ARTICLE OPEN

Check for updates

Epigenome-wide analysis identifies methylome profiles linked to obsessive-compulsive disorder, disease severity, and treatment response

Rafael Campos-Martin[1,9], Katharina Bey[2,3,9], Björn Elsner[4], Benedikt Reuter[4,5], Julia Klawohn[4,5], Alexandra Philipsen[2], Norbert Kathmann[4], Michael Wagner[2,3,6,10] and Alfredo Ramirez[1,3,6,7,8,10]✉

Obsessive-compulsive disorder (OCD) is a prevalent mental disorder affecting ~2–3% of the population. This disorder involves genetic and, possibly, epigenetic risk factors. The dynamic nature of epigenetics also presents a promising avenue for identifying biomarkers associated with symptom severity, clinical progression, and treatment response in OCD. We, therefore, conducted a comprehensive case-control investigation using Illumina MethylationEPIC BeadChip, encompassing 185 OCD patients and 199 controls recruited from two distinct sites in Germany. Rigorous clinical assessments were performed by trained raters employing the Structured Clinical Interview for DSM-IV (SCID-I). We performed a robust two-step epigenome-wide association study that led to the identification of 305 differentially methylated CpG positions. Next, we validated these findings by pinpointing the optimal set of CpGs that could effectively classify individuals into their respective groups. This approach identified a subset comprising 12 CpGs that overlapped with the 305 CpGs identified in our EWAS. These 12 CpGs are close to or in genes associated with the *sweet-compulsive brain hypothesis* which proposes that aberrant dopaminergic transmission in the striatum may impair insulin signaling sensitivity among OCD patients. We replicated three of the 12 CpGs signals from a recent independent study conducted on the Han Chinese population, underscoring also the cross-cultural relevance of our findings. In conclusion, our study further supports the involvement of epigenetic mechanisms in the pathogenesis of OCD. By elucidating the underlying molecular alterations associated with OCD, our study contributes to advancing our understanding of this complex disorder and may ultimately improve clinical outcomes for affected individuals.

Molecular Psychiatry (2023) 28:4321–4330; https://doi.org/10.1038/s41380-023-02219-4

INTRODUCTION

Obsessive-compulsive disorder (OCD) is a psychiatric disorder that affects around 2–3% [1, 2] of the general population and can result in severe psychosocial impairment if untreated. The disorder is characterized by excessive, unwanted thoughts (obsessions) and/or repetitive behaviors (compulsions) [3]. Despite OCD's large burden on affected individuals and the health care system, up to date, no biomarker has been found to classify the disorder in a clinical setting or to aid clinicians to predict response to pharmacological or psychological treatment.

OCD is considered a multifactorial disorder in which the risk to develop the disease is defined by the complex interaction of genetics, epigenetics, and environmental factors. From a genetic perspective, twin studies have estimated that the heritability of OCD is 47–61% [4–7]. Despite this high heritability, genome-wide association studies in OCD have identified only one genetic locus reaching genome-wide significance [8]. This might be explained by the current lack of statistical power to identify genetic variants of small effects. An alternative explanation is that the missing heritability is due to gene x environment interactions contributing to the etiology of OCD [5, 9, 10]. For example, evidence from retrospective and longitudinal studies has shown that childhood trauma and other environmental risk factors may predispose to OCD, presumably in combination with genetics [11, 12]. Although it is widely accepted that environmental factors play a role in OCD pathophysiology, evidence is not unequivocal, and specific mechanisms are largely unclear. Also, many studies lack methodological quality so far [13, 14]. Interestingly, environmental factors are known to exert their effects on disease susceptibility through epigenetic modifications leading to the modulation of expression and co-expression of several genes [15–19]. In humans, the most studied epigenetic modification is the methylation of DNA (DNAm). The development of high throughput array technology enabled genome-wide assessment of DNAm for many

[1]Division of Neurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, University of Cologne, Medical Faculty, 50937 Cologne, Germany. [2]Department of Psychiatry and Psychotherapy, University Hospital Bonn, Bonn, Germany. [3]German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany. [4]Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany. [5]Department of Medicine, MSB Medical School Berlin, Berlin, Germany. [6]Department for Neurodegenerative Diseases and Geriatric Psychiatry, University Hospital Bonn, Bonn, Germany. [7]Department of Psychiatry and Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, San Antonio, TX, USA. [8]Cluster of Excellence Cellular Stress Responses in Aging-associated Diseases (CECAD), University of Cologne, Cologne, Germany. [9]These authors contributed equally: Rafael Campos-Martin, Katharina Bey. [10]These authors jointly supervised this work: Michael Wagner, Alfredo Ramirez. ✉email: alfredo.ramirez-zuniga@uk-koeln.de

Received: 23 February 2023 Revised: 27 July 2023 Accepted: 4 August 2023
Published online: 16 August 2023

individuals at a moderate cost [20, 21]. Epigenome-wide association studies (EWAS) have shed light on many psychiatric disorders such as depression [10], anorexia nervosa, Alzheimer's disease [22, 23], or schizophrenia [24], complementing genetic research. Given their dynamic and modifiable nature, DNAm can be acquired or lost over the lifespan depending on environmental influences. Thus, epigenetic modifications may serve as biomarkers for gene x environment interactions, providing further insights into the molecular basis of OCD [25].

Exposure and response prevention (ERP) based cognitive-behavioral therapy (CBT) constitutes a first-line treatment for OCD with effect sizes being among the largest of all psychological treatments for mood, anxiety, and related disorders [26]. Still, a substantial proportion of OCD patients show an insufficient response to CBT [27], stressing the need to identify predictive markers of CBT response.

We recruited 384 participants from two German cities, Berlin and Bonn, to investigate the relationship between blood DNAm and OCD, which makes it the largest study to date. We first search for genomic loci showing differentially methylated sites between cases and healthy controls. We then computed a methylation profile score (MPS) to assess its classification power to differentiate cases from controls (Fig. 1), as well as its association with treatment response and symptom severity (Y-BOCS scale).

MATERIALS AND METHODS

Patients and controls

Biological samples were obtained from 185 patients with OCD and 199 healthy individuals who participated in the Endophenotypes of OCD study [28, 29]. The two recruitment centers, the Department of Psychology of Humboldt-University in Berlin and the Department of Psychiatry and Psychotherapy of the University Hospital in Bonn enrolled and evaluated all participants according to the same protocols (Table 1). Healthy individuals from the general population were recruited through public advertisements. All participants came from European ancestry. Before recruitment, written informed consent was given by all participants, and monetary compensation was paid for their time. The study was performed following the revised Declaration of Helsinki and approved by the local ethics committees of Humboldt University and the University Hospital Bonn.

Clinical evaluation

All participants were examined by trained psychologists using the Structured Clinical Interview for DSM-IV (SCID-I) [30]. The severity of OCD symptoms was evaluated using the German version of the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) [31, 32]. Patients with OCD were included if they: (a) were free of any psychotic, bipolar, or substance-related disorder in the past or present (b) had not been treated with any neuroleptic drug during the past 4 weeks, and (c) had not used benzodiazepines 2 weeks before the study examination. Moreover, healthy participants were excluded if they (a) had taken any psychoactive drug in the past 3 months, (b) reported any Axis I disorder, or (c) had a relative with OCD.

Current or previous treatments were assessed in the patients' group, in which ~50% had received pharmacotherapy, predominantly with SSRI. 79 OCD patients reported treatment with psychotropic medication in the past 4 weeks. A total of 25 patients had their treatments discontinued several weeks before baseline and did not take any specific medications at the time of assessment. Another 98 patients were medication-naive, reporting no priory psychotropic medication. Four patients did not provide a medication status report. The majority of patients had one or more comorbid Axis I disorders, with major depressive disorder being the most common comorbidity ($n = 41$).

