Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Epigenetic signatures of recurrent suicide attempts with cross-ethnic and cross-tissue validation highlight the NEK4-ITIH1 locus

Jong Bhak

jongbhak@genomics.org

Ulsan National Institute of Science and Technology https://orcid.org/0000-0002-4228-1299

Yoonsung Kwon

Ulsan National Institute of Science and Technology https://orcid.org/0000-0002-7242-2734

Yeonsu Jeon

Ulsan National Institute of Science and Technology https://orcid.org/0000-0003-4560-4142

Sungwon Jeon

Ulsan National Institute of Science and Technology https://orcid.org/0000-0002-2729-9087

Hyojung Ryu

Kyungwhan An

Yeo Jin Kim

Asta Blazyte

Ulsan National Institute of Science and Technology

Byoung-chul Kim

Hyung-Tae Jung

Yong Min Ahn

Seoul national University Hospital https://orcid.org/0000-0002-4458-797X

Jeong Hun Yang

Sang Jin Rhee

Seoul National University College of Medicine https://orcid.org/0000-0001-6221-5953

Laima Ambrozaitytė

Algirdas Utkus

Robertas Strumila

Lina Vencevičienė

Robertas Badaras

Edgaras Dlugauskas

https://orcid.org/0000-0002-0475-5713

Byung-Joo Ham

Korea University

Eun-Seok Shin

Ulsan University Hospital https://orcid.org/0000-0002-9169-6968

Article

Keywords:

Posted Date: September 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10605383/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Yes YJ, SJ, HR, YJK, and BCK were employed by Clinomics Inc. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be constructed as a potential conflict of interest.

Abstract

Suicide is a major public health concern, yet objective biological markers for risk assessment remain lacking. Prior blood-based epigenomic studies have largely returned immune-related signals of uncertain specificity to suicidality. Here, we performed differential methylation analysis on whole blood from 69 Korean suicide attempters and 207 matched controls, weighting each sample by the number of prior attempts to prioritize epigenetic changes associated with recurrent suicidal behavior. Combined with iterative random sampling to retain only markers reproducible across sample compositions, this approach identified 365 CpG sites enriched in neurobiological pathways such as dopaminergic synapses and dendritic spines rather than inflammatory processes. Consistent with the underlying hypothesis, 28 of these sites showed methylation levels that scaled with attempt frequency, mapping to genes governing circadian regulation (*CLOCK*), synaptic remodeling (*MMP9*), and serine biosynthesis (*PHGDH*). The marker panel achieved consistent classification (AUROC 0.733 to 0.760) across three independent cohorts of Korean blood, UK postmortem brain, and Lithuanian blood. The *NEK4-ITIH1* locus ranked as the top marker in both blood and brain and showed suicidality-associated methylation changes across all cohorts. This locus lies within 3p21.1, a robustly replicated risk region for bipolar disorder in which genetic variation has been directly associated with suicide attempt, and *NEK4* itself regulates dendritic spine density and circadian emotional behavior. These findings identify *NEK4-ITIH1* locus as a cross-tissue epigenetic marker of suicidality detectable in peripheral blood.

Introduction

Suicide claims more than 700,000 lives annually and ranks among the leading causes of death in young people worldwide (1). Yet current risk assessments rely heavily on clinical interviews and self-report instruments, which suffer from subjectivity and low predictive accuracy (2-5), and meta-analysis research indicate that no clinical factor or combination thereof substantially improves longitudinal prediction (6). This predictive gap is particularly consequential in high-burden settings such as South Korea, which recorded the highest suicide rate among OECD countries in 2021 (1), underscoring the need for objective, biologically informed markers that can complement clinical judgment.

Epigenetic biomarkers, particularly DNA methylation, represent a promising avenue for stratifying suicide risk. Aberrant methylation patterns have been associated with a wide range of psychiatric conditions, including bipolar disorder, major depressive disorder, post-traumatic stress disorder, and autism spectrum disorder (7-10). Critically, childhood stress is associated with methylation differences that remain detectable in adulthood, indicating that epigenetic changes can persist for years to decades as a stable record of past experiences (11, 12). These properties make DNA methylation well-suited for capturing the long-term biological burden of suicidal behavior. However, previous blood-based studies of suicide have predominantly reported immune- and inflammation-related alterations (13-17), which may reflect nonspecific stress responses confounded by infection, smoking, and diet (18), and have lacked cross-ethnic or cross-tissue validation.

Individuals who have attempted suicide are clinically heterogeneous, yet previous epigenomic studies have treated them as a uniform category through binary comparison with controls (13-17). Notably, the number of prior suicide attempts is one of the strongest predictors of future attempts and eventual death by suicide (19-21), suggesting that attempt frequency captures a meaningful dimension of suicidal vulnerability. Because each attempt represents an episode of extreme psychological crisis, the concept of allostatic load predicts that repeated exposure produces cumulative biological consequences, including progressive epigenetic remodeling (22-24). We therefore hypothesized that incorporating attempt frequency as a quantitative weighting factor would preferentially identify CpG sites sensitive to this cumulative process.

Here, we applied this recurrence-weighted approach to whole blood DNA methylation data from Korean suicide attempters and evaluated the risk stratification performance of resulting markers through cross-ethnic and cross-tissue classification in independent cohorts. We report that this approach identifies neurobiologically relevant CpG sites with consistent predictive performance across independent cohorts, and that the *NEK4-ITIH1* locus consistently emerges as a key marker across multiple datasets.

Materials and Methods

Data collection and study population

We enrolled 122 suicide attempters from three hospitals in Korea (Korea University Anam Hospital, Ulsan Medical Center, and Seoul National University Hospital) and used 296 controls from the Korean Genome Project (KGP) (25–27) (Supplementary Table 1). Samples from Korea University Anam Hospital and Ulsan Medical Center focused on fatal suicide attempts and recorded the number of prior attempts; the 69 attempters from these two hospitals therefore constituted the discovery set, while the 53 attempters from Seoul National University Hospital served as the validation set. The KGP controls, comprising individuals without a history of mental illness based on lifestyle and mental health questionnaires, were split between discovery and validation sets in a 7:3 ratio. Age and sex distributions did not differ significantly between the discovery and validation sets (Supplementary Table 1). Written informed consent was obtained from all participants, and all procedures were approved by the institutional review boards of Ulsan National Institute of Science and Technology (UNISTIRB-15-19-A), Korea University Anam Hospital (ED15006), and Ulsan Medical Center (USH.20.013).

Targeted bisulfite sequencing and raw data processing

Blood samples were collected in EDTA tubes and frozen, and genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen, 69506). Sequencing libraries were prepared with the SureSelectXT Methyl-Seq Target Enrichment System Kit (Agilent, G9651) and sequenced on the Illumina NovaSeq 6000 and HiSeq 2500 platforms, generating 150-bp and 100-bp paired-end reads, respectively. Adapters were trimmed with fastp v0.23.1 (28), and bisulfite-treated reads were aligned to the hg38 reference genome using Bismark v0.23.1 with Bowtie2 v2.5.1 (29). Methylation calling at CpG sites was performed with Bismark.

