



This is a repository copy of *Meta-analysis of genome-wide DNA methylation identifies shared associations across neurodegenerative disorders*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/173063/>

Version: Published Version

---

**Article:**

Nabais, MF, Laws, SM, Lin, T et al. (67 more authors) (2021) Meta-analysis of genome-wide DNA methylation identifies shared associations across neurodegenerative disorders. *Genome Biology*, 22. 90. ISSN 1474-760X

<https://doi.org/10.1186/s13059-021-02275-5>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:  
<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

RESEARCH

Open Access



# Meta-analysis of genome-wide DNA methylation identifies shared associations across neurodegenerative disorders

Marta F. Nabais<sup>1,2</sup>, Simon M. Laws<sup>3</sup>, Tian Lin<sup>1</sup>, Costanza L. Vallergera<sup>1,4</sup>, Nicola J. Armstrong<sup>5</sup>, Ian P. Blair<sup>6</sup>, John B. Kwok<sup>7</sup>, Karen A. Mather<sup>8,9</sup>, George D. Mellick<sup>10</sup>, Perminder S. Sachdev<sup>8,11</sup>, Leanne Wallace<sup>1</sup>, Anjali K. Henders<sup>1</sup>, Ramona A. J. Zwamborn<sup>12</sup>, Paul J. Hop<sup>12</sup>, Katie Lunnon<sup>2</sup>, Ehsan Pishva<sup>2</sup>, Janou A. Y. Roubroeks<sup>2</sup>, Hilka Soininen<sup>13</sup>, Magda Tsolaki<sup>14</sup>, Patrizia Mecocci<sup>15</sup>, Simon Lovestone<sup>16</sup>, Iwona Kloszewska<sup>17</sup>, Bruno Vellas<sup>18</sup>, the Australian Imaging Biomarkers and Lifestyle study, the Alzheimer's Disease Neuroimaging Initiative, Sarah Furlong<sup>19</sup>, Fleur C. Garton<sup>1</sup>, Robert D. Henderson<sup>20,21,22</sup>, Susan Mathers<sup>23</sup>, Pamela A. McCombe<sup>21,22</sup>, Merrilee Needham<sup>24,25,26</sup>, Shyuan T. Ngo<sup>20,21,27</sup>, Garth Nicholson<sup>28</sup>, Roger Pamphlett<sup>29</sup>, Dominic B. Rowe<sup>19</sup>, Frederik J. Steyn<sup>22,30</sup>, Kelly L. Williams<sup>19</sup>, Tim J. Anderson<sup>31,32</sup>, Steven R. Bentley<sup>33</sup>, John Dalrymple-Alford<sup>31,34</sup>, Javed Fowder<sup>10</sup>, Jacob Gratten<sup>35,36</sup>, Glenda Halliday<sup>37</sup>, Ian B. Hickie<sup>37</sup>, Martin Kennedy<sup>38</sup>, Simon J. G. Lewis<sup>37</sup>, Grant W. Montgomery<sup>1</sup>, John Pearson<sup>39</sup>, Toni L. Pitcher<sup>31,32</sup>, Peter Silburn<sup>20</sup>, Futao Zhang<sup>1</sup>, Peter M. Visscher<sup>1</sup>, Jian Yang<sup>1,40,41</sup>, Anna J. Stevenson<sup>42</sup>, Robert F. Hillary<sup>42</sup>, Riccardo E. Marioni<sup>42</sup>, Sarah E. Harris<sup>43</sup>, Ian J. Deary<sup>43</sup>, Ashley R. Jones<sup>44</sup>, Aleksey Shatunov<sup>44</sup>, Alfredo Iacoangeli<sup>44</sup>, Wouter van Rheenen<sup>12</sup>, Leonard H. van den Berg<sup>12</sup>, Pamela J. Shaw<sup>45</sup>, Cristopher E. Shaw<sup>44</sup>, Karen E. Morrison<sup>46</sup>, Ammar Al-Chalabi<sup>44,47</sup>, Jan H. Veldink<sup>12</sup>, Ellis Hannon<sup>2</sup>, Jonathan Mill<sup>2,48</sup>, Naomi R. Wray<sup>1,20</sup> and Allan F. McRae<sup>1\*</sup>

\* Correspondence: [a.mcrae@uq.edu.au](mailto:a.mcrae@uq.edu.au)

Data used in the preparation of this article was obtained from the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO). Unless otherwise listed, AIBL researchers contributed data but did not participate in analysis or writing of this report. AIBL researchers are listed at [www.aibl.csiro.au](http://www.aibl.csiro.au)

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://adni.loni.usc.edu)). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http://adni.loni.usc.edu/wp-content/](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf)

[uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](#)

<sup>1</sup>Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia  
Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Abstract

**Background:** People with neurodegenerative disorders show diverse clinical syndromes, genetic heterogeneity, and distinct brain pathological changes, but studies report overlap between these features. DNA methylation (DNAm) provides a way to explore this overlap and heterogeneity as it is determined by the combined effects of genetic variation and the environment. In this study, we aim to identify shared blood DNAm differences between controls and people with Alzheimer's disease, amyotrophic lateral sclerosis, and Parkinson's disease.

**Results:** We use a mixed-linear model method (MOMENT) that accounts for the effect of (un)known confounders, to test for the association of each DNAm site with each disorder. While only three probes are found to be genome-wide significant in each MOMENT association analysis of amyotrophic lateral sclerosis and Parkinson's disease (and none with Alzheimer's disease), a fixed-effects meta-analysis of the three disorders results in 12 genome-wide significant differentially methylated positions. Predicted immune cell-type proportions are disrupted across all neurodegenerative disorders. Protein inflammatory markers are correlated with profile sum-scores derived from disease-associated immune cell-type proportions in a healthy aging cohort. In contrast, they are not correlated with MOMENT DNAm-derived profile sum-scores, calculated using effect sizes of the 12 differentially methylated positions as weights.

**Conclusions:** We identify shared differentially methylated positions in whole blood between neurodegenerative disorders that point to shared pathogenic mechanisms. These shared differentially methylated positions may reflect causes or consequences of disease, but they are unlikely to reflect cell-type proportion differences.

**Keywords:** Neurodegenerative disorders, DNA methylation, Mixed-linear models, Methylation profile score, Out-of-sample classification, Inflammatory markers

## Background

Neurodegenerative disorders are a heterogeneous group of disorders that cause progressive disruption of structure and function of the central or peripheral nervous system. Considerable genetic heterogeneity is often observed across patients [1, 2], who can display diverse clinical syndromes that mostly relate to specific brain regions affected by pathology [3]. Nonetheless, studies have also reported overlap between genetic risk factors, mechanisms, and pathological features of these disorders [3, 4]. Common neuronal pathways altered in multiple neurodegenerative disorders include protein quality control, the autophagy–lysosome pathway, mitochondria homeostasis, protein seeding and propagation of stress granules, synaptic toxicity, network dysfunction, and altered immune responses [3]. Importantly, the combination of unique and overlapping clinical and pathological features can lead to difficulty in the diagnosis of individual cases and perhaps calls for an overview of the biological processes that are shared or unique, to allow better classification of disease.

Genetic studies have been widely employed in relation to neurodegenerative disorders, with trait architecture mostly unique to each disease [5–7]. Although the heritability of neurodegenerative disorders ranges from 40 to 80% [8–11], a substantial fraction of the variance in liability is non-genetic, with robust evidence for environmental exposures as important contributors to disease pathogenesis [12]. DNA methylation which in mammals primarily refers to the reversible addition of a methyl group to a

cytosine residue at a CpG dinucleotide is the most widely studied chemical modification of DNA. DNA methylation can repress transcription directly, by inhibiting the binding of specific transcription factors, or indirectly, by recruiting methyl-CpG-binding proteins and their repressive chromatin remodeling activities. There has been increasing evidence that alterations in DNA methylation play an important role in neurodegenerative and other brain disorders, with reports of significant associations with Alzheimer's disease (AD) [13–16], Parkinson's disease (PD) [17, 18], amyotrophic lateral sclerosis (ALS) [19, 20], and schizophrenia [21, 22], both in the brain and blood. Additionally, DNA methylation data can capture signatures of unmeasured environmental exposures. In this context, DNA methylation changes often show large effects; for example, composite DNA methylation scores explained 61% and 12.5% of the phenotypic variance of smoking status (current/ever/never) and alcohol intake (units per week), respectively [23]. Imputation of unmeasured environmental exposures could therefore help stratify patients across diagnostic boundaries, which may provide stimuli for additional analyses and clinical follow-up.

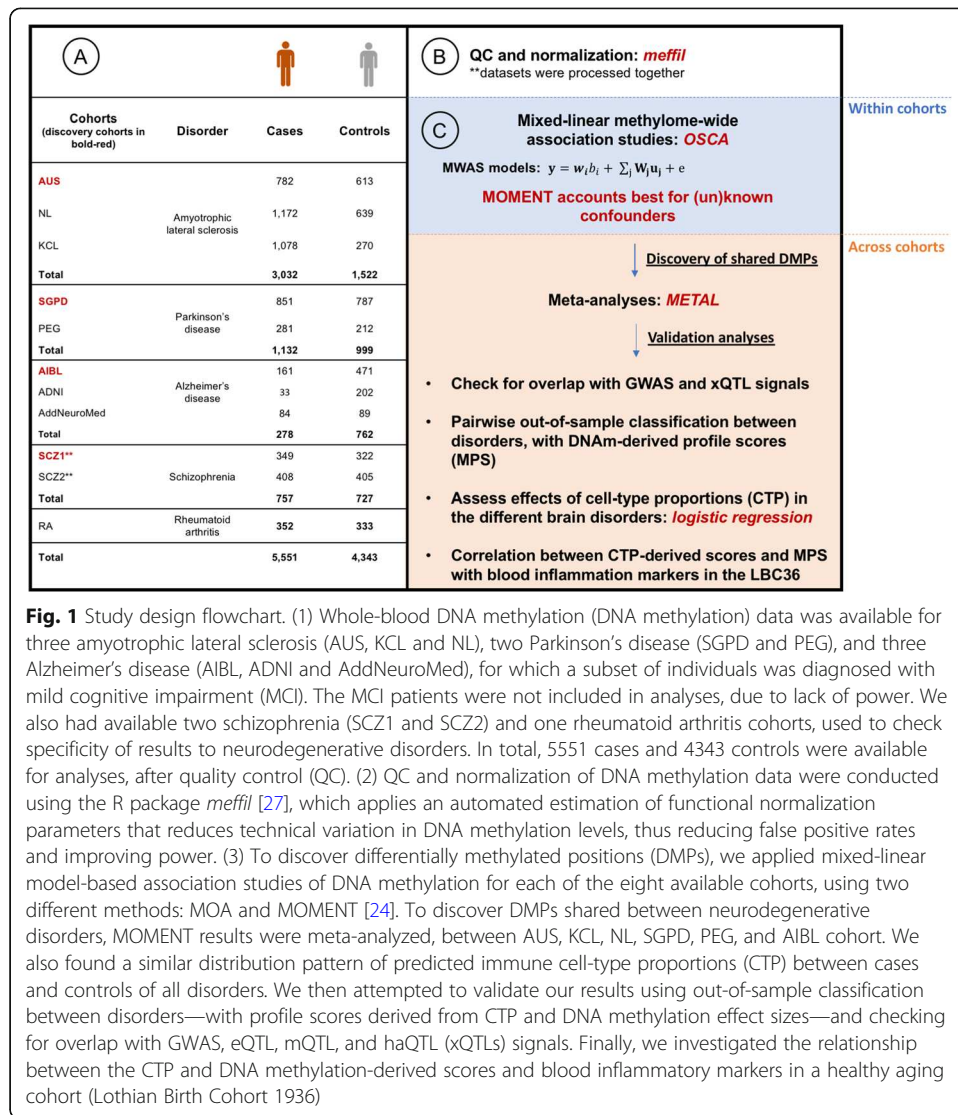
Using a similar concept to genome-wide association studies, methylome-wide association studies (MWAS) methods have started to emerge in order to address the effect of differentially methylated positions (DMPs) on complex diseases. Recently, the OmicS-data-based Complex trait Analysis (OSCA) software has been developed [24]. OSCA implements two reference-free mixed-linear model approaches that model different genome-wide architecture of DNA methylation: MOA and MOMENT (see the “Methods” section). Both methods have been shown, through extensive simulations [24], to better account for known (cell type proportion, smoking, age, batch effects) and unknown confounders than other methods. We have previously applied both MOA and MOMENT to two independently collected ALS case-control cohorts [20]. Both methods showed higher out-of-sample classification accuracy compared to linear regression, with MOMENT showing the best performance, despite detection of fewer significantly associated probes.

In this study, we investigated blood DNA methylation differences between patients and healthy controls across neurodegenerative disorders, including AD, ALS, and PD. We used both MOA and MOMENT to test for association between each DNA methylation site and the traits. We included schizophrenia, because of the previously reported genetic correlation between schizophrenia and ALS [25], and rheumatoid arthritis, a long-term autoimmune disorder, with a clearly defined pathogenic role of peripheral immune cells [26]. Analyses of schizophrenia and rheumatoid arthritis demonstrate if differences we find are specific to neurodegenerative diseases.

## Results

### Study design

Figure 1 shows an overview of the study design and analyses we used to investigate the shared DNA methylation alterations across brain disorders. After data preprocessing, quality control (QC), and normalization of DNA methylation data conducted in each cohort (see the “Methods” section and Additional file 1: Supplementary Note), 5551 genetically confirmed unrelated (except in the The Alzheimer's Disease Neuroimaging



**Fig. 1** Study design flowchart. (1) Whole-blood DNA methylation (DNA methylation) data was available for three amyotrophic lateral sclerosis (AUS, KCL and NL), two Parkinson’s disease (SGPD and PEG), and three Alzheimer’s disease (AIBL, ADNI and AddNeuroMed), for which a subset of individuals was diagnosed with mild cognitive impairment (MCI). The MCI patients were not included in analyses, due to lack of power. We also had available two schizophrenia (SCZ1 and SCZ2) and one rheumatoid arthritis cohorts, used to check specificity of results to neurodegenerative disorders. In total, 5551 cases and 4343 controls were available for analyses, after quality control (QC). (2) QC and normalization of DNA methylation data were conducted using the R package *meffil* [27], which applies an automated estimation of functional normalization parameters that reduces technical variation in DNA methylation levels, thus reducing false positive rates and improving power. (3) To discover differentially methylated positions (DMPs), we applied mixed-linear model-based association studies of DNA methylation for each of the eight available cohorts, using two different methods: MOA and MOMENT [24]. To discover DMPs shared between neurodegenerative disorders, MOMENT results were meta-analyzed, between AUS, KCL, NL, SGPD, PEG, and AIBL cohort. We also found a similar distribution pattern of predicted immune cell-type proportions (CTP) between cases and controls of all disorders. We then attempted to validate our results using out-of-sample classification between disorders—with profile scores derived from CTP and DNA methylation effect sizes—and checking for overlap with GWAS, eQTL, mQTL, and haQTL (xQTLs) signals. Finally, we investigated the relationship between the CTP and DNA methylation-derived scores and blood inflammatory markers in a healthy aging cohort (Lothian Birth Cohort 1936)

Initiative (ADNI) and AddNeuroMed cohorts) cases and 4343 controls were available for analyses, across 11 different cohorts (Fig. 1).

