Epigenetic profile of the immune system associated with symptom severity and treatment response in schizophrenia

Yuanhao Tang, PhD; Yunlong Tan, PhD; Lena Palaniyappan, PhD; Yin Yao, PhD; Qiang Luo, PhD; Yanli Li, PhD

Background: Environmental modification of genetic information (epigenetics) is often invoked to explain interindividual differences in the phenotype of schizophrenia. In clinical practice, such variability is most prominent in the symptom profile and the treatment response. Epigenetic regulation of immune function is of particular interest, given the therapeutic relevance of this mechanism in schizophrenia.

Methods: We analyzed the DNA methylation data of immune-relevant genes in patients with schizophrenia whose disease duration was less than 3 years, with previous lifetime antipsychotic treatment of no more than 2 weeks total. Results: A total of 441 patients met the inclusion criteria. Core symptoms were consistently associated with 206 methylation positions, many of which had previously been implicated in inflammatory responses. Of these, 24 methylation positions were located either in regulatory regions or near the CpG islands of 20 genes, including the SRC gene, which is a key player in glutamatergic signalling. These symptom-associated immune genes were enriched in neuronal development functions, such as neuronal migration and glutamatergic synapse. Compared with using only clinical information (including scores on the Positive and Negative Syndrome Scale), integrating methylation data into the model significantly improved the predictive ability (as indicated by area under the curve) for response to 8 weeks of antipsychotic treatment.

Limitations: We focused on a small number of methylation probes (immune-centred search) and lacked nutritional data and direct brain-based measures.

Conclusion: Epigenetic modifications of the immune system are associated with symptom severity at onset and subsequent treatment response in schizophrenia.

Introduction

Schizophrenia is characterized by its heterogeneity in both clinical symptoms and treatment response. It affects 24 million individuals around the world, resulting in substantially reduced life expectancy and a poor recovery rate.1 Although symptoms generally improve, with reduced variability, after treatment with antipsychotics,2 a substantial number of people do not experience a clinically relevant response in the early treatment period. In exploring the genetic risk of schizophrenia, it has been found that altered inflammatory processes may be reflected in the clinical phenotypes, such as positive or negative valence, affective dysregulation, and cognitive impairment.3 The heterogeneity of the immune response under the influence of epigenetic modification is a valid explanation for the phenotypic heterogeneity of patients with schizophrenia.4

The immune system has been implicated in both the pathogenesis and the treatment outcomes of schizophrenia. In this disease, the severity of clinical symptoms has been associated with both peripheral and central inflammation, and poorer response to antipsychotics has been associated with increased inflammation.5 Although some trials have failed,6 many clinical studies have shown that anti-inflammatory drugs (e.g., celecoxib, acetylsalicylic acid) can improve the treatment effect of antipsychotics,7 with many others still being tested. However, it is still unknown how peripheral immune processes can affect the clinical severity and treatment response in this illness.
As a method of investigating the role of the immune system in schizophrenia, epigenetics, especially DNA methylation, provides a unique opportunity. Epigenome-wide association studies comparing healthy controls with patients who have schizophrenia have identified changes in many methylation markers associated with immune functions, including interleukins (ILs), cluster of differentiation antigens (CDs), and many immune-related genes. Antipsychotic therapy has also been affected by methylation-related changes of schizophrenia-associated genes, including the 5-HT1A receptor gene (HTR1A) and genes in the dopaminergic pathway. Some studies have suggested that an individual’s unique methylation profile may predict the efficacy of antipsychotics.

One pioneering study, albeit with a limited sample size, reported decreased methylation of neurotransmission-related genes in patients who had poor treatment response in schizophrenia. Another study of 98 patients with schizophrenia found that 11 CpG sites associated with inflammatory response were significantly correlated with reality distortion. However, this line of research is still in its infancy. Identifying the association between differential methylation and symptom severity in schizophrenia would be the first step toward the discovery of methylation predictors for treatment outcomes in this disease. Therefore, we hypothesized that the symptom severity of schizophrenia is associated with differential methylation of genes involved in immune functions and that this association could be used to predict treatment outcome.

In a previous study, we first identified 4277 sites with differential methylation between patients with schizophrenia (disease duration < 3 yr) and healthy controls and then investigated their biological significance in the pathogenesis of schizophrenia. However, the correlation between DNA methylation and clinical presentation, as well as treatment response, has not been explored. The DNA methylation and symptom severity data were collected at recruitment, whereas response to antipsychotics was assessed after 8 weeks of treatment (n = 441). In the study reported here, we first tested the associations between DNA methylation of a set of immune-relevant genes and symptom severity at baseline. Next, we used bioinformatic tools to characterize the functional significance of the observed associations. Finally, we built a machine learning model to predict the antipsychotic treatment response at week 8 using the DNA methylation profile at baseline.

Methods

Study sample

The DNA methylation data used in the current study were obtained from 469 Han Chinese patients with schizophrenia recruited for our previous study. These patients were recruited by the Beijing Huilongguan Hospital, Chongqing Three Gorges Central Hospital, and Zhumadian Psychiatry Hospital in 2017 and 2018. The inclusion criteria were age 14–50 years, total disease duration less than 3 years, and previous lifetime antipsychotic treatment of no more than 2 weeks total. Patients with a history of smoking were excluded. The study was approved by the Ethics Committee of Beijing Huilongguan Hospital following the Helsinki Declaration (2004 edition). After receiving a detailed explanation of the study procedures, each participant provided written informed consent. For all patients, schizophrenia was diagnosed according to criteria in the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV), based on consensus within the clinical team providing care.

