Identification of 34 genes conferring genetic and pharmacological risk for the comorbidity of schizophrenia and smoking behaviors

Yunlong Ma¹, Jingjing Li¹, Yi Xu¹, Yan Wang¹, Yinghao Yao¹, Qiang Liu¹, Maiqiu Wang¹, Xinyi Zhao¹, Rongli Fan¹, Jiali Chen¹, Bin Zhang¹, Zhen Cai¹, Haijun Han¹, Zhongli Yang¹, Wenji Yuan¹, Yigang Zhong², Xiangning Chen³, Jennie Z. Ma⁴, Thomas J. Payne⁵, Yizhou Xu², Yuping Ning⁶, Wenyan Cui³, Ming D. Li¹,⁷

¹State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China
²Department of Cardiology, Affiliated Hangzhou First People’s Hospital, Zhejiang University School of Medicine, Hangzhou, China
³Institute of Personalized Medicine, University of Nevada at Las Vegas, Las Vegas, NV 89154, USA
⁴Department of Public Health Sciences, University of Virginia, Charlottesville, VA 22904, USA
⁵Department of Otolaryngology and Communicative Sciences, University of Mississippi Medical Center, Jackson, MS 39216, USA
⁶The Affiliated Brain Hospital of Guangzhou Medical University, Guangzhou, China
⁷Research Center for Air Pollution and Health, Zhejiang University, Hangzhou, China

Correspondence to: Ming D. Li, Wenyan Cui; email: ml2km@zju.edu.cn, cuiwy@zju.edu.cn

Keywords: schizophrenia, tobacco smoking, multi-omics, susceptibility genes, pathways, GWAS

Received: November 11, 2019 Accepted: January 2, 2020 Published: February 3, 2020

Copyright: Ma et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

The prevalence of smoking is significantly higher in persons with schizophrenia (SCZ) than in the general population. However, the biological mechanisms of the comorbidity of smoking and SCZ are largely unknown. This study aimed to reveal shared biological pathways for the two diseases by analyzing data from two genome-wide association studies with a total sample size of 153,898. With pathway-based analysis, we first discovered 18 significantly enriched pathways shared by SCZ and smoking, which were classified into five groups: postsynaptic density, cadherin binding, dendritic spine, long-term depression, and axon guidance. Then, by using an integrative analysis of genetic, epigenetic, and expression data, we found not only 34 critical genes (e.g., PRKZ, ARHGEF3, and CDKN1A) but also various risk-associated SNPs in these genes, which convey susceptibility to the comorbidity of the two disorders. Finally, using both in vivo and in vitro data, we demonstrated that the expression profiles of the 34 genes were significantly altered by multiple psychotropic drugs. Together, this multi-omics study not only reveals target genes for new drugs to treat SCZ but also reveals new insights into the shared genetic vulnerabilities of SCZ and smoking behaviors.

INTRODUCTION

Cigarette smoking is a common brain disorder that is extremely harmful to the individual and society. Smoking prevalence is much higher among individuals with mental disorders than in the general population [1]. In developed countries, the smoking rate among the general population has decreased over recent decades, whereas there has been no decrease among mental health patients. In particular, smoking shows a significant
association with schizophrenia (SCZ) [2], with around 80% of schizophrenic patients being smokers [3, 4]. There are three primary hypotheses intended to elucidate the comorbidity of smoking and SCZ [5–8]. One dominant hypothesis contends that smoking is able to remedy, at least partly, the symptoms of SCZ [6]. There are two main lines of evidence for this hypothesis. One is that nicotine intake enhances the metabolizing of anti-psychotic drugs [7]; the other is that nicotine promotes the release of several neurotransmitters (e.g., dopamine, glutamate, and serotonin) and improves patient performance in memory and attention [4, 5]. The second hypothesis is that given that both diseases are highly influenced by genetics [9, 10], there might exist shared genetic components predisposing to both SCZ and smoking behaviors. Recently, many genetics-based studies have lent support to this hypothesis [11–14]. The third hypothesis is that smoking leads to the onset of SCZ in view of the fact that smoking initiation typically predates the appearance of SCZ [8]. A meta-analysis of cross-sectional and prospective studies reported that daily cigarette smoking was associated with an earlier age of onset of a psychosis [15]. Consistently, a large prospective study of Swedish registry data demonstrated that both light and heavy smoking were highly associated with a greater risk for SCZ [16]. To some extent, these hypotheses are mutually non-exclusive and may collectively contribute to the correlation of SCZ with smoking.

Increasing neuroimaging evidence supports the presence of an association between SCZ and smoking [17]. An fMRI study [18] showed that nicotine can restore deficient sensorimotor gating, which is associated with activation of the limbic regions and striatum in both SCZ patients and healthy controls. Compared with non-smoking SCZ patients and healthy subjects, SCZ patients with concurrent nicotine addiction have reduced grey matter volumes [19]. Furthermore, it is well documented that nicotine increases the release of dopamine, acetylcholine, glutamate, norepinephrine, and serotonin, which are all implicated in the etiology of SCZ [4].

Accumulating evidence has revealed shared genetic components for SCZ and smoking [11–14]. One of the most promising findings is that variants in the CHRNA5/A3/B4 cluster on chromosome 15q24 are associated not only with nicotine dependence (ND) [12] but also with SCZ [11]. By using experimental evidence from animal models of addiction and SCZ, Koukouli et al. [20] found that in rodents, nicotine addiction-associated polymorphisms in CHRNA5 provoke a decrease in neuronal activity, which mirrors the hypofrontality detected in SCZ or addictive patients. Furthermore, there were multiple susceptibility genes reported to be associated with both SCZ and smoking risk, including DRD2 [11, 21–23], BDNF [24], and COMT [25, 26], to name a few. As we know, psychiatric disorders including SCZ are highly comorbid diseases [27–30]; thus there exists a great comorbidity on the relations between SCZ and substance addictions [31, 32]. For example, a good number of reports [33–37] have concentrated on the genetic effect of the Val158Met polymorphism in COMT on the comorbidity of SCZ and substance addictions. In addition, Chen et al. [13] reported that ND is positively correlated with a polygenic risk score for SCZ whereas SCZ is positively correlated with the polygenic risk score for cotinine concentration. A recent study [38] revealed a statistically significant genetic correlation between SCZ and several smoking-related phenotypes.

To make further progress in the prevention and treatment of SCZ and smoking, it is essential to identify the etiologic biological pathways and susceptibility genes underlying the comorbidity of both disorders. Earlier genetics association studies or pathway-based studies focused primarily on either SCZ or smoking [39–42]. To date, only two reports [13, 14] investigated the genetic relations between SCZ and smoking based on pathway analysis results for a limited number of candidates or significant genes. To the best our knowledge, there has been no study providing an integrative genomics analysis based on multi-omics data and biological pathways for both SCZ and smoking. Therefore, the primary objective of this study was to identify susceptibility SNPs, genes, and pathways for the comorbidity of SCZ and smoking with the use of multi-omics data from various sources (Supplementary Figure 1).

RESULTS

GWAS-based enrichment analysis for SCZ and smoking behaviors

Our pathway analysis of GWAS summary statistics on SCZ and smoking phenotypes revealed 175, 172, 233, 225, and 158 significantly enriched pathways (q value < 0.1) for SCZ, CPD, ever smoking, former smoking, and age at smoking initiation, respectively. There were 84 pathways with a q value < 0.1 for SCZ that were in common with that for at least one smoking phenotype (Figure 1A and Supplementary Table 2). Of them, 18 showed significant enrichment in SCZ and all four smoking phenotypes (Table 1 and Figure 1A). For SCZ, the most significant pathway was postsynaptic density (P < 3.04 × 10^-14), which is consistent with a previous report [43]. This postsynaptic density pathway also showed highly significant enrichment in all smoking-related behaviors, with P values from 3.09 × 10^-6 for age at smoking initiation to 1.43 × 10^-13 for CPD (Table 1),
confirming an important role for postsynaptic density in these two disorders. In contrast, we found that the enriched pathways in smoking behaviors had a much higher overlap with SCZ than that in the traits of null and height (Figure 1B, Supplementary Figures 2 and 3). Next, by using multidimensional scaling (MDS) for shared genes, we found that the 18 common pathways were categorized into five clusters (Figure 1C and Table 1): postsynaptic density (Cluster #1), cadherin binding (Cluster #2), dendritic spine (Cluster #3), long-term depression (Cluster #4), and axon guidance (Cluster #5).

Following the pathway analysis of GWAS summary data on SCZ and smoking, we performed gene-based analysis on the same dataset. For SCZ, we found 590 significantly associated genes after correction for multiple testing. Of them, 236 were located in 108 previously identified loci [11]. Even though numerous genes were significantly associated with CPD and former smoking, no one gene reached significance for ever-smoking or age at smoking initiation. We found 208 genes significantly associated with SCZ that also were associated with at least one of the four smoking phenotypes (Supplementary Table 3). Of these 208 shared genes, 70 were located in 108 previously reported loci for SCZ (Supplementary Figure 4). In addition, by performing computer permutation analysis, we found that the genes identified in our gene-based analysis were significantly overrepresented in the identified common pathways (P < 0.003; Figure 1D and Supplementary Figures 5 and 6).

**Brain gene expression of common pathways**

By analyzing brain expression data, we found 1,443 genes from the 84 common pathways to be coexpressed

---

**Figure 1. Shared pathways of SCZ and smoking phenotypes.** (A) Venn diagram of significantly enriched pathways (q value < 0.1) for SCZ, CPD, ever smoking, former smoking, and age at smoking initiation. (B) Heatmap of the correlation among SCZ, CPD, ever smoking, former smoking, age at smoking initiation, null, and height based on the Z score of pathway enrichment. (C) Multidimensional scaling plot of 18 shared pathways for SCZ and smoking behaviors. Circular ring sizes reflect number of genes in the pathway (range 18–284). Color indicates the significance of the pathway (red marks the significant pathways with lowest P values). Arabic numerals are the pathway numbers as shown in Table 1. (D) Computer permutation analysis of 590 genes associated with SCZ in 84 shared pathways.
Table 1. Top 18 pathways shared between schizophrenia and smoking behaviors.