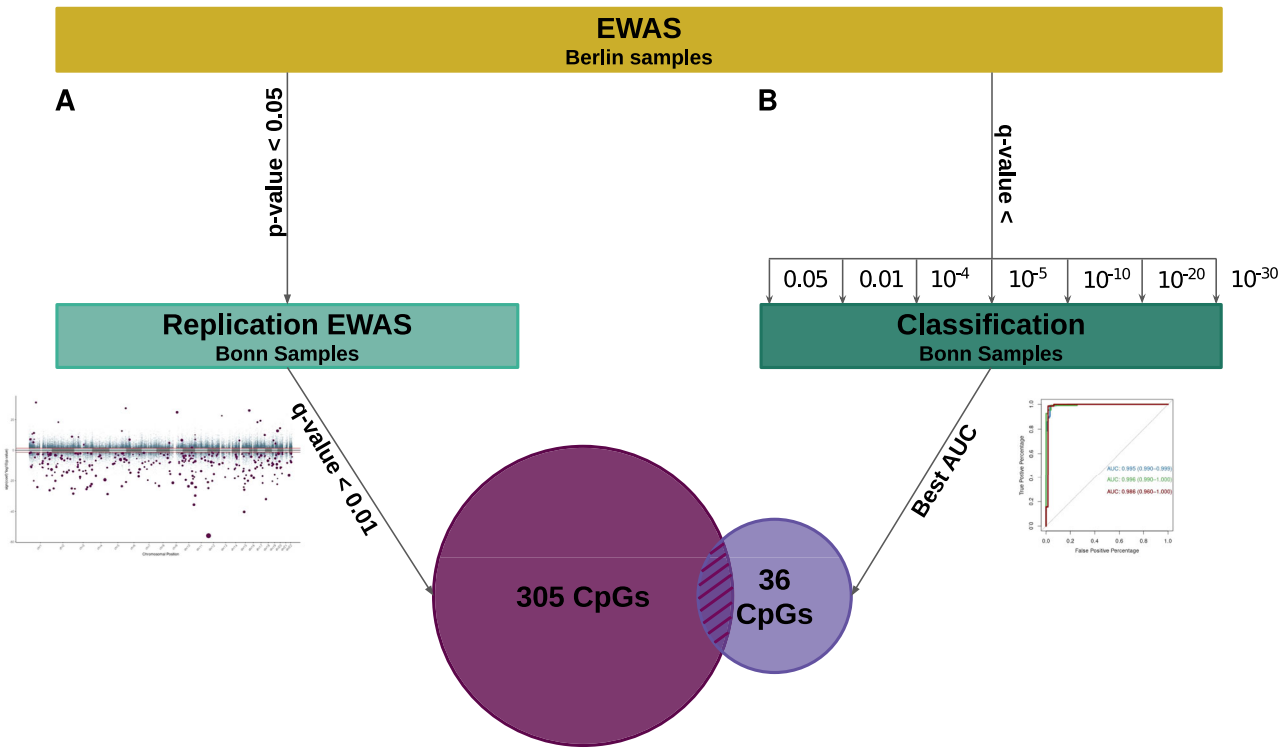

Fig. 1 A schematic representation of our analysis. After the first EWAS only on the Berlin data set (discovery stage), probes are filtered based on two different approaches and replicated in the Bonn data set. **A** all probes with p value > 0.05 in the discovery stage were removed and the remaining CpGs were further explored in the Bonn samples. The probes reaching a q value ≤ 0.01 (BH correction) in the Bonn samples were considered differentially methylated (DMP). **B** probes with q values equal or lower to 0.05, 0.01, 10^{-4}, 10^{-5}, 10^{-10}, 10^{-20}, and 10^{-30} were used to compute different MPS values. Next, each MPS was used as an independent variable to classify the Bonn samples and to select the best treshold based on the AU-ROC metric. The set of CpGs used to build the best MPS was selected. The intersection of both methods (12 CpGs) was then selected as the actual signals.

Table 1. Cohort demographics.

	Berlin		Bonn	
	OCD	**Control**	**OCD**	**Control**
N	112 (45.16%)	136 (54.84%)	73 (53.68%)	63 (46.32%)
Age	32.04 ± 9.63	32.88 ± 10.43	34.55 ± 12.3	37.95 ± 15.81
Gender (% males)	51 (45.54%)	57 (41.91%)	28 (38.36%)	17 (26.98%)
Smoking (%)	10.11	11.75	5.19	2.73
Y-BOCS	22.9 ± 5.48	-	21.33 ± 8.28	-
OCD onset (years)	21.77 ± 10.05	-	20.81 ± 12.23	-

OCD obsessive-compulsive disorder, *Y-BOCS* yale-brown obsessive-compulsive scale.

Treatment subsample

A subsample of OCD patients completed individual CBT at a university outpatient unit at the Berlin study site (Hochschulambulanz für Psychotherapie und Psychodiagnostik der Humboldt-University). The CBT sessions were administered by licensed psychotherapists and conformed to the general conditions for psychotherapy in the public German health care system, typically consisting of 25 or more individual 50-min sessions per week. Details about the treatment can be found in Bey et al. [33] and Kathmann et al. [34]. For $n = 100$ patients ($n = 54$ female, $n = 46$ male), Y-BOCS data were available at pre- and post-treatment [35].

Methylation arrays

Blood aliquots were obtained from all participants. Genomic DNA was isolated from whole blood and DNA concentration and purity were determined using the NanoDrop ND1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). All samples were of sufficient quantity and quality. 500 ng genomic DNA was used as input for the bisulfite conversion reaction using the EZ-96 DNAm Methylation-Lightning MagPrep Kit (Zymo Research Europe GmbH, Freiburg, Germany) with an elution volume of 15 µl. Bisulfite-treated DNA was vacuum concentrated and resuspended in 10 µl. A total of 4 µl of the resuspension was used as input for the Infinium Methylation EPIC BeadChip (Illumina Inc, San Diego, CA, USA). All analysis steps were performed following the manufacturer's instructions. The Illumina iScan was used for imaging the array and data were exported in .idat format.

Data acquisition and quality control

The R (Bioconductor) Meffil [36] package was used throughout our pipeline to analyze the complete data set. All raw idat files were pooled together to run the quality control and normalization steps. Samples were removed if there was a mismatch between the estimated methylation sex and the gender provided by the participant, deviations from the mean value for control probes, or the median intensity for the methylated or unmethylated signal deviated more than three standard deviations (s.d.).

Probes were removed for further analysis if they mapped to a sex-chromosome, had a detection *p* value below 0.05, beadcount lower than three, or were aligned to multiple locations in the genome according to Nordlund et al. [37]. In addition, we removed the 10% of probes with the lowest variability to reduce the number of probes and multiple tests [38]. In the end, 366 samples (189 controls and 177 cases) and 632,997 probes passed all our quality control filters and were used to normalize the methylation intensities.

Functional normalization [36, 39] was applied to remove technical variation using 15 principal components (PCs) and an assessment center (Berlin/Bonn) as a fixed effect. Blood cell proportion was imputed using functionalities from meffil for each individual and used in the linear models to correct the methylation effect.

Two-step EWAS

To analyze our data set, the two cohorts were initially kept separated (Berlin and Bonn). While the larger cohort from Berlin served as a discovery cohort in the EWAS, the Bonn cohort was used for replication.

Meffil uses the Independent Surrogate Variable Analysis (ISVA) method which allows for estimating confounding factors (CF) in methylation studies [36, 40, 41]. Briefly, the ISVA uses the independent component analysis method to model CFs as statistically independent variables in each probe analysis [41]. Thus, ISVA provides a non-supervised framework for accounting for any CF.

Methylation status was compared between controls and OCD cases using a linear regression model. Adjustments were made for age [42], sex [43], smoking [44], cell composition [45], and surrogate variables calculated by meffil.

The current strategy for selecting CpGs for further analysis aims first to remove the maximum number of probes in the discovery step optimizing the minimum number of false negatives (p value < 0.05). The replication step follows with a more restrictive adjusted p value (q value) threshold to select CpGs that are truly associated with the phenotype (Holm-Bonferroni q value < 0.01). A similar strategy has been applied to genetics [46] and methylation [10, 23] studies.

We estimated the false discovery rate (FDR) for our approach following the method suggested by Jiang et al.[47]:

$$\widehat{FDR}(a_1, a_2) = \frac{\hat{P}(p_{i2} \le d_2 | D_2 = 0, D_1 = 0, p_{i1} \le d_1)}{\hat{P}(p_{i2} \le d_2 | p_{i1} \le d_1)} \cdot FDR_1 + \frac{\hat{P}(p_{i2} \le d_2 | D_2 = 0, D_1 = 1, p_{i1} \le d_1)}{\hat{P}(p_{i2} \le d_2 | p_{i1} \le d_1)} \cdot \hat{\pi}_{o2} \quad (1)$$

Briefly, a probe i with $p_{i1} \le c_1$ will pass to the second stage, where p_{i1} is the p value in the first stage and c_1 is the threshold for the first stage. Following similar arguments for the second stage, $p_{i2} \le c_2$, then we say that this probe has a significant difference in methylation values between cases and controls. At stage j, d_j is the smallest p value for the probes that $p_{ij} > c_j$, and the D_j is a binary variable that indicates whether there are actual differences between the cases and controls; $D_j = 0$ for no differences, and $Dj = 1$ for actual differences. The probability that a probe is significant after our two-stage approach when there are no real differences, $\hat{P}(p_{i2} \le d_2 | D_2 = 0, D_1 = 0, p_{i1} \le d_1)$, was estimated by permutating for 100 times the samples. The proportion of the true null hypothesis ($\hat{\pi}_{o2}$) was estimated following the Storey method [48] and $\hat{P}(p_{i2} \le d_2 | D_2 = 0, D_1 = 1, p_{i1} \le d_1)$ equals d_2. Finally, $\hat{P}(p_{i2} \le d_2 | p_{i1} \le d_1)$ is the proportion of significant probes in the second stage. Our calculation for our setup yielded an FDR of 3.26×10^{-5}.