DNA methylation was measured in whole blood, with 450k Illumina arrays, except for the Australian Imaging, Biomarkers and Lifestyle (AIBL) and ADNI cohorts, which were measured with EPIC Illumina arrays. Prior to analysis, we removed probes that failed QC, probes linked to sex-chromosomes, probes overlapping with SNPs, and probes with non-unique hybridization and extension, following recommendations described elsewhere [28]. We also excluded remaining probes on a per-cohort basis for which the standard deviation (s.d.) of measurements was < 0.02. This decision is justified, because power to detect an association depends, in part, on the variance between individuals and (standardized) effect sizes. Excluding these DNA methylation sites also reduces the multiple testing burden in association studies. The number of DNA methylation sites used for analyses in each cohort ranged from 206K to 254K, except for the AIBL cohort (EPIC array) for which 373 K sites passed QC (Additional file 2: Table S1). A description of the

sample characteristics can be found in Additional file 1: Supplementary Note, Additional file 1: Figure S1 and Additional file 2: Table S2.

### Meta-analysis of MOMENT mixed-linear model association studies identifies differentially methylated positions across neurodegenerative disorders

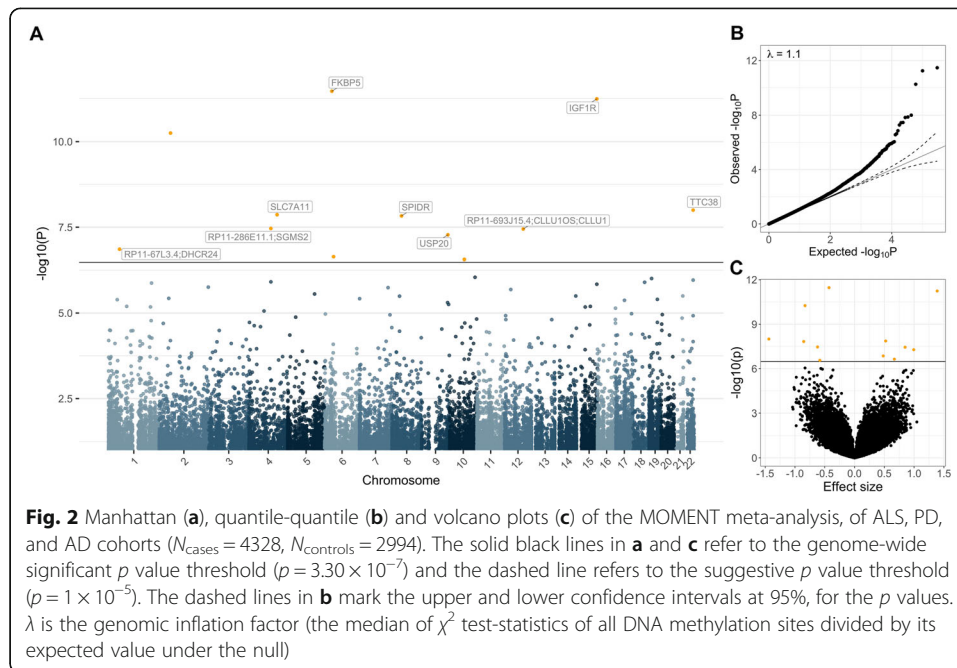
To discover differentially methylated positions (DMPs) between cases and controls, we conducted MOMENT and MOA MWAS for each available cohort (Additional file 1: Figures S2, and S3, respectively). We have previously applied both MOA and MOMENT to two independently collected ALS case-control cohorts [20]. Both methods showed higher out-of-sample classification accuracy compared to linear regression, with MOMENT showing the best performance, despite detection of fewer significantly associated probes. Thus, our focus is the MOMENT results, but the MOA analyses aid with interpretation given potential confounding factors. We did not find evidence of genomic inflation with either method (i.e., the median of  $\chi^2$  test-statistics of all DNA methylation sites divided by its expected value under the null:  $\lambda = [0.98-1.04]$ ) (Additional file 1: Figures S2, and S3).

The only probe found to be significantly associated with schizophrenia in the MOMENT MWAS was cg05575921 (annotated to *AHRR*;  $p = 2.79 \times 10^{-27}$ ) (Additional file 1: Figure S2), a well-replicated DMP that has been previously associated with cigarette smoking [23, 29–31]. Due to extensive epidemiological evidence that showed elevated smoking rates and intensity in patients with schizophrenia [32, 33], we fitted predicted smoking scores in the schizophrenia MWAS, to adjust for its confounding effect in all downstream analyses.

Next, we applied fixed-effects inverse-variance-weighted [34] meta-analyses to the MOMENT results of the disease-specific cohort MWAS, i.e., within ALS ( $N_{\text{cases}} = 3032$  and  $N_{\text{controls}} = 1522$ ) (Additional file 1: Figure S4) and within Parkinson's disease (PD) ( $N_{\text{cases}} = 1132$  and  $N_{\text{controls}} = 999$ ) (Additional file 1: Figure S5). We then meta-analyzed results across Alzheimer's disease (AD), ALS and PD ( $N_{\text{cases}} = 4325$ ,  $N_{\text{controls}} = 2992$ ). The results for  $m = 151,506$  probes (the low number of probes reflects different probes with s.d.  $< 0.02$ , between cohorts) included in the meta-analysis of the three neurodegenerative disorders can be found in Fig. 2a and show no evidence of genomic inflation (Fig. 2b,  $\lambda = 1.1$ ). Compared to the very few genome-wide significant hits in the individual MWAS (Additional file 1: Figure S2), 12 CpGs pass the Bonferroni corrected genome-wide significance threshold ( $p = 3.30 \times 10^{-7}$ ) (Table 1). In contrast, the MOA meta-analysis shows 41 genome-wide significant results (Additional file 1: Figure S6A), with higher genomic inflation ( $\lambda = 1.15$ ) (Additional file 1: Figure S6B). Importantly, only five of the 12 DMPs were genome-wide significant in the ALS and PD MOMENT meta-analyses whereas the AD MOMENT MWAS (Additional file 1: Figure S2) showed no genome-wide significant associations (Additional file 2: Table S3). As expected, the significantly associated probes show the same direction of effect across all cohorts, but three probes show significant heterogeneity in effect size between cohorts [35] ( $I^2 = 60.2\%$ , probe cg06690548 in *SLC7A11*,  $I^2 = 73.2\%$ , for probe cg17901584 in RP11-67 L3.4;DHCR24 and  $I^2 = 81.1\%$  probe cg26033520) (Table 1).

Adding the schizophrenia cohorts to the analysis results in twelve additional genome-wide significant CpGs (Additional file 1: Figure S7A), with effect sizes highly correlated





between analyses ( $r_{\text{Pearson}} = 0.89$ ,  $p < 2.2 \times 10^{-16}$ , Additional file 1: Figure S7B). Lastly, we added the rheumatoid arthritis cohort to the meta-analysis with AD, ALS, and PD, as a positive control, i.e., to demonstrate the specificity of results to neurodegenerative or brain disorders. Interestingly, we found two additional genome-wide significant probes: cg01447828, annotated to gene *PRX* ( $p_{\text{META\_NDs\_RA}} = 1.48 \times 10^{-7}$ ) and cg03785076, annotated to gene *SNED1* ( $p_{\text{META\_NDs\_RA}} = 3.1 \times 10^{-7}$ ) (Additional file 1: Figure S7C), with effect sizes highly correlated between analyses ( $r_{\text{Pearson}} = 0.95$ ,  $p < 2.2 \times 10^{-16}$ , Additional file 1: Figure S7D).

### GWAS signals do not overlap with loci centered at the 12 differentially methylated positions

We next investigated if the 12 DMPs overlapped with brain [36] or blood [37] methylation quantitative trait loci (mQTL) regions ( $p < 5 \times 10^{-8}$ ) and SNPs ( $p < 5 \times 10^{-8}$ ) from a sample-size weighted meta-analysis of publicly available GWAS summary statistics for AD ( $N = 368,440$ ) [38], ALS ( $N = 80,610$ ) [39] and PD (minimum  $N = 520$ , maximum  $N = 482,730$ , excluding the 23andMe cohort) [40] (Additional file 1: Figure S8). Three of the 12 DMPs overlapped with blood mQTLs (number of mQTLs;  $m = 266$ ), and another four overlapped with brain mQTLs ( $m = 86$ ) with the top mQTLs presented in Additional file 2: Table S4. We found no evidence for mQTL overlap with GWAS hits (Additional file 2: Table S4), which could reflect lack of power in both MWAS and GWAS, as has previously been observed for body-mass index (BMI) [41].

Since lack of power could hide potential causal genetic relationships to disease with marginal signals in GWAS that may be present in the DNA methylation data as well, we further investigated whether the loci located  $\pm 500$  kb of each of the 12 DMPs overlapped with the previously mentioned GWAS signals from each disorder and our meta-analysis (Additional file 1: Figures S9-S12). For AD, we additionally compared with

**Table 1** DNA methylation sites significantly associated with the traits at  $p < 3.3 \times 10^{-7}$ , in a MOMENT meta-analyses of AD, ALS and PD. Chr—chromosome number; Probe—probe identification number as provided by Illumina; bp—base pair position in the genome; Gene—closest genes the probe is annotated to, based on distance to transcription starting site, following the method described elsewhere [28]; Orient—DNA strand orientation, F = forward, R = Reverse [28];  $b_{\text{META}}$ —effects sizes (increase (positive sign) or decrease (negative sign) of methylation between cases and control per standard deviation unit) of meta-analysis results;  $p_{\text{META}}$ — $p$  values of meta-analysis models; s.e.<sub>META</sub>—standard errors from meta-analysis; pMETA— $p$  values from meta-analysis; Direction—direction of effect sizes, within each cohort (AUS, KCL, NL, SGPD, PEG, AIBL, respectively);  $I^2$ —proportion of total variation in study estimates that is due to heterogeneity between the six cohorts in the meta-analysis [35]

Chr	Probe	bp	Gene	Orient	$b_{\text{META}}$	s.e. <sub>META</sub>	$p_{\text{META}}$	Direction	$I^2$	Q	$p_Q$
6	cg03546163	35686586	FKBP5	R	-0.43	0.06	$3.42 \times 10^{-12}$	-	46.2	9.29	0.10
15	cg26272088	98900110	IGF1R	F	1.39	0.20	$5.74 \times 10^{-12}$	++++++	0	2.11	0.83
2	cg24166814	55840142	-	F	-0.83	0.13	$5.60 \times 10^{-11}$	-	36.8	7.91	0.16
22	cg04431254	46288875	TTC38	F	-1.44	0.25	$1 \times 10^{-8}$	-	10.4	5.58	0.35
4	cg06690548	138241654	SLC7A11	R	0.52	0.09	$1.36 \times 10^{-8}$	++++++	60.2	12.58	0.03
8	cg14195992	47353350	SPIDR	R	-0.85	0.15	$1.48 \times 10^{-8}$	-	40.4	8.38	0.14
4	cg17786255	107893233	RP11-286E11.1; SGMS2	R	-0.62	0.11	$3.45 \times 10^{-8}$	-	37.9	8.05	0.15
12	cg11881599	92420308	RP11-693J15.4; CLLU1OS; CLLU1	R	0.85	0.15	$3.57 \times 10^{-8}$	++++++	0	2.25	0.81
9	cg13953978	129838509	USP20	F	0.99	0.18	$5.28 \times 10^{-8}$	++++++	0	0.93	0.97
1	cg17901584	54,888033	RP11-67L3.4; DHCR24	F	0.48	0.09	$1.39 \times 10^{-7}$	+++++-	73.2	18.69	$2.20 \times 10^{-3}$
6	cg18120259	43926902	-	F	0.67	0.13	$2.29 \times 10^{-7}$	++++++	0	1.30	0.94
10	cg26033520	72244313	-	F	-0.58	0.11	$2.74 \times 10^{-7}$	-	81.1	26.40	$7.47 \times 10^{-5}$

GWAS results from clinically diagnosed AD cases, from Kunkle et al. ( $N = 63,926$ ) [42] (Additional file 1: Figure S13). As for the mQTL analysis, we found no evidence for overlap with GWAS signals (i.e., with  $p < 5 \times 10^{-8}$ ) in the pre-defined loci. The strongest overlapping signal was from our meta-analysis of AD, ALS, and PD GWAS, albeit non-significant: SNP rs112184630 ( $p_{\text{META\_GWAS}} = 5.91 \times 10^{-7}$ ), located in chromosome 9q, which has been shown to be a blood eQTL for genes TOR1A ( $p = 6.74 \times 10^{-16}$ ), FBP1 ( $p = 3.04 \times 10^{-15}$ ), C9orf78 ( $p = 9.52 \times 10^{-12}$ ) [43], and also PTGES ( $p = 7.92 \times 10^{-6}$ ) [44].

Finally, we expanded our query within these loci to brain eQTL ( $N = 1433$ ), mQTL ( $N = 411$ ), and haQTL ( $N = 411$ ) summary statistics from the AMP-AD consortium [45, 46]. Six of the twelve loci showed significant signals across all xQTLs, suggesting these results could be relevant to brain tissue in a panQTL manner (Additional file 1: Figure



S14). The top eQTL per gene in the queried 1 Mb window can be found in Additional file 2: Table S5.