<table>
<thead>
<tr>
<th>Pathway Number</th>
<th>Pathway ID</th>
<th>Description</th>
<th>Category</th>
<th>SCZ P-value</th>
<th>CPD P-value</th>
<th>Ever smoking P-value</th>
<th>Former smoking P-value</th>
<th>Age at smoking initiation P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GO:0014069</td>
<td>Postsynaptic density</td>
<td>Cluster #1</td>
<td>3.04 x 10^-14</td>
<td>1.43 x 10^-13</td>
<td>1.21 x 10^-9</td>
<td>1.50 x 10^-6</td>
<td>3.09 x 10^-6</td>
</tr>
<tr>
<td>2</td>
<td>GO:0045211</td>
<td>Postsynaptic membrane</td>
<td>Cluster #3</td>
<td>7.44 x 10^-13</td>
<td>4.97 x 10^-11</td>
<td>1.10 x 10^-10</td>
<td>5.92 x 10^-9</td>
<td>5.52 x 10^-6</td>
</tr>
<tr>
<td>3</td>
<td>GO:0045202</td>
<td>Synapse</td>
<td>Cluster #4</td>
<td>3.01 x 10^-9</td>
<td>2.76 x 10^-13</td>
<td>6.88 x 10^-15</td>
<td>1.91 x 10^-9</td>
<td>4.88 x 10^-8</td>
</tr>
<tr>
<td>4</td>
<td>GO:0030425</td>
<td>Dendrite</td>
<td>Cluster #1</td>
<td>9.46 x 10^-9</td>
<td>1.34 x 10^-7</td>
<td>0.0043</td>
<td>0.0011</td>
<td>0.0001</td>
</tr>
<tr>
<td>5</td>
<td>GO:0007268</td>
<td>Synaptic transmission</td>
<td>Cluster #1</td>
<td>3.22 x 10^-8</td>
<td>8.03 x 10^-10</td>
<td>1.85 x 10^-11</td>
<td>3.22 x 10^-12</td>
<td>9.84 x 10^-6</td>
</tr>
<tr>
<td>6</td>
<td>GO:0007411</td>
<td>Axon guidance</td>
<td>Cluster #5</td>
<td>5.89 x 10^-8</td>
<td>8.28 x 10^-6</td>
<td>6.04 x 10^-12</td>
<td>8.10 x 10^-12</td>
<td>3.64 x 10^-9</td>
</tr>
<tr>
<td>7</td>
<td>GO:0005001</td>
<td>Transmembrane receptor protein tyrosine phosphatase activity</td>
<td>Cluster #2</td>
<td>7.21 x 10^-7</td>
<td>0.0010</td>
<td>5.50 x 10^-8</td>
<td>0.0086</td>
<td>7.84 x 10^-5</td>
</tr>
<tr>
<td>8</td>
<td>GO:0043197</td>
<td>Dendritic spine</td>
<td>Cluster #3</td>
<td>4.34 x 10^-6</td>
<td>0.0009</td>
<td>0.0007</td>
<td>0.0011</td>
<td>0.0048</td>
</tr>
<tr>
<td>9</td>
<td>hsa05412</td>
<td>Arrhythogenic right ventricular cardiomyopathy (ARVC)</td>
<td>Cluster #1</td>
<td>1.06 x 10^-5</td>
<td>0.0025</td>
<td>4.17 x 10^-6</td>
<td>1.36 x 10^-5</td>
<td>0.0008</td>
</tr>
<tr>
<td>10</td>
<td>hsa04912</td>
<td>GnRH signaling pathway</td>
<td>Cluster #4</td>
<td>7.76 x 10^-5</td>
<td>0.0056</td>
<td>8.06 x 10^-5</td>
<td>0.0042</td>
<td>0.0001</td>
</tr>
<tr>
<td>11</td>
<td>GO:0045296</td>
<td>Cadherin binding</td>
<td>Cluster #2</td>
<td>0.00014</td>
<td>0.0028</td>
<td>0.0008</td>
<td>0.0004</td>
<td>5.83 x 10^-5</td>
</tr>
<tr>
<td>12</td>
<td>GO:0007156</td>
<td>Homophilic cell adhesion</td>
<td>Cluster #5</td>
<td>0.00049</td>
<td>0.0011</td>
<td>1.48 x 10^-5</td>
<td>2.00 x 10^-7</td>
<td>1.50 x 10^-7</td>
</tr>
<tr>
<td>13</td>
<td>GO:0005216</td>
<td>Ion channel activity</td>
<td>Cluster #1</td>
<td>0.00051</td>
<td>2.95 x 10^-13</td>
<td>0.0001</td>
<td>1.29 x 10^-5</td>
<td>0.0001</td>
</tr>
<tr>
<td>14</td>
<td>GO:0051056</td>
<td>Regulation of small GTPase mediated signal transduction</td>
<td>Cluster #1</td>
<td>0.00053</td>
<td>0.0002</td>
<td>1.14 x 10^-6</td>
<td>4.88 x 10^-5</td>
<td>0.0002</td>
</tr>
<tr>
<td>15</td>
<td>hsa04020</td>
<td>Calcium signaling pathway</td>
<td>Cluster #1</td>
<td>0.0006</td>
<td>0.0056</td>
<td>2.74 x 10^-10</td>
<td>2.38 x 10^-6</td>
<td>2.99 x 10^-6</td>
</tr>
<tr>
<td>16</td>
<td>GO:0006813</td>
<td>Potassium ion transport</td>
<td>Cluster #1</td>
<td>0.00074</td>
<td>0.0038</td>
<td>2.73 x 10^-6</td>
<td>0.0017</td>
<td>9.14 x 10^-5</td>
</tr>
<tr>
<td>17</td>
<td>GO:0007626</td>
<td>Locomotory behavior</td>
<td>Cluster #2</td>
<td>0.0012</td>
<td>9.36 x 10^-14</td>
<td>0.0032</td>
<td>3.10 x 10^-5</td>
<td>0.0030</td>
</tr>
<tr>
<td>18</td>
<td>hsa04730</td>
<td>Long-term depression</td>
<td>Cluster #4</td>
<td>0.0042</td>
<td>7.10 x 10^-3</td>
<td>0.0012</td>
<td>2.85 x 10^-6</td>
<td>0.0009</td>
</tr>
</tbody>
</table>

in six modules (Supplementary Figures 7–9), and the top 10 “hub” genes in each module were demonstrated to have highly intramodular and intermodular connections (Supplementary Figures 10 and 11). Interestingly, we observed two distinct and dynamic expression patterns in different brain regions and at different developmental stages (Figure 2A). For example, one module (marked in yellow) has a greater than twofold difference between brain regions, and two modules (brown and red) showed twofold temporal changes (Supplementary Table 4).

Common methylated genes in both smoking and schizophrenia

By comparison with the previously reported findings [44], we found 149 module genes that overlapped with smoking-associated DNA-methylated genes; and 38 of these genes had at least two independent pieces of supporting evidence (Supplementary Table 5). Next, we performed a differential methylation analysis of 8,236 CpG loci in 149 genes related to SCZ and found that 822 CpG loci, mapped to 124 genes (124/149; binomial test: P < 2.2 x 10^-16), showed significant association with SCZ (P < 6.07 x 10^-6; Supplementary Table 5). For the 38 genes, 413 significantly methylated CpG loci were annotated in 34 of the genes, and there existed a high consistency of shared methylated genes between SCZ and smoking (34/38; binomial test: P = 6.04 x 10^-7). Of note, 16 of these 34 candidate genes have been reported extensively to be associated with SCZ (Table 2 and Supplementary Table 6).

Through using these 34 genes to construct a subnetwork, we found 23 of 34 genes significantly enriched in that subnetwork (PPI enrichment: P = 1.3 x 10^-4; Figure 2E). Many of these genes had several significantly SCZ-associated CpG loci (Figure 2B–2D and Supplementary Figures 12 and 13). For example, SCZ-associated cg10334053 (P = 8.66 x 10^-21), cg02481000 (P = 1.7 x 10^-19), and cg18188739 (P = 3.56 x 10^-14) were all located in PRKCZ (Figure 2B). Of note, cg18188739 appeared to be associated with smoking at a genome-wide significance level (P = 2.86 x 10^-8) [45].

Differential expression profiles of 34 candidate genes

We next examined the difference in the influence of smoking on gene expression between SCZ patients and healthy controls and found that 9 of 34 genes showed significantly different expression in SCZ subjects and...
controls stratified by smoking status (Table 2; Supplementary Table 7; Figure 3; Supplementary Figures 14–16). Furthermore, several genes showed distinct expression profiles in schizophrenic hiPSC-derived neurons (Supplementary Figure 17) compared with the control neurons. Consistent with previous studies [11, 39], most of these candidate genes were highly expressed in human brain regions (Supplementary Figures 18–26).

The pharmacological effects of 34 candidate genes

Based on the DGIdb database, we found that 13 of the 34 genes (38.24%) were targeted by at least one drug. For example, HRH1, CDKN1A, and TNF are targeted by various psychotropic drugs. Also, 25 (73.53%) belong to one or more potentially “druggable” gene categories (Supplementary Figure 27). Consistently, the expression of several genes in the mouse brain was significantly modulated by quetiapine (Supplementary Figure 28) and nicotine (Supplementary Figure 29) in a dose-dependent pattern. Further, the count of significantly quetiapine- and nicotine-induced genes among these 34 genes were prominently higher than that of genes in 18 (Figure 3C, 3G) and 84 common pathways (Figure 3D, 3H) or all background genes (Figure 3E, 3I). For example, the expression of Arhgef3 was downregulated by both quetiapine (Figure 3F: ANOVA P = 0.00018) and nicotine (Figure 3J: ANOVA P = 0.0084). In addition, we found that many genes showed different expression

Figure 2. Brain co-expression modules and common methylated genes. (A) Regional and temporal patterns of gene expression mean-centered by the extent of gene expression within each module. (B–D) Regional plot of association between genetic and epigenetic variants of the PRK2, PTPRN2, GNA12 loci, and SCZ, respectively. Circular symbols indicate the association of CpG loci with SCZ (red represents loci significantly associated with SCZ with P ≤ 6.07 × 10^-6; orange indicates loci with 6.07 × 10^-6 < P ≤ 0.05; blue marks loci with P > 0.05). Triangular symbols indicate association of SNPs with SCZ (red represents top-ranked SNPs associated with SCZ; orange indicates SNPs associated with SCZ with P ≤ 0.05; blue marks SNPs with P > 0.05). (E) Gene subnetwork constituted from the 34 common genes. The protein–protein interactions are according to the database of STRING (v. 10.5). We used Cytoscape software to visualize the subnetwork. The color of a node indicates the co-expression module of the genes.
Table 2. Multiple lines of evidence support involvement of 34 genes in the comorbidity of both SCZ and smoking.