Weighted correlation network analysis

Weighted correlation Network Analysis (WCNA) uses the pairwise correlation between variables to define clusters within the variables and to associate these clusters with other phenotypes.

The R package WGCNA was used for this purpose [49, 50]. Once the network was constructed, module detection was achieved by unsupervised clustering. WGCNA uses the dynamic tree-cut method to select the number of clusters given the hierarchical clustering for the adjacency matrix.

Case–control classification based on methylation profile score

The MPS is a numerical value computed for each individual using a set of DMPs. Like polygenic risk scores in genetic studies [51], MPS improves classification capacity by leveraging methylation information on DNAm differences between cases and controls. The MPS for an individual i can be computed as

$$MPS_i = \sum_{j=1}^{P} \beta_j m_{ji} \quad (2)$$

where P is the number of CpGs, β_j is the coefficient for the association of the probe j to the phenotype, and $m_{j,i}$ is the methylation status of probe j. Herein, a set of q values thresholds was used to select the number of probes: 0.05, 0.01, 1×10^{-4}, 1×10^{-5}, 1×10^{-10}, 1×10^{-20}, 1×10^{-30}, and 1×10^{-40}. For each threshold, an MPS was computed using the selected CpGse and then its classification capacity was tested using the Bonn sample as an independent dataset. AU-ROC for each threshold was computed using the R-package p-ROC.

Clinical correlates and treatment analysis

To assess whether the OCD-related methylation profile is associated with symptom severity, we correlated our most reliable MPS (i.e., MPS_{common}, see Results) with the Y-BOCS scores of all patients. In the treatment subsample, we also examined whether the MPS_{common} predicts treatment response by performing linear regression analysis with Y-BOCS baseline score, and the Y-BOCS baseline score × MPS_{common} interaction as independent variables, and pre-to-post change in Y-BOCS score as the dependent variable. Age, gender, and medication were included as covariates.

Dimensionality reduction

A linear transformation algorithm and a non-linear transformation algorithm were used to reduce dimensionality. Principal component analysis (PCA) is the most popular linear transformation for dimensionality reduction. PCA estimates new coordinates that preserve the maximum variance of the dataset and projects the data points into the new orthogonal coordinate system. The base function prcomp in R was used to estimate the new PCs and projections. On the other hand, uniform manifold approximation and projection (UMAP) has become one of the most popular non-linear transformation algorithms. By using a framework that combines geometry and algebraic topology, UMAP can project a data set into two dimensions and reflect distances between points. The function umap in R was used to obtain the new coordinates.

RESULTS

Epigenome-wide association study

To identify potential loci associated with OCD, we conducted a two-step case-control EWAS using samples recruited in Berlin for discovery and samples originating from Bonn for replication. This approach rendered in the Berlin sample a total of 188,488 DMPs with a nominal p value < 0.05. These sites were moved forward to the replication stage using the Bonn samples. We identified 310 DMPs discriminating cases and controls with a corrected p value for multiple testing $q < 0.01$ (Fig. 2).

We explored the correlation between the coefficients of the probes analyzed in the discovery and the replication stage (Fig. 2).

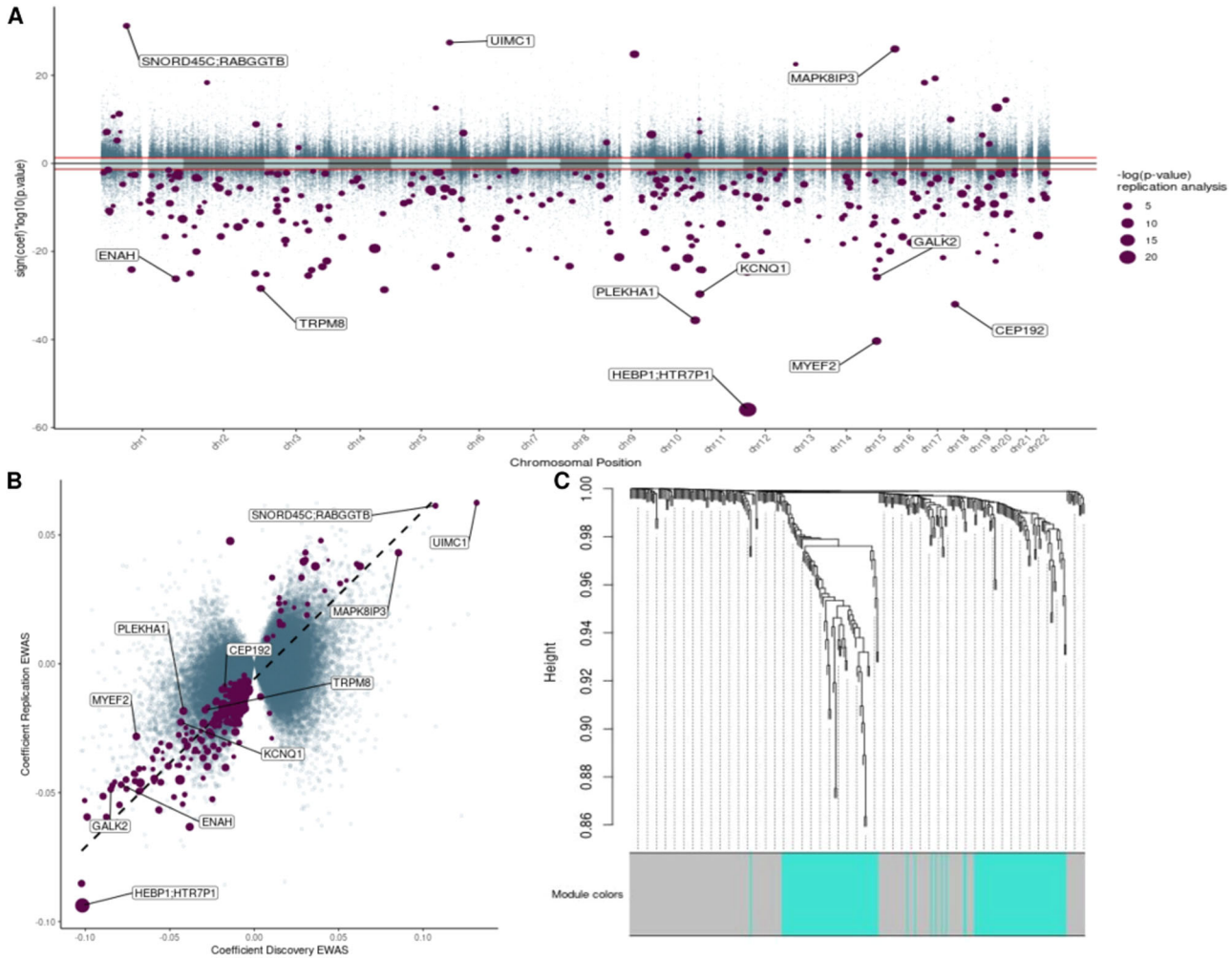

Fig. 2 Two-step EWAS results. A Miami Plot for the two-stage analysis. The X-axis is the genome position. Y-axis is the nominal p-value for the discovery EWAS on a logarithmic scale multiplied by the sign of the coefficient in the same analysis. Horizontal red lines define the threshold 0.05 of the discovery analysis. Purple dots are the 310 CpGs that were significant at the replication stage, the dot size is equivalent to the adjusted p-value in the replications stage on a logarithmic scale. **B** Correlation of the discovery and replication stage. The *X*-axis shows the discovery coefficient, and *Y*-axis shows the replication coefficient. The purple dots represent the 310 CpGs that were significant at the end of the two-step approach; the dot size is equal to the adjusted p value in the replication stage. The dashed line shows the trend of the linear model based on the purple dots. **C** Cluster dendrogram. Branches refer to highly interconnected clusters of CpGs. Modules are represented by the colors in the horizontal bar.