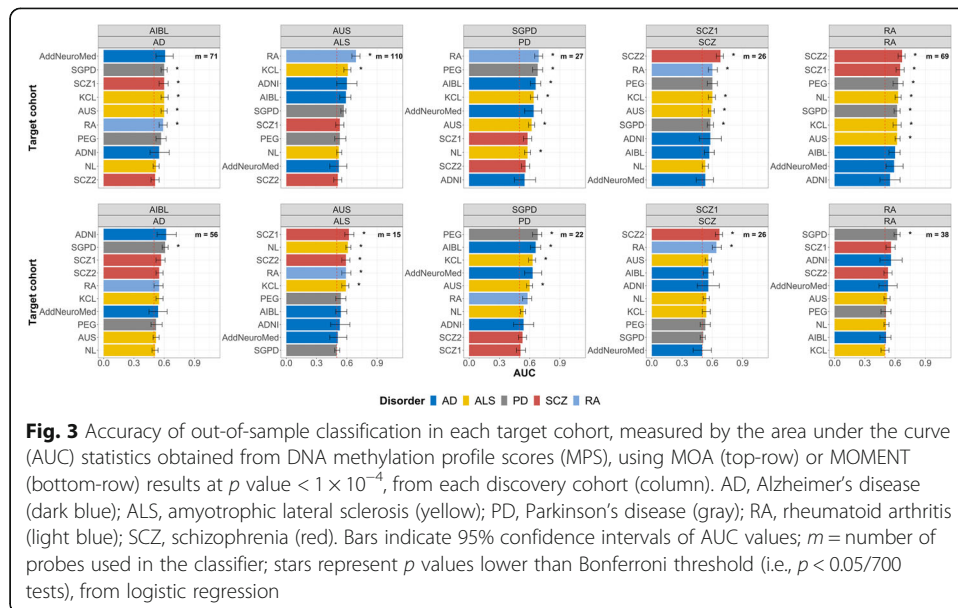
#### **Out-of-sample classification accuracy within- and across-disorders from DNA methylation-derived profile scores**

Out-of-sample classification provides independent evidence that differences in DNA methylation between cases and controls reflect differences associated with disease status rather than technical confounding effects, although they could also reflect shared disease-associated confounders (e.g., smoking status and schizophrenia). It can also leverage DNA methylation differences between cases and controls that do not achieve statistical significance. Thus, we performed pairwise out-of-sample classification using DNA methylation-derived profile scores (MPS), with DNA methylation effect sizes as weights multiplied by each corresponding site in the target cohort. MPS were calculated keeping effect sizes that passed different  $p$  value thresholds in each MOA or MOMENT MWAS. Classification accuracy of the MPS was evaluated by the area under the receiver-operator characteristic (ROC) curves (AUC) (see the “Methods” section). AUC ranges from 0.5 (random classification) to 1 (perfect classification) and can be interpreted as the probability that a case ranks higher than a control (either in the sample or in the population from which the sample was drawn). We use the notation  $AUC_{X,Y}$  to denote a predictor based on probes and effect sizes estimated in data set of disorder X and used to classify cases and controls of disorder Y.

The maximum AUC obtained with MOA- and MOMENT-MPS is summarized in Additional file 1: Figure S15. The maximum AUC is obtained when classifying rheumatoid arthritis cases from controls with any other brain disorder used as the discovery MWAS, using MOA-MPS calculated from DNA methylation sites associated at less stringent  $p$  value thresholds. These results may sound surprising, since one would expect brain disorders to share more similarly disrupted DNA methylation patterns with each other than with rheumatoid arthritis. However, in the majority of cases, and particularly obvious at lower  $p$  value thresholds (Fig. 3), these AUC patterns differed for MOA- and MOMENT-MPS, with MOMENT-MPS giving higher AUC values within-disorders. For instance, when we use the System Genomics of Parkinson’s disease (SGPD) cohort DNA methylation effect sizes with  $p < 1 \times 10^{-4}$  to calculate the MPS, the maximum MOA-MPS AUC is obtained when classifying rheumatoid arthritis patients from controls ( $AUC_{SGPD:RA} = 0.69$ ,  $m = 27$ ,  $p = 9.1 \times 10^{-16}$ ), whereas the maximum MOMENT-MPS AUC is obtained when classifying PD patients from controls ( $AUC_{SGPD:PEG} = 0.68$ ,  $m = 26$ ,  $p = 1 \times 10^{-9}$ ) and  $AUC_{SGPD:RA}$  is much reduced with MOMENT-MPS ( $AUC_{SGPD:RA} = 0.58$ ,  $m = 27$ ,  $p = 6.8 \times 10^{-4}$ ) (Fig. 3). Since rheumatoid arthritis is a long-term autoimmune disorder, with a clearly defined pathogenic role of peripheral immune cells [26], we hypothesized that the effects of cell-type composition, particularly at higher  $p$  value thresholds, were driving the high AUC values.

#### **Analysis of DNA methylation-derived immune cell-type proportions**

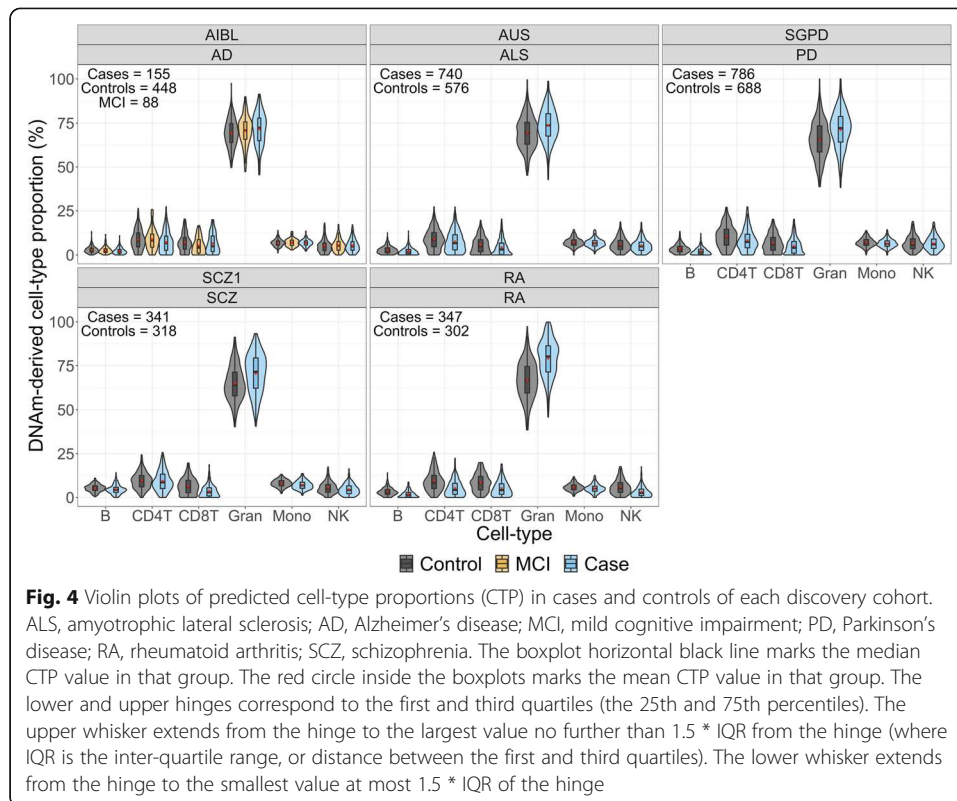
We used the EpiDISH algorithm [47] to predict immune cell-type proportions (CTP) of B lymphocytes, CD4+ T lymphocytes, CD8+ T lymphocytes, granulocytes, monocytes, and natural killer cells, as empirical cell-type measurements were not available to



us. Visual inspection of the CTP distributions between cases and controls showed very similar patterns between all disorder (Fig. 4). A subset of individuals in the AD cohorts was diagnosed with mild-cognitive impairment (MCI). Although we did not include these individuals in any analyses due to lack of power, the upwards trend in granulocytes (and in turn a decrease in other CTP) is worthy of note, with MCI patients showing a midway value between controls and AD patients (Fig. 4). These results were replicated across all AD cohorts (Additional file 1: Figure S16), suggesting a potential link with disease progression (that can be cause or consequence of disease).

To assess the association between each CTP and disorder we used multiple logistic regression models, with case-control status of the discovery cohorts fitted as response variable and CTP (excluding CD8+ T cells, due to redundancy in proportion data and their lower abundancy), sex, DNA methylation age [48] and DNA methylation-derived smoking score [23] fitted as covariates. We excluded individuals with outlying CTP values ( $N = 464$ ) prior to fitting the models (see the “Methods” section). We found no significant associations of estimated proportion of cell-types for the AIBL and the Australian ALS cohort (AUS), after Bonferroni correction (i.e.,  $p < 0.01$ ) (Additional file 2: Table S6). The odds ratio (OR) of being classified as a case given an increase in granulocytes proportions (per one-point percentage increase) ranged from 1.02–1.12, after controlling for the other covariates. The highest OR were observed for the rheumatoid arthritis (OR = 1.12,  $CI_{95\%} = [1.07–1.18]$ ,  $p = 4.4 \times 10^{-7}$ ) and the SCZ1 cohort (OR = 1.12,  $CI_{95\%} = [1.04–1.19]$ ,  $p = 1.67 \times 10^{-3}$ ) and the SGPD cohort (OR = 1.04,  $CI_{95\%} = [1.01–1.07]$ ,  $p = 2.82 \times 10^{-3}$ ). Notably, we found the direction of effect sizes to be highly consistent between disorders for some of the CTPs (e.g., granulocytes), but in different directions for other cell-types (e.g., B lymphocytes) (Additional file 2: Table S6).

We also assessed the pairwise out-of-sample classification accuracy of CTP-derived scores (sum scores of effect sizes estimated from the previously described logistic regression models, multiplied by the corresponding CTP in the target cohort) (Additional file 1: Figure S17). Similar to the MPS, the maximum AUC using CTP-scores is



obtained when classifying rheumatoid arthritis cases from controls, strengthening our hypothesis that these results are mainly driven by CTP. Indeed, if we adjust the MPS that give maximum AUC by the CTP in a linear model and use the residuals to determine the AUC of the resulting ROC curves, the overall maximum accuracy of both MOA and MOMENT MPS goes down (Additional file 1: Figure S18).

#### Significant correlation of CTP profile scores with blood protein markers of inflammation in a healthy aging cohort

To determine if CTP-scores and MPS were capturing immune-related signals, we calculated correlations between these scores and blood inflammatory markers in the Lothian Birth Cohort 1936 (LBC1936). The LBC1936 is a healthy aging elderly cohort for whom we had access to whole blood DNA methylation data measured from 450k Illumina arrays ( $N = 980$ ), 92 blood inflammatory protein markers measured with the Olink® inflammation panel (pg/mL) ( $N = 1048$ ), and real cell-counts ( $10^9/L$ ), including lymphocytes, monocytes, and granulocytes measured at wave 1 ( $N = 909$ ). After quality control (see the “Methods” section), 823 individuals had blood DNA methylation, measured cell counts and inflammatory markers measures available, and were included in the correlation analyses.

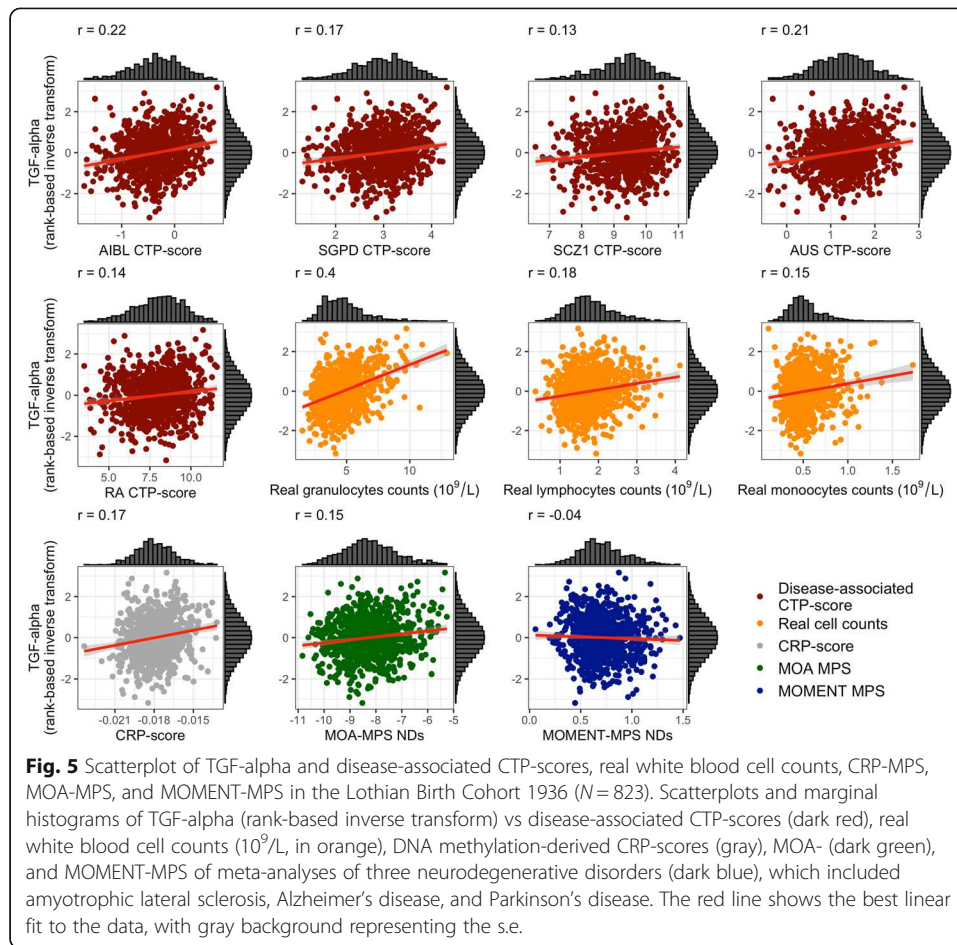
As expected, many blood inflammation markers are highly correlated with measured white cell counts (Additional file 1: Figure S19), although the observed absolute magnitude of correlation coefficients was low to moderate ( $|r_{\text{Pearson}}| = 0$  to 0.4). Some markers, such as transforming-growth factor alpha (TGF-alpha), showed moderate

positive correlations with all disease-associated CTP-scores: AIBL CTP-scores ( $r_{\text{Pearson}} = 0.22$ ,  $p = 2.78 \times 10^{-10}$ ), AUS CTP-scores ( $r_{\text{Pearson}} = 0.21$ ,  $p = 1.81 \times 10^{-9}$ ), SGPD CTP-scores ( $r_{\text{Pearson}} = 0.17$ ,  $p = 9.88 \times 10^{-7}$ ), rheumatoid arthritis CTP-scores ( $r_{\text{Pearson}} = 0.14$ ,  $p = 9.41 \times 10^{-5}$ ), and schizophrenia CTP-scores ( $r_{\text{Pearson}} = 0.13$ ,  $p = 2.21 \times 10^{-4}$ ) (Fig. 5). To the contrary, markers such as TNF-beta were negatively correlated with CTP-scores from neurodegenerative disorders specifically: SGPD CTP-scores ( $r_{\text{Pearson}} = -0.23$ ,  $p = 2.38 \times 10^{-11}$ ), AIBL CTP-scores ( $r_{\text{Pearson}} = -0.18$ ,  $p = 2.08 \times 10^{-7}$ ), and AUS CTP-scores ( $r_{\text{Pearson}} = -0.11$ ,  $p = 3.52 \times 10^{-3}$ ) (Additional file 1: Figure S20). All disease-associated CTP-scores were highly positively correlated with granulocyte counts ( $r_{\text{Pearson}} = [0.49-0.62]$ ) and negatively correlated with lymphocyte counts ( $r_{\text{Pearson}} = [-0.32 \text{ to } -0.43]$ ) (Additional file 1: Figure S21), with the difference in sign of correlation likely a consequence that the CTP-scores were estimated from proportion data.