<table>
<thead>
<tr>
<th>Gene names</th>
<th>Chromosome</th>
<th>Brain coexpression modules</th>
<th>Smoking-associated methylation genes</th>
<th>SCZ-associated methylation genes</th>
<th>cis-eQTLs in brain samples</th>
<th>Genes enriched in the PPI subnetwork</th>
<th>Differential expressed genes in SCZ patients stratified by smoking: Anova P values</th>
<th>Differential expression genes treated by quetiapine or nicotine: Anova P values</th>
<th>Genes risk to SCZ in previous studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRIO</td>
<td>Chr5</td>
<td>Turquoise</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Not significant</td>
<td>0.031 (Quetiapine)</td>
<td>Yes (PMID: 21422296)</td>
</tr>
<tr>
<td>GNA12</td>
<td>Chr7</td>
<td>Brown</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Not significant</td>
<td>0.049 (Quetiapine)</td>
<td>Yes (PMID: 22792057)</td>
</tr>
<tr>
<td>KCNQ1</td>
<td>Chr11</td>
<td>Blue</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Not significant</td>
<td>0.0073 (Nicotine); 0.3 (Quetiapine)</td>
<td>Yes (PMID: 26971948; 28188958)</td>
</tr>
<tr>
<td>AKT3</td>
<td>Chr1</td>
<td>Turquoise</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.0395</td>
<td>0.018 (Nicotine)</td>
<td>Yes (PMID: 25056061; 28467426; 29173281; 23974872; 25599223)</td>
</tr>
<tr>
<td>ARHGGEF3</td>
<td>Chr3</td>
<td>Yellow</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.00723</td>
<td>0.00018 (Quetiapine); 0.0084 (Nicotine)</td>
<td>No data</td>
</tr>
<tr>
<td>CACNA1D</td>
<td>Chr3</td>
<td>Turquoise</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.052</td>
<td>Not significant</td>
<td>Yes (PMID: 29214423; 26255836; 24996399)</td>
</tr>
<tr>
<td>RGS12</td>
<td>Chr4</td>
<td>Brown</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.0058</td>
<td>0.053 (Nicotine)</td>
<td>Yes (PMID: 25420024)</td>
</tr>
<tr>
<td>RPS6KA2</td>
<td>Chr6</td>
<td>Brown</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Not significant</td>
<td>0.05 (Nicotine)</td>
<td>No data</td>
</tr>
<tr>
<td>NOTCH1</td>
<td>Chr9</td>
<td>Yellow</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Not significant</td>
<td>0.0166 (Quetiapine)</td>
<td>Yes (PMID: 26232790)</td>
</tr>
<tr>
<td>PARD3</td>
<td>Chr10</td>
<td>Yellow</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.054</td>
<td>0.018 (Nicotine)</td>
<td>Yes (PMID: 22969987)</td>
</tr>
<tr>
<td>AP2A2</td>
<td>Chr11</td>
<td>Brown</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Not significant</td>
<td>0.0032 (Nicotine)</td>
<td>Yes (PMID: 23811784)</td>
</tr>
<tr>
<td>NOS1AP</td>
<td>Chr1</td>
<td>Brown</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Not significant</td>
<td>Not significant</td>
<td>Yes (PMID: 26861996; 25542305; 20605702; 16146415; 15065015; 12116186; 19077434)</td>
</tr>
<tr>
<td>HRH1</td>
<td>Chr3</td>
<td>Brown</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Not significant</td>
<td>Not significant</td>
<td>Yes (PMID: 28400155; 27855565)</td>
</tr>
<tr>
<td>HTT</td>
<td>Chr4</td>
<td>Brown</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Not significant</td>
<td>0.0085 (Quetiapine); 0.075 (Nicotine)</td>
<td>No data</td>
</tr>
<tr>
<td>CDKN1A</td>
<td>Chr6</td>
<td>Turquoise</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Not significant</td>
<td>0.00085 (Quetiapine); 0.23 (Nicotine)</td>
<td>Yes (PMID: 23549417)</td>
</tr>
<tr>
<td>TIAM2</td>
<td>Chr6</td>
<td>Turquoise</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.062</td>
<td>Not significant</td>
<td>No data</td>
</tr>
<tr>
<td>TNF</td>
<td>Chr6</td>
<td>Blue</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Not significant</td>
<td>0.06 (Nicotine)</td>
<td>Yes (PMID: 29499967; 29706448)</td>
</tr>
<tr>
<td>CNTNAP2</td>
<td>Chr7</td>
<td>Brown</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>0.087</td>
<td>0.0085 (Quetiapine)</td>
<td>Yes (PMID: 29610457)</td>
</tr>
</tbody>
</table>
profiles in the mouse striatum at four time points of treatment with 18 major psychotropic drugs (Supplementary Figures 30–32).

Cis-regulatory effects of SNPs in 34 candidate genes

Based on two large GWAS datasets, there were 3,483 risk-suggested SNPs shared by SCZ and at least one smoking phenotype (Figure 4A and Supplementary Figure 33), and 228 of these were common to SCZ and all four smoking phenotypes (Figure 4B). A great number of SNPs among these identified 3,483 risk SNPs were located within different types of regulatory elements in brain tissues and neuroblastoma cell lines (Supplementary Figures 34 and 35). We performed a cis-meQTL analysis in human brain samples, which showed 7,558 significant SNP–CpG methylation pairs with 1,145 SNPs in 21 genes (FDR < 0.01; Figure 4A and Supplementary Figure 36). A number of the 392 variants in the 21 genes had cis-regulatory roles in both DNA methylation and gene expression (Figure 4A, 4C).

DISCUSSION

Cigarette smoking is highly concurrent with SCZ [2, 18, 46]. Individuals with mental disorders such as SCZ are at higher risk for developing smoking-related diseases, which include cardiovascular and respiratory diseases and various cancers [5]. Based on a national cohort study including about 6 million Swedish adults, Crump et al. [47] found a significant portion of the morbidity and premature death in persons with SCZ was ascribable to ischemic heart disease and cancers. Considering that the co-occurrence of SCZ and smoking has greatly impacted public health, it is important to understand the pathogenesis of the comorbidity of the two diseases. Previous studies concentrated largely on the investigation of the genetic mechanisms of either SCZ [11] or smoking [12]. Similarly, studies employing pathway-based enrichment analysis so far have focused mainly on either SCZ or smoking [39–42].

Recently, multiple large-scale GWAS studies [27–30, 38] consistently revealed that there exists a considerable genetic correlation between SCZ and smoking-related traits. However, susceptibility variants, genes, and biological pathways for the comorbidity remain largely unknown. In the present study, by conducting an integrative genomics analysis of large-scale GWAS data with multi-omics data, we intended to identify the risk SNPs, genes, and pathways implicated in the etiology of these two comorbid diseases.

We first conducted pathway-based enrichment analysis of GWAS summary data. For this hypothesis-free genome-wide approach, we utilized four commonly used resources (i.e., KEGG, GO, BioCarta, and Reactome) to determine the number of genes included in the GWAS pathway enrichment analyses [40], which could overcome the bias of previous studies for arbitrary gene

<table>
<thead>
<tr>
<th>MAD1L1</th>
<th>Chr7</th>
<th>Brown</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>No</th>
<th>Not significant</th>
<th>0.024 (Nicotine)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTPRN2</td>
<td>Chr7</td>
<td>Red</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>0.00142</td>
<td>0.0028 (Nicotine)</td>
</tr>
<tr>
<td>CLCN6</td>
<td>Chr1</td>
<td>Green</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Not significant</td>
<td>Not significant</td>
</tr>
<tr>
<td>PRKCZ</td>
<td>Chr1</td>
<td>Brown</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Not significant</td>
<td>Not significant</td>
</tr>
<tr>
<td>ALPI</td>
<td>Chr2</td>
<td>Blue</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Not significant</td>
<td>Not significant</td>
</tr>
<tr>
<td>ALPL2</td>
<td>Chr2</td>
<td>Blue</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Not significant</td>
<td>No data</td>
</tr>
<tr>
<td>ARHGAP2 5</td>
<td>Chr2</td>
<td>Turquoise</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>0.00026</td>
<td>Not significant</td>
</tr>
<tr>
<td>PLEK</td>
<td>Chr2</td>
<td>Turquoise</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>0.043</td>
<td>0.018 (Nicotine)</td>
</tr>
<tr>
<td>ZMIZ1</td>
<td>Chr10</td>
<td>Turquoise</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.041</td>
<td>0.0128 (Quetiapine)</td>
</tr>
<tr>
<td>TBC1D14</td>
<td>Chr4</td>
<td>Turquoise</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Not significant</td>
<td>No data</td>
</tr>
<tr>
<td>SORBS1</td>
<td>Chr10</td>
<td>Turquoise</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Not significant</td>
<td>No data</td>
</tr>
<tr>
<td>TGFBR3</td>
<td>Chr1</td>
<td>Turquoise</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Not significant</td>
<td>Not significant</td>
</tr>
<tr>
<td>CPOX</td>
<td>Chr3</td>
<td>Brown</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Not significant</td>
<td>Not significant</td>
</tr>
<tr>
<td>TIGIT</td>
<td>Chr3</td>
<td>Blue</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Not significant</td>
<td>Not significant</td>
</tr>
<tr>
<td>GPSM3</td>
<td>Chr6</td>
<td>Blue</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>0.0014</td>
<td>Not significant</td>
</tr>
<tr>
<td>DUSP4</td>
<td>Chr8</td>
<td>Turquoise</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Not significant</td>
<td>Not significant</td>
</tr>
</tbody>
</table>
This analysis revealed 18 significantly enriched pathways that were shared by SCZ and all four smoking phenotypes. These pathways were clustered into postsynaptic density, cadherin binding, dendritic spine, long-term depression, and axon guidance; all of them have been implicated in various psychiatric disorders [11, 13, 39, 43, 48–51]. Compared with the negative controls, the degrees of correlation between pathways across SCZ and smoking phenotypes were much higher, indicating that overlap between SCZ and smoking was nonrandom at any pathway level. These findings strongly indicate that the common pathways identified for both SCZ and smoking behaviors were attributable to the shared genetic vulnerability.

