GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia


Cannabis use is a heritable trait that has been associated with adverse mental health outcomes. In the largest genome-wide association study (GWAS) for lifetime cannabis use to date (N = 184,765), we identified eight genome-wide significant independent single nucleotide polymorphisms in six regions. All measured genetic variants combined explained 11% of the variance. Gene-based tests revealed 35 significant genes in 16 regions, and S-PrediXcan analyses showed that 21 genes had different expression levels for cannabis users versus nonusers. The strongest finding across the different analyses was CADM2, which has been associated with substance use and risk-taking. Significant genetic correlations were found with 14 of 25 tested substance use and mental health–related traits, including smoking, alcohol use, schizophrenia and risk-taking. Mendelian randomization analysis showed evidence for a causal positive influence of schizophrenia risk on cannabis use. Overall, our study provides new insights into the etiology of cannabis use and its relation with mental health.
Here we present a GWAS of lifetime cannabis use on a substantially larger sample, providing more power to identify genetic variants. As mentioned above, cannabis use has been linked to a variety of mental health outcomes, including substance abuse and dependence and psychiatric disorders. In particular, the relationship between cannabis use and schizophrenia has been the subject of intensive research and debate. It has long been established that the prevalence of cannabis use is higher in patients with schizophrenia, but other hypotheses (namely, schizophrenia increases the use of cannabis, or the association is due to genetic pleitropy) have also been posed. Previous studies have shown that genetic risk factors for cannabis use and schizophrenia are positively correlated. However, a genetic correlation does not provide insight in the direction of causation. With Mendelian randomization it is possible to examine the causality of the association between cannabis use and schizophrenia, and recently it has become possible to apply this method using summary statistics from GWASs. Previous Mendelian randomization studies have investigated the link between lifetime cannabis use and schizophrenia, but findings were inconsistent. Vaucher et al. tested for causal effects from cannabis use to schizophrenia and found evidence for a causal influence of cannabis use on schizophrenia risk. Gage et al. tested bidirectional effects and found weak evidence for a causal effect of cannabis use on schizophrenia and much stronger evidence for a causal effect in the other direction. The results from our GWAS provide more power to examine the causal association between cannabis use and schizophrenia.

Here we report the largest GWAS yet for lifetime cannabis use. We increased the sample size substantially by meta-analyzing GWAS results from the ICC study (N = 35,297), along with new data from the UK Biobank (N = 126,785) and 23andMe (N = 22,683). The combined sample size of this study was N = 184,765, five times as large as the previous largest GWAS on lifetime cannabis use. We tested the association of millions of SNPs with lifetime cannabis use and estimated the heritability of lifetime cannabis use based on all SNPs. Tests of association for individual genetic variants were complemented with gene-based tests of association and S-PrediXcan analysis. The latter was used to identify genes with differential expression levels in cannabis users versus nonusers. We further estimated the genetic correlation of lifetime cannabis use with other traits, including use of other substances and mental health traits, such as schizophrenia. Lastly, we performed bidirectional two-sample Mendelian randomization analysis to examine whether there was evidence for a causal relationship from cannabis use to schizophrenia and vice versa.

Results
Genome-wide association meta-analysis. The meta-analysis resulted in eight independent genome-wide significant SNP associations (linkage disequilibrium (LD) R^2 < 0.1, window size 250 kb) on chromosomes 3, 7, 8, 11, 16 and 17 (Fig. 1, Table 1 and Supplementary Table 1). The top SNP and two other independent associations were located in CADM2 on chromosome 3 (rs2875907, P = 9.38 × 10^{-17}; rs1448602, P = 6.55 × 10^{-14}; rs7651996, P = 2.37 × 10^{-9}). Other hits were located in ZNF704, SDK1, NCAM1, RABEP2 or ATP2A1 and SMG6 (Fig. 2). All SNPs combined explained 11% (h^2 = 0.11, s.e. = 0.01) of the individual differences in lifetime cannabis use. Supplementary Figs. 1–3 and Supplementary Table 2 provide information on results of the individual GWASs (ICC, UK Biobank and 23andme).

Gene-based test of association and expression. Gene-based tests of associations in MAGMA identified 35 genes genome-wide significantly associated with lifetime cannabis use (Fig. 3, Table 2, Supplementary Fig. 4 and Supplementary Table 3). These genes were located in 5 regions that were already identified in the SNP-based analysis (including those containing CADM2 and NCAM1) and in 11 other regions (Supplementary Fig. 5).

S-PrediXcan analysis revealed 133 Bonferroni-corrected significant associations across tissues targeting 21 unique genes (Supplementary Tables 4 and 5). Eight genes were also significant in the gene-based test, whereas 13 were newly identified. For genes identified in multiple tissues, directions of effects were largely consistent across tissues (Supplementary Fig. 6). Again, the most significant finding was CADM2; genetic variants associated with increased liability to use cannabis are predicted to upregulate expression levels of CADM2 in eight nonbrain tissues, including whole blood (z = 5.88, P = 4.17 × 10^{-9}). Of note, although CADM2 is expressed more widely in brain than in other tissues (Supplementary Fig. 7), the top SNP, rs2875907, regulates the expression of CADM2 only in nonbrain tissues (Supplementary Fig. 8). Exploration of S-PrediXcan results in UK Biobank data (https://imlab.shinyapps.io/gene2pheno_ukb_neale/) showed that CADM2 expression is significantly associated with multiple traits, including increased risk-taking, body mass index and reduced feelings of anxiety. Like the SNP- and gene-based tests of association, the S-PrediXcan analysis detected a strong signal in a high-LD region at 16p11.2. Supplementary Table 3 provides an overview of all genes that were identified in the gene-based test of association and the S-PrediXcan analyses, along with information about the gene product and previously identified associations with the gene.

Genetic correlations with other traits. Using our GWAS results and those of other GWASs, we estimated the genetic correlation of lifetime cannabis use with 25 traits of interest, including substance use, personality and mental health phenotypes. Fourteen traits were significantly genetically correlated with lifetime cannabis use after correction for multiple testing (Fig. 4 and Supplementary Table 6). Positive genetic correlations were found with substance use phenotypes, including smoking and alcohol use and dependence, as well as with mental health phenotypes, including ADHD and schizophrenia. Furthermore, positive genetic correlations were found with risk-taking behavior, openness to experience, and educational attainment, as well as a negative correlation with conscientiousness.

