.* A gene expression pattern in blood for the early detection of Alzheimer's disease. *J. Alzheimers Dis.* **23**, 109–119 (2011).
31. Lunnon, K. *et al.* A blood gene expression marker of early Alzheimer's disease. *J. Alzheimers Dis.* **33**, 737–753 (2013).
32. Fehlbaum-Beurdeley, P. *et al.* Validation of AclarusDx, a blood-based transcriptomic signature for the diagnosis of Alzheimer's disease. *J. Alzheimers Dis.* **32**, 169–181 (2012).
33. Wu, Z.C. *et al.* Association of DAPK1 genetic variations with Alzheimer's disease in Han Chinese. *Brain Res.* **1374**, 129–133 (2011).
34. Laumet, G. *et al.* Systematic analysis of candidate genes for Alzheimer's disease in a French, genome-wide association study. *J. Alzheimers Dis.* **20**, 1181–1188 (2010).
35. Chapuis, J., Vingtdeux, V., Campagne, F., Davies, P. & Marambaud, P. Growth arrest–specific 1 binds to and controls the maturation and processing of the amyloid-beta precursor protein. *Hum. Mol. Genet.* **20**, 2026–2036 (2011).
36. Lunnon, K. *et al.* Mitochondrial dysfunction and immune activation are detectable in early Alzheimer's disease blood. *J. Alzheimers Dis.* **30**, 685–710 (2012).
37. Meaburn, E.L., Schalkwyk, L.C. & Mill, J. Allele-specific methylation in the human genome: implications for genetic studies of complex disease. *Epigenetics* **5**, 578–582 (2010).
38. Shoemaker, R., Deng, J., Wang, W. & Zhang, K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res.* **20**, 883–889 (2010).

ONLINE METHODS

Subjects and samples. Brain tissue was obtained from three independent sample cohorts, enabling us to take a powerful cross-tissue sequential-replication approach to identifying DNA methylation differences in AD. Our discovery cohort comprised of EC, STG, PFC and CER tissue, in addition to whole blood where available, obtained from 122 individuals archived in the MRC London Neurodegenerative Disease Brain Bank (http://www.kcl.ac.uk/iop/depts/cn/research/MRC-London-Neurodegenerative-Diseases-Brain-Bank/MRC-London-Neurodegenerative-Diseases-Brain-Bank.aspx). Ethical approval for the study was provided by the NHS South East London REC 3. Matched blood samples collected before death were available for a subset of individuals (**Supplementary Tables 1** and **2**) as part of the Alzheimer's Research UK funded study "Biomarkers of AD Neurodegeneration", with informed consent according to the Declaration of Helsinki (1991). For validation purposes STG and PFC tissue was obtained from 144 individuals archived in the Mount Sinai Alzheimer's Disease and Schizophrenia Brain Bank (http://icahn.mssm.edu/research/labs/neuropathology-and-brain-banking)[19] and EC, STG and PFC samples from an additional 62 individuals archived in the Thomas Willis Oxford Brain Collection (http://www.medsci.ox.ac.uk/optima/information-for-patients-and-the-public/the-thomas-willis-oxford-brain-collection)[20]. All samples were dissected by trained specialists, snap-frozen and stored at −80 °C. Further information about the samples is given in **Supplementary Tables 1** and **2**. Genomic DNA was isolated from ~100 mg of each dissected brain region or whole blood stored in EDTA collection tubes using a standard phenol-chloroform extraction method, and tested for degradation and purity before analysis.

Power. Power calculations for EWAS analyses are difficult given the paucity of existing data for brain DNA methylation and limited information about the extent of inter-individual variation occurring at individual CpG sites[9]. As we have previously discussed, studies investigating the role of epigenetic dysfunction in complex brain diseases such as AD are in their infancy, and no real precedents have yet been set about the optimal sample-sizes needed to detect them[9]. Conventional methods for multiple-test correction, such as those used in genome-wide association studies, are likely to be overly stringent and inappropriate given the non-independence of DNA methylation across multiple CpG sites and lack of inter-individual variation at many loci; in this study, we therefore report nominal *P* values. However, a conservative power calculation using methylome data from this and other ongoing studies in our laboratory[11,39–41] suggests that we are well-powered to identify DNA methylation differences of ~5% between groups for the majority of probes on the Illumina 450K array based conservatively on a case-control *t* test with an array-wide Bonferroni threshold and the observed distribution of beta-value variances for the entorhinal cortex data set. Notably, our study represents the largest cross-tissue study of AD using DNA from both affected and unaffected brain regions, and the first to employ a sequential replication design incorporating three independent study cohorts and two independent technologies (Illumina 450K array and bisulfite-pyrosequencing). The marked overlap between DMPs identified across our sample cohorts (**Fig. 2a**), and with those identified by De Jager *et al.*[21] (**Fig. 2g** and **Supplementary Fig. 2**), suggests that our study was adequately powered to detect robust AD-associated differences.

Methylomic profiling. 500 ng DNA from each sample was sodium bisulfite-treated using the Zymo EZ 96 DNA methylation kit (Zymo Research) according to the manufacturer's standard protocol. Samples were assessed using the Illumina Infinium HumanMethylation450K BeadChip (Illumina) using a Illumina HiScan System (Illumina). All samples were assigned a unique code for the purpose of the experiment and grouped by tissue and randomized with respect to sex and disease status to avoid batch effects, and processed in batches of four BeadChips. Illumina Genome Studio software was used to extract the raw signal intensities of each probe (without background correction or normalization).