Given that leveraging multi-omics datasets is a better way to understand the molecular mechanism of complex diseases, we here provide robust evidence to explain the

Figure 3. The differential expression patterns of 34 candidate genes. (A) Pattern of PTPRN2 in SCZ patients and controls divided by smoking status. (B) Pattern of ARHGA25 in SCZ patients and controls divided by smoking status. (C–E) Permutation analysis of 34 candidate genes in 18 common pathways (N = 1,588 genes), 84 common pathways (N = 3,334 genes), and background genes (N = 45,037) for quetiapine treatment. (F) Plot summarizing Arhgef3 expression changes in quetiapine (at doses of 10 or 100 mg/kg)-treated mice. (G–I) Permutation analysis of 34 candidate genes in 18 common pathways (N = 1,588), 84 common pathways (N = 3,334), and background genes (N = 31,047) for nicotine treatment. (J) Plot summarizing Arhgef3 expression alterations in nicotine (at doses of 8 μg, 15 μg, and 23 μg/L)-treated mice.
underlying mechanism of the comorbidity between SCZ and smoking from genetic, epigenetic, and expression points of view. Based on the gene co-expression data, we found a considerable number of genes in common pathways that showed differential co-expression patterns in different developmental time points and brain regions, providing supportive evidence that shared genetic risk for SCZ and smoking phenotypes has a vital influence on neurodevelopmentally regulated pathways. Similar to a previous study [43], by integrating gene co-expression data with pathway-based analyses on large-scale GWAS, we could identify even greater specificity of candidate genes conferring risk for the comorbidity.

Previous studies [11, 44, 52–54] have shown that aberrant methylated DNAs (DNAm) were involved in the etiology of SCZ and smoking. For example, our previous study, based on a sophisticated data mining of published papers [44], showed that 320 genes exhibited robust methylation evidence of involvement in the etiology of smoking. In the current study, we highlighted 34 genes with significant methylation evidence that

---

**Figure 4.** The cis-acting regulatory effects of risk-associated SNPs in 34 common genes on DNA methylation and gene expression. (A) Schematic of risk-associated SNPs (P < 0.05) compiled from two large-scale meta-GWASs on SCZ (N = 79,845) and smoking behaviors (N = 74,053) and 3,483 risk-associated SNPs in 34 genes shared by SCZ and smoking for cis-meQTL and cis-eQTL analysis. The SNPs in strong LD with risk-associated SNPs (LD cutoff r² ≥ 0.5) were generated according to the 1000 Genome European Phase 3 panel as reference. (B) Venn diagram of risk-associated SNPs (P < 0.05) for SCZ, CPD, ever smoking, former smoking, and age at smoking initiation. (C) Plot shows the 21 promising genes with SNPs had cis-regulatory roles in both DNA methylation and gene expression.
were associated with both SCZ and smoking. More interestingly, we found that 23 of these 34 genes were located in a molecular subnetwork, which is consistent with the previous notion that function-related genes may collectively contribute the risk of the pathology of complex diseases [44]. Notably, 16 of the 34 methylated genes have often been reported to be associated with SCZ and smoking. For example, AKT3 is a critical molecule underlying psychiatric-related behaviors, including SCZ [11, 55], cognitive function [56], and smoking behaviors [44]. Multiple lines of evidence from genetic association studies [11, 57, 58] have indicated that MAD1L1 is significantly associated with SCZ. Based on network-assisted investigation of combined causal signals from GWAS studies [59], CNAI2 showed a significant association with SCZ. PRKACZ and PTPRN2 were reported to be SCZ-associated differential methylation regions [54]. The KCNQ1 gene belongs to the potassium channel gene family and plays a vital role in signal transduction within the central nervous system. It may contribute to the shared risk of diminished processing speed, diminished white matter integrity, and increased risk of SCZ [60, 61]. In addition, we found that a large proportion of the 34 genes were targeted by various psychotropic drugs, and the expression patterns of these drug-responsive genes were significantly regulated in mouse brain by treatment with 18 major psychotropic drugs in a dose- or time-dependent pattern. Thus, these genes likely represent targets for pharmacotherapeutic intervention in SCZ and smoking behaviors.

Genetic variants influencing DNAm and mRNA expression can modulate gene transcript levels and thereby exert risk effects on diseases, as evidenced by meQTL and eQTL analyses of top GWAS risk-associated SNPs for various human phenotypes, including smoking [54, 62, 63] and SCZ [64, 65]. Thus, we first explored the vital influence of regulatory genomic elements (i.e., cis-meQTLs and cis-eQTLs) in neurodevelopmental processes of the comorbidity of smoking and SCZ. In human brain samples, we identified 21 of 34 candidate genes with a great number of SNP-CpG pairs. Interestingly, there was a large proportion, 39.4%, of these SNPs that showed allele-specific gene expression. By downloading SNPs from the GWAS Catalog (on 03 May 2019), some of the identified SNPs were found to be associated with psychiatric disorders in GWAS studies. For example, rs1107592 (P = 2.0 × 10−6) and rs802568 (P = 2.0 × 10−7) were suggested to be associated with bipolar disorder and SCZ [66]. Rs1107592 (P = 5.0 × 10−6), rs4721295 (P = 6.0 × 10−10), and rs12666575 (P = 2.0 × 10−9) were significantly associated with SCZ and other psychiatric disorders, such as autism spectrum, bipolar, and major depressive disorders [67–69]. Rs1403174 (P = 3.0 × 10−10) showed significant association with age at smoking initiation [70]. These findings were consistent with earlier reports where it was discovered that the majority of non-coding risk-associated SNPs for brain disorders influence gene expression via DNAm [71, 72].

Some limitations of the current study warrant comment. First, those SCZ- and smoking-related genes identified were prioritized from GWAS. In view of the inherent defects of the GWAS approach, some of these genes might not be truly associated with both of the disorders. Second, a great number of genes have not been characterized or mapped to computationally predicted or manually curated pathways. Thus, the contributions of these genes could not be delineated in pathway enrichment analysis. Third, although we have provided very strong evidence from genetic association, gene expression, and DNA methylation to support the 18 common pathways as being linked to both smoking and SCZ, we could not determine to what extent each specific pathway contributes to smoking or SCZ, nor could we quantify the relative influences of each pathway for the two disorders.

In sum, by employing a comprehensive bioinformatics analysis of genomic and pharmacogenomics data at the DNA, methylation, or expression levels, we first identified 18 biological pathways that are significantly associated with both SCZ and smoking phenotypes. Subsequently, we discovered 34 novel and promising susceptibility genes and variants within these genes with robust genetic, epigenetic, expression, and pharmacological evidence for both SCZ and smoking. Our findings lend considerable weight to the hypothesis that shared genetic vulnerabilities create a propensity for the comorbidity of SCZ and smoking. By pinpointing which and how SNPs in these candidate genes affect these disorders, this study provides novel insights into the biological mechanism and a solid foundation for understanding the shared disease biology of SCZ and smoking behaviors.

MATERIALS AND METHODS

Multi-omics datasets used in current study

In current investigation, we performed a series of multi-omics data analyses, which include large-scale meta-GWAS, BrainSpan exon array data on brain development and aging, methylation data, expression data, and pharmacogenomics data. The following is a brief summary of datasets used in our analysis. For details on these datasets, please refer to Supplementary Table 1.

1. GWAS data on SCZ: This dataset was obtained from a published GWAS [11] on SCZ from the
Psychiatric Genomics Consortium (PGC), of which the data from case-control samples (N = 79,845) were used for both gene and pathway analysis.

2. GWAS data on smoking: This dataset was from a published meta-GWAS on smoking behaviors (including smoking status, quantity smoked in ever-regular smokers, smoking cessation, and age at initiation) by the Tobacco and Genetics Consortium (TAG), which contained the genotype data for a total sample size of 74,053 [12].

3. GWAS data on height and null: To demonstrate that the identified pathways were attributable to shared biology between SCZ and smoking, we performed the GWAS enrichment analysis on two other independent data sets as negative controls. One of them was a GWAS of human height with a sample size of 183,727 [73], and the other was a null GWAS based on randomly distributed phenotypes that we constructed from a real GWAS with a sample size of 3,960 [74].

4. Expression data on brain development and aging: The BrainSpan exon array data related to brain development and aging were downloaded from the NCBI’s Gene Expression Omnibus (GEO: Accession No. GSE25219), with a sample size of 1,340.

5. Methylation data on smoking and SCZ: Three datasets were used for this part of the analysis, with the first one from our previous study on smoking and consisting of 18,677 samples [44], the second one on 847 SCZ blood samples from NCBI GEO (Accession No. GSE84727) [75], and the third one on brain samples (N=258) from NCBI GEO (Accession No. GSE74193) [76].

6. Expression data on smoking and SCZ: Two expression datasets were used here, with the first one based on olfactory epithelium tissues (N = 31) and the second one on induced pluripotent human stem cells (hiPSCs; N = 8). Both were downloaded from NCBI GEO (Accession Nos. GSE73129 and GSE25673).

7. Pharmacogenomics data: To reveal potential druggable genes with therapeutic effects, we downloaded the psychotropic drug-treated gene expression data from NCBI’s GEO with Accession Nos. of GSE45229, GESS0254, GSE48954, GSE15774, and GSE48951. We first explored the dosage influence of quetiapine and nicotine on gene expression changes in mouse striatum, then concentrated on the time-course (1, 2, 4, and 8 hours) of gene expression alterations in mouse striatum that were induced by 18 major psychotropic drugs.

Pathway- and gene-based analysis

We combined gene-set data from four sources: KEGG [77], GO [78], BioCarta [79], and Reactome [80], which were downloaded from their respective sources on or before May 12, 2017. Because of concerns that the gene set with > 300 genes or < 10 genes is considered to be either less specific and computationally inefficient or over-dispersed [40], we confined our analysis to those pathways with 10–300 genes.