This analysis showed that while the overall correlation for all 188,488 CpG sites was moderate ($r = 0.42$, $p < 2.2 \times 10^{-16}$), it was much stronger for the 310 DMPs in the replication stage ($r = 0.88$, $p < 2.2 \times 10^{-16}$). Only five DMPs showed opposite effect directions between discovery and replication, therefore they were removed from further analysis (Supplementary Table 1, Supplementary Fig. 1).

Of the 305 probes identified by our analysis, 241 were annotated to 233 genes based on the Illumina annotation. Gene Ontology (GO) analysis, using the R package missMethyl [49, 50], for the same probes did not show any term enriched after multiple test corrections. Of note, five terms from the GO analysis showed a nominal p value < 0.05 (Supplementary Table 2).

Network analysis identifies two different submodules

Given the complex nature and many pathways involved in OCD, we sought to search whether common patterns of methylation emerge among the 305 DMPs. Thus, we used WCNA which exploits correlations among probes and groups them into modules using network topology. After fitting several powers (β), we found that a power of ten approximated the best scale-free network for our co-methylation network (Supplementary Fig. 2). The adjacency matrix was then computed by using the optimal β and the methylation values. Based on the TOM dissimilarity measure, the hierarchical clustering yielded two consensus network modules, i.e., gray ($n = 169$, Supplementary Table 2) and turquoise ($n = 136$) (Fig. 2, Supplementary Tables 2 and 4).

Then we examined whether each module was associated with other phenotypes. To this end, we looked at the Pearson correlation coefficient and p value of the association of the eigenvector of each module with OCD status, age, sex, city, smoking, and Y-BOCS. While both modules were highly correlated with OCD phenotype (turquoise: $r = -0.88$, $p = 4 \times 10^{-122}$; gray: $r = -0.79$, $p = 3 \times 10^{-78}$), only the turquoise module was associated significantly with the Y-BOCS ($r = -0.2$, $p = 2 \times 10^{-4}$) (Supplementary Fig. 3). Interestingly, the gray module better captured the differences in the origin of the samples (Supplementary Fig. 3, 4, and 5).

The methylation profile score offers predictive performance for sample classification

Considering that both submodules and the full set showed a strong correlation with OCD status, we attempted to derive a MPS by following a similar strategy to developing polygenic risk scores [52]. To this end, we first constructed an MPS using only the 305 DMPs, which were confirmed in the replication stage ($MPS_{two\text{-}step}$). We also computed an MPS for each module, i.e., turquoise ($MPS_{turquoise}$) and gray (MPS_{gray}).

The $MPS_{two\text{-}step}$ was indeed statistically different between OCD patients and controls for both, the Berlin ($p < 2.2 \times 10^{-16}$) and the Bonn samples ($p < 2.2 \times 10^{-16}$), whereas the difference of $MPS_{two\text{-}step}$ values between both cities for the control group ($p = 0.269$) and the OCD patients ($p = 0.057$) was not significant (Fig. 3).

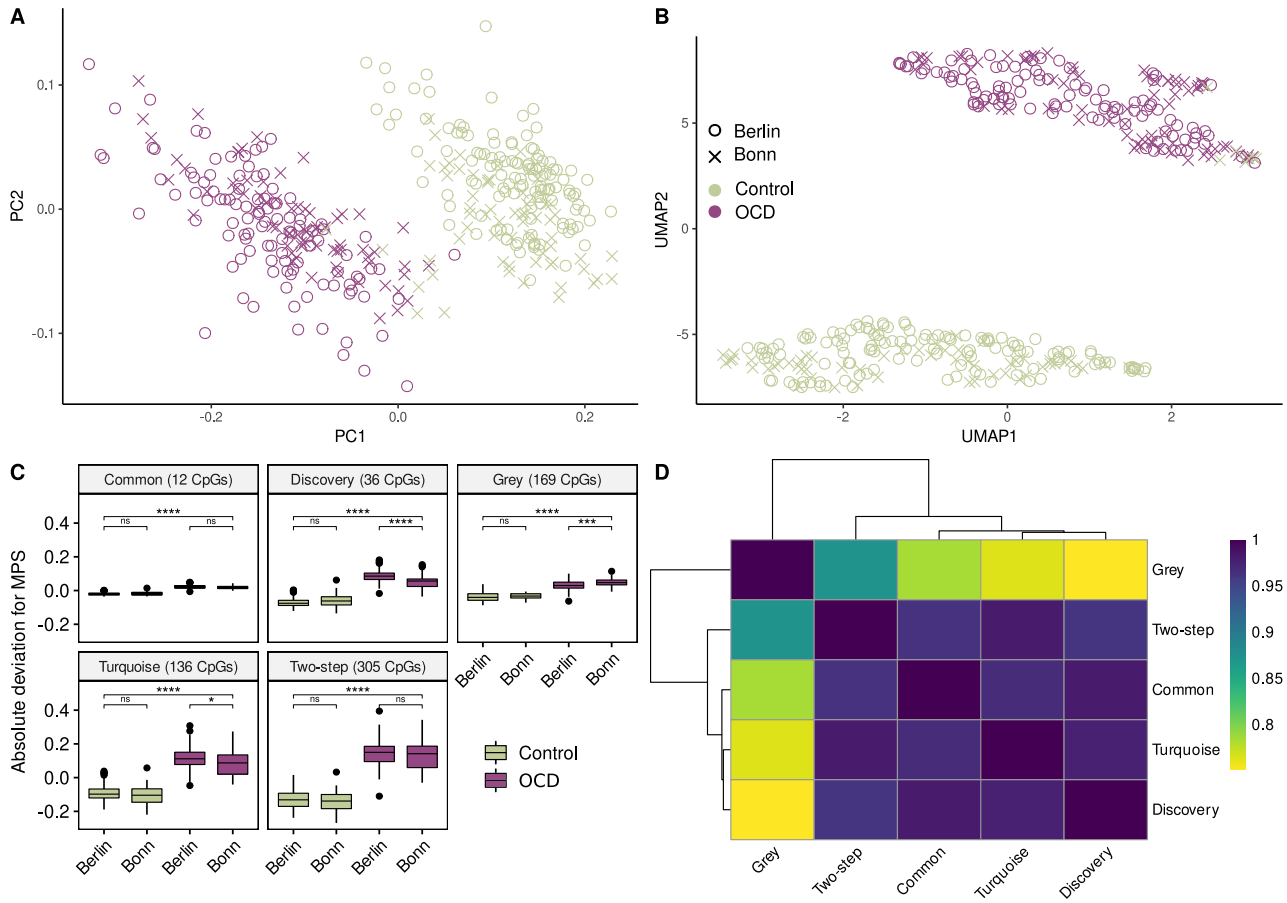

Fig. 3 Classification power of the resulting CpGs. Projection of the samples into a two-dimensional space using (**A**) PCA, and (**B**) UMAP. The 12 CpGs found as result of our analysis were used as input features. Purple data points are OCD patients and green are Controls. **C** Each facet represents the deviation from the mean for each MPSs. The number of CpGs that were used to calculate the MPS is shown in parentheses. Horizontal brackets display the results of the t-test for the set. **D** MPS correlation matrix. PC: Principal Component; UMAP: Uniform manifold approximation and projection; ns: not significant; * : p.value < 0.05; ** : p value < 0.01; ***: p value < 0.001; ****: p value < 2×10^{-16}.

The lack of an independent third validation cohort to test the $MPS_{two\text{-}step}$ independently prompted us to consider an alternative strategy for constructing the MPS. Herein, we constructed several MPSs using DMPs based on an a priori set of 8 corrected p value (q values) thresholds (P_T) obtained from the EWAS performed in the discovery stage (Berlin samples only). Finally, classification accuracy for each calculated MPS was examined in the Bonn data set, which did not contribute to this MPS and could be used to test out-of-sample classification accuracy. The best classification accuracy for the Bonn sample is achieved using probes with q values $< 1 \times 10^{-20}$ ($AU\text{-}ROC_{Berlin} = 0.991$, $AU\text{-}ROC_{Bonn} = 0.968$; Table 2). The MPS obtained for this threshold ($MPS_{discoverey}$) contains 36 DMPs (Supplementary Table 5 and Supplentary Fig. 6), from which 12 are shared with the $MPS_{two\text{-}step}$ and the $MPS_{turquoise}$ (Table 2). For this reason, we also constructed an MPS containing only the common CpGs (MPS_{common}) which also showed a good classification power (Fig. 3 and Table 2).