Identifying DNA methylation biomarkers by weighting suicide attempts

Using methylKit (R v4.1.3; methylKit v1.20.0) (30), we retained CpG sites with a depth of at least 10 across all samples, yielding 835,694 sites. Suicidality-associated CpG sites were identified with an in-house script using weighted logistic regression, in which each sample's weight was the product of its sequencing depth and a suicide-attempt-based weight, following the methylKit formulation. Three weighting scenarios were compared: no weighting (weight = 1), linear weighting by the number of attempts (weight = n), and squared weighting (weight = n^2). Log-likelihood ratio tests compared a null model with sex and age as covariates against a model additionally including methylation level, with overdispersion corrected as in methylKit (31). P-values were adjusted by the Benjamini-Hochberg FDR method, and between-group methylation differences were quantified as the weighted mean difference. CpG sites with FDR < 0.05 and an absolute weighted mean difference > 5% were considered significant.

To minimize sampling bias, this procedure was repeated 1,000 times on random 70% subsamples of the discovery set, and CpG sites significant in over 90% of iterations were retained. To assess confounding by blood cell type composition, we performed a sensitivity analysis in which blood cell type proportions estimated by UXM deconvolution (32) (https://github.com/nloyfer/UXM_deconv) were added as covariates, and the differential methylation analysis was repeated under this adjusted model.

Correlation between methylation levels and the number of suicide attempts

For each of the 365 CpG sites, the correlation between methylation level and the number of suicide attempts was assessed using Spearman's correlation coefficient within the discovery cohort. The resulting p-values were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate (FDR) method, and CpG sites with FDR < 0.05 were considered significantly correlated with attempt frequency.

Application of random forest model for validation

Random forest models were trained using scikit-learn (v0.23.2), with hyperparameters tuned by GridSearchCV under 3-fold cross-validation, optimizing max_depth, min_samples_split, min_samples_leaf, and n_estimators. Feature importance was computed using Shapley Additive exPlanations (SHAP) (shap v0.41.0) (33) on the discovery set. For the UK brain tissue and Lithuanian blood datasets, SHAP values were computed on the training set of each fold during 3-fold cross-validation and averaged across folds to determine overall feature-importance rankings.

To evaluate the contribution of each methodological component, all six combinations of weighting strategy (no weight, linear, squared) and random sampling (with and without) were applied to the

discovery set. For each condition, a random forest classifier was trained on the resulting marker set and validated on the independent Korean cohort. Pairwise comparisons of AUROC values were performed using DeLong's test with the fast DeLong algorithm (34), using code adapted from https://github.com/yandexdataschool/roc_comparison. All analyses were conducted in Python 3.7.16.

Cross-validation of CpG sites with external datasets

Brain tissue dataset

To validate the discovered CpG sites, we used a normalized brain methylation dataset from the Gene Expression Omnibus (GSE88890), comprising methylation array data from Brodmann area 11 (BA11) tissue of depressed suicide decedents (n = 20) and non-psychiatric, sudden-death controls (n = 20). Of the discovered CpG sites, 46 matched target positions on the Illumina methylation array and were retained for validation. Suicide decedents and controls were randomly divided into three groups for 3-fold cross-validation using a random forest model based on these 46 CpG sites, with model training and validation conducted as described above.

Lithuanian dataset

To assess generalizability across ethnic populations, we recruited 47 Lithuanian suicide attempters and 45 controls. Attempters were individuals admitted to the Toxicology Intensive Care Unit of Republican Vilnius University Hospital within 48 h of a suicide attempt, regardless of attempt type or clinical diagnosis. Controls were individuals without a psychiatric diagnosis, suicide attempt history, or significant suicidal ideation or depressive symptomatology, recruited during prophylactic check-ups at the Center of Family Medicine of Vilnius University Hospital Santaros Klinikos. The study was approved by the Vilnius Regional Biomedical Research Ethics Committee of Lithuania (No. 2020/9-1266-747). DNA extraction, library preparation, sequencing (Illumina NovaSeq 6000, 150-bp paired-end reads), and data preprocessing followed the same protocols as the Korean blood samples.

Of the discovered CpG sites, 315 without missing values in the Lithuanian data were used for validation. As with the brain tissue dataset, the cohort was randomly partitioned into three groups for 3-fold cross-validation using a random forest model, following the procedures described above. Because chr3:52773754 lies in the *NEK4-ITIH1* intergenic region, we extended the search scope to the *NEK4* and *ITIH1* gene bodies and assessed differential methylation at individual CpG sites within this region using the Wilcoxon rank-sum test.

Gene annotation and enrichment analysis

We annotated each CpG site using Annotatr (35) with the options "hg38_cpgs", "hg38_basicgenes", and "hg38_enhancers_fantom" to identify genes associated with interested CpG sites. Utilizing ShinyGO (36),

we conducted gene enrichment analysis in Gene Ontology (GO, http://geneontology.org/) (37, 38) and Kyoto Encyclopedia of Genes and Genomes (KEGG, https://www.genome.jp/kegg/) (39). The p-values were adjusted using the false discovery rate (FDR) method. Terms with FDR < 0.05 were considered statistically significant, and even in cases where FDR was not substantial, terms with high fold enrichment were also considered.

Results

Recurrence-weighted analysis identifies 365 suicidality-associated CpG sites

To identify CpG sites associated with suicidal behavior, we performed differential methylation analysis on whole blood from 69 Korean suicide attempters and 207 matched controls (discovery set), incorporating the number of suicide attempts as a sample-level weighting factor (see Methods) (Supplementary Fig. 1). To ensure marker robustness against sample composition effects in this small cohort, we implemented an iterative random sampling procedure of 1,000 subsampled analyses, retaining only CpG sites that reached significance (FDR < 0.05, weighted mean difference > 5%) in at least 90% of iterations. Combining squared recurrence weighting with this procedure yielded 365 consistently identified CpG sites (Fig. 1C, F) (Supplementary Table 2). To validate this design, we trained random forest classifiers on marker sets derived under different conditions and tested them on an independent Korean validation cohort from a separate hospital (Fig. 1G).

Recurrence weighting was a prerequisite for marker discovery within this framework. Without weighting, only two CpG sites reached significance in the full discovery set, and none survived random sampling (Fig. 1A, D), indicating that conventional unweighted analysis lacks the sensitivity to detect suicidality-associated methylation differences in a cohort of this size.

The impact of random sampling was most pronounced under squared weighting, where applying the 90% consistency threshold reduced the marker set from 9,954 to 365 CpG sites, yet this reduction was accompanied by a substantial gain in validation performance (AUROC from 0.621 to 0.760; DeLong's test, P = 0.0064) (Fig. 1C, F, G). The unsampled set, despite containing over 25 times more markers, performed worse than any other condition, confirming that the expanded signal from aggressive weighting contains considerable noise that the sampling procedure effectively removes.

Weighting intensity also affected performance. After sampling, the squared-weighted set achieved significantly higher AUROC than the linear-weighted set (0.760 vs. 0.654; DeLong's test, P = 0.0169) (Fig. 1G), while capturing a broader signal (365 vs. 14 CpG sites) (Fig. 1E, F). The 14 linear-weighted sites were entirely contained within the 365, indicating that squared weighting expands the signal around a shared core rather than identifying different markers.

We further confirmed that the identified markers were not driven by blood cell type composition (Supplementary Fig. 2). When estimated cell type proportions were included as covariates, significance levels were generally attenuated across the genome (Supplementary Fig. 2A). However, all 365 CpG sites retained significance (FDR < 0.05) under this more conservative model (Supplementary Fig. 2B), indicating that the iterative sampling procedure inherently selects markers robust to cellular heterogeneity.

Together, these results establish that squared weighting and iterative sampling operate synergistically, with the former amplifying recurrence-associated biological signals and the latter ensuring their reproducibility, yielding a robust set of 365 CpG sites for downstream analysis.