The severity and characteristics of clinical symptoms of schizophrenia were assessed at baseline by psychiatrists at the Beijing Huilongguan Hospital using the Positive and Negative Syndrome Scale (PANSS). Demographic information (sex, age [in years], education level, and age at first onset of schizophrenia [in years]) was recorded, and blood samples were collected for methylation studies. After 8 weeks of treatment, patients’ medication status and PANSS score were reassessed by the same psychiatrists who carried out the baseline measurements. Of the 469 patients, 441 had complete follow-up information.

Each patient’s schizophrenia medication was recorded, and the daily olanzapine equivalent dose determined, where the equivalent conversion was based on defined daily dose using the conversion scheme developed by the World Health Organization Collaborating Centre for Drug Statistics Methodology. Any patient with a decrease from baseline of more than 30% in the total PANSS score after 8 weeks of antipsychotic treatment was considered to have a treatment response.

Genome-wide methylation data and preprocessing

To measure genome-wide methylation, we used the Infinium MethylationEPIC BeadChip kit (Illumina) with 850000 probes, made available by a case-control epigenome-wide association study through the National Omics Data Encyclopedia (NODE), accession no. OEP001178 (www.biosino.org/node). In total, the data of 469 patients with schizophrenia were downloaded from this database. Quality control and beta mixture quantile dilation (BMIQ) preprocessing operations were implemented using the minfi package (Appendix 1, Figures S1 and S2, available at jpn.ca/lookup/doi/10.1503/jpn.230099/tab-related-content). In addition, cell composition was estimated and corrected using the Reference-Free Adjustment for Cell Type Composition (ReFACTor) algorithm in the GLINT software. Sex, age, education, illness course, and age at onset were also used as covariates for methylation analysis. The effect of this correction is shown in Appendix 1, Figure S3.

Selection of immune-related DNA methylation probes

In our previous epigenome-wide association study of schizophrenia (n = 945), these immune-related probes did not reach the epigenome-wide significance level, but such probes were reported in a meta-analysis with a much larger sample.
size (n = 4483). To increase the statistical power of the current study, we restricted our search to the immune-related genes. We downloaded a list of genes that have been associated with immune functions from the ImmPort database (https://www.immport.org), which contains 1793 genes, categorized according to 9 functions: antigen processing and presentation (146 genes), antimicrobial functions (494 genes), B-cell antigen receptor signalling pathway (266 genes), chemokines (42 genes), chemokine receptors (37 genes), cytokines (301 genes), cytokine receptors (241 genes), natural killer cell cytotoxicity (41 genes), and T-cell receptor signalling pathway (225 genes). Then, according to the platform annotation information (https://static-content.springer.com/esm/art%3A10.1038%2Fs41380-020-00968-0/MediaObjects/41380_2020_968_MOESM9_ESM.xls), we selected 18985 methylation probes corresponding to these genes as research objects.

**Statistical analysis**

**Association analysis**

We used linear regression models to determine the association between DNA methylation and symptom severity. Here, we considered 4 PANSS scores: the positive, negative, and general subscale scores, and the total score. Similar to what was done for the methylation data, we adjusted for the essential covariates, specifically sex, age, education level, duration of illness, age at first onset, and cell type composition. Patients with a history of smoking were excluded, so we did not consider smoking as a covariate. All participants were Han Chinese; as such, there was no population stratification in the samples through principal component analysis of methylation, and a population correction was not required (Appendix 1, Figure S4).

We used the DMPfinder function in the minfi package to build a regression model of immune-relevant DNA methylation for each symptom score separately.

Next, we tested the associations between the associated methylation probes (AMPs) and treatment response (designated as 0 for nonresponse and 1 for response) by linear regression models. In addition to the covariates considered in the model described above, this model for treatment response included extra covariates relevant to the treatment, specifically, the antipsychotic type, the dose equivalent, and the baseline PANSS total score (week 0, before treatment). Significant associations were identified after a false discovery rate correction for multiple comparisons (adjusted p value < 0.05).

**Enrichment analysis**

We performed enrichment analysis with the Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA GWAS) database (https://fuma.ctglab.nl). Gene sets of cell types were designated according to single-cell sequencing of a human brain, including representative gene sets for 56 cell types.

**Random forest models for predicting patient response to drug treatment**

We built a random forest model using demographic information and disease variables to predict treatment response. The area under the curve (AUC) of the model was established by a 5-fold cross-validation. The mlr3 package (machine learning in R) was used to build the models. We included preselected confounding factors for treatment response, following Carbon and Correll:21 patient variables (education level and sex), disease variables (age at onset, duration of disease, PANSS subscale, and total scores at baseline), and treatment variables (type and dose of antipsychotics). Additionally, because patient age was significantly associated with drug response in our sample, we included age as a predictor in this model.

Next, we tested whether including the significant AMPs identified above could improve the prediction accuracy of the model. The significance of this improvement was assessed by the net reclassification index (NRI). A significant value for this index would suggest that including the AMPs significantly increased the AUC of the model. To illustrate the significance of AMPs for drug response, independent of baseline total PANSS, we also built a model based on demographic information and candidate AMPs.

