




DNA Methylation Analysis in Monozygotic Twins Discordant for ALS in Blood Cells

Volkan Yazar^{1*}, Wolfgang P Ruf^{2*}, Antje Knehr², Kornelia Günther², Ole Ammerpohl³, Karin M Danzer^{1,2} and Albert C Ludolph^{1,2}

¹Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), German Center for Neurodegenerative Diseases, Ulm, Germany. ²Department of Neurology, Ulm University, Ulm, Baden-Württemberg, Germany. ³Institute of Human Genetics, Ulm University & Ulm University Medical Center, Ulm, Germany.

Epigenetics Insights
Volume 16: 1–13
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/25168657231172159


ABSTRACT: ALS is a fatal motor neuron disease that displays a broad variety of phenotypes ranging from early fatal courses to slowly progressing and rather benign courses. Such divergence can also be seen in genetic ALS cases with varying phenotypes bearing specific mutations, suggesting epigenetic mechanisms like DNA methylation act as disease modifiers. However, the epigenotype dictated by, in addition to other mechanisms, DNA methylation is also strongly influenced by the individual's genotype. Hence, we performed a DNA methylation study using EPIC arrays on 7 monozygotic (MZ) twin pairs discordant for ALS in whole blood, which serves as an ideal model for eliminating the effects of the genetic-epigenetic interplay to a large extent. We found one CpG site showing intra-pair hypermethylation in the affected co-twins, which maps to the Glutamate Ionotropic Receptor Kainate Type Subunit 1 gene (*GRIK1*). Additionally, we found 4 DMPs which were subsequently confirmed using 2 different statistical approaches. Differentially methylated regions or blocks could not be detected within the scope of this work. In conclusion, we revealed that despite a low sample size, monozygotic twin studies discordant for the disease can bring new insights into epigenetic processes in ALS, pointing to new target loci for further investigations.

KEYWORDS: Amyotrophic lateral sclerosis (ALS), methylome, epigenetics, peripheral blood, monozygotic twins

RECEIVED: January 30, 2023. **ACCEPTED:** April 9, 2023.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Deutsche Forschungsgemeinschaft (DFG) Emmy Noether Research Group DA 1657/2-1.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Albert C Ludolph, Medical Faculty, Department of Neurology, Ulm University, Albert-Einstein-Allee 23, Oberer Eselsberg 45, Ulm, Baden-Württemberg 89081, Germany. Emails: Albert.Ludolph@uni-ulm.de; albert.ludolph@dzne.de

Introduction

Studies investigating DNA methylation in blood cells in ALS have been accumulating in the last few years, indicating the importance of epigenetic studies in the field of motor neuron diseases (MND).¹⁻³ Despite the presence of numerous genomic sites that show static patterns of DNA methylation, there are also methylation sites in the genome that are regulated dynamically and influenced by lifestyle and environmental factors. Furthermore, comorbidities that are not associated with the neurodegenerative disease can have an impact on the DNA methylome, which makes it rather difficult to detect true changes in methylation associated with neurodegenerative diseases. Additionally, a complex interplay between genetic and epigenetic factors makes it even more difficult to detect the subtle changes in DNA methylation in blood cells associated with neurodegeneration. Our design, given the genetic and clinical nature of the study subjects, provides an ideal setup for matching both genotype and rearing environment within each pair of MZ twins, with the unaffected co-twins being the control group. This study design accounts for not only known potential confounding factors, such as genetics, age, and sex, but latent sources of unwanted variability, as well.⁴ Overall, the approach we used in this work, also called the case co-twin

design, is superior to the ordinary case-control design in terms of the power of a significance test,⁵ providing a unique opportunity to detect subtle DNA-methylation changes associated with the motor neuron disease.

Methods

Study design and sampling of monozygotic twins

Sixteen participants (eg, 8 twin pairs) were initially enrolled in the study. One twin pair was excluded because they were dizygotic. So, the present study was composed of 7 monozygotic (MZ) twin pairs discordant for ALS, all genotyped at the University Clinic of Ulm. All individuals were of European descent without known diabetes or cancer history. The affected co-twins were diagnosed according to the revised El-Escorial criteria⁶ for definite ALS. No unaffected co-twins had signs of MND. Zygosity information was further verified using SNP information using targeted sequencing of the known ALS genes and the SNP information provided on the Illumina EPIC DNA methylation arrays.

Ethical approval

All procedures were performed in accordance with the relevant guidelines and regulations stated in the Helsinki II declaration and approved by the local medical ethics committees

* These authors contributed equally to this work.



(ethics committees of Ulm University, approval no 19/12) medical ethics committees. Informed written consent from all study subjects was obtained before enrollment in the study.

Blood collection

Collection of human peripheral venous blood was performed in a non-fasting state using a Monovette™ blood drawing system (Sarstedt, Germany) and EDTA tubes (Sarstedt, Germany) by a standardized procedure.

DNA isolation

5ml EDTA blood was used for DNA extraction. Erythrocytes were lysed in 20ml lysis buffer (155mM NH₄Cl, 10mM KHCO₃, 0.1mM Na₂EDTA, pH=7.4) for 15 minutes on ice. The leukocyte pellet was resuspended in 2.5 ml SE buffer (75 mM NaCl, 25 mM EDTA). To release the DNA, Proteinase K digestion was performed using 100 µl 20% SDS + 10 mg/ml Proteinase K (Roche, Switzerland) overnight at 37°C. 900 µl of saturated sodium chloride solution (6 M NaCl) was added. The DNA was precipitated in 7 ml of 100% ethanol. Then, the DNA was washed in 70% ethanol, dried at room temperature (RT), and resuspended in 100 µl H₂O.

DNA quality control and quantification

Nanodrop (Thermo Fisher Scientific, MA, USA) was used to determine the 260/280 ratio and the 260/230 ratio. All samples had a 260/280 ratio between 1.7 and 2 and a 260/230 ratio greater than 1.5. Nanodrop was used to get a rough estimate of the DNA concentration, then DNA quantification was carried out using a DNA Qubit fluorometer (Thermo Fisher Scientific, MA, USA).

Genotyping

C9orf72 genotyping in all samples was carried out by fragment analysis and repeat-primed PCR (RP-PCR). All normal homozygotes and expanded alleles were confirmed with Southern Blot. For the targeted gene sequencing, a custom panel from Illumina with 36 ALS genes (ie, *ALS2*, *ANG*, *ARHGEF28*, *CCNF*, *CFAP410*, *CHCHD10*, *CHMP2B*, *DCTN1*, *ERBB4*, *FIG4*, *FUS*, *GLE1*, *GRN*, *HNRNPA1*, *HNRNPA2B1*, *MAPT*, *MATR3*, *NEFH*, *NEK1*, *OPTN*, *PFN1*, *PRPH*, *SETX*, *SIGMAR1*, *SOD1*, *SPG11*, *SQSTM1*, *TAF15*, *TARDBP*, *TBK1*, *TUBA4A*, *UBQLN2*, *VAPB*, *VCP*, *VEGFA*, *VPS54*) was used. The subsequent sequencing was performed for 150 cycles by generating 2× 74bp paired-end reads on the MiSeq (Illumina, CA, USA). The coverage of all genes was at least 20×. Enrichment for targeted gene sequencing was performed with the Nextera Rapid Capture_Custom Kit (Illumina, CA, USA).

Sodium bisulfite conversion

500 ng of DNA from each sample underwent sodium bisulfite conversion using the EZ-96 DNA Methylation kit (Zymo Research, CA, USA).

Illumina EPIC DNA methylation arrays

Samples were measured at the single nucleotide resolution across the human genome for CpG methylation levels using Illumina EPIC DNA methylation arrays (Illumina, CA, USA) at the University Clinic of Ulm. This platform offers simultaneous interrogation of CpG methylation levels of over 850 000 different sites spanning CpG islands, genes, enhancers, and beyond. Sample processing for the assay was done according to the manufacturer's protocol and finally, fluorescence imaging of the arrays using an Illumina HiScan SQ scanner yielded raw Intensity Data files (.idat) for all samples, following the standard Illumina scanning protocol. To reduce likely batch-to-batch variation in intra-pair CpG methylation differences, samples from co-twins in a pair were loaded together on the same array.