Neurobiological pathway enrichment and recurrence-associated methylation changes support biological relevance

To assess whether the 365 CpG sites reflect biologically coherent processes, we performed pathway enrichment analysis on the 390 genes annotated to these sites (Supplementary Table 3). KEGG analysis revealed significant enrichment in neurobiological pathways, including inflammatory mediator regulation of TRP channels and dopaminergic synapse (Fig. 2A). Long-term depression, a signaling cascade that persistently decreases synaptic strength primarily in the cerebellum, showed a notably high fold enrichment (fold enrichment of 6.84; FDR = 0.139) but did not reach statistical significance after multiple testing correction. Gene Ontology cellular component analysis identified significant enrichment in neuronal structures, including voltage-gated potassium channel complexes, dendritic spines, neuron spines, dendrites, and synapses (Fig. 2B), while biological process terms further supported involvement in synaptic function, including regulation of synaptic plasticity and trans-synaptic signaling (Supplementary Fig. 3). These findings align with epigenetic modifications previously reported in the brain tissue of suicide decedents (40, 41), and contrast with the predominantly immune- and inflammation-related pathways identified in prior blood-based studies (13–17). Using a recurrence-weighted approach, the present analysis identified neurobiologically relevant epigenetic changes in peripheral blood that more closely resemble those observed in postmortem brain tissue, suggesting that incorporating attempt frequency as a weighting factor may improve the biological specificity of blood-based marker discovery.

As hypothesized, a subset of the 365 CpG sites exhibited methylation changes that scaled with the number of suicide attempts. Specifically, 28 sites (7.67%) showed a significant correlation between methylation level and attempt frequency (Spearman's correlation, FDR < 0.05) (Fig. 2C) (Supplementary Table 4). The direction of all 28 correlations was concordant with the direction of differential methylation between attempters and controls, indicating that methylation at these sites progressively diverges from the control profile as suicide attempts accumulate.

Among the CpG sites showing the strongest correlations (|ρ| > 0.4), several mapped to genes with established roles in brain function and psychiatric disorders. *CLOCK*, a circadian clock gene, has been directly associated with suicide attempts in bipolar disorder (42). *MMP9* mediates dendritic spine remodeling, and its elevated levels are linked to suicidal behavior and bipolar disorder (43, 44). *PHGDH*, a rate-limiting enzyme for brain L-serine biosynthesis, is implicated in schizophrenia and Alzheimer's disease (45, 46). The CpG sites annotated to *ZNF767P*, which harbored the highest positive correlations, are located within a genomic region linked by regulatory annotation to *ZNF746*, a gene implicated in dopamine neuron degeneration (47, 48).

The convergence of pathway enrichment and recurrence-associated methylation changes at brain-relevant genes provides orthogonal evidence that the 365 CpG sites capture neurobiologically grounded aspects of suicidal behavior rather than nonspecific inflammatory or stress-related signals. The recurrence-correlated subset suggests that a subset of these CpG sites may serve as molecular indicators of the cumulative biological impact of repeated suicidal episodes.

Cross-cohort validation and convergence at the NEK4-ITIH1 locus

To evaluate whether the 365 CpG sites generalize beyond the discovery population, we tested their classification performance in three independent cohorts spanning different ethnicities, tissue types, and profiling platforms (Fig. 3). In the independent Korean blood validation cohort, the random forest model trained on the discovery set achieved an AUROC of 0.760 (Fig. 3A). For the UK postmortem brain tissue dataset (GSE88890, Brodmann area 11), comparing individuals who died by suicide with sudden-death controls, 3-fold cross-validation using the 46 CpG sites available on the array platform yielded a mean AUROC of 0.733 (Fig. 3B). In the Lithuanian blood cohort, 3-fold cross-validation using the 315 available CpG sites produced a mean AUROC of 0.756 (Fig. 3C).

This consistency across cohorts differing in sample size, tissue type, ethnic background, and available CpG sites argues that the observed signals reflect genuine biological associations rather than overfitting. Notably, the reduced subset of 46 CpG sites sufficed to maintain comparable performance in brain tissue, suggesting that the core discriminative signal is carried by a subset of consistently informative sites.

To identify which markers drove this performance, we examined SHAP-based feature importance in each cohort (Fig. 4A-C). A single CpG site at the *NEK4-ITIH1* intergenic region (chr3:52773754) ranked as the most important feature in both the Korean blood validation cohort and the UK brain tissue dataset (Fig. 4A, B). Classification using this site alone yielded AUROCs of 0.676 in Korean blood and 0.838 in UK brain (Supplementary Fig. 4), indicating that a substantial proportion of the model's discriminative capacity is attributable to this single position. This CpG site also showed increased significance after cell-type adjustment, emerging as the most significant site in the adjusted model (Supplementary Fig. 2B), supporting its independence from cellular heterogeneity.

In the Lithuanian cohort, chr3:52773754 was not a top feature and showed no significant case-control difference (Fig. 4C, Supplementary Fig. 5H). Instead, three adjacent CpG sites within the *ITIH1* gene body, approximately 5.7 kb downstream, were significantly hypomethylated in attempters (Fig. 4D). The signal was therefore complementary across cohorts. chr3:52773754 was significant in Korean blood and UK brain, whereas the downstream *ITIH1* region was significant only in Lithuanian blood, and this region was not assessable in UK brain due to array coverage (Supplementary Fig. 5). Although the precise CpG position carrying the strongest signal varies across populations, the broader *NEK4-ITIH1* locus consistently harbors suicidality-associated alterations.

Despite its strong discriminative power, chr3:52773754 did not show a significant correlation with the number of suicide attempts in the discovery cohort (Supplementary Fig. 6). Methylation levels at this site were consistently elevated in suicide attempters compared to controls, regardless of attempt frequency, suggesting that the epigenetic alteration at this locus may represent an early or threshold-dependent change associated with suicidality rather than a cumulative modification that scales with repeated attempts.

Taken together, these results identify the *NEK4-ITIH1* locus as a convergent epigenetic marker of suicidality, with cross-tissue and cross-ethnic evidence pointing to this genomic region as a central node in the epigenetic architecture of suicidal behavior. The partially overlapping top 10 feature importance profiles across cohorts, including shared markers such as *PHGDH*, *DNAJC7*/*NKIRAS2*, and *SH3BP5*, further support the broader generalizability of the 365-marker panel, while the *NEK4-ITIH1* locus stands out as the most consistently prominent signal across independent datasets.

Discussion

The most notable finding of this study is the convergence of epigenetic signals at the *NEK4-ITIH1* locus across tissues and populations. A single intergenic CpG site (chr3:52773754) ranked as the most important feature in both the Korean blood and UK brain tissue datasets, demonstrating that this locus carries substantial discriminative information across both tissue types. In the Lithuanian cohort, this specific CpG site did not replicate, but three CpG sites within the *ITIH1* gene body, approximately 5.7 kb downstream, showed significant differential methylation between suicide attempters and controls, though the direction of the difference differed. This directional discordance may reflect population-specific genetic differences such as meQTL structures or allele frequencies. Unfortunately, as the downstream *ITIH1* gene body region could not be assessed in the UK brain tissue dataset due to array platform limitations, this study could not validate it in the European brain dataset. Nevertheless, the co-occurrence of significant suicidality-associated methylation alterations within the same narrow genomic region across three independent cohorts is unlikely to be coincidental.