Finally, we used the randomForestExplainer package of R to compute interactions between drug categories and other variables. To further validate the interaction, we established 3 logistic regression models with interaction terms, using the drugs with the largest sample sizes (> 90) — risperidone, aripiprazole, and olanzapine — as dummy variables, and selected the top 30 methylated sites with the strongest interaction effects in the random forest model. We tested the significance of the interaction terms (adjusted p value < 0.05).

**Results**

**Clinical characteristics**

Of the 441 patients included in this study, 242 had a response to treatment (mean age ± standard deviation 26.64 ± 6.64 yr, 93 males, 149 females) and 199 had no response (mean age 25.05 ± 5.68 yr, 85 males, 114 females) after 8 weeks of antipsychotic treatment (Table 1). Compared with the nonresponse group, those in the response group were older (t = 2.01, p = 0.045), had higher PANSS scores at week 0 (t = 9.14, p < 0.001), and had lower PANSS scores at week 8 (t = -13.13, p < 0.001). Significant differences between the 2 groups were observed for overall drug use (p = 0.038) and ziprasidone (p = 0.004), with no significant differences in drug dosage (Table 1).
Symptom association with immune-relevant methylation

Considering attrition in follow-up, we had 747,372 methylation probes with good-quality methylation data for the 441 patients. Of the 18,985 probes corresponding to immune-related genes that were selected for our analysis, we found that 6,575 were associated with the PANSS total score after correction for false discovery rate (Figure 1A). In terms of the PANSS subscales, we found that 1,464, 946, and 4,697 probes were associated with the positive, negative, and general subscale scores, respectively (Figure 1B–1D). Among these associations, 206 were shared by all 4 symptom scores (Figure 1E and 1F). These AMPs were located on a total of 158 genes, among which both increased methylation of 105 genes and decreased methylation of 99 genes were associated with symptom severity (Appendix 1, Table S1). As expected, 75.5% of AMPs were distributed in nonregulatory regions (i.e., body) of the genes. Compared with the AMPs that were negatively associated with the PANSS total score, the positively associated AMPs were located more in regulatory regions (nonbody) of the genes ($\chi^2 = 1, p = 0.004$; Figure 1G) and were more abundant near CpG islands ($\chi^2 = 1, p = 0.024$; Figure 1H). The same patterns were observed for the 3 PANSS subscales (Appendix 1, Figures S5–S8).

We found that 24 of the 206 AMPs were located either in regulatory regions or near CpG islands of 20 genes (Table 2). We also identified several new AMPs on genes that are involved in the development of nerves or synapses, including SEMA4B, SRC, CXCR4, FGF11, and NDRG1.

Associated genes enriched in both immune function and neural development

Enrichment analysis was performed on the 158 genes related to the 206 AMPs found to be associated with all 4 symptom scores, as described above. This analysis identified a total of 28 enriched gene ontology (GO) biological process (BP) terms (Figure 2A) and 17 enriched GO cellular component (CC) terms (Figure 2B), with adjusted $p$ values less than 0.05. The complete enrichment results are shown in Appendix 1, Table S2. Although our study focused on immune-related genes, we obtained enrichment for 2 neurologically related GO terms: neuron migration among the BP terms and glutamate synapse among the CC terms. In other enriched