Association between MPS, clinical variables, and treatment response

As indicated by Pearson correlation, the MPS_{common} was significantly associated with Y-BOCS scores across all OCD patients ($r = 0.17$, $p = 0.023$), indicating that a more severe symptom severity goes along with a higher epigenetic profile score. In the regression model assessing treatment response, we observed effects at the trend level for the Y-BOCS baseline score ($\beta = -3.108$, $t = -1.96$, $p = 0.053$) and the Y-BOCS baseline by MPS_{common} interaction ($\beta = -2.78$, $t = -1.74$, $p = 0.086$). To follow up on this interaction, we ran separate analyses for patients with high and low MPS_{common} (median split: $n = 56$ low-scorers, $n = 44$ high-scorers). In MPS_{common} high-scorers, we found a significant effect of Y-BOCS baseline ($\beta = -0.44$, $t = -3.09$, $p = 0.004$) and a trend level association of the MPS_{common} ($\beta = -0.28$, $t = -1.86$, $p = 0.070$) with treatment response, indicating that patients with a higher score might show a better treatment response independently of baseline symptom severity. In MPS_{common} low-scorers, there were significant effects of Y-BOCS baseline ($\beta = -0.32$, $t = -2.43$, $p = 0.019$) and medication ($\beta = 0.35$, $t = 2.67$, $p = 0.010$). Notably, we did not observe any significant effects of age or gender in all analyses ($p > 0.05$). Moreover, there was no significant association between MPS_{common} and Y-BOCS baseline score in the treatment subsample ($r = 0.09$, $p = 0.37$), potentially due to sample size reduction.

Functional Annotation

Since the GO analysis did not reveal clear supporting evidence for functional terms that may be relevant or previously associated with OCD, we conducted a focused literature search on the 12 common CpGs identified in both MPS approaches because they may still represent true signals involved in the disease process operating in OCD. Consequently, we first mapped each CpG to the closest gene and gene position (Table 3).

The highest association was found for the CpG cg17232014, which shows a substantial hypomethylation in OCD patients compared to controls. This CpG maps to a transcription start site (TSS) for two genes: Heme Binding Protein 1 (*HEBP1*) and the 5-Hydroxytryptamine Receptor 7 Pseudogene 1 (*HTR7P1*), most commonly known as serotonin receptor pseudogene (Supplementary Fig. 7). Although the functional consequence of the decreased methylation at this TSS is not fully understood yet, it likely results in an elevated gene expression of either *HEBP1* or *HTR7P1* or both.

Next, we observed that some of the associated CpGs were located close to genes linked to glucose metabolism. Thus, the cg01647172 is mapped to the 5′ untranslated region of the gene Pleckstrin Homology Domain Containing A1 (*PLEKHA1*) and is found hypomethylated in OCD patients. Likewise, we observed that the hypomethylated CpG cg00382572 position is assigned to the *KCNQ1* gene coding for the KCNQ1 potassium channel, which is located in the pancreas and has been also associated with diabetes [53–58]. Finally, the cg19069918 is located near the gene *TRPM8*, which has been long studied as a cancer biomarker, particularly in pancreatic cancer [59].

The next set of CpGs was annotated to genes involved in different processes related to resident cells of the brain. Thus, the probe cg06215939 is found hypermethylated at the TSS of the Mitogen-Activated Protein Kinase 3 gene predicting a reduction in gene expression. For cg21812670, the methylation was found to be increased in OCD patients. This position is located at the TSS of the gene coding for the Rab geranylgeranyl transferase which is essential for synaptic vesicle release [60]. Along these lines, the cg13959110 is located in the gene coding for the brain myelin expression factor 2. This gene is a transcriptional repressor of the myelin basic protein gene that has been involved in myelin homeostasis. Another CpG, cg25195309 is located in the Enable Homolog gene. The function of this gene has been linked to actin polymerization in neurons [61]. Herein, neurons lacking these proteins cannot perform neuritogenesis in the developing cortex [61].

Table 2. MPS properties.

		AU-ROC		Y-BOCS Correlation		nCpGs
		Berlin	Bonn	*r*	*p*	
q values threshold for top-down Analysis	0.05	1	0.911	-	-	11,998
	0.01	1	0.908	-	-	9,430
	0.0001	1	0.909	-	-	4,902
	1×10^{-5}	1	0.915	-	-	3,568
	1×10^{-10}	1	0.88	-	-	808
	1×10^{-20} (Discovery)	0.999	0.974	0.200	0.009	36
	1×10^{-30}	1	0.944	-	-	2
Two-step analysis		0.990	0.984	0.229	0.003	305
Gray		0.963	0.994	0.113	0.142	169
Turquoise		0.993	0.981	0.223	0.003	136
12 common CpGs		0.998	0.986	0.146	0.058	12

AU-ROC area under the receiver operating curve, *Y-BOCS* yale-brown obsessive-compulsive scale, *nCpGs* number of CpGs under the threshold selected, *q value* Bonferroni adjusted the *p* value for the discovery study, *r* Pearson correlation coefficient, *p* p-value for the test.

Table 3. Biological annotation and summary statistics for the 12 common CpGs.

	Position	Gene	*q* value discovery	Coefficient discovery	*q* value replication	Coefficient replication
cg17232014	chr12:13153193	*HEBP1*; *HTR7P1*	7.38×10^{-51}	−0.1	7.47×10^{-24}	−0.09
cg01647172	chr10:124146007	*PLEKHA1*	1.45×10^{-30}	−0.04	1.68×10^{-6}	−0.02
cg13959110	chr15:48466199	*MYEF2*	2.78×10^{-35}	−0.07	5.65×10^{-6}	−0.03
cg00382572	chr11:2574042	*KCNQ1*	1.32×10^{-24}	−0.04	1.08×10^{-5}	−0.02
cg06215939	chr16:1755402	*MAPK8IP3*	6.45×10^{-21}	0.09	1.55×10^{-5}	0.04
cg20469575	chr4:169122189		1.26×10^{-23}	−0.07	3.63×10^{-5}	−0.05
cg19069918	chr2:234921635	*TRPM8*	2.53×10^{-23}	−0.03	5.43×10^{-5}	−0.02
cg25195309	chr1:225766155	*ENAH*	4.12×10^{-21}	−0.08	8.89×10^{-5}	−0.05
cg07397958	chr15:49476141	*GALK2*	8.86×10^{-21}	−0.08	9.63×10^{-5}	−0.05
cg16449667	chr18:13024185	*CEP192*	6.44×10^{-27}	−0.02	1.14×10^{-4}	−0.01
cg21812670	chr1:76251636	*SNORD45C*; *RABGGTB*	3.66×10^{-26}	0.11	6.43×10^{-4}	0.06
cg19755108	chr5:176434079	*UIMC1*	2.30×10^{-22}	0.13	7.54×10^{-4}	0.06

q value, bonferroni adjusted the *p* value.

DISCUSSION

In the present study, our primary goal was to identify changes in DNAm associated with OCD status. Following a discovery and replication strategy, we identified 305 CpGs that were differentially methylated between cases and controls. Using these 305 DMPs, or a subset of them, allowed us to classify cases and controls accurately. Importantly, similarly, high classification accuracy was reached when we applied a different analytical strategy using the strongest disease-related DMP signals of the Berlin sample to predict caseness in the independent sample from Bonn. Both analytical strategies converged on 12 common CpGs deserving further scrutiny. Finally, we found a significant association of a methylation score based on these common 12 CpGs with OCD symptom severity, as well as a trend level association with treatment response to CBT in OCD patients with high MPS, indicating that patients with larger values show better treatment response. This latter result might allow MPS to be used as a biomarker for predicting treatment response in OCD from a translational perspective.

While EWAS has already led to important advances in other neurological and psychiatric disorders, it is still early days for OCD epigenetics [31, 62]. For example, a study on the Chinese Han population reported 8417 DMPs in the blood of 65 cases and 96 controls [62]. In addition, the comparison of DNA methylation in the saliva of 59 patients with OCD and 54 controls of European origin identified nine genes with methylation changes related to OCD and ADHD which however did not survive multiple testing correction [31]. In 2022, Shiele et al. reported nine genome-wide significant DMPs mapping to several microRNAs and pseudogenes in the saliva of 68 OCD patients and 68 controls of European origin [63]. Importantly, we could not identify any overlapping signal in our datasets.