Additionally, we investigated the relationship between the protein inflammatory markers, with three DNA methylation-derived scores: a C-reactive protein score (CRP-MPS), derived from a chronic low-grade inflammation MWAS [49, 50] and MOA- and MOMENT-MPS, calculated using genome-wide significant probes that were present in the LBC1936, from the MOMENT (Fig. 2) or MOA (Additional file 1: Figure S3) meta-analyses of neurodegenerative disorders, as weights ( $m = 11$  and  $m = 38$ , respectively). It is important to note that the LBC1936 was used to estimate the effect sizes used as weights in the CRP-MPS and thus these may be slightly inflated. We note both CRP-MPS and MOA-MPS are highly positively correlated with many markers of inflammation, in contrast to MOMENT-MPS (Additional file 1: Figure S22). This is also reflected in the correlation between scores: MOA-MPS are strongly positively correlated with all CTP-scores, CRP-MPS, and granulocytes counts ( $r_{\text{Pearson}} = [0.50-0.77]$ ), whereas MOMENT-MPS are strongly negatively correlated ( $r_{\text{Pearson}} = [-0.29 \text{ to } -0.59]$ ), suggesting MOMENT could be over-correcting for CTPs (Additional file 1: Figure S21).

## Discussion

In this study, we aimed to identify shared effects of blood DNA methylation in three neurodegenerative disorders: ALS, AD, and PD. Here, we used MOA and MOMENT [24], two mixed-linear model methods, to test for association of each DNA methylation site with the traits on a per-cohort basis (Fig. 1). In simulations, MOMENT has been shown to be more powerful and generate fewer false positive associations than standard methods, given known and unknown confounding effects that can generate high correlation of DNA methylation levels across the genome. Our focus is thus the MOMENT results, but the MOA analyses aid with interpretation given potential confounding factors. While only three probes were found to be genome-wide significant in each MOMENT association analyses of ALS and PD (and none with AD), in the MOMENT meta-analysis of the three disorders, 12 DMPs passed the Bonferroni-corrected genome-wide significance threshold ( $p = 3.30 \times 10^{-7}$ , Table 1). This result is not expected by chance and implies that neurological disorders may have similarly affected biological mechanisms in blood, although whether this is part of the causal pathway to disease or as a result of disease cannot be shown from these data. We did not find any overlap with the top signals from a previous AD blood MWAS [15]. Five and 6 DMPs from our meta-analysis overlapped with the top signals from a blood MWAS of PD



[18], and ALS [20], respectively (Additional file 1: Figure S23), but it is important to note the PD and ALS cohorts from these studies were included in our meta-analysis.

The 12 DMPs were annotated to genes which encoded proteins involved in intracellular trafficking/protein quality control, such as co-chaperone activities involved in glucocorticoid signaling, deubiquitination involved in beta-2 adrenergic receptor recycling and anionic amino-acid transport (*FKBP5*, *USP20*, and *SLC7A11*, respectively), tyrosine kinase activity involved in cell growth and survival control (*IGF-1R*), double-stranded DNA damage repair (*SPIDR*), and lipid metabolism (*SGMS2* and *DHCR24*). Lastly, some of these probes were also annotated to intergenic regions and to a gene with unknown function (*CLLUI*) that has mainly been studied in the context of chronic lymphocytic leukemia [51].

The probe cg03546163, located at the promoter of gene *FKBP5*, showed the strongest association in the MOMENT meta-analysis of neurodegenerative disorders ( $p_{\text{META}} = 3.42 \times 10^{-12}$ , decreased blood DNA methylation in cases compared to controls). Demethylation at this site has previously been reported in patients with Cushing's syndrome (marked by chronic excess and attenuation of the endogenous diurnal variation in cortisol secretion) [52] and Behçet's disease compared to controls, in the same direction of effect as for neurodegenerative disorders. Especially relevant to neurodegenerative disorders is the finding that *FKBP5* expression has been shown to progressively increase with normal aging, concomitant with reduced *FKBP5* DNA methylation [53],



which correlated with Braak staging in human brains and increased tau pathology both in vitro and in mouse models of AD [53, 54]. Additionally, a recent study suggested a model whereby aging and stress-related phenotypes synergize to decrease DNA methylation at selected enhancer-related FKBP51 sites, epigenetically upregulating FKBP5 in whole blood and in distinct immune cell subtypes. Higher FKBP5 in turn would promote NF- $\kappa$ B (nuclear factor kappa-light-chain-enhancer of activated B cells)-driven peripheral inflammation [55]. Nonetheless, it is important to note that in all these studies, the authors did not adjust for cell type composition, a widely recognized confounder in DNA methylation studies [56, 57]. Although we find reassurance in these past findings, we also point out a negative correlation between predicted granulocyte proportions and DNA methylation levels (and positive correlations with other cell subtypes due to the nature of proportion data) at cg03546163, which does not exclude confounder effects induced by shifts in CTP between cases and controls (Additional file 1: Figure S24). Due to space restrictions we provide a more thorough discussion of the remaining associated probes in Additional file 1: Supplementary Note.

The addition of schizophrenia and rheumatoid arthritis cohorts to the meta-analysis of neurodegenerative disorders resulted in additional genome-wide significant associations, a result that is also not expected by chance and implies that neurological disorders may have similarly affected biological mechanisms in blood with psychiatric and auto-immune disorders. For example, the association with probe cg01447828 ( $p_{\text{META\_NDS\_SCZ}} = 1.17 \times 10^{-7}$ ,  $p_{\text{META\_NDS\_RA}} = 1.48 \times 10^{-7}$ , decreased methylation blood DNA methylation in cases compared to controls), annotated to the gene *PRX* encoding periaxin, is significant only when adding schizophrenia or rheumatoid arthritis to the MOMENT meta-analyses. Periaxin is a structural membrane-associated protein required for maintenance of the myelin sheath of peripheral Schwann cells and normal remyelination after nerve injury. We note that mutations in *PRX* have been previously shown to cause peripheral neuropathies [58, 59], which are not uncommon in Parkinson's disease and rheumatoid arthritis patients, but do not speculate on the relationship between these observations.

Some case-control DNA methylation differences, driven by aging, cell type composition, or environmental exposures, medications, or complications of the disease, could be considered confounders to the trait of primary interest, or alternatively, could be of primary interest themselves depending on the context of the scientific question. While the effects of aging in neurodegenerative disorders are well-recognized and widely studied [60], associations with other risk factors such as smoking or heavy metals are still highly debated. In the context of cell-type composition, there is accumulating evidence for an active role of immune cells, and inflammation in general, in neurodegenerative disorders (both as cause and consequence of disease) [61, 62]. We have found CTP to be major drivers of high MPS classification accuracy (as measured by AUC), given by MPS calculated at lower  $p$  value thresholds. Although at more stringent  $p$  value thresholds, AUC values given by MOMENT-MPS were generally higher when classifying within-disorders, in contrast to MOA-MPS, we also note the overlap of confidence intervals between AUC, and thus disease-specificity cannot be completely assumed (Fig. 3).

To help us determine if the associations with the 12 DMPs in the MOMENT meta-analysis were due to confounding with CTP, we assessed the relationship between MPS



derived from the 12 DMPs and well-known blood inflammatory markers, in a healthy aging cohort. We then compared the results obtained when utilizing disease-associated scores derived from CTP, low-grade chronic inflammation-related CRP-scores, or with measured cell counts. As expected, we found measured cell counts, CTP-scores, and low-grade chronic inflammation-related CRP-scores to be highly positively correlated with blood protein inflammatory markers, whereas MOMENT-MPS (derived from genome-wide significant results of a meta-analysis of ALS, AD and PD) showed no statistically significant correlations (Additional file 1: Figure S19), suggesting these MPS are not strongly confounded by CTP.

Our study has several limitations. Firstly, although growing in number, the current scarcity of homogeneously collected DNA methylation datasets, largely from small cohorts, makes it difficult to perform replication analyses, particularly in a mixed-linear model setting due to lack of power, as we previously performed for ALS and PD [18, 20]. Secondly, a perceived limitation of our study is that blood is not a relevant tissue for understanding the biological mechanisms underlying brain disorders, due to the tissue-specificity of most DNA methylation patterns. However, in the context of diagnosis/prognosis or biomarker discovery, more accessible peripheral tissues (e.g., whole blood), will always prove more useful, especially if the concordance of genome-wide DNA methylation of the tissue analyzed with DNA methylation in live brain tissue is taken into consideration [36, 63, 64]. This hypothesis is further supported by findings that variation between people in DNA methylation controlled by SNP variation has been shown to have high correlation between brain and blood [36, 64]. Lastly, as real cell-type composition measures were unavailable we studied predicted cell-type proportions from DNA methylation. These CTP are inherently estimated with error (Additional file 1: Figure S25), which also depends on DNA methylation measurement noise that may be cohort-dependent. Furthermore, CTP-scores are highly correlated with granulocyte proportions (Additional file 1: Figure S21), because these are the most abundant cell types in whole blood potentially masking biologically relevant actions from other cell-types.

Although it is economically unfeasible to collect DNA methylation data at single-cell resolution for these sample sizes, these results advocate that it would be good practice to collect single-cell data in at least a subset of individuals to validate results from whole-blood analysis. A similar conclusion is supported by a recent study assessing the co-variability of DNA methylation across peripheral cells and tissue [65]. We annotated our 12 DMPs from the MOMENT meta-analysis to characteristic scores supplied by the authors, to determine which cell types are potentially affected by the significant differences reported (Additional file 2: Table S7). The scores for each DNA methylation site and cell type were calculated by fitting a one-sided Levene's test comparing the variation of a single cell type against the variation across all samples from the other four cell types, specifically testing for a larger variance in that cell type (i.e., one-tailed test). DNA methylation sites were determined to be characteristic of single cell type if the  $P$  value from Levene's test was  $< 9 \times 10^{-8}$ . We have found only cg04431254 and cg11881599 reflected variation in a single cell type, CD8T and B cells, respectively. None of the remaining probes reflected significant variation in any of the cell types, although it is possible that the cell-type driving variation at this site have not been interrogated in this study.

Despite some limitations, taken together these results show the presence of aberrant peripheral DNA methylation differences and CTP distribution patterns that point to shared pathogenic mechanisms between neurodegenerative disorders, which are likely a reflection of neuroinflammatory dysregulation. Larger samples with blood collected prior to diagnosis and with deep clinical phenotyping are needed to allow investigation of the potential predictive ability of DNA methylation/CTP-based biomarkers in neurodegenerative disorders and to distinguish which of these mechanisms are cause or consequence of disease.

## Conclusions

In this study, we identified shared DMPs, in whole blood, and similar CTP distribution patterns between neurodegenerative disorders that point to shared pathogenic mechanisms, which are likely a reflection of neuroinflammatory dysregulation. Larger samples with blood collected prior to diagnosis and with deep clinical phenotyping are needed to allow investigation of the potential predictive ability of DNA methylation/CTP-based biomarkers in neurodegenerative disorders and to distinguish which of these mechanisms are cause or consequence of disease.

## Methods

### Cohorts description

#### *Amyotrophic lateral sclerosis (ALS) cohorts*

**Australian ALS cohort** The Australian ALS cohort (AUS) and DNA methylation assays have been previously described elsewhere [20]. Part of the Australian sample comprised patients and controls that were ascertained from the University of Sydney as part of the Australian MND DNA bank, which recruited participants from April 2000 to June 2011. Cases were white Australians older than 25 years recruited from around Australia via state-based MND associations with diagnosis verified by a neurologist. Control individuals were either partners or friends of patients with ALS or community volunteers. The remainder of Australian cases were recruited from clinics across Australia between 2015 and 2017 diagnosed with definite or probable ALS according to the revised El Escorial criteria [66]. Control subjects were healthy individuals free of neuromuscular diseases, recruited as either partners or friends of patients with ALS or community volunteers or from the Older Australian Twin Study (OATS) [67]. ALS cases with a recorded family history of ALS were excluded. The DNA methylation was measured using Illumina Infinium HumanMethylation450 BeadChip.

**Netherlands ALS cohort—Project MinE** The Netherlands ALS cohort was collected under Project MinE [68]. The participants of this study consisted of 1866 Netherlands individuals (1222 ALS cases and 644 control individuals) [69]. All ALS cases were diagnosed with definite or probable ALS according to the revised El Escorial criteria [66], and those with a recorded family history for ALS were excluded. All participants gave written informed consent, and the institutional review board of the University Medical Center Utrecht approved this study. DNA methylation data for the NL sample were

measured using Illumina Infinium HumanMethylation450 BeadChip and generated under similar protocols to the AUS cohort.

**King's College London ALS cohorts** The King's College London (KCL) ALS cohort was collected under Project MinE [68]. The participants of this study consisted of a subset of 1433 individuals of UK nationality from the UK National DNA Bank for MND Research who were put forward for DNA methylation profiling. Cases were diagnosed with ALS in one of 20 UK hospitals by neurologists specialized in motor neuron diseases; patients had no family history for ALS and were of self-reported European descent. All cases and controls gave written informed consent. The national Integrated Research Approval System (IRAS) approved the study, reference: 08/H0405/60. DNA was extracted by use of standard methods at three centers within 1 week of the blood being drawn (usually on the same day) and was stored centrally at the UK DNA banking network in Manchester. We used a barcode-based sample tracking system to minimize the risk of clerical error. DNA methylation status of the participants was extracted from whole blood samples using Illumina Infinium HumanMethylation450 BeadChip array following manufacturer's protocol. These samples were run in two separate batches at two different time points (batch1  $N=666$ ; batch2  $N=767$ ). Both batches followed the same quality control pipeline and DNA methylation data were normalized together. Since we did not observe major batch effects, we analyze them as a unique cohort.