Table 1: Characteristics of study groups

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>With response</th>
<th>No response</th>
<th>Statistical test result†</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex, no. male/female</td>
<td>93/149</td>
<td>85/114</td>
<td>$\chi^2 = 0.66$</td>
<td>0.42</td>
</tr>
<tr>
<td>Age, yr</td>
<td>26.64 ± 6.64</td>
<td>25.05 ± 5.68</td>
<td>$t_{438} = 2.01$</td>
<td>0.045</td>
</tr>
<tr>
<td>Education level, yr</td>
<td>11.52 ± 2.32</td>
<td>11.33 ± 2.37</td>
<td>$t_{438} = 0.85$</td>
<td>0.40</td>
</tr>
<tr>
<td>Age at first onset, yr</td>
<td>25.43 ± 6.51</td>
<td>24.41 ± 5.70</td>
<td>$t_{438} = 1.76$</td>
<td>0.08</td>
</tr>
<tr>
<td>Duration of disease, yr</td>
<td>7.11 ± 5.82</td>
<td>6.90 ± 6.77</td>
<td>$t_{438} = 0.35$</td>
<td>0.73</td>
</tr>
<tr>
<td>PANSS at week 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>88.43 ± 12.99</td>
<td>77.37 ± 12.356</td>
<td>$t_{438} = 9.14$</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Positive</td>
<td>25.13 ± 4.89</td>
<td>22.33 ± 4.94</td>
<td>$t_{438} = 5.96$</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Negative</td>
<td>21.28 ± 6.50</td>
<td>19.31 ± 5.96</td>
<td>$t_{438} = 3.31$</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>General</td>
<td>42.02 ± 7.46</td>
<td>35.73 ± 6.56</td>
<td>$t_{438} = 9.40$</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>PANSS at week 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>49.73 ± 8.84</td>
<td>63.89 ± 12.94</td>
<td>$t_{438} = −13.13$</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Positive</td>
<td>11.35 ± 3.59</td>
<td>16.01 ± 5.59</td>
<td>$t_{438} = −10.16$</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Negative</td>
<td>14.08 ± 4.58</td>
<td>18.43 ± 4.96</td>
<td>$t_{438} = −9.49$</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>General</td>
<td>24.30 ± 4.20</td>
<td>29.46 ± 6.61</td>
<td>$t_{438} = −9.54$</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Drug, no. of patients‡</td>
<td></td>
<td></td>
<td>$\chi^2 = 13.32$</td>
<td>0.038</td>
</tr>
<tr>
<td>Aripiprazole</td>
<td>50</td>
<td>41</td>
<td>$\chi^2 = 0.00$</td>
<td>0.96</td>
</tr>
<tr>
<td>Fluphenazine</td>
<td>11</td>
<td>15</td>
<td>$\chi^2 = 1.48$</td>
<td>0.22</td>
</tr>
<tr>
<td>Haloperidol</td>
<td>9</td>
<td>8</td>
<td>$\chi^2 = 0.00$</td>
<td>&gt; 0.99</td>
</tr>
<tr>
<td>Olanzapine</td>
<td>46</td>
<td>48</td>
<td>$\chi^2 = 1.90$</td>
<td>0.17</td>
</tr>
<tr>
<td>Quetiapine</td>
<td>31</td>
<td>30</td>
<td>$\chi^2 = 0.48$</td>
<td>0.49</td>
</tr>
<tr>
<td>Risperidone</td>
<td>54</td>
<td>36</td>
<td>$\chi^2 = 0.64$</td>
<td>0.42</td>
</tr>
<tr>
<td>Ziprasidone</td>
<td>38</td>
<td>12</td>
<td>$\chi^2 = 8.53$</td>
<td>0.004</td>
</tr>
<tr>
<td>Olanzapine dose equivalent, mg/d§</td>
<td>15.14 ± 7.15</td>
<td>14.72 ± 7.55</td>
<td>$t_{438} = 0.59$</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Note: PANSS = Positive and Negative Syndrome Scale; SD = standard deviation.
†Differences in baseline variables, PANSS score, and medication information were compared between responder and nonresponder groups.
‡For each group, the sum of values is less than the total number of patients in the group because data were omitted for drugs with small numbers of patients.
§Olanzapine dose equivalent determined with conversion scheme of World Health Organization Collaborating Centre for Drug Statistics Methodology. 17
Figure 1: Epigenome-wide association studies of symptom severity in patients with schizophrenia. Volcano plots were used to illustrate associations with (A) total, (B) positive, (C) negative, and (D) general scores on the Positive and Negative Syndrome Scale (PANSS), where blue indicates a negative association, red a positive association, and grey a nonsignificant (NS) association; dashed line represents $-\log_{10}(0.05)$. The curves on the horizontal and vertical coordinates were probability density distributions of the $\Delta\beta$ and $-\log_{10}$ (adjusted $p$). (E) Venn plot for the associated methylation probes (AMPs) of the 4 symptom scores. (F) Chord diagram, where the coordinates represent the number of genes, the coloured outer circle represents all genes associated with the 4 symptom scores, and its length represents the number of genes. The functional genomic regions (G) and the locations relative to CpG islands (H) are shown for the AMPs associated with all 4 symptom scores. BCR = B-cell receptor; TCR = T-cell receptor; TSS1500 = 200–1500 bases upstream of the transcriptional start site; TSS200 = 0–200 bases upstream of the transcriptional start site; 5’ UTR = sequence between transcriptional start site and ATG start site; 1st exon = first exon; body = sequence between the ATG and stop codon; 3’ UTR = sequence between the stop codon and poly A; N_Shelf = 0–2 kb upstream of island; OpenSea = more than 4 kb from a CpG island; S_Shelf = 2–4 kb downstream of island; S_Shore = 0–2 kb downstream of island.
results of BP terms, we found many functions related to blood vessels, such as vascular development, tube morphogenesis, tube development, and blood vessel morphogenesis. Furthermore, cell-type enrichment revealed that the 105 genes with increased methylation had specific expression patterns in microglia of the brain (Figure 3A). Similarly, tissue-specific enrichment results showed that these genes were associated with gene expression patterns in the brain, particularly the prefrontal lobe, the anterior cingulate cortex, and the striatum (Figure 3B).

The protein–protein interaction network suggested that these genes were interacting closely with one another (Figure 4). Both the SRC and LCK genes mentioned above were also hubs of the interaction network. Another hub protein, RAC1, has been associated with schizophrenia, showing hypomethylation in patients with this disease. Dysregulation of LI-10 and CDC42 proteins and their respective families has also been associated with a risk of schizophrenia.

**Improved prediction of treatment response with methylation**

The treatment responses (change scores) among all PANSS domains were highly correlated, as has been shown repeatedly across many trials. This feature, called pseudospecificity of treatment response, speaks to the fact that antipsychotics have a broad-spectrum effect on all symptom domains. As a result, we hypothesized that response-relevant immune genes represent a common process. We expected the aberrant immune response or inflammation to have a multifaceted effect on brain function (i.e., affecting all 4 PANSS scores), and the methylation sites associated with all 4 scores are key to this process. Treatment of schizophrenia aims to ameliorate the effects brought about by these methylation patterns. Using a univariate approach, we found that none of the probes were significantly associated with treatment response. However, using a multivariate approach (i.e., the random forest model) with 5-fold cross-validation, we found that including the 206 AMPs identified above significantly improved prediction accuracy, from 0.66 to 0.71 (NRI = 0.32, p = 0.003; Figures 5A and 5B). The model using demographic information and AMPs achieved an AUC of 0.68, further supporting the correlation between AMPs and drug response (Figure 5B). According to the feature importance score of the random forest model (Appendix 1, Table S3), the top 20 features were the 18 methylation probes, the PANSS total score, and an item score (PANSS G2, i.e., anxiety) from the PANSS general subscale (Figure 5C). Results of the gene enrichment analysis using the ranking of methylation importance are shown in Appendix 1, Figures S9 and S10.