This epigenetic convergence is reinforced by extensive genetic evidence implicating 3p21.1 in psychiatric disease. This locus is among the most robustly replicated GWAS risk regions for bipolar disorder and schizophrenia (49–53), with eQTL-integrated analyses identifying *NEK4* as a primary risk

driver (54) and the strongest nonsynonymous signal, rs1042779, mapping to *ITIH1* (55). Genetic variation in this region has also been directly associated with suicide attempt (56). Our study adds an epigenetic dimension, showing that both sequence-level variation and methylation state at this locus are associated with suicidality.

Because chr3:52773754 lies in the intergenic region between *NEK4* and *ITIH1*, we cannot determine from methylation data alone which gene or genes it regulates. However, the functional literature on *NEK4* provides a compelling mechanistic framework for interpreting this signal. *NEK4* overexpression in the mouse hippocampus reduces the density of mushroom-type dendritic spines (57). Also, conditional overexpression of *NEK4* in forebrain pyramidal neurons disrupts circadian fluctuations of emotional behaviors and synaptic structure (58), a phenotype of relevance given the well-documented circadian patterns of suicidal behavior. Both the behavioral and synaptic effects of *NEK4* overexpression were ameliorated by lithium (58), one of the few pharmacological agents with established anti-suicidal efficacy. Cell-type specificity analyses have further revealed that *NEK4* is expressed on the surface of glutamatergic neurons (54), placing it within the excitatory neurotransmission system targeted by emerging anti-suicidal interventions such as ketamine (59).

These functional data position *NEK4* at the intersection of three neurobiological pathways that are established as relevant to suicidality: dendritic spine dynamics, circadian regulation of emotion, and glutamatergic neurotransmission. Two of these pathways independently converge with findings from the recurrence-correlated CpG analysis in the present study, where CpG sites near *CLOCK* (a core circadian gene associated with suicide attempts in bipolar disorder) and *MMP9* (a mediator of dendritic spine remodeling linked to suicidal behavior) both showed significant recurrence-associated methylation changes. This convergence across independently identified markers strengthens the hypothesis that disruption of synaptic plasticity and circadian regulatory networks represents a core epigenetic signature of recurrent suicidal behavior. While the mechanistic role of *ITIH1* in the brain is less well characterized, the localization of the Lithuanian signal within the *ITIH1* gene body, together with the direct genetic association between *ITIH1* coding variants and bipolar disorder (55), suggests that epigenetic variation here may affect multiple genes within the 3p21.1 region. Expression studies will be necessary to identify the regulatory targets of the identified CpG sites.

The convergence of the *NEK4-ITIH1* signal in both peripheral blood and postmortem brain tissue suggests that the epigenetic alteration at this locus is not brain-restricted. While the mechanisms underlying this cross-tissue concordance remain unclear, the practical implication is that the *NEK4-ITIH1* locus may serve as an accessible, blood-based surrogate for brain-specific changes associated with suicidality, supporting the feasibility of minimally invasive risk monitoring.

Notably, chr3:52773754 showed consistently elevated methylation in attempters regardless of attempt frequency, without a significant correlation with the number of attempts. This pattern may indicate an early or threshold-dependent alteration that distinguishes individuals with any history of suicidal

behavior from controls, rather than a cumulative change, although limited statistical power cannot be excluded. Distinguishing these possibilities will require larger longitudinal studies.

Several major limitations of this study should be acknowledged. First, the sample size is small, with fewer than 70 individuals in each blood-based cohort and only 20 postmortem brain samples. Although the consistency of classification performance (AUROC > 0.73) across three independent cohorts argues against simple overfitting, the high single-marker performance in the brain dataset (AUROC = 0.838 for chr3:52773754 alone) should be interpreted with caution given this limited size. Second, the cross-sectional design precludes determining whether the observed methylation changes are causes, consequences, or correlates of suicidal behavior, which will require prospective studies tracking methylation before and after suicide attempts. Third, while recurrence was a central hypothesis of this study, attempt frequency data were unavailable for the UK brain and Lithuanian cohorts. The consistent classification performance demonstrates that the selected CpG sites carry discriminative information for suicidality in general, but it remains open whether the specific methylation-frequency relationship generalizes beyond the Korean cohort. Finally, the directional and positional differences in the *NEK4-ITIH1* signal across populations remain incompletely understood, and together with the known influence of ancestry on methylation (60, 61), indicate that translational application of this locus will require validation in larger, ancestrally diverse cohorts before clinical utility can be established.

This study demonstrates that recurrence-weighted epigenomic analysis can identify blood-based methylation markers with neurobiological relevance that maintain predictive utility across independent cohorts, tissue types, and ethnic backgrounds. The consistent identification of the *NEK4-ITIH1* locus as the top-ranked signal in both blood and brain provides a rare example of cross-tissue convergence in suicidal behavior and positions this locus as a candidate for mechanistic investigation. Clinically, the ability to detect neurobiologically grounded epigenetic signatures of suicidality in peripheral blood offers a pathway toward objective, accessible biomarkers that, if validated in larger prospective studies, could complement existing risk assessment and inform long-term monitoring.

Declarations

Conflict of Interest

YJ, SJ, HR, YJK, and BCK were employed by Clinomics Inc. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be constructed as a potential conflict of interest.

Acknowledgments

We thank all the genome-donating participants. The biospecimens for this study were provided by Korea University Anam Hospital, Ulsan Medical Center, Seoul National University Hospital, and the members of

the National Biobank of Korea, which is supported by the Ministry of Health, Welfare and Family Affairs. This study also included data from 696 participants provided by the Ulsan University Hospital Biobank, including DNA samples and clinical information (60SA2016001-010-5). We thank the Korea Institute of Science and Technology Information (KISTI), which provided us with the Korea Research Environment Open NETwork (KREONET). We also appreciate the Ulsan ICT Promotion Agency (UIPA), which provided us with the BioDataFarm system, which supports the storage, analysis, and management of the BioBigData. AI-assisted tools were used for English language editing and proofreading. The authors reviewed and take full responsibility for the content of the manuscript

This work was supported by the Promotion of Innovative Business for Regulation-Free Special Zones funded by the Ministry of SMEs and Startups (MSS, Korea) (grant number [P0016195, P0016193] (1425156792, 1425157301) (2.220035.01, 2.220036.01)). This work was also supported by the Establishment of Demonstration Infrastructure for Regulation-Free Special Zones fund (MSS, Korea) (grant number [P0016191] (2.220037.01) (1425157253)) by the Ministry of SMEs and Startups. This work was also supported by the Development of prediction system for depression and stress states based on machine learning using multi-omics data funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea) (1415170577). This work was also supported by the Civil-Military Dual-Use Technology Development Program (14-BR-SS-03) through the Agency for Defense Development. This work was also supported by the U-K BRAND Research Fund [1.200108.01] of UNIST (Ulsan National Institute of Science and Technology). This work was also supported by the Research Project Funded by the Ulsan City Research Fund [1.200047.01] of UNIST(Ulsan National Institute of Science & Technology). This work was also supported by the Korea Evaluation Institute of Industrial Technology (KEIT) with funding from the Ministry of Trade, Industry and Energy (1415187694, Materials and Parts Technology New Development Project (Heterogeneous Technology Convergence Type)).

Data and materials availability

All data, along with the code used for analysis and visualization, are available in the paper, the Supplementary Materials, and/or on a GitHub repository (https://github.com/korean-genomics-center/DNA-Methylation-Weighted-Suicide-Biomarkers). The datasets supporting the conclusions of this article are available on the Korean Genome Project website (https://koreangenome.org). All other relevant data related to this study are available from the corresponding author upon request.