Causal association between cannabis use and schizophrenia: two-sample Mendelian randomization. A positive genetic correlation was found between genetic risk factors for cannabis use and schizophrenia (r^2 = 0.24, s.e. = 0.03, P < 0.01). To examine whether there was evidence for a causal effect of cannabis use on schizophrenia risk, we performed bidirectional two-sample Mendelian randomization analysis. In our main analysis, inverse-variance-weighted (IVW) regression analysis, we found some weak (nonsignificant) evidence for a causal influence of lifetime cannabis use on schizophrenia risk, but only for the genetic instrument containing SNPs associated with cannabis use under the P-value threshold 1 × 10^{-5}. The IVW regression odds ratio was 1.10 (95% confidence interval (CI) 0.99–1.21, P = 0.074). We found stronger evidence for a causal positive influence of schizophrenia risk on lifetime cannabis use, the IVW regression odds ratio being 1.16 (95% CI 1.06–1.27, P = 0.001; see Table 3, Supplementary Figs. 9 and 10, and Supplementary Tables 7–9).

To determine the robustness of these findings, we performed four sensitivity analyses that rely on distinct assumptions regarding instrument validity. The sensitivity analyses showed a consistent pattern supporting weak evidence for a causal effect of cannabis use on schizophrenia and strong evidence for a causal effect of schizophrenia on cannabis use (Table 3). As an exception, the evidence provided by MR-Egger SIMEX (Mendelian randomization Egger simulation extrapolation) for a causal relation from schizophrenia risk to cannabis use was very weak. However, since the Egger intercept
Articles Nature Neuroscience was not significantly different from 0 (Supplementary Table 10), indicating no pleiotropic effects for the SNPs included in the genetic instruments27, it is likely that this method simply lacked power to be able to reject the null hypothesis of no causal effect28.

Discussion
SNP- and gene-based tests revealed several SNPs and genes strongly associated with lifetime cannabis use. Overall, 11% of the variation in the phenotype was explained by the combined effect of SNPs, which amounts to approximately 25% of twin-based heritability estimates4.

**Table 1** | Association results of eight independent SNPs that are significantly associated with lifetime cannabis use

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Gene</th>
<th>BP</th>
<th>A1</th>
<th>A2</th>
<th>Freq A1</th>
<th>N</th>
<th>β</th>
<th>SE</th>
<th>P-value</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2875907</td>
<td>3p21.1</td>
<td>CADM2</td>
<td>85,518,580</td>
<td>A</td>
<td>G</td>
<td>0.352</td>
<td>181,675</td>
<td>0.070</td>
<td>0.009</td>
<td>9.38 × 10⁻¹⁷</td>
<td>+++</td>
</tr>
<tr>
<td>rs1448602</td>
<td>3p21.1</td>
<td>CADM2</td>
<td>85,780,454</td>
<td>A</td>
<td>G</td>
<td>0.756</td>
<td>184,765</td>
<td>-0.062</td>
<td>0.010</td>
<td>6.55 × 10⁻⁻⁵</td>
<td>- - -</td>
</tr>
<tr>
<td>rs7651996</td>
<td>3p21.1</td>
<td>CADM2</td>
<td>85,057,349</td>
<td>T</td>
<td>G</td>
<td>0.477</td>
<td>184,765</td>
<td>0.049</td>
<td>0.008</td>
<td>2.37 × 10⁻⁻⁹</td>
<td>+++</td>
</tr>
<tr>
<td>rs10085617</td>
<td>7p22.2</td>
<td>SDK1</td>
<td>3,634,711</td>
<td>A</td>
<td>T</td>
<td>0.416</td>
<td>184,765</td>
<td>0.046</td>
<td>0.008</td>
<td>2.93 × 10⁻⁻⁸</td>
<td>+++</td>
</tr>
<tr>
<td>rs9773390</td>
<td>8q21.1</td>
<td>ZNF704</td>
<td>81,565,692</td>
<td>T</td>
<td>C</td>
<td>0.933</td>
<td>44,595</td>
<td>-0.171</td>
<td>0.029</td>
<td>5.66 × 10⁻⁻⁵</td>
<td>- - ?</td>
</tr>
<tr>
<td>rs9919557</td>
<td>11q23.2</td>
<td>NCAM1</td>
<td>112,877,408</td>
<td>T</td>
<td>C</td>
<td>0.614</td>
<td>180,428</td>
<td>-0.055</td>
<td>0.009</td>
<td>9.94 × 10⁻⁻¹</td>
<td>- - -</td>
</tr>
<tr>
<td>rs10499</td>
<td>16p11.2</td>
<td>RABEP2, ATP2A1</td>
<td>28,915,527</td>
<td>A</td>
<td>G</td>
<td>0.651</td>
<td>179,767</td>
<td>0.053</td>
<td>0.009</td>
<td>1.13 × 10⁻⁻³</td>
<td>+++</td>
</tr>
<tr>
<td>rs17761723</td>
<td>17p13.3</td>
<td>SMG6</td>
<td>2,107,090</td>
<td>T</td>
<td>C</td>
<td>0.346</td>
<td>184,765</td>
<td>0.047</td>
<td>0.009</td>
<td>3.24 × 10⁻⁻⁹</td>
<td>+++</td>
</tr>
</tbody>
</table>

Independent hits were defined as $R^2 < 0.01$, window size 250 kb. The threshold was set at $P < 5 × 10^{-8}$ (conventional genome-wide significant threshold; significance was tested two-sided). Table gives chromosomal region (Chr), gene the SNP is located in or the nearest gene (within 500 kb), base pair (BP) location SNP on Hg19, allele 1 (A1), allele 2 (A2), frequency of allele 1 (Freq A1), number of individuals for which variant was included (N), β coefficient of the effect allele A1, standard error (SE) of the β coefficient, and direction for each sample: allele A1 increases (+) or decreases (-) liability for cannabis use, or sample did not contribute to this SNP (?). Order of samples within the Direction column, from left to right: ICC, 23andMe, UK Biobank. Independent SNPs were selected as SNPs with linkage disequilibrium $R^2 < 0.1$ using a window size of 250 kb. SNP rs9773390 was not present in the UK Biobank sample and its effect is rather isolated (see Figs. 1b and 2); it might not represent a robust association.
Cannabis use has previously been associated with related personality traits, including high levels of impulsivity and novelty seeking13,14. Taken together, these findings suggest that risk variants in CADM2 are associated with a broad profile of a risk-taking, optimistic and care-free personality15. CADM2 showed that was associated with various personality traits, with the risk variant being associated with reduced anxiety, neuroticism and conscientiousness and with increased risk-taking14. Taken together, these findings suggest that risk variants in CADM2 are associated with related personality traits, including high levels of impulsivity and novelty seeking13,14.