**Table 2: Intersection set associated methylation probes (AMPs) located in important genomic function and CpG regions**

<table>
<thead>
<tr>
<th>Probe ID†</th>
<th>RefGene</th>
<th>Status‡</th>
<th>CHR</th>
<th>Position</th>
<th>Region</th>
<th>CpG content</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg00110171</td>
<td>NFATC1</td>
<td>Positive</td>
<td>18</td>
<td>35668450</td>
<td>5'UTR</td>
<td>N_Shore</td>
</tr>
<tr>
<td>cg01141721</td>
<td>SRC</td>
<td>Positive</td>
<td>20</td>
<td>51703444</td>
<td>5'UTR</td>
<td>N_Shore</td>
</tr>
<tr>
<td>cg01525376</td>
<td>LCK</td>
<td>Negative</td>
<td>1</td>
<td>71685381</td>
<td>TSS1500</td>
<td>S_Shelf</td>
</tr>
<tr>
<td>cg05417950</td>
<td>NUDT6</td>
<td>Positive</td>
<td>4</td>
<td>37734310</td>
<td>5'UTR</td>
<td>N_Shore</td>
</tr>
<tr>
<td>cg05904013</td>
<td>IRF5</td>
<td>Positive</td>
<td>7</td>
<td>67128314</td>
<td>5'UTR</td>
<td>S_Shore</td>
</tr>
<tr>
<td>cg09722826</td>
<td>CYLD</td>
<td>Positive</td>
<td>16</td>
<td>60645748</td>
<td>5'UTR</td>
<td>S_Shore</td>
</tr>
<tr>
<td>cg10865498</td>
<td>SRC</td>
<td>Positive</td>
<td>20</td>
<td>46705343</td>
<td>5'UTR</td>
<td>S_Shore</td>
</tr>
<tr>
<td>cg11218434</td>
<td>TPT1</td>
<td>Negative</td>
<td>13</td>
<td>20788391</td>
<td>TSS200</td>
<td>N_Shelf</td>
</tr>
<tr>
<td>cg11804928</td>
<td>THRA</td>
<td>Positive</td>
<td>17</td>
<td>26600475</td>
<td>5'UTR</td>
<td>S_Shore</td>
</tr>
<tr>
<td>cg12595667</td>
<td>CXCRI4</td>
<td>Positive</td>
<td>2</td>
<td>45761310</td>
<td>3'UTR</td>
<td>N_Shore</td>
</tr>
<tr>
<td>cg12948771</td>
<td>LTBR</td>
<td>Positive</td>
<td>14</td>
<td>47718318</td>
<td>5'UTR</td>
<td>N_Shore</td>
</tr>
<tr>
<td>cg13492227</td>
<td>FGFI1</td>
<td>Positive</td>
<td>17</td>
<td>58742439</td>
<td>TSS1500</td>
<td>N_Shore</td>
</tr>
<tr>
<td>cg14349538</td>
<td>IRF5</td>
<td>Positive</td>
<td>7</td>
<td>11637376</td>
<td>TSS1500</td>
<td>N_Shore</td>
</tr>
<tr>
<td>cg14424363</td>
<td>IKBKB</td>
<td>Positive</td>
<td>8</td>
<td>14798410</td>
<td>TSS1500</td>
<td>N_Shore</td>
</tr>
<tr>
<td>cg14614416</td>
<td>TNFRS9</td>
<td>Negative</td>
<td>1</td>
<td>40634465</td>
<td>TSS1500</td>
<td>N_Shore</td>
</tr>
<tr>
<td>cg17356733</td>
<td>IFNGR2</td>
<td>Positive</td>
<td>21</td>
<td>32691445</td>
<td>TSS1500</td>
<td>N_Shore</td>
</tr>
<tr>
<td>cg18037033</td>
<td>IL12B</td>
<td>Negative</td>
<td>5</td>
<td>16734379</td>
<td>1st Exon</td>
<td>N_Shore</td>
</tr>
<tr>
<td>cg20100745</td>
<td>NDRG1</td>
<td>Positive</td>
<td>8</td>
<td>14744315</td>
<td>5'UTR</td>
<td>N_Shore</td>
</tr>
<tr>
<td>cg2166213</td>
<td>IFITM1</td>
<td>Negative</td>
<td>11</td>
<td>30726412</td>
<td>3'UTR</td>
<td>N_Shore</td>
</tr>
<tr>
<td>cg22595920</td>
<td>PIK3CD</td>
<td>Positive</td>
<td>1</td>
<td>26603349</td>
<td>5'UTR</td>
<td>S_Shelf</td>
</tr>
<tr>
<td>cg23374992</td>
<td>CXCR4</td>
<td>Positive</td>
<td>2</td>
<td>17607400</td>
<td>3'UTR</td>
<td>N_Shelf</td>
</tr>
<tr>
<td>cg24182521</td>
<td>PLTP</td>
<td>Positive</td>
<td>20</td>
<td>39709366</td>
<td>TSS1500</td>
<td>S_Shore</td>
</tr>
<tr>
<td>cg25913761</td>
<td>SEMA4B</td>
<td>Negative</td>
<td>15</td>
<td>55701356</td>
<td>TSS1500</td>
<td>N_Shore</td>
</tr>
<tr>
<td>cg26955540</td>
<td>SEMA4B</td>
<td>Negative</td>
<td>15</td>
<td>50803471</td>
<td>TSS1500</td>
<td>N_Shore</td>
</tr>
</tbody>
</table>