We used the GSA-SNP program [81] to perform GWAS-based enrichment analysis. The SNPs of interest were assigned to genes if they lie within 20 kb upstream or downstream of the gene, and each SNP was assigned to only one gene. When multiple SNPs were mapped to the same gene, the GSA-SNP chose the most significantly associated SNP. Additionally, we used the MAGMA (https://ctg.cnqr.nl/software/magma) for gene-based analysis of GWAS summary data. Because LD could result in a number of genes rather than one being counted as significant when genes are physically close in one region, we used the LD-pruned method to calculate the LD for the published GWAS data with the 1000 Genome European Phase 3 panel as the reference.

Computer permutation analysis

There were 3,334 genes (named Gene set 1) from 84 common pathways with a q value < 0.1 from our pathway enrichment analysis. To determine whether these genes were indeed significantly overrepresented with identified genes from our gene-based analysis, we conducted permutation analysis by randomly selecting 3,334 genes in the 84 common pathways from the total genes (N = 17,385) in all 2,532 pathways for 10^6 times. We then calculated how many times the counts of genes overlapped with the genes from gene-based analysis that were larger than the observed number among 10^6 trials. The probability of the observation was treated as the P-value, with a P value < 0.05 being considered significant.

Co-expression network analysis

We then conducted further analysis for 3,334 identified genes in Gene Set 1 to determine how these pathways were related to brain development and aging with the BrainSpan exon array data. RNA expression profiles for 1,340 samples were analyzed by using weighted gene co-expression network analysis (WGCNA), an R package used for clustering genes into modules according to co-expression data. The 10 most highly
connected genes within each module were displayed in the network plot by using Cytoscape (v. 3.5.1) (http://www.cytoscape.org/).

**Smoking- and SCZ-associated differentially methylated loci and regions**

To determine whether the identified genes from co-expression modules were involved in the smoking- and SCZ-associated methylation process, we first selected a list of 1,429 smoking-associated methylated genes from our previous study [44], and then examined whether these smoking-associated methylation module genes were significantly methylated in SCZ patients [75]. Bonferroni correction was used to determine significant association. An exact binomial analysis was applied to examine whether there existed a significant excess of consistency of observed methylated genes between smoking and SCZ more often than was expected by chance.

**Candidate gene expression profiles in SCZ and smoking**

We performed an ANOVA analysis to explore the differential expression profiles of the 34 genes in olfactory epithelium tissues among SCZ patients and controls grouped by smoking status. Turkey HSD test was used for multiple comparisons. Considering that human induced pluripotent stem cells (hiPSCs) provide a novel strategy for defining characteristics of schizophrenic neurons, we further used the first cell-based human model of SCZ by directly reprogramming fibroblasts from schizophrenic patients into hiPSCs and subsequently differentiating these disorder-specific hiPSCs into neurons in vitro to explore the different expression profiles of 34 genes between controls and schizophrenic patients.

**Candidate genes in response to various psychotropic drugs**

To identify potential druggable targets, we searched for these identified genes in the Drug-Gene Interaction database (DGIdb) (v. 3.0; http://www.dgidb.org/). Firstly, we searched the 20 databases for drug–gene interactions with FDA-approved pharmaceutical compounds according to 51 known interaction types with 34 common genes for both SCZ and smoking phenotypes. Secondly, we searched 10 databases for the potential drug ability for gene targets to reveal genes that might form targets for novel therapies in addition to existing medicines. To explore whether these genes had therapeutic effects, we applied the dosage treatment of quetiapine and nicotine, and the time-course (1, 2, 4, and 8 hours) treatment of 18 major psychotropic drugs on gene expression changes in mouse striatum. Furthermore, using the computer permutation analysis of 10² times, we determined whether these 34 candidate genes were more prone to drug-induced action than other 1,588 genes in 18 common pathways, 3,334 genes in 84 common pathways, as well as more than 30,000 background genes.

**Cis-meQTLs/eQTLs of SNPs within 34 genes**

To explore the relations between genotype and methylation status, we conducted cis-meQTL analysis of SNPs within 34 candidate genes. Based on the two large GWAS used in our pathway-based analysis, we collected suggested SNPs with a P value of < 0.05 in 34 candidate genes for SCZ and smoking behaviors. Considering that multiple lines of evidence have suggested that identified tag SNPs were more likely to be in LD with casual variants [82], we generated a list of SNPs that were in strong LD (LD cutoff r² ≥ 0.5) with each tag SNP using the 1000 Genome European Phase 3 panel for reference genotyping.

By employing these identified SNPs, we downloaded cis-meQTL data from human brain samples (N = 258) [76] and used the Matrix eQTL (v. 2.1.1) R package [83] to examine the associations between SNPs and methylation loci with linear regression under an additive model. We restricted methylation loci to 20 kb upstream and downstream of each SNP. The intervals for nearby SNPs were combined if they overlapped. There were 7,426,085 common genotyped and imputed SNPs from the 1,000 Genomes reference panel and 477,636 qualified CpGs used for cis-meQTL analysis with a maximum distance of 20 kb between each SNP and CpG analyzed, resulting in 47,675,913 tests. A total of 4,107,214 significant SNP-CpG methylation associations at FDR < 0.01 were identified from this dataset. To further explore the cis-regulatory effects of SNPs on expression of the 34 candidate genes, we performed a cis-acting eQTL analysis in human tissues by using a web-based tool of GTEX PROTAL (https://gtexportal.org/home/).

**Abbreviations**

SCZ: schizophrenia; ND: nicotine dependence; GWAS: genome-wide association study; CPD: cigarette smoking per day; MDS: multidimensional scaling; PPI: protein-protein interaction; eQTL: expression quantitative trait loci; meQTL: methylation quantitative trait loci; PGC: the psychiatric genomics consortium; TAG: the tobacco and genetics consortium; GEO: Gene Expression Omnibus; hiPSCs: human induced pluripotent stem cells; KEGG: Kyoto Encyclopedia of Genes and Genomes; GO: gene ontology; WGCNA:
weighted gene co-expression network analysis; DGIdb: the Drug-Gene Interaction database; LD: linkage disequilibrium.

AUTHOR CONTRIBUTIONS

YM participated sample collection, data analysis and wrote the manuscript. JL, YX, and YY performed literature and public database searches. MW, QL, and XZ conducted the cis-meQTL analysis. RF, JC, and JL performed co-expression network analysis and other expression analysis. RF, JC, HH, BZ, YW, ZC, ZY, WC, WY, XC, YN, JZM, and TJP participated sample collection and/or reviewed the manuscript. MDL conceived the study and was involved in every step of the research. All authors approved the final manuscript.

ACKNOWLEDGMENTS

We thank Dr. Patrick F. Sullivan from the Departments of Genetics and Psychiatry in University of North Carolina for his critical comments and suggestions on the paper. We also thank Dr. David L. Bronson for excellent editing of this manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

FUNDING

This study was supported in part by the China Precision Medicine Initiative (2016YFC0906300), China Postdoctoral Science Foundation (2018M630667), Research Center for Air Pollution and Health of Zhejiang University, and the State Key Laboratory for Diagnosis and Treatment of Infectious Diseases of the First Affiliated Hospital of Zhejiang University. All authors declare no competing interest in relation to the work reported in this communication.

REFERENCES

   https://doi.org/10.1001/jama.284.20.2606
   PMID:11086367

2. de Leon J, Diaz FJ. A meta-analysis of worldwide studies demonstrates an association between schizophrenia and tobacco smoking behaviors. Schizophr Res. 2005; 76:135–57.
   https://doi.org/10.1016/j.schres.2005.02.010
   PMID:15949648

   https://doi.org/10.11919/j.issn.1002-0829.216044
   PMID:28638206

   https://doi.org/10.1176/ajp.155.11.1490
   PMID:9812108

   https://doi.org/10.1177/070674370905400603
   PMID:19527556

   https://doi.org/10.1016/S0893-133X(02)00324-X
   PMID:12225705

   https://doi.org/10.2165/00023210-200115060-00005
   PMID:11524025

   https://doi.org/10.1016/S2215-0366(15)00333-8
   PMID:26236007

   https://doi.org/10.1046/j.1360-0443.2003.00295.x
   PMID:12492752

   https://doi.org/10.1001/archpsyc.60.12.1187
   PMID:14662550

   https://doi.org/10.1038/nature13595
   PMID:25056061

Nicotine reverses hypofrontality in animal models

https://doi.org/10.1016/S2215-0067(15)00152-2 PMID: 26249303


Ma Y, Wang M, Yuan W, Su K, Li MD. The significant association of Taq1A genotypes in DRD2/ANKK1 with smoking cessation in a large-scale meta-analysis of Caucasian populations. Transl Psychiatry. 2015; 5:e686. https://doi.org/10.1038/tp.2015.176 PMID: 26624925


Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic


42. Liu C, Bousman CA, Pantelis C, Skafidas E, Zhang D, Yue


PMID: 29666375.