#### ***Parkinson's disease (PD) cohorts***

**System Genomics of Parkinson's Disease (SGPD) cohort** The SGPD case-control cohort comprises genotype, phenotype and DNA methylation data for a total of 2333 participants (1292 PD cases, 1041 controls) recruited from three different studies across Australia and New Zealand: (1) the Queensland Parkinson's Project (QPP), (2) the New Zealand Brain Research Institute PD case-control cohort (NZBRI), and (3) the Sydney PD case-control cohort (SYD). The study design, diagnostic criteria, and DNA methylation assays have been described elsewhere [18], but briefly, the QPP cohort includes 1791 participants (867 PD cases, 924 controls), mostly of European ancestry (individuals with genetically confirmed non-European ancestry were excluded prior to analysis). The NZBRI cohort comprises 210 participants (151 PD cases, 59 matched-controls) recruited by the NZBRI. Exclusion criteria for PD patients were prior history of learning disability, severe head injury, stroke, or other neurological impairment and major psychiatric complications at the point of study entry. Whole-blood samples were collected at the same time as phenotypic measurements, which included demographic, medical, and environmental exposure information for all participants. In the SYD cohort, 332 participants (274 PD cases, 57 matched-controls) were recruited from the Parkinson's Disease Research Clinic, Brain and Mind Research Centre at the University of Sydney. The DNA methylation data were measured using Illumina Infinium HumanMethylation450 BeadChip.

**Parkinson's disease, environment, and genes (PEG) cohort** The PEG study is a large population-based study of PD of mostly rural and township residents of California's central valley [70]. The PEG study comprises of 508 European (289 PD cases, 219 controls) and 63 Hispanic individuals (45 PD cases, 18 controls) for a total of 334 PD cases and 237 controls. Study design, diagnostic criteria, and DNA methylation assays have been described elsewhere [70, 71]. The DNA methylation data measured using Illumina Infinium HumanMethylation450 BeadChip.

#### ***Alzheimer's disease (AD) cohorts***

**The Australian Imaging, Biomarkers and Lifestyle (AIBL) cohort** The AIBL cohort is a prospective longitudinal study of aging aimed to recruit 1000 individuals aged over 60 to assist with prospective research into AD. Data was collected by the AIBL study group. AIBL study methodology has been reported previously [72]. Participants with AD ( $N = 211$ ) had neuropsychological profiles which were consistent with AD and were more impaired than participants with mild cognitive impairment (MCI) ( $N = 133$ ) or healthy controls ( $N = 768$ ), who performed within expected norms for age on neuropsychological testing. A subset of the AIBL cohort ( $N = 162$  AD cases,  $N = 94$  MCI cases, and  $N = 485$  controls) were subjected to a similar DNA methylation assay protocol as the AUS ALS cohort [20], but bisulfite DNA samples were hybridized to the 8 sample, HumanMethylationEPIC BeadChip Array.

**The Alzheimer's Disease Neuroimaging Initiative (ADNI)** ADNI is a consortium of universities and medical centers in the United States and Canada. They launched in 2003 as a public-private partnership, led by principal investigator Michael W. Weiner, MD. ADNI was established to develop standardized imaging techniques and biomarker procedures in normal subjects, subjects with MCI, and subjects with mild AD [73]. The main goal of ADNI is to characterize cross-sectionally and longitudinally clinical measures in normal controls, subjects with MCI, and subjects with mild Alzheimer disease (AD) to enable the assessment of the utility of neuroimaging and chemical biomarker measures. The study design, enrolment process, neuropsychological assessments, and diagnostic criteria have been previously described elsewhere [73]. Briefly, a total of 819 subjects (229 cognitively normal, 398 with MCI, and 192 with AD) were enrolled at baseline and followed for 12 months using standard cognitive and functional measures typical of clinical trials [73]. Whole-genome DNA methylation profiling was done from blood samples of ADNI participants. DNA was isolated and plated out at NCRAD and DNA methylation profiling was performed at AbbVie Inc. for a total of 1920 samples, including 1719 unique samples and 201 technical replicates (653 unique individuals). Longitudinal DNA samples at baseline, + 1 and + 2 years were obtained from all subjects. The Illumina Infinium HumanMethylationEPIC BeadChip Array was used for methylation profiling. Samples were randomized using a modified incomplete balanced block design, whereby all samples from a subject were placed on the same chip, with remaining chip space occupied by age- and sex-matched samples. Subjects from different diagnosis groups were placed on the same chip to avoid confounding. Unused chip

space was leveraged for technical reproducibility assessment via replicated DNA samples.

#### **AddNeuroMed—the European collaboration for the discovery of novel biomarkers for Alzheimer’s disease**

AddNeuroMed is part of the Innovative Medicines in Europe initiative (InnoMed), and was established with the goal of biomarker discovery for Alzheimer’s disease (AD) [74, 75]. Participants were recruited at six centers throughout Europe (Kuopio, Finland; Łódź, Poland; London, UK; Perugia, Italy; Thessaloniki, Greece; Toulouse, France), following standardized procedures. Ethical approval was obtained at each site, and informed consent was obtained according to the declaration of Helsinki (1991). Recruited participants included individuals with AD diagnosed according to the NINCDS-ADRDA criteria [76], individuals with mild cognitive impairment (MCI) according to Petersen’s criteria [77], and controls who showed no symptoms of dementia and had a mini mental state examination (MMSE) score of 28 or higher. Participants were excluded if they had depression or any other neurological syndrome.

A subset of 301 samples were selected from the cohort for DNA methylation profiling, including 96 controls, 111 individuals with MCI, and 94 AD cases. DNA extraction methods have previously been described elsewhere [15, 78]. DNA methylation was measured using Illumina HumanMethylation450 BeadChip arrays and the Illumina HiScan System. Samples were randomized by sex, diagnostic status, and recruitment center. Clinical data for this cohort can be accessed according to the data terms of use for AddNeuroMed <https://www.synapse.org/#!Synapse:syn2790911/wiki/235389>.

#### **Schizophrenia cohorts**

**University College London schizophrenia cohort (SCZ1)** The University College London schizophrenia case–control sample study design, diagnostic criteria, DNA methylation collection details, and assay characteristics have been thoroughly described elsewhere [21, 79], but briefly comprises of unrelated ancestrally matched cases ( $N = 353$ ) and controls ( $N = 322$ ) from the UK. Case participants were recruited from UK National Health Service (NHS) mental health services with a clinical International Classification of Diseases 10th edition (ICD-10) diagnosis of schizophrenia. The DNA methylation data were measured using Illumina Infinium HumanMethylation450 BeadChip.

**Aberdeen schizophrenia cohort (SCZ2)** The Aberdeen schizophrenia case–control sample study design, diagnostic criteria, DNA methylation collection details and assay characteristics have been thoroughly described elsewhere [21, 80], but briefly contains patients with schizophrenia ( $N = 414$ ) and controls ( $N = 433$ ) who have self-identified as born in the British Isles (95% in Scotland). All cases met the Diagnostic and Statistical Manual for Mental Disorders fourth edition (DSM-IV) and ICD-10 criteria for schizophrenia. The DNA methylation data were measured using Illumina Infinium HumanMethylation450 BeadChip.

### ***Rheumatoid arthritis cohort***

The raw methylation data from Illumina Infinium HumanMethylation450 BeadChip arrays and phenotypic data for rheumatoid arthritis cases and controls were obtained from the publicly available Gene Expression Omnibus submitted dataset GSE42861, which was part of the Epidemiological Investigation of Rheumatoid Arthritis (EIRA) study [81, 82]. Only incident rheumatoid arthritis cases were invited for the study within the years 1996–2009 from middle Sweden. The controls matched by sex, age, smoking status, and residence area were selected from the same population. The cohort and DNA methylation collection details and assay characteristics have been thoroughly described elsewhere [81].

### ***Lothian Birth Cohort 1936 (LBC1936)***

LBC1936 [83, 84] was used for out-of-sample DNA methylation classification and correlation analyses with protein inflammatory markers. LBC1936 is a cohort comprising individuals born in 1936, who were aged approximately 70 years at recruitment. Whole blood DNA methylation was measured using the measured using Illumina Infinium HumanMethylation450 BeadChip arrays in 1004 participants from samples collected at mean age 70 years. The DNA methylation collection details and assay characteristics have been thoroughly described elsewhere [85, 86].

### **Quality control (QC) and normalization of DNA methylation data**

Data QC and normalization were conducted using the *meffil* R package [27]. The same pipeline for DNA methylation data processing and QC was applied to all samples. QC threshold parameters (Additional file 1: Supplementary Note) determined samples and DNA methylation sites to exclude prior to normalization. Functional normalization was performed to remove technical variation, as described elsewhere [87]. Briefly, probe intensity quantiles were normalized between samples by fitting linear models with these quantiles to the top principal components of the control probe matrix. After normalization, the most variable probes ( $m = 20,000$ ) were extracted, decomposed into principal components, and each component regressed against slide, chip column, chip row, and sex to test for batch effects. The association detection  $p$  value threshold was set to 0.01. Sex-chromosome linked probes, probes overlapping with SNPs, and probes with non-unique hybridization and extension were also removed prior analysis, following general masking recommendations described elsewhere [28]. Afterwards, we removed remaining probes with  $s.d. < 0.02$ . This decision is justified, because power to detect an association depends in part on the variance between individuals and (standardized) effect sizes. Excluding these DNA methylation sites also reduces the multiple testing burden in MWAS.

### **Protein measurements with Olink® inflammation panel, measured at wave 1 of the Lothian Birth Cohort 1936 (LBC1936) and subsequent quality control measures**

Plasma was extracted from 1047 blood samples and collected in lithium heparin tubes at mean age  $69.8 \pm 0.8$  years (wave 1). Plasma samples were analyzed using a 92-plex proximity extension assay (Olink® Bioscience, Uppsala Sweden). The proteins assayed comprise the Olink® inflammatory biomarker panel. Briefly, 1  $\mu$ L of sample was



incubated in the presence of proximity antibody pairs linked to DNA reporter molecules. Upon appropriate antigen-antibody recognition, the DNA tails form an amplicon by proximity extension which is quantified by real-time PCR. Data pre-processing was performed by Olink® using NPX Manager software. We retained all protein measures with at least 80% of individuals above the limit of detection (below 3 times standard deviation over background). The remaining 73 proteins were transformed by rank-based inverse normalization prior to analysis, to ensure normally distributed values. One protein from the panel, BDNF, failed quality control and was removed from the study.

### Statistical analyses

Mixed linear model-based omics association (MOA) and multi-component mixed linear model-based omics association excluding the target (MOMENT) MWAS analyses.

One of the most well-recognized challenge in MWAS (and other omics-based analyses) is how to better control the false-positive rate (FPR) in the presence of confounding factors (e.g., cell type proportion, smoking, age, batch effects), since failing to account for their effects may lead to spurious associations [57, 88]. To address this issue, the software OSCA has been recently developed [24]. OSCA implements two reference-free mixed-linear model approaches: MOA and MOMENT. Briefly, the MOA method fits a random genome-wide DNA methylation factor per person with variance-covariance matrix between individuals built from genome-wide DNA methylation sites (equivalent to a model of fitting all DNA methylation sites as random effects); this model is analogous to the MLM association method implemented in EMMAX [89] and GCTA [90] for SNP data. The MOMENT method fits an MLM with two random-effect components for each probe tested with the DNA methylation sites grouped by their associations with the trait (leaving out the DNA methylation sites in a window around the target probe being tested for association) [24], allowing a different genome-wide architecture of DNA methylation compared to MOA; in MOMENT DNA methylation effect sizes genome-wide are drawn from two distributions with the variances of the distributions estimated from the data. Both methods have been shown, through extensive simulations [24], to have lower FPR than other methods. MOMENT has slightly less power compared to MOA when a single distribution of effect sizes is appropriate for the trait under study. We conducted mixed-linear model MWAS using both MOA and MOMENT [24]. The MOA MWAS model is:

$$y = w_i b_i + Wu + e \quad (1)$$

where  $y$  is an  $n \times 1$  vector of phenotype values of  $n$  individuals,  $w_i$  (a  $n \times 1$  vector of DNA methylation measures ( $\beta$  values) of a probe  $i$ , i.e., the target probe) and  $b_i$  (the effect of probe  $i$  on the phenotype; fixed effect),  $W$  is an  $n \times m$  matrix of  $m$  standardized DNA methylation values, where  $m$  is the number of DNA methylation sites,  $u$  is an  $m \times 1$  vector of the joint random probe effects on the phenotype, and  $e$  is an  $n \times 1$  vector of residuals. The variance of  $y$  is  $\text{var}(y) = V = WW' \sigma_u^2 + I \sigma_e^2$ . We can re-write this equation as  $V = A \sigma_o^2 + I \sigma_e^2$  with  $A = WW'/m$  and  $\sigma_o^2 = m \sigma_u^2$ , where  $A$  is then the omics-data-based relationship matrix and  $\sigma_u^2$  is the variance between individuals attributed to genome-wide DNA methylation differences. The null hypothesis ( $H_0: b_i = 0$ ) can

then be tested by a two-sided t test given  $\hat{b}_i$  and its s.e. The variance components can be estimated by REML.

In this model, the probe being tested is fitted twice, once as a fixed and also as a random effect, which results in slightly reduced power compared to a (hypothetical) model in which the focal probe is excluded from  $W$ , but this would be computationally very demanding. It is also assumed that all probe effects follow a single distribution, which may not reflect the true distribution. In the MOMENT model, DNA methylation probe effect sizes are drawn from two effect size distributions for different probes sets, selected according to their association statistics in an initial linear regression model, with each group then fitted as a random-effect:

$$y = w_i b_i + \sum_j W_j u_j + e \quad (2)$$

where  $W_j$  is an  $n \times m_j$  matrix of standardized DNA methylation probe values in the  $j$ th group, and  $m_j$  is the number of DNA methylation sites in the group (excluding the DNA methylation sites in a 100 Kb region of centered at the probe to be tested). When the number of probes in the first group is too large, the analysis becomes slow because the program needs to re-estimate the variance components whenever one or more probes are removed from the first group (to avoid proximity contamination) [24]. This may also cause convergence problem because of too much variation explained by the first random-effect component. However, OSCA implements a version of MOMENT where an additional stepwise selection procedure to reduce the number of probes in the first group. Simulation shows that this method has approximately the same level of false positive rate, but slightly higher power compared to the initial MOMENT implementation (<https://cnsgenomics.com/software/osca/#EWAS>). We thus performed MWAS for the AUS, NL, KCL, SGPD, PEG, AIBL, SCZ1, SCZ2, and rheumatoid arthritis cohorts separately and for consistency always used the `--moment2-beta` function for MOMENT analyses, as implemented in the OSCA software. DNA methylation sites were then mapped to the latest GRCh38/hg38 genome build [28] and annotated to genes, based on GENCODE v22.