Data analysis. All computations and statistical analyses were performed using R 3.0.2 (ref. 42) and Bioconductor 2.13 (ref. 43). Signal intensities were imported into R using the methylumi package[44] as a methylumi object. Initial quality control checks were performed using functions in the methylumi package to assess concordance between reported and genotyped gender. Non-CpG SNP probes on the array were also used to confirm that all four brain regions and matched bloods were sourced from the same individual in the London Cohort and two brain regions in the Mount Sinai cohort where expected. Data was pre-processed in the R package *wateRmelon* using the *dasen* function as previously described[11]. Array data for each of the tissues was normalized separately and initial analyses were performed separately by tissue. The effects of age and sex were regressed out before subsequent analysis. For identification of DMPs specifically altered with respect to neuropathological measures of AD, we performed a quantitative analysis in which samples were analyzed using linear regression models in respect to Braak stage (London n = 104 (EC), 113 (STG), 110 (PFC) and 108 (CER); Mount Sinai n = 144 (STG) and 142 (PFC)) and amyloid burden (Mount Sinai n = 144 (STG) and 142 (PFC)). We used a two-level strategy for avoiding spurious signals due to SNPs rather than DNA methylation differences. Probes with common (MAF > 5%) SNPs in the CG or single base extension position or probes that are nonspecific or mismapped were flagged and disregarded in the evaluation of our results[45]. To also clean up rarer SNPs while discarding minimum data, in each tissue, and for each probe, we discarded beta values lying more than four times the interquartile range from the mean; these extreme outliers are generally the result of polymorphisms. Data was analyzed separately in each brain region using linear regression with probes ranked according to *P* value and Q-Q plots assessed to check for *P* value inflation (**Supplementary Fig. 5**). To identify differentially methylated regions (DMRs), we identified spatially correlated *P* values in our data using the Python module *comb-p* (ref. 18) to group ≥4 spatially correlated CpGs in a 500-bp sliding window. The CETS package in R[17] was used to check whether our top-ranked DMPs were mediated by the effect of differential neuronal cell proportions across samples. To identify probes with consistent associations between Braak stage and methylation across the three cortical regions, we employed a meta-analysis of EC, STG and PFC. *P* values from the individual region results for each site were generated using Fisher's method and (as a way of controlling for the covariance of the samples which come from the same individuals) Brown's method. Raw data has been deposited in GEO under accession number GSE59685.

Targeted replication using bisulfite-pyrosequencing. Bisulfite pyrosequencing was used to quantify DNA methylation across eight individual *ANK1* CpG sites, including cg05066959 and cg11823178, spanning from 41519302 to 41519420 in chromosome 8 (hg19). A single amplicon (246 bp) was amplified using primers designed using the PyroMark Assay Design software 2.0 (Qiagen) (**Supplementary Table 11a**), and sequenced using two sequencing primers to maximize coverage across eight CpG sites. DNA methylation was quantified in 62 samples in the Oxford replication cohort using the Pyromark Q24 system (Qiagen) following the manufacturer's standard instructions and the Pyro Q24 CpG 2.0.6 software. Data was adjusted for the effects of age and sex. An analysis was performed to compare samples with Braak scores 0-II to samples with Braak scores V-VI at individual CpGs and amplicon-averaged DNA methylation.

Transcript variant analysis. A subset of samples from the London cohort was selected for RNA analyses. RNA was extracted from 30 mg of brain tissue using the Qiagen RNeasy mini kit and those with a concentration >90 ng μl^{-1} and an RNA integrity number (RIN) >7 (N = 36) were used for subsequent quantitative PCR (qPCR) (**Supplementary Table 2**). 20 µl cDNA was synthesized from 1,300 ng total RNA using the SuperScript VILO cDNA Synthesis Kit according to the manufacturer's protocol and diluted five to tenfold for qPCR, depending on the downstream assay. Off the shelf TaqMan Gene Expression assays (Life Technologies) were purchased for the five housekeeping genes (*EIF4A2*, *GAPDH*, *ACTB*, *SF3A1*, *UBC*) identified as being most stably expressed in the brain using GeNORM (Primer Design). At least ten known protein coding splice variants for *ANK1* have been characterized (**Supplementary Fig. 3**), and we were able to design three custom TaqMan Gene Expression assays to target variants 1–4, 5, 7 and 10, and 9 (**Supplementary Table 11b**). qPCR was performed using TaqMan Gene Expression Mastermix (Life Technologies) for each sample in duplicate on an ABI7900HT according to the manufacturer's protocol. The abundance of *ANK1* transcript variants was determined by relative quantification to the geometric mean of the five housekeeping genes. Data was adjusted for the effect of age and sex and linear models used to analyze variant levels with respect to Braak score.

A **Supplementary Methods Checklist** is available.

39. Davies, M.N. *et al.* Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biol.* **13**, R43 (2012).

doi:10.1038/nn.3782

40. Dempster, E.L. *et al.* Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. *Hum. Mol. Genet.* **20**, 4786–4796 (2011).
41. Wong, C.C. *et al.* Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Mol. Psychiatry* **19**, 495–503 (2014).
42. R Development Core Team. R: a language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna, Austria, 2012).
43. Gentleman, R.C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
44. Davis, S., Du, P., Bilke, S., Triche, J. & Bootwalla, M. *Methylumi: Handle Illumina Methylation Data, v2.10.0* (R Graphic Manual, 2014).
45. Chen, Y.A. *et al.* Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* **8**, 203–209 (2013).