PMID: 26220977.

https://doi.org/10.1038/s41398-018-0130-x

PMID: 24166486

https://doi.org/10.1016/j.srep03075 PMID: 24166486

https://doi.org/10.1038/mp.2012.73 PMID: 22688191

https://doi.org/10.1038/ncomms14519 PMID: 28240266

https://doi.org/10.1038/nemm00595.

https://doi.org/10.1038/nemm00595.

https://doi.org/10.1038/nemm00595.

https://doi.org/10.1038/nemm00595.

https://doi.org/10.1038/nemm00595.

https://doi.org/10.1038/nemm00595.

https://doi.org/10.1038/nemm00595.

https://doi.org/10.1038/nemm00595.

https://doi.org/10.1038/nemm00595.

https://doi.org/10.1038/nemm00595.

https://doi.org/10.1038/nemm00595.

https://doi.org/10.1038/nemm00595.

https://doi.org/10.1038/nemm00595.
2010; 467:832–38. https://doi.org/10.1038/nature09410 PMID:20881960


SUPPLEMENTARY MATERIALS

Supplementary References


15. Levenga J, Wong H, Milstead RA, Keller BN, LaPlante LE, Hoeffer CA. AKT isoforms have distinct...


42. Goldsmith DR, Haroon E, Miller AH, Strauss GP, Buckley PF, Miller BJ. TNF-α and IL-6 are associated with the deficit syndrome and negative symptoms in patients with chronic schizophrenia. Schizophr Res. 2018; 199:281–84. https://doi.org/10.1016/j.schres.2018.02.048 PMID: 29499967


Supplementary Figure 1. Overview of statistical approaches for integrative pathway- and gene-based analysis of GWAS summary statistics. Using GSA-SNP and MAGMA software, pathway- and gene-based analyses were carried out. A number of novel genes were identified from MAGMA analysis. There was a total of 208 shared genes between SCZ and smoking behaviors. According to pathway-based enrichment analysis, 18 common pathways were identified. We performed a simulation analysis and found these 208 genes to be significantly enriched in common pathways identified from GSA-SNP pathway analysis. Further, to explore the underlying biological mechanism of the comorbidity of SCZ and smoking, we then integrated multi-dimensional omics datasets, including human brain and blood transcriptome data, on smoking and SCZ with well-established tools for interrogating such datasets. Finally, we identified 34 candidate genes conferring risk of the comorbidity of SCZ and smoking and revealed the regulatory mechanism of SNP-methylation expression of the comorbidity.
**Supplementary Figure 2. Overlapped pathways among smoking behaviors, null, and height.** (A) Overlapped pathways among CPD, null, and height; (B) Overlapped pathways among ever-smoking, null, and height; (C) Overlapped pathways among former smoking, null, and height; (D) Overlapped pathways among age at smoking initiation, null, and height; (E) Heatmap of the proportion of significantly enriched pathways in CPD, ever smoking, former smoking, and age at smoking initiation with SCZ, null, and height. Heatmap plots were generated using the *pheatmap* package in R.
Supplementary Figure 3. Overlapped pathways among SCZ and smoking behaviors, null, and height. (A) Venn diagram of significantly enriched pathways for SCZ with the pathways for null and height. (B) Venn diagram of significantly enriched pathways among 84 shared pathways between SCZ and at least one of four smoking phenotypes with the pathways for null and height. (C) Venn diagram of significantly enriched pathways among 18 pathways shared by SCZ and all four smoking phenotypes with the pathways for null and height. (D) The proportion of significantly enriched pathways in CPD, ever smoking, former smoking, age at smoking initiation, null, and height with SCZ.
Supplementary Figure 4. Circus plot of results of gene-based analysis of SCZ and smoking phenotypes. The outer ring shows the 22 autosomal human chromosomes. Gene symbols marked in the second ring are the 70 of 208 common genes located in the previously reported 108 loci for SCZ. The third ring demonstrates the results of gene-based analysis of SCZ and 590 genes significantly associated with it ($P < 2.73 \times 10^{-6}$), which are marked with orange points. The orange bars represent 236 genes located in 108 previously identified SCZ-related loci. The grey bars mark novel genes associated with SCZ in the current study. The fourth through the seventh rings mark CPD, ever smoking, former smoking, and smoking initiation, respectively. The yellow bars indicate the distribution of the 208 genes shared by SCZ and smoking.
Supplementary Figure 5. Computer permutation analysis of 208 genes associated with SCZ and smoking behaviors for genes in 84 shared pathways.

Supplementary Figure 6. Computer permutation analysis of 70 genes associated with SCZ and smoking behaviors for genes in 84 shared pathways.
Supplementary Figure 7. WGCNA module-based analysis for genes in all shared pathways with q value < 0.1. The hierarchical clustering dendrogram is for all 1,644 expressed and highly connected genes. Each line represents an individual gene. Genes were clustered on the basis of a dissimilarity measure (1-TOM). The branches are related to modules of highly interconnected gene groups. Below the dendrogram of each gene group is a color to indicate the six module colors (i.e., blue, brown, green, red, turquoise, and yellow). Genes within grey boxes were not assigned to a module.
Supplementary Figure 8. Scatterplot of gene significance for brain regions (y axis) vs. module membership (x axis) in different modules. (A) For blue module; (B) for brown module; (C) for green module; (D) for red module; (E) for turquoise module; (F) for yellow module.
Supplementary Figure 9. Scatterplot of gene significance for developmental time points (stages; y axis) vs. module membership (x axis) in different modules. (A) For blue module; (B) for brown module; (C) for green module; (D) for red module; (E) for turquoise module; (F) for yellow module. For the most significant modules (i.e., blue, brown, green, turquoise, and yellow), genes with high module membership often also show high gene significance. Gene significance and module membership have a highly significant correlation, indicating that hub genes of these modules also predispose to high correlation with weight.
Supplementary Figure 10. Network plot of 10 hub genes from each module showing clustering across different brain regions and developmental time points. The nodes (genes) are colored by module, while the edges reflect positive correlations across brain regions and developmental epochs. The rectangular nodes represent genes clustered in the 18 shared pathways, and the circular nodes mark genes not so clustered.