#### ***Meta-analyses of MOMENT MWAS within and between neurodegenerative disorders to identify DMPs***

We conducted inverse-variance weighted meta-analyses using the MOMENT results from each individual cohort, using METAL [91]. We only kept probes in common between all datasets in the analyses. We performed meta-analyses within ALS (AUS, NL, and KCL cohorts,  $N_{\text{cases}} = 3035$ ,  $N_{\text{controls}} = 1524$ ), within PD (SGPD and PEG,  $N_{\text{cases}} = 1133$ ,  $N_{\text{controls}} = 998$ ), and between neurodegenerative disorders (ALS, PD, and AD,  $N_{\text{cases}} = 4329$ ,  $N_{\text{controls}} = 2993$ ). We conducted meta-analyses to identify shared DMPs using MOMENT results, since this method as shown to be more robust to (un)observed confounders [24].

#### ***Meta-analyses of publicly available meta-GWAS summary statistics between neurodegenerative disorders***

To assess the potential overlap of the MWAS meta-analyses results with methylation quantitative trait loci (mQTL) and GWAS SNPs, we first performed a meta-analysis of

three publicly available meta-GWAS summary statistics, for AD ( $N = 368,440$ ) [38], ALS ( $N = 80,610$ ) [39], and PD (minimum  $N = 520$ , maximum  $N = 482,730$ , excluding 23andMe SNPs) [40]. We conducted sample-size weighted meta-analyses, using METAL [91], for all SNPs in common between datasets. Sex-linked SNPs, SNPs with minor allele frequency (MAF)  $< 0.01$ , and SNPs with incongruent MAF or base pair position between datasets were excluded.

### ***IGAP summary statistics***

When looking for overlap with GWAS signals, we analyzed summary statistics provided by the International Genomics of Alzheimer's Project (IGAP), in addition to the summary statistics from the GWAS above. IGAP is a large three-stage study based upon GWAS on individuals of European ancestry. In stage 1, IGAP used genotyped and imputed data on 11,480,632 single nucleotide polymorphisms (SNPs) to meta-analyze GWAS datasets consisting of 21,982 Alzheimer's disease cases and 41,944 cognitively normal controls from four consortia: The Alzheimer Disease Genetics Consortium (ADGC); The European Alzheimer's disease Initiative (EADI); The Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE); and The Genetic and Environmental Risk in AD Consortium Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERADES). In stage 2, 11,632 SNPs were genotyped and tested for association in an independent set of 8362 Alzheimer's disease cases and 10,483 controls. Meta-analysis of variants selected for analysis in stage 3A ( $n = 11,666$ ) or stage 3B ( $n = 30,511$ ) samples brought the final sample to 35,274 clinical and autopsy-documented Alzheimer's disease cases and 59,163 controls.

### ***Multiple logistic regression to estimate disorder-specific effect sizes of predicted cell-type proportions (CTP)***

We used the EpiDISH algorithm [47] to predict DNA methylation-derived proportions of B lymphocytes (Bcell), CD4+ lymphocytes (CD4T), CD8+ T lymphocytes (CD8T), eosinophils (Eosino), monocytes (Mono), neutrophils (Neu), and natural killer cells (NK). We then used multiple logistic regression models with case-control status (excluding MCI) as response variable, to estimate effect sizes of CTPs associated with each disorder. Predicted CTP, predicted DNA methylation age [48], DNA methylation-derived smoking scores [23], and study site were included as covariates in the models. We summed the Neu and Eosino proportions, since these are biologically classified as granular leukocytes. We excluded CD8T proportions from analyses due to redundancy in proportion data. Prior fitting the models, we excluded outlying CTP values that were larger than  $mean(CTP) \pm 3 \times SD(CTP)$  ( $N = 524$ , including MCI), as such extreme CTP may be indicative of current sickness in these elderly participants. A summary of the models can be found in (Additional file 2: Table S7). To fit the regression models, we used the *glm* function in the *stats* R package, with a binomial error distribution and logit link function. We calculated Wald 95% confidence intervals for the exponentiated log odds using the *confint.default* function in the *MASS* R package [92].

### ***Out-of-sample classification using cell-type-proportion (CTP) and DNA methylation profile scores (MPS)***

A profile score is calculated for each individual in the target sample as the sum of CTP or DNA methylation values (MPS) weighted by their effect sizes, estimated in a discovery sample. Classification efficacy of the profile scores was evaluated by the area under the receiver-operator characteristic curve (AUC) that relates the false positive rate (specificity) to the true positive rate (sensitivity), from logistic regression, with case-control status as dependent variable and MPS or CTP-scores as independent variable. We used the R package *pROC* [93] to plot the receiver-operator characteristic curves and calculate AUC for each profile score. The  $CI_{95\%}$  for the AUC was calculated using the *ci.auc* function, using the *DeLong* method. We conducted out-of-sample classification using the MOA/MOMENT results of the AUS, SGPD, SCZ1, and AIBL cohorts, as discovery samples. We conducted both within-trait and cross-trait classification. We calculated MPS using DNA methylation probes that passed significance at the following  $p$  value thresholds:  $p < 0.5$ ,  $p < 0.2$ ,  $p < 0.1$ ,  $p < 1 \times 10^{-2}$ ,  $p < 1 \times 10^{-3}$ ,  $p < 1 \times 10^{-4}$ ,  $p < 1 \times 10^{-5}$ . We only kept probes in common between all cohorts in the analyses. The CTP effect sizes, used to calculate the CTP-scores were estimated from multiple logistic regression models, as described above.

### ***Correlation of CTP-scores and MPS in the Lothian Birth Cohort 1936 (LBC1936), a healthy aging cohort***

To assess if disease associated CTP-scores and MPS were capturing inflammation signals, we calculated correlations with blood protein inflammatory markers, as measured by the Olink® panel, in the LBC1936, a healthy aging cohort (see above). We calculated MPS using DNA methylation probes effect sizes from the MOA/MOMENT meta-analyses of neurodegenerative disorders (AD, ALS, and PD). Additionally, we calculated disease-associated CTP-scores, with effect sizes estimated from multiple logistic regression models, described above. As before, we excluded outlying CTP values in the LBC1936 that were larger than  $mean(CTP) \pm 3 \times SD(CTP)$  ( $N = 46$ ). Finally, we calculated an inflammation-related profile score for each individual in the LBC1936, as described by Barker et al. [49]. Briefly, CRP-related probes were selected based on a recent methylome-wide association study by Ligthart et al. [50]. This selection was limited to 7 CpG probes (spanning a total of 9 genes) that showed the strongest evidence for a functional association with CRP levels, including cg06126421 (standardized effect size =  $-0.0052$ ), cg06690548 ( $-0.0048$ ), cg10636246 ( $-0.0069$ ), cg18181703 ( $-0.0053$ ), cg19821297 ( $-0.0051$ ), cg25325512 ( $-0.0031$ ), and cg27023597 ( $-0.005$ ). All analyses were conducted in R version 3.6.0, Rstudio v1.2.1335, and OSCA version 0.45.

## **Supplementary Information**

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-021-02275-5>.

**Additional file 1.** Supplementary notes and supplementary figures.

**Additional file 2.** Supplementary tables.

**Additional file 3.** Review history.

## **Acknowledgements**

We would like to acknowledge the people who have directly contributed their data, for their participation in this project. We would like to thank the clinical nurses and lab personnel that directly handle and process the data. MFN

would like to acknowledge the University of Queensland/University of Exeter (QUEX) PhD studentship. AJŚ and RFH are supported by funding from the Wellcome Trust 4-year PhD in Translational Neuroscience—training the next generation of basic neuroscientists to embrace clinical research [108890/Z/15/Z]. REM is supported by Alzheimer's Research UK major project grant ARUK-PG2017B-10. SJGL was supported by an NHMRC-ARC Dementia Fellowship (1110414) and GH was supported by an NHMRC Fellowship (1079679). We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for these analyses. The investigators within IGAP contributed to the design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report. IGAP was made possible by the generous participation of the control subjects, the patients, and their families. The i-Select chips were funded by the French National Foundation on Alzheimer's disease and related disorders. EADI was supported by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant, Inserm, Institut Pasteur de Lille, Université de Lille 2, and the Lille University Hospital. GERAD/PERADES was supported by the Medical Research Council (grant no. 503480), Alzheimer's Research UK (grant no. 503176), the Wellcome Trust (grant no. 082604/2/07/Z), and German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) grant no. 01GI0102, 01GI0711, 01GI0420. CHARGE was partly supported by the NIH/NIA grant R01 AG033193 and the NIA AG081220 and AGES contract N01-AG-12100, the NHLBI grant R01 HL105756, the Icelandic Heart Association, and the Erasmus Medical Center and Erasmus University. ADGC was supported by the NIH/NIA grants: U01 AG032984, U24 AG021886, U01 AG016976, and the Alzheimer's Association grant ADGC-10-196728.

### Review history

The review history is available as Additional file 3.

### Peer review information

Anahita Bishop was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

### Authors' contributions

NRW, MFN, and AFM designed the study. MFN performed the main analyses. SML, TL, CLV, NJA, RAJZ, PJH, KL, EP, JAYR, SF, FCG, TJA, SRB, JD-A, JF, GH, IBH, MK, SJGL, GWM, JP, TLP, PAS, ARJ, AS, AI, WvR, LvdB, PJS, CES, KEM, and EH performed preliminary analyses, collected data, and/or helped with the quality control of the DNA methylation cohorts. HS, MT, PM, SL, IK, and BV were the clinical leads responsible for data collection of the AddNeuroMed cohort. NRW, AFM, FZ, JY, JG, PMV, EH, and JM gave analyses advice. KLW, STN, FJS, KAM, MN, and PSS facilitated access to blood the AUS ALS blood samples. SML, NJA, IPB, JBK, KAM, GN, GDM, PSS, JHV, SM, DBR, RDH, PAM, MN, RP, KLW, AA-C, JM, NWR, and AFM are the cohort leaders and overviewed blood sample collection and participant recruitment. AJŚ, RFH, SEH, REM, and IID helped with phenotype curation of the LBC cohort blood protein inflammatory data and facilitated access to the DNA methylation data. FZ and JY developed the OSCA software. AKH and LW coordinated and overviewed the lab quality control of blood samples to generate DNA methylation data, for the AUS ALS, SGPD, and AIBL cohorts. MFN, AFM, NRW, JM, and EH wrote the manuscript. All authors read, commented and approved the final manuscript.

### Funding

Australian ALS cohort: This work was directly funded by EU Joint Programme - Neurodegenerative Disease Research (JPND), co-funded through the Australian National Health and Medical Research (NHMRC) Council: 1151854 as well as NHMRC 1083187, 1173790, 1078901, 1113400. Additional funding was provided by the Motor Neurone Disease Research Institute of Australia Ice Bucket Challenge grant for the SALSA-SGC consortium. The Older Australian Twins Study (OATS, used for controls) was facilitated through Twins Research Australia, a national resource in part supported by a Centre for Research Excellence from the Australian National Health and Medical Research Council (NHMRC). Funding for this study was awarded by the (NHMRC)/Australian Research Council Strategic Award (Grant 401162) and the NHMRC Project grant 1405325. We would also like to thank Prof. Nigel Laing for facilitating access to blood samples in West Australia.

KCL ALS cohort: The project is supported through the following funding organizations under the aegis of JPND - [www.jpnd.eu](http://www.jpnd.eu) (United Kingdom, Medical Research Council (MR/L501529/1; MR/R024804/1) and Economic and Social Research Council (ES/L008238/1)) and through the Motor Neurone Disease Association. The cohort collection was partly funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The work leading up to this publication was funded by the European Community's Health Seventh Framework Programme (FP7/2007–2013; grant agreement number 259867) and Horizon 2020 Programme (H2020-PHC-2014-two-stage; grant agreement number 633413). Samples used in this research were in part obtained from the UK National DNA Bank for MND Research, funded by the MND Association and the Wellcome Trust.

NL ALS cohort: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 772376 - ESCORIAL<sup>TM</sup>). The collaboration project is co-funded by the PPP Allowance made available by Health-Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships. This is in part an EU Joint Programme - Neurodegenerative Disease Research (JPND) project. The project is supported through the following funding organizations under the aegis of JPND - [www.jpnd.eu](http://www.jpnd.eu) (United Kingdom, Medical Research Council (MR/L501529/1; MR/R024804/1, the Netherlands, ZONMW, grant no. 733051071) and through the Motor Neurone Disease Association. This study represents independent research part funded by the National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre. Samples used in this research were in part obtained from the UK National DNA Bank for MND Research, funded by the MND Association and the Wellcome Trust. We would like to thank people with MND and their families for their participation in this project.

SGDP cohort: This research was supported by the National Health and Medical Research Council (NHMRC: 1078037, 1078901, 1103418, 1107258, 1127440, 1113400), the Australian Research Council (ARC: DP160102400 and FT180100186),

and the Mater Foundation. Support also came from ForeFront, a large collaborative research group dedicated to the study of neurodegenerative diseases and funded by the NHMRC (Program Grant 1132524, Dementia Research Team Grant 1095127, NeuroSleep Centre of Research Excellence 1060992) and ARC (Centre of Excellence in Cognition and its Disorders Memory Program CE10001021). The Queensland Parkinson's Project (QPP) was supported by a grant from the Australian National Health and Medical Research Council (1084560) to GWM. The New Zealand Brain Research Institute (NZBRI) cohort was funded by a University of Otago Research Grant, together with financial support from the Jim and Mary Carney Charitable Trust (Whangarei, New Zealand). We thank Allison Miller for processing and handling of NZBRI samples.

**AIBL cohort:** Funding for the AIBL study was provided in part by the study partners [Commonwealth Scientific Industrial and research Organization (CSIRO), Edith Cowan University (ECU), Mental Health Research institute (MHRI), National Ageing Research Institute (NARI), Austin Health, CogState Ltd.]. The AIBL study has also received support from the National Health and Medical Research Council (NHMRC) and the Dementia Collaborative Research Centres program (DCRC2), as well as funding from the Science and Industry Endowment Fund (SIEF) and the Cooperative Research Centre (CRC) for Mental Health—funded through the CRC Program (Grant ID:20100104), an Australian Government Initiative. We thank all those who took part as a participant in the AIBL study for their commitment and dedication to helping advance research into the early detection and causation of AD.