Data preprocessing

Preprocessing of raw data was adapted from the workflow commonly used for the analysis of array-based methylation data.⁷ An initial round of quality control (QC) checks, filtering, within- and between-library normalizations, and data exploration were performed using multiple R v4.1.2⁸ packages before moving on to any statistical tests for differentially methylated cell types (DMCTs), -probes (DMPs), -regions (DMRs), and -blocks (DMBs). Briefly, detection *P*-values calculated per CpG in all samples separately were used to check global signal reliabilities, with a cutoff of 0.05 for filtering out samples and a cutoff of 0.01 for filtering out probes that performed poorly. Also, many other QC plots were generated using the "qcReports()" function in the R package minfi v1.40.0.⁹ As we did not expect global differences between affected and unaffected co-twins, data normalization was done using the "preprocessQuantile()" function.¹⁰ To confirm for both sample sexes and twin IDs reported in the sample metadata sheet using the methylation data, we used the "getSex()" together with the "plotSex()" functions and the "getSnpBeta()" function in the R package minfi, respectively. Cross-reactive probes and probes associated with SNPs were removed with the "dropCrossReactiveProbes()" function.¹¹ Multiple MDS plots were generated to cluster the samples by kinship, batch, disease status, sex, chronological age, and genotype (ie, *C9orf72* hexanucleotide expansion (HRE) status) using the "plotMDS()" function in the R package limma v3.50.3,¹² which helped us figure out which covariates to include in the regression analysis. Twin correlation on global CpG methylation levels was plotted as a means of QC at the sample level using the "chart.Correlation()" function in the R package

PerformanceAnalytics v2.0.4¹³ considering the notion that a high degree of intra-pair correlation on global CpG methylation can be assumed for genetically-matched twins and that any unusual intra-pair dissimilarity might indicate a confounding factor (e. g. altered blood cell distribution).⁴ M- and β -values were calculated from preprocessed signal intensities for downstream analyses, ranging from 0 (unmethylated) to 1 (fully methylated), using the “getM()” and the “getBeta()” functions, respectively. Finally, a multi-set Venn diagram was plotted to identify shared probes among all twin pairs with an intra-pair absolute difference in β -value of ≥ 0.1 using the “ggVennDiagram()” function in the R package ggVennDiagram v1.2.0.¹⁴

Differentially methylated cell type (DMCT) analyses

Reference-based estimations of blood cell type compositions from ALS and non-ALS whole blood samples were made using the “epidish()” function in the R package EpiDISH v2.10.0,¹⁵ with the “centDHSbloodDMC.m” matrix being the whole blood reference of 333 cell-type specific DNase Hypersensitive Site (tsDHS)-DMCs and 7 blood cell subtypes, and in “constraint projection” mode with “inequality” normalization constraint as implemented in the Houseman et al method¹⁶ for epigenetic dissection of intra-pair cell type heterogeneity. First, a boxplot of estimates grouped by blood cell subtypes with 2 levels (ie, affected and unaffected) was made using the R package ggplot2 v3.3.6¹⁷ for a visual assessment of the estimates as a function of individual cell types and disease status. Then, the “CellDMC()” function from EpiDISH was used to 1. identify differentially methylated cell types, 2. find their directions of change, 3. check if and which specific cell type(s) drives methylation changes observed between ALS and non-ALS samples. By including the statistical interaction terms between disease status and the proportions of underlying cell types within PBMCs in the linear regression (ie, the Robust Partial Correlations (RPCs)) framework normally used to identify DMCs, this method allows for the detection of differentially methylated cytosines in a cell-type specific manner.¹⁵ Eosinophils were excluded from the analysis because their fraction in the estimates was extremely low. Preceding the analysis, all cell type estimates were incremented by 0.001 to deal with zero observations before log transformation of the counts. Lastly, after confirming the non-normality of the data through “shapiro.test()” and “qqPlot()” functions in the R package stats v4.1.2,⁸ pairwise comparisons of each cell type between ALS and non-ALS samples were made using the “wilcox.test()” function in the same R package for a non-parametric alternative to the t-test.

Differentially methylated probe (DMP) analyses

For the differentially methylated probe analysis, we used the case co-twin design to capture intra-pair differences in CpG methylation. β -values were logit-transformed and converted

into intra-pair differences (ie, values of unaffected co-twin in a pair be subtracted from those of the affected in the same pair for each probe across the twin set). This adjustment assigns a critical task to the intercept in the mixed-effects model (LMM) we used for probe-level differential methylation analysis: the mean level of intra-pair methylation differences per probe, when all fixed effect variables are set to zero, is represented by the intercept. In R⁸ we used the lme4 v1.1.29¹⁸ and the lmerTest v3.1.3¹⁹ packages to conduct a linear mixed-effects analysis for the detection of DMPs through the “lmer()” function. As the fixed effects, sex, age, *C9orf72* status, CD4+ T cells, and neutrophils were used as covariates while as the random effects, the batch was the covariate incorporated into our random-intercept-fixed-slopes regression model, leading to the best outcome regarding singularity issues. In an attempt to prevent overfitting, underrepresented cell types in the data set and other low-priority predictors identified based on the MDS plots, including BMI, have been excluded from the regression analysis. The predictors such as different cell types in the model that vary between co-twins have also been converted into intra-pair differences as before, and subsequently mean-centered to improve the numerical stability of the regression analysis. The intercepts, together with various statistics including the *P*-values calculated per probe, have then been collected as model fitting which was accomplished iteratively for the entire data set. Nominal *P*-values were adjusted using a cutoff of FDR ≤ 0.05 for multiple testing corrections through the “p.adjust()” function.²⁰ The “getAnnotation()” function has been used together with the R package IlluminaHumanMethylationEPICanno.ilm10b2.hg19 v0.6.0⁹ for genomic annotation. To assess the distribution of all DMPs across the genome at the chromosome resolution, a Manhattan plot was generated using the “mhtplot()” function in the R package ENmix v1.30.3.²¹ Finally, a selected subset of probes was used to assess indices of model performances (the “compare_performance()” function) and the underlying assumptions, including normality of residuals and random effects (the “check_distribution()” and the “shapiro.test ()” functions), homogeneity of variance (the “check_heterogeneity_bias()” and the “leveneTest()” functions), linear relationship and multicollinearity (the “compare_model()” function), and autocorrelation (the “compare_autocorrelation()” and the “durbinWatsonTest()” functions), using the R packages performance v0.9.1,²² stats v4.1.2, lmtest v0.9.40,²³ and car v3.1.0.¹⁹ Also, it has been assumed that the data have been generated exogenously and that the twin sampling, as well as the entire set of probe-level observations on the EPIC arrays, were made independently. The “pheatmap()” function in the R package pheatmap v1.0.12, together with the color palettes in the package RColorBrewer v1.1.3, was used to generate heatmaps. The annotation track option of the UCSC Genome Browser²⁴ was enabled to illustrate a selected DMP along the CpG islands and gene bodies.

Functional enrichment analysis

After the DMP-annotated genes were extracted, a STRING v11.5²⁵ protein association network analysis was done to investigate functional interactions in this final list of genes with a minimal interaction score of 0.400. A gene list enrichment platform, EnrichR v15.06.2022²⁶ at <https://amp.pharm.mssm.edu/Enrichr/> has produced the pathway- and the ontology-level functional annotation results using the following libraries: KEGG Human PA (2021) and GO Biological Process (2021). The enrichment *P*-values were calculated using a Fisher's Exact Test or a hypergeometric test, which were then corrected for multiple testing using the Benjamini-Hochberg (BH) FDR-controlling method.²⁰

Differentially methylated region (DMR) and -block (DMB) analyses

The data have been grouped by disease status, instead of the twin structure used for the DMP analysis, to estimate "regions" of the genome binned into 1.5 kb windows for which a CpG methylation profile deviates from its baseline value between 2 (ie, affected and unaffected) populations. Detection of DMRs was done on the β -values through the "bumphunter()" function in the R package minfi v1.40.0 with default settings, except a genomic profile cutoff of .2 to identify candidate regions and a resampling number of 1000 to assess uncertainty. The "DMR.plot()" function in the R package DMRcate v2.8.5²⁷ was used to visualize the potential regions within the scope of this analysis. On the other hand, to explore the methylation status of genomic "blocks," which is based on a metric devised to segregate hundreds of neighboring intergenic CpG sites into large-scale genomic regions with similar methylation profiles,²⁸ the probes associated only with OpenSea clusters (ie, the CpG-sparse regions outside the CpG islands) were used per the developer's guidelines.²⁹ In this context, a differentially methylated block (DMB) analysis was performed using the "champ.Block()" function in the R package ChAMP v2.24.0²⁹ between affected and unaffected samples with maximum block and between-cluster gap sizes of 250 kb, a threshold of 10 as a minimum number of probes per block, and a resampling number of 500.