Supplementary Figure 11. Enrichment of the five clusters of 18 shared pathways in six co-expression modules (red = cluster with the highest proportion in a module). More than 15% of the genes in the five clusters among 18 shared pathways were in the turquoise module.
Supplementary Figure 12. Regional plot of association between genetic and epigenetic variation at the four genomic loci and SCZ. (A) For NOTCH1; (B) for AP2A2; (C) for DUSP4; (D) for RPS6KA2. Circular symbols indicate the association of CpG loci with SCZ (red represents these loci significantly associated with SCZ with P ≤ 6.07 × 10^{-6}; orange represents these loci with 6.07 × 10^{-6} < P ≤ 0.05; blue represents these loci with P > 0.05). Triangular symbols indicate the association of SNPs with SCZ (red represents the top-ranked SNPs associated with SCZ; orange represents these SNPs associated with SCZ with P ≤ 0.05; blue represents these SNPs with P > 0.05).
Supplementary Figure 13. Regional plot of association between genetic and epigenetic variation at the six genomic loci and SCZ. (A) For AKT3; (B) for CDKN1A; (C) for TNF; (D) for PLEK; (E) for KCNQ1; (F) for PARD3. Circular symbols indicate the association of CpG loci with SCZ (red represents these loci significantly associated with SCZ with $P \leq 6.07 \times 10^{-6}$; orange represents these loci with $6.07 \times 10^{-6} < P \leq 0.05$; blue represents these loci with $P > 0.05$). Triangular symbols indicate the association of SNPs with SCZ (red represents the top-ranked SNPs associated with SCZ; orange represents these SNPs associated with SCZ with $P \leq 0.05$; blue represents these SNPs with $P > 0.05$).
Supplementary Figure 14. Heatmap of significant genes from ANOVA by comparing the difference between SCZ patients and controls divided by smoking status. Colored rectangles represent transcript abundance indicated above the gene labeled on the right. The intensity of the color is proportional to the standardized values between -3 and 3 for each gene, as indicated on the bar on the right of the heat map plot. Clustering was performed using Euclidean distance according to the scale on the left. Major gene transcription patterns are arbitrarily described as A, B, and C. Heatmap plots were generated using the pheatmap package in R.
Supplementary Figure 15. Boxplots of 3 genes’ expression patterns in SCZ patients and controls divided by smoking status. (A) For HTT; (C) for NOTCH1; (D) for PARD3.
Supplementary Figure 16. Boxplots of 3 genes’ expression patterns in SCZ patients and controls divided by smoking status. (A) For ZMIZ1; (B) for TBC1D14; (C) for ARHGEF3.
Supplementary Figure 17. Plots summarizing significant gene expression alterations in schizophrenic hiPSC-derived neurons. Student’s t-test was used for comparing the difference between SCZ hiPSC neurons and controls. Gene expression data from the GEO dataset (Accession No. GSE25673). (A) For NOTCH1; (B) for DUSP4; (C) for SORBS1; (D) for RGS12; (E) for CACNA1D; (F) for GPSM3; (G) for PTQRN2; (H) for ZMIZ1; (I) for GNA12; (J) for NOS1AP; (K) for AP2A2; (L) for CDKN1A; (M) for CNTNAP2; (N) for TBC1D14; (O) for MAD1L1; (P) for CPOX. These plots were generated using the beeswarm package in R.
Supplementary Figure 18. An example of PRKCZ expression in 53 human tissues, which were obtained from GTEx analysis release V7 (dbGaP Accession No. phs000424.v7.p2). Expression values are shown in transcripts per million (TPM), calculated from a gene model with isoforms collapsed into a single gene. Box plots describe median and 25th and 75th percentiles; points are displayed as outliers if they are above or below 1.5 times the interquartile range.
Supplementary Figure 19. Expression of (A) AP2A2 and (B) CACNA1D in 53 human tissues. The expression data were obtained from GTEx analysis release V7 (dbGaP Accession No. phs000424.v7.p2). Expression values are shown in transcripts per million (TPM), calculated from a gene model with isoforms collapsed to a single gene. Box plots are shown as median and 25th and 75th percentiles; points are displayed as outliers if they are above or below 1.5 times the interquartile range.
Supplementary Figure 20. Expression of (A) GNA12 and (B) TRIO in 53 human tissues. The expression data were obtained from GTEx analysis release V7 (dbGaP Accession No. phs000424.v7.p2). Expression values are shown in transcripts per million (TPM), calculated from a gene model with isoforms collapsed to a single gene. Box plots are shown as median and 25th and 75th percentiles; points are displayed as outliers if they are above or below 1.5 times the interquartile range.
Supplementary Figure 21. Expression of (A) HTT and (B) RPS6KA2 in 53 human tissues. The expression data were obtained from GTEx analysis release V7 (dbGaP Accession No. phs000424.v7.p2). Expression values are shown in transcripts per million (TPM), calculated from a gene model with isoforms collapsed to a single gene. Box plots are shown as median and 25th and 75th percentiles; points are displayed as outliers if they are above or below 1.5 times the interquartile range.
Supplementary Figure 22. Expression of (A) AKT3 and (B) CNTNAP2 in 53 human tissues. The expression data were obtained from GTEx analysis release V7 (dbGaP Accession No. phs000424.v7.p2). Expression values are shown in transcripts per million (TPM), calculated from a gene model with isoforms collapsed to a single gene. Box plots are shown as median and 25th and 75th percentiles; points are displayed as outliers if they are above or below 1.5 times the interquartile range.
Supplementary Figure 23. Expression of (A) TIAM2 and (B) ARHGEF3 in 53 human tissues. The expression data were obtained from GTEx analysis release V7 (dbGaP Accession phs000424.v7.p2). Expression values are shown in transcripts per million (TPM), calculated from a gene model with isoforms collapsed to a single gene. Box plots are shown as median and 25th and 75th percentiles; points are displayed as outliers if they are above or below 1.5 times the interquartile range.
Supplementary Figure 24. Expression of (A) RGS12 and (B) MAD1L1 in 53 human tissues. The expression data were obtained from GTEx analysis release V7 (dbGaP Accession No. phs000424.v7.p2). Expression values are shown in transcripts per million (TPM), calculated from a gene model with isoforms collapsed to a single gene. Box plots are shown as median and 25th and 75th percentiles; points are displayed as outliers if they are above or below 1.5 times the interquartile range.
Supplementary Figure 25. Expression of (A) PTPRN2 and (B) TBC1D14 in 53 human tissues. The expression data were obtained from GTEx analysis release V7 (dbGaP Accession No. phs000424.v7.p2). Expression values are shown in transcripts per million (TPM), calculated from a gene model with isoforms collapsed to a single gene. Box plots are shown as median and 25th and 75th percentiles; points are displayed as outliers if they are above or below 1.5 times the interquartile range.
Supplementary Figure 26. Expression of (A) CLCN6 and (B) PRKCZ in 53 human tissues. The expression data were obtained from GTEx analysis release V7 (dbGaP Accession phs000424.v7.p2). Expression values are shown in transcripts per million (TPM), calculated from a gene model with isoforms collapsed to a single gene. Box plots are shown as median and 25th and 75th percentiles; points are displayed as outliers if they are above or below 1.5 times the interquartile range.
Supplementary Figure 27. Summary of candidate genes in potentially druggable categories. The numbers of these genes in potentially druggable categories, and the numbers of genes in these categories that are targeted by a known drug.
Supplementary Figure 28. Plots summarizing significant gene expression alterations in quetiapine (at doses of 10 or 100 mg/kg)-treated mice. One-way ANOVA was applied, and Tukey-Kramer HSD test was used for multiple comparisons. Gene expression data from the GEO dataset (Accession No. GSE45229). (A) For Htt; (B) for Cttnap2; (C) for Zmiz1; (D) for Trio; (E) for Notch1; (F) for Gna2. These plots were generated using the beeswarm package in R.
Supplementary Figure 29. Plots summarizing significant gene expression alterations in nicotine (at doses of 8 μg nicotine/l, 15 μg nicotine/l, and 23 μg nicotine/l)-treated mice. One-way ANOVA was applied, and Tukey-Kramer HSD test was used for multiple comparisons. Gene expression data from the GEO dataset (Accession No. GSE50254). (A) For Ptprn2; (B) for Ap2a2; (C) for Kcnq1; (D) for Akt3; (E) for Pard3; (F) for Plek; (G) for Mad1l1; (H) for Rps6ka2; (I) for Rgs12; (J) for Tnf; (K) for Htt; (L) for Cdkn1a. These plots were generated using the beeswarm package in R.
Supplementary Figure 30. Heatmap of the identified candidate genes treated by psychotropic drugs. Colored rectangles represent transcript abundance indicated above the gene labeled on the right. The intensity of the color is proportional to the standardized values between -3 and 3 for each gene, as indicated on the bar on the right of the heat map plot. Clustering was performed using Euclidean distance according to the scale on the left. Major gene transcription patterns are arbitrarily described as (A–C) Gene expression data from the GEO dataset (Accession No. GSE50254). Three antidepressants (bupropion 20 mg/kg, tranylcypromine 20 mg/kg, mianserin 20 mg/kg, i.p.), three anxiolytics (diazepam 5 mg/kg, buspirone 10 mg/kg, hydroxyzine 10 mg/kg, i.p.), and three antipsychotics (clozapine 3 mg/kg, risperidone 0.5 mg/kg, haloperidol 1 mg/kg) were selected for the comparison. To analyze dynamics of early, intermediate, and relatively late changes of mRNA abundance, the experiment was performed at four time points (1, 2, 4, and 8h after drug administration). To exclude influence of drug injection and circadian rhythm on gene expression profile, control groups of saline or Tween (1% Tween 80)-treated and naive animals were prepared for each time point. Heatmap plots were generated using the pheatmap package in R.
Supplementary Figure 31. Heatmap of the identified candidate genes treated by psychotropic drugs. Colored rectangles represent transcript abundance indicated above the gene labeled on the right. The intensity of the color is proportional to the standardized values between -3 and 3 for each gene, as indicated on the bar on the right of the heat map plot. Clustering was performed using Euclidean distance according to the scale on the left. Major gene transcription patterns are arbitrarily described as (A–D) Gene expression data from the GEO dataset (Accession No. GSE48851). Three antidepressants (imipramine 10 mg/kg, fluoxetine 20 mg/kg, and tianeptine 20 mg/kg, i.p.) were selected for the comparison. To analyze dynamics of early, intermediate, and relatively late changes of mRNA abundance, the experiment was performed at four time points (1, 2, 4, and 8h after drug administration). To exclude influence of drug injection and circadian rhythm on gene expression profile, control groups of saline or Tween (1% Tween 80)-treated and naïve animals were prepared for each time point. Heatmap plots were generated using the pheatmap package in R.
Supplementary Figure 32. Heatmap of the identified candidate genes treated by psychotropic drugs. Colored rectangles represent transcript abundance indicated above the gene labeled on the right. The intensity of the color is proportional to the standardized values between -3 and 3 for each gene, as indicated on the bar on the right of the heatmap plot. Clustering was performed using Euclidean distance according to the scale on the left. Major gene transcription patterns are arbitrarily described as (A–C). Gene expression data from the GEO dataset (Accession No. GSE15774). Six of the most addictive and harmful drugs of abuse (morphine 20 mg/kg, heroin 10 mg/kg, ethanol 2 g/kg, nicotine 1 mg/kg, methamphetamine 2 mg/kg, or cocaine 25 mg/kg, i.p.) were selected for comparison. To analyze dynamics of early, intermediate, and relatively late changes of mRNA abundance, the experiment was performed at four time points (1, 2, 4, and 8 h after drug administration). To exclude influence of drug injection and circadian rhythm on gene expression profile, control groups of saline-treated and naïve animals were prepared for each time point. Heatmap plots were generated using the pheatmap package in R.
Supplementary Figure 33. Variants annotation of 3,483 SNPs shared by SCZ and at least one phenotype of smoking behavior. We used snpEff (http://snpeff.sourceforge.net/) to carry out the annotation analysis.

Supplementary Figure 34. 3843 SNPs shared by SCZ and smoking located within different types of regulatory elements in brain tissues and neuroblastoma cell lines. (A) For brain tissues. (B) for SK-N-MC cells. (C) for SK-N-SH. The data of different types of regulatory elements, including H2AFZ, H3F3A, H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me2, H3K4me3, H3K79me2, H3K9ac, H3K9me2, H3K9me3, and H4K20me1, were downloaded from ENCODE database (https://www.encodeproject.org/).
Supplementary Figure 35. 933 of 3843 SNPs shared by SCZ and smoking having 4313 eQTLs in 31 human tissues. (A) 4313 eQTLs distributed in 31 human tissues. (B) Variant annotation of 933 SNPs shared by SCZ and at least one phenotype of smoking behavior. We used the tool snpEff (http://snpeff.sourceforge.net/) to carry out the annotation analysis. Based on the tissue-specific eQTL association data from the GTEx database, 933 of 3,483 SNPs have cis-regulatory effects on gene expression with 4,313 SNP-eQTLs pairs in 31 human tissues.
Supplementary Figure 36. Volcano plot for significant SNP–CpG methylation associations in brain samples. Green represents these significant SNP-CpG pairs with negative beta value. Red represents these significant SNP-CpG pairs with positive beta value.
Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1, 2, 3, 5, 7.

Supplementary Table 1. Multiple omics datasets used in current study downloaded from public databases.

Supplementary Table 2. 84 significantly enriched pathways shared between schizophrenia and smoking behaviors.

Supplementary Table 3. Shared genes between schizophrenia and smoking behaviors.

Supplementary Table 4. Module-level statistics.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Modules</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Blue</td>
</tr>
<tr>
<td>Orbital prefrontal (OFC)</td>
<td>4.93</td>
</tr>
<tr>
<td>Dorsolateral prefrontal (DFC)</td>
<td>4.95</td>
</tr>
<tr>
<td>Ventrolateral prefrontal (VFC)</td>
<td>4.93</td>
</tr>
<tr>
<td>Medial prefrontal (MFC)</td>
<td>4.94</td>
</tr>
<tr>
<td>Primary motor (M1C)</td>
<td>4.95</td>
</tr>
<tr>
<td>Primary somatosensory (S1C)</td>
<td>4.97</td>
</tr>
<tr>
<td>Inferior parietal (IPC)</td>
<td>4.97</td>
</tr>
<tr>
<td>Primary auditory (A1C)</td>
<td>4.96</td>
</tr>
<tr>
<td>Superior temporal (STC)</td>
<td>4.97</td>
</tr>
<tr>
<td>Inferior temporal (ITC)</td>
<td>4.97</td>
</tr>
<tr>
<td>Primary visual (V1C)</td>
<td>4.96</td>
</tr>
<tr>
<td>Hippocampus (HIP)</td>
<td>4.95</td>
</tr>
<tr>
<td>Amygdala (AMY)</td>
<td>4.96</td>
</tr>
<tr>
<td>Striatum (STR)</td>
<td>4.94</td>
</tr>
<tr>
<td>Thalamus (MD)</td>
<td>4.95</td>
</tr>
<tr>
<td>Cerebellum (CBC)</td>
<td>4.93</td>
</tr>
<tr>
<td>max(log2(expression))</td>
<td>4.97</td>
</tr>
<tr>
<td>min(log2(expression))</td>
<td>4.93</td>
</tr>
<tr>
<td>max(log2(FC))</td>
<td>0.05</td>
</tr>
<tr>
<td>8-10 weeks (stage2)</td>
<td>4.92</td>
</tr>
<tr>
<td>10-13 weeks (stage3)</td>
<td>5.01</td>
</tr>
<tr>
<td>13-16 weeks (stage4)</td>
<td>5.01</td>
</tr>
<tr>
<td>16-19 weeks (stage5)</td>
<td>4.97</td>
</tr>
<tr>
<td>19-24 weeks (stage6)</td>
<td>4.88</td>
</tr>
<tr>
<td>24-38 weeks (stage7)</td>
<td>5.13</td>
</tr>
<tr>
<td>0-0.5 years (stage8)</td>
<td>5.16</td>
</tr>
<tr>
<td>0.5-1 years (stage9)</td>
<td>5.25</td>
</tr>
<tr>
<td>1-6 years (stage10)</td>
<td>5.26</td>
</tr>
<tr>
<td>6-12 years (stage11)</td>
<td>5.04</td>
</tr>
<tr>
<td>12-20 years (stage12)</td>
<td>5.09</td>
</tr>
<tr>
<td>20-40 years (stage13)</td>
<td>5.08</td>
</tr>
<tr>
<td>40-60 years (stage14)</td>
<td>5.21</td>
</tr>
<tr>
<td>60+ years (stage15)</td>
<td>5.19</td>
</tr>
<tr>
<td>max(log2(expression))</td>
<td>5.26</td>
</tr>
<tr>
<td>min(log2(expression))</td>
<td>4.88</td>
</tr>
<tr>
<td>max(log2(FC))</td>
<td>0.38</td>
</tr>
</tbody>
</table>