References

1. World Health O. Suicide rates2021 (Accessed on 27 May 2025)]. Available from: https://www.who.int/data/gho/data/themes/mental-health/suicide-rates.
2. Beck AT, Brown GK, Steer RA. Psychometric characteristics of the Scale for Suicide Ideation with psychiatric outpatients. Behaviour Research and Therapy. 1997;35(11):1039-46.
3. Deming CA, Harris JA, Castro-Ramirez F, Glenn JJ, Cha CB, Millner AJ, et al. Inconsistencies in self-reports of suicidal ideation and attempts across assessment methods. Psychological Assessment.

2021;33(3):218-29.

4. Millner AJ, Lee MD, Nock MK. Single-Item Measurement of Suicidal Behaviors: Validity and Consequences of Misclassification. PLOS ONE. 2015;10(10):e0141606.
5. Cochrane-Brink KA, Lofchy JS, Sakinofsky I. Clinical rating scales in suicide risk assessment11Information about the rating scales can be obtained from the author. General Hospital Psychiatry. 2000;22(6):445-51.
6. Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, et al. Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. Psychol Bull. 2017;143(2):187-232.
7. Mirza S, Lima CNC, Del Favero-Campbell A, Rubinstein A, Topolski N, Cabrera-Mendoza B, et al. Blood epigenome-wide association studies of suicide attempt in adults with bipolar disorder. Translational Psychiatry. 2024;14(1):70.
8. Aberg KA, Dean B, Shabalin AA, Chan RF, Han LKM, Zhao M, et al. Methylome-wide association findings for major depressive disorder overlap in blood and brain and replicate in independent brain samples. Molecular Psychiatry. 2020;25(6):1344-54.
9. Katrinli S, Wani AH, Maihofer AX, Ratanatharathorn A, Daskalakis NP, Montalvo-Ortiz J, et al. Epigenome-wide association studies identify novel DNA methylation sites associated with PTSD: a meta-analysis of 23 military and civilian cohorts. Genome Medicine. 2024;16(1):147.
10. Andrews SV, Sheppard B, Windham GC, Schieve LA, Schendel DE, Croen LA, et al. Case-control meta-analysis of blood DNA methylation and autism spectrum disorder. Molecular Autism. 2018;9(1):40.
11. Peng H, Zhu Y, Strachan E, Fowler E, Bacus T, Roy-Byrne P, et al. Childhood Trauma, DNA Methylation of Stress-Related Genes, and Depression: Findings From Two Monozygotic Twin Studies. Biopsychosocial Science and Medicine. 2018;80(7):599-608.
12. McGowan PO, Sasaki A, D'Alessio AC, Dymov S, Labonté B, Szyf M, et al. Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. Nature Neuroscience. 2009;12(3):342-8.
13. Clive ML, Boks MP, Vinkers CH, Osborne LM, Payne JL, Ressler KJ, et al. Discovery and replication of a peripheral tissue DNA methylation biosignature to augment a suicide prediction model. Clinical Epigenetics. 2016;8(1):113.
14. Mirza S, Lima CNC, Del Favero-Campbell A, Rubinstein A, Topolski N, Cabrera-Mendoza B, et al. Blood epigenome-wide association studies of suicide attempt in adults with bipolar disorder. Transl Psychiatry. 2024;14(1):70.
15. Brundin L, Bryleva EY, Thirtamara Rajamani K. Role of Inflammation in Suicide: From Mechanisms to Treatment. Neuropsychopharmacology. 2017;42(1):271-83.
16. Daray FM, Chiapella LC, Grendas LN, Casiani RIÁ, Olaviaga A, Robetto J, et al. Peripheral blood cellular immunophenotype in suicidal ideation, suicide attempt, and suicide: a systematic review and meta-analysis. Molecular Psychiatry. 2024;29(12):3874-92.

17. Neupane SP, Daray FM, Ballard ED, Galfalvy H, Itzhaky L, Segev A, et al. Immune-related biomarkers and suicidal behaviors: A meta-analysis. European Neuropsychopharmacology. 2023;75:15-30.
18. Yuan N, Chen Y, Xia Y, Dai J, Liu C. Inflammation-related biomarkers in major psychiatric disorders: a cross-disorder assessment of reproducibility and specificity in 43 meta-analyses. Translational Psychiatry. 2019;9(1):233.
19. Fushimi M, Sugawara J, Saito S. Comparison of completed and attempted suicide in Akita, Japan. Psychiatry and Clinical Neurosciences. 2006;60(3):289-95.
20. Lim M, Lee SU, Park J-I. Difference in suicide methods used between suicide attempters and suicide completers. International Journal of Mental Health Systems. 2014;8(1):54.
21. DeJong TM, Overholser JC, Stockmeier CA. Apples to oranges?: A direct comparison between suicide attempters and suicide completers. Journal of Affective Disorders. 2010;124(1):90-7.
22. Danese A, McEwen BS. Adverse childhood experiences, allostasis, allostatic load, and age-related disease. Physiology & Behavior. 2012;106(1):29-39.
23. O'Neill KA, van der Veer BK, Charvet L, Azmy N, Friedman S, Hu J, et al. Childhood adversity, allostatic load and epigenetic signatures in paediatric and adult-onset multiple sclerosis. Brain Commun. 2026;8(1):fcaf512.
24. Chamberlain JD, Ackermann D, Bochud M, Booth T, Chapatte L, Corley J, et al. Development and validation of an epigenetic signature of allostatic load. Biosci Rep. 2025;45(4):247-62.
25. Jeon S, Bhak Y, Choi Y, Jeon Y, Kim S, Jang J, et al. Korean Genome Project: 1094 Korean personal genomes with clinical information. Science advances. 2020;6(22):eaaz7835.
26. Jeon Y, Jeon S, Blazyte A, Kim YJ, Lee JJ, Bhak Y, et al. Welfare genome project: a participatory Korean personal genome project with free health check-up and genetic report followed by counseling. Frontiers in genetics. 2021;12:633731.
27. Jeon S, Choi H, Jeon Y, Choi W-H, Choi H, An K, et al. Korea4K: whole genome sequences of 4,157 Koreans with 107 phenotypes derived from extensive health check-ups. Gigascience. 2024;13:giae014.
28. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884-i90.
29. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27(11):1571-2.
30. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biology. 2012;13(10):R87.
31. McCullagh P. Generalized linear models: Routledge; 2019.
32. Loyfer N, Magenheim J, Peretz A, Cann G, Bredno J, Klochendler A, et al. A DNA methylation atlas of normal human cell types. Nature. 2023;613(7943):355-64.

33. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence. 2020;2(1):56-67.
34. Sun X, Xu W. Fast Implementation of DeLong's Algorithm for Comparing the Areas Under Correlated Receiver Operating Characteristic Curves. IEEE Signal Processing Letters. 2014;21(11):1389-93.
35. Cavalcante RG, Sartor MA. annotatr: genomic regions in context. Bioinformatics. 2017;33(15):2381-3.
36. Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics. 2019;36(8):2628-9.
37. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25-9.
38. Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, et al. The gene ontology knowledgebase in 2023. Genetics. 2023;224(1):iyad031.
39. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30.
40. Policicchio S, Washer S, Viana J, Iatrou A, Burrage J, Hannon E, et al. Genome-wide DNA methylation meta-analysis in the brains of suicide completers. Translational Psychiatry. 2020;10(1):69.
41. Kouter K, Zupanc T, Videtič Paska A. Genome-wide DNA methylation in suicide victims revealing impact on gene expression. Journal of Affective Disorders. 2019;253:419-25.
42. Benedetti F, Riccaboni R, Dallaspezia S, Locatelli C, Smeraldi E, Colombo C. Effects of CLOCK gene variants and early stress on hopelessness and suicide in bipolar depression. Chronobiol Int. 2015;32(8):1156-61.
43. Dickerson F, Katsafanas E, Origoni A, Rowe K, Khan S, Mukhtar F, et al. Levels of matrix metalloproteinase 9 (MMP-9) are elevated in persons with major depressive disorder who have had a recent suicide attempt. Brain, Behavior, & Immunity - Health. 2026;51:101163.
44. Dickerson F, Vaidya D, Liu Y, Yolken R. Levels of Matrix Metalloproteinase 9 Are Elevated in Persons With Schizophrenia or Bipolar Disorder: The Role of Modifiable Factors. Biological Psychiatry Global Open Science. 2023;3(4):766-72.
45. Chen X, Calandrelli R, Girardini J, Yan Z, Tan Z, Xu X, et al. PHGDH expression increases with progression of Alzheimer's disease pathology and symptoms. Cell Metab. 2022;34(5):651-3.
46. Yang JH, Wada A, Yoshida K, Miyoshi Y, Sayano T, Esaki K, et al. Brain-specific Phgdh deletion reveals a pivotal role for L-serine biosynthesis in controlling the level of D-serine, an N-methyl-D-aspartate receptor co-agonist, in adult brain. J Biol Chem. 2010;285(53):41380-90.
47. Shin J-H, Ko Han S, Kang H, Lee Y, Lee Y-I, Pletinkova O, et al. PARIS (ZNF746) Repression of PGC-1α Contributes to Neurodegeneration in Parkinson's Disease. Cell. 2011;144(5):689-702.
48. Alieva AK, Filatova EV, Karabanov AV, Illarioshkin SN, Slominsky PA, Shadrina MI. Potential Biomarkers of the Earliest Clinical Stages of Parkinson's Disease. Parkinson's Disease. 2015;2015(1):294396.

49. Brandl EJ, Lett TA, Chowdhury NI, Tiwari AK, Bakanidze G, Meltzer HY, et al. The role of the ITIH3 rs2535629 variant in antipsychotic response. Schizophrenia Research. 2016;176(2):131-5.
50. Gadd DA, Hillary RF, McCartney DL, Shi L, Stolicyn A, Robertson NA, et al. Integrated methylome and phenome study of the circulating proteome reveals markers pertinent to brain health. Nature Communications. 2022;13(1):4670.
51. Szczepankiewicz A. Evidence for single nucleotide polymorphisms and their association with bipolar disorder. Neuropsychiatric Disease and Treatment. 2013;9(null):1573-82.
52. Scott LJ, Muglia P, Kong XQ, Guan W, Flickinger M, Upmanyu R, et al. Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. Proc Natl Acad Sci U S A. 2009;106(18):7501-6.
53. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nature Genetics. 2021;53(6):817-29.
54. Zhang C, Yang Z, Li X, Zhao L, Guo W, Deng W, et al. Unraveling NEK4 as a Potential Drug Target in Schizophrenia and Bipolar I Disorder: A Proteomic and Genomic Approach. Schizophrenia Bulletin. 2024;50(5):1185-96.
55. Scott LJ, Muglia P, Kong XQ, Guan W, Flickinger M, Upmanyu R, et al. Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. Proceedings of the National Academy of Sciences. 2009;106(18):7501-6.
56. Finseth PI, Sønderby IE, Djurovic S, Agartz I, Malt UF, Melle I, et al. Association analysis between suicidal behaviour and candidate genes of bipolar disorder and schizophrenia. Journal of Affective Disorders. 2014;163:110-4.
57. Yang Z, Zhou D, Li H, Cai X, Liu W, Wang L, et al. The genome-wide risk alleles for psychiatric disorders at 3p21.1 show convergent effects on mRNA expression, cognitive function, and mushroom dendritic spine. Molecular Psychiatry. 2020;25(1):48-66.
58. Yang Z-H, Cai X, Zhang C-Y, Zhang Q, Li M, Ding Z-L, et al. NEK4 modulates circadian fluctuations of emotional behaviors and synaptogenesis in male mice. Nature Communications. 2024;15(1):9180.
59. Larkin GL, Beautrais AL. A preliminary naturalistic study of low-dose ketamine for depression and suicide ideation in the emergency department. International Journal of Neuropsychopharmacology. 2011;14(8):1127-31.
60. Elliott HR, Burrows K, Min JL, Tillin T, Mason D, Wright J, et al. Characterisation of ethnic differences in DNA methylation between UK-resident South Asians and Europeans. Clinical Epigenetics. 2022;14(1):130.
61. Galanter JM, Gignoux CR, Oh SS, Torgerson D, Pino-Yanes M, Thakur N, et al. Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. Elife. 2017;6.