NCAM1 (neural cell adhesion molecule 1) also encodes a cell adhesion protein and is member of the immunoglobulin superfamily. The encoded protein is involved in cell–matrix interactions and cell differentiation during development16,17. NCAM1 is located in the NCAM1–TTC12–ANKK1–DRD2 gene cluster, which is related to neurogenesis and dopaminergic neurotransmission. This gene cluster has been associated with smoking, alcohol use and illicit drug use13,15,16,18 and has been implicated in psychiatric disorders, such as schizophrenia and mood disorders16,17.

A putatively novel finding comprises the 16p11.2 region (identified in the SNP and gene-based tests of association and in S-PrediXcan analysis). Deletions and duplications in this region have previously been reported to be associated with autism and schizophrenia19,20, while a common 16p11.2 inversion underlies susceptibility to asthma and obesity20. The inversion explains a substantial proportion of variability in expression of multiple genes in this region, including TUFM and SH2B120. Given the high LD in this region and high levels of coexpression of the differentially expressed genes, follow-up studies will be needed to determine which genes are functionally driving the association with cannabis use.

Several of the top genes from the gene-based and/or S-PrediXcan analyses have previously shown an association with other traits, including schizophrenia (for example, TUFM, NCAM1), body mass index or obesity (for example, SH2B1, APOBR, ATXN2L),
alcohol use (for example, ALDH2), intelligence and cognitive performance (CNNM2, CCDC101) and externalizing and impulsive phenotypes (for example, CADM2; see Supplementary Table 3). Also of note is the association with \( HTR1A \); this gene has been implicated in alcohol and nicotine codependence\(^{41} \), body mass index\(^{42} \), psychiatric disorders\(^{43,44} \) and antipsychotic pharmacological treatment response\(^{45} \). At the phenotypic level, associations between cannabis use and psychiatric disorders\(^2 \) and use of other substances\(^{10} \) are well established.

There are two previous studies that found significant SNP associations for a cannabis use phenotype. Sherva et al.\(^6 \) found three SNPs significantly associated with cannabis dependence. In our results only one of the SNPs was available (rs77378271) and was not significantly associated with lifetime cannabis use (\( P = 0.144 \)). The other two SNPs (rs143244591 and rs146091982) or their high-LD proxies were not available in our data. The SNPs rs77378271 and rs146091982 were located in genes CSDM1 and SLC35G1, respectively, and neither of those were significant in our gene-based results (\( P = 0.96 \) and \( P = 0.49 \), respectively). Demontis et al.\(^{11} \) found one independent significant signal on chromosome 8 to be associated with cannabis dependence (with top SNP rs56372821, a strong expression quantitative trait locus (eQTL) for CHRNA2). Neither the SNP (\( P = 0.55 \)) nor the gene (\( P = 0.52 \)) was significantly associated with lifetime cannabis use in our study. The protein encoded by CHRNA2 is a subunit of certain nicotinic acetylcholine receptors, and Demontis et al.\(^{11} \) offer three potential biological explanations for the link between cannabis intake and CHRNA2. However, it is possible that while CHRNA2 is associated with cannabis dependence, it does not act in the initial stages of cannabis use, which are more related to personality and risk-taking behaviors and less to the actual effects of cannabis intake on the brain.

The genetic correlation analyses revealed genetic overlap of cannabis use with a broad range of traits, including positive associations with substance use and mental health phenotypes. Furthermore, positive genetic correlations were found with risk-taking behavior, openness to experience, and educational attainment, as well as a negative correlation with conscientiousness. The range of correlations suggests that genetic liability to lifetime cannabis use should be viewed in a broader context of personality and mental health traits. Specifically, the substantial genetic correlations with risk-taking behavior and openness to experience may indicate that liability to start using cannabis is an indication of one’s personality. The positive genetic correlation between lifetime cannabis use and educational attainment was unexpected and in contrast to a previous study that found a negative genetic correlation between cannabis dependence and educational attainment\(^11 \). We therefore investigated phenotypic associations of cannabis use with household income and fluid intelligence using UK Biobank data. Within Caucasian participants of UK Biobank (\( N = 438,870 \)), categorically rated household income was higher among lifetime cannabis users compared to nonusers (\( \chi^2(4) = 2,243, P = 2.2 \times 10^{-16} \)). Cannabis users also scored higher on fluid intelligence (\( t(50,856) = 25.13, P < 2 \times 10^{-16} \)). The findings are in agreement with observations by Patrick et al.\(^{47} \), who showed that cannabis use is associated with higher childhood family social economic status in a survey of US families. Possibly, environments more often experienced by those with backgrounds of higher social economic status, such as universities, increase accessibility to cannabis, explaining how a positive
**Table 2** | Genes significantly associated with lifetime cannabis use, as identified in the MAGMA and/or S-PrediXcan analyses