Copy number variations (CNV) analysis

To detect copy number variations between affected and unaffected co-twins using genome-wide CpG methylation profiles,³⁰ we used the "champ.CNA()" function in the same R package ChAMP with default parameters. Two sets of plots were returned, including the aberrations of each sample group and the aberrations of each affected sample with respect to the mean of the unaffected sample group.

DNA methylation (DNAm) age analysis

DNA methylation age estimates were generated using the online calculator³¹ at <https://dnamage.genetics.ucla.edu/>, as described by Horvath (2013). Briefly, any probes missing in our data set that was originally used in the 27 K arrays to develop the underlying algorithm were appended row-wise to obtain the input data set for the analysis, with any missing value replaced by "NA." A matching sample annotation file was generated for an in-depth analysis of DNAm age. After the analysis, the samples that fail to correlate well (ie, $r \leq .80$) with the standard samples were excluded from further consideration. Plots for data visualization were generated using base R graphics or the R package ggplot2 v3.3.6.

Power of a sample analysis

Power analysis was done using the "pwrEWAS_shiny()" function in the R package pwrEWAS v1.8.0³² based on 500 simulations and a target maximal difference in DNA methylation between 20% and 50% at 5% FDR, with limma¹² being the method for DM analysis.

Receiver operating characteristic curve (ROC) analysis

With discriminatory power expressed as the area under the curve (AUC), ROC analysis with sensitivity and specificity calculations were all done using the "plot.roc()" function in the R package pROC v1.18.0³³ to assess how well each DMP identified in this study is capable of discriminating between affecting and unaffected samples. Efficacy evaluation: AUC = 0.5 means non-efficiency, $0.5 \leq \text{AUC} \leq 0.7$ means a modest level of efficiency, $\text{AUC} \geq 0.7$ means high efficiency.

Results

Demographics

The current study is composed of 7 pairs of MZ twins ($n = 14$). Though the initial sample size at the time of study design was 16 (ie, 8 twin pairs), one of the pairs turned out to be dizygotic (DZ) during the verification stage of the zygosity information, and was thereby excluded from the downstream analyses. Of these 7 pairs (Table 1), 3 were female and 4 were male with chronological ages ranging from 38 to 74 years ($\bar{x} = 57.43$, $s = 10.64$, $\tilde{x} = 58$). The age at onset ranged from 36 to 73 years ($\bar{x} = 55.84$, $s = 11.46$, $\tilde{x} = 56.6$). The *C9orf72* hexanucleotide repeat expansion (HRE) was identified in 3 twin pairs (1 male and 2 female twin pairs), 1 twin pair (female twin pair) showed a variant in the *TBK1* (*TBK1*:c.217A>C, p.I73L) gene, 1 twin pair (male twin pair) was identified with a variant in the *NEK1* (*NEK1*:c.1634T>G, p.M545T) gene and 1 twin pair (male twin pair) was identified with a variant in the *NEFH* gene

Table 1. Demographics for the MZ twin set.

TWIN PAIR ID	DISEASE STATUS	GENDER	AGE	MUTATION	AGE OF ONSET	BMI
Twin1	Unaffected	female	62	C9Orf72 HRE	0	22.9
	Affected	female	62	C9Orf72 HRE	60	19.2
Twin2	Unaffected	male	53	NEK1: p.M545T	0	23.9
	Affected	male	53	NEK1: p.M545T	53	20.2
Twin3	Unaffected	male	38	wt	0	23.3
	Affected	male	38	wt	37	19.5
Twin4	Unaffected	male	54	NEFH: p.R352S	0	23.1
	Affected	male	54	NEFH: p.R352S	52	22.3
Twin5	Unaffected	female	58	TBK1: p.I73L	0	28.3
	Affected	female	58	TBK1: p.I73L	57	27.6
Twin6	Unaffected	female	63	C9Orf72 HRE	0	23.1
	Affected	female	63	C9Orf72 HRE	62	18.7
Twin7	Unaffected	male	74	C9Orf72 HRE	0	22.8
	Affected	male	74	C9Orf72 HRE	73	21

Abbreviations: HRE, hexanucleotide repeat expansion; wt, wild type for known ALS genes; BMI, body mass index.

(NEFH: c.1054C>A, p.R352S). In total there were 3 spinal ALS cases, 3 bulbar ALS cases, and one ALS-FTD case.

Intra-pair methylation profiles of the mutated ALS genes in the corresponding twin pairs

To determine whether there is any discrepancy in the methylation status of an upstream CpG island spanning the promoter region and exon 1 of each of the mutated ALS genes between affected and unaffected co-twins in the corresponding twin pair, we used the associated probes in the EPIC data and observed a highly similar pattern in the methylation profile of each of these regions per intra-pair comparison (Figure 1). As expected, there is a lack of DNA methylation at CpG islands encompassing the first exons of the *TBK1* and *NEK1* genes (Figure 1b and c). Importantly, as the probes mapping to CpG islands associated with the *C9orf72* gene were filtered out during preprocessing (Supplemental Figure 1), we confined our analysis to this subgroup of ALS genes mutated in our MZ twin set (Table 2). Overall, there appears to be a clear consistency in the methylation profiles of the mutated ALS genes between co-twins in the mutation-positive twin pair, in line with the previous literature.³⁴

Global changes in CpG methylation across the twin pairs

After confirming the twin correlation on global CpG methylation levels using the methylation data, as well as sample sex and

twin IDs (Supplemental Figure 2), we explored any relationship between genome-wide DNA methylation and chronological age within- or between MZ twin pairs (Figure 2a). With the mean levels of global DNA methylation being almost the same across all 14 samples, there appeared to be no particular trend associating genomic methylation on a broad scale with age, which also seems to be independent of disease status. Then, we sorted the data by sex and repeated the experiment, only to see the same outcome (Figure 2b): there is likely no relationship between sex and CpG methylation at the genome level, with or without disease status involved. Taken together, our observations on global alterations in CpG methylation across the twin pairs are mostly in agreement with the current literature.

*Intra-pair variability analysis revealed altered DNA methylation associated with the *GRIK1* gene*

A consistent body of previous literature revealed that intra-pair methylation differences in disease-discordant MZ twins are generally low (≤ 0.20) at the probe level due to the matched genetic make-up and shared rearing environment between co-twins.^{1,4,5,34} Based on this observation, we tested for probes with a sufficient (ie, $\Delta\beta$ -value ≥ 0.10) intra-pair variability shared by all twin pairs in our data set (Figure 2c). We identified cg26023019 with a $\Delta\beta$ -value above the threshold, which maps to the *GRIK1* gene, which was previously associated with ALS in a group of independent ALS-CNV studies.^{35,36} Lastly, after estimating blood cell type compositions from ALS and