Figures

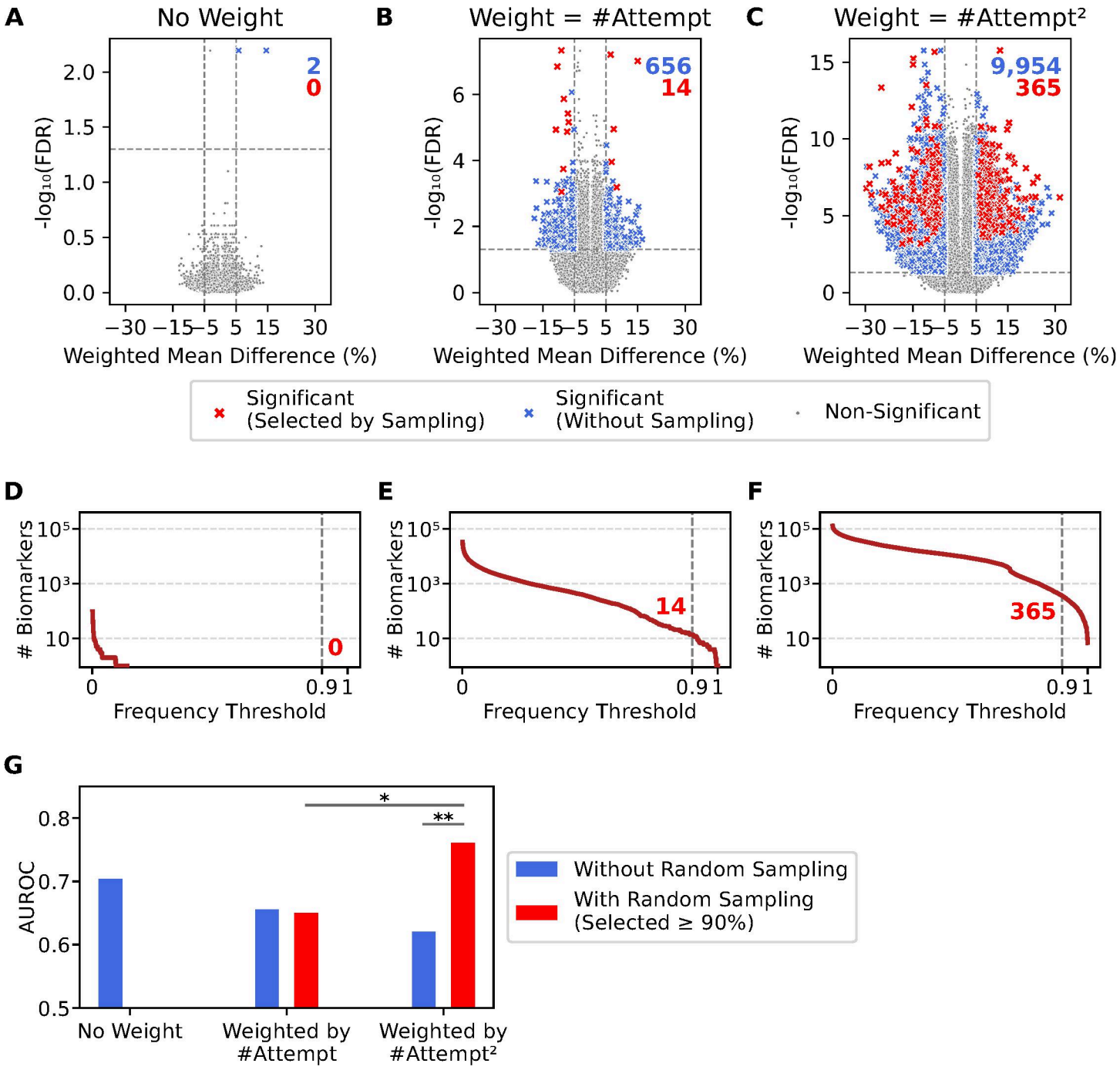


Figure 1

Recurrence weighting and iterative random sampling jointly optimize suicidality-associated CpG site discovery. (A to C) Volcano plots of differential methylation analysis under three weighting conditions: no weight (w = 1) (A), linear weight by attempt frequency (w = n) (B), and squared weight (w = n^2) (C). Blue points indicate CpG sites reaching significance (FDR < 0.05, weighted mean difference > 5%) in the full discovery set without random sampling; red points indicate those consistently selected in at least 90% of 1,000 random sampling iterations. Numbers denote the count of significant CpG sites for each condition. (D to F) Number of CpG sites selected as a function of the consistency threshold across 1,000 random sampling iterations, corresponding to the three weighting conditions in (A) to (C). Vertical

dashed lines mark the 90% threshold used for final marker selection; the number of CpG sites meeting this threshold is indicated. (G) Classification performance (AUROC) of random forest models trained on marker sets derived from each condition and validated on an independent Korean cohort. Blue bars represent models using all significant CpG sites from the full discovery set (without random sampling); red bars represent models using CpG sites passing the 90% consistency threshold (with random sampling). No red bar is shown for the no-weight condition, as zero CpG sites met the threshold. Asterisks denote significant differences by DeLong's test (*: 0.01≤p<0.05; **: 0.001≤p<0.01).

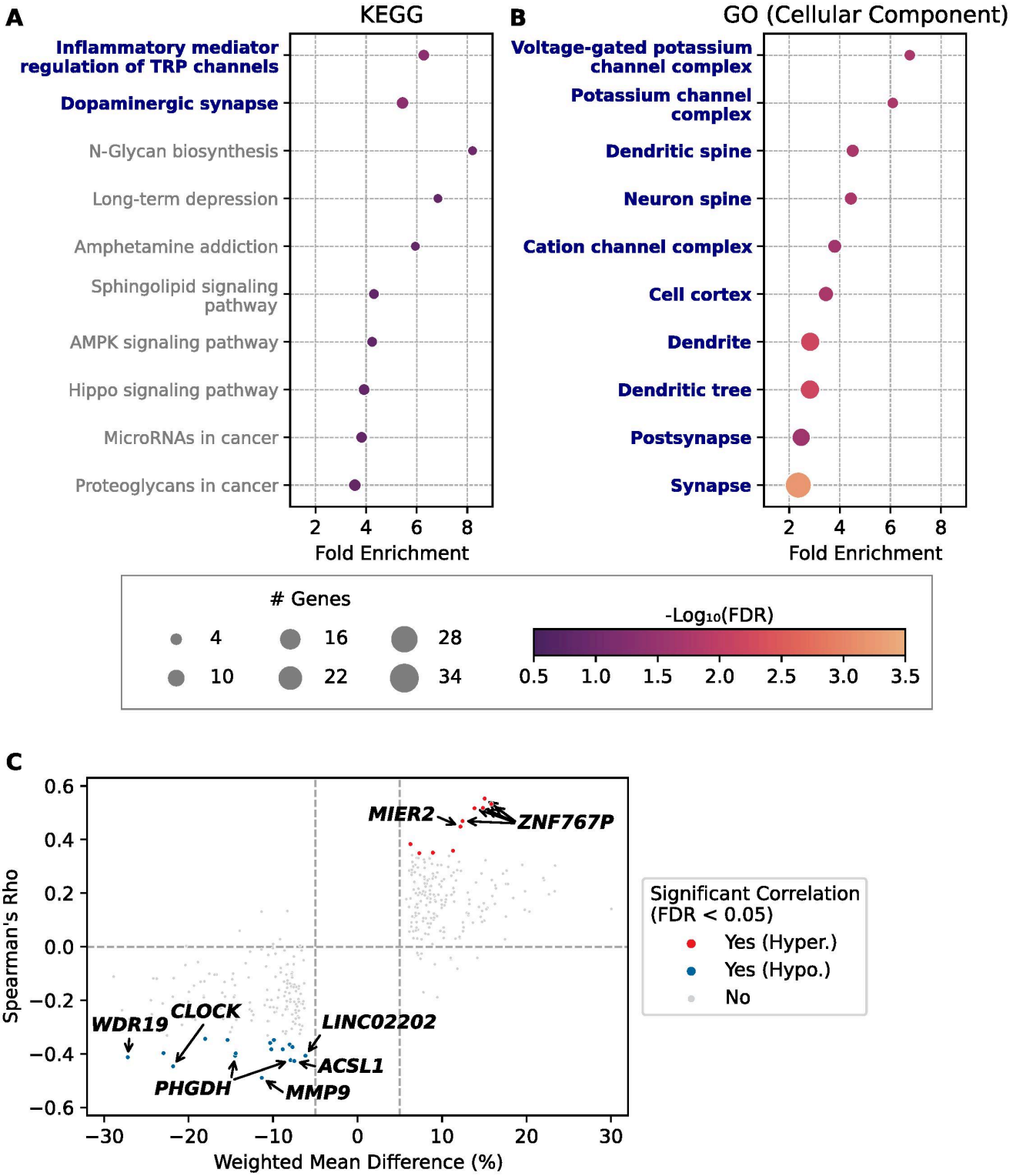


Figure 2

Pathway enrichment and recurrence-associated methylation changes among the 365 suicidality-associated CpG sites. (A) KEGG pathway enrichment analysis of 390 genes annotated to the 365 CpG sites. (B) Gene Ontology cellular component enrichment analysis. In both (A) and (B), dot size represents the number of genes in each term, and color indicates statistical significance (-log10 FDR). Terms reaching significance (FDR < 0.05) are labeled in blue; nonsignificant terms are shown in gray. (C) Relationship between weighted mean methylation difference (attempter vs. control, x axis) and Spearman's correlation coefficient between methylation level and number of suicide attempts (y axis) for all 365 CpG sites. Among CpG sites with significant correlations (FDR < 0.05), red dots indicate hypermethylated sites and blue dots indicate hypomethylated sites in attempters relative to controls. Gray dots indicate nonsignificant correlations. Gene names are annotated for CpG sites with |ρ| > 0.4.