<table>
<thead>
<tr>
<th>Locus</th>
<th>Top genes</th>
<th>BP start</th>
<th>BP stop</th>
<th>SNPs</th>
<th>z</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1p36.31</td>
<td>KLHL21</td>
<td>6,640,784</td>
<td>6,672,958</td>
<td>96</td>
<td>4.81</td>
<td>7.65 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>PHF13</td>
<td>6,663,756</td>
<td>6,694,093</td>
<td>84</td>
<td>4.61</td>
<td>1.99 × 10⁻⁸</td>
</tr>
<tr>
<td>2p12</td>
<td>LRRTM4</td>
<td>76,969,849</td>
<td>77,754,502</td>
<td>3,621</td>
<td>5.19</td>
<td>1.03 × 10⁻⁷</td>
</tr>
<tr>
<td>3p12.1</td>
<td>CADM2</td>
<td>85,003,133</td>
<td>86,128,579</td>
<td>4,287</td>
<td>8.96</td>
<td>1.59 × 10⁻⁹</td>
</tr>
<tr>
<td>4p16.3</td>
<td>MSANTD1</td>
<td>3,240,766</td>
<td>3,283,465</td>
<td>231</td>
<td>4.59</td>
<td>2.22 × 10⁻⁶</td>
</tr>
<tr>
<td>5q12.3</td>
<td>HTR1A</td>
<td>63,245,875</td>
<td>63,268,119</td>
<td>64</td>
<td>4.57</td>
<td>2.41 × 10⁻⁴</td>
</tr>
<tr>
<td>6p12.1</td>
<td>BEND5</td>
<td>56,814,773</td>
<td>56,897,450</td>
<td>252</td>
<td>5.22</td>
<td>2.60 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>KIAA1586</td>
<td>56,906,343</td>
<td>56,925,023</td>
<td>58</td>
<td>5.09</td>
<td>1.75 × 10⁻²</td>
</tr>
<tr>
<td></td>
<td>RAB23</td>
<td>57,046,790</td>
<td>57,092,112</td>
<td>86</td>
<td>5.86</td>
<td>2.32 × 10⁻⁹</td>
</tr>
<tr>
<td>6q21</td>
<td>REV3L</td>
<td>111,610,234</td>
<td>111,814,421</td>
<td>539</td>
<td>4.61</td>
<td>1.99 × 10⁻⁴</td>
</tr>
<tr>
<td>6q25.3</td>
<td>ARIDB</td>
<td>157,093,980</td>
<td>157,356,913</td>
<td>1,344</td>
<td>5.59</td>
<td>1.15 × 10⁻⁸</td>
</tr>
<tr>
<td>8q24.3</td>
<td>ADGRB1</td>
<td>143,535,377</td>
<td>143,636,369</td>
<td>275</td>
<td>4.71</td>
<td>1.23 × 10⁻⁴</td>
</tr>
<tr>
<td>10q24.32-33</td>
<td>NEURL</td>
<td>103,493,890</td>
<td>103,592,552</td>
<td>17</td>
<td>5.22</td>
<td>1.83 × 10⁻⁷</td>
</tr>
<tr>
<td>112q12</td>
<td>BRAP</td>
<td>112,069,950</td>
<td>112,133,790</td>
<td>97</td>
<td>4.87</td>
<td>5.48 × 10⁻⁷</td>
</tr>
<tr>
<td>112q24</td>
<td>ACAD10</td>
<td>112,118,857</td>
<td>112,199,911</td>
<td>141</td>
<td>5.22</td>
<td>8.96 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>ALDH2</td>
<td>112,199,691</td>
<td>112,252,789</td>
<td>112</td>
<td>4.96</td>
<td>3.61 × 10⁻⁹</td>
</tr>
<tr>
<td></td>
<td>MAPKAPK5</td>
<td>112,275,032</td>
<td>112,336,228</td>
<td>195</td>
<td>4.87</td>
<td>5.58 × 10⁻⁷</td>
</tr>
<tr>
<td></td>
<td>TMEM116</td>
<td>112,364,086</td>
<td>112,456,023</td>
<td>222</td>
<td>4.94</td>
<td>3.96 × 10⁻⁸</td>
</tr>
<tr>
<td>16p11.2-16q12.1</td>
<td>SBKI</td>
<td>28,303,840</td>
<td>28,335,170</td>
<td>23</td>
<td>5.47</td>
<td>4.52 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>NPIPB7</td>
<td>28,467,693</td>
<td>28,481,868</td>
<td>10</td>
<td>5.44</td>
<td>5.46 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>CLN3</td>
<td>28,483,600</td>
<td>28,510,897</td>
<td>62</td>
<td>5.84</td>
<td>2.56 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>APOBR</td>
<td>28,500,970</td>
<td>28,515,291</td>
<td>49</td>
<td>5.66</td>
<td>7.56 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>IL27</td>
<td>28,505,683</td>
<td>28,523,155</td>
<td>57</td>
<td>5.66</td>
<td>7.48 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>CCDC101</td>
<td>28,560,249</td>
<td>28,608,111</td>
<td>181</td>
<td>4.90</td>
<td>4.87 × 10⁻²</td>
</tr>
<tr>
<td></td>
<td>SULT1A2</td>
<td>28,603,264</td>
<td>28,608,391</td>
<td>25</td>
<td>5.40</td>
<td>6.66 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>SULT1A1</td>
<td>28,605,196</td>
<td>28,623,625</td>
<td>51</td>
<td>5.30</td>
<td>1.14 × 10⁻⁷</td>
</tr>
<tr>
<td></td>
<td>CDC37P1</td>
<td>28,700,176</td>
<td>28,701,611</td>
<td>14</td>
<td>5.37</td>
<td>8.08 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>EIF3C</td>
<td>28,722,782</td>
<td>28,747,053</td>
<td>23</td>
<td>5.47</td>
<td>4.55 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>EIF3CL</td>
<td>28,722,785</td>
<td>28,747,053</td>
<td>14</td>
<td>5.37</td>
<td>8.08 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>NPIPB9</td>
<td>28,742,728</td>
<td>28,772,850</td>
<td>8</td>
<td>5.41</td>
<td>6.29 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>ATXN2L</td>
<td>28,829,369</td>
<td>28,853,558</td>
<td>89</td>
<td>5.85</td>
<td>2.50 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>TUFM</td>
<td>28,848,732</td>
<td>28,862,729</td>
<td>55</td>
<td>5.83</td>
<td>2.83 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>SH2B1</td>
<td>28,867,939</td>
<td>28,890,534</td>
<td>71</td>
<td>5.72</td>
<td>5.46 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>ATP2A1</td>
<td>28,884,192</td>
<td>28,920,830</td>
<td>89</td>
<td>5.97</td>
<td>1.20 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>NAFATC2IP</td>
<td>28,962,318</td>
<td>28,977,767</td>
<td>8</td>
<td>5.35</td>
<td>8.82 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>RABEP2</td>
<td>28,910,742</td>
<td>28,942,339</td>
<td>71</td>
<td>5.43</td>
<td>2.84 × 10⁻⁸</td>
</tr>
<tr>
<td>17q13.3</td>
<td>SRR</td>
<td>2,202,244</td>
<td>2,233,553</td>
<td>121</td>
<td>5.33</td>
<td>5.03 × 10⁻⁸</td>
</tr>
<tr>
<td>18q11.2</td>
<td>C18orf8</td>
<td>21,078,434</td>
<td>21,118,311</td>
<td>132</td>
<td>5.30</td>
<td>5.65 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td>NPC1</td>
<td>21,081,148</td>
<td>21,171,581</td>
<td>79</td>
<td>5.30</td>
<td>5.87 × 10⁻⁸</td>
</tr>
</tbody>
</table>