Note: CHR = chromosome number.

*All of the 206 intersection set AMPs are located in both the non-body and the non-open-sea regions.
†Probe ID corresponds to the CpG site, where each probe on the methylation chip represents identification of the specified CpG site.
‡“Status” refers to methylation status, where “positive” means a positive association with symptom severity and “negative” means a negative association with symptom severity.
Figure 2: Functional characteristics of methylation probes that were associated with all 4 symptom types (negative, positive, general, and total) simultaneously. The size of each data point indicates the relative number of genes contained in the item, with larger points indicating more genes. (A) Enrichment of gene ontology (GO) biological process terms. (B) Enrichment of GO cellular component terms. Labels shown in red represent functions related to neurons and synapses, which are of importance in schizophrenia.
Figure 3: Functional characteristics of methylation probes that were associated with all 4 symptoms (negative, positive, general, and total) simultaneously. The size of each data point indicates the relative number of genes contained in the item, with larger points indicating more genes; red indicates brain-related tissues or cell types with significant enrichment; blue indicates non–brain-related tissues or cell types with significant enrichment. (A) Cell type enrichment analysis. (B) Tissue type enrichment analysis. adjP = adjusted p value; EBV = Epstein–Barr virus.
Figure 4: Protein–protein interaction network. Upward-pointing triangles indicate a positive association, downward-pointing triangles indicate a negative association, and circles indicate that both positive and negative associations were found in the same gene. BCR = B-cell receptor; TCR = T-cell receptor.

To validate the importance of the 24 core sites located in regulatory regions and near CpG islands, we employed random forest modelling to predict drug responses by combining these 24 sites with clinical data. As a comparison, we randomly selected the same number of sites (i.e., n = 24) from all 206 sites and conducted the same modelling and prediction process, repeating it 100 times. The results showed that in only 3 of 100 instances did the AUC of the random sites surpass that of the core sites (p = 0.03).

Using the randomForestExplainer package, we detected 30 possible interaction effects between antipsychotics (aripiprazole, olanzapine, and risperidone) and methylation on the treatment response (Appendix 1, Figure S11). Using the logistic regression model, we found that methylation probes cg24591861 and cg17382048 exhibited significant interactions with all 3 drugs, whereas cg12750431 showed significant interactions with aripiprazole and olanzapine (Appendix 1, Figure S12). Probes cg24591861, cg17382048, and cg12750431 are located in the MAP2K1, CXCR5, and CXCL12 genes, respectively. The MAP2K1 gene is part of the mitogen-activated protein (MAP) kinase signalling pathway. It has been reported that membrane receptors targeted by antipsychotics activate the MAP kinase signalling pathway, which is an important mechanism of activity for these drugs.27 CXCR5 and CXCL1 are, respectively, a chemokine receptor and a chemokine. Research has shown that risperidone induces neutrophil apoptosis by modulating chemokine release.28 Our results suggest that aripiprazole and olanzapine may have a similar mechanism of action.

Discussion

To our knowledge, the current study is the first to uncover the relationship between DNA methylation of immune-relevant genes and symptom severity among Chinese Han patients with schizophrenia with a disease duration of less
Figure 5: Predictions of response to antipsychotic treatment using the methylation profile at baseline. Comparisons of (A) area under the curve (AUC) and (B) receiver operating characteristic (ROC) curves for the 3 random forest models, including the PMRF model built from scores on the Positive and Negative Syndrome Scale (PANSS) and baseline demographic information. The PMRF model additionally included the 206 associated methylation probes (AMPs) shared by all 4 symptom scores, and the MRF model was built from demographic information and AMPs. (C) Top 20 features of the MRF model, ranked by importance score. MRF = methylation random forest model; NRI = net reclassification index; PMRF = PANSS and methylation random forest model; PRF = PANSS random forest model.
than 3 years. The study was based on methylation in peripheral blood samples, an approach widely used as a surrogate measure for methylation in the brain. There is evidence of a significant correlation between blood and brain in terms of tissue methylation. The similarity could be attributed to DNA methylation being a conserved process that regulates similar genes in various tissues with shared environmental factors, common cell types, common diseases, and common genetic influences. Also, cytokines can cross the blood–brain barrier and interact with the peripheral and central immune systems. Additionally, the inflammatory response can influence the synthesis, metabolism, and function of dopamine.