**AddNeuroMed cohort:** The AddNeuroMed data are from a public-private partnership supported by EFPIA companies and SMEs as part of InnoMed (Innovative Medicines in Europe), an Integrated Project funded by the European Union of the Sixth Framework program priority FP6 2004-LIFESCIHEALTH-5. Clinical leads responsible for data collection are Iwona Kloszewska (Lodz), Simon Lovestone (London), Patrizia Mecocci (Perugia), Hilka Soininen (Kuopio), Magda Tso-laki (Thessaloniki), and Bruno Vellas (Toulouse) and imaging leads are Andy Simmons (London), Lars-Olad Wahlund (Stockholm) and Christian Spenger (Zurich) and bioinformatics leads are Richard Dobson (London) and Stephen Newhouse (London).

**ADNI cohort:** Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; Cere-Spir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

#### Availability of data and materials

All code used for analyses and generate figures in this manuscript can be found at [https://github.com/m-nabais/202001\\_cross\\_disorder\\_DNA\\_methylation\\_and\\_Zenodo](https://github.com/m-nabais/202001_cross_disorder_DNA_methylation_and_Zenodo) [94].

The DNA methylation data sets are available as follows:

- Australian ALS cohort: The DNA methylation data have been submitted to dbGAP under accession number: phs002068.v1.p1, excluding the OATS samples ( $N = 84$ ) [20]. For OATS samples, data are available to researchers by request to the OATS Governance Committee.
- SGDP Parkinson's cohort: The DNA methylation data are publicly available at GEO under accession number GSE145361 [95].
- PEG Parkinson's cohort: The DNA methylation data are publicly available at GEO under accession number GSE111629 [96].
- KCL ALS cohort: The DNA methylation data have been submitted to EGA under accession number: EGAC00001000703 [97].
- Netherlands ALS cohort: The DNA methylation data have been submitted to EGA under accession number: EGAC00001000703 [97].
- AIBL cohort: The DNA methylation data are publicly available at GEO under accession number GSE153712 [98].
- ADNI cohort: The DNA methylation data are available to researchers by request as outlined in the ADNI access policy [73].
- AddNeuroMed cohort: The DNA methylation data are publicly available at GEO under accession number GSE144858 [99].
- UCL schizophrenia cohort: The DNA methylation data are publicly available at GEO under accession number GSE80417 [100].
- Aberdeen schizophrenia cohort: The DNA methylation data are publicly available at GEO under accession number GSE84727 [101].
- Epidemiological Investigation of Rheumatoid Arthritis (EIRA): The DNA methylation data are publicly available at GEO under accession number GSE42861 [102].
- Lothian Birth Cohort: The DNA methylation data are available to researchers by request as outlined in the Lothian Birth Cohort access policy [84].

#### Ethics approval and consent to participate

This study was approved by the University of Queensland Human Research Ethics Committee (HREC): HREC/14/QRBW/495 (RBWH) and 2015000022 (UQ) for bodpod samples and HREC/06/QRBW/22 (2006/047) for the MUNE samples.



Participants were consented for the use of the DNA samples and records by the HREC at each site. All experimental methods comply with the Helsinki Declaration.

#### Consent for publication

All authors have consented to publication.

#### Competing interests

The authors declare that there are no competing interests.

#### Author details

<sup>1</sup>Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia. <sup>2</sup>University of Exeter Medical School, RILD Building, RD&E Hospital Wonford, Barrack Road, Exeter EX2 5DW, UK. <sup>3</sup>School of Medical and Health Sciences, Edith Cowan University, 270 Joondalup Dr, Joondalup, WA 6027, Australia. <sup>4</sup>Department of Internal Medicine, Erasmus MC, University Medical Center, 3015GD Rotterdam, The Netherlands. <sup>5</sup>Murdoch University, 90 South St, Murdoch, WA 6150, Australia. <sup>6</sup>Australian Centre for Precision Health, University of South Australia Cancer Research Institute, School of Health Sciences, University of South Australia, Adelaide, SA 5001, Australia. <sup>7</sup>Brain and Mind Centre, Sydney Medical School, The University of Sydney, Sydney, Australia. <sup>8</sup>Centre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales, Sydney, NSW 2031, Australia. <sup>9</sup>Neuroscience Research Australia Institute, Randwick, NSW 2031, Australia. <sup>10</sup>Griffith Institute for Drug Discovery (GRIDD), Griffith University, Brisbane, Australia. <sup>11</sup>Neuropsychiatric Institute, The Prince of Wales Hospital, UNSW, Randwick, NSW 2031, Australia. <sup>12</sup>Department of Neurology, University Medical Center Utrecht Brain Center, Utrecht University, Utrecht, The Netherlands. <sup>13</sup>Institute of Clinical Medicine, Neurology, University of Eastern Finland, Kuopio, Finland. <sup>14</sup>1st Department of Neurology, Memory and Dementia Unit, Aristotle University of Thessaloniki, Thessaloniki, Greece. <sup>15</sup>Department of Medicine, Institute of Gerontology and Geriatrics, University of Perugia, Perugia, Italy. <sup>16</sup>Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford, UK. <sup>17</sup>Medical University of Lodz, Lodz, Poland. <sup>18</sup>INSERM U 558, University of Toulouse, Toulouse, France. <sup>19</sup>Centre for Motor Neuron Disease Research, Macquarie University, Sydney, NSW 2109, Australia. <sup>20</sup>Queensland Brain Institute, The University of Queensland, Brisbane, QLD 4072, Australia. <sup>21</sup>Centre for Clinical Research, The University of Queensland, Brisbane, QLD 4019, Australia. <sup>22</sup>Department of Neurology, Royal Brisbane and Women's Hospital, Brisbane, QLD 4029, Australia. <sup>23</sup>Calvary Health Care Bethlehem, Parkdale, VIC 3195, Australia. <sup>24</sup>Fiona Stanley Hospital, Perth, WA 6150, Australia. <sup>25</sup>Notre Dame University, Fremantle, WA 6160, Australia. <sup>26</sup>Institute for Immunology and Infectious Diseases, Murdoch University, Perth, WA 6150, Australia. <sup>27</sup>The Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Brisbane, QLD 4072, Australia. <sup>28</sup>ANZAC Research Institute, Concord Repatriation General Hospital, Sydney, NSW 2139, Australia. <sup>29</sup>Discipline of Pathology and Department of Neuropathology, Brain and Mind Centre, The University of Sydney, Sydney, NSW 2050, Australia. <sup>30</sup>School of Biomedical Sciences, The University of Queensland, Brisbane, QLD 4072, Australia. <sup>31</sup>New Zealand Brain Research Institute, Christchurch, New Zealand. <sup>32</sup>Department of Medicine, University of Otago, Christchurch, New Zealand. <sup>33</sup>Eskitis Institute for Drug Discovery, Griffith University, Brisbane, Australia. <sup>34</sup>School of Psychology, Speech and Hearing, University of Canterbury, Christchurch, New Zealand. <sup>35</sup>Mater Research, Translational Research Institute, Brisbane, Australia. <sup>36</sup>Mater Research Institute, The University of Queensland, Brisbane, Australia. <sup>37</sup>Brain and Mind Research Centre, Sydney Medical School, The University of Sydney, Sydney, Australia. <sup>38</sup>Department of Pathology and Biomedical Science, University of Otago, Christchurch, New Zealand. <sup>39</sup>Department of Pathology, University of Otago, Christchurch, New Zealand. <sup>40</sup>School of Life Sciences, Westlake University, Hangzhou, Zhejiang, China. <sup>41</sup>Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang, China. <sup>42</sup>Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK. <sup>43</sup>Department of Psychology, Lothian Birth Cohorts group, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK. <sup>44</sup>Department of Basic and Clinical Neuroscience, King's College London, Institute of Psychiatry, Psychology and Neuroscience, London SE5 9RX, UK. <sup>45</sup>SiTraN, University of Sheffield, Sheffield, UK. <sup>46</sup>Queen's University Belfast, Belfast, Northern Ireland, UK. <sup>47</sup>King's College Hospital, London SE5 9RS, UK. <sup>48</sup>Institute of Psychiatry, Psychology & Neuroscience, King's College London, London SE5 8AF, UK.

Received: 13 July 2020 Accepted: 20 January 2021

Published online: 26 March 2021

#### References

- De Jager PL, Yang H-S, Bennett DA. Deconstructing and targeting the genomic architecture of human neurodegeneration. *Nat Neurosci*. 2018;21:1310–7.
- Pihlström L, Wiethoff S, Houlden H. Chapter 22 - Genetics of neurodegenerative diseases: an overview. In: Kovacs GG, Alafuzoff I, editors. *Handbook of Clinical Neurology*. Volume 145: Elsevier; 2018. p. 309–23. <https://doi.org/10.1016/b978-0-12-802395-2.00022-5>.
- Gan L, Cookson MR, Petrucelli L, La Spada AR. Converging pathways in neurodegeneration, from genetics to mechanisms. *Nat Neurosci*. 2018;21:1300–9.
- Armstrong RA, Lantos PL, Cairns NJ. Overlap between neurodegenerative disorders. *Neuropathology*. 2005;25:111–24.
- Gratten J, Visscher PM. Genetic pleiotropy in complex traits and diseases: implications for genomic medicine. *Genome Med*. 2016;8:78.
- Ibanez L, Farias FHG, Dube U, Mihindukulasuriya KA, Harari O. Polygenic risk scores in neurodegenerative diseases: a review. *Curr Genet Med Rep*. 2019;7:22–9.
- Maier RM, Zhu Z, Lee SH, Trzaskowski M, Ruderfer DM, Stahl EA, Ripke S, Wray NR, Yang J, Visscher PM, Robinson MR. Improving genetic prediction by leveraging genetic correlations among human diseases and traits. *Nat Commun*. 2018; 9:989.
- Al-Chalabi A, Fang F, Hanby MF, Leigh PN, Shaw CE, Ye W, Rijdsdijk F. An estimate of amyotrophic lateral sclerosis heritability using twin data. *J Neurol Neurosurg Psychiatry*. 2010;81:1324–6.

9. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, Fiske A, Pedersen NL. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry*. 2006;63:168–74.
10. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry*. 2003;60:1187–92.
11. Wirdefeldt K, Gatz M, Reynolds CA, Prescott CA, Pedersen NL. Heritability of Parkinson disease in Swedish twins: a longitudinal study. *Neurobiol Aging*. 2011;32:1923.e1921–1923.e19238.
12. Cannon JR, Greenamyre JT. The role of environmental exposures in neurodegeneration and neurodegenerative diseases. *Toxicol Sci*. 2011;124:225–50.
13. De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, Eaton ML, Keenan BT, Ernst J, McCabe C, et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat Neurosci*. 2014;17:1156–63.
14. Lunnon K, Smith R, Hannon E, De Jager PL, Srivastava G, Volta M, Troakes C, Al-Sarraj S, Burrage J, Macdonald R, et al. Methyloomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat Neurosci*. 2014;17:1164–70.
15. Roubroeks JAY, Smith AR, Smith RG, Pishva E, Ibrahim Z, Sattler M, Hannon EJ, Kloszewska I, Mecocci P, Soininen H, et al. An epigenome-wide association study of Alzheimer's disease blood highlights robust DNA hypermethylation in the HOXB6 gene. *Neurobiol Aging*. 2020;95:26–45.
16. Roubroeks JAY, Smith RG, van den Hove DLA, Lunnon K. Epigenetics and DNA methylomic profiling in Alzheimer's disease and other neurodegenerative diseases. *J Neurochem*. 2017;143:158–70.
17. Chuang Y-H, Paul KC, Bronstein JM, Bordelon Y, Horvath S, Ritz B. Parkinson's disease is associated with DNA methylation levels in human blood and saliva. *Genome Med*. 2017;9:76.
18. Vallerga CL, Zhang F, Fowdar J, McRae AF, Qi T, Nabais MF, Zhang Q, Kassam I, Henders AK, Wallace L, et al. Analysis of DNA methylation associates the cystine–glutamate antiporter SLC7A11 with risk of Parkinson's disease. *Nat Commun*. 2020;11:1238.
19. Figueroa-Romero C, Hur J, Bender DE, Delaney CE, Cataldo MD, Smith AL, Yung R, Ruden DM, Callaghan BC, Feldman EL. Identification of epigenetically altered genes in sporadic amyotrophic lateral sclerosis. *PLoS One*. 2012;7:e52672.
20. Nabais MF, Lin T, Benyamin B, Williams KL, Garton FC, Vinkhuyzen AAE, Zhang F, Vallerga CL, Restuadi R, Freydenzon A, et al. Significant out-of-sample classification from methylation profile scoring for amyotrophic lateral sclerosis. *NPJ Genomic Med*. 2020;5:10.
21. Hannon E, Dempster E, Viana J, Burrage J, Smith AR, Macdonald R, St Clair D, Mustard C, Breen G, Therman S, et al. An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biol*. 2016;17:176.
22. Viana J, Hannon E, Dempster E, Pidsley R, Macdonald R, Knox O, Spiers H, Troakes C, Al-Sarraj S, Turecki G, et al. Schizophrenia-associated methylomic variation: molecular signatures of disease and polygenic risk burden across multiple brain regions. *Hum Mol Genet*. 2017;26:210–25.
23. McCartney DL, Hillary RF, Stevenson AJ, Ritchie SJ, Walker RM, Zhang Q, Morris SW, Birmingham ML, Campbell A, Murray AD, et al. Epigenetic prediction of complex traits and death. *Genome Biol*. 2018;19:136.
24. Zhang F, Chen W, Zhu Z, Zhang Q, Nabais MF, Qi T, Deary IJ, Wray NR, Visscher PM, McRae AF, Yang J. OSCA: a tool for omic-data-based complex trait analysis. *Genome Biol*. 2019;20:107.
25. McLaughlin RL, Schijven D, van Rheenen W, van Eijk KR, O'Brien M, Kahn RS, Ophoff RA, Goris A, Bradley DG, Al-Chalabi A, et al. Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. *Nat Commun*. 2017;8:14774.
26. Yap H-Y, Tee SZ-Y, Wong MM-T, Chow S-K, Peh S-C, Teow S-Y. Pathogenic role of immune cells in rheumatoid arthritis: implications in clinical treatment and biomarker development. *Cells*. 2018;7:161.
27. Min JL, Hemani G, Davey Smith G, Relton C, Suderman M. Meffil: efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics*. 2018;34(23):3983–9.
28. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res*. 2017;45:e22.
29. Elliott HR, Tillin T, McArdle WL, Ho K, Duggirala A, Frayling TM, Davey Smith G, Hughes AD, Chaturvedi N, Relton CL. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clin Epigenetics*. 2014;6:4.
30. Tsaprouni LG, Yang T-P, Bell J, Dick KJ, Kanoni S, Nisbet J, Viñuela A, Grundberg E, Nelson CP, Meduri E, et al. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics*. 2014;9:1382–96.
31. Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One*. 2013;8:e63812.
32. de Leon J, Diaz FJ. A meta-analysis of worldwide studies demonstrates an association between schizophrenia and tobacco smoking behaviors. *Schizophr Res*. 2005;76:135–57.
33. McClave AK, McKnight-Eily LR, Davis SP, Dube SR. Smoking characteristics of adults with selected lifetime mental illnesses: results from the 2007 National Health Interview Survey. *Am J Public Health*. 2010;100:2464–72.
34. de Bakker PIW, Ferreira MAR, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet*. 2008;17:R122–8.
35. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21:1539–58.
36. Qi T, Wu Y, Zeng J, Zhang F, Xue A, Jiang L, Zhu Z, Kemper K, Yengo L, Zheng Z, et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat Commun*. 2018;9:2282.
37. McRae AF, Marioni RE, Shah S, Yang J, Powell JE, Harris SE, Gibson J, Henders AK, Bowdler L, Painter JN, et al. Identification of 55,000 replicated DNA methylation QTL. *Sci Rep*. 2018;8:17605.
38. Marioni RE, Harris SE, Zhang Q, McRae AF, Hagenaars SP, Hill WD, Davies G, Ritchie CW, Gale CR, Starr JM, et al. GWAS on family history of Alzheimer's disease. *Transl Psychiatry*. 2018;8:99.
39. Nicolas A, Kenna KP, Renton AE, Ticozzi N, Faghri F, Chia R, Dominov JA, Kenna BJ, Nalls MA, Keagle P, et al. Genome-wide analyses identify KIF5A as a novel ALS gene. *Neuron*. 2018;97:1268–1283.e1266.
40. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, Tan M, Kia DA, Noyce AJ, Xue A, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2019;18:1091–102.