For the gene-based test, the P-value was set at P < 2.74 × 10⁻⁶. Bonferroni corrected threshold of P < 0.05 adjusted for 18,293 tests. For the S-PrediXcan analysis, P < 1.92 × 10⁻⁹. Bonferroni corrected threshold of P < 0.05 adjusted for 259,825 tests. The MAGMA statistical test is based on multiple regression. Significance was tested two-sided in both analyses. Genes that were significant in both analyses are bolded; the others were significant in the S-PrediXcan analysis alone. Table gives location in base pairs (hg19) at beginning and end of gene (BP start and BP stop, respectively), number of SNPs included in the gene, and test statistic for the test of association (z). The CDC37P1 gene was significant in two different tissues; information presented here is based on the association with the smallest P-value. For full results, see Supplementary Table 4.
correlation between lifetime cannabis use and educational attainment in our study could arise.

We also found a significant genetic correlation between cannabis use and schizophrenia ($r = 0.24$), which is in line with previous findings, indicating that genetic risk factors for cannabis use and schizophrenia are positively correlated. As for the causal direction of this correlation, we found weak evidence for a causal link from cannabis use to schizophrenia and much stronger evidence for a causal link from schizophrenia to cannabis use. This suggests that individuals with schizophrenia have a higher risk to start using cannabis. These results are in contrast with results from a Mendelian randomization study by Vaucher et al., who found strong evidence for a causal effect from cannabis use to schizophrenia (causality in the other direction was not tested). However, our findings are in line with a Mendelian randomization study by Gage et al., who used genetic instruments similar to ours and also found weak evidence for a causal effect of cannabis use to schizophrenia and much stronger evidence for a causal effect in the other direction. Our findings may indicate that individuals at risk for developing schizophrenia experience prodromal symptoms or negative affect that make them more likely to start using cannabis to cope or self-medicate. The lack of strong evidence of a causal influence of cannabis use on schizophrenia may be due to the lower power of the instrumental variables. The instrumental variable based on schizophrenia SNPs explained 3.38% of variance in liability to schizophrenia. For cannabis use, the genetic instruments explained 1.12% and 0.15% of the variance in cannabis use for SNPs included with $P < 1 \times 10^{-3}$ and $P < 5 \times 10^{-4}$, respectively.

The results of our study should be interpreted in view of its strength and limitations. Important strengths of this study include the analyses of the largest population sample to date, which has led to a substantial increase in power to identify genetic variants associated with lifetime cannabis use. The association analyses were complemented with several follow-up analyses to further investigate the genetic basis of cannabis use and the extent to which the genetic etiology of cannabis use overlaps with that of other complex phenotypes. Strong genetic correlations across a wide spectrum of traits are observed, confirming that lifetime cannabis use is a relevant measure of an individual's vulnerability.

Our study also has several limitations. First, lifetime cannabis use was analyzed as a dichotomous measure combining experimental and regular users in a single group. Additionally, the different samples varied substantially regarding the age of the participants, the prevalence of cannabis use, and the country's policies regarding cannabis use. All these factors may introduce heterogeneity that may reduce the power to detect genetic associations. Second, power of some analyses may have been limited. For example, the Mendelian randomization analysis from cannabis to schizophrenia was based on an instrument of only five SNPs, and the summary statistics of some traits used for the genetic correlation analyses in LD-score regression (for example, cannabis dependence) were based on a small sample size. Finally, some regions identified in the SNP-based analyses did not appear in the gene-based analyses. In particular, inspection of the region around rs9777390 (in ZNF704) showed that the top SNP in this region was isolated and that the SNP was only available in two of the three datasets (not in UK Biobank). SNPs in LD with the top SNP that were included in all three datasets were not genome-wide significant. Thus, this result may not represent a robust association.

In summary, our GWAS of lifetime cannabis use, which is the largest to date, revealed significant SNP and gene associations in 16

---

**Fig. 4** | Genetic overlap between lifetime cannabis use and other phenotypes. Blue dots represent point estimates of the genetic correlation, blue error bars represent 95% confidence intervals and red asterisks indicate significant associations after correction for multiple testing (two-sided $P < 0.002$, Bonferroni corrected threshold of $P < 0.05$ adjusted for 25 tests). MDD, major depressive disorder; ADHD, attention deficit hyperactivity disorder; BMI, body mass index.

---

**Table 3** | Results of the bidirectional two-sample Mendelian randomization analysis between lifetime cannabis use and schizophrenia, including results of four sensitivity analyses