In this regard, a strength of our study is the two-step approach in which we treated Berlin and Bonn samples as independent cohorts. As a result, we were able to avoid the "winner's course" in our analysis, i.e., overestimation of small effect sizes in underpowered cohorts. Although our sample size might seem underpowered, we defined the expected number of false positive signals that will arise from our study design following the methodology described by Jiang et al.[47]. Thus, after permuting the samples to estimate the FP rates, on average, 32,711 probes would be significant after the discovery step, which is in agreement with the theoretically expected ($0.05 \times 632{,}997 \approx 31650$). In addition, the second step would not report any significant probe under the threshold imposed. Consequently, the overall FDR was 0.003%, which corresponds to approximately 21 false DMPs after the replication step. Therefore, we assume that genuine signals among the 305 CpGs identified in our study are included. Supporting this assumption, our analytical strategy converged on 12 common probes out of the 305 DMPs that may represent true pathophysiological processes involved in OCD.

Pathway search did not lead to the identification of obvious candidate pathways for OCD, but the disgenet [64, 65] tool and literature search revealed that genes near the 12 CpGs have been linked to diseases like diabetes, Parkinson's disease, ADHD, and multiple sclerosis. Interestingly, the pathogenic processes involving these genes are also linked to OCD, including glucose metabolism, the dopaminergic/serotonin system, and neuronal function. For glucose metabolism, we found that the PLEKHA1 locus has been associated with type 1 and type 2 diabetes mellitus and age-related macular degeneration (AMD) [66, 67]. In AMD, previous research has shown that TAPP1, a PLEKHA1 protein product, works as an activator of lymphocytes, indicating that PLEKHA1 plays a role in inflammation. Interestingly, increasing evidence has shown that inflammatory pathways are common pathogenetic mediators in the natural course of both types of diabetes that involve the activity of PLEKHA1 [68]. For KCNQ1, research has shown that overexpression of the ion channel in mouse-derived pancreatic β-cells leads to an impairment in insulin secretion stimulated by glucose and pyruvate [53]. Lastly, rats with deletion of the *TRPM8* gene showed reduced insulin levels in serum due to enhanced insulin clearance in the liver. This was caused by afferent fibers innervating the hepatic portal vein, which is critical for metabolic homeostasis [69]. Importantly, this latter mechanism also seems to be the intersection connecting the nervous system with the metabolism of glucose and insulin. Hence, our data suggest that an underlying dysregulation in insulin/glucose metabolism may drive, at least in part, the symptoms and the disease processes occurring in OCD patients. Unfortunately, we did not have serum samples from patients before and after therapy to analyze whether glucose and insulin homeostasis changed after treatment.

Besides insulin and glucose metabolism, we also identified several genes involved in brain function. For example, both genes near cg17232014 on chromosome 12 *HEBP1* and *HTR7P1* have been associated with brain phenotypes. Thus, increased expression of HEBP1 in the brain has been linked to neurotoxicity [70] and neuroinflammation [71]. *HTR7P1*, although this is a pseudogene that does not translate into protein, genetic variants in *HTR7P1* have been associated with neurological and growth phenotypes in children [72].

In our study, we identified several signals that support the sweet-compulsive brain hypothesis [73]. This hypothesis states that abnormal dopaminergic transmission in the striatum may perturb insulin signaling sensitivity in OCD patients. Deep brain stimulation in patients with OCD supports the hypothesis that dopamine transmission affects glucose and insulin metabolism in the brain. Interestingly, non-diabetic OCD patients seem to have an increased hepatic and peripheral insulin sensitivity [74], supporting our findings on *PLEKHA1*, *KCNQ1*, and *TRPM8*. Further reinforcing our brain-related genes and their connection with glucose and insulin homeostasis, research on insulin receptor signaling in the central nervous system showed that insulin receptor signaling regulates the maintenance of synapses. In addition, insulin receptor signaling contributes to the processing of sensory information, as well as structural plasticity triggered by external experience [75].

From a molecular perspective, our findings on insulin signaling receive further support from previous genetic studies using gene enrichment tools on published OCD GWASs. Herein, gene enrichment analysis using suggestive genetic signals from these GWASs showed that 40 out of 89 of the GWAS suggestive signals clustered in insulin and insulin-related signaling cascades [76]. Furthermore, using polygenic risk score-based analysis, Bralten et al. reported shared genetic etiology between OCD or Obsessive-compulsive symptoms (OCS) and type 2 Diabetes Mellitus and fasting insulin levels. Noteworthy, a significant association with OCS was found for a gene set containing central nervous system insulin signaling genes. It is interesting to note that a CpG site in our study mapped to *KCNQ1*, one of the genes contained in this gene set analyzed by Bralten et al. [77]. *KCNQ1* is an obesity susceptibility gene that shows differentially methylated CpG sites between obese and lean women [78]. *KCNQ1* is also an imprinted gene (a parental-specific epigenetic modification) expressed exclusively from the maternal allele during fetal development [79]. These combined findings emphasize the significance of investigating the intricate interplay between genetics and environmental factors in the etiology of OCD and how epigenetic modifications may serve as a bridge connecting both. Although there is still a long road ahead, exploring the relationship between OCD and the methylation status of *KCNQ1* and other insulin-related genes might open new avenues for potential therapeutic or prevention strategies involving non-pharmacological dietary intervention.

In supporting our findings, three of our twelve most significant DMPs were found in a recent study comparing people with generalized anxiety disorder (GAD), or OCD, with healthy controls of Chinese Han origin [80]. These probes map to *RABGGTB*, *MPK8IP3*, and *ENAH* genes. To our knowledge, this is the first time that two different studies on methylation in OCD replicated each other's results using populations of different ethnic backgrounds. Of note, Guo et al. used a similar approach and methodologies to analyze their data as in our study. Herein, DNAm is highly sensitive to batch effect and other factors that might increase the variability. Therefore, it is crucial to account for confounding factors when analyzing this kind of data set.

The correlation between MPS and OCD symptom severity highlights the potential clinical utility of epigenetic measures. Future studies should examine whether changes in symptom severity also go along with epigenetic modifications. Interestingly, we observed a trend-level association between MPS and treatment response in OCD patients, indicating that patients with the highest MPS showed better treatment response independent of baseline symptom severity. Among MPS-low scorers, there was no association with treatment response. Although we interpret this preliminary finding with caution, it may show that patients with high MPS exhibit features that make them benefit more from CBT than others, e.g., a larger environmental component contributing to their OCD.

Our results should be interpreted considering some important limitations. First, DNA extraction was done in Bonn for all samples including those derived from the Berlin sample. Consequently, Berlin blood samples were transported uncooled before DNA extraction, which may contribute to variation in the methylation analysis. However, our study considered this source of bias including the fact that we initially analyzed both samples independently. To avoid this source of technical bias, future studies should include cool transport of blood samples to the processing center or proceed locally with the DNA extraction before frozen transport to the analyzing center.

In summary, we identified 12 epigenome-wide significant CpGs for OCD using a robust statistical analysis of two German samples. The clinical validity of these CpGs is supported by the significant associations of our methylation profile score with OCD diagnosis, symptoms severity, and—at trend level—treatment response to CBT. Furthermore, genetic annotation contemplates a strong interaction of insulin and the dopaminergic system with OCD. Our findings thus support the role of epigenetic mechanisms in OCD and may help pave the way for biologically-informed individualized treatment options.