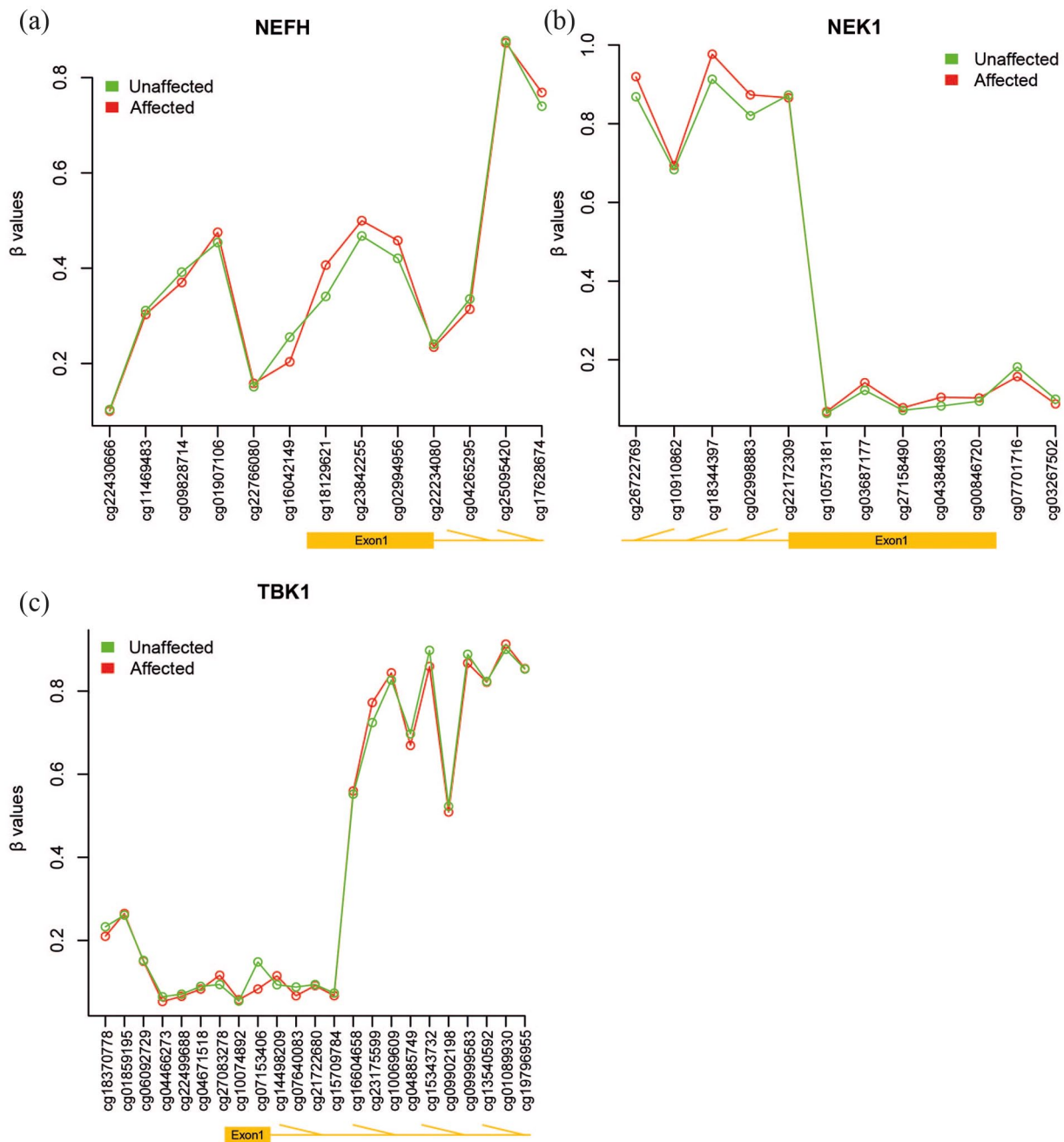


Figure 1. The promoter proximal sites of neither NEFH nor NEK1 or TBK1 are differentially methylated between mutation-positive ALS-discordant twins. The methylation β -values of both the relative location of the promoter proximal regions and exon 1 for NEFH ((a), upper left), NEK1 ((b), upper right), and TBK1 ((c), bottom) plotted as a function of corresponding probe IDs. Methylation status of each gene was identified using EPIC arrays. Methylation profile of the promoter regions and exon 1 of NEFH, NEK1, and TBK1 does not indicate any pattern of differential methylation between affected and unaffected co-twins in the corresponding twin pairs, concordant for NEFH p.R352S, NEK1 p.M545T, and TBK1 p.I73L, respectively.

non-ALS whole blood samples, we performed a pair of DMCT analyses (Figure 2d). Using neither a particular linear regression framework described in the Methods section nor individual pairwise comparisons of each cell type between ALS and non-ALS samples we could detect DMCT by disease status. Nevertheless, elevated levels of B cells and granulocytes were observed in ALS-affected co-twins, as published before.³ As a final point, we checked to see whether the intra-pair correlations computed on CpG methylation decline with biological age, as previously reported in the literature^{4,37} for MZ twin

pairs. As expected, a downtrend line suggesting a decrease in the correlation with biological aging was observed (Supplemental Figure 2c), implying the importance of adjusting for age in estimating the relationships between CpG methylation patterns and disease status.

Probe-level differential methylation analysis

To analyze the methylation data from ALS-discordant MZ twins at the probe level using the case co-twin design, we used

Table 2. Pairwise correlations between individual samples and the gold standard.

TWIN ID	PEARSON'S R^2
Twin1 (Unaff.)	0.905
Twin1 (ALS)	0.916
Twin2 (Unaff.)	0.920
Twin2 (ALS)	0.919
Twin3 (Unaff.)	0.918
Twin3 (ALS)	0.919
Twin4 (Unaff.)	0.914
Twin4 (ALS)	0.914
Twin5 (Unaff.)	0.920
Twin5 (ALS)	0.920
Twin6 (Unaff.)	0.910
Twin6 (ALS)	0.915
Twin7 (Unaff.)	0.916
Twin7 (ALS)	0.915

The gold standard is defined by averaging the beta values across the samples from the largest blood data set. The accepted threshold of $R^2 = 0.800$.

a mixed-effects model with fixed effects of age, sex, *C9orf72* status, and cell type estimates, with random effects of batch (Supplemental Figure 3). As indicated above for biological age, though discordant MZ twins are matched for biological factors specific to individual pairs, the effects of these factors on the within-pair difference in CpG methylation patterns should be accounted for,^{4,5} which is addressed in the Discussion section. Considering that the intercept is the key term of the analysis as the mean level of intra-pair methylation differences per probe when all fixed effect variables are set to zero, we plotted the frequency distribution of the intercept estimates obtained from the regression analysis. In line with the literature,³⁸ they followed a clear pattern of normal distribution (Figure 3a), just as the M-values. Ultimately, a total of 35 probes were found significant at an $FDR \leq 0.05$ with varying levels of significance (Supplemental Figure 4). In an attempt to visually assess their degree of variation in methylation levels by disease status, we observed a heterogeneous profile of CpG methylation for the DMPs (Figure 3b). Subsequently, each of the underlying assumptions of our linear model was evaluated using a selected subset of probes for the quality of the model fit, performance characteristics, and diagnostic accuracy. From a technical aspect, a remarkable robustness of LMMs has been confirmed in the literature,^{39,40} except for small biases in estimated parameters stemming from missing predictors or random effect components, even if the majority of the assumptions are violated. Our results revealed, though using a subset of probes, that the

majority of the assumptions were met (Supplemental Figures 5 and 6). Next, we checked to see whether these DMPs are capable of distinguishing between ALS and non-ALS samples. The results of different methodologies converged on a low power of discrimination between sample groups given just these significant probes (Figure 3c and d). When the DMP-annotated genes were clustered based on the direction of the change in methylation, 2 groups emerged: (a) *ATG9A*, *BCAT1*, *BARX2*, *C1orf103*, *CYGB*, *FES*, *MYH9*, *PARVB*, and *RNF4* as hypomethylated genes, (b) *ATP6VOB*, *BOC*, *BUD31*, *C3orf75*, *FEZ2*, *IL17D*, *KILLIN*, *MEGF6*, *OSBPL10*, *PDAP1*, *PTEN*, *RSPO4*, *SLX4*, *UBLCP1*, and *USP20* as hypermethylated. Though not significant (FET P -value: .0699), protein association network analysis of these 24 genes performed using the STRING platform revealed the enrichment of a part of the ubiquitin (Ub)-proteasome pathway (UPP) of protein degradation, which is known to be dysfunctional in ALS.^{41,42} On the other hand, “autophagy” and “axon guidance” were identified as the top deregulated signaling pathways (q -value = 0.012 and 0.020, respectively), and “actin cytoskeleton reorganization” was identified as the top enriched biological process (q -value = 0.003) in these DMP-annotated genes which has been strongly associated with ALS pathogenesis.⁴³⁻⁴⁵ To detect differences up to 20% and 50% in CpG methylation between affected and unaffected co-twins with a minimum number of 7 subjects in a sample group based on the power of sample calculations, the calculated power of making a statistically significant comparison at an $FDR \leq 0.05$ turned out to be low, as expected. Yet, it was also observed for target differences of 20% and 50% that the probability of detecting at least 1 true positive significant probe out of 35 DMPs identified between affected and unaffected co-twins is about 10% and 61%, respectively given the current sample size (Supplemental Figure 7). In other words, approximately 4 of the DMPs identified within the scope of this study are expected to be true positive results with $\geq 20\%$ methylation difference between sample groups. Interestingly, our ROC analysis revealed using a classical AUC threshold of 0.70 that there are exactly 4 probes (Figure 4a) out of 35 DMPs (Supplemental Figure 8) that appear to classify the sample groups well. While 3 of these candidate probes are located in the “OpenSea clusters” of the genome, one of them with an AUC value of 0.755 is located near the TSS of the *IL17D* gene, which is implicated in positive regulation of cytokine production during inflammatory response⁴⁶ (Figure 4b). Overall, our probe-level differential methylation analysis appears to provide a small but biologically relevant group of genes that are differentially methylated within the context of ALS.