<table>
<thead>
<tr>
<th></th>
<th>Cannabis-schizophrenia ($P &lt; 5 \times 10^{-4}$, 5 SNPs)</th>
<th>Cannabis-schizophrenia ($P &lt; 1 \times 10^{-5}$, 69 SNPs$^*$)</th>
<th>Schizophrenia–cannabis ($P &lt; 5 \times 10^{-4}$, 109 SNPs$^*$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$B$ ($SE(B)$) OR $P$-value</td>
<td>$B$ ($SE(B)$) OR $P$-value</td>
<td>$B$ ($SE(B)$) OR $P$-value</td>
</tr>
<tr>
<td>IVW</td>
<td>0.039 (0.158) 1.04 0.806</td>
<td>0.091 (0.051) 1.10 0.074</td>
<td>0.151 (0.046) 0.166</td>
</tr>
<tr>
<td>Weighted median</td>
<td>-0.048 (0.015) 0.95 0.649</td>
<td>0.069 (0.049) 1.07 0.156</td>
<td>0.163 (0.049) 0.17</td>
</tr>
<tr>
<td>MR-Egger SIMEX</td>
<td>-0.044 (0.190) 0.96 0.827</td>
<td>0.106 (0.110) 1.11 0.340</td>
<td>0.071 (0.293) 1.07</td>
</tr>
<tr>
<td>Weighted mode</td>
<td>-0.084 (0.125) 0.92 0.536</td>
<td>0.016 (0.071) 1.02 0.823</td>
<td>0.315 (0.178) 1.37</td>
</tr>
<tr>
<td>GSMR after HEIDI filtering</td>
<td>- - -</td>
<td>0.192 (0.080) 1.21 0.017</td>
<td>0.237 (0.038) 1.27</td>
</tr>
</tbody>
</table>

Significant results ($P < 0.05$, tested two-sided) are shown in bold. IVW, inverse-variance-weighted regression analysis; MR-Egger SIMEX, Mendelian randomization Egger simulation extrapolation; GSMR, generalized summary-data-based Mendelian randomization (HEIDI outlier analysis detects and eliminates from the analysis instruments that show significant pleiotropic effects on both risk factor and disease); $B$, risk coefficient representing the change in outcome for a one-unit increase in the exposure variable; $SE(B)$, standard error of the $B$ coefficient; OR, odds ratios representing the odds of schizophrenia for lifetime cannabis users versus nonusers (when cannabis is the exposure) or the odds of lifetime cannabis use for those with a schizophrenia diagnosis versus those without (when schizophrenia is the exposure). *Number of SNPs in instrument was 74 for the GSMR analysis. †Number of SNPs in instrument was 102 for the GSMR analysis.
regions, 14 of which have not been previously implicated in cannabis use. The most promising candidates for future functional studies are CADM2, NCAM1 and multiple genes located at 16p11.2. Our findings further indicated a causal influence of schizophrenia on cannabis use and substantial genetic overlap between cannabis use mental-health-related traits, personality traits and use of other substances, including smoking and alcohol use, schizophrenia, ADHD and risk-taking.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41593-018-0206-1.

Received: 12 January 2018; Accepted: 28 May 2018; Published online: 27 August 2018

References


Acknowledgements

We would like to thank the research participants and employees of 23andMe for making this work possible. We gratefully acknowledge the Psychiatric Genomics Consortium contributing studies and the participants in those studies without whom this effort
were partly carried out on the Genetic Cluster Computer (http://www.geneticcluster.org)
hosted by SURFsara and financially supported by the Netherlands Organization for
Scientific Research (NWO; 480-05-003 Pl. Posthuma) along with a supplement from the
Dutch Brain Foundation and the VU University Amsterdam. M.G.N. is supported by
Royal Netherlands Academy of Science Professor Award to D.I.B. (PAH/6635). Part of
the computation of this project was funded by NWO Exact Sciences for the application:
“Population scale Genetic Analysis” awarded to M.G.N. The genome-wide association
analysis on the UK Biobank dataset has been conducted using the UK Biobank resource
under application numbers 9905, 16406 and 25331.

The Substance Use Disorders Working Group of the Psychiatric Genomics
Consortium (PGC-SUD) is supported by funds from NIDA and NIMH to MH109532
and, previously, with analyst support from NIAAA to U01AA088401 (COGA). I.M.S.
contributions were partially supported by the Peter Boris Chair in Addictions Research.

Contributors to the International Cannabis Consortium. The PGC-SUD group provided
supervised genome-wide association analyses. J.A.P. performed the quality control
and meta-analysis of genome-wide association studies, under supervision of K.J.H.V.,
N.G.M. contributed to data acquisition of the samples in the International Cannabis
Consortium. S.L.E., H.d.W., L.K.D. and J.M.K. contributed to data acquisition and
analysis for the 23andMe dataset. All authors provided critical revision of the
manuscript for important intellectual content.

Contributors to the 23andMe Research Team: M. Agee, B. Alipanahi, A. Auton, R. K. Bell, K. Bryc,
S. L. Elson, P. Fontanillas, N. A. Furlotte, D. A. Hinds, K. E. Huber, A. Kleinman, N. K. Litterman,
J. C. McCreight, M. H. McIntyre, J. L. Mountain, E. S. Noblin, C. A. M. Northover, S. J. Pitts,
J. Fah Sathirapongsasuti, O. V. Sazonova, J. F. Shelton, S. Shringarpure, C. Tian, J. Y. Tung,

Contributors to the International Cannabis Consortium: S. Stringer, C. C. Minica, K. J. H. Verweij,
H. Mbarek, M. Bernard, J. Derringer, K. R. van Eijk, J. D. Isen, A. Loukola, D. F. Maciejewski,
E. Mihailov, P. J. van der Most, C. Sánchez-Mora, L. Roos, R. Sherva, R. Walters, J.J. Ware,
A. Abdellaoui, T. B. Bigdeli, S. J. T. Branje, S. A. Brown, M. Bruinenberg, M. Casas, T. Esko,
M. Hickman, C. J. Hopfer, J. J. Hottenga, A. C. Huizink, D. E. Irons, R. S. Kahn, T. Korhonen,
H. R. Kranzler, K. Krauter, P. A. C. van Lier, G. H. Lubke, P. A. F. Madden, R. Mägi, M. K. McGue,
S. E. Medland, W. H. J. Meeus, M. B. Miller, G. W. Montgomery, M. G. Nivard, I. M. Nolte,
A. J. Oldehinkel, Z. Pausova, B. Qaiser, L. Quaye, J. A. Ramos-Quiroga, V. Richarte, R. J. Rose,
J. Shin, M. C. Stallings, A. I. Stiby, T. L. Wall, M. J. Wright, H. M. Koot, T. Paus, J. K. Hewitt,
M. Ribasés, J. Kaprio, M. P. M. Boks, H. Snieder, T. Spector, M. R. Munafò, A. Metspalu, J. Gelernter,