REFERENCES

1. Schulze D, Kathmann N, Reuter B. Getting it just right: a reevaluation of OCD symptom dimensions integrating traditional and Bayesian approaches. J Anxiety Disord. 2018;56:63–73.
2. Ruscio AM, Stein DJ, Chiu WT, Kessler RC. The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. Mol Psychiatry. 2010;15:53–63.
3. Stein DJ, Costa DLC, Lochner C, Miguel EC, Reddy YCJ, Shavitt RG, et al. Obsessive–compulsive disorder. Nat Rev Dis Prim. 2019;5:52.
4. Monzani B, Rijsdijk F, Harris J, Mataix-Cols D. The structure of genetic and environmental risk factors for dimensional representations of DSM-5 obsessive-compulsive spectrum disorders. JAMA Psychiatry. 2014;71:182–9.
5. Hudziak JJ, Beijsterveldt CEM, van, Althoff RR, Stanger C, Rettew DC, Nelson EC, et al. Genetic and environmental contributions to the child behavior checklistobsessive-compulsive scale: a cross-cultural twin study. Arch Gen Psychiatry. 2004;61:608–16.
6. Mahjani B, Bey K, Boberg J, Burton C. Genetics of obsessive-compulsive disorder. Psychol Med. 2021;51:2247–59.
7. Alemany-Navarro M, Cruz R, Real E, Segalàs C, Bertolín S, Rabionet R. et al. Looking into the genetic bases of OCD dimensions: a pilot genome-wide association study. Transl Psychiatry.2020;10:151 https://doi.org/10.1038/s41398-020-0804-z.
8. Strom NI, Yu D, Gerring ZF, Halvorsen MW, Abdellaoui A, Rodriguez-Fontenla C, et al. Genome-wide association study identifies new locus associated with OCD. Medrxiv. 2021;2021.10.13.21261078.
9. Rodriguez N, Martinez-Pinteño A, Blázquez A, Ortiz AE, Moreno E, Gassó P, et al. Integrative DNA methylation and gene expression analysis of cognitive behavioral therapy response in children and adolescents with obsessive-compulsive disorder; a pilot study. Pharmacogenom Pers Med. 2021;14:757–66. https://www.dovepress.com/integrative-dna-methylation-and-gene-expression-analysis-of-cognitive–peer-reviewed-fulltext-article-PGPM.
10. Jovanova OS, Nedeljkovic I, Spieler D, Walker RM, Liu C, Luciano M, et al. DNA methylation signatures of depressive symptoms in middle-aged and elderly persons: meta-analysis of multiethnic epigenome-wide studies. JAMA Psychiatry. 2018;75:949–59. https://jamanetwork.com/journals/jamapsychiatry/fullarticle/2687369.
11. Grisham JR, Fullana MA, Mataix-Cols D, Moffitt TE, Caspi A, Poulton R. Risk factors prospectively associated with adult obsessive–compulsive symptom dimensions and obsessive–compulsive disorder. Psychol Med. 2011;41:2495–506.
12. Destrée L, Brierley MEE, Albertella L, Jobson L, Fontenelle LF. The effect of childhood trauma on the severity of obsessive-compulsive symptoms: a systematic review. J Psychiatr Res. 2021;142:345–60.
13. Brander G, Pérez-Vigil A, Larsson H, Mataix-Cols D. Systematic review of environmental risk factors for obsessive-compulsive disorder: a proposed roadmap from association to causation. Neurosci Biobehav Rev. 2016;65:36–62.
14. Raposo-Lima C, Morgado P. The role of stress in obsessive-compulsive disorder: a narrative review. Harv Rev Psychiatry. 2020;28:356–70.
15. Pisco AO, d'Hérouël AF, Huang S. Conceptual Confusion: The case of Epigenetics. Biorxiv. 2016;053009.
16. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nat Genet. 2017;49:131–8.
17. Mehta D, Klengel T, Conneely KN, Smith AK, Altmann A, Pace TW, et al. Childhood maltreatment is associated with distinct genomic and epigenetic profiles in posttraumatic stress disorder. Proc Natl Acad Sci. 2013;110:8302–7.

18. Thumfart KM, Jawaid A, Bright K, Flachsmann M, Mansuy IM. Epigenetics of childhood trauma: long term sequelae and potential for treatment. Neurosci Biobehav Rev. 2022;132:1049–66.
19. Woo HI, Lim SW, Myung W, Kim DK, Lee SY. Differentially expressed genes related to major depressive disorder and antidepressant response: genome-wide gene expression analysis. Exp Mol Med. 2018;50:1–11.
20. Mansell G, Gorrie-Stone TJ, Bao Y, Kumari M, Schalkwyk LS, Mill J, et al. Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. BMC Genom. 2019;20:366.
21. Pidsley R, Wong CCY, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genom. 2013;14:293 http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-14-293.
22. Lardenoije R, Roubroeks JAY, Pishva E, Leber M, Wagner H, Iatrou A, et al. Alzheimer's disease-associated (hydroxy)methylomic changes in the brain and blood. Clin Epigenet. 2019;11:164.
23. Jager PLD, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat Neurosci. 2014;17:1156–63. https://www.nature.com/articles/nn.3786.
24. Hannon E, Dempster E, Viana J, Burrage J, Smith AR, Macdonald R, et al. An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. Genome Biol. 2016;17:176.
25. Freytag V, Vukojevic V, Wagner-Thelen H, Milnik A, Vogler C, Leber M, et al. Genetic estimators of DNA methylation provide insights into the molecular basis of polygenic traits. Transl Psychiatry. 2018;8:31.
26. Hunsley J, Elliott K, Therrien Z. The efficacy and effectiveness of psychological treatments for mood, anxiety, and related disorders. Can Psychol. 2014;55:161–76.
27. Hirschtritt ME, Bloch MH, Mathews CA. Obsessive-compulsive disorder: advances in diagnosis and treatment. JAMA. 2017;317:1358–67.
28. Riesel A, Klawohn J, Grützmann R, Kaufmann C, Heinzel S, Bey K, et al. Error-related brain activity as a transdiagnostic endophenotype for obsessive-compulsive disorder, anxiety and substance use disorder. Psychol Med. 2019;49:1207–17.
29. Bey K, Weinhold L, Grützmann R, Heinzel S, Kaufmann C, Klawohn J, et al. The polygenic risk for obsessive-compulsive disorder is associated with the personality trait harm avoidance. Acta Psychiatr Scand. 2020;142:326–36.
30. Wittchen HU, Zaudig M, Fydrich T. Skid. Strukturiertes klinisches Interview f{\"u}r DSM-IV. Achse I und II. Handanweisung. Hogrefe; 1997.
31. Goodman SJ, Burton CL, Butcher DT, Siu MT, Lemire M, Chater-Diehl E, et al. Obsessive-compulsive disorder and attention-deficit/hyperactivity disorder: distinct associations with DNA methylation and genetic variation. J Neurodev Disord. 2020;12:23 https://jneurodevdisorders.biomedcentral.com/articles/10.1186/s11689-020-09324-3.
32. Jacobsen D, Kloss M, Fricke S, Hand I, Moritz S. Reliabilität der deutschen version der yale-brown obsessive compulsive scale. Verhaltenstherapie. 2003;13:111–3.
33. Bey K, Campos-Martin R, Klawohn J, Reuter B, Grützmann R, Riesel A, et al. Hypermethylation of the oxytocin receptor gene (OXTR) in obsessive-compulsive disorder: further evidence for a biomarker of disease and treatment response. Epigenetics. 2022;17:642–52. https://www.tandfonline.com/doi/abs/10.1080/15592294.2021.1943864.
34. Kathmann N, Jacobi T, Elsner B, Reuter B. Effectiveness of individual cognitive-behavioral therapy and predictors of outcome in adult patients with obsessive-compulsive disorder. Psychother Psychosom. 2022;91:123–35.
35. Grützmann R, Klawohn J, Elsner B, Reuter B, Kaufmann C, Riesel A, et al. Error-related activity of the sensorimotor network contributes to the prediction of response to cognitive-behavioral therapy in obsessive–compulsive disorder. Neuroimage Clin. 2022;36:103216.
36. Min JL, Hemani G, Smith GD, Relton C, Suderman M. Meffil: efficient normalization and analysis of very large DNA methylation datasets. Bioinformatics. 2018;34:3983–9. https://pubmed.ncbi.nlm.nih.gov/29931280/.
37. Nordlund J, Bäcklin CL, Wahlberg P, Busche S, Berglund EC, Eloranta ML, et al. Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. Genome Biol. 2013;14:r105.
38. Hackstadt AJ, Hess AM. Filtering for increased power for microarray data analysis. BMC Bioinform. 2009;10:11.
39. Fortin JP, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol. 2014;15:503.
40. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. Plos Genet. 2007;3:e161.
41. Teschendorff AE, Zhuang J, Widschwendter M. Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. Bioinformatics. 2011;27:1496–505. https://academic.oup.com/bioinformatics/article/27/11/1496/217145.
42. Johnson AA, Akman K, Calimport SRG, Wuttke D, Stolzing A, Magalhães JPde. The role of DNA methylation in aging, rejuvenation, and age-related disease. Rejuvenation Res. 2012;15:483–94.
43. Davegårdh C, Wedin EH, Broholm C, Henriksen TI, Pedersen M, Pedersen BK, et al. Sex influences DNA methylation and gene expression in human skeletal muscle myoblasts and myotubes. Stem Cell Res Ther. 2019;10:26.
44. Tsai PC, Glastonbury CA, Eliot MN, Bollepalli S, Yet I, Castillo-Fernandez JE, et al. Smoking induces coordinated DNA methylation and gene expression changes in adipose tissue with consequences for metabolic health. Clin Epigenet. 2018;10:126.
45. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol. 2014;15:R31 http://genomebiology.biomedcentral.com/articles/10.1186/gb-2014-15-2-r31.
46. Pattaro C, Grandi AD, Vitart V, Hayward C, Franke A, Aulchenko YS, et al. A meta-analysis of genome-wide data from five European isolates reveals an association of COL22A1, SYT1, and GABRR2with serum creatinine level. BMC Med Genet. 2010;11:41.
47. Jiang H, Doerge RW. A two-step multiple comparison procedure for a large number of tests and multiple treatments. Stat Appl Genet Mol. 2006;5:Article28.
48. Storey JD. A direct approach to false discovery rates. J R Stat Soc Ser B Stat Methodol. 2002;64:479–98.
49. Yan S, Wu G. WGCNA revisited: module indetification. J Phys Conf Ser. 2021;1955:012108.
50. Zhao W, Langfelder P, Fuller T, Dong J, Li A, Hovarth S. Weighted gene coexpression network analysis: state of the art. J Biopharm Stat. 2010;20:281–300.
51. Choi SW, Mak TSH, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc. 2020;15:2759–72. https://www.nature.com/articles/s41596-020-0353-1.
52. Hüls A, Czamara D. Methodological challenges in constructing DNA methylation risk scores. Epigenetics 2019;15:1–11.
53. Yamagata K, Senokuchi T, Lu M, Takemoto M, Fazlul Karim M, Go C, et al. Voltage-gated K+ channel KCNQ1 regulates insulin secretion in MIN6 β-cell line. Biochem Biophys Res Commun. 2011;407:620–5.
54. Zhou Z, Gong M, Pande A, Lisewski U, Röpke T, Purfürst B, et al. A missense KCNQ1 mutation impairs insulin secretion in neonatal diabetes. Biorxiv. 2021;2021.08.24.457485.
55. Olgar Y, Durak A, Bitirim CV, Tuncay E, Turan B. Insulin acts as an atypical KCNQ1/KCNE1-current activator and reverses long QT in insulin-resistant aged rats by accelerating the ventricular action potential repolarization through affecting the β3-adrenergic receptor signaling pathway. J Cell Physiol. 2022;237:1353–71.
56. Sun Q, Song K, Shen X, Cai Y. The association between KCNQ1 gene polymorphism and type 2 diabetes risk: a meta-analysis. Plos One. 2012;7:e48578.
57. Vliet-Ostaptchouk JV, van, Haeften TW, van, Landman GWD, Reiling E, Kleefstra N, Bilo HJG, et al. Common variants in the type 2 diabetes KCNQ1 gene are associated with impairments in insulin secretion during hyperglycaemic glucose clamp. Plos One. 2012;7:e32148.
58. Been LF, Ralhan S, Wander GS, Mehra NK, Singh J, Mulvihill JJ, et al. Variants in KCNQ1 increase type II diabetes susceptibility in South Asians: a study of 3310 subjects from India and the US. BMC Med Genet. 2011;12:18.
59. Yee NS. Chapter four TRPM8 ion channels as potential cancer biomarker and target in pancreatic cancer. Adv Protein Chem Struct Biol. 2016;104:127–55.
60. Maycox PR, Kelly F, Taylor A, Bates S, Reid J, Logendra R, et al. Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function. Mol Psychiatry. 2009;14:1083–94.
61. Kwiatkowski AV, Rubinson DA, Dent EW, Veen JE, van, Leslie JD, Zhang J, et al. Ena/VASP is required for neuritogenesis in the developing cortex. Neuron. 2007;56:441–55.
62. Yue W, Cheng W, Liu Z, Tang Y, Lu T, Zhang D, et al. Genome-wide DNA methylation analysis in obsessive-compulsive disorder patients. Sci Rep. 2016;6:31333. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4985637/.
63. Schiele MA, Lipovsek J, Schlosser P, Soutschek M, Schratt G, Zaudig M, et al. Epigenome-wide DNA methylation in obsessive-compulsive disorder. Transl Psychiat. 2022;12:221.
64. Bauer-Mehren A, Rautschka M, Sanz F, Furlong LI. DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene–disease networks. Bioinformatics. 2010;26:2924–6.
65. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res. 2017;45:D833–9.
66. Kaur S, Mirza AH, Overgaard AJ, Pociot F, Størling J. A dual systems genetics approach identifies common genes, networks, and pathways for type 1 and 2 diabetes in human islets. Front Genet. 2021;12:630109.
67. Hu Y, Tan LJ, Chen XD, Greenbaum J, Deng HW. Identification of novel variants associated with osteoporosis, type 2 diabetes and potentially pleiotropic loci using pleiotropic cFDR method. Bone. 2018;117:6–14.