Region- and block-level differential methylation analyses

After the data have been regrouped according to their disease status, ignoring the twin structure used for the DMP analysis, we also conducted differential methylation analyses at 2 different

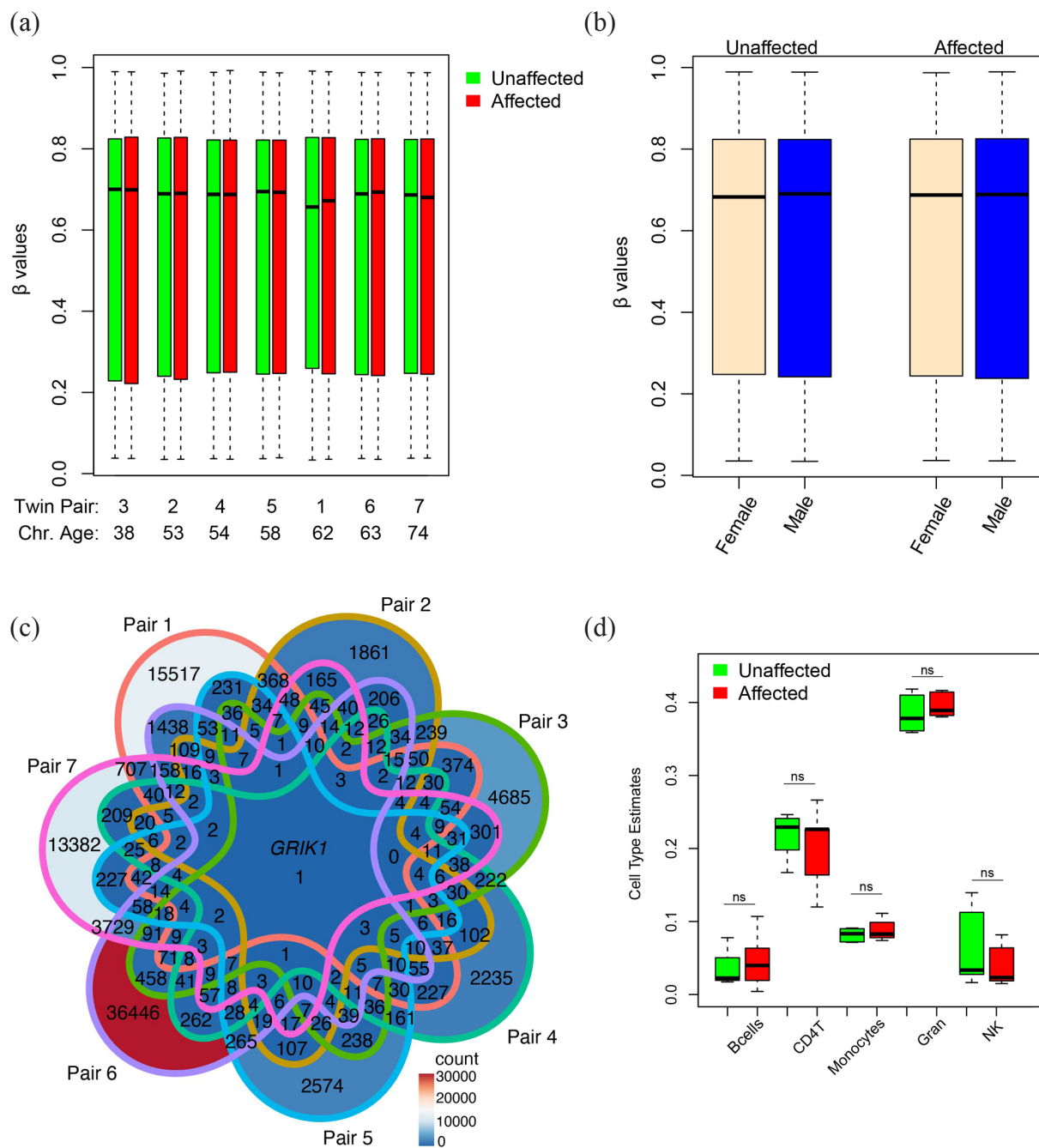


Figure 2. Global changes in CpG methylation across the twin pairs. (a) An attempt to explore any relationship between genome-wide DNA methylation and chronological age within- or between MZ twin pairs revealed that there is no particular trend associating genomic methylation on a broad scale with aging regardless of disease status. All twin pairs are ranked by chronological age. $N(\text{un})\text{affected}=7$. (b) It is also clear that there is no relationship between gender and CpG methylation at the genome level, with or without disease status involved. $N(\text{un})\text{affected, female}=3$, $N(\text{un})\text{affected, male}=4$. (c) Intra-pair variability analysis performed to identify probes with a sufficient (ie, $\Delta\beta$ -value ≥ 0.10) intra-pair variability shared by all twin pairs in our data set revealed altered DNA methylation associated with the GRIK1 gene. Also, the twin pairs 1, 6, and 7 accommodate the highest number of pair-specific probes with sufficient intra-pair variability, with 6 being the most extreme. (d) Comparisons of blood cell type composition estimates using neither basic univariate (ie, wilcoxon test) nor more advanced multivariate statistics (ie, the Robust Partial Correlations (RPCs)) approaches reported differentially methylated cell types by disease status. Yet, an increase in B cell and granulocyte levels is apparent in ALS-affected co-twins. ns: not significant. $N(\text{un})\text{affected}=7$.

levels, namely genomic region (ie, a 1.5 kb genomic window) and genomic block (ie, a genomic bin up to 250 kb). In this context, each block contains hundreds of neighboring intergenic CpG sites with similar methylation profiles. As can be seen in Supplemental Figure 9, the lack of significant deviations from

the baseline value between affected and unaffected populations is apparent even when the smoothed group means for the top-ranking region are compared. The same holds for the DMB analysis: no genomic blocks were identified differentially methylated between the sample groups. Overall, ALS-discordant