41. Shah S, Bonder Marc J, Marioni Riccardo E, Zhu Z, McRae Allan F, Zhernakova A, Harris Sarah E, Liewald D, Henders Anjali K, Mendelson Michael M, et al. Improving phenotypic prediction by combining genetic and epigenetic associations. *Am J Hum Genet.* 2015;97:75–85.
42. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, Boland A, Vronskaya M, van der Lee SJ, Amlie-Wolf A, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A $\beta$ , tau, immunity and lipid processing. *Nat Genet.* 2019;51:414–30.
43. Jansen R, Hottenga J-J, Nivard MG, Abdellaoui A, Laport B, de Geus EJ, Wright FA, Penninx BWJH, Boomsma DI. Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Hum Mol Genet.* 2017;26:1444–51.
44. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, van Iterson M, van Dijk F, van Galen M, Bot J, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet.* 2017;49:131–8.
45. Ng B, White CC, Klein H-U, Sieberts SK, McCabe C, Patrick E, Xu J, Yu L, Gaiteri C, Bennett DA, et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat Neurosci.* 2017;20:1418–26.
46. Carrasquillo MM, Allen M, Reddy JS, Hoffman GE, Dang KK, Calley J, Ebert PJ, Eddy J, Wang X, Greenwood AK, Mostafavi S, CommonMind Consortium (CMC), The AMP-AD Consortium, Omberg L, Peters MA, Logsdon BA, De Jager PL, Ertekin-Taner N, Mangravite LM. Large eQTL meta-analysis reveals differing patterns between cerebral cortical and cerebellar brain regions. *Sci Data.* 2020;7(1):340. <https://doi.org/10.1038/s41597-020-00642-8>.
47. Teschendorff AE, Breeze CE, Zheng SC, Beck S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. *BMC Bioinformatics.* 2017;18:105.
48. Zhang Q, Vallegger CL, Walker RM, et al. Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. *Genome Med.* 2019;11:54. <https://doi.org/10.1186/s13073-019-0667-1>.
49. Barker ED, Cecil CAM, Walton E, Houtepen LC, O'Connor TG, Danese A, Jaffee SR, Jensen SKG, Pariante C, McArdle W, et al. Inflammation-related epigenetic risk and child and adolescent mental health: a prospective study from pregnancy to middle adolescence. *Dev Psychopathol.* 2018;30:1145–56.
50. Ligthart S, Marzi C, Aslibekyan S, Mendelson MM, Conneely KN, Tanaka T, Colicino E, Waite LL, Joehanes R, Guan W, et al. DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biol.* 2016;17:255.
51. Buhl AM, Jurlander J, Geisler CH, Pedersen LB, Andersen MK, Josefsson P, Petersen JH, Leffers H. CLLU1 expression levels predict time to initiation of therapy and overall survival in chronic lymphocytic leukemia. *Eur J Haematol.* 2006;76:455–64.
52. Glad CAM, Andersson-Assarsson JC, Berglund P, Berghorsdottir R, Ragnarsson O, Johannsson G. Reduced DNA methylation and psychopathology following endogenous hypercortisolism - a genome-wide study. *Sci Rep.* 2017;7:44445.
53. Blair LJ, Nordhues BA, Hill SE, Scaglione KM, O'Leary JC III, Fontaine SN, Breydo L, Zhang B, Li P, Wang L, et al. Accelerated neurodegeneration through chaperone-mediated oligomerization of tau. *J Clin Invest.* 2013;123:4158–69.
54. Jinwal UK, Koren J 3rd, Borysov SI, Schmid AB, Abisambra JF, Blair LJ, Johnson AG, Jones JR, Shults CL, O'Leary JC 3rd, et al. The Hsp90 cochaperone, FKBP51, increases Tau stability and polymerizes microtubules. *J Neurosci.* 2010;30:591–9.
55. Zannas AS, Jia M, Hafner K, Baumert J, Wiechmann T, Pape JC, Arloth J, Ködel M, Martinelli S, Roitman M, et al. Epigenetic upregulation of FKBP5 by aging and stress contributes to NF- $\kappa$ B-driven inflammation and cardiovascular risk. *Proc Natl Acad Sci.* 2019;116:11370.
56. Houseman E, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics.* 2012;13:86. <https://doi.org/10.1186/1471-2105-13-86>.
57. Michels KB, Binder AM, Dedeurwaerder S, Epstein CB, Grealley JM, Gut I, Houseman EA, Izzì B, Kelsey KT, Meissner A, et al. Recommendations for the design and analysis of epigenome-wide association studies. *Nat Methods.* 2013;10:949.
58. Boerkoel CF, Takashima H, Stankiewicz P, Garcia CA, Leber SM, Rhee-Morris L, Lupski JR. Periaxin mutations cause recessive Dejerine-Sottas neuropathy. *Am J Hum Genet.* 2001;68:325–33.
59. Guilbot AL, Williams A, Ravisé N, Verry C, Brice A, Sherman DL, Brophy PJ, LeGuern E, Vr D, Bareil C, et al. A mutation in periaxin is responsible for CMT4F, an autosomal recessive form of Charcot-Marie-Tooth disease. *Hum Mol Genet.* 2001;10:415–22.
60. Hung C-W, Chen Y-C, Hsieh W-L, Chiou S-H, Kao C-L. Ageing and neurodegenerative diseases. *Ageing Res Rev.* 2010;9: S36–46.
61. Amor S, Puentes F, Baker D, van der Valk P. Inflammation in neurodegenerative diseases. *Immunology.* 2010;129:154–69.
62. McCombe PA, Henderson RD. The role of immune and inflammatory mechanisms in ALS. *Curr Mol Med.* 2011;11:246–54.
63. Braun PR, Han S, Hing B, Nagahama Y, Gaul LN, Heinzman JT, Grossbach AJ, Close L, Dlouhy BJ, Howard MA, et al. Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals. *Transl Psychiatry.* 2019;9:47.
64. Hannon E, Lunnon K, Schalkwyk L, Mill J. Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics.* 2015;10:1024–32.
65. Hannon E, Mansell G, Burrage J, Kepa A, Best-Lane J, Rose A, Heck S, Moffitt T, Caspi A, Arseneault L, Mill J. Assessing the co-variability of DNA methylation across peripheral cells and tissues: implications for the interpretation of findings in epigenetic epidemiology. *bioRxiv.* 2020; 2020.2005.2021.107730.
66. Brooks BR, Miller RG, Swash M, Munsat TL. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Other Motor Neuron Disord.* 2000;1:293–9.
67. Sachdev PS, Lammel A, Trollor JN, Lee T, Wright MJ, Ames D, Wen W, Martin NG, Brodaty H, Schofield PR. A comprehensive neuropsychiatric study of elderly twins: the older Australian twins study. *Twin Res Hum Genet.* 2009;12:573–82.
68. Project Min EALSSC. Project MinE: study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. *Eur J Hum Genet.* 2018;26:1537–46.
69. Huisman MHB, de Jong SW, van Doormaal PTC, Weinreich SS, Schelhaas HJ, van der Kooij AJ, de Visser M, Veldink JH, van den Berg LH. Population based epidemiology of amyotrophic lateral sclerosis using capture–recapture methodology. *J Neurol Neurosurg Psychiatry.* 2011;82:1165.

70. Costello S, Cockburn M, Bronstein J, Zhang X, Ritz B. Parkinson's disease and residential exposure to maneb and paraquat from agricultural applications in the central valley of California. *Am J Epidemiol*. 2009;169:919–26.
71. Horvath S, Ritz BR. Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging (Albany NY)*. 2015;7:1130–42.
72. Ellis KA, Bush AJ, Darby D, De Fazio D, Foster J, Hudson P, Lautenschlager NT, Lenzo N, Martins RN, Maruff P, et al. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *Int Psychogeriatr*. 2009;21:672–87.
73. Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, Jack CR Jr, Jagust WJ, Shaw LM, Toga AW, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology*. 2010;74:201–9 <http://adni.loni.usc.edu/data-samples/access-data/>.
74. Lovestone S, Francis P, Kloszewska I, Mecocci P, Simmons A, Soininen H, Spenger C, Tsolaki M, Vellas B, Wahlund L-O, et al. AddNeuroMed—the European collaboration for the discovery of novel biomarkers for Alzheimer's disease. *Ann N Y Acad Sci*. 2009;1180:36–46.
75. Lovestone S, Francis P, Strandgaard K. Biomarkers for disease modification trials—the innovative medicines initiative and AddNeuroMed. *J Nutr Health Aging*. 2007;11(4):359–61.
76. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease. *Neurology*. 1984;34:939.
77. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol*. 1999;56:303–8.
78. Furney SJ, Simmons A, Breen G, Pedroso I, Lunnon K, Proitsi P, Hodges A, Powell J, Wahlund LO, Kloszewska I, et al. Genome-wide association with MRI atrophy measures as a quantitative trait locus for Alzheimer's disease. *Mol Psychiatry*. 2011;16:1130–8.
79. Datta SR, McQuillin A, Rizig M, Blaveri E, Thirumalai S, Kalsi G, Lawrence J, Bass NJ, Puri V, Choudhury K, et al. A threonine to isoleucine missense mutation in the pericentriolar material 1 gene is strongly associated with schizophrenia. *Mol Psychiatry*. 2008;15:615.
80. The International Schizophrenia C, Stone JL, O'Donovan MC, Gurling H, Kirov GK, Blackwood DHR, Corvin A, Craddock NJ, Gill M, Hultman CM, et al. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*. 2008;455:237.
81. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinius L, Acevedo N, Taub M, Ronninger M, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol*. 2013;31:142.
82. Padyukov L, Silva C, Stolt P, Alfredsson L, Klareskog L. A gene–environment interaction between smoking and shared epitope genes in HLA–DR provides a high risk of seropositive rheumatoid arthritis. *Arthritis Rheum*. 2004;50:3085–92.
83. Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol*. 2011;41:1576–84.
84. Taylor AM, Pattie A, Deary IJ. Cohort profile update: the Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol*. 2018;47:1042–1042r <https://www.lothianbirthcohort.ed.ac.uk/content/collaboration>.
85. Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, et al. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol*. 2015;16:25.
86. Shah S, McRae AF, Marioni RE, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, Pattie A, Corley J, et al. Genetic and environmental exposures constrain epigenetic drift over the human life course. *Genome Res*. 2014;24:1725–33.
87. Fortin J-P, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, Greenwood CMT, Hansen KD. Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol*. 2014;15:503.
88. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol*. 2014;15:R31.
89. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-Y, Freimer NB, Sabatti C, Eskin E. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet*. 2010;42:348.
90. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet*. 2014;46:100–6.
91. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England)*. 2010;26:2190–1.
92. Venables WN, Ripley BD. *Modern Applied Statistics with S*. 4th ed. New York: Springer; 2002.
93. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77. <https://doi.org/10.1186/1471-2105-12-77>.
94. Nabais MF. Meta-analysis of genome-wide DNA methylation identifies shared associations across neurodegenerative disorders. Github. 2020. <https://doi.org/10.5281/zenodo.4287177>.
95. Gratten J. Epigenome analysis of Parkinson's disease and control samples from the System Genomics Parkinson's disease (SGPD) consortium. GSE145361. *Gene Expression Omnibus*. 2020. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE145361>.
96. Ritz B, Horvath S. Genome wide DNA methylation study of Parkinson's disease in whole blood samples. GSE111629. *Gene Expression Omnibus*. 2018. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111629>.
97. Veldink JH: ProjectMine. European Genome-phenome Archive. 2020. <https://www.ebi.ac.uk/ega/dacs/EGAC00001000703>.
98. Nabais MF, Laws SM, Wray NR, Henders AK, Wallace L, McRae AF. Meta-analysis of genome-wide DNA methylation identifies shared associations across neurodegenerative disorders. GSE153712. *Gene Expression Omnibus*. 2020. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE153712>.
99. Lunnon K, Roubroeks JAY. An epigenome-wide association study of Alzheimer's disease blood highlights robust DNA hypermethylation in the HOXB6 gene. GSE144858. *Gene Expression Omnibus*. 2020. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144858>.
100. Hannon E, Dempster E, Viana J, Burrage J, Smith AR, Macdonald R, St. Clair D, Mustard C, Breen G, Therman S, et al. An integrated genetic-epigenetic analysis of schizophrenia: Evidence for co-localization of genetic associations and

- differential DNA methylation. GSE84727. Gene Expression Omnibus. 2016. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE84727>.
101. Hannon E, Dempster E, Viana J, Burrage J, Smith AR, Macdonald R, St. Clair D, Mustard C, Breen G, Therman S, et al: An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. GSE80417. Gene Expression Omnibus. 2016. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE80417>.
102. Liu Y, AP F: Differential DNA methylation in rheumatoid arthritis. GSE42861. Gene Expression Omnibus. 2013. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse42861>.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