Note: For both regional expression patterns and temporal expression patterns, we used the method of mean log₂(expression) of all genes in each module.
Supplementary Table 5. 149 coexpression module genes involved in smoking- and SCZ-associated methylation.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Description</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKT3</td>
<td>Genetic variation in the AKT3 locus (chr1:243503719–244002945) is a top GWAS signal in schizophrenia and pathway analysis identified 50 single nucleotide polymorphisms (SNPs) within the AKT3 gene that contribute to four of the top pathways associated with risk for schizophrenia and bipolar disorder. AKT3 shows prenatal enrichment during human neocortical development and recurrent copy number variations involving the 1q43-44 locus are associated with cortical malformations and intellectual disability, implicating an essential role in early brain development.</td>
<td>[1, 14–17]</td>
</tr>
<tr>
<td>AP2A2</td>
<td>Schizophrenia-based genetic association study shows the involvement of the clathrin-mediated endocytosis (CME)-related protein enthropin encoded by the clathrin interactor 1 (CLINT1) gene. The expression of genes encoding adaptor and clathrin assembly proteins, AP2A2, AP2B1, AP180, CLINT1, HIP1, ITSN2, and PICALM, increased relative to the control in SH-SY5Y cells incubated with 5–10 µmol/l clozapine for 24–72 h.</td>
<td>[18]</td>
</tr>
<tr>
<td>CACNA1D</td>
<td>The animal model of neonatal lesion in ventral hippocampus (NLVH) is a recognized animal model for schizophrenia. NLVH influenced change expressions in various genes involved in Ca^{2+} homeostasis, including Cacna1d, Atp2a2, Atdc2, Ppp3cb, and Ptk2b.</td>
<td>[19–21]</td>
</tr>
<tr>
<td>CDKN1A</td>
<td>Quetiapine is an atypical neuroleptic with a pharmacological profile distinct from classic neuroleptics that function primarily via blockade of dopamine D2 receptors. In the United States, quetiapine is currently approved for treating patients with schizophrenia, major depression and bipolar I disorder. Despite its widespread use, its cellular effects remain elusive. To address possible mechanisms, we chronically treated mice with quetiapine, haloperidol or vehicle and examined quetiapine-specific gene expression change in the frontal cortex. Through microarray analysis, we observed that several groups of genes were differentially expressed upon exposure to quetiapine compared with haloperidol or vehicle; among them, Cdkn1a, the gene encoding p21, exhibited the greatest fold change relative to haloperidol. The quetiapine-induced downregulation of p21/Cdkn1a was confirmed by real-time polymerase chain reaction and in situ hybridization. Consistent with single gene-level analyses, functional group analyses also indicated that gene sets associated with cell cycle/fate were differentially regulated in the quetiapine-treated group. In cortical cell cultures treated with quetiapine, p21/Cdkn1a was significantly downregulated in oligodendrocyte precursor cells and neurons, but not in astrocytes. We propose that cell cycle-associated intervention by quetiapine in the frontal cortex may underlie a unique efficacy of quetiapine compared with typical neuroleptics.</td>
<td>[12]</td>
</tr>
<tr>
<td>CNTNAP2</td>
<td>Contactin associated protein-like 2 (CNTNAP2) has emerged as a prominent susceptibility gene implicated in multiple complex neurodevelopmental disorders, including autism spectrum disorders (ASD), intellectual disability (ID), and schizophrenia (SCZ). Based on genomic rearrangements and copy number variations, the contactin-associated protein-like 2 gene (CNTNAP2) has been implicated in neurodevelopmental disorders such as Gilles de la Tourette syndrome, intellectual disability, obsessive compulsive disorder, cortical dysplasia-focal epilepsy syndrome, autism, schizophrenia, Pitt-Hopkins syndrome, and attention deficit hyperactivity disorder.</td>
<td>[22–24]</td>
</tr>
<tr>
<td>GNA12</td>
<td>GNA12 is significantly associated with schizophrenia based on network-assisted investigation of combined causal signals from GWAS studies in schizophrenia.</td>
<td>[25]</td>
</tr>
<tr>
<td>GPSM3</td>
<td>GPSM3 is significantly associated with schizophrenia based on a multi-stage schizophrenia GWAS.</td>
<td>[1]</td>
</tr>
<tr>
<td>HRH1</td>
<td>HRH1 is significantly associated with schizophrenia based on genetic association study.</td>
<td>[26, 27]</td>
</tr>
<tr>
<td>KCNQ1</td>
<td>Patients with schizophrenia show decreased processing speed on neuropsychological testing and decreased white matter integrity as measured by diffusion tensor imaging, two traits shown to be both heritable and genetically associated indicating that there may be genes that influence both traits as well as schizophrenia disease risk. The potassium channel gene family is a reasonable candidate to harbor such a gene given the prominent role potassium channels play in the central nervous system in signal transduction, particularly in myelinated axons. KCNQ1 may contribute to the shared risk for diminished processing speed, diminished white matter integrity and increased risk of</td>
<td>[28, 29]</td>
</tr>
</tbody>
</table>
MAD1L1

Multiple lines of evidence from genetic association studies indicate MAD1L1 confer risk to schizophrenia. [1, 30, 31]

NOS1AP

NOS1AP is a protein implicated in schizophrenia. Several independent studies reported linkage of schizophrenia to chromosome 1q21–22, containing NOS1AP. Many molecular functional-based and genetic-based studies have been identified NOS1AP gene as a schizophrenia susceptibility gene. [32–38]

NOTCH1

The antipsychotic and myelin protective effects of quetiapine are mediated by Notch signaling in a mouse model of cuprizone-induced demyelination associated with schizophrenia-like behaviors. The Notch pathway might therefore be a novel target for the development of antipsychotic drugs. [39]

PARD3

Based on a genetic-based association study, PARD3 is associated with susceptibility to schizophrenia in a Korean population. [40]

RGS12

RGS12 is putative candidate genes for sporadic schizophrenia. [41]

TNF

TNF-α is associated with the deficit syndrome and negative symptoms in patients with chronic schizophrenia. [42, 43]

TRIO

Disrupted-in-Schizophrenia 1-mediated axon guidance involves TRIO-RAC-PAK small GTPase pathway signaling. [44]

Supplementary Table 7. Differential expressed genes in SCZ patients and controls divided by smoking status.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>SCZ VS CTRL</th>
<th>Smoking VS Nonsmoking</th>
<th>Anova P value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTPRN2</td>
<td>0.0000518</td>
<td>0.316</td>
<td>0.00142</td>
</tr>
<tr>
<td>ARHGAP25</td>
<td>0.0001</td>
<td>0.0028</td>
<td>0.00026</td>
</tr>
<tr>
<td>ARHGEF3</td>
<td>0.000281</td>
<td>0.11</td>
<td>0.00723</td>
</tr>
<tr>
<td>AKT3</td>
<td>0.00111</td>
<td>0.381</td>
<td>0.0395</td>
</tr>
<tr>
<td>RGS12</td>
<td>0.00202</td>
<td>0.347</td>
<td>0.0058</td>
</tr>
<tr>
<td>SORBS1</td>
<td>0.0134</td>
<td>0.471</td>
<td>0.108</td>
</tr>
<tr>
<td>TIA1M2</td>
<td>0.0161</td>
<td>0.954</td>
<td>0.062</td>
</tr>
<tr>
<td>CACNA1D</td>
<td>0.0162</td>
<td>0.403</td>
<td>0.052</td>
</tr>
<tr>
<td>HTT</td>
<td>0.0179</td>
<td>0.259</td>
<td>0.132</td>
</tr>
<tr>
<td>NOTCH1</td>
<td>0.0239</td>
<td>0.301</td>
<td>0.156</td>
</tr>
<tr>
<td>PLEK</td>
<td>0.0256</td>
<td>0.0383</td>
<td>0.043</td>
</tr>
<tr>
<td>RPS6KA2</td>
<td>0.0268</td>
<td>0.815</td>
<td>0.112</td>
</tr>
<tr>
<td>PARD3</td>
<td>0.0296</td>
<td>0.033</td>
<td>0.054</td>
</tr>
<tr>
<td>TRIO</td>
<td>0.0326</td>
<td>0.324</td>
<td>0.147</td>
</tr>
<tr>
<td>CNTNAP2</td>
<td>0.0344</td>
<td>0.062</td>
<td>0.087</td>
</tr>
<tr>
<td>GPSM3</td>
<td>0.0369</td>
<td>0.353</td>
<td>0.0014</td>
</tr>
<tr>
<td>ZMIZI</td>
<td>0.0708</td>
<td>0.499</td>
<td>0.041</td>
</tr>
<tr>
<td>TBC1D14</td>
<td>0.845</td>
<td>0.0389</td>
<td>0.030</td>
</tr>
</tbody>
</table>

Note: anova statistical test was used for comparing the significant differences between SCZ patients and controls grouped by smoking status.