Competing interests
P.F., S.L.E. and members of the 23andMe Research Team are employees of 23andMe Inc.
J.A.R.-Q. was on the speakers’ bureau and/or acted as consultant for Eli Lilly, Janssen-
Cilag, Novartis, Shire, Lundbeck, Almirall, BRAINGAZE, Sincrolab and Rubió in the
last 5 years. He also received travel awards (air tickets and hotel) for taking part in
psychiatric meetings from Janssen-Cilag, Rubió, Shire and Eli Lilly. The Department of
Psychiatry chaired by him received unrestricted educational and research support from
the following pharmaceutical companies in the last 5 years: Eli Lilly, Lundbeck, Janssen-
Cilag, Actelion, Shire, Ferrer and Rubió.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/
s41593-018-0206-1.
Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to J.M.V.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in
published maps and institutional affiliations.
1Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands. 2Amsterdam UMC, University of Amsterdam, Department of Psychiatry, Amsterdam, The Netherlands. 3Genetic Epidemiology, Statistical Genetics, and Translational Neurogenomics Laboratories, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. 4Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. 5Department of Psychiatry, University of California San Diego, La Jolla, CA, USA. 6MRC Integrative Epidemiology Unit (IEU), University of Bristol, Bristol, UK. 7Department of Biological Psychology/Netherlands Twin Register, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. 8MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge, UK. 23andMe, Inc., Mountain View, CA, USA. 9A list of 23andMe Research Team members appears at the end of the paper. 10Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL, USA. 11Vanderbilt Genetics Institute, Division of Genetic Medicine, Department of Medicine, Vanderbilt University, Nashville, TN, USA. 12Pamela Boris Centre for Addictions Research and Michael G. DeGroote Centre for Medicinal Cannabis Research, McMaster University/St. Joseph’s Healthcare Hamilton, Hamilton, Ontario, Canada. 13Psychiatric Genomics Consortium. 14A list of International Cannabis Consortium members appears at the end of the paper. 15Department of Psychology, University of Illinois Urbana-Champaign, Champaign, IL, USA. 16Department of Youth and Family, Utrecht University, Utrecht, the Netherlands. 17Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion Regulation, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. 18Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. 19Department of Developmental Psychology and EMGO Institute for Health and Care Research, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. 20Estonian Genome Center, University of Tartu, Tartu, Estonia. 21Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia. 22Hospital for Sick Children, Toronto, Ontario, Canada. 23Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d’Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. 24Department of Psychiatry, Hospital Universitari Vall d’Hebron, Barcelona, Spain. 25Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Barcelona, Spain. 26Department of Psychiatry and Legal Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain. 27Rotman Research Institute, Baycrest, Toronto, Ontario, Canada. 28Departments of Psychology and Psychiatry, University of Toronto, Toronto, Ontario, Canada. 29Institute for Molecular Medicine Finland FIMM, HiLIFE Unit, University of Helsinki, Helsinki, Finland. 30Brain Center Rudolf Magnus, Department of Psychiatry, University Medical Center Utrecht, Utrecht, The Netherlands. 31Department of Twin Research and Genetic Epidemiology, King’s College London, London, UK. 32Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA. 33Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA. 34UK Centre for Tobacco and Alcohol Studies and School of Experimental Psychology, University of Bristol, Bristol, UK. 35Department of Psychiatry, Virginia Institute for Psychiatric and Behavior Genetics, Virginia Commonwealth University, Richmond, VA, USA. 36These authors contributed equally: Joëlle A. Pasman, Karin J. H. Verweij. 37These authors jointly supervised: Nathan A. Gillespie, Eske M. Derks, Jacqueline M. Vink. *e-mail: j.vink@bsi.ru.nl
SNP-based heritability analysis. The proportion of variance in liability to cannabis use that could be explained by the aggregated effect of the SNPs ($h^2_{SNP}$) was estimated using LD-score regression analysis. The method is based on the premise that an estimated SNP effect size includes effects of all SNPs in LD with that SNP. A SNP that tags many other SNPs will have a higher probability of tagging a causal genetic variant compared to a SNP that tags few other SNPs. The LD score estimates the amount of genetic variation tagged by a SNP within a specific population. Accordingly, assuming a trait with a polygenic architecture, SNP's with a higher LD score have on average stronger effect sizes than SNPs with lower LD scores. When regressing the effect size from the association analysis against the LD score for each SNP, the slope of the regression line provides an estimate of the proportion of variance accounted for by all analyzed SNPs. For this analysis, we included 1,179,898 SNPs that were present in all cohorts and the HapMap 3 reference panel. Standard LD scores were used as provided by Bulik-Sullivan et al. based on the Hapmap 3 reference panel, restricted to European populations.

Genetic correlations with other substances and mental health phenotypes. We used cross-trait LD-score regression to estimate the genetic correlation between lifetime cannabis use and 25 other traits using GWAS summary statistics. The genetic covariance is estimated using the slope from the regression of the product of z-scores from two GWAS on the LD score. The estimate represents the genetic covariance between the two traits based on all polygenic effects captured by SNPs. Summary statistics from well-powered GWAS were available for 25 relevant substance use and mental health traits, including nicotine, alcohol and caffeine use, schizophrenia, depression, bipolar disorder and loneliness (Supplementary Table 6). To correct for multiple testing we adopted a Bonferroni corrected $P$-value threshold of significance of 0.002 (0.05/25). LD scores were based on the Hapmap 3 reference panel, restricted to European populations.

Causal association between cannabis use and schizophrenia: two-sample Mendelian randomization. We performed two-sample Mendelian randomization analyses (MR) to examine whether there was evidence for a causal relationship from cannabis use to schizophrenia. Analyses were performed with the R package of database and analytical platform MR-Base and with the gsmr R package, which implements the GSRM (generalized summary-data-based Mendelian randomization) method.

MR utilizes genetic variants strongly associated with an exposure variable as an ‘instrument’ to test for causal effects of the exposure on an outcome variable. This approach minimizes the risk of spurious findings due to confounding or reverse causation present in observational studies, provided that the following assumptions are met: (i) the genetic instrument is predictive of the exposure
variable, (ii) the genetic instrument is independent of confounding effects, and (iii) the genetic instrument is not directly associated with the outcome variable, other than by its potential causal effect through the exposure (i.e., there is no directional pleiotropy). Two-sample MR refers to the application of MR methods to well-powered summary association results estimated in non-overlapping sets of individuals in order to reduce instrument bias toward the exposure–outcome estimate.