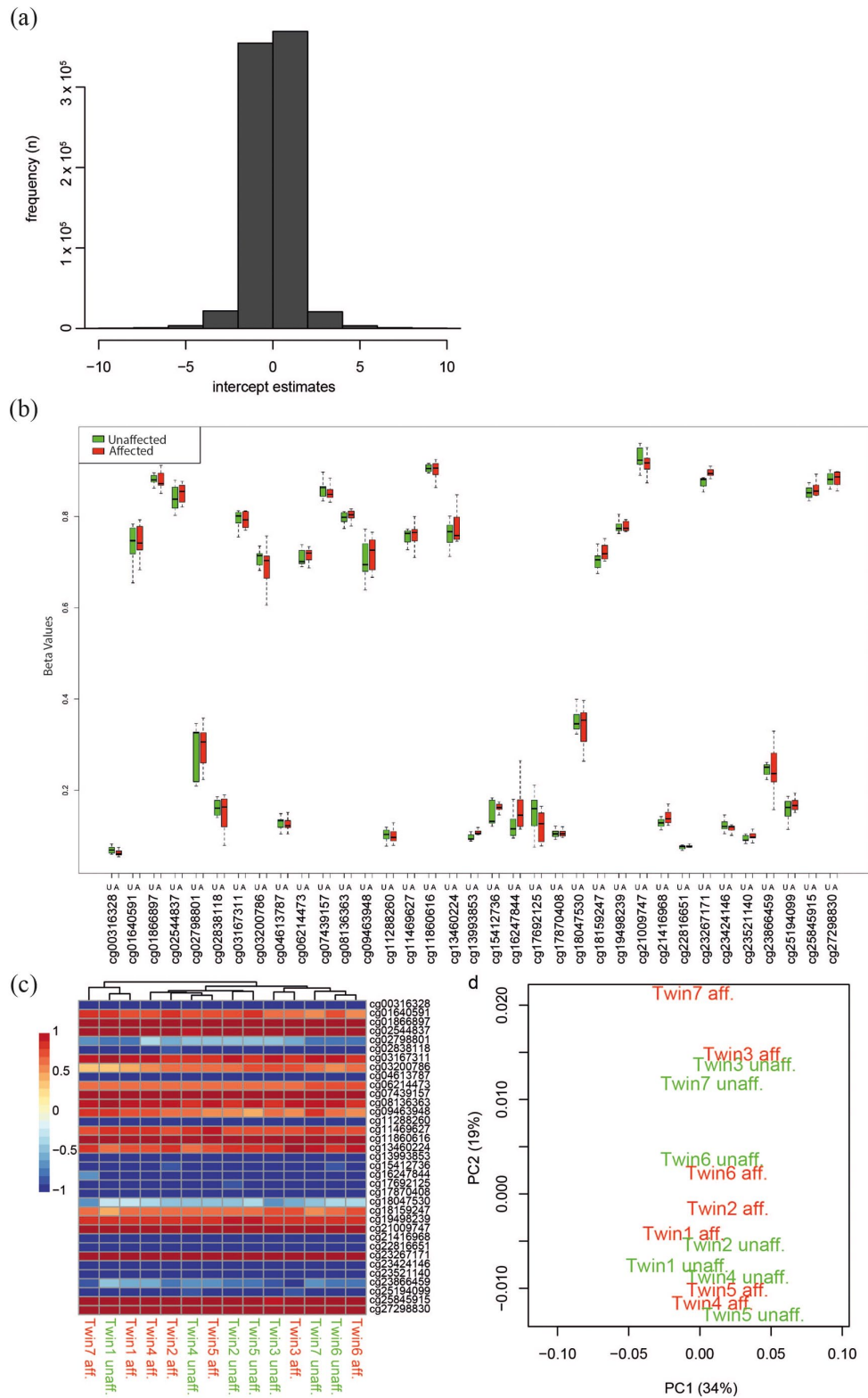


Figure 3. Probe-level differential methylation analysis identified 35 DMPs with varying levels of methylation changes between conditions. (a) As a QC step for the statistical approach (ie, a linear mixed-effects model) used in the differentially methylated probe analysis, the frequency distribution of the intercept estimates obtained from the regression analysis follows a clear pattern of normal distribution. (b) The degree of variation by disease status in methylation levels of the 35 DMPs identified within the scope of this work is highly variable: while there are DMPs with clear differences in methylation levels between ALS and non-ALS samples, there are also DMP that have roughly similar levels of methylation by disease status when raw β values are used for visual assessment, implying the importance of adjusting for key covariates such as age and gender in estimating the relationships between CpG methylation patterns and disease status. $N(\text{un})\text{affected}=7$. (c and d) When we tested whether these DMPs are capable of distinguishing between ALS and non-ALS samples, the results of different methodologies (ie, a heatmap (c) or a PCA (d) of raw methylation levels) indicated a low power of discrimination between sample groups given just the significant probes. In other words, methylation was roughly similar across samples for most DMPs, and samples did not seem to cluster by disease status. PC1 in (d) explains 34% of the variation in the input data of DMPs.

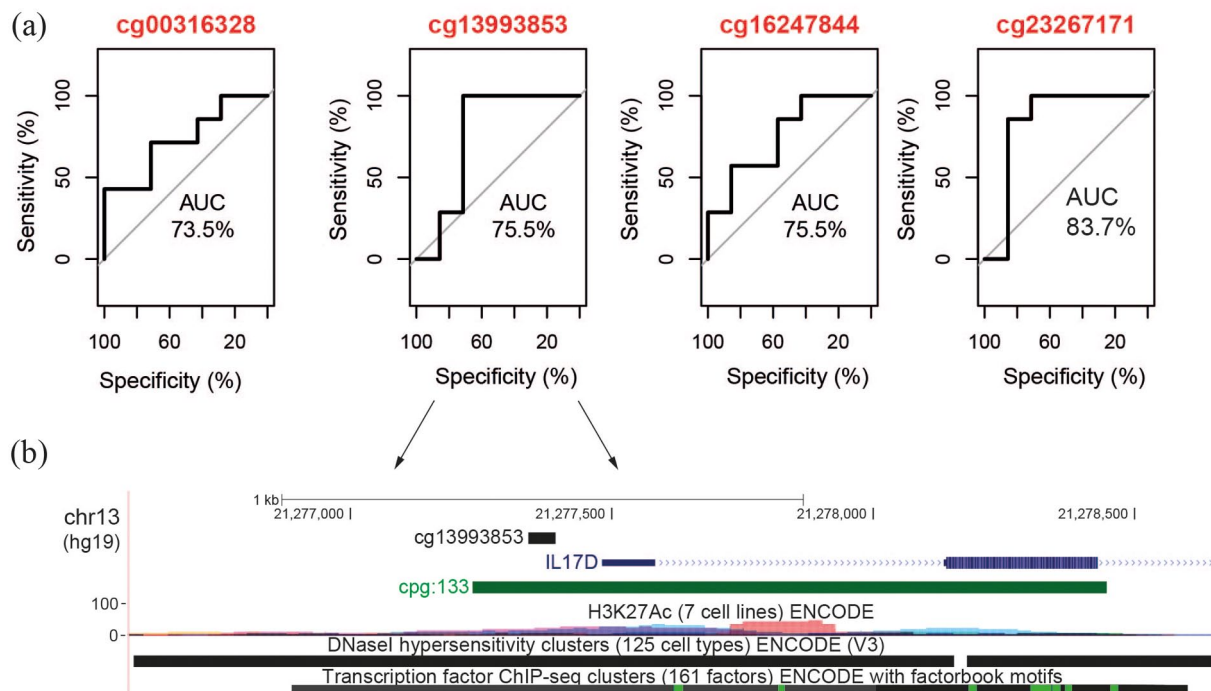


Figure 4. In silico validation of DMP analysis results indicate differential methylation of IL17D in ALS blood. (a) The ROC analysis results of 35 DMPs using a classical cutoff value at $AUC \geq 0.70$ identified 4 probes that appear to classify the sample groups well, just as expected after the results of the power of sample calculation. The corresponding probe IDs are shown in red at the top of each plot. $N(\text{un})\text{affected}=7$. (b) Graphical representation of the DMP with the ID number “cg13993853” overlapping the IL17D promoter region, as indicated by the adjacent activating epigenetic marks (eg, CpG islands, H3K27Ac, and DNaseI hypersensitivity), as well as the neighboring TF binding sites. Unlike the other 3 candidate probes that are located in the “OpenSea clusters” of the genome, this probe with an AUC value of 0.755 is the only candidate probe that is located near the TSS of a gene.

twin studies with relatively small sample sizes seem to work better for probe-level than region-level methylation analysis.

Copy number variation analysis

Finally, we performed a copy number variation (CNV) analysis using the genome-wide CpG methylation profiles at the probe resolution between ALS and non-ALS samples. The copy number variations (gains or losses) both between sample groups and of each affected sample with respect to the mean of the unaffected sample group failed to yield a consistent pattern for disease status (Supplemental Figure 10).

Discussion

In this study, our goal was to provide further biological insight into ALS pathogenesis at the DNA methylation level using whole blood samples of ALS-discordant MZ twins. We have shown that a few CpG locations in the genome carry differential methylation marks associated with ALS. In addition, we observed differential methylation of neither genomic bins of different sizes nor blood cell types associated with ALS in our data set. Nevertheless, global patterns of genomic methylation across the twin pairs, including an increasing intra-pair discrepancy in methylation age and a decreasing intra-pair correlation on CpG methylation with chronological aging, a low degree of variation in within-pair CpG methylation, and elevated levels of B cells and neutrophils in ALS-affected

co-twins, were all in agreement with the current literature. Our results imply that the methylation landscape of whole blood in ALS is a dynamic and highly heterogeneous network at the genome level, with only some commonly found perturbations and affected pathways contributing to the disease.