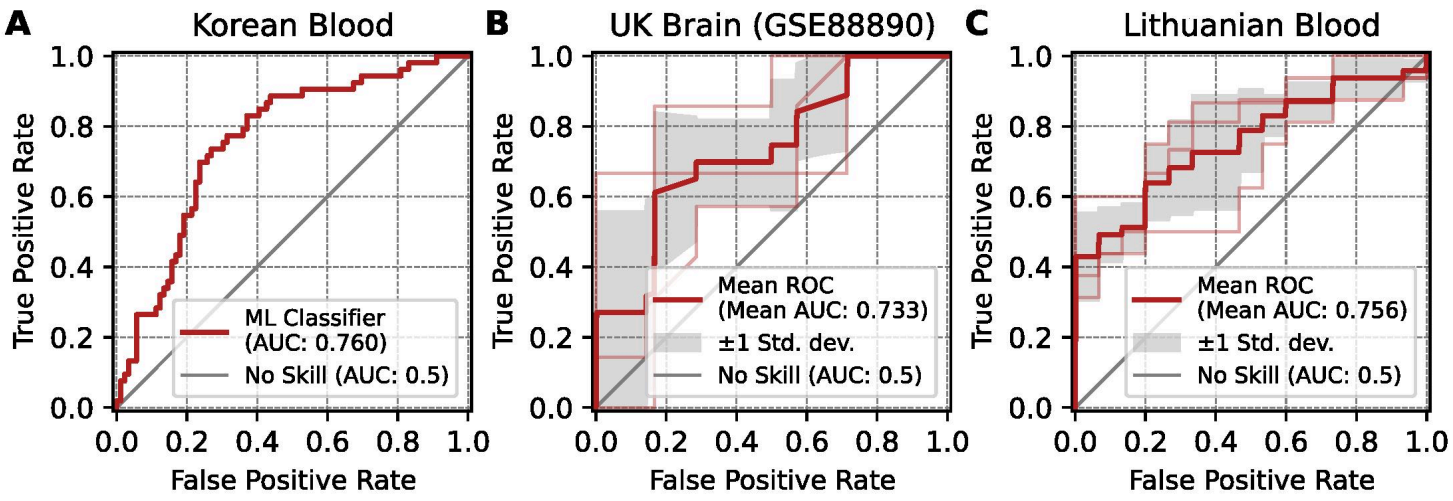


Figure 3

Classification performance of the 365 CpG marker panel across three independent validation cohorts. (A) ROC curve for the Korean blood validation cohort, in which the random forest model trained on the full discovery set was directly applied to classify suicide attempters versus matched controls. (B) ROC curve for the UK postmortem brain tissue dataset (GSE88890, Brodmann area 11). Performance was evaluated by 3-fold cross-validation within this dataset. The red line indicates the mean ROC across folds, and the shaded area represents ±1 standard deviation. (C) ROC curve for the Lithuanian blood cohort. Performance was evaluated by 3-fold cross-validation as in (B).

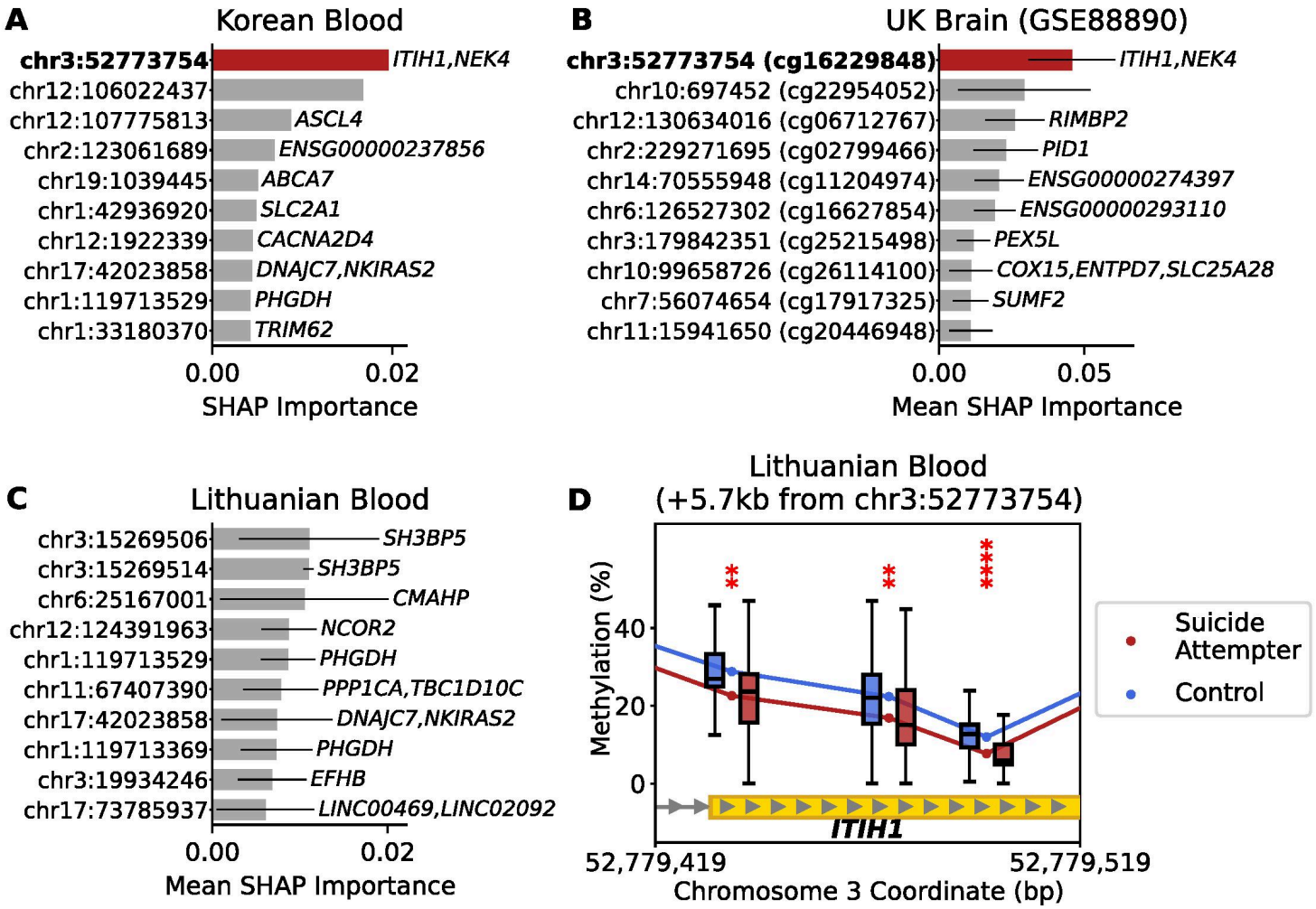


Figure 4

Feature importance across three validation cohorts reveals convergence at the *NEK4-ITIH1* locus. (A to C) SHAP-based feature importance for the 10 top-ranked CpG sites in the (A) Korean blood validation cohort, (B) UK postmortem brain tissue, and (C) Lithuanian blood cohort. Bar height represents the mean absolute SHAP value for each CpG site, with annotated gene names shown alongside. In (B) and (C), error bars indicate ±1 standard error derived from 3-fold cross-validation. The intergenic CpG site chr3:52773754 at the *NEK4-ITIH1* locus is highlighted in red. (D) Boxplots of methylation levels at three adjacent CpG sites within the *ITIH1*gene body in the Lithuanian cohort, located approximately 5.7 kb downstream of chr3:52773754. Asterisks above individual CpG sites indicate significant differences between groups by Wilcoxon rank-sum test (*: 0.01≤p<0.05; **: 0.001≤p<0.01; ***: 0.0001≤p<0.001; ****: p<0.0001)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable2.xlsx

- SupplementaryTable3.xlsx
- MPsupplementarymaterials.suicide.docx
- SupplementaryTable4.xlsx
- SupplementaryTable1.xlsx