Bidirectional causal effects were tested between lifetime cannabis use and schizophrenia. We used genetic variants from our cannabis GWAS, as well as those from the largest schizophrenia GWAS, to serve as instruments (gene–exposure association). For lifetime cannabis use we used two genetic instruments: (i) an instrument including all independent genetic variants that were genome-wide significantly associated with lifetime cannabis use (P < 5 × 10^-8; 5 SNPs) and (ii) an instrument including independent variants with a more lenient significance threshold (P < 1 × 10^-6; 69 SNPs). For schizophrenia we used one genetic instrument, including independent genetic variants that were genome-wide significantly associated with schizophrenia (instrument P < 5 × 10^-10; 109 SNPs).

Information on the included SNPs in the genetic instruments is provided in Supplementary Table 7.

Genetic variants were pruned (R^2 < 0.001) and the remaining genetic variants (or proxies (R^2 ≥ 0.8), when an instrumental SNP was not available in the other GWAS) were then identified in GWAS summary-level data of the outcome variable (gene–outcome association). Note that not all independent SNPs identified in the exposure dataset have been included in the analyses because not all exposure SNPs or their proxies were also available in the outcome dataset and because some SNPs were palindromic (see Supplementary Table 7).

Evidence for both a gene–exposure and a gene–outcome association suggests a causal effect, provided that the MR assumptions are met. To combine estimates from individual genetic variants, we applied inverse-variance-weighted (IVW) linear regression. In addition, four sensitivity analyses more robust to horizontal extrapolation, SIMEX, were palindromic (see Supplementary Table 7).

In addition to the genetic instrument to obtain an estimate of the total R^2 explained by the instrument.

**References**


Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

☐ n/a Confirmed

☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement

☐ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

☐ The statistical test(s) used AND whether they are one- or two-sided.

☐ *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*

☐ A description of all covariates tested

☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

☐ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

☐ For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted

☐ Give P values as exact values whenever suitable.

☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

☐ Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

☐ Clearly defined error bars

☐ *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

For this study, we meta-analyzed genetic and phenotype data from 3 different cohorts. Data collection procedures differed per cohort and are summarized in the Methods section (for details on data collection in the ICC cohorts see Stringer et al., 2016). In general, DNA was extracted from blood or saliva samples, genotyped, and imputed using European reference data. Phenotype information was collected using paper-and-pencil or online surveys.

Data analysis

PLINK 2.0 - genomewide association tool; MAGMA v1.06 - gene-based tests; S-PrediXcan - gene expression analysis; LD score regression - genetic correlations and heritability; R qqqman - visualisation of GWAS results; LocusZoom - creation of regional plots; GTEx Analysis Release V7 - eQTL; METASOFT - eQTL forest plot; MR-Base - mendelian randomization; R gsmr - mendelian randomization; METAL - meta-analysis

*all software mentioned here is publicly available

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Summary statistics based on the UK-Biobank and ICC samples and full results from the top 10,000 SNPs based on all three subsamples (i.e. including the 23andMe sample) will be available via LDhub (http://ldsc.broadinstitute.org/gwashare/). Codes and scripts are available upon reasonable request. Full summary statistics can only be provided after permission by 23andMe.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | A previous GWAS of the International Cannabis Consortium with a sample size of ~33,000 had limited power to detect genomewide significant hits. We expanded this sample with all available data that we had access to, resulting in a 5-fold increase in sample size, providing sufficient power to detect genomewide signals. |
| Data exclusions | Extensive quality control procedures were used to select valid SNPs and individuals using pre-established criteria. These have been described in Supplementary Table S12 and include exclusion of related individuals and individuals with missing data, variants with a low HWE, a low minor allele frequency, a low imputation quality score, or high missingness rates, and variants whose alleles and allele frequency differ from those in reference panels. For secondary analysis, sometimes a subset of the genome-wide data was used (i.e., SNPs that could be mapped to a gene in gene-based tests, SNPs that were present in reference files that were used by LocusZoom, LDscore regression, or S-PrediXcan). |
| Replication | We did not divide our sample into a discovery and replication sample, so that we had one large sample with sufficient power to detect a genome-wide signal. We have been as transparent as possible about our methodology and summary statistics will be made available, so that replication can be attempted by other research groups. |
| Randomization | N/A; we did not use an experimental design. |
| Blinding | N/A; we did not use an experimental design. Analysts were not blind to case-control status. |

Reporting for specific materials, systems and methods

<table>
<thead>
<tr>
<th>Materials &amp; experimental systems</th>
<th>n/a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Involved in the study</td>
<td></td>
</tr>
<tr>
<td>Unique biological materials</td>
<td>X</td>
</tr>
<tr>
<td>Antibodies</td>
<td>X</td>
</tr>
<tr>
<td>Eukaryotic cell lines</td>
<td>X</td>
</tr>
<tr>
<td>Palaeontology</td>
<td>X</td>
</tr>
<tr>
<td>Animals and other organisms</td>
<td>X</td>
</tr>
<tr>
<td>Human research participants</td>
<td>X</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Methods</th>
<th>n/a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Involved in the study</td>
<td></td>
</tr>
<tr>
<td>ChIP-seq</td>
<td>X</td>
</tr>
<tr>
<td>Flow cytometry</td>
<td>X</td>
</tr>
<tr>
<td>MRI-based neuroimaging</td>
<td>X</td>
</tr>
</tbody>
</table>

Human research participants

Policy information about studies involving human research participants

Population characteristics | The sample included N=184,765 individuals, with 55.5% females and a mean age of 35.7 (range 16-87). Only individuals from
Population characteristics

European ancestry were included. Using principal components for population stratification, we controlled for systematic ancestry differences within this European cohort.

Recruitment

Recruitment strategies differed for the 3 cohorts (ICC, 23andMe, UKB) and within the different cohorts making up the ICC cohort (Stringer et al., 2016). 23andMe is a commercial platform where individuals can have their DNA genotyped at their own costs, and can provide permission to make their material available for research. UKB and ICC participants are volunteer samples. Information about recruitment can be found in our supplementary material and in the supplementary material of Stringer et al., 2016.