As rarely as these study subjects can be found, our twin set presumably harbors invaluable information on the molecular etiology of ALS, independent of the genetic mutations in the ALS gene each twin pair carries. More specifically, the ultimate reason why we chose a study design as such, that is, with identical twins that have remarkably different phenotypes, was because co-twins in a pair are matched for familial factors, including but not limited to baseline genetic sequence, age, sex, and rearing environment, thereby reducing the variability between pairs of subjects, but are still discordant for the disease. This leaves room for epigenetic components that might be involved in the disease pathogenesis. We observed rather consistent methylation profiles of the mutated ALS genes between co-twins in the mutation-positive twin pair, implying causative mutations detected in each twin pair are independent of differential CpG methylations identified within the scope of this study.

Another limitation of the current work relates to the LMM that was found to produce heavily inflated P -values (likely a high number of false positives⁴⁷) within the context of the probe-level differential methylation analysis. As statistically

precarious as this situation may be, such inconsistencies can be explained in the context of our experimental design, as well as our analysis workflow, independent of modeling errors. However, it should also be noted that as we preferred the case co-twin design to the classical twin design for DMP analysis chiefly due to our initial experimental design, the whole procedure of data preparation and processing mentioned in the “Differentially methylated probe (DMP) analyses” part of the Methods section, composed mainly of transformations, value adjustments, selection of predictor variables (both fixed and random effect covariates) in the LMM, model fitting, and extraction of key metrics for downstream analysis, has been planned and performed based on the developers’ guidelines.⁵ This is also evident by the total number of DMPs identified in this work, which is in line with many other discordant twin studies in the literature.^{34,48,49} Back to the potential reasons for inflated P -values, outlier probes in a data set are widely known to cause instability in statistical tests, which can result in exceptionally high standard errors, and thereby, inflated P -values. The `preprocessQuantile()` function used for data normalizations during preprocessing stage of the analysis workflow also fixes outliers, but only those that have low (ie, close to zero) signal intensities, not high. In this regard, it is likely that the significant results are, to some extent, driven by technical artifacts or outliers. In fact, only 4 of the 35 DMPs identified in this study could be validated using ROC analysis, just as predicted using the power of sample calculations as the total number of true positive findings from DMP analysis with $\geq 20\%$ methylation difference between sample groups. This is why we decided to prioritize the genomic locations associated with these 4 DMPs over those associated with all DMPs while making biological interpretations of our results. Another point is that we already performed a comprehensive search trying most of the combinations of predictor variables to find the subset with the best evaluation criterion using not only simple linear model but also mixed-effects model, with the MDS/PCA results in mind, before constructing the final version of our multivariate regression model. In this respect, singularity issues and multicollinearity were also taken into account as selection criteria. Lastly, our power calculations demonstrated that the small sample size we had for the EWAS presented here made it rather difficult to obtain any significant results (Supplemental Figure 7). So, there is also a chance, though not the most likely scenario, that small sample sizes lead to such statistical biases as an inflated type-1 error, as reported in many different clinical trial studies before.⁵⁰ Taken together, the statistical approach used in this work, though not minimizing the risk of false positive findings being reported together with true positive results, allowed us to identify a small subset of DMPs that were later validated bioinformatically.

An overly underestimated concept in discordant MZ twin studies is the intra-pair dynamics of epigenetic drift. Epigenetic drift can be defined as the divergence of the epigenome with

biological age due to stochastic alterations in methylation.⁵¹ This paradigm has often been overlooked especially when the methylation changes are investigated, with chronological age being part of the equation, in identical twins who do not share the disease phenotype between co-twins. It is critical to note that the effect of the age-associated intra-pair difference in DNA methylation status must not be disregarded in such studies though the biological factors specific to individual pairs are matched out in MZ twins.⁵² Empirical evidence confirming this effect came from the observation that twin correlation on genome-wide DNA methylation declines with age (Supplemental Figure 2). If this increasing divergence of the epigenome with age actually roots from stochastic alterations, then co-twins in identical twin pairs should be affected differently along the natural course of epigenetic drift, leading to diminishing within-pair twin correlations on global DNA methylation. From a technical perspective, this approach might as well serve as a QC check in MZ twin studies and be used to monitor sample compliance.

Considering our analysis approach for probe-level differential methylation, one could easily claim that age and sex are matched out in the case co-twin design and that including them as predictor variables in the final model seems excessive. Yet, such an argument does not apply to the EWAS as: (a) intra-pair differences in the epigenome are likely to be higher in older twin pairs, just as mentioned above, (b) it has been known that intra-pair epigenetic differences in particular regions of the genome vary by sex.⁵ As a result, fixed effect covariates as such can improve the regression analysis considerably when included in the model. From a practical standpoint, the regression models we tested in the absence of such predictors did not seem to yield any better results, either. Further examples can be found in other studies with similar experimental designs in the literature.^{4,34,48}

The only probe with sufficient (ie, $\Delta\beta$ -value ≥ 0.10) intra-pair variability shared by all twin pairs in our data set, cg26023019, was found to be hypermethylated in affected co-twins and map to *GRIK1* which is involved in the excitatory neurotransmitter receptors in the mammalian brain. The glutamate signaling pathway is known to be intimately related to both acute and chronic neurodegenerative disorders, including ALS.⁵³ In this regard, several lines of evidence also suggest that excessive activation of the glutamate receptors results in excitotoxicity.⁵³⁻⁵⁵ With the *GRIK1* gene being hypermethylated (or presumably transcriptionally suppressed) in ALS samples, our findings implicate the activation of a compensatory mechanism responsible for restoring the overly active glutamate signaling back to normal levels in ALS. Further research is needed to confirm this hypothesis and to understand the potential role of *GRIK1* in ALS biology.

Finally, 1 of 4 DMPs identified within the scope of this study and later verified by ROC was found to be hypermethylated in affected co-twins and located near the TSS of the

IL17D gene, which is implicated in positive regulation of cytokine production during an inflammatory response. Interestingly, recent evidence has shown that *IL17* directly affects human motor neuron survival and directly contributes to motor neuron degeneration in ALS.^{56,57} Considering that *IL17D* is vastly expressed in various brain regions⁵⁸ and that blood is a reasonable surrogate for brain tissue in DNA methylation studies,⁵⁹ *IL17D* has the potential to serve as a blood-based biomarker.

Conclusion

Despite the small sample size of the current study, we identified a group of DMPs that might serve as potential epigenetic loci for future ALS studies. Especially for the studies investigating different phenotypes (eg, fast vs slow ALS progressors) irrespective of the genetic cause underlying the disease, testing for these DMPs might provide meaningful biological outcomes and could further validate the findings of the current study.

Acknowledgments

We thank Ms. Julia Kühlwein (M.Sc.) for the assistance with preparation of patient samples and data generation.

Authors' Contributions

VY—methodology, data analysis in all aspects, visualization, data interpretation, writing original draft (major), writing review and editing. WPR—medical diagnosis of patients, study design, data generation, data interpretation, writing original draft (minor), writing review and editing. AK—acquisition and preparation of patient samples. KG—acquisition and preparation of patient samples, study design. OA—genotyping of patients, conceptualization, intellectual input. KMD—intellectual input, funding acquisition, study design. ACL—conceptualization, funding acquisition, intellectual input, study design. All authors reviewed and approved the manuscript.