68. Tsalamandris S, Antonopoulos AS, Oikonomou E, Papamikroulis GA, Vogiatzi G, Papaioannou S, et al. The role of inflammation in diabetes: current concepts and future perspectives. Eur Cardiol Rev. 2019;14:50–9.
69. Uchida K, Tominaga M. The role of thermosensitive TRP (transient receptor potential) channels in insulin secretion [review]. Endocr J. 2011;58:1021–8.
70. Yagensky O, Kohansal-Nodehi M, Gunaseelan S, Rabe T, Zafar S, Zerr I, et al. Increased expression of heme-binding protein 1 early in Alzheimer's disease is linked to neurotoxicity. Elife. 2019;8:e47498.
71. Devosse T, Dutoit R, Migeotte I, Nadai PD, Imbault V, Communi D, et al. Processing of HEBP1 by Cathepsin D gives rise to F2L, the agonist of formyl peptide receptor 3. J Immunol. 2011;187:1475–85.
72. Uechi L, Jalali M, Wilbur JD, French JL, Jumbe NL, Meaney MJ, et al. Complex genetic dependencies among growth and neurological phenotypes in healthy children: towards deciphering developmental mechanisms. Plos One. 2020;15:e0242684.
73. Grassi G, Figee M, Pozza A, Dell'Osso B. Obsessive-compulsive disorder, insulin signaling and diabetes—a novel form of physical health comorbidity: the sweet compulsive brain. Compr Psychiatry. 2022;117:152329.
74. Horst KW, ter, Lammers NM, Trinko R, Opland DM, Figee M, Ackermans MT, et al. Striatal dopamine regulates systemic glucose metabolism in humans and mice. Sci Transl Med. 2018;10:442.
75. Beattie EC, Carroll RC, Yu X, Morishita W, Yasuda H, Zastrow Mvon, et al. Regulation of AMPA receptor endocytosis by a signaling mechanism shared with LTD. Nat Neurosci. 2000;3:1291–300.
76. van de Vondervoort I, Poelmans G, Aschrafi A, Pauls DL, Buitelaar JK, Glennon JC, et al. An integrated molecular landscape implicates the regulation of dendritic spine formation through insulin-related signalling in obsessive–compulsive disorder. J Psychiatry Neurosci. 2016;41:280–5.
77. Bralten J, Widomska J, Witte WD, Yu D, Mathews CA, Scharf JM, et al. Shared genetic etiology between obsessive-compulsive disorder, obsessive-compulsive symptoms in the population, and insulin signaling. Transl Psychiatry. 2020;10:121.
78. Kvaløy K, Page CM, Holmen TL. Epigenome-wide methylation differences in a group of lean and obese women—a hunt study. Sci Rep. 2018;8:16330.
79. Asahara SI, Etoh H, Inoue H, Teruyama K, Shibutani Y, Ihara Y, et al. Paternal allelic mutation at the Kcnq1 locus reduces pancreatic β-cell mass by epigenetic modification of Cdkn1c. Proc Natl Acad Sci. 2015;112:8332–7.
80. Guo L, Ni Z, Wei G, Cheng W, Huang X, Yue W, et al. Methylation analysis of whole blood cells derived from patients with GAD and OCD in the Chinese han population. Transl Psychiatry. 2022;12:465.

ACKNOWLEDGEMENTS

This work was supported by the German Research Foundation (DFG) grants numbers: [KA815/6-1] to NK, [WA731/10-1], [WA731/15-1] to MW, and [RA1971/8-1], [RA1971/7-1] to AR.

AUTHOR CONTRIBUTIONS

Drafting of the manuscript: RCM, AR, KB, AP, NK, and MW; epigenetic and phenotype data acquisition: BE, BR, and JK.; bioinformatics analysis: RCM. Treatment analysis: KB, and NK.; Study design: AR, MW.; Obtaining funding: AR, MW, NK, and, AP Critical revision of the manuscript: all authors.

FUNDING

Open Access funding enabled and organized by Projekt DEAL.

COMPETING INTERESTS

All authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41380-023-02219-4.

Correspondence and requests for materials should be addressed to Alfredo Ramirez.

Reprints and permission information is available at http://www.nature.com/reprints

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