Our current findings need validation in future methylation studies using brain tissues. Among 18,985 immune-relevant probes, we found 206 probes that were associated with all 4 PANSS scales. Among these, 24 probes were located both in regulatory regions of genes and near CpG islands. Regulatory regions encompass first exon, 3'UTR, 5'UTR, TSS200, and TSS1500 sequences, and methylation within these sequences often affects the binding of transcriptional complexes. Methylation sites located near CpG islands (specifically island, N shore, N shelf, S shore, and S shelf) are more likely to form continuous methylation within the region, and their impact on transcription as a group is greater than the impact at individual sites. Therefore, these 24 AMPs are likely to have important regulatory effects. Previous literature has already shown that many of these genes are associated with schizophrenia, including NFATC1, LCK, IRF5, IFTMT1, and PIK3CD.

We also identified several new AMPs on genes that are involved in the development of nerves or synapses. For example, the SEMA4B gene affects both chemorepellent activity and semaphorin receptor binding activity and is associated with synaptic elongation. The SRC gene is the key player in various signalling mechanisms that affect the phosphorylation of receptor subunit 2 (GluN2), and low expression of sarcoma tyrosine kinase (SRC) will affect N-methyl-D-aspartate (NMDA) receptor signalling. The CXCR4 gene regulates both migration and regional distribution of cortical interneurons, with evidence that the promoter region of CXCR4 is hypermethylated in patients who have schizophrenia. The FGF11 and NDRG1 genes are associated with neurodevelopment. The numerous AMPs located within the gene body may also have potential biological significance. Although the biological significance of these methylation sites is not known for certain, they may be involved in positive regulation of gene transcription and alternative splicing, among other processes. Finally, using a multivariate approach, the baseline methylation profiles significantly improved the accuracy of predicting patients' responses to antipsychotic treatments. These findings provide new epigenetic evidence for the immune hypothesis of schizophrenia and have revealed the potential of epigenetic tools for monitoring symptoms and predicting treatment response in schizophrenia.

The current finding of significant associations between symptom burden and immune-relevant methylation adds new epigenetic evidence to the immune hypothesis of schizophrenia. It has been hypothesized that inflammatory dysregulation, including chronic macrophage and T-cell activation, imbalance of T helper cells (Th1 and Th2), and enhanced microglial activation, can disrupt neurotransmitters, synaptic plasticity, and cortisol concentrations, which in turn may contribute to the burden of schizophrenia. Indeed, in our study on DNA methylation of immune-related genes, we identified numerous important loci where genes are involved in regulation of the T-cell receptor signalling pathway, cellular immunity, and other related functions. With regard to the potential role of Th1/Th2 imbalance in schizophrenia, we found in the current study that the Th1 regulatory-related genes, including SLClIA1, TNFSF4, IL27, and IL1R1, had significant associations with symptom severity, whereas Th2-related genes were not directly implicated. In the absence of functional markers, we could not determine if disruption in the Th1 regulatory pathway is related to Th1/Th2 imbalance, but our findings provide new empiric evidence for the role of Th1 in treatment response. However, not all of the case-control epigenetic studies of schizophrenia have identified methylation probes associated with immune-relevant genes. This may be because other studies sought disease associations, not symptom-level relations explaining phenotype variability, given our current finding of new epigenetic evidence of symptom-associated methylation in cytokine-, cytokine receptor-, and chemokine-related genes. These findings are also supported by previous reports of the association of more severe symptoms of schizophrenia with both higher concentrations of plasma IL-6, IL-8, and IL-17α and enhanced expression of the chemokines MCP-1 and CCL11.

In the GO enrichment analysis, we used the selected immune-related genes as background, so immune-related functions did not appear in the results. We noticed that many vascular-related terms, such as vascular development, tube morphogenesis, tube development, and blood vessel morphogenesis, were significantly enriched in the BP category. It is possible that abnormalities in the regulatory functions of the nervous system may be due to the presence of DNA methylation during embryonic development in affected patients, which could lead to developmental abnormalities in the neural tube. However, we cannot confirm if this pattern of DNA methylation was a feature of the embryonic period in affected individuals. Nonetheless, convincing evidence exists to support a significantly increased risk of cardiovascular disease among individuals with schizophrenia, and our findings lend support to the relevance of vascular pathogenesis for schizophrenia from the perspective of DNA methylation. Notably, the symptom-associating and immune-relevant genes identified in the current study were also enriched in neural terms. The glutamate synapse term was particularly noteworthy. Glutamatergic dysregulation has been hypothesized as a key pathogenic mechanism of schizophrenia. More specifically, hypoactivity of SRC can lead to hypofunction of NMDA receptors in mice models of schizophrenia, consistent with postmortem brain studies from human patients. Our finding of an association between hypermethylation of the cg0114721 probe and more severe symptoms generates a new hypothesis of epigenetic dysregulation of SRC in schizophrenia. This probe is located in both the 5'UTR region of SRC and the N shore of the CpG.
island, suggesting its regulatory role in the activity of SRC. Another notable gene is C-X-C chemokine receptor type 4 (CXCR4), which is also involved in regulating NMDA receptor signalling.\textsuperscript{34} Our finding that hypermethylation of this gene was associated with more severe symptoms supports reports of below-normal CXCR4 expression and a hypermethylated promoter region in patients with schizophrenia.\textsuperscript{42}