ORCID iDs

Volkan Yazar  <https://orcid.org/0000-0003-3861-0181>

Wolfgang P Ruf  <https://orcid.org/0000-0002-8105-6123>

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

- Ruf WP, Hannon E, Freischmidt A, et al. Methylome analysis of ALS patients and presymptomatic mutation carriers in blood cells. *Neurobiol Aging*. 2022;116:16-24.
- Hop PJ, Zwamborn RAJ, Hannon E, et al. Genome-wide study of DNA methylation shows alterations in metabolic, inflammatory, and cholesterol pathways in ALS. *Sci Transl Med*. 2022;14:eabj0264.
- Nabais MF, Lin T, Benyamin B, et al. Significant out-of-sample classification from methylation profile scoring for amyotrophic lateral sclerosis. *NPJ Genomic Med*. 2020;5:1-9.
- Tan Q, Frost M, Heijmans BT, et al. Epigenetic signature of birth weight discordance in adult twins. *BMC Genomics*. 2014;15:1062.
- Tan Q, Christiansen L, von Bornemann Hjelmberg J, Christensen K. Twin methodology in epigenetic studies. *J Exp Biol*. 2015;218:134-139.
- Ludolph A, Drory V, Hardiman O, et al. A revision of the El Escorial criteria - 2015. *Amyotroph Lateral Scler Frontotemporal Degener*. 2015;16:291-292.
- Maksimovic J, Phipson B, Oshlack A. A cross-package bioconductor workflow for analysing methylation array data. *F1000Res*. 2016;5:1281.
- R: The R Project for Statistical Computing. Accessed July 14, 2022. <https://www.r-project.org/>
- Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30:1363-1369.
- Touleimat N, Tost J. Complete pipeline for Infinium® human methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics*. 2012;4:325-341.
- Chen YA, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 2013;8:203-209.
- Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43:e47.
- R: Econometric tools for performance and risk analysis. Accessed July 14, 2022. <http://braverock.com/brian/R/PerformanceAnalytics/html/PerformanceAnalytics-package.html>
- Gao CH, Yu G, Cai P. ggVennDiagram: an intuitive, Easy-to-Use, and highly customizable R package to generate Venn diagram. *Front Genet*. 2021;12:1-8. doi:10.3389/fgene.2021.706907.
- Zheng SC, Breeze CE, Beck S, Teschendorff AE. Identification of differentially methylated cell types in epigenome-wide association studies. *Nat Methods*. 2018;15:1059-1066.
- Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86.
- Ggplot2. Accessed July 14, 2022. <https://link.springer.com/book/10.1007/978-3-319-24277-4>
- Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67:48.
- Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: Tests in linear mixed effects models. *J Stat Softw*. 2017;82:1-26.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 1995;57:289-300.
- Xu Z, Niu L, Li L, Taylor JA. ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip. *Nucleic Acids Res*. 2016;44:e20.
- Lüdecke D, Ben-Shachar M, Patil I, Waggoner P, Makowski D. Performance: an R package for assessment, comparison and testing of statistical models. *J Open Source Softw*. 2021;6:3139.
- Zeileis A, Hothorn T. Diagnostic Checking in Regression Relationships. Published online 2015. Accessed July 16, 2022. <https://www.semanticscholar.org/paper/Diagnostic-Checking-in-Regression-Relationships-Zeileis-Hothorn/58b4d02ec546c30e31a3ead4cb956fc451e6e95>
- Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res*. 2002;12:996-1006.
- Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47:D607-D613. doi:10.1093/nar/gky1131.
- Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14:128.
- Peters TJ, Buckley MJ, Statham AL, et al. De novo identification of differentially methylated regions in the human genome. *Epigenet Chromatin*. 2015;8:6.
- Liang N, Li B, Jia Z, et al. Ultrasensitive detection of circulating tumour DNA via deep methylation sequencing aided by machine learning. *Nat Biomed Eng*. 2021;5:586-599.
- Tian Y, Morris TJ, Webster AP, et al. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics*. 2017;33:3982-3984.
- Feber A, Guilhamon P, Lechner M, et al. Using high-density DNA methylation arrays to profile copy number alterations. *Genome Biol*. 2014;15:R30.
- Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14:3156.
- Graw S, Henn R, Thompson JA, Koestler DC. pwrEWAS: a user-friendly tool for comprehensive power estimation for epigenome wide association studies (EWAS). *BMC Bioinformatics*. 2019;20(1):218. doi:10.1186/s12859-019-2804-7
- Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
- Tarr IS, McCann EP, Benyamin B, et al. Monozygotic twins and triplets discordant for amyotrophic lateral sclerosis display differential methylation and gene expression. *Sci Rep*. 2019;9:8254.

35. Wain LV, Pedroso I, Landers JE, et al. The role of copy number variation in susceptibility to amyotrophic lateral sclerosis: genome-wide association study and comparison with published loci. *PLoS One*. 2009;4:e8175.
36. Cronin S, Blauw HM, Veldink JH, et al. Analysis of genome-wide copy number variation in Irish and Dutch ALS populations. *Hum Mol Genet*. 2008;17:3392-3398.
37. Talens RP, Christensen K, Putter H, et al. Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs. *Aging Cell*. 2012;11:694-703.
38. Kruppa J, Sieg M, Richter G, Pohrt A. Estimands in epigenome-wide association studies. *Clin Epigenetics*. 2021;13:98.
39. Mansell G, Gorrie-Stone TJ, Bao Y, et al. Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. *BMC Genomics*. 2019;20:366.
40. Schielzeth H, Dingemans NJ, Nakagawa S, et al. Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods Ecol Evol*. 2020;11:1141-1152.
41. Farrowell NE, McAlary L, Lum JS, et al. Ubiquitin homeostasis is disrupted in TDP-43 and FUS cell models of ALS. *iScience*. 2020;23:101700.
42. Bendotti C, Marino M, Cheroni C, et al. Dysfunction of constitutive and inducible ubiquitin-proteasome system in amyotrophic lateral sclerosis: implication for protein aggregation and immune response. *Prog Neurobiol*. 2012;97:101-126.
43. Vicencio E, Beltrán S, Labrador L, Manque P, Nassif M, Woehlbier U. Implications of selective autophagy dysfunction for ALS pathology. *Cells*. 2020;9:381.
44. Hensel N, Claus P. The actin cytoskeleton in SMA and ALS: how does it contribute to motoneuron degeneration? *Neuroscientist*. 2018;24:54-72.
45. Schmidt ER, Pasterkamp RJ, van den Berg LH. Axon guidance proteins: novel therapeutic targets for ALS? *Prog Neurobiol*. 2009;88:286-301.
46. Starnes T, Broxmeyer HE, Robertson MJ, Hromas R. Cutting edge: IL-17D, a novel member of the IL-17 family, stimulates cytokine production and inhibits hemopoiesis. *J Immunol*. 2002;169:642-646.
47. Appasani K. *Genome-Wide Association Studies: From Polymorphism to Personalized Medicine*. Cambridge University Press; 2016.
48. Bainomugisa CK, Sutherland HG, Parker R, et al. Using monozygotic twins to dissect common genes in posttraumatic stress disorder and migraine. *Front Neurosci*. 2021;15:1-13. doi:10.3389/fnins.2021.678350.
49. Rodríguez-Cortez VC, Del Pino-Molina L, Rodríguez-Ubreva J, et al. Monozygotic twins discordant for common variable immunodeficiency reveal impaired DNA demethylation during naïve-to-memory B-cell transition. *Nat Commun*. 2015;6:7335.
50. Rom DM, McTague JA. Exact critical values for group sequential designs with small sample sizes. *J Biopharm Stat*. 2020;30:752-764.
51. Shah S, McRae AF, Marioni RE, et al. Genetic and environmental exposures constrain epigenetic drift over the human life course. *Genome Res*. 2014;24:1725-1733.
52. Yazar V, Dawson VL, Dawson TM, et al. DNA Methylation signature of aging: Potential impact on the pathogenesis of parkinson's disease. *J Parkinson's Dis*. 2023;13(2):1-20. doi:10.3233/jpd-223517.
53. Foran E, Trotti D. Glutamate transporters and the excitotoxic path to motor neuron degeneration in amyotrophic lateral sclerosis. *Antioxid Redox Signal*. 2009;11:1587-1602.
54. Yuan S, Zhang ZW, Li ZL. Cell death-autophagy loop and glutamate-glutamine cycle in amyotrophic lateral sclerosis. *Front Mol Neurosci*. 2017;10:1-7. doi:10.3389/fnmol.2017.00231.
55. Schiel KA. A beneficial role for elevated extracellular glutamate in amyotrophic lateral sclerosis and cerebral ischemia. *Bioessays*. 2021;43:e2100127.
56. Jin M, Akgün K, Ziemssen T, Kipp M, Günther R, Hermann A. Interleukin-17 and Th17 lymphocytes directly impair motoneuron survival of wildtype and FUS-ALS mutant human iPSCs. *Int J Mol Sci*. 2021;22:8042.
57. Fu J, Huang Y, Bao T, Liu C, Liu X, Chen X. The role of Th17 cells/IL-17A in AD, PD, ALS and the strategic therapy targeting on IL-17A. *J Neuroinflammation*. 2022;19:98.
58. Karlsson M, Zhang C, Méar L, et al. A single-cell type transcriptomics map of human tissues. *Sci Adv*. 2021;7:eabh2169.
59. Ozyurt T, Gautam M. Differential epigenetic signature of corticospinal motor neurons in ALS. *Brain Sci*. 2021;11:754.