Our findings have shown the potential of using a multivariate methylation profile to predict patients’ responses to antipsychotic treatment in schizophrenia. Each of these methylation sites was associated with treatment response with small effect size, but collectively they significantly improved the prediction accuracy in our sample of 441 patients with schizophrenia. Among the 24 core methylation sites located in regulatory regions, 3 sites (cg11218434, cg09722826) ranked within the top 10 in terms of importance. We noticed that the top few AMPs with strong contributions to the model included genes for several pro-inflammatory molecules: IL12B,\textsuperscript{55} IL27,\textsuperscript{56} S100A12,\textsuperscript{57} and ZAP70.\textsuperscript{58} Hypermethylation of these genes was positively associated with better treatment response. At the same time, HTR3A\textsuperscript{60} and AP3B1\textsuperscript{60} also had high rankings. These 2 genes contribute to the normal transmission of neural signals, and their methylation levels were negatively associated with treatment response. In addition, there have been studies indicating that the gene polymorphism of HTR3A is related to a patient’s response to antipsychotics,\textsuperscript{60} which also supports our conclusion. The evidence outlined above may suggest that lower levels of inflammation and a healthier level of neuronal signalling may predict superior therapeutic effects. Interestingly, the type of antipsychotics used seemed to be insignificant in terms of the prediction model. These findings can be taken to suggest that agents targeting the immune genes (e.g., anti-inflammatory drugs), as implicated in our multivariate methylation profile, may improve the efficacy of antipsychotic treatment. Some support comes from prior trials demonstrating that anti-inflammatory drugs such as celecoxib and acetylsalicylic acid improve the treatment effect of antipsychotics.\textsuperscript{7} The multivariate methylation profile established by our study suggests new candidate genes for drug targets.

Limitations

Our study had some limitations. The drug response rate, about 55%, was relatively low within our sample, probably because we defined response as a 30% reduction in PANSS total score, instead of the 20% threshold used in some other studies.\textsuperscript{61} Although methylation in brain tissues is strongly correlated with methylation in peripheral blood cells, the neurobiologic implications of peripheral measurements require further validation. We did not collect information about patients’ nutritional status, which could be a confounding factor influencing the results.\textsuperscript{62} DNA methylation probes represent a small proportion of the functional methylation sites in the genome and cannot be assumed to fully represent methylation at all relevant sites.

Conclusion

Epigenetic analyses of patients with schizophrenia who had disease duration less than 3 years showed associations between methylation of immune-relevant genes and symptom severity. The multivariate immune-centred methylation profile significantly improved the prediction of response to antipsychotic treatment.

Acknowledgement: The authors express appreciation for Professor Yun Liu’s valuable suggestions for this study.

Affiliations: From the National Clinical Research Center for Aging and Medicine at Huashan Hospital, State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Institutes of Brain Science and School of Life Sciences, Human Phenome Institute, Fudan University, Shanghai, China (Yunlong Tan); the Peking University Huilongguan Clinical Medical School, Beijing Huilingguan Hospital, Beijing, China (Tang, Yao); the Douglas Mental Health University Institute, Department of Psychiatry, McGill University, Montreal, Que. (Palaniyappan); Robarts Research Institute and Department of Medical Biophysics, Western University, London, Ont. (Palaniyappan); the Lawson Health Research Institute, London, Ont. (Palaniyappan); the MOE Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China (Luo).

Competing interests: Lena Palaniyappan has received grants from the Graham Boeckh Foundation (Douglas Research Centre, McGill University) and the Canadian Institutes of Health Research; royalties from Oxford University Press and SPMM Course, UK; and speaker’s honoraria from Janssen Canada and Otsuka Canada. He is the co-editor in chief for the Journal of Psychiatry and Neuroscience; he had no role in editorial decision-making for this article. No other competing interests were declared.

Contributors: Yunlong Tan, Yin Yao, Qiang Luo, and Yanli Li contributed to the conception and design of the project. Yunlong Tan, Yin Yao, and Yanli Li provided the experimental materials and collected the data. Yuanhao Tang and Lena Palaniyappan contributed to analysis and interpretation of the data. Yuanhao Yang, Lena Palaniyappan, and Qiang Luo wrote the manuscript. All of the authors critically reviewed the manuscript, approved the final version to be published, agreed to be accountable for all aspects of the work, and can certify that no other individuals not listed as authors have made substantial contributions to the work.

Funding: This study was partially supported by grants from the National Key Research and Development Program of China (no. 2023YFE0100371), the National Natural Science Foundation of China (nos. 81873909 and 12072236), the Beijing Natural Science Foundation (no. 7212054), the Program of Shanghai Academic Research Leader (no. 23XD1423400), the Shanghai Municipal Science and Technology Major Project (nos. 2018SH2DZX01 and 2021SH2DZX0103), and a start-up capital grant to Yin Yao from the School of Life Sciences of Fudan University and the Beijing Huilingguan Hospital. Lena Palaniyappan received salary support from the Fonds de Recherche du Québec – Santé and research support from the Monique H. Bourgeois Chair in Developmental Disorders. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability: DNA methylation data are accessible through the following URL: http://www.biosino.org/node/project/detail/OEP001178

Code availability: The R code for this study is available at the following URL: https://github.com/Yuanhao-Tang/SZ_severity_methylation.git
Content licence: This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY-NC-ND) licence, which permits use, distribution and reproduction in any medium, provided that the original publication is properly cited, the use is noncommercial (i.e., research or educational use), and no modifications or adaptations are made. See: https://creativecommons.org/licenses/by-nc-nd/4.